GABI – The German Plant Genome Research Program

Progress Report 1999 – 2004

SPONSORED BY THE

Federal Ministry
of Education
and Research

Imprint
GABI – The German Plant Genome Research Program · Progress Report 1999 – 2004
Published by: GABI managing office · Dr. Jens Freitag · Dr. Saskia Dombrowski
c/o Max Planck Institute of Molecular Plant Physiology
Am Mühlenberg 1 · 14476 Potsdam-Golm · Germany
Telephone +49.331.567 8301 · Telefax +49.331.56789 8301
eMail freitag@mpimp-golm.mpg.de · www.gabi.de

ISBN: 3-00-011587-0

Federal Ministry
of Education
and Research

Gestaltung und Layout: Dirk Biermann
Photos: Josef Bergstein, Dirk Biermann
Druck: sd:k Satz Druck GmbH, Teltow

GABI – The German Plant Genome Research Program

Progress Report 1999 – 2004

Foreword

Plants are the basis of all animal and human life and their well-being. Plants produce, directly or indirectly, all food or feed and provide medically-active substances as well as renewable raw materials. More than 99% of the total terrestrial biomass consists of plants of a large variety – from unicellular algae up to giant sequoia trees. Plants are the key to maintaining and regenerating complex ecosystems and the biosphere. We cannot live or survive without them.

Plant genome research aims at a better understanding and use of the "biological system plant". In the past years, plant genome research has become one of the most innovative branches of science. It has the aim to discover and understand and in addition to comprehensively use the molecular correlations in a biological system, the plant. Furthermore, plant genome research with its interdisciplinary and multi-parallel approaches enables the registration and molecular characterization of the genetic diversity to be found in nature. This field of research has become one of the most integrative scientific disciplines and underlines the change of paradigms in the modern life sciences. It combines classical biological thinking with cutting edge analytical methodology and the rapidly growing possibilities of data processing to a level resembling the system biological approach. The results of the plant genome research are an important key to the solution of upcoming global tasks and will change future production to higher sustainability. Agriculture, the 'engine of change' for the developing countries, will also play an important role in the transformation of our hydrocarbon dependent economies within the 21st century into bio-based industries. With rapid world growth and continuing changes in consumer demands, there is a need to find additional and therefore preferably renewable, resources for industrial production and for our energy needs. We believe that plant genomics research is a key to provide continued economic growth, healthy standards of living, and through the development of plant-based renewable resources security for our society. All this is related to the transition of our society towards a technology and knowledge based one.

GABI, the German plant genome program brings together the German expertise in this research area. GABI is proving to be a functioning "public-private" partnership. The immense task to identify the biological functions of important plant genes and to transfer them into real applications can only be achieved in a concerted effort. The research projects realized in GABI include the research of basic processes on the model organism *Arabidopsis thaliana* and in the crop plants barley, wheat, rye, maize, rapeseed, sugar beet and potato. In addition, three resource centers and two bioinformatics centers are anchored within GABI.

After the completion of the first GABI program phase, we already know more about disease and stress resistance in plants, the physiological processes in seeds and storage organ development. We know numerous specific "gene switches" in crops and can use those specifically. We know details about the creation and the function of complex genetic networks and know new signal and transport systems in plants. Moreover, GABI clarifies the molecular structures of important chromosomal regions in crops. These are the foundations for **marker-assisted plant breeding**. "Molecular Breeding" is already supporting and accelerating plant breeding significantly. It is the key to a broad use of the natural genetic diversity to improve our crop varieties.

The second project phase of GABI already started at the end of 2004. During this phase GABI will concentrate on a so called **'bridging concept'** between model and crop plants in order to guarantee a prompter realization of relevant research results. GABI was successful during the first program phase in regaining lost German competence in the modern biosciences. One of the most important tasks of following funding periods will be to maintain and extend those.

GABI provides, and will continue to provide, an example for the building of the **European Research Area**. During 2004 the bilateral cooperation with the French partner initiative Génoplante, which was started in the first program phase, will be strengthened furthermore. In addition, multinational plant genome projects between the Spanish Ministry of Education and Science, Génoplante and GABI started in 2004. Up to 25% of the financial support for GABI 2 will be available for international co-operations to develop synergies and to avoid unnecessary duplications in the different countries as there is a need for international co-operation in genome research alongside the creation of national networks. Such joint developments and the shared use of resources or technologies guarantee a more efficient handling of the available resources and avoid fruitless doublings. This internationalization of the European plant genome research was initiated by GABI together with other national plant genomics research programs, national funding agencies as well as the European Commission. For this reason GABI is one of the supporting pillars of the creation of a **"European Research Area Net Plant Genomics"** (ERA Net PG). By means of this network, pan-European activities shall be brought together and the European research area shall be created and designed in a sustainable and trendsetting way. The immense task of gene function clarification and utility can only be realized through co-operation and networking.

To get an impression of ongoing GABI projects, structures and partners as well as on our outreach activities and connections please refer to www.gabi.de.

The following GABI Progress Report gives an overview on what has been achieved during the first funding period of GABI. Beyond these results, the German plant genome research community has an optimistic outlook and is prepared to make a significant contribution to the further development of the German and European research landscape. Functional genome research offers immense opportunities to shape our future. Let's rise to the challenge.

Dr. Jens Freitag
Head of the GABI Managing Office

Content

5 **Foreword**
6 **Content**
10 **Sitemap**
12 **Introduction**
16 **Patent and Licensing Agency (PLA for GABI)**
17 **Economic Network of Plant Genome Research** (WPG)

Resourcecenters and Bioinformatics

22 **The GABI resourcecenters**
Overview: GABI KAT, GABI-LAPP, GABI Plant, GABI Bioinformatics

GABI-KAT

24 **Sequencing of T-DNA flankings for "*in silico*" detection of tagged alleles (GABI-KAT – Kölner Arabidopsis T-DNA lines)**

24 **A Génoplante-GABI bi-lateral project for sharing sequence data and Computer resources on T-DNA transformants of *Arabidopsis thaliana***
Bernd Weisshaar, Mario Rosso, Yong Li, Sabine Steiner-Lange, Koen Dekker, Bernd Reiss, Heinz Saedler

GABI-LAPP

26 **Generation of Arabidopsis and Barley protein expression libraries and chips**
Birgit Kersten, Tanja Feilner, Armin Kramer, Alexandra Poßling, Silke Wehrmeyer, Hans Lehrach, Dolores J. Cahill

28 **2-D gel electrophoresis and mass spectrometer for the analysis of plant proteomes**
Johan Gobom, Patrick Giavalisco, Niklas Gustavsson, Eckhard Nordhoff, Joachim Klose, Hans Lehrach

30 **Identification of protein-protein interactions using an automated yeast two hybrid system** (Subproject 3)
Lukas Bürkle, Marion Amende, Rita Fischer, Eckehard Kuhn, Uli Schneider, Hans Lehrach, Alexander Heyl

32 **RNA Expression analysis**
Wilfried Nietfelt, Stefanie Albrecht, Holger Eickhoff, Antje Krüger, LajosNyarsik, Hans Lehrach, CATMA consortium

GABI Plant

34 **Management of germplasm and mapping poulations – Construction of a high density transcript map of barley**
Manoj Prasad, Nils Stein and Andreas Graner

36 **A core collection of barley cDNAs and expressed sequence tags (ESTs)**
Hangning Zhang, Nils Stein and Andreas Graner

38 **Development of a DNA array based method for efficient identification of plant genetic resources**
Konrad Bachmann, Frank Blattner, Jörg Geistlinger, Dirk Fischer

GABI Trafo

40 **Overview: Qualitative and Quantitative Improvement for Genetic Transformation of Barley**

42 **Qualitative and Quantitative Improvement for Genetic Transformation of Barley** (Subproject A)
Liliya Serazetdinova, Horst Lörz, Martina Brumm-Scholz

44 **Qualitative and Quantitative Improvement for Genetic Transformation of Barley** (Subproject B)
Jochen Kumlehn, Rajiv Vishnoi, Götz Hensel

46 **Qualitative and Quantitative Improvement for Genetic Transformation of Barley** (Subproject C)
Silke Mühlenbeck, Jens Weyen

Barley Transposon

48 **Transposon-Mutated barley – A Powerful Tool for Genome Analysis**
Thomas Koprek

GABI Info

50 **Bioinformatic Resourcecenter for Plant Genome Research**
Hans-Werner Mewes, Klaus Mayer

GABI-PD

54 **GABI Primary Database**
Svenja Meyer, Axel Nagel, Steffen Schulze-Kremer

Dicotyledones

Arabidopsis

58 **Overview**

60 **The GABI Arabidopsis Community Overview: Genetic Diversity, Protein Community, Gauntlets**

Genetic Diversity

62 **Creation of Novel Genetic Variants of Arabidopsis**
Rhonda Meyer, Carsten Müssig, Thomas Altmann

64 **Large-scale SNP identification and analysis in the Arabidopsis genome**
K. Schmid, Thomas Mitchell-Olds

66 **Establishment of high-efficiency mapping tools**
Dieter Berger†, Otto Törjek, Thomas Altmann

68 **Establishment of high-efficiency SNP-based mapping tools and development of methods for genome-wide mutation detection**
Ralf Stracke, Thomas Rosleff Sörensen, Bernd Weisshaar

70 **Ligation detection reaction – TaqMan procedure for SNP detection on genomic DNA**
Aleksey Soldatov, Tatiana Borodina, Thomas Altmann, Bernd Weisshaar, Thomas Mitchell-Olds

Protein Community

72 **Genomics, Expression and Membranes: Identification and functional analysis of transport proteins**
Reinhard Kunze, Marcelo Desimone, Anja Schneider, Rainer Schwacke, Anja Schmidt, Giorgio Gonella, Eric van der Graaff, Ulf-Ingo Flügge, Wolf B. Frommer

74 **Transcription factor function search: How do individual factors regulate agronomical important processes in plants?** (Subproject A)
Isabell Witt, Maria Ines Zanor, Bernd Müller-Röber

76 **Transcription factor function search: How do individual factors regulate agronomically important processes in plants?** (Subproject B)
Gieta Dewal, Martin Sagasser, Illona Zimmermann, Bernd Weisshaar

Gauntlets

78 **Carbon and Nutrient Signalling: Test Systems, and Metabolite and Transcript Profiles**
Mark Stitt

82 **Analysing cell wall biosynthesis to study its role in biotic and abiotic stress reactions**
Nicolai Obel, Björn Usadel, Tze Siang Choo, Markus Pauly

84 **Spatio/temporal analysis of gene expression profiles within single cells and specific tissue types of developing Arabidopsis plants**
Elke Lieckfeldt, Ulrike Simon Rosin, Daniela Zöller, Berit Ebert, Julia Kehr, Joachim Fisahn

86 **Searching for signals: stress-induced changes in Arabidopsis metabolite, peptide and protein patterns**
Stephan Clemens, Jürgen Schmidt, Ludger Wessjohann, Dierk Scheel

Rapeseed

88 **Overview: Genome Analysis in Rapeseed – Genetic, physical and functional characterization using Arabidopsis resources**

GABI-GARS

90 **Biochemical investigations of seed development in rapeseed functional gene analysis (*Brassica napus*)**
Roland Baetzel, Wilfried Lühs, Wolfgang Friedt

92 **A strategy to identify regulatory genes important for the regulation of metabolic pathways in seeds during rapeseed development** (Subproject B)
Ute Heim, Michael Geiger, Karin Herbers

94 **Development of intervarietal substitution lines, QTL mapping, and synteny analysis**
Wolfgang Ecke, Rubens Marschalek, Heiko Becker

96 **Seed colour in oilseed rape (*Brassica napus*): Classical genetics, QTL analysis and candidate genes for map-based cloning and marker development**
Ana Gloria Badani, Rod Snowdon, Roland Baetzel, Tina Lotz, Renate Horn, Wilfried Lühs, Wolfgang Friedt

Content

98 **Oilseed rape (*Brassica napus* L.) – Establishment of Segregating Progenies and Field Evaluation**
Christoph Lüdecke, Heinrich Busch, Dieter Stelling, Konstanze Richter

100 **Development and phenotypic description of basic material for genome analysis of oilseed rape**
Lutz Brahm, Jutta Böhm

Sugar Beet

102 **Overview: Genome Analysis of Sugar Beet**

104 **The Sugar Beet Community**
Overview: GABI Beet, GABI BOLT, Sweet GABI

GABI Beet

106 **Genome analysis of sugar beet** (Subproject A)
Christian Jung, Uwe Hohmann, Maike Niemann, Thomas Schmidt, Tim Thurau, Hans-Werner Mewes, Stephen Rudd, Georg Büttner, Bernward Märländer, Lothar Frese

108 **Genome analysis of sugar beet** (Subproject B)
Silke Möhring, Ralf-Schäfer-Pregl, Jörg Wunder, Thomas Rosleff-Sörensen Charlotte Bulich, ADIS (DNA core facility headed by Bernd Weisshaar), Katharina Schneider

110 **Genome analysis of sugar beet** (Subproject C)
Dagmar Kulosa, Andreas Polley, Susanne Reich, Carina Kehling and Martin Ganal

112 **Genome analysis of sugar beet** (Subproject D)
Georg Koch

GABI Bolt

114 **Cloning of the bolting gene B from sugar beet**
Christian Jung, Uwe Hohmann, Britta Schulz, Georg Koch

Sweet-GABI

116 **Crosslinking Structural and Functional Genomics with Regard to Metabolism and Development in Sugar Beet**
Britta Schulz, Katharina Schneider

Potato

118 **Overview: GABI CONQUEST**

GABI CONQUEST

120 **Identification and characterization of genes controlling quantitative agronomic characters in potato by a candidate gene approach**
CONQUEST (Genes **CON**trolling **QU**antitativ**E** traits of **S**olanum **T**uberosum)
Christiane Gebhardt, Hans-Reinhard Hofferbert, Eckart Tacke, Jens Lübeck

Poplar

122 **Overview: GABI-POP**

GABI POP

124 **Isolation of tree-specific genes and promoters by a transposon tagging approach**
Matthias Fladung, Frank Deutsch, Sandeep Kumar

Monocotyledones

Cereals

126 **Overview: Cereals**

128 **Overview: Development of new DNA marker systems and utilization of genetic resources for Barley**

Barley Diversity

130 **AB-QTL analysis in barley: Detection of favorable genes for quality and yield traits from barley wild species and their introgression into the elite germplasm of cultivated barley by means of DNA markers**
Klaus Pillen, Maria v. Korff Schmising, Huajun Wang, Jens Léon

132 **Development of single nucleotide polymorphism (SNP) markers for Barley (*Hordeum vulgare* L.)** (Subproject A)
Frank Ordon, Gisela Neuhaus, Jörg Schondelmaier, Jens Weyen

134 **Development of single nucleotide polymorphism (SNP) markers for Barley (*Hordeum vulgare* L.)** (Subproject B)
Raja Kota, Nils Stein and Andreas Graner

136 **Development of new DNA marker systems and the use of genetic resources for Barley**
Ludmilla Malysheva, Martin Ganal and Marion Röder

138 **Overview: Functional genomics of developing and germinating barley seeds**

GABI Seed

140 **Functional Genomics of Barley Seed Development** (Subproject A)
Winfriede Weschke, Nese Sreenivasulu, Volodymyr Radchuk, Sabine Gubatz, Ulrich Wobus

142 **Functional genomics of germinating barley seeds** (Subproject B)
Elena Potokina, Nils Stein, Andreas Graner

144 **Functional Genomics of Malting Quality** (Subproject C)
Markus Herz, Sabine Mikolajewski, Günther Schweizer

GABI Map

146 **Comparative sequencing of two regions of the Barley and Rice genome**
Roberto Cossu, Dragan Perovic, Nils Stein, Martin Ganal, Andreas Graner and Marion Röder

GABI Nonhost

148 **A Consortium-Based Functional Genomics Initiative on Plant Nonhost Disease Resistance**
Jane Parker, Paul Schulze-Lefert, Patrick Schweizer, Dierk Scheel, Thorsten Nürnberger, Markus Frank

GABI Agrotec

150 **Functional Analysis of the Barley Genome: Functional Characterization of Agronomically Relevant Genes and their Use to Improve Disease Resistance to the Genus Fusarium**
Carin Janssen, Sibylle von Rüden, Rebekko Fensch, Karl-Heinz Kogel, Sophia Biemelt, Jochen Kumlehn, Uwe Sonnewald, Angelika Felk, Simon Henning, Joe-Gmem Ekanem, Wilhelm Schäfer

GABI Rye

152 **Overview: Assessment and Functional Analysis of Agronomically Important Genome Regions in Hybrid Rye**

154 **Structural and Functional Comparison of Disease Resistance Genes in Rye and Other Plants**
Thomas Bringezu, Bernd Hackauf, Bodo Linz, Peter Wehling

156 **Assessment and Functional Analysis of Agronomically Important Genome Regions in Hybrid Rye and Assessment of Agronomically Important Genome Regions from Exotic Germplasm in Hybrid Rye**
Subproject 1: Bernd Hackauf, Stephen Rudd, Peter Wehling
Subproject 2a: Zoran Susic, Thomas Miedaner, Hartwig H. Geiger
Subprojects 2b, 2c: Heinrich Wortmann, Jörg Schondelmaier, Peer Wilde, Viktor Korzun

Wheat AB-QTL

160 **AB-QTL analysis in wheat: Detection of favorable genes for quality and yield traits from wheat wild species and their introgression into the elite germplasm of cultivated wheat by means of DNA markers**
Klaus Pillen, Antje Kunert, Anne Schubert, Jens Léon

162 **GABI Maize Projects**
Overview: Maize Cool, Maize Promoter

GABI Cool

164 **Genomics of Chilling Tolerance in Maize**
Thomas Presterl, Karin Ernst, Evelyn M. Möller, Milena Ouzunova, Hartwig H. Geiger, Peter Westhoff

Promoter

166 **Isolation and analysis of monocotyledonous promoters from *Zea mays* L.**
René Lorbiecke, Katja Müller, Carina Paul, Reinhold Brettschneider, Günter Donn, Udo Wienand

168 **Authors Index**
170 **GABI Publications**
180 **GABI Contacts**

Sitemap

GABI Progress Report

General Information

Introduction
Page 12

PLA for GABI
Page 16

WPG
Page 17

Resource Centers

GABI-KAT
Page 24

GABI-LAPP
Page 26

GABI Plant
Page 34

Gabi Trafo
Page 40

Barley Transposon
Page 48

GABI Info
Page 50

GABI-PD
Page 54

Dicotyledones

Arabidopsis

Genetic Diversity
Page 62

Protein Community
Page 72

Gauntlets
Page 78

Rapeseed

GABI-GARS
Page 90

Sugar Beet

GABI Beet
Page 106

GABI Bolt
Page 114

Sweet GABI
Page 116

Potato

GABI CONQUEST
Page 120

Monocotyledones

Poplar

GABI POP
Page 124

Barley

Barley Diversity
Page 130

GABI Seed
Page 140

GABI Map
Page 146

GABI Nonhost
Page 148

GABI Agrotec
Page 150

Rye

GABI Rye
Page 154

Wheat

Wheat AB-QTL
Page 160

Maize

GABI Cool
Page 164

Promoter
Page 166

Appendix

Authors Index
Page 168

Publications
Page 170

Contacts
Page 180

Introduction: Plant Genomics Research – A Must for our Societies

1 Jens Freitag & 2 Thomas Altmann
1 GABI Managing Office c/o Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm
2 Institute of Biochemistry and Biology– Genetik, Universität Potsdam; Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
(Head of the GABI Scientific Coordinating Committee during the first funding period 1999 – 2004)

Definition

Plant Genomics refers to the analysis of plant genomes, which comprises physical mapping of the genome, genome sequencing and annotation (prediction and identification of genes), and the elucidation of the (biochemical) functions of the identified genes and gene products and their roles in determining the characteristics of the organisms.

Characteristics and relations to human needs

Plant genomics aims to provide a detailed and comprehensive knowledge of the fundamental molecular processes of plant life. To identify all genes of a plant, to determine its biochemical functions and the corresponding gene products as well as uncovering the functions of those genes and their effect on the plant metabolism, growth and development or their interaction with the environment lies at the heart of plant genome analysis as it would for any other organism. In that respect the major principles of plant genomics do not differ from those in human, animal or microbial genomics. However, some features do set the characteristics of plant genome analysis apart from the other fields. A main difference is that a broad range of different and important plant species has to be studied, that plant genomes have an enormous size and complexity, that the establishment of (large) experimental populations including immortalized segregating populations is possible, and that genetic transformation (at least for a number of model plant species) is easily achieved.

Justification of plant genomics

In representing more than 99% of the total terrestrial biomass plants are the dominating life form on our planet. Plants can be found in all regions of the earth with an enormous diversity of properties as for example their (biochemical) composition. They constitute the basis and essential component of any ecosystem, and represent the essential prerequisite for all human and animal life and welfare. As naturally growing renewable resources with an adjusted ecological balance, plants are not only absolutely indispensable for the production of bulk nutrients for human food and animal feed but they also provide us with essential vitamins and medicine, with fibers for clothing and many raw materials for industrial use and for building and construction. Due to the numerous and various uses of plants and plant products, plant genome analysis is interested in a number of different plant species. Just as model organisms, such as *Arabidopsis thaliana*, are a focus in understanding the principles of plant life in as short a time as possible, major crops, trees or medicinal plants are of equal interest. Through this broad approach genome analysis will not only deliver knowledge on the molecular mechanisms with which plants function or interact with their environment. The achievements of our research will also have a direct impact on the improvements of plants, increase or improve their utility for humans and animals and will show us better and more sustainable methods of using them. These considerations and expectations led to the initiation and implementation of significant plant genome analysis programs world-wide. Next to their importance to humanity, plants have a number of traits that make them ideal as model systems and which further our basic understanding of life. Despite the particular goal of a project, be it purely scientific or application oriented, the advances in plant genomics make plant research one of the most exciting research branches of all.

Front runners in plant genomics

Noteworthy is the fact that such a variety of species is examined in plant genomics, ranging from model species, as front runners, to crops with highly complex genomes. Due to its favorable properties the most widely used plant model species is *Arabidopsis thaliana*, a member of the *Brassicaceae* family related to oilseed rape and cabbage (8). Its small size, numerous seeds produced by a single individual via self-pollination and the ease of chemical mutagenesis made it an ideal object for genetic studies in plants. The observation of the very small size of ca. 135 Million base pairs (Mbp) distributed among the five chromosomes and the low complexity of its genome made it the first choice for plant genome analysis. The achievements of the international Arabidopsis Genome Initiative (AGI) comprise the first complete BAC-based physical map established for any organism and the first complete plant genome sequence at the end of the year 2000 (9). Arabidopsis is the second multi-cellular organism (after the nematode *Caenorhabditis elegans*) that has had its genome fully sequenced, and it therefore represents one of the central biological model systems for basic research. With the genome sequenced and a comprehensive inventory of the genes of this model species at hand, the Arabidopsis (genome) researchers have now taken the challenge to decipher the function of all genes. For every gene this means the determination of its site, condition, and level of activity (expression), the (sub-cellular) localization and the biochemical function of the gene product, its position in regulatory or metabolic networks and its interaction with other factors and its role in plant growth and development, metabolism and interaction with the environment (functional genome analysis) (1). The multinational Arabidopsis steering committee (MASC; http://www.arabidopsis.org/info/2010_projects/MASC_Info.jsp) co-ordinates the worldwide Arabidopsis functional genome analysis projects. MASC has been founded according to the excellent experience of the predecessor committee that co-ordinated the international Arabidopsis genome sequencing project. A wide range of important resources and novel techniques strongly support this work. They include

○ 1. J. Chory, J.R. Ecker, S. Briggs, M. Caboche, G.M. Coruzzi, D. Cook, J. Dangl, S. Grant, M.L. Guerinot, S. Henikoff, R. Martienssen, K. Okada, N.V. Raikhel, C.R. Somerville and D. Weigel **National Science Foundation-Sponsored Workshop report: "The 2010 Project" – Functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them.** Plant Physiology (2000) 123(2):423-425. ○ 2. O. Fiehn, S. Kloska and T. Altmann **Integrated studies on plant biology using multiparallel techniques.** Current Opinion in Biotechnology (2001) 12(1):82-86. ○ 3. O. Fiehn **Metabolomics – the link between genotypes and phenotypes.** Plant Molecular Biology (2002) 48(1-2):155-171. ○ 4. S.A. Goff, D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchison, C. Martin, F. Katagiri, B.M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W.I. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant and S. Briggs **A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. japonica).** Science (2002) 296:92-100. ○ 5. A. Hohe, T. Egener, J.M. Lucht, H. Holtorf, C. Reinhard, G. Schween and R. Reski **An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, *Physcomitrella patens*.** Current Genetics. (2004) 44(6):339-347.

very comprehensive sequence-indexed gene k.o. populations (see: http://signal.salk.edu/cgi-bin/tdnaexpress, http://www.mpiz-koeln.mpg.de/GABI-Kat, http://flagdb-genoplante-info.infobiogen.fr/projects/fst, http://www.arabidopsis.org/abrc), full-length cDNA collections (see: http://pfg-web.gsc.riken.go.jp/projects/raflcdna.html, http://www.arabidopsis.org/abrc/ssp_clone.jsp), gene-chips for genome-wide expression analysis (see: http://www.affymetrix.com/products/arrays/specific/arab.affx, http://www.catma.org) and other novel techniques such as proteome analysis and metabolic profiling, the latter of which had been originally developed in Arabidopsis and is now adopted for many other organisms (3). It is supported by a comprehensive database, which serves as the major information resource (TAIR, http://www.arabidopsis.org). Still Arabidopsis research accumulates the most extensive information on plant gene/genome function.

As mentioned earlier, Arabidopsis is „only" the front runner of plant genomics and several crop species genomes are now also undergoing very extensive analysis. The *Arabidopsis thaliana* genome sequencing has served as an excellent example of how well the tremendous workload can be distributed and how the high financial expenses can be shared among various national genome programs. This led to the formation of new international sequencing consortia, concentrating on the sequencing of the rice genome as a model for grasses (http://rgp.dna.affrc.go.jp/IRGSP/index.html) and of the poplar genome, as a model for trees (http://www.ornl.gov/sci/ipgc). Similar to the situation in the dicots, where Arabidopsis was selected for sequencing, rice was chosen for sequencing according to its relatively compact genome as the first monocot plant species and as a representative of the grasses. These include other important crops such as wheat, barley, maize, rye and sorghum. While the rice genome has a size of ca. 430 Mbp distributed among 12 chromosomes, the evolutionary related genomes of the other cereal crops are much larger: sorghum (750 Mbp), maize (2,400 Mbp), barley (5,000 Mbp) and wheat (16,000 Mbp). Contrary to Arabidopsis, however, rice does not only serve as a model for monocots / grasses but is a very important crop itself. It feeds half of the world's population with a production area of more than 150 million hectares and more than 580 million metric tones in yield (FAO statistics). For many countries in Africa, Asia/Pacific, Latin America and the Caribbean rice is the predominant staple. In developing countries, rice accounts for 715 kcal/capita/day; 27% of dietary energy supply, 20% of dietary protein and 3% of dietary fat (6).

Chinese researchers generated a draft rice-genome sequence of the "Indica-variety" in spring 2002 (10). Simultaneously the Researchers of Syngenta assembled a draft version of the almost complete sequence of the "Japonica-variety" (4). The completely finished high quality genomic sequence of the rice genome, established by the International Rice Genome Initiative should be available at the end of this year. Rice research will be equally boosted through this as much as the availability of the genome sequence did in Arabidopsis and similar resources for functional analysis (including sequence indexed k.o. populations are currently being established by the International Functional Genomics Consortium (http://www.iris.irri.org/IRFGC). The third plant genome sequence finished was that of the poplar tree. The Populus genus includes poplars, cottonwoods, and aspens – fast-growing trees widely used in forestry research. These fast growing trees produce seeds prolifically, and can be genetically modified in the laboratory. The tree sequenced was a female *Populus balsamifera* (black cottonwood, or balsam poplar). With a size of 550 Mbp the genome of this model tree is only 4 times as big as that of *Arabidopsis thaliana*. Organized in 19 chromosomes the complete sequencing and annotation was also finished at the end of 2004 (http://genome.jgi-psf.org/poplar0/poplar0.info.html).

Another plant being developed as a model system for plant biological studies is the moss *Physcomitrella patens* , which will support plant functional genomics. Bryophytes are the oldest living branch in land plant evolution, which are separated by approx. 450 million years of evolution from seed plants. By adding mosses to the current list of model plants, researchers hope to fully understand and employ land plant evolution and plant diversity. The dominance of the haploid gametophyte in the life cycle of this moss facilitates genetic analysis. To date, this moss is the only known terrestrial plant with an efficient system for homologous recombination in its nuclear DNA, making gene targeting strategies as easy as in yeast. In contrast to seed plants, the dominating generation in the moss life cycle is the haploid gametophyte. Therefore *Physcomitrella* is an ideal model for gene-function-analysis. Loss-of-function mutants created by a targeted gene knock-out approach (homologous recombination) can be analyzed immediately without complex back-crosses (5). The *Physcomitrella* genome is about 3.5 times as big (~480 Mbp) as the Arabidopsis genome and is distributed among 27 chromosomes. Today more than 100.000 EST entries exist in databases and represent approximately 25,000 genes. Preliminary analysis of genes and ESTs strongly indicate that these sequences are highly similar to those of the corresponding genes of seed plants.

With the availability of the genome sequences, the multi-parallel analysis techniques ("x-omics" such as transcriptomics, proteomics and metabolomics) can be applied as efficiently in these plant species as in the model plant Arabidopsis (2). Applied to functional genome analysis the massively increasing amount of available data to be stored, analyzed and interpreted will be further enhanced. As a consequence, bioinformatics, developed in parallel to the analytical techniques, will play the same essential role in plant genomics as in human, animal, or microbe genomics. As

can be witnessed in the other genomic research fields, plant biology is becoming more and more an information and computable science research and thus has taken the lead into plant systems biology.

Beside this shift of paradigms and the broad increase in basic knowledge, plant genomics are affecting the production methods of our staple crops today and will continue to do so in future.

Genomics of other plant species

Humans are dependent on plants for numerous reasons. A wide range of crop plants provide bulk nutrients for us such as carbohydrates (cereals), starch (potato, cassava or yams), and proteins (legumes). Plants also provide fat, oil, wax and fiber (cotton, soy or rapeseed). Furthermore, we cultivate plants as fruits or vegetables and as fodder plants for our livestock. But plants also contain natural stimulants (cacao, coffee) and provide wood, cork, caoutchouc, dyes and tanning agents (indigo, henna, and eucalyptus). Some also supply insecticides (neem tree, dalmatian pyrethrum or tobacco). And many plants have a positive influence on our health and many substances from medicinal plants are either directly used or serve as lead compounds for the development of new drugs. Ten of the world's 25 top-selling drugs in the late nineties derived from natural sources. It is for this reason that the number of plant species with very large genomes that are being subjected to genome analysis is far greater than that of animals (including models such as mouse, rat, and primates and our farm animals). This challenge is being met in several ways.

While general and basic gene functions will mainly be determined in the principal plant model systems such as Arabidopsis and rice, it can be said that the corresponding genes in crops and other important plant species will carry out the same or very similar functions. Therefore the knowledge gained through the functional genomics / systems biology approaches in the principal models will be highly useful to directly target analysis in the other plants species. This "bridging concept" requires a good sequence information basis of the crops which is being provided by comprehensive Expressed Sequence Tag (EST) libraries. These tags represent snapshots of the genes expressed in a given tissue and/or at a given developmental stage and provide a good representation of the expressed regions of the genome if a sufficient number of different libraries from a range of tissues, developmental stages and after various environmental challenges are analyzed. As only genic sequences are analyzed independent of the genome size, this approach is highly cost efficient, especially for species with very large genomes. Up to now, more than 20 million ESTs (from all kind of species) are known in public databases, and platforms for comparative plant genomics are being developed at different places in the world including the "reconstructomics" platform using the SPUTNIK program (http://mips.gsf.de/proj/sputnik/). SPUTNIK is a tool to analyze "genome-less genomes", i.e. to estimate the gene's content represented by the ESTs and to establish unigene sets for plant species for which no genome sequence information is available.

Another tool which proves to be highly important for the verification of the proposed gene function in crops is TILLING (Targeted Identification of Local Lesions In Genomes) (7). TILLING was first used in Arabidopsis and stands for a procedure for the targeted identification of mutant alleles. It is now applied to many other species including animals and other plants. In this procedure, selected genes in a crop plant are tested for their functions through the identification of corresponding mutants and their characterization. The selection of these specific genes that are to be analyzed in detail was made on the basis of the information provided by the model system Arabidopsis.

However, as many gene functions specific to certain plant families will remain largely unknown, the base of model species will need to be broadened step by step to include representatives from other families. These will comprise tomato for the Solanaceae and *Lotus japonicus* or *Medicago truncatula* for legumes. The latter play an important role in providing protein and are able to fertilize themselves by fixation of atmospheric nitrogen through symbiosis with soil bacteria. Other genome sequencing programs that have recently begun or will start soon are on tomato, soy bean, Medicago, maize, potato, grape vine etc. In contrast to Arabidopsis or the rice genome sequences these programs currently do not aim to achieve the same level of completion. Instead, so-called draft sequence versions will be established (low coverage) or sequencing will be concentrated on gene-rich regions. These cost saving methods, such as the enrichment of weakly methylated DNA or the separation from repetitive sequences based on re-association kinetics, have been developed to give first insights into crop species genomes. Both methods have just proven their functionality on the occasion of the sequencing of gene-rich regions in the maize genome. The sequences thus obtained for a wider range of plant species will be anchored to the genomes of the principal models Arabidopsis and rice and large emphasis will be put on comparative genomics in plants in the future.

How plant genomics are applied to crop improvement

Plant genomics and also the closely related genome analysis in farm animals radically differ from the genome analysis in humans and model animal. Plant genomics is directly connected to practical application. Crop genome analysis provides breeders with a wealth of information and novel tools to enhance and speed up the breeding process. Molecular markers such as microsatellites (simple sequence length polymorphisms, SSLPs) or Single Nucleotide Polymorphisms (SNPs) provide the means to identify

○ 6. G. Kennedy, B. Burlingame and V.N. Nguyen **Nutritional contribution of rice: impact of biotechnology and biodiversity in rice-consuming countries. The International Rice Commission (Twentieth Session)** (2002); Bangkok, Thailand, 23-26 July. ○ 7. C.M. McCallum, L. Comai, E.A. Greene and S. Henikoff **Nature Biotechnology 1**(2000) 8(4):455-457. ○ 8. D.W. Meinke, J.M. Cherry, C. Dean, S.D. Rounsley and M. Koornneef **Arabidopsis thaliana: a model plant for genome analysis.** Science (1998) 282: 662-682. ○ 9. **The Arabidopsis Genome Initiative Nature** (2000) 408, 796-815. ○ 10. J. Yu, S. Hu, J. Wang, G.K.S. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, X. Huang, W. Li, J. Li, Z. Liu, L. Li, J. Liu, Q. Qi, J. Liu, L. Li, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Zhang, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zeng, L. Tao, J. Ye, J. Tan, X. Ren, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, J. Wang, W. Zhao, P. Li, W. Chen, X. Wang, Y. Zhang, J. Hu, J. Wang, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, L. Guojie, S. Liu, M. Tao, J. Wang, L. Zhu, L. Yuan and H.A. Yang **Draft Sequence of the Rice Genome (*Oryza sativa* L. *ssp. indica*).** Science (2002) 296:79-92.

favorable genome segments as quantitative trait loci (QTL) in experimental populations or breeding populations. Marker-assisted selection (MAS) is increasingly applied e.g. in back-crossing programs used to introduce favorable genes/alleles from exotic germplasm.

The (classical) breeding process is increased in efficiency through these measures and the necessary boost in high quality food production for the global human population is achieved through the use of new knowledge gained through plant genome research via genetic engineering. This is taking place against the backdrop of a still growing world population and the reduction of available resources like water, arable land and energy per capita. The plant genome research is an essential part in supporting the development of a sustainable and ecologically balanced agriculture. Moreover, it also supports the production of new and improved products. The potential benefits in the uses of plant genomics for humans and the environment is evident in the diversity of the uses of plants mentioned above. Through the possible use of exotic breeding material and the efficient use of such germplasm via MAS assisted back-crossing it becomes also clear that plant genomics research provides the means to enhance the use of natural genetic diversity and to support its preservation. These natural resources offer great potentials for further crop improvement plants and the information gained through plant genome analysis will strongly support its protection.

A number of other plant genome programs that also focus on crops have commenced due to the potentials and the need for enhancement of the crop improvement procedures. In the year 1998 the „National Plant Genome Initiative" was founded in the USA (http://www.nsf.gov/bio/dbi/dbi_pgr.htm). In contrast to its first period, which mainly focused on building up resources and establishing technologies, the second phase (2003–2008) of this strategy plan announced in January 2003 now focuses on the use of these tools to address a wide range of important biological questions in plant science. The second phase will involve a continued elucidation of genome structure and organization of crops like rice, maize, tomato and Medicago as key plant species. In addition we will further understand the complex processes like tuber or bulb development, wood formation, fruit development, nutrition uptake and use in plants with the help of the transfer of knowledge about gene functions from key plants to other crops. Plant genomics will expand into the analysis of biodiversity, studies of ecology and ecosystems, and the development of sustainable production systems, renewable resources and novel biomaterials. It will strongly benefit from advances in bioinformatics, the integration and interaction of plant genome databases, the development of standard operating principles and the development of new algorithms to analyze plant genomics data. Equally important will be the reaching out to our societies to further their recognition of the importance of plant genomics and the education and training of students and young researchers. National plant genomics research programs focusing on rice, maize, wheat, soybean, rapeseed and sugarcane also exist in Canada, Brazil, China, Japan, and Australia. In Europe, several larger national plant genomics programs with major emphasis on model and crop plants, including Arabidopsis, rice, wheat, barley, maize, rapeseed, grapes, tomato, potato, Medicago, sugar beet etc. have been started in the UK (GARNET, 2000, http://www.york.ac.uk/res/garnet/garnet.htm), in France (Génoplante, 1999, http://www.genoplante.org), in Germany (GABI, 1999, www.gabi.de), in the Netherlands (BioSystems Genomics, 2002, http://www.biosystemsgenomics.nl), in Sweden (SCTFG, 1999, http://www.upsc.nu), and – since 2003 – in Spain, (http://www.mcyt.es). Additionally, efforts are made to bundle national activities in Denmark, Switzerland, Slovenia, Finland, Hungary and Italy. Others are to follow. It will be necessary to make existing resources, technologies, and information freely available to enhance the output of these various initiatives, to further increase the co-operation across programs and to set cross-links between programs. In this manner the efficiency of plant genome research will be enhanced and the output achieved will be increased in spite of the limited available financial resources. The European Research Area Net Plant Genomics (http://www.erapg.org) is a European wide activity directed towards these goals.

A vision of plant genomics

Plant genomics bridges the gap between the classical molecular biology and the holistic view into complex systems. Recently the term "plant systems biology" thus emerged as it did for several other groups of species. Like in the general concepts of genomics, many commonalties exist also at this level among the studies of various species. The different kingdoms share between them the use of systematic genomics, transcriptomics, proteomics, and metabolomics and the increasing use of bioinformatics to construct models of complex biological systems. But plant systems biology will surpass the limits of the analysis of properties and mechanisms of the organisms itself much faster to the point where it will incorporate the interaction of organisms with their environment (both, biotic and abiotic influences). Environmental conditions very strongly limit and affect plant performance and particularly that of our field crops. Therefore the tolerance or resistance to adverse conditions offers the greatest potentials to increase and secure yields in sustainable production systems. Thus, while plant systems biology as a logical expansion from plant functional genomics will probably develop similarly to that of other systems from analysis of simple cellular processes to studies at the organ and entire plant, the long term vision may be a widening into the study of plants as interacting partners within ecological systems.

Patent and Licensing Agency (PLA for GABI)

One of the major strategic objective within GABI is to ensure an efficient transfer of knowledge and technology between all cooperating parties and to secure comprehensive protection of the results by intellectual property rights. In order to meet these criteria the Patent and Licensing agency (PLA for GABI) has been established and is financially supported by the member companies of the Business Platform promoting GABI Plant Genome Research (Wirtschaftsverbund Pflanzengenomforschung GABI e.V. WPG).

Since 1998 the German Plant Breeders Association (Bundesverband Deutscher Pflanzenzücher BDP e.V.) has established its subsidiary GVS (Gesellschaft für Erwerb und Verwertung von Schutzrechten GVS mbH) for the acquisition and exploitation of property rights within the field of 'green' biotechnology. Taking advantage of these facilities the PLA for GABI was established at the GVS in the House of Plant Breeding in Bonn.

One of the main tasks of the PLA for GABI is to check manuscripts before publication and to give advice for securing intellectual property rights. This holds true not only for manuscripts of scientific articles but also for all kinds of posters and internet presentations, abstracts and lectures. Within six weeks the PLA for GABI has to decide to either pass the paper or to secure the IPR. In addition the PLA for GABI gives advice and support to scientists on how to modify manuscripts in order not to endanger subsequent patent potentials. If the decision is made to file a patent on GABI results, the PLA for GABI supports the applicants by conciliating a competent patent attorney and by the absorption of charges for the basic IPR protection. Concerning licence offers for these patents, the PLA for GABI acts as a mediator between the patent assignee and the members of WPG.

In order to guarantee a productive and trustful cooperation between the different partners, groups and bodies within GABI comprehensive agreements had to be arranged and signed. Supported by a legal task group established by the WPG, the PLA for GABI keeps a central role in designing, negotiating and signing contracts, agreements, and material transfer agreements within the context of GABI.

As one major link between scientific research groups and industry the PLA for GABI takes an active part within the boards and panels of GABI, as e.g. the GABI Steering Committee and the Scientific Coordination Committee. Furthermore the PLA for GABI keeps close contacts to corresponding exploitation agencies of universities and research institutes participating in GABI, in particular to Garching Innovation GmbH (GI), the technology transfer organisation of the Max-Planck-Gesellschaft (MPG).

When the idea came up to start a cooperation between GABI and the French plant genome initiative Génoplante, needs for legal agreements became obvious. With Génoplante Valor, the exploitation company of Génoplante, as partner the PLA for GABI takes central positions within the legal affairs concerning the Cooperation between the French plant genome initiative Génoplante and GABI. In this context regulations for material exchange as well as cooperation contracts for the already running joint projects have been established by support of a legal task group of WPG (JurAG) and have been negotiated with Génoplante. Thereby the PLA for GABI contributes to the successful and exemplary cooperation between both plant genome initiatives. With the experiences in the establishment of the fruitful cooperation between GABI and Génoplante, the PLA for GABI will take an active role in further intensifying this cooperation and in initiating further international efforts for cooperations and joint programs aiming to raise a European Plant Genome Network.

Business Platform promoting GABI Plant Genome Research
(WPG)

At an early stage, German companies with interests in plant breeding, plant protection, and the plant processing industry recognized the potential of plant genome research for dynamic social and economic development. Therefore, economic enterprises have had a defining role since the planning stages of GABI.
Guidelines have been established that make GABI attractive for companies acting on different levels of the value-producing chain:

- Bringing together Germany's existing research capabilities and coordinately transforming its research infrastructure into an integrated network
- Connecting the interests of science and industry
- Establishing the promotion of quality research and creating favorable conditions for direct participation by industry
- Creating technology transfer strategies promoting the efficient transformation of research results into innovative products
- Guaranteeing comprehensive patent coverage
- Preserving the autonomy of participating companies: Every individual company may decide freely the way and extent to which it participates in GABI projects.

Once these guidelines were established, the registered association Wirtschaftsverbund Pflanzengenomforschung GABI (Business Platform promoting GABI Plant Genome Research, WPG) was founded simultaneously with the public advertisement of the Genome Analysis of the Plant Biological System (GABI) promotion focus by the BMBF on September 18, 1998.

The WPG is a registered association seated in Bonn. The purpose of the association is the promotion of research – especially GABI. Initiated by industry, this organization has succeeded in achieving two essential aims:

1. Establishment of a central contact partner for industry on matters of policy, science, and economy as well as for all other GABI bodies.
2. Coordination of the activities of the companies participating in GABI and joint definition of binding proceedings and rules for members within the framework of the promotion regulations of the BMBF.

Thus, members of the WPG jointly assume the responsibility of enforcing the above guidelines and are therefore a supporting pillar of the whole GABI concept. In return, members have the opportunity to carry through their own projects and gain access to results generated by GABI.

12 companies founded the WPG but within the first year, the number of members increased to 24 and is currently 29. The five-member Managing Committee is responsible for all association matters. All positions are assigned on an honorary basis. The management is located at the Haus der Pflanzenzüchtung (House of Plant Breeding) in Bonn. The WPG Managing Committee acts as the central communication partner for industry. The Managing Committee selects delegates for the GABI control bodies from among its members. Two WPG members are represented in the Steering Committee. Furthermore, nearly half of the members of the Scientific Advisory Board come from the WPG. Thus, the WPG assures a balanced influence on research political as well as scientific decisions within GABI. The Managing Committee elaborates proposals for a contribution scheme and submits these proposals to the WPG members for discussion. The contribution scheme takes into account the very heterogeneous structure of the WPG members consisting of small and medium as well as large enterprises. The membership fees are primarily spent to finance the activities of the WPG and of the Patent and Licensing Agency (PLA for GABI). Furthermore, the Managing Committee decides upon the admission of new WPG members. An informal WPG membership application may be submitted to the Managing Office at any time. All enterprises in the field of the plant breeding, plant protection, and plant processing industries as well as in biotechnology may become members. These members (and their parental companies) must have their headquarters in Germany and carry out the major part of their research and development activities in Germany. WPG membership opens numerous possibilities. Together with experts from research and industry, an enterprise may establish research networks for special problem fields. Thus, the enterprise gets access to extensive expertise.

Furthermore, the activities of the PLA for GABI, guarantee an extensive information flow especially from the projects of the Research Area 1 to all WPG members. WPG members are informed on a regular basis about research projects aims as well as about the results. They receive publications in advance and have a time-limited exclusive reading right for primary data. Thus, each company may check the data in terms of usefulness for its own entrepreneurial activities. Furthermore, each member is counseled on the application for and granting of patents coming from the Research Area 1 and has the possibility to take out a license on a most-favored-nation basis. The complicated and cost-intensive analysis of the whole range of data can normally not be managed by individual members. By means of the network structure and the establishment of the PLA for GABI, the patent protection results from Research Area 1 are guaranteed - WPG members can access these results and extend their use for their own activities within or outside of GABI.

The general concept of the WPG does not aim solely at offering individual members ready solutions: It requires active co-operation in addition to financial contribution. This also applies to the international interlinking of GABI with other genome projects such as the French initiative GENOPLANTE. It is industry's responsibility to launch such pilot projects. National borders cannot be allowed to limit efficient research.

Business Platform promoting GABI Plant Genome Research (WPG)

Managing Committee of the WPG (state of April 2003)

Chairman
Andreas J. Büchting, KWS SAAT AG

Deputy
Dieter Berg, Bayer Crop Science GmbH
Gisbert Kley, Deutsche Saatveredelung Lippstadt- Bremen GmbH

Members
Wolf v. Rhade, Nordsaat Saatzuchtgesellschaft mbH,
Hans Kast, BASF AG

Managing Director
Carl Bulich, Wirtschaftsverbund Pflanzengenomforschung GABi e.V., Kaufmannstrasse 71, D-53115 Bonn

WPG-Member-Companies

1. Plant Breeding
- Kartoffelzucht Böhm KG
- Böhm-Nordkartoffel Agrarproduktion GbR
- W. von Borries-Eckendorf GmbH & Co
- Saatzucht Josef Breun GbR
- Pflanzenzucht Dr. h.c. Carsten - Inh. Erhardt Eger KG
- Deutsche Saatveredelung Lippstadt-Bremen GmbH
- A. Dieckmann-Heimburg Saatzucht Sülbeck
- Hybro Saatzucht GmbH & Co. KG
- KWS SAAT AG
- Saatzucht Fritz Lange KG
- Lochow-Petkus GmbH
- Norddeutsche Pflanzenzucht Hans-Georg Lembke KG
- Nordkartoffel Zuchtgesellschaft mbH
- Nordsaat Saatzuchtgesellschaft mbH
- NORIKA Nordring-Kartoffelzucht- und Vermehrungs-GmbH
- Raps GbR, Saatzucht Lundsgaard
- Saaten-Union Resistenzlabor GmbH
- Saka-Ragis Pflanzenzucht GbR
- Saatzucht Hans Schweiger & Co. oHG
- Saatzuchtgesellschaft Streng's Erben GmbH & Co. KG
- Fr. Strube Saatzucht KG
- Südwestdeutsche Saatzucht GmbH & Co. KG

2. Plant Protection/Biotechnology
- Bayer Crop Science GmbH
- BASF AG

3. Processing industry
- Südzucker AG

4. Start-up companies
- Genetic Concepts GmbH
- G.A.G. Bioscience
- Trait Genetics GmbH
- Genetic Concepts GmbH


F11
8
)


3
4
%
5
A
S
Y

The GABI Resourcecenters

GABI-KAT Overview

The GABI-KAT (**K**ölner **A**rabidopsis **T**-DNA lines) project that is executed at the Max-Planck-Institute for Plant Breeding Research provides easy access to knock-out (KO) mutants in preselected genes for the study of gene/phenotype relationships in *Arabidopsis thaliana*.

International efforts culminated in the first sequenced genome, 120 Mb, of a flowering plant in the year 2000. So far, about 1,000 genes with a loss-of-function phenotype are known, and an ambitious plan is underway to obtain information about the biological- and agronomical significance of all 26,500 *A. thaliana* genes until 2010. For such studies of gene-function-relationships in all sorts of organisms, random insertional mutagenesis, in which the foreign DNA not only disrupts a gene but also provides a tag for the molecular characterization of the mutation, has been used.

In *A. thaliana*, transposon and T-DNA mutagenised populations have been exploited to unravel gene function, e.g. by searching for mutants in DNA pools by PCR-based screens. Now, the availability of the A. thaliana genome sequence enables the identification of each *A. thaliana* gene by sequence and genome position.

GABI-KAT early on started to exploit these new possibilities for the establishment of a Flanking Sequence Tag (FST) based T-DNA population in the accession Columbia-0 (the sequenced accession) where sequences of PCR-amplified insertion sites from individual mutants are stored in a database. GABI-KAT created and optimized a pipeline for the generation of FSTs from T-DNA mutagenised *A. thaliana* lines. The pipeline included robotized extraction of genomic DNA in 96-well format, an adapter-ligation PCR method for amplification of plant sequences adjacent to T-DNA borders, automated purification and sequencing of PCR fragments, and computational processing and annotation of the generated sequence files.

Users can apply BLAST- or keyword-searches to easily select lines with KO alleles according to their special interest. Also, the availability of data on the location of the insertion site, e.g. in an intron or an exon, allows the prioritization of candidate lines.

Currently, the GABI-KAT population contains insertions in more than 60% of the predicted *A. thaliana* genes, and more than 1500 confirmed insertion mutant lines have been distributed to GABI partners, as well as to labs in Germany and all over the world. GABI-KAT is generally recognized as an important resource for the scientific community.

GABI-LAPP Overview

Large-scale and high throughput approaches increasingly play an essential role in the study of biological systems. These need to be examined by extensive methods to reach the generation of large data sets of genomic and proteomic networks, which is ultimately necessary to understanding the complex system that is a living organism. In plant biology, complex system analysis is currently most accessible in the model plant *Arabidopsis thaliana*, whose genome is completely sequenced.

Sequence information alone allows only an incomplete picture of a biological system. However, sequence information is invaluable as the sole frame for further investigation. In general, life depends on the selective readout of individual genes from the genome. These are converted into transcripts and translated into a protein sequence. Ultimately, post-translational modifications are carried out, which together with alternative splicing multiply the information coded by DNA. Furthermore, protein cleavage and multi-protein complex formation may occur. All these processes influence the function of proteins. Despite the ever-increasing volume of genomic data available, a relatively small percentage of the generated sequence can be assigned a function with certainty. One of the challenges of the post-genomic era lies in the realm of protein studies that utilize these data for better understanding.

The central task of the GABI-LAPP (Large-scale Automated Plant Proteomics) groups is to develop different technology platforms within GABI. But it is also the creation of genomic resources and tools for the scientific work with plants to cover and elucidate the above mentioned biological processes. Within GABI-LAPP the focus lies on protein chips (subproject 1) and DNA microarrays (subproject 4) as well as 2-DE and MS methods (subproject 2) and the yeast two hybrid system (subproject 3), which are the major technologies for plant proteomics and genomics. These tools are developed and adapted for high throughput using *Arabidopsis thaliana* as a model organism. But the technology is applicable to other organisms of interest, i.e. crop plants like barley. Underlying is the idea that within three years a basic toolbox will be available to analyze RNA as well as protein samples including interaction studies on a genome-wide level.

The groups working in the GABI-LAPP subprojects are located at the Max-Planck-Institute for Molecular Genetics in Berlin in the Department of Vertebrate Genomics headed by Prof. Hans Lehrach.

GABI Plant Overview

The genomes of most agriculturally important cereal species are characterized by a large size, exceeding that of the human genome by up to one order of magnitude. Nevertheless, ongoing improvements in DNA-sequencing and bioinformatics together with the availability of ample information on the rice genome make cereal genomes increasingly amenable to a systematic structural and functional analysis. The overriding goal of the GABI-Plant project is the generation of key resources required for the structural and functional analysis of the barley genome (*Hordeum vulgare* L.). This species was chosen because of its agricultural importance, its experimental suitability and the central position of its genome within the *Triticeae* tribe, which includes other agriculturally important species like wheat and rye.

The GABI-Plant resources centre generates resources and provides services for a series of related projects within the GABI consortium in the following areas:

Germplasm and Mapping Populations. A comprehensive transcript map of EST-based markers was established as a basic resource for anchoring functional and genetic information to the barley genome. Moreover, the transcript map serves as a connecting point to the rice genome. Three major mapping populations are available for distribution as seeds, DNA samples or hybridisation filters.

BAC libraries trimmed for high throughput screening procedure. DNA pools of a BAC library, constructed form the cultivar Morex, were established for high-throughput screening. In this context, the assignment of randomly selected genes (ESTs) to individual BAC clones provides an estimate of the gene distribution and the size of the gene space of the barley genome.

A core collection of barley cDNAs and expressed sequence tags (ESTs). A comprehensive collection of more than 150,000 ESTs from a diverse range of tissues and growth stages was established. It provides the basis for high-throughput EST-based marker development, unigene definition and cDNA-array development. Regarding the latter, a 10,000 unigene cDNA array was developed in collaboration with the IPK Plant Genome Resources Center (PGRC).

Development of a DNA-array based method for efficient identification of plant genetic resources. An on-chip oligonucleotide-extension assay system was developed as a generic resource that can be deployed for high throughput SNP genotyping in barley as well as in other genomes. A corresponding service is available from the start-up company 'ArrayOn'.

Bioinformatics Activities within the GABI Project

The genomics era in plant biology has fundamentally changed the perspective on the understanding of molecular and cellular processes within plants. Systematic, genome based and large scale analyses employing high-throughput analytical methodologies are the state-of-the-art. Complete, up-to-date, well structured, and comprehensive data and tools are prerequisites to get the maximum out of genome-based experiments. To a large extent the interpretation of experimental genome data relies on the application of bioinformatics methods, on the consideration of contextual information and the linkage with analytical genome data from heterogeneous origins. Bioinformatics is a key enabling technology to process and structure experimental data derived information. Bioinformatics is obliged to render data resources and technologies for the interpretation of data and the detection of functional interactions. Thus bioinformatics based projects develop indispensable strategic resources for transcending functional information from model genomes to crops.

Within the GABI network various bioinformatics activities have been established, tight connections to the GABI research community have been developed and bioinformatics activities significantly contribute to the international visibility of the GABI project.

GABI Info is carried out at MIPS at the Research Center for Environment and Health (GSF), Neuherberg. GABIInfo's objective is to match the challenges of genome- and function oriented research. It develops and supports a wide range plant genomics resources and aims to systematically apply and develop state-of-the-art bioinformatics tools. A specific focus is devoted to comparative and combinatorial analysis of heterogeneous datasets.

The GABI primary database (Gabi-PD) has been developed at the German Resourcecenter for Genome Research (RZPD), Berlin. GabiPD collects, integrates and links all relevant data from the GABI projects and makes them accessible via internet. Interfaces for text based retrieval of sequence, snp, mapping data etc. are provided and GabiPD acts as a repository for GABI generated data and proprietary data deposition. GabiPD also offers bioinformatics services and tools to the gabi community.

At the University of Cologne efforts have been put into developing into a database, Aramemnon, with the aim to summarize and structure data on Arabidopsis transmembrane proteins. Aramemnon contains data on sequences, relationships among different proteins, transmembrane protein predictions and signal peptide predictions of Arabidopsis membrane proteins. (More information about Aramemnon, please find in the chapter Arabidopsis Protein Communities).

Also included:

Sequencing of T-DNA flankings for "*in silico*" detection of tagged alleles (GABI-KAT – Kölner Arabidopsis T-DNA lines)

A Génoplante-GABI bi-lateral project for sharing sequence data and Computer resources on T-DNA transformants of *Arabidopsis thaliana*

Resourcecenter
GABI-KAT

Bernd Weisshaar, Mario Rosso, Yong Li, Sabine Steiner-Lange, Koen Dekker, Bernd Reiss, Heinz Saedler
Max-Planck-Institute for Plant Breeding Research (MPIZ), ADIS – "DNA core facility", Cologne

Goals/Aims

To build up resources for efficient progress in plant genomics, a set of T-DNA mutagenised lines with insertions sites identified by DNA sequencing is being generated. The output is a database describing, for the lines in the collection, which gene(s) have been disrupted, so that users can select KO alleles for their needs. Finally, mutant seeds are delivered for physiological and gene-function relationship studies. The ultimate goal is to use the new knowledge obtained from the model system *Arabidopsis thaliana* for the further improvement of today's food crops.

Arabidopsis thaliana (Col-0) was transformed with a special Ti-based binary vector designed to allow selection of transformed plants in the green house. Transgenic seeds to create a population of at least 70,000 lines are available. KO alleles from these 70,000 lines will make a significant contribution to saturate the *A. thaliana* genome, which contains more than 26,500 genes, with NULL mutations. The project profited from the previous experiences gained in ZIGIA (Zentrum zur Identifikation von Genfunktionen durch Insertionsmutagenese bei *Arabidopsis thaliana*, project number 3117519). In 2003, the integration of ZIGIA materials into GABI-KAT is initiated.

The actual start of the GABI-Kat project was June 2000. This report covers data and developments until December 2003. In a collaboration project with Génoplante (title: "Gemeinsame Nutzung von FST-Sequenzinformationen über T-DNA-transformierte *Arabidopsis thaliana*-Linien und den entsprechenden Bioinformatik-Ressourcen (Kooperation Génoplante/GABI)"; project number 0312480), FST data are exchanged and KO mutants are made accessible to the respective partner program. Results from this project are incorporated into this report.

Project status and Further work

The following service is offered to other GABI projects and scientists who signed a "Verpflichtungserklaerung" to stick to GABI rules:

- a searchable database containing flanking sequence tag (FST) information,
- distribution of a seed sample of "*in silico*"– selected individual T-DNA lines,
- confirmation of the insertion-site by gene-specific PCR,
- access to new FSTs from the corresponding Génoplante project.

The steps required to reach the above goal are:

A: Selection of finally 70000 T-DNA tagged lines of A. thaliana Col-0.
B: Isolation of DNA of individual lines, PCR amplification of flanking fragments, and sequencing.
C: Construction of a FST database for "*in silico*" detection of KO alleles, including the evaluation of FST data from Génoplante.
D: Distribution of confirmed T2 seeds (seeds from T1 plants) from a requested individual T-DNA line for studies on particular candidate genes.

Below, the four areas are discussed in more detail. A summary "in numbers" is included at the end of the document. These data are taken from the "GABI-Kat LIMS" (LIMS: laboratory information management system) which stores important data generated within GABI-Kat, along with organisational information required to manage seed distribution.

A) Selection of T-DNA lines

The growth and the selection of transgenic lines population is carried out continuously. About 600 transgenic plants were selected each week for herbicide resistance. Leaf material was harvested in 96-well blocks for high-throughput preparation of genomic DNA. Individual seed material was harvested after ripening and prepared for storage.

B) Isolation of DNA from individual lines and sequencing

From the harvested leaf material genomic DNA was prepared and used for PCR-based amplification of the genomic sequences adjacent to the T-DNA left border. The resulting PCR fragments were purified and sequenced; the methods applied have been optimised and are running since May 2001 in full production speed.

C) Construction of sequence-based database

The initial output of step (B) is a collection of sequence data trace files. These data were processed by a bioinformatics analysis pipeline which was tightly integrated with the existing ADIS LIMS (designated SeqOrderDB). To manage GABI-Kat, the primary plant-related data as well as the sequence-related data produced by the analysis pipeline are stored in a database designated GK-LIMS.

○ Y. Li, M. G. Rosso, N. Strizhov, P. Viehoever and B. Weisshaar **GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in *Arabidopsis thaliana*.** Bioinformatics (2003) 19, 1441-1442. ○ M. G. Rosso, Y. Li, N. Strizhov, B. Reiss, K. Dekker and B. Weisshaar **An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics.** Plant Molecular Biology (2003) 53, 247-259. ○ N. Strizhov, Y. Li, M. G. Rosso, P. Viehoever, K. A. Dekker and B. Weisshaar. **High-throughput generation of sequence indexes from T-DNA mutagenized *Arabidopsis thaliana* lines.** BioTechniques (2003) 35, 1164-1168.

To allow access from the GABI community to the FST data, a web interface was created (designated "SimpleSearch"). The database searchable via the web interface can answer the following question: has a hit in gene "YFG" (your favourite gene identified by AGI GeneCode or by its sequence) been found? Finally, a line-ID is identified.

Secure and password-protected access to the SimpleSearch tool was opened in July 2001. The first outside request was placed on 09.08.2001 (GABI project "Arabidopsis Protein Communities"). Since then, an increasing number of GABI researchers is using the resource (1508 send outs until December 03). In addition, access options for other researchers from Germany and also from the world-wide scientific community have been implemented. Since June 2002, public access to GABI-Kat lines is possible (website: http://www.mpiz-koeln.mpg.de/GABI-Kat/).

D) Distribution of seeds

After a request has been placed on the basis of search results from using SimpleSearch, the respective FST hit is confirmed at GABI-Kat by PCR using material from young T2 plants and a gene-specific primer from the gene corresponding to the FST. After confirmation, about 50 T2 seeds per line are delivered. T3 seeds from the segregating T2 generation are stored at GABI-Kat.

The confirmation rate is currently about 78%, and we are improving this rate by determining the reasons for confirmation failure and implementing measures to avoid the respective reasons.

We are working on an improvement of the T-DNA vector used to at least partially overcome the technical limitations of the current protocol to generate T-DNA insertion site flanking fragments. The planned improvement is based on the frequently cutting restriction enzyme used to restrict the genomic DNA close to the insertion site which must not cut within the T-DNA border sequence. Plants containing this T-DNA are used in 2003 for FST generation.

GABI-Kat in Numbers:

A) Timetable (dates):

01.06.2000:	Start of the project
01.01.2001:	All positions filled, including database manager
01.05.2001:	start of the production phase of FST sequence generation
09.07.2001:	Opening of the GABI-Kat website (GABI-version of SimpleSearch)
09.08.2001:	First order of a GABI-Kat line
15.08.2001:	Approval of the GABI-Kat MTA by PTJ
16.11.2001:	Delivery of the first line to a GABI user
01.04.2002:	SimpleSearch updated to allow BLAST searches on FST sequences
01.06.2002:	Opening of the public GABI-Kat website (public version of SimpleSearch)
01.08.2002:	SimpleSearch updated to include graphic display of insertion sites
01.09.2002:	Start of the Génoplante-FST/GABI-Kat exchange program
23.09.2002:	First request from Génoplante for a GABI-Kat line Extension of project approved
15.02.2003	New "gene hit" definition (major update of Simple Search)

B) Numbers (by end of December 2003):

Resistant lines germinated and selected (incl. ZIGIA):	83,700
Lines for which analysis has been started:	64,200
Lines of which seeds are harvested:	82,400
Lines with "genome hit" detected:	55,561
Number of genes with at least one FST hit (non-redundant gene hits):	16,338
Number of "genome hits" exchanged with Génoplante:	9,650
Total number of non-redundant gene hits available:	14,395
Percentage of gene coverage:	60.2%
Lines requested in total:	2,282
Requests with GABI rules:	1,144
Requests from GABI projects:	571
Requests from Génoplante:	50
Requests from the public:	517
Lines sent out:	1,508
Percentage of requests from Germany	77.5%
Lines which are publicly available:	54,000
Publicly available non-redundant gene hits:	12,470

Generation of Arabidopsis and Barley protein expression libraries and chips

Resourcecenter
GABI-LAPP

Birgit Kersten, Tanja Feilner, Armin Kramer, Alexandra Poßling, Silke Wehrmeyer, Hans Lehrach, Dolores J. Cahill
Max-Planck-Institute for Molecular Genetics (MPI-MG), Berlin

Introduction

Protein array technology emerged as a new tool to enable ordered screening of proteins for expression and molecular interactions in high-throughput (Cahill, 2001; Kersten *et al.* 2002). For the generation of protein arrays with recombinant proteins, cDNAs have to be cloned into an appropriate expression vector. After protein expression and purification, the proteins have to be arrayed onto an immobilising surface. Besides more classical solid phase substrates, such as microtiter plates and membrane filters, protein arrays have recently been developed in chip formates. Several applications of protein chips have been described (e.g. antibody-antigen, enzyme-substrate, and protein-protein interactions) but no applications using plant protein chips were published so far.

The aim of this subproject was to develop these new technologies for the analysis of plant proteins, starting with *Arabidopsis thaliana* as a model organism followed by Barley, an important crop in Germany and northern Europe.

Cloning Strategies

We used different *E. coli* expression vectors which enable the expression of the proteins as fusion proteins with a N-terminal His-tag. High-throughput cloning was performed either by the construction of ordered cDNA expression libraries (library-approach) or by recombinational cloning of full-length cDNAs using GATEWAY™-technology (ORF-approach). The first approach is restricted by following reasons: the expressed proteins are not all full-length, but if full-length, proteins may contain artificial parts of the expressed 5'-UTR which may lead to artefacts in functional analysis. The second approach depends on the availability of a sequenced and annotated genome.

After the completion of the Arabidopsis genome sequencing and in silico annotation in 2000, we were able to use both approaches to clone Arabidopsis cDNAs. For Barley the library-approach was the method of choice.

Generation of cDNA expression libraries

For Arabidopsis we constructed a cDNA library from inflorescence meristem (~41,000 clones) in the *E. coli* expression vector pQE-30NAS*TattB* in cooperation with Richard Immink (Plant Research International, The Netherlands) (Feilner *et al.* 2002). Using robot technologies, this library was arrayed in microtiter plates. To check the quality of this library 431 randomly chosen clones were sequenced. 426 sequences (99%) were found to match Arabidopsis proteins in the database. 192 of these sequences (45%) matched the beginning of an Arabidopsis protein, suggesting that the coding region had been cloned in full-length.

Two Barley cDNA libraries were generated from pericarp and embryo sack in the *E. coli* expression vector pQE-30NST in cooperation with Winfriede Weschke and Volodymyr Radchuk (IPK, Gatersleben). Both libraries were picked into 384 well microtitre plates. The library sizes are ~21,500 (embryo sack) and ~23,800 clones (pericarp). The average insert size was ~1kB for the constructed libraries.

To select putative protein expression clones from the constructed libraries (Arabidopsis and Barley) all clones were gridded onto high-density filters. The filters were screened using a monoclonal anti-His antibody. Positve clones were rearrayed into sublibraries:

~5,000 from Arabidopsis mersitem
~4,100 from Barley embryo sack
~2,900 from Barley pericarp.

The expression clones from Arabidopsis meristem and Barley embryo sack were sequenced using 5'-vector primers.

Furthermore we constructed a pooled cDNA library from Arabidopsis pistils in a GATEWAY™-entry vector (~660,000 clones) in cooperation with Richard Immink (Plant Research International, The Netherlands), which can be shuttled in expression vectors.

Generation of full-length expression clones

Using the ORF-approach we created 230 full-length cDNA clones from Arabidopsis in *E. coli* expression vectors, in cooperation with Isabell Witt (MPI-MP, Golm) and Bernd Weisshaar (MPIZ, Köln). 185 of them were identified as expression clones as indicated by SDS-PAGE following overexpression in *E. coli*.

Protein purification in high-throughput

We developed an automated method for the purification of His-tagged proteins. Protein expression is performed in small volume (1.5ml) and in high-throughput (96-well microtitre plates). The tagged proteins are purified in high-throughput by Biorobot 8,000 (Qiagen). Using this technology, we are able to perform up to three purifications runs (96 proteins each) a day. We got a yield of 25µg protein on average (70µg/ml) under denaturing conditions. 384 different Arabidopsis (Figure 1) and Barley proteins were purified for the generation of protein chips or MS analysis.

The development of automated purification methods to get soluble or refolded proteins is in progress.

Generation of Arabidopsis protein chips for different applications

For the generation of protein chips the purified recombinant Arabidopsis proteins were robotically arrayed onto glass slides coated either with a nitrocellulose membrane (FAST™-slides, Schleicher & Schüll) or a polyacrylamide membrane (PAA-slides). We are able to spot up to 3,200 proteins onto one slide. All His-tagged Arabidopsis proteins tested up to now were specifically detected on the chips using an anti-His antibody. The detection limit was approximately 2 to 4fmol protein on FAST™-slides or 0.1

○ L. Bürkle, T. Feilner, W. Brenner, P. Giavalisco, H. Lehrach and B. Kersten **Genomic and proteomic approaches in plant research (book article).** In Pandalai S.G., ed., Recent Research Developments in Plant Molecular Biology (2003) 1, 239-263, Research Signpost, Kerala, India. ○ T. Feilner **Hochdurchsatz Generierung und Analyse von *Arabidopsis thaliana*-Proteinen (Dissertation, Theses).** Fachbereich Biologie, Chemie Pharmazie; Freie Universität Berlin; September 2004. ○ T. Feilner, J. Kreutzberger, B. Niemann, A. Kramer, A. Possling, H. Seitz and B. Kersten **Proteomic studies using microarrays (review).** Current Proteomics (2004) 1 (4), in press. ○ T. Feilner, C. Hultschig, J. Lee, S. Meyer, R. Immink, A. Koenig, A. Possling, A. Beveridge, D. Scheel, D.J. Cahill, H. Lehrach, J. Kreutzberger and B. Kersten **High-troughput identification of Arabidopsis MAP kinases substrates.** (2004) submitted. ○ B. Kersten, A. Possling, F. Blaesing, E. Mirgorodskaya, J. Gobom and H. Seitz **Protein microarray technology and UV-crosslinking combined with mass spectrometry for the analysis of protein-DNA interactions.** Analytical Biochemistry (2004b) 331, 303-313. ○ B. Kersten, T. Feilner, P. Angenendt, P. Giavalisco, W. Brenner and L. Bürkle **Proteomic approaches in plant biology (review).** Current Proteomics (2004a) 1, 131-144. ○ B. Kersten, T. Feilner, A. Kramer, S. Wehrmeyer, A. Possling, I. Witt, M.I. Zanor, R. Stracke, A. Lueking, J. Kreutzberger, H. Lehrach and D.J. Cahill **Generation of Arabidopsis protein chips for antibody and serum screening.** Plant Molecular Biology (2003) 52, 999-1010. ○ B. Kersten, L. Bürkle, E.J. Kuhn, P. Giavalisco, Z. Konthur, A. Lueking, G. Walter, H. Eickhoff and U. Schneider **Large-scale plant proteomics (review).** Plant Molecular Biology (2002) 48, special issue, 133-141. ○ A. Kramer **Herstellung von Gerste-Protein-Chips mit Proteinen aus cDNA-Expressions-Bibliotheken und Testung ihrer Anwendbarkeit für Phosphorylierungsstudien in vitro (Diplomarbeit, Theses).** Fachbereich Biologie, Chemie Pharmazie; Freie Universität Berlin; 24.9.2003. ○ A. Kramer, T. Feilner, A. Possling, V. Radchuk, W. Weschke, L. Bürkle and B. Kersten **Identification of barley CK2a targets by using protein microarray technology.** Phytochemistry (2004) 65, 1777-1784. ○ H. Seitz, M. Krause, C. Gotthold, U. Borgmeier, B. Kersten, C. Luebbert and V. Heiser **Application of protein array technology to study protein-protein interactions and protein-DNA-interactions.** Molecular & Cellular Proteomics (2003) 2, No. 9, 939.

to 2fmol on PAA slides for the proteins tested. The Arabidopsis protein chips were used for the characterization of the specificity and cross-reactivity of monoclonal antibodies and new polyclonal sera (Kersten *et al.* 2003). The sera were generated against recombinant Myb6 and Dof43 in cooperation with Isabell Witt (MPI-MP, Golm) and Bernd Weisshaar (MPIZ, Köln). We demonstrated that a monoclonal anti-TCP1 antibody (Figure 2) as well as the anti-Myb6 and anti-Dof43 sera bound specifically to their respective antigens and did not cross-react with the other 94 proteins on the chips. Furthermore we started with protein phosphorylation studies (Figure 3) in cooperation with Ulrich Wobus (IPK, Gaterleben) and protein-DNA interaction studies in cooperation with Ralf Stracke (MPIZ, Köln) using protein chips.

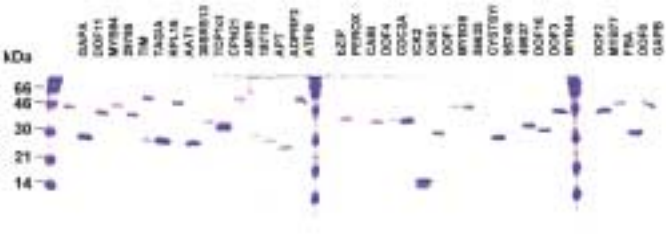


Fig 1: *This figure shows 37 different Arabidopsis proteins (full-length) following separation by SDS-PAGE and Coomassie staining. The proteins were expressed and purified in high-throughput.*

Outlook

All expression clones and libraries from Arabidopsis and Barley are a resource for the GABI-community. They can be used as a source to get cDNAs or to express recombinant proteins. Purified proteins of Arabidopsis and Barley may be used for several applications, such as antibody generation in addition to the generation of protein chips. The setup of the protein chip technology for plant proteins and the development of several applications for Arabidopsis and Barley protein chips and filters is a valuable resource for the GABI-community.

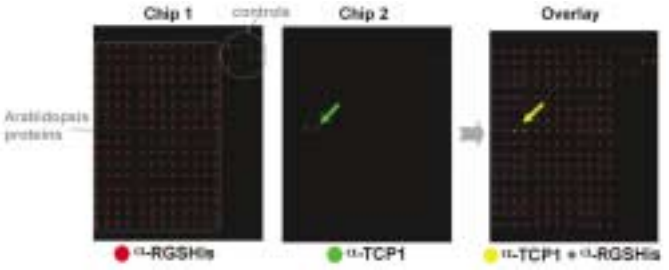


Fig 2: *The figure shows protein chips with 95 different Arabidopsis proteins. The purified proteins and controls were spotted in duplicates on FAST™-slides and screened with an anti-His (chip 1) or anti-TCP1 (chip 2) antibody. Proteins detected with both antibodies appear in yellow in the overlay.*

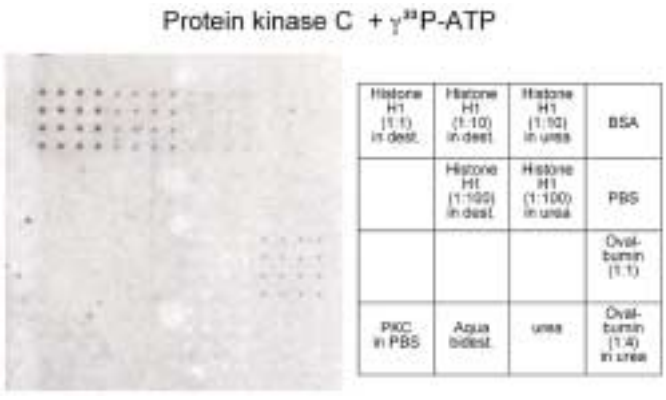


Fig 3: *This figure shows a pilot study on phosphorylation of histone H1 (bovine) on chip by protein kinase C (rat). Histone H1, protein kinase C (PKC) and several controls were spotted on FAST™-slides. The chip was incubated for one hour with PKC in the presence of γ33P-ATP. Signals were visualised by x-ray film.*

2-D gel electrophoresis and mass spectrometry for the analysis of plant proteomes

Resourcecenter
GABI-LAPP

1 Johan Gobom, 1 Patrick Giavalisco, 1 Niklas Gustavsson, 1 Eckhard Nordhoff, 2 Joachim Klose, 1 Hans Lehrach
1 Max Planck Institute for Molecular Genetics (MPI-MG), Berlin
2 Institute of Human Genetics, Humboldt University, Berlin

Introduction

The combination of two-dimensional gel electrophoresis and mass spectrometry has become an important analytical technique for the characterization of complex protein populations extracted from tissue, cells or subcellular fractions. The large-gel 2-D electrophoresis (2DE) technique, developed by Klose and coworkers over the past 25 years, provides the resolving power necessary to separate crude proteome extracts of higher eukaryotes. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) provides the sample throughput necessary to identify thousands of different protein species in an adequate time period. Adaptation of the 2DE technique to separation and detection of proteins in crude proteome extracts from plant tissues and the development of technology, experimental protocols and software for identification of the fractionated proteins by the use of MALDI-TOF MS is part of the research effort, Large Scale Automated Plant Proteomics (LAPP).

Results

Plant protein extraction and 2DE

A protocol was developed for preparation of protein extracts from plant material for 2DE analysis. Proteins are extracted from the plant tissue using different solubilizing agents, applied in sequence. From one plant tissue sample, two or three fractions are obtained containing partially different proteins. Using this protocol, the number of different proteins that can be detected by 2DE significantly increased in comparison to what has been previously reported. Using this protocol, reproducible 2DE patterns of *Arabidopsis thaliana* leaf, stem, seedling, flower, rosetta leaf, and silique were established (Giavalisco *et al.* 2003). For example, in *Arabidopsis thaliana* stem tissue, 2600 protein species were detected in the first fraction, and 3200 in the second. Image analysis revealed 900 spots to be common to the two fractions, resulting in a total of 4600 unique detected protein species (fig 1 a, b). In comparison, only 1700 protein spots were detected when using the precipitation-based extraction (fig 1 c).

Gel spot excision and protein sample processing

The ability to mass spectrometrically analyze the thousands of proteins separated and detected in a single gel puts stress on sample processing routines. Accurate excision of protein spots, conditioning of the excised gel fractions, in situ reduction of disulfide bonds, alkylation of cysteine residues, in situ proteolysis, extraction of the digestion products from the gel matrix, and peptide purification/concentration as well as the mass spectrometric sample preparation, are the necessary steps that interface large-gel 2-DE and MS. We developed an automatic gel exciser (fig 2 a), which was interfaced with commercial software for spot detection. Excised gel samples are transferred to 96-well pierced microtitre plates (MTPs). All subsequent sample processing steps are conducted in this format. A solenoid-valve liquid dispenser has been designed for rapid and accurate supplement of reagents and solvents to the MTPs (fig 2 b). Liquid is removed from the MTP by centrifugation. A method has been developed that integrates purification, concentration, and preparation of the protein digests for MALDI MS (Gobom *et al.* 2001). The protocol is based on the observation that a microcrystalline layer of hydrophobic MALDI matrix compounds exhibits a strong binding affinity for peptides in acidified aqueous solutions. Protein digests can thus be prepared for MALDI MS, without preceding purification. Prepared samples are homogeneous, facilitating automatic spectra acquisition. Using a liquid handler, this method allows preparation of large numbers of samples with little effort and in short time (fig 2 c).

Data analysis and visualization

The large amount of data accumulated from 2DE-MALDI-MS experiments requires routines for data management, algorithms for data analysis, and software tools for visualization. We have developed new computer algorithms for protein identification by peptide mass fingerprinting (PMF) (Egelhofer *et al.* 2000). These algorithms have recently been further developed (Egelhofer *et al.* 2002) and implemented in a computer software. Evaluating and extracting useful information from the large amount of data generated by the combination of 2-DE and MALDI-TOF-MS requires efficient software tools. Software tools have been developed, which allow visualization of information extracted from the lab database on demand, for example displaying the protein identification results graphically onto the 2DE image (fig 2 d). The 2DE-MS technology described here was recently reviewed (Nordhoff *et al.* 2001). Descriptions, laboratory protocols and illustrations are also available in the Internet (www.molgen.mpg.de/~massspec/index/technologies/proteomics/proteomics.html).

Proteome analysis of *Arabidopsis thaliana*

In the final stage of the project, the developed technology and experimental protocols were applied to the proteomic characterization of *A. thaliana*. The resulting collection of standardized 2DE images of *A. thaliana*, annotated with mass spectrometric protein identification data for over 2,900 proteins, is documented in a publication (Giavalisco *et al.* 2005) and serves as a reference for further studies. The data is available to the research community through the GABI Primary Database of the Resource Centre in Berlin at the following URL: https://gabi.rzpd.de/projects/Arabidopsis_Proteomics.

○ P. Giavalisco *et al.* **Extraction of proteins from plant tissues for two-dimensional electrophoresis analysis.** Electrophoresis (2003) 24, 207-216. ○ P. Giavalisco *et al.* **Proteome analysis of *Arabidopsis thaliana* by 2-D polyacrylamide gel electrophoresis and MALDI-TOF mass spectrometry.** Proteomics (2005) in press. ○ J. Gobom *et al.* **Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics.** Analytical Chemistry (2001) 73, 434-438. ○ V. Egelhofer *et al.* **Improvements in protein identification by MALDI-TOF-MS peptide mapping.** Analytical Chemistry (2000) 72, 2741-2750. ○ V. Egelhofer *et al.* **Protein identification by MALDI-TOF-MS peptide mapping: a new strategy.** Analytical Chemistry (2002) 74, 1760-71. ○ E. Nordhoff *et al.* **Large-gel two-dimensional electrophoresis-matrix assisted laser desorption/ionization-time of flight mass spectrometry: an analytical challenge for studying complex protein mixtures.** Electrophoresis (2001) 22, 2844-55.

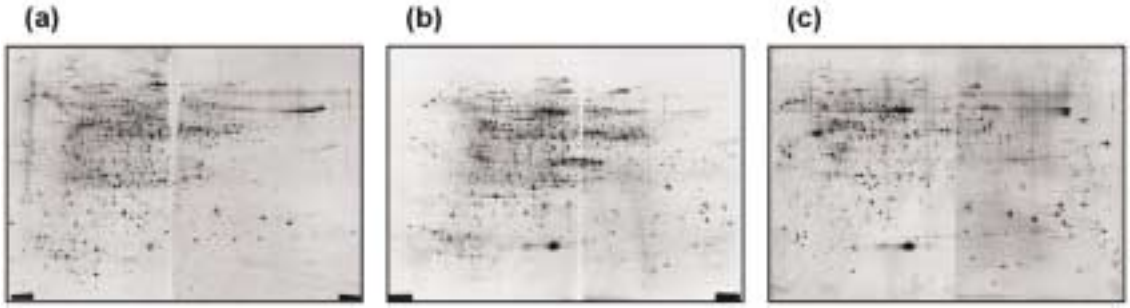


Fig 1: *Proteome of Arabidopsis thaliana stem tissue. The newly developed protein extraction procedure allowed detection of (a) 3,200 protein species in the first fraction and (b) 2,600 in the second fraction, resulting in a total of 4,600 unique protein spots. (c) Only 1,700 spots were detected with the standard extraction protocol based on protein precipitation.*

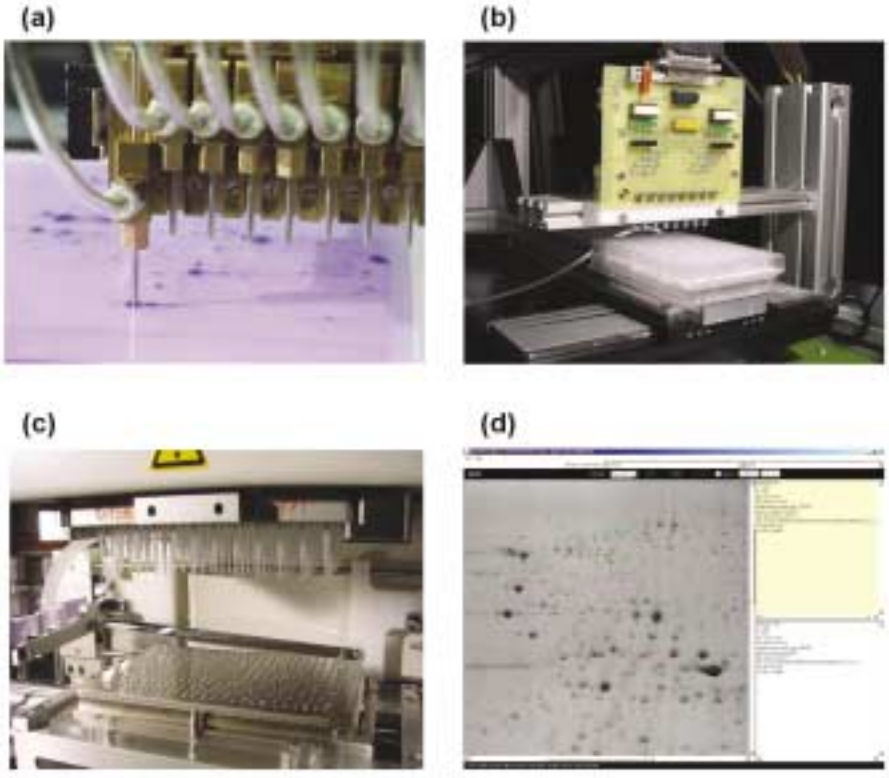


Fig 2: *Technology development: (a) Automatic gel exciser. (b) Solenoid-valve liquid dispenser. (c) Automated MALDI sample preparation. (d) Data analysis and visualization.*

Identification of protein-protein interactions using an automated yeast two hybrid system
(Subproject 3)

Resourcecenter GABI-LAPP

Lukas Bürkle, Marion Amende, Rita Fischer, Eckehard Kuhn, Uli Schneider, Hans Lehrach, Alexander Heyl
Max Planck Institute for Molecular Genetics (MPI-MG), Berlin

Introduction

The completion of the *Arabidopsis thaliana* sequencing project in 2000 has provided the community with a resource that greatly facilitates research efforts. However it also clearly demonstrated that the majority of the predicted open reading frames (ORF) have an unknown function. In Arabidopsis, as in other sequenced genomes (e.g. yeast) research has been focused on a relatively small subset of genes and their proteins, while the largest part of the genes remains uncharted territory in respect to their role in the organism. To change this situation and to unlock the potential of the uncharacterized genes and their respective proteins, high throughput technologies aimed to determine different aspects of protein function have to be developed and applied on a large scale.

One way to characterize the function of proteins is to identify interacting partners. If the function of a partner protein is known, the function of an unknown protein can often be proposed. The development of the yeast two-hybrid (Y2H) system was a decisive step towards the efficient identification of such protein-protein interactions (Fields and Song, 1989). This genetic procedure allows the rapid identification of *in vivo* protein-protein interactions and the simple isolation of corresponding nucleic acid sequences, encoding the interacting partners. The Y2H system has its basis in the reconstitution of a transcription factor in yeast by the interaction of two proteins. In a typical Y2H experiment, a hybrid protein consisting of a DNA-binding domain (DB) and a protein of interest ("bait") is assayed against a ("prey") library of proteins expressed as fusions with a transcriptional activation domain (AD) (Fig 1).

The reconstruction of such an artificial transcription factor in a suitable yeast strain leads to the activation of a reporter system. Within the GABI-LAPP subproject Y2H a yeast strain is used that contains three independent reporter genes to minimize false positives during the screening.

Goals

The main objective of this subproject is to construct a map containing protein-protein interactions of the Arabidopsis proteome by the development of a sustainable, automated high-throughput yeast two-hybrid system. This will contribute significantly to the functional annotation of still unknown genes.

There are two different systems available to establish the high throughput Y2H screening: (i) Co-transformation of the yeast with the bait vector and the prey library, (ii) Transformation of different haploid yeast mating types with either the bait vector or the prey library that will form diploid strains after mating containing both vectors.

i.) Co-transformation is easier to perform but requires high amounts of the prey library and is very difficult to adapt for high-throughput screening. The prey library has to be renewed by several steps of amplification that might lead to reduction in complexity. This method is used to test the libraries for quality.

ii.) The mating approach is ideal for automated high throughput adaptation (Fig 2) because the prey library can be arrayed in 96- or 384-well format. This method is more sustainable because the clones are amplified separately and the complexity is kept at a constant level.

Status

We have chosen a system, which uses high-copy number 2-µ-based shuttle vectors. Both the lexA-DB- and Gal4-AD-fusions are expressed from the constitutive, moderate strength alcohol dehydrogenase promotor (PADH). Each vector has independent selection markers for the yeast. To facilitate the isolation of the prey plasmid from bacterial host, two independent antibiotic marker are used. Both vectors are adapted to the GATEWAY™ system to increase the throughput. The vectors have been specifically designed for our system to be compatible with the yeast strain L40ccU (a and α), which carries three independent reporter genes enabling a more stringent three-phenotype screening. Using the lacZ-reporter gene the bait proteins are tested for autoactivation before entering the screening procedure.

The key of a effective Y2H system is the quality of the prey cDNA libraries. We are using oligo-(dT) primed cDNA libraries based on RNA from seedlings, primary leaves, developing seeds, silliques and mature inflorences. These libraries have been constructed in collaboration with Bernd Korn (RZPD, Heidelberg). They have been cloned in the pENTR1A™ and are therefore compatible with many different downstream applications. The quality of the libraries was determined by restriction analysis and sequenc-

○ B. Kersten *et al.* **Large-scale plant proteomics.** Plant Molecular Biology (2002) 48, 133-141. ○ E.J. Kuhn and L. Bürkle **Functional Genomics – Methodensprung für die Ertrags- und Stressphysiologie: Ein automatisiertes Yeast-two-Hybrid-System für *Arabidopsis thaliana*.** Vortr. Pflanzenzüchtg. (2001) 52, 43-49. ○ S. Fields and O. Song **A novel genetic system to detect protein-protein interactions.** Nature (1989) 340, 245-246. ○ L. Bürkle, S. Meyer, H. Dortay, H. Lehrach and A. Heyl **In vitro recombination cloning of entire cDNA libraries in *Arabidopsis thaliana* and its application to the yeast two hybrid system.** Funct. Integr. Genomics (2005), in press.

ing of 200 random clones per library, demonstrating that all the clones carry an insert in the sense orientation. The libraries were shuttled into the pGAD10 vector, which we adapted to the GATEWAY™ system. This recombination technology greatly facilitates collaborations with other GABI groups using the same system.

To further evaluate our libraries we used the co-transformation approach with different bait proteins, which belong to the cytokinin signal transduction pathway. These tests have demonstrated that our libraries are of good quality. Therefore, we are using bait proteins originating from GABI-LAPP subproject 1 (protein expression) and from numerous collaborators within and also outside the GABI community.

In parallel we plan to pick and array about 200,000 yeast prey clones of our libraries. These will be integrated into our already developed automated yeast two-hybrid mating technology platform. To increase the throughput and the efficiency of the mating system it is advantageous to create pools of prey clones. Our preliminary studies showed that a bait protein can still find its interaction partner within a pool of more than 30 different prey clones. To ensure that even weak interactions are detected within that system we use only pools of 20 prey clones.

Conclusion

Within this project four high quality cDNA libraries from different tissues of *Arabidopsis thaliana* have been created in the GATEWAY™ system. The libraries have been tested in the yeast two-hybrid system using baits from the cytokinin signaling pathway. Published and previously unknown interactions have been detected. Technology for automated yeast two hybrid screens have been developed.

The utilization of the GATEWAY™ technology greatly increases compatibility with other projects within GABI and facilitates our efforts to providing an interaction screening service for the GABI community. Our screening capacity until the end of project funding will be around 80 interaction screens per month. The results of our screens are currently stored in a local database, but will be transferred to GABI Primary Database.

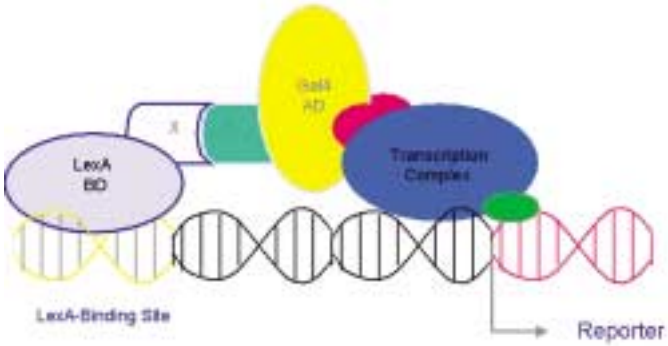


Fig 1: *Scheme of the yeast two hybrid system.*

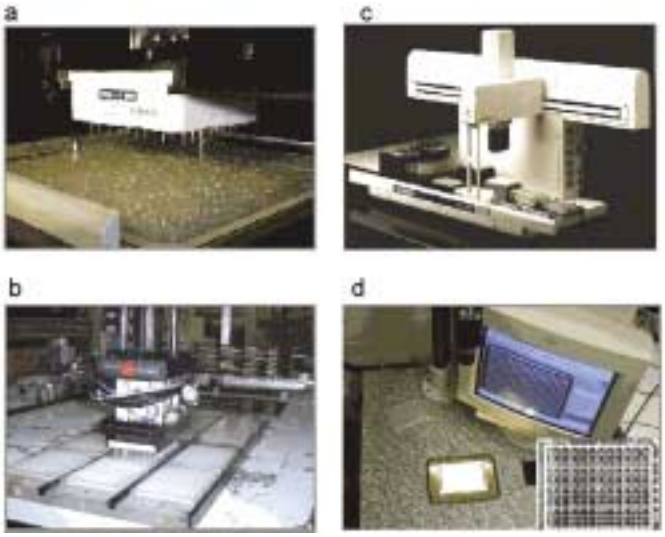


Fig 2: *Automation of the yeast two hybrid system: (a) Robot picking yeast clones (b) Rearraying robot (c) Liquid handling (d) Documentation center.*

RNA Expression Analysis

Resourcecenter
GABI-LAPP

1 Wilfried Nietfeld, 1 Stefanie Albrecht, 1 Holger Eickhoff, 1 Antje Krüger, 1 Lajos Nyarsik, 1 Hans Lehrach
2 CATMA consortium
1 Max Planck Institute for Molecular Genetics (MPI-MG), Berlin
2 CATMA partner see: www.catma.org

Summary
The generation and production of full-genome *Arabidopsis thaliana* DNA chips containing gene specific tags (GSTs) of all Arabidopsis genes was the main goal of our research in this project. For the generation the GSTs the necessary primer design, the generation of GSTs by PCR and establishing of a database was done in collaboration within the CATMA (Complete Arabidopsis Transcript MicroArray) consortium. During this report period we successfully continued on the generation of GSTs for all Arabidopsis genes, the optimisation of the microarray production technology as well as optimising protocols for RNA isolation, labeling and hybridisation. We have produced the first batches of CATMA chips containing about 21,000 GSTs and performed initial test-hybridisation experiments for quality control. Very recently we also could update the first CATMA collection by adding about 3,500 new GSTs to the collection. Thus our current CATMA collection version 2 consists of about 24,500 GSTs, which in total covers about 80 percent of the whole *Arabidopsis thaliana* transcriptome.

Annotation of gene-specific tags
The annotation (using the EuGène software package) for all chromosomes is completed and comparative analysis with the AGI annotation is done. In more detail, EuGène (29,787 genes) predicts over 4,000 genes more than AGI (25,470). Although the difference is partly due to the availability of novel cDNA or ESTs published in the meantime, EuGène identifies more genes not seen by AGI than the reverse, and EuGène splits more genes than AGI. Out of those, 12,445 correspond to the subset of GST designed from predicted genes. A new analysis is underway to match the whole GST set with AGI defined genes to which are added putative 3'UTR not included in the AGI annotation. It will provide a more better – although not perfect – evaluation of the portion of AGI genes tagged with CATMA GSTs. The annotation is continuously updated in Pierre Rouzé's group (VIB-PSB, Genth, Belgium) and has been recently compared to the Ricken EST set (>14,000 clones) and Ricken full-length cDNA set (about 9,000 clones).

GST database
The CATMA GST database was continuously developed at John Innes Centre (Norwich, UK) and is publicly accessible since 22nd June 2002 (http://www.catma.org/Database/login.html). The database has integrated GSTs with MIPS/TIGR gene information and also a flag indicate for faulty GSTs in the database is included. A manuscript, which describes the CATMA database, has been recently published (Crowe *et al.* 2003).

Generation of GSTs
The first round PCR of 21,120 GST using gene specific primer pairs was finished in March/April 2002 including all corrections (CATMA version 1). Aliquots of this PCR reactions have been distributed to all participating CATMA partners. The PCR performance for all labs was good and consistent, indicating a very good GST selection and primer prediction. The overall success rate was over 90%. In May/June 2002 our group started the second round of amplification of all 21,120 GSTs with tag-specific primer pairs. After having optimized the conditions for amplification we got good results in the yield of PCR products even after purification. Purification of small PCR fragments is always related to a high loss of those fragments. Therefore we had to increase the reactions volume, after purification we obtained an average yield of about 1.5-2.5 microgram of DNA per reaction as measured randomly with the Agilent Bioanalyser 2100. Recently about 3,500 additional first round GST PCRs have been distributed within the CATMA consortium leading to a collection of about 24,500 CATMA GSTs in total (CATMA version 2). A new round of the second round PCRs for all 24,500 GSTs has been done. The purified GSTs are currently spotted in collaboration with the MPI of Molecular Plant Physiology (Potsdam-Golm).

Array production
Optimising the DNA microarrayer performance for high-density spotting (> 20,000 spots per array) was done in collaboration with the group of Claus Hultschig (MPI for Molecular Genetics). Conditions for high throughput printing of microarrays in high-density format were optimised and established. Currently we are able to spot more than 22,000 spot per glass chip on more than 80 slides per batch accomplish the demands for the GABI community. First batches of CATMA microarrays have been produced and used in complex cDNA hybridisation experiments for quality control. An example of such a scan-file obtained after background hybridisation is shown in false colour in figure 1. The complete CATMA set was spotted onto glass slides and randomly selected arrays were tested in a so-called background hybridisation experiment using Cy3 or Cy5 labelled suitable primer. Other CATMA arrays were used in complex cDNA hybridisation experiments and an example of such an obtained scan-file is shown in figure 2. *A. thaliana* leaf and flower mRNA have been used as templates for labelling with Cy3 (flower mRNA) and Cy5 (leaf mRNA) in a reverse transcription reaction. The CATMA microarrays, which contained the about 21,000 GSTs were hybridised with the purified Cy3 and Cy5 mixed samples followed by washing and scanning (Affymetrix 428 Array Scanner). These results indicate that we have achieved a very high spotting quality level. Nevertheless additional experiments and repetitions are necessary as well as a careful statistical analysis has to be performed to proof the high quality of the CATMA arrays. Arabidopsis tissue sample were kindly provided by Thomas Altmann and Dirk Hincha (MPI of Molecular Plant Physiology, Potsdam-Golm).

○ The Arabidopsis Genome Initiative **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** Nature (2000) 408, 796 - 815. ○ N. Pavy *et al.* **Evaluation of gene prediction software using a genomic data set: application to *Arabidopsis thaliana* sequences.** Bioinformatics (1999) 15, 887-899. ○ T. Schieix, A. Moisan and P Rouzé. **EuGène: an eukaryotic gene finder that combines several sources of evidence.** In "JOBIM 2000, LNCS 2066" (2001), O. Gascuel, M.F. Sagot (Eds.), pp 111-125. ○ Thareau *et al.* **Automatic design of gene specific tags for transcriptome studies.** Proc. Of JOBIM'2001 (Journees Ouvertes Biologie Informatique Mathematiques). Toulouse, France (2001). ○ M. Crowe *et al.* **CATMA – A complete Arabidopsis GST database.** Nucleic Acids Res. (2003) Jan 1;31(1):156-8. ○ P. Hilson, I. Small and M. T. Kuiper **European consortia building integrated resources for Arabidopsis functional genomics.** Curr Opin Plant Biol (2003) 6, 426-429. ○ P. Hilson, J. Allemeersch, T. Altmann, S., Avon A. Aubourg, J. Beynon, R.P. Bhalerao, F. Bitton, M. Caboche, B. Cannoot, V. Chardakov, C. Cognet-Holliger, V. Colot, M. Crowe, C. Darimont, S. Durinck, H. Eickhoff, A.F. de Longevialle, E.E. Farmer, M. Grant, M.T. Kuiper, H. Lehrach, C. Leon, A. Leyva, J. Lundeberg, C. Lurin, Y. Moreau, W. Nietfeld, J. Paz-Ares, P. Reymond, P. Rouze, G.Sandberg, M.D. Segura, C. Serizet, A. Tabrett, L. Taconnat, V. Thareau, P. Van Hummelen, S. Vercruysse, M. Vuylsteke, M. Weingartner, P.J. Weisbeek, V. Wirta, F.R. Wittink, M. Zabeau and I. Small **Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications.** Genome Res. (2004) 14(10B), 2176-89.

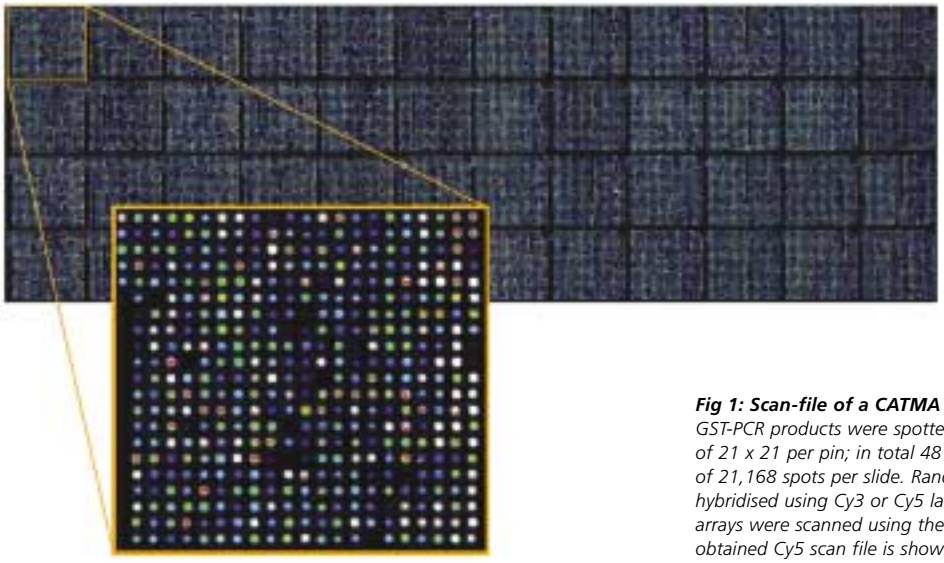


Fig 1: Scan-file of a CATMA chip background hybridisation *Purified CATMA GST-PCR products were spotted onto microarray surfaces by a spotting pattern of 21 x 21 per pin; in total 48 blocks were spotted, resulting in a total number of 21,168 spots per slide. Random samples of the produced CATMA arrays were hybridised using Cy3 or Cy5 labelled suitable primer and washed thereafter. The arrays were scanned using the Affymetrix 428 Array Scanner. An example of an obtained Cy5 scan file is shown in false colour representation.*

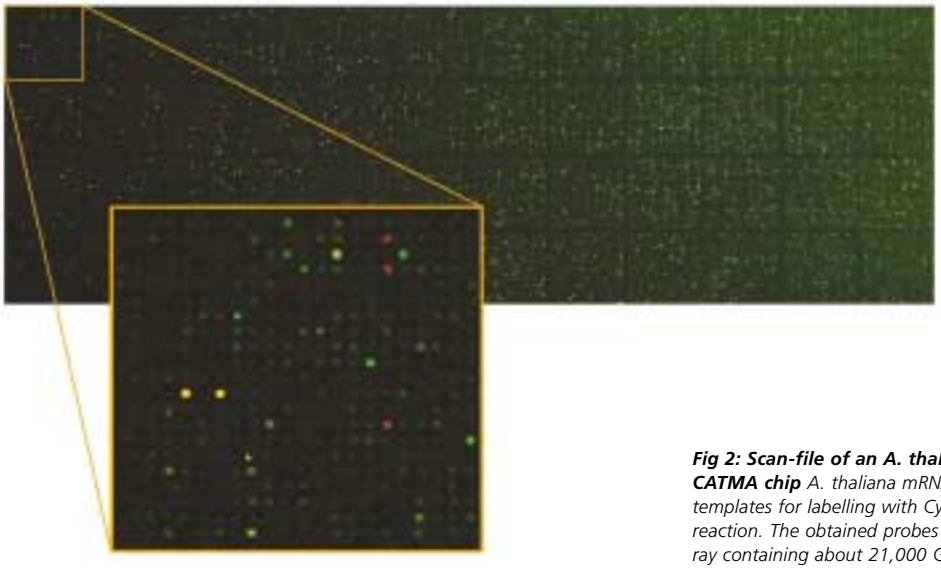


Fig 2: Scan-file of an A. thaliana complex cDNA hybridisation using a CATMA chip *A. thaliana mRNA was isolated from flowers and leafs and used as templates for labelling with Cy3 and Cy5 respectively in a reverse transcription reaction. The obtained probes were purified and hybridised to CATMA microarray containing about 21,000 GSTs.*

Management of germplasm and mapping populations – Construction of a high density transcript map of barley

Resourcecenter
GABI Plant

Manoj Prasad, Nils Stein and Andreas Graner
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

Introduction

One of the focal points of both conventional recombination genetics and modern genome analysis is the correlation between genotypes and phenotypes. New ways to associate phenotypic and structural information emerge from the availability of large numbers of Expressed Sequence Tags (ESTs). Mapping information will provide a clue to correlate the presence of an EST at a specific map position with a trait present in the same chromosomal region. Therefore, one of the overriding goals of systematic genome research is the determination of map coordinates for any given cDNA. This can be achieved by mapping on the physical level, i.e. by placing the cDNAs on YACs or BACs. To integrate this information on the genetic level, this project aims at the generation of a barley transcript map comprising in total 1,200 cDNAs, a figure that corresponds to about one marker per map unit. These cDNAs will serve to anchor BACs, YACs, or contigs of these, on the genetic map and will provide a platform to integrate the results from trait analysis in order to identify candidate genes for distinct traits or quantitative characters.

Development of a transcript map

For the construction of a high density transcript map of barley, three DH mapping populations and a resource of over 110,000 barley ESTs obtained from more than 20 different cDNA libraries (GABI-PLANT C, GABI-SEED) were used. The genetic mapping of the EST-based probes is carried out in the DH populations Igri x Franka (IF: 71 lines), Steptoe x Morex (SM: 150 lines) and Oregon Wolfe Barley Dom x Rec (OWB: 94 lines). All the above three populations have been used previously in other mapping studies and a large number of reference markers will allow to anchor the obtained mapping results to other genetic maps of barley and other Triticeae.

So far, more than 1,350 ESTs were screened for polymorphisms among the parental genotypes by Southern analysis of genomic DNA digested with six different enzymes (Bam HI, Hind III, Eco RI, Eco RV, Xba I and Dra I). Out of these, 1,113 (82.0%) produced satisfying hybridisation signals. A total of 274 probes was polymorphic in IF (24.6%), 442 (39.7%) in SM and 510 (45.8%) in OWB, respectively. Considering all the three mapping populations, 695 probes detected a polymorphism (62.0%). Out of 695 polymorphic probes, 485 ESTs that detect 506 loci have been mapped (Tab 1). On each of the 7 barley chromosomes 48 to 96 loci were identified, respectively. In addition to the RFLP mapping efforts the barley ESTs have been scanned for the presence of simple sequence repeats (SSR) that allow for the generation of EST-based SSR-markers (Thiel *et al.* 2003). 191 of these markers were developed and included together with RFLP- and SNP-markers (GABI-DIVERSITY C) into a consensus barley transcript map comprising almost 950 loci (Fig 1).

The initial source of candidates for marker development and mapping of single or low copy number genes in either one of the three mapping populations was initially a 3,737 unigene set derived from cluster analysis of 13,109 ESTs (Michalek *et al.* 2002). Currently this dataset has increased to roughly 20,000 unigenes from 110,981 ESTs (GABI-PLANT C, Zhang *et al.* 2004). Within this project, cDNA clones were used as RFLP-probes after assessing their putative function via BLASTX sequence comparison or via annotation to the MIPS catalogue of functional classification (GABI-PLANT C, GABI-INFO). In addition, mapping has been performed as a service for collaborating partners like GABI-SEED (164 EST-based probes) and GABI1b-TRANSPOSON MUTATED BARLEY. Seeds of the individuals of the mapping populations have been supplied (GABI1b-TRANSPOSON MUTATED BARLEY, GABI-DIVERSITY) as well as parent screen membranes (GABI1b-TRANSPOSON MUTATED BARLEY).

Comparative EST mapping in barley & rice

An initial attempt has been made to predict the probability that a given barley EST exhibits an orthologous map position (genetic or physical) or marker synteny between barley and rice. BLASTN analysis of the non-redundant mapped barley ESTs (849 EST) against the rice EST database yielded in a total number of 567 hits with expected value less than e-20 (sequence identity is more than 80%). For 463 of the detected rice genes a genetic map position was available. Out of these, 294 displayed a syntenic relationship to the corresponding barley ESTs. To extrapolate the potential of this procedure in a chromosome-wide (and finally genome-wide) manner our 110,981 barley ESTs were analysed by BLASTN against the sequence of rice chromosome 1 (available from http://rgp.dna.affrc.go.jp/). A total of 1,637 non-redundant unigenes fulfilled the selection criteria. Therefore, altogether about 230 barley ESTs of the current unigene dataset are likely to share a syntenic map position to rice chromosome 1. Based on this data, the direct selection of candidate syntenic markers would allow for the saturation of a target region in barley with cDNA markers.

Development of an RFLP database

The same set of unigenes has been used as a source for RFLP-marker, SNP-marker (GABI-DIVERSITY C) and SSR-marker (IPK) development. To avoid the redundant mapping of the same probe by different marker technologies a local Microsoft Access database was set up and is used extensively for the coordination of the diverse mapping efforts. It has a flexible design and allows for the extension of the source unigene set for marker development after new EST sequences have been included into the clustering procedure. The database stores essential information on ESTs selected for RFLP analysis accumulated during the evaluation procedure of this probe together with the possible mapping results. Currently, the database can be accessed only through the IPK intranet as it is mainly a working database. However, in close communication with the PLA of GABI, public access to the database via the IPK web page is planned to be provided in the long term. In prospect to provide this access, the porting of the database to a Oracle database platform is undergoing.

○ T. Thiel *et al.* **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).** Theor Appl Genet (2003) 106, 411-422. ○ W. Michalek *et al.* **Towards the identification of a unigene set in barley.** Theor. Appl. Gen. (2002) 104, 97-103. ○ R.K. Varshney, T. Thiel, N. Stein, P. Langridge and A. Graner **In silico analysis on frequency and distribution of microsatellites in ests of some cereal species.** Cellular & Molecular Biology Letters (2002) 7, 537-546. ○ R.K. Varshney, R. Sigmund, A. Boerner, V. Korzun, N. Stein, M.E. Sorrells, P. Langridge and A. Graner **Interspecific transferability and comparative mapping of barley est-ssr markers in wheat, rye and rice.** Plant Science (2004), in press. ○ H. Zhang, N. Sreenivasulu, W. Weschke, N. Stein, S. Rudd, V. Radchuk, E. Potokina, U. Scholz, P. Schweizer, U. Zierold, P. Langridge, R.K. Varshney, U. Wobus and A. Graner **Large-scale analysis of the barley transcriptome based on expressed sequence tags.** The Plant Journal (2004) 40, 276-290.

Conclusion

A comprehensive transcript map of the barley genome has been established comprising currently a set of 506 loci detected via EST-based RFLP marker. Together with 200 EST-SSR markers (this project) and 230 barley SNP markers (GABI-DIVERSITY C) developed from the same resource of barley ESTs (GABI-PLANT C) this map includes 950 loci and will contain more than 1,100 after the finishing of the projects. This map represents a new and highly valuable resource for functional and structural genome analysis of a major cereal crop species and it is highly desirable to be significantly extended within the future.

DH-Population	Barley-Chromosome							Total
	1H	2H	3H	4H	5H	6H	7H	
IF	16	13	22	10	11	15	26	113
SM	25	28	27	6	25	14	23	148
OWB	34	55	34	32	38	24	28	245
Total	75	96	83	48	74	53	77	506

Tab 1: *Chromosomal localisation and number of RFLP loci mapped in the mapping populations Igri x Franka (IF), Steptoe x Morex (SM) and Oregon Wolfe Barley DOM x REC (OWB).*

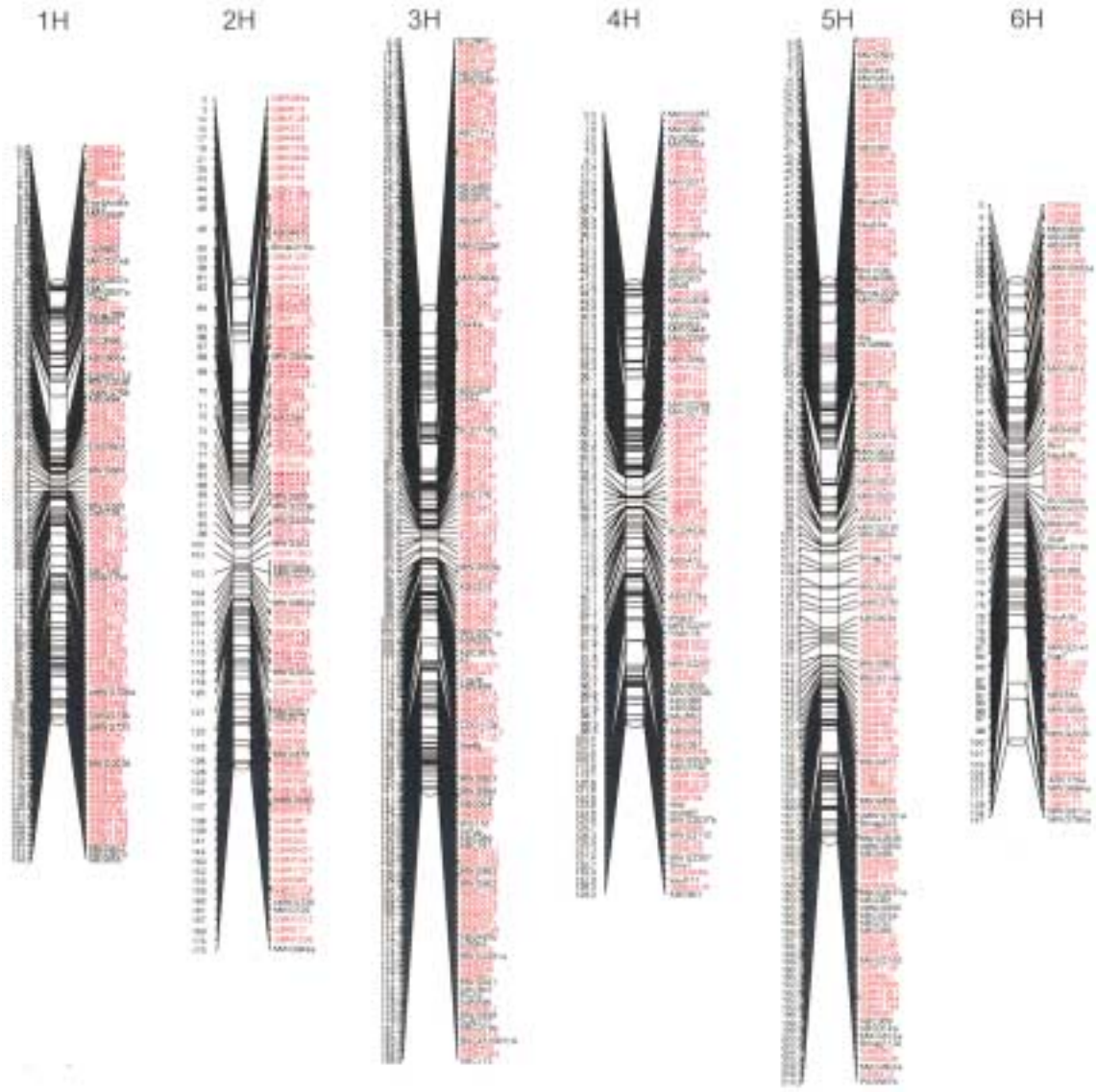


Fig 1: A consensus transcript map of the seven barley chromosomes. *IPK barley ESTs mapped as RFLP (GBR), SSRs (GBS, ESSR) and SNPs (GBS) are highlighted in red.*

A core collection of barley cDNAs and expressed sequence tags (ESTs)

Resourcecenter
GABI Plant

Hangning Zhang, Nils Stein and Andreas Graner
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

Introduction

The major goal of this project is to define a barley unigene set based on large scale cDNA sequencing. ESTs are supposed to be generated from a number of different cDNA libraries with an emphasis on those representing different stages of the barley seed development and germination. The defined unigene set will serve as a resource for marker development regarding the construction of a high density transcript map of barley (GABI Plant A, GABI Diversity C) and for the production of a high-density cDNA macroarray. Such an array will serve as an invaluable tool for studies of global gene expression in seed development and germination (GABI Seed), abiotic stress (GABI Stress) and the characterization of mutants, ecotypes and transgenic plants.

cDNA sequencing, EST generation and data analyses

110,981 ESTs were generated from 22 barley cDNA libraries including 85,859 5'-sequences and 25,122 3'-sequences (tab 1). The 5' sequences allowed for a more comprehensive functional assignment whereas the 3' sequences were of superior value for the development of SNPs and for the differentiation of gene families. Annotation of the ESTs was performed by blastx2 similarity search against the NRPEP database available at HUSAR (Heidelberg Unix Sequence Analysis Resources; http://genius.embnet.dkfz.heidelberg.de). As a result, about 50% of the ESTs hit well-characterized genes, 24% hit putative or predicted genes and 26% did not resolve significant sequence similarity to any known amino acid sequence present in public databases.

Sequence cluster analysis and definition of a barley unigene set

The software package StackPACK v2.1.1 (SANBI, South Africa) was applied for EST cluster analysis: 110,981 barley ESTs were clustered into 14,151 consensi comprising multi-EST sequences and 11,073 singletons, respectively. These were nominated as TUCs (tentative unigene consensi), which should represent unique genes in the barley genome. Besides of the IPK barley EST collection, there were 224,197 barley ESTs present in dbEST/NCBI (Sept. 2002). A BLASTN search revealed that 17.5% of the 25,224 IPK TUCs did not have a significant expect value $\leq$ E-5 at dbEST and thus were assumed to be new genes present in the IPK barley EST resources only. Annotation of the IPK TUCs was conducted by BLASTX comparison to a protein database NRPEP. Under the threshold of expect value $\leq$ E-10, 34.6% TUCs hit well characterized genes, 30.9% hit putative, hypothetical or unknown genes and 34.5% had no significant hits. The 25,224 TUCs were compared to MIPs (Munich Information Center for Protein Sequences, http://mips.gsf.de) gene catalogue for functional categorisation (Funcat). With the criteria expect value $\leq$ E-25, 26.1% of the TUCs could be functionally classified. Apart from the unclassified TUCs, the most abundant genes belong to the categories of "CELLULAR ORGANIZATION" and "METABOLISM".

Ubiquitious vs. library specific gene expression

Cluster analysis of ESTs from different cDNA libraries enabled us to define the ubiquitously expressed genes and more importantly putative library specific genes. Thirty barley TUCs were found consisting of ESTs from all fifteen libraries that are represented by more than 5,000 ESTs each. Apart from three unclassified genes, the functional roles of the remaining may be attributed to metabolism, DNA synthesis, transcription, protein synthesis, cellular organization and transport facilitation. For this group of genes homologs could also be found in rice as well as in *Arabidopsis thaliana* (expect value $\leq$ E-60). These 30 genes may represent for housekeeping function. In contrast, 132 TUCs were found consisting of at least five ESTs exclusively from one specific library and therefore were assumed to indicate library-specificity. Fifty-five of these TUCs corresponded to well-characterized genes. The library specific data may provide a resource for the identification of tissue specific promoters.

Estimating the gene content of barley

In order to achieve a rough estimation of the gene content of barley in comparison to rice, our current set of TUCs was clustered together with 46,154 barley TUCs taken from HarvEST release 0.86 (http://harvest.ucr.edu/, cluster analysis of 258,311 public ESTs) to obtain a most comprehensive barley unigene set comprising 42,073 TUCs. In a next step approximately 1 Mbp barley genomic DNA present at Genbank has been analysed for gene prediction by utilizing the GenScan software (Burge and Karlin 1997). One hundred-fiftyfour (91.6%) of 168 predicted genes with more than 100 amino acid residues could be found in the set of 42,073 barley TUCs. On the other hand, 380 (92.9%) out of 409 non-redundant annotated barley genes from NRPEP could be found in the same barley unigene dataset (TBLASTN, expect value $\leq$ E-5). Both results suggested that the TUCs represent about 92% of all barley genes. Taking account of about 20% redundancy remaining in the barley TUCs due to non-overlapping 5' and 3' ESTs, the total number of barley genes is likely to be higher than 37,000.

Production of a barley cDNA macroarray

A major goal of this project is to provide a unigene cDNA array. Therefore, an intermediate unigene set has been obtained after StackPack cluster analysis of 41,600 EST. Individual cDNA clones representing the current set of tentative unigenes were selected and re-arrayed and the amplification of roughly 10,000 clone inserts utilising standard sequencing primers is currently undergoing. The representation of clones from the individual libraries on the 10,000 unigene array as well as the suggested functional classification of the genes is given in table 1 and figure 1, respectively. The anticipated time of availability of the 10,000 unigene macroarray is April 2003. An intermediate array consisting of almost 6,000 unigenes from 17 libraries has already been produced and is currently under evaluation.

○ W. Michalek *et al.* **Towards the identification of a unigene set in barley**. Theor. Appl. Gen. (2002) 104, 97-103. ○ H. Zhang, N. Sreenivasulu, W. Weschke, N. Stein, S. Rudd, V. Radchuk, E. Potokina, U. Scholz, P. Schweizer, U. Zierold, P. Langridge, R.K. Varshney, U. Wobus and A. Graner **Large-scale analysis of the barley transcriptome based on expressed sequence tags.** The Plant Journal (2004) 40, 276-290. ○ R. Varshney, H. Zhang, E. Potokina, N. Stein, P. Langridge and A. Graner **A simple hybridization-based strategy for the generation of non-redundant est collections – a case study in barley (*Hordeum vulgare* L.).** Plant Science (2004) 167, 629-634.

Since this array was produced to provide a resource for the experiments to be carried out in GABI-SEED, major emphasis has been put on clones selected from seed development- (1,633 clones) and seed germination- (1,127 clones) specific cDNA libraries.

Outlook

Future EST sequencing will be focused on the generation of 3' sequences in order to increase the resolution of the cluster analysis of genes belonging to gene families. Furthermore, additional ESTs will be generated from a cDNA library subtracted against the most current unigene set with the aim of discovering new genes. A bioinformatics tool has been developed for automation of the unigene definition and is currently under testing. It will greatly facilitate the whole procedure of the unigene selection.

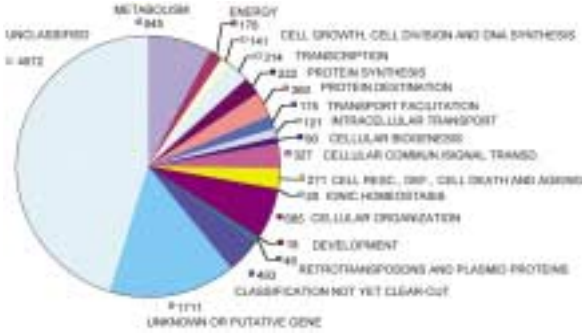


Fig 1: *Functional classification of the clones of the planned 10,000 barley unigene cDNA macroarray based on the MIPS functional catalogue.*

Library	Origin	Cultivars	Tissue	No. clones	No. ESTs	
Unigenes for						
ID			sequenced	5'-seq	3'-seq	total 10,000 array
HK	IPK	Barke etiolated leaves(old)	1031	492	891	1383 406
HW	IPK	Barke roots (old)	3020	2764	2639	5403 1116
HY	IPK	Barke developing caryopsis (old)	3516	3246	3068	6314 1133
HG	GABI Plant	Barke Green leaves	344	328	237	565 66
HE	GABI Plant	Barke etiolated leaves (new)	335	309	241	550 61
HR	GABI Plant	Barke roots (new)	334	302	238	540 69
HP	GABI Plant	Barke epidermis	268	224	239	463 66
HM	GABI Plant	Barke male inflorescences	4350	4228	810	5038 449
HI	GABI Plant	Barke female inflorescences	4256	4150	790	4940 313
HD	GABI Plant	Golden Promise callus (GoldenPromise)	4578	4447	877	5324 593
HX	GABI Plant	Barke apex	4760	4631	937	5568 358
HS	GABI Seed	Barke embryo+scutellum	4822	4547	2533	7080 278
HT	GABI Seed	Barke endosperm (early)	4926	4781	2051	6832 291
HU	GABI Seed	Barke germinating seeds 16h-48h	4631	4401	2499	6900 475
HV	GABI Seed	Barke germinating seeds 48h-96h	4462	4173	2527	6700 406
HZ	GABI Seed	Barke pericarp, 0-7DAP	9773	9745	295	10040 63
HA	GABI Seed	Barke embryosac, 0-7DAP	10044	9748	3015	12763 805
HB	GABI Seed	Barke developing caryopsis (8-15 DAP)	10049	10016	244	10260 171
HF	GABI Seed	Barke developing caryopsis (16-25 DAF)	7418	7410	279	7689 235
HC	GRDC	Sloop coleoptile (3 days old)	399	399	0	399 0
HH	GRDC	Sloop coleoptile (days old)	1072	500	712	1212 194
HO	PGRC	Ingrid BC mlo5 Epidermis (infected by Bgh & Bgt)	5018	5018	0	5018 2535
total			**89406**	**85859**	**25122**	**110981** **10083**

Tab 1: *Summary of cDNA libraries & ESTs available at IPK, and unigene clones selected for spotting on a 10,000 unigene array.*

Development of a DNA array based method for efficient identification of plant genetic resources

Resourcecenter
GABI Plant

Konrad Bachmann, Frank Blattner, Jörg Geistlinger, Dirk Fischer
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

Objective

This project aimed at the in-silicio development of SNPs from publicly available Hordeum-ESTs by means of minimal sequencing efforts. On this basis a microarray analyses platform with high throughput potential was established. On-chip single-base extension was utilized for reproducible SNP-marker patterns in Barley germplasm.

Working Programm

1. In-silicio-detection of SNP-candidates by aligning cDNA sequences.
2. Primer design and resequencing of SNP-regions in 5 cultivars.
Verified SNP markers to be analyzed in barley populations on primer extension microarrays.
3. Development of a SNP microarray platform to test verified SNP loci.
4. Evaluation of the potential of SNP-microarrays for cultivar identification using the Barley Core Collection at the IPK.
5. Evaluation of genetic uniformity in pure breeding lines.
6. Evaluation of parental lines and DH descendants on SNP microarrays.
7. Development of a commercial SNP analysis service as a result of the project.

The identification of plant genetic resources at and below the species level is increasingly becoming a matter of interest. Questions of the physical or intellectual ownership and the geographic origin of plant species, varieties or cultivars may lead to legal disputes which require an accurate identification based on objective criteria. Since all heritable characters are eventually encoded in the genome, character differences will be reflected by sequence polymorphisms. Methods based on the identification and interpretation of sequence polymorphisms may provide a general and reliable technique to identify plants in a universal and standardized way. The success of this approach will depend on the selection and interpretation of relevant sequence polymorphisms. SNPs (single-nucleotide polymorphisms) appear to be the best possible marker system for the development of diagnostic methods and SNPs influencing the protein level will be linked to physiology and morphology to recognize adaptive phenotypes. Moreover our method to identify SNPs on microarrays can be used across species borders for many other applications in plant biology and cultivar identification to protect owners rights (fig 3), namely high-throughput genetic mapping, genomics assisted breeding, point mutation screening (fig 2). These aims are reached and besides plant science and biodiversity analyses our method holds a great potential for medical applications in human and veterinarian genetics.

With these "Polydimensional SNP Microarrays" which enable us to do simultaneous analyses of many loci in many individuals on one chip excluding cross-hybridization between homologous loci in different individuals, we applied for patent rights (see references). Further know-how about verified SNP markers is protected as a company secrets. About 20,000 extensions can be performed in one reaction chamber, which saves lots of expensive materials and chemicals. A pilot study with diagnostic analyses was successful and 48 cultivars were typed at 48 loci (fig 1). In the course of this study we were able to observe wrong gene bank entries when comparing germplasm from different origins. Additionally, known relationships have been verified and the independence of the SNP markers used has been shown through normal distribution of marker correlations (fig 4).

The utilization of self–made oligonucleotide microarrays for routine analysis of point mutations or SNPs was developed for barley as a model organism. Compared to various allele-specific hybridization techniques our newly developed method "on-chip single-base primer extension by DNA polymerase" is superior. Free access to solid phase primers and stringent washing options yield clear results. The method was optimized for accurate multiparallel SNP-typing using fluorescently labeled dideoxynucleotides (ddNTPs) and barley template DNA. SNPs were developed from multiple alignments of expressed sequence tags (ESTs) of different barley accessions and resequenced in at least five divergent lines. After this SNP verification step oligonucleotides for microarraying (solid phase detection primers) are designed with their 3'-ends one base adjacent to identified SNP positions. These oligonucleotides are synthesized with 5'-amino modifications and immobilized on epoxy-silanized glass slides to form covalent bonds for irreversible attachment. DNA chips are manufactured by using a split pin microarrayer (Biorobotics Microgrid II) producing dots 100µm in diameter with a center-to-center distance of 250µm. After PCR-amplification of SNP-loci and production of single-stranded fragments by a specific exonuclease, template DNAs are hybridized and template dependent primer extension with labeled ddNTPs is performed on the chip. A novel multiplexing technique developed by the authors allows to screen numerous individuals at the same locus on a single chip. Beyond that, multiparallel analyses ensure high reproducibility and accuracy by uniform conditions. Tests spanning such a large scale can not be achieved by conventional methods in due time.

The template strand (single-stranded PCR fragment) is removed from the extended and thereby ddNTP labeled oligonucleotide by a hot wash, whereas the covalently bonded oligonucleotide stays on the chip surface. A following laser scan of the microarrays with the Genomic Solutions LS IV Scanner yields valuable results.

Results can be expressed as a binary identification code (DNA bases: red spot =AA, green spot=GG, Yellow Spot=GA) with 48 positions that can potentially discriminate 248 (= 2.8 x 1014) individuals, where allele frequencies are close to 50%. Allele frequencies were evaluated for unequivocal high-throughput typing of barley germplasm, and have been found to be normally distributed around a mean of 0.331 with a maximum of 0.489 Furthermore, we started segregation analyses of SNPs on a mapping population (Steptoe x Morex). Pedigree analyses to identify relationships in cultivated lines and the evaluation of genetic pureness in single-seed versus bulked descendants are successful, too.

○ **Patent application at the German Patent Office**, (DPMA), registered at Munich, Number dp10245145.1, priority founding date: 27th of September 2002.

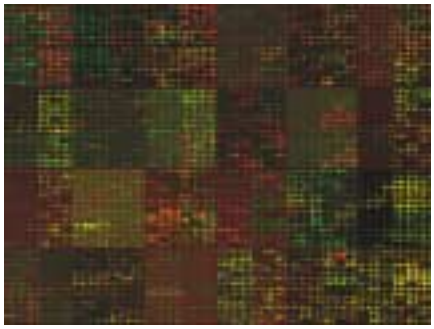


Fig 1: *The image shows analyses of 48 barley accessions at 48 SNP loci using the on-chip primer extension biochip and fluorescent detection as described in the text. These spots have a diameter of about 100µm, a center-to-center distance of 250µm and are produced using 6nl of a spotting solution 40µM in the oligonucleotide.*

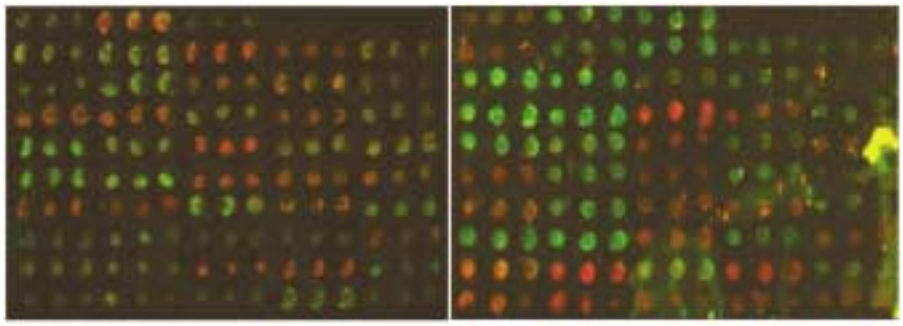


Fig 2: *This picture shows genotyping/identification chips: Two accessions of barley analyzed at 48 SNPs with high allele frequency. Germplasm can be unequivocally genotyped.*

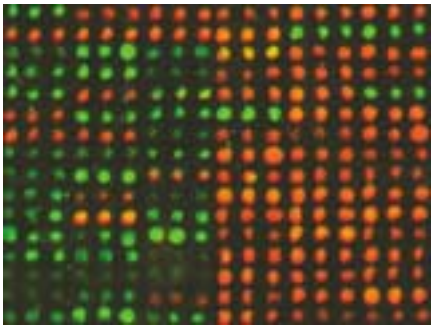


Fig 3: *This picture shows a chip setup with 48 individuals (barley cultivars) analyzed at two loci (separated by dashed line) spotted in threefold redundancy. Applications are biodiversity studies at agro-nomically relevant loci, population dynamics, point mutation screening and potential disease risk assessment chip.*

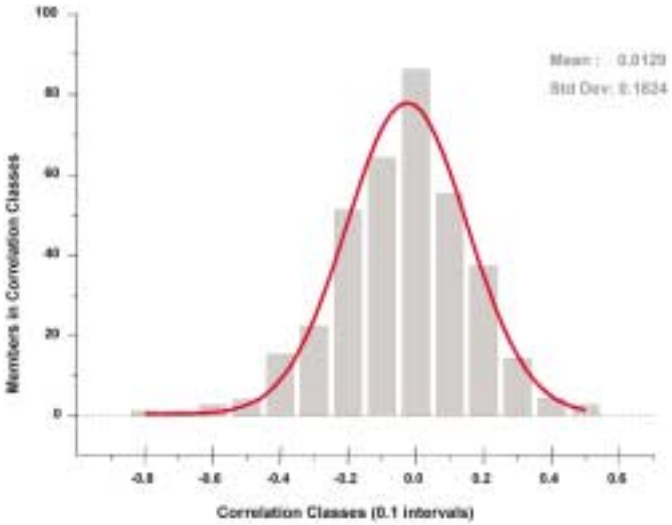


Fig 4: *This figure shows the normal distribution of coefficients from pairwise correlation of all markers from the analysis shown in figure 1. They have been put into 14 classes with intervals of 0.1. This was performed to check for dependencies (due to close linkage) between markers which turned out to occur only by chance.*

Qualitative and Quantitative Improvement for Genetic Transformation of Barley

Genetic transformation contributes significantly to plant genome research. The majority of the scientific approaches within GABI, e.g. those related to stress, seed development as well as the structural and functional analysis of the barley genome, will therefore profit from the improvement of genetic transformation. In this context, it was necessary to develop a transformation technology which is efficient, reproducible, applicable to different barley genotypes, suitable also for large constructs, and implementable to transgenic plants for field release and the consumer market. Three main partners were involved in the project.

Agrobacterium-mediated transformation of immature embryos

A method for transformation of immature embryos was established and further improved. Five transgenic plants were reproducibly obtained per 100 isolated embryos. T-DNA insertion patterns are comparable with what was found in dicotyledonous plants. Plants showed normal phenotypes and were fertile. More than 50% of the transgenic plants had a single copy insertion. By analysis of sexual progenies we confirmed that the transgene inheritance mostly followed the Mendelian rules. A comprehensive screen of breeding material for general tissue culture performance and transformation revealed strong genotype dependence. No other line than the model Golden Promise turned out to be amenable to this transformation method.

Agrobacterium-mediated gene transfer in barley pollen cultures

Immature barley pollen at the microspore or early bicellular stage can deviate from the normal process of pollen formation to undergo embryogenesis. Microspores are available in large amount and posses a high potential to form embryogenic structures and doubled haploid plants. Thanks to the haploid nature of pollen, direct formation of homozygous transgenic lines from pollen cultures can be expected. Microspores pre-cultured for 6-8 days were used for co-culture with Agrobacteria. Transgenic calli were formed under selective conditions and showed expression of ß-glucuronidase or gfp, respectively. Candidate transgenic plantlets developed after 4 weeks of selection in Bialaphos- or Hygromycin-containing liquid medium. More than 2 transgenic plants per donor spike were reproducibly obtained, which corresponds to 15 transgenic plants per 100 embryos. Stable integration of transferred genes was proven by Southern blot analysis. Some 50% of the transgenic plants had a single copy integration of the T-DNA, whereas 21% had 2 or 3 copies. Transgene expression was detected in 93% of the lines. About 69% of the transgenic lines did not set seed, which is likely due to the failing of genome doubling in a portion of the microspore-derived calli. 44% of the fertile transgenic plants turned out to be homozygous in the reporter gene, which was revealed by segregation analysis of the T1-generation. Transgenic plants were obtained from the winter barley cv. 'Igri' as well as from the spring barley cv. 'Gimpel'.

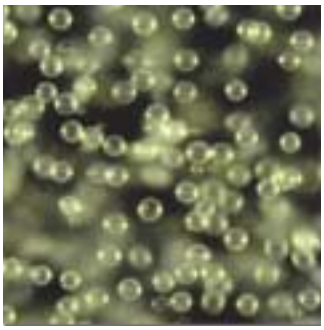


Fig 1a: *Isolated barley microspores.*

Fig 1b: *Gus-expression in pollen-derived calli.*

Elimination of selectable marker genes

Co-transformation by use of two strains of AGL1 carrying different binary vectors was conducted to enable elimination of unlinked selectable marker genes from transgenic plants by segregation in the sexual progeny. The transformation efficiency was 7 transgenic plants per 100 embryos. More than 40% turned out to be co-transformed, i.e. 3 co-transformed plants were obtained per 100 embryos. Experiments on segregation of the unlinked genes are still underway.

Delivery of large intact DNA fragments into barley

Modifications of binary BAC-vectors were undertaken to make this technology applicable to barley. Existing promoters and marker genes were replaced by appropriate ones. Fragments of 30 to 200 kb from different BAC-libraries of Oryza sativa 'Nipponbare', Beta vulgaris and *Arabidopsis thaliana* are used in ongoing experiments on T-DNA transfer into barley.

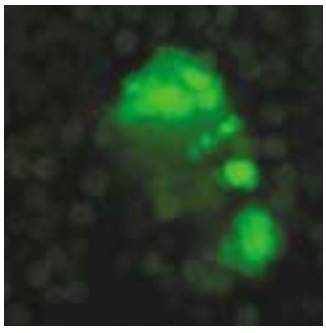


Fig 1c: *gfp-expression in pollen-derived calli.*

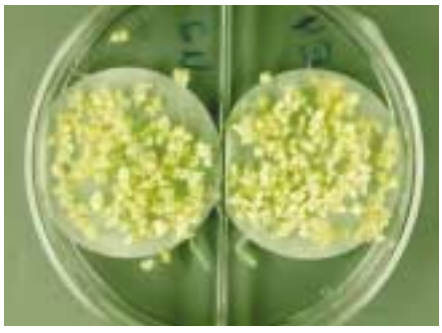


Fig 1d: *Survival of PPT-resistant plants upon spraying.*

Fig 1e: *Shoot regeneration under selective pressure.*

Qualitative and Quantitative Improvement for Genetic Transformation of Barley (Subproject A)

Barley Resources
GABI Trafo

Liliya Serazetdinova, Horst Lörz, Martina Brumm-Scholz
University of Hamburg, Institute of General Botany

Initial situation

Genetic transformation essentially contributes to plant genome research.

Genetic transformation of barley were considered for a long time as difficult because of the extreme recalcitrance to manipulation in vitro. The majority of the transgenic plants were produced by microprojectile-mediated gene transfer in immature embryos (Wan and Lemaux 1994) and isolated microspores (Jähne and Lörz 1995), by electroporation (Gürel and Gözürkirmizi 2000) and transformation of protoplasts (Lazerri *et al.* 1991). All these methods have several limitations and difficulties in general application. In the last years the Agrobacterium-mediated transformation method has been applied for cereals. The first report on Agrobacterium-mediated transformation of cereals was published for rice (Raineri *et al.* 1990). Successful transformation of barley using Agrobacterium as a vector was reported first in 1997 (Tingay *et al.* 1997) and later reports date to 2001-2002 (Fang *et al.* 2002; Patel *et al.* 2001; Stahl *et al.* 2002). In comparison to methods of direct DNA- transfer, Agrobacterium-mediated transformation has several advantages: a low number of integrated gene copies, the integration of transgenes in transcriptionally active regions of the chromosome, the potential to transfer large DNA fragments and considerably lower costs for production of transgenic plants. Thanks to this technology it is possible to produce transgenic plants with stable transgene expression, which are more preferable for an application in genome research programs and high-throughput production of transgenic plants. Although diverse protocols for production of transgenic barley plants by use of Agrobacterium were published, transformation efficiency and reproducibility still has to be improved.

Concept

The objective of the project was to establish a reproducible transformation protocol for barley pollen cultures, that provides efficient production of transgenic plants from different genotypes.

Using different explants as a target for transformation and different transformation techniques, it is possible to further improve the efficiency and reproducibility of gene transfer. In above-mentioned reports immature embryos and embryo-derived callus cultures were used for transformation. Barley pollen culture represents a promising system for genetic transformation, since it demonstrates a high potential for formation of embryogenic structures and regeneration of plants. The methods of isolation, cultivation and regeneration were already established in our group for the winter genotype "Igri" (Jähne and Lörz 1995). In addition, thanks to the haploid nature of the pollen, direct formation of homozygous transgenic lines from pollen cultures can be expected.

Work plan-Milestones

For successful realisation of the project the following tasks must be accomplished:

1. Optimisation of parameters relevant for Agrobacterium-mediated transformation: pollen pre-culture, co-culture with Agrobacteria and elimination of Agrobacteria after the co-culture.
2. Transformation of pollen cultures with different marker genes (uidA, gfp) and selection genes (bar, hpt). Optimisation of the selection protocols for the herbicide BASTA and the antibiotic Hygromycin. Development of appropriate tests for gfp and hpt integration.
3. Transfer of genes of interest.
4. Optimisation of transformation protocol for different Agrobacteria strains and different barley genotypes.
5. Analysis of integration, inheritance and expression stability of transgenes.
6. Design of a PCR- based method for the analysis of the T-DNA border junctions after integration into the plant genome.

Results

Parameters relevant for successful transformation of pollen cultures via Agrobacteria were identified and optimised. The utilisation of these improvements led to the establishment of a reliable system for transformation of pollen cultures, which was used for production of transgenic barley plants integrated different marker and selection genes (gfp, uidA, hyg, bar). Conditions for selection of transgenic plants resistant to herbicide BASTA and antibiotic Hygromycin were optimised. It was obsereved, that hygromycin selection, in contrast to Bialophos selection, which led to regeneration of high number of "false positive" plants, was more stringent in selection, thereby allow the regeneration of high numbers of "true positive" transgenic plants. In addition one target gene, provided by the group of Prof. W. Schäfer (University of Hamburg), coding for Chitosanase from Trichoderma harzianum was transferred to barley. Several transgenic plants harbouring

this gene were recovered, thus we proved that the established transformation system can be applied for genes of interest.

Two Agrobacteria strains, the hypervirulent strain LBA 4404 from Japan Tobacco Company and the non-hypervirulent strain GV3101, were used for transformation. Although we were able to produce transgenic plants using GV3101, it showed lower capacity for gene transfer compared with LBA 4404. On average 1.31 transgenic plants per barley donor spike were produced using GV3101 and 2.26 transgenic plants per donor spike by use of LBA 4404. The differences of these two vector systems were already noticed in earlier stages of the selection procedure, where the pollen cultures transformed with LBA 4404 revealed superior growth and regeneration.

Stable integration of transferred genes was proven by Southern-Blot analysis. In plants analysed so far 58% had a single copy, 21% had 2 or 3 copies, 12% had 3 or 4 copies and 2% carried more then 4 copies of the transgene.

As it was expected, a significant portion of pollen-derived transgenic plants was sterile (69% on average), which represents a general problem in regenerating plants from immature pollen. The seed set in fertile plants varied from 0.1 to 36.6g per transgenic barley plant. The inheritance of transgenes were analysed by activity of the marker and selection genes in the T1 generation. The lines transformed with uidA and bar genes were analysed by histochemical GUS assay and BASTA spray test. In order to analyse the activity of the gfp and hpt genes in transgenic plants, we established activity tests for these genes. Plants with integration of the gfp gene were analysed for fluorescence under UV-light. Root tips turned out to be the best material for fluorescence microscopy.

So far, the T1 of 23 selected transgenic lines with 1 or 2 integrated copies of the transgenes were analysed. The segregation analysis revealed that 5 lines were homozygous and 18 were heterozygous. This result can be explained by the fact, that integration of the transgene into the genomic DNA can occur before as well as after the doubling of the haploid genome during the culture period.

To test the activity of the hpt gene we developed a novel test based on infiltration of leaf segment with a water solution of hygromycin B (300mg/l) or alternatively by germination of seed on MS-medium with 300mg/l hygomycin.

The first prerequisite for application of the established Agrobacterium-mediated transformation method for a wide range of genotypes is optimization of in vitro culture and regeneration protocols for certain genotypes. All above described experiments were first carried out with a model barley genotype "Igri". In order to prove whether the established protocol is applicable for other barley genotypes we tested two spring genotypes "Gimpel" and "Golden Promise". The parameters relevant for isolation, culture and regeneration of microspores from these genotypes were optimised. The genotype "Gimpel" was comparable with the genotype "Igri" regeneration efficiency and was used in transformation. We were able to produce several transgenic plants of this genotype. The genotype "Golden Promise" proved to be not suitable for pollen transformation because of low amount of microspors per processed spike and failure to show the activity of any marker gene after the transformation. The conditions for regeneration and transformation of this genotype should be further improved.

To analyse the T-DNA border junctions after integration into the plant genome we currently optimised the method of inverse PCR.

Utilisation perspectives

The developed method for Agrobacterium-mediated transformation of barley pollen cultures posses hight transformation efficiency and allows direct regeneration of homozygous transgenic lines. This method was developed for two different genotypes and can be applied in projects for functional genomics as well as for biotechnological improvement of barley.

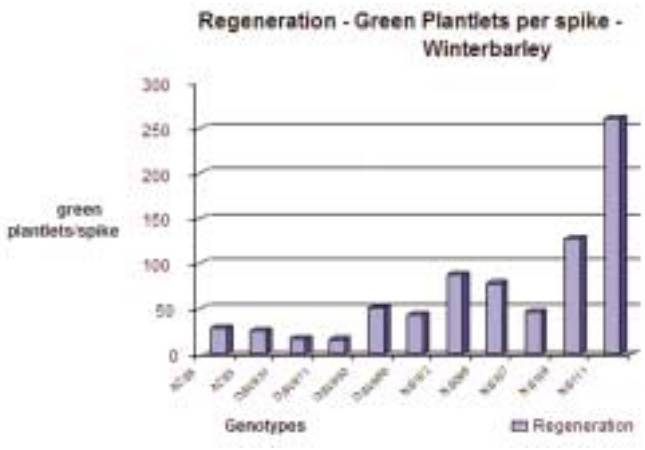


Fig 2: *Response of barley genotypes in isolated microspore culture.*

Qualitative and Quantitative Improvement for Genetic Transformation of Barley (Subproject B)

Resources
GABI Trafo

Jochen Kumlehn, Rajiv Vishnoi, Götz Hensel
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

The available methods to genetically transform barley did not keep pace with the increasing demand for overexpression or knock-down analyses of newly identified genes of scientific and/or economic interest. The development of DNA-array techniques further deteriorated this situation. Immature embryos have been the most frequently used target for genetic transformation of barley. Alternatively, succesful gene transfer was accomplished by use of immature embryo-derived callus, immature pollen cultures or protoplasts. Transgenic barley has been obtained by biolistic as well as Agrobacterium-mediated gene transfer protocols. Biolistic gene transfer has shown to be strongly limited in terms of efficiency. Furthermore, it causes a number of serious problems regarding the quality of the transgenic plants due to degradation of DNA-fragments prior to integration, predominant multi-copy insertions, and random integration into chromosomal regions irrespective of its transcriptive activity. Therefore, the majority of the transgenic plants obtained by this method are eventually not really valuable for further use. By contrast, Agrobacterium-mediated gene transfer has the advantages that relatively low copy numbers will be obtained, that there is hardly any degradation of T-DNA prior to integration into the recipient's genome and that the transgenes are integrated preferably into actively transcribed chromosomal regions. Moreover, co-transformed T-DNAs are often inserted at different loci, thereby enabling elimination of marker genes by segregation after the transgenic plants have been identified.

At the IPK Gatersleben, an efficient and reproducible method for genetic transformation of barley has been developed. The high regeneration capacity of immature embryo cultures was combined with the potential efficiency and high-quality transformation mediated by Agrobacteria. Factors involving preculture medium for Agrobacteria, presence of antibiotics during preculture of Agrobacteria, population density of Agrobacteria during inoculation, orientation of the embryos during preculture, inoculation, co-culture and callus induction, co-culture with contact to filter paper, and gas exchange during co-culture and callus induction by modification of petri dish sealing and/or ventilation were investigated. By use of the resulting protocol, the transformation efficiency repeatedly exceeded 5 transgenic plants per 100 isolated embryos. An analysis of the T-DNA border junctions and the flanking genomic regions of the transgenics revealed insertion patterns which are comparable with what was found upon transformation of dicotyledonous plants, i.e. the sequences at the right border showed significantly higher conservation compared with those at the left border (Fang *et al.* 2002). This finding is in accordance with the commonly accepted model of directed T-DNA insertion by Agrobacteria. Sexual progenies of transgenic plants have been analysed for inheritance of the transgenes as well as for general performance. In the vast majority of cases the genes transferred were inherited as expected according to the Mendelian rules. Transformation of reporter and selectable marker genes did not influence the phenotype and general performance of the model line 'Golden Promise'. A more detailed analysis of the transgenic plants and their progenies based on Southern hybridization is still to be completed.

The method for Agrobacterium-mediated transformation of pollen cultures developed by the Hamburg partner was established in the Gatersleben lab as well. Some modifications turned out to be necessary to get the method working under the particular conditions in the Gatersleben lab. On the other hand, we showed that the method is generally reproducible and applicable in other labs. Moreover, we have obtained some transgenic plants of the spring cultivar 'Gimpel', i.e. we have widened the still limited range of barley genotypes transformable by Agrobacteria.

The genotype-dependency for tissue culture performance and transformation capability of barley is well documented in the literature. Only few genotypes have been shown to be genetically transformable and none of these lines is listed in the German federal cultivar list. Because of this situ-

ation it is necessary to widen the range of barley genotypes amenable to genetic transformation. In particular, actual breeding material has to be screened to enable direct utilization of the potential of gene transfer for crop improvement. We have accomplished a comprehensive screen of actual breeding material for their potential in immature embryo culture and Agrobacterium-mediated transformation to identify lines showing both reasonable regeneration capacity in immature embryo culture and amenability to Agrobacterium-mediated transformation. In these experiments, lines were identified which generated even more shoots from immature embryos than the responsive model 'Golden Promise'. Unfortunately, thus far we were not able to produce transformed plants from any of these lines.

The T-DNA insert size appears to be limited to some 20 kb when common binary vectors are used. Hamilton and coworkers showed that up to 150 kb inserts can be transferred to tobacco by use of binary BAC-like vectors. The transfer of large DNA fragments is extremely valuable for positional cloning approaches as well as for pathway engineering strategies. To facilitate map-based cloning approaches or pathway engineering, BAC-like binary vectors have to be tested for their potential to enable the transfer of large T-DNAs to barley. In the Gaterleben lab, immature embryos have been transformed with three different constructs based upon the vector pBIBAC2. The constructs contain inserts of 20 to 30 kb. The vectors were transferred to the Agrobacteria strains AGL1 and GV3101, respectively. The transformed cultures are currently undergoing regeneration under selective conditions. Binary BAC vectors formerly used for dicotyledonous plant species have been modified to be tested for the transfer of large DNA fragments to monocots, namely by use of appropriate promoters and selectable marker genes. The use of unknown DNA-fragments of tobacco as T-DNA insert, as was initially planned and conducted, has led to serious problems in vector stability in E. coli. To tackle the unexpected problems encountered, we were trying to clone 30 to 200 kb fragments from different BAC-libraries (*Oryza sativa* 'Nipponbare', *Beta vulgaris*) into pBIBAC, pYLTAC7 and pYLTAC17. As an alternative approach, eight different constructs of pYLTAC7 carrying inserts from an *Arabidopsis thaliana* BAC-library (insert sizes of 25 to 115 kb, provided by the JIC Norwich) have been transfered into the Agrobacterium strains AGL1 and LBA4404 to be currently used for barley transformation experiments.

To make the production of marker-free transgenic plants feasible, co-transformation of separate binary vectors carrying different genes has to be evaluated with regard to the potential to eliminate selectable marker genes by segregation in following generations. Inoculation of immature embryos with a mixture of two different stocks of AGL1 which contain pYF133 (hpt under control of the 35S promotor and gfp under control of the ubiquitin promoter) and pYF134 (hpt under control of the 35S promotor and gus-Intron under control of the ubiquitin promoter), respectively, succeeded in formation of plants co-transformed with both T-DNAs. Out of about 200 immature embryos, we obtained 14 independent plants which were transgenic for the hpt gene. Six plants out of these 14 turned out to be co-transformed as was revealed by sequence-specific amplification of gfp as well as gus-Intron. The introduced genes were stably inherited to the T1 generation. A detailed segregation analysis of the T2 generation is yet to be performed.

The methods developed in the scope of the project will have a significant impact on further research on functional genomics as well as for crop improvement in barley. Furthermore, the successful establishment of two barley transformation systems along with all the experience made will facilitate the development of a reproducible method to genetically transform wheat by means of Agrobacteria which constitutes a big challenge yet to be resolved.

Qualitative and Quantitative Improvement for Genetic Transformation of Barley (Subproject C)

Resources
GABI Trafo

Silke Mühlenbeck, Jens Weyen
Saaten-Union Biotech-Laboratory (Saaten-Union Resistenzlabor), Leopoldshöhe

Introduction

Goal of the project was the establishment of a efficient and cheap genetic transformation procedure of barley microspores by the agrobacterium approach. Spring and winter barley genotypes were used in the experiments. It was not the goal to improve the genetic transformation of model genotypes as 'Igri' or 'Golden Promise', but high yielding genotypes with superior agronomic characters were used in our experiments. F1-donor-plants and varieties were chosen for experiments. 'Igri' was used as a control variety with extremely high androgenic response. In a further step it was planned to establish marker gene free transformation technologies. This approach was not yet started due to technical problems after co-culture of microspores and agrobacteria. The established barley microspore culture will have also an positive influence on similar experiments in wheat, triticale, rye and oat.

Results

The isolated barley microspore culture could be improved significantly by modification of isolation technology, media composition and donor plant growth conditions. All three criteria could be optimized. Therefore the isolated barley microspore culture is now a routine application in barley breeding programmes. Several hundreds of genotypes of spring and winter barley (2 and 6-rowed, fodder and malting types) were tested and the level of androgenic response could be increased by application of different stress treatments. Isolated microspore culture is now preferred against anther culture. In some series of genotypes the level of response is even so high that it is impossible to differentiate between low and high responsive genotypes. One to three isolations are necessary to produce several hundreds of plantlets. Fortunately the spontaneous doubling of these plantlets is round about 70%. No chromosome doubling induced by toxic chemicals is necessary. Nevertheless when genetic transformation is performed the rate of chromosome doubling may decrease and colchicine application may become an effective tool for generating a higher rate of fertile transgenic plants. In opposite to our partner in Hamburg we optimized the induction of immature pollen cells on solidified media. In experiments with liquid induction media it could be shown that the induction of modern varieties and F1-genotypes was significantly higher on solid media. 'Igri' showed in our hands also a better induction and following regeneration on solid medium. For the co-culture and the time after the co-culture liquid media are preferred due to better contact and distribution of media components to the plant tissue and the agrobacteria colonies.

While in winter barley up to 260 green plantlets per harvested spike could be produced in spring barley 37,1 plantlets per spike were possible to reach. In spring barley spikes the average number of immature pollen is of course lower so that the regeneration rate in spring barley may be similar to winter barley on the basis of 100 microspores. Due to workload no exact analysis of the number of microspores per spike was done.

Different induction media were tested (Hunter, MSO20) and significant differences were observed.

Co-cultivation of agrobacteria was performed simultaneously with the model genotype 'Igri'. The agrobacteria strain AGL1 (TINGAY *et al.* 1997) and a second agrobacterium strain was becoming available for us (GV3101, without and with the binary vector pUAGB7) and we used both strains for co-culture experiments of microspore cultures of 'Igri'. After some days of co-culture it was tried to remove the bacteria or even to stop the growth of agrobacteria and this experiments were performed during the hole reporting period.

Unfortunately it was not possible to remove the bacteria completely from the growing microspores which finally died in all experiments. Different chemical agents (calcium, glutamine, potassium hydrogen buffers, MES buffer) were applicated to decrease the concentration of bacteria and to protect the plant cells against aggregation. Currently we try to get access to so called "auxotroph" agrobacterium strains, which can be easily controlled by the concentration of essential amino acids histidine and arginine in the media.

○ Y. Fang , C. Akula and F. Altpeter, JIPlant Physiol (2002) 159: 1131-1138 ○ F. Gürel, N. Gözürkirmizi, Plant Cell Rep (2000) 19, 787-791. ○ A. Jähne, H. Lörz Plant Science (1995)109, 1-12. ○ Lazzeri *et al.* Theor Appl genet (1991) 81: 437-444. ○ D.M. Raineri, P. Boritto, M. P. Gordon, E.W. Nester, Bio/Technology (1990) 8: 33-37. ○ Stahl *et al.* PNAS (2002) 99:2146-2151. ○ S. Tingay, D. McElroy, R. Kalla, S. Fieg, M. Wang, S. Thornton and R. Brettell, The Plant Journal (1997) 11(6): 1369-1376. ○ Y. Wan, P. G. Lemaux, J Plant Physiol (1994) 104, 37-48.

Conclusion and future prospects

Isolated microspore culture is now a routine application in spring and winter barley breeding. Breeding progress is significantly accelerated due to the characteristics of doubled haploid plant material.

Genetic transformation of isolated microspores was becoming possible by the partnership and will be improved in the near future. Although the transformation procedure is still not functional in the laboratory of the industrial partner we are quite optimistic to establish it through the close interaction with our academic partners. Nevertheless, genetic transformation of microspores will be a useful application only for specialized laboratories due to the characteristics and sensitivity of the in vitro cultures. While the project was running, it was becoming evident, that all steps of the method and all materials must be harmonized. That is very important for the different agrobacteria strains, because they are very sensitive to different environmental conditions, including pH value, antibiotics, amino acids and many more. Unfortunately it was not possible for us to receive the agrobacteria strains which were used by our partners due to high licensing costs. New negotiations were started in winter 02/03 and we hope to get access to the desired strains while the project is running.

It seems also clear that DH-production and genetic transformation experiments will base on totally different in vitro technologies. The main difference will be the application of solid media in DH-production and of liquid culture media in genetic transformation experiments. Therefore the culture in liquid culture media must be optimized in our laboratory.

Due to this special characteristics of the technology the genetic transformation of immature embryos seems to be the easier method, but the success of that method is also dependent from many circumstances.

Unfortunately no barley accession was identified so far which shows a good androgenic response and a good regeneration in the immature embryo method. Therefore we could not compare both methods.

Transposon-Mutated barley – A Powerful Tool for Genome Analysis

Thomas Koprek
Max Planck Institute for Plant Breeding Research, (MPIZ) Cologne

Scientific goals

Insertional mutagenesis has proven to be a very efficient tool in plant genomics. There are several populations available in *Arabidopsis thaliana* in which the genome is saturated with T-DNA or transposon tags. Due to the much larger genome of barley, saturation mutagenesis in this important model system would require unreasonable resources. Therefore, the scientific goal of the proposed project is to develop a transposon based mutagenesis system in barley that provides in defined locations a high degree of saturation with inserts. The objective of the project is to establish a versatile system for insertional mutagenesis using the maize transposons Activator and Dissociation (*Ac*/*Ds*) that allows the application of different forward and reverse genetics strategies to study gene function, to create knock-out mutants, to isolate genes and that also contributes to mapping programmes.

The concept of the project is based on the generation of numerous independent transgenic barley plants carrying single copy non-autonomous elements (*Ds*). After mapping of *Ds* insertion sites the element will be activated by crossing the *Ds*-carrying plant with *Ac*Transposase expressing plants. Due to the transpositional behaviour of *Ac*/*Ds*, the *Ds* element will preferentially reinsert into nearby locations. This will create locally a high degree of saturation with inserts. The resulting plants can efficiently be used for different forward and reverse genetic screens and insertional mutants can be stabilized by out crossing of the *Ac*Transposase gene.

Working programme and progress

During the first funding period of the project the generation of new single copy *Ds* containing plants was the focus of the work. Two different approaches were used to increase the number of new single copy lines.

In the first approach already existing lines containing single copy *Ds* elements were crossed with *Ac*Transposase expressing plants, which were generated in earlier experiments. *Ac*Transposase activated the non-autonomous *Ds* element and lead to transpositions of *Ds*.

Resulting plants of the F2 generation were analyzed for transpositions of *Ds*. Plants carrying only one new and independent insertion site were selfed. If the transposed *Ds* elements prove to be stably transmitted to the next generation these plants will be integrated in the collection of single copy *Ds* plant lines. In addition to the already existing 43 single copy *Ds* lines 39 more plants carrying a single transposed *Ds* have been identified. This increased the number to a total of 82 single copy *Ds* lines.

In the second approach new transgenic plants were generated using *Agrobacterium* mediated transformation of barley immature embryos. Earlier studies have shown that our *Agrobacterium* mediated transformation system produces a high proportion of stable and fertile single copy transgenic plants.

Two new binary vectors were made for *Agrobacterium* mediated transformation. Vector pWBV-Ubi-Ds-bar carries a modified 650 bp *Ds* element containing the terminal inverted repeats of the transposon and *Ac*Tranposase binding sites as an insert between the Ubiquitin promoter and first intron from maize and the bar selectable marker gene. This will facilitate the identification of transposition events after activation of *Ds* by applying the herbicide BASTA to the plants. Two rare-cutting I-SceI restriction sites were added (one in the plasmid backbone the other one within the *Ds* element) to facilitate the determination of the physical distance of *Ds* transpositions. This information will eventually help to make better estimates about the efficiency of the system. Vector pWBV-Ac contains the full length *Ac*Transposase element and its putative promoter.

Nine independent transformation experiments have been carried out with the *Ds* containing plasmid. The transformation frequencies were on average 23% (tab 1). In total 240 independent single copy *Ds* containing lines have been generated by *Agrobacterium* mediated transformation and have been transferred to the greenhouse. Fifty-six of these carry the new vector pWBV-Ubi-Ds-bar. At present the number of available single copy lines is 322. Large numbers of independent transgenic lines are currently being analyzed for the copy number of *Ds* elements or are still in tissue culture.

Transformations with the binary vector pWBV-Ac started later and transgenic callus lines are still in the regeneration process. Stable transgenic lines which express functional *Ac*Transposase will be used in later experiments to activate *Ds* elements.

Simultaneously, mapping of already existing single copy *Ds* containing lines has started. The method of choice for the mapping of independent *Ds* insertion events is RFLP mapping. RFLP mapping can easily be performed by using *Ds* flanking genomic DNA as hybridization probes in well characterized mapping populations. Ten of these *Ds* flanks were tested for polymorphisms in different mapping populations. Eight *Ds* flanking sequences proved to be polymorphic in the mapping populations Chebec X Harrington und Halcyon X Sloop and their genetic position was determined by P. Langridge (University of Adelaide, Australia). After the feasibility of this mapping procedure had successfully been tested more *Ds* flanking regions were isolated and analyzed in the three mapping populations Igri X Franka, Steptoe X Morex and Oregon Wolfe Barley (A. Graner, N. Stein and M. Prasad, IPK Gatersleben). Isolated genomic borders which were polymorphic in at least one of these mapping populations were used as probes for

RFLP mapping. 38 *Ds* insertion sites have been mapped so far in various mapping populations. Because the RFLP mapping procedure turned out to be very time consuming and allows the mapping of only a part of the single copy *Ds* elements other mapping methods have to be taken into account and tested. Mapping of *Ds* insertion sites will have priority during the remaining two years of the project.

In order to make a first test of the targeted gene tagging strategy, mapped *Ds* elements close to morphological markers or genes of interest have been activated. The corresponding *Ds* containing lines have been crossed with plants containing the *Ac*Transposase gene either under control of the Ubiquitin promoter from maize or the putative *Ac* promoter. The resulting F1 generation has been planted and progeny of these will be analyzed for insertional mutants.

The increasing number of single copy insertion sites allows the systematic isolation and sequencing of *Ds* flanking regions. The resulting data of tagged and mapped insertion sites will be put together in a data base and made available to other GABI participants.

# of experiment (# of embryos)	Transgenic lines	Transformation frequency (%)	Fertile transgenic lines	Fertile copy lines single
1 (250)	50	20	49	12
2 (280)	71	25	71	21
3 (350)	85	24	84	24
4 (325)	49	15	49	20
5 (200)	66	33	66	23
6 (200)	48	24	48	13
7 (350)	70	20	70	19
8 (300)	78	26	78	22
9 (220)	52	24	52	15
10 (400)	68	17	66	19
11 (325)	8	2	8	2
12 (445)	73	16	71	34
13 (350)	51	15	50	16
Total (3995)	**769**	**19.2**	**762**	**240**

Tab 1: *Transformations were done with the binary vector pWBV-Ubi-Ds-bar in Hordeum vulgare L. cv Golden Promise.*

Bioinformatic Resourcecenter for Plant Genome Research

Bioinformatics
GABI-Info

Hans-Werner Mewes, Klaus Mayer
National Research Center for Environment and Health (GSF), Institute for Bioinformatics, Neuherberg

Objectives

GABI-Info's aim is the collaboration and interaction with experimental groups within GABI and to integrate GABI generated data on a genome-wide scale to enable the complex analysis on heterogenous datasets and to summarize the analysis results in adequate comprehensible representations. A further goal of GABI-Info is to develop bioinformatics resources for plant genome data and to provide w3 accessible information. This objective comprises maintainance of existing resources, development of bioinformatic tools and applications as well as structuring, analysing and integrating the drastically increasing amount of genome associated data available from a wide range of different plant species.

Genome backbones

The MIPS *Arabidopsis thaliana* database (MAtDB) has been internationally accepted as a valuable resource for genomics and Arabidopsis research. The functionality of the web access is continually expanded, and the utility of the data is reflected in the continuous increase in the number of web accesses to MAtDB. In the past year, the daily average has risen to over 20000 hits during weekdays, accumulating to approximately 7,000,000 accesses per year. The requests originate worldwide and indicate a widespread visibility of the GABI project.

The Arabidopsis genome sequence and a first analysis has been published in late 2000. Although the published Arabidopsis analysis data are generally considered to be of superior quality in comparison to other annotated higher eucaryotic genomes, the sequence analysis interpretation of the genome can not be considered as perfect or final by any means. Internationally huge efforts are carried out to functionally characterize the individual genes and large scale full length cDNA projects. Large amounts of experimental data related functional information flow into sequence databases.

To provide information to the scientific community that is valuable for the design of new experiments as well as for the interpretation of complex data, it is of upmost importance to keep genome databases up-to-date to interpret large scale experimental data in correlation to the genomic backbone containing the latest information available. To fulfil this task we established a tool which on a regular basis scans public and, if specified, proprietary databases for new entries, evaluates for new information content and automatically integrates these into MAtDB. Beside the data integration of exon and intron coordinates information on transcription start and end as well as literature references are components within this tool. Large scale combination of genomic and experimental data are the fundamental resource for any complex genome focused plant genomics study as well as for any attempts to train intelligent systems on the recognition of biologically significant patterns and genome encoded specificities. The rich experimental resources available for *Arabidopsis thaliana* allow these developments which then can also be applied and transferred to other plants resulting in higher quality analysis standards.

Integration of functional genomic data

To a large extent the function of individual genes within Arabidopsis is not yet known. A widely applied technique to elucidate the function of individual genes is to apply reverse genetic approaches such as transposon or T-DNA tagging and analyse lines harboring knock-outs for individual genes for apparent defects which point towards the functional role of individual genes. All flanking sequence tags (FST) generated by GABI-KAT have been integrated into MAtDB, insertion sites are graphically displayed in the context of gene structure and direct electronic links to GABI-KAT for line ordering are provided. As these efforts are embedded in an international environment, MAtDB is aiming to integrate all transposon/T-DNA tag data worldwide and pointing towards the respective laboratories for clone ordering.

Expanding the focus: The MIPS rice database (MOsDB) a reference for grass genomes

Although of wide use as a model system for plants in general, Arabidopsis has limitations to serve as a model for monocotyledonous plants and genomes. Thus the efforts undertaken in public consortia to sequence and analyse the rice genome are of high importance as a genome reference system for monocotyledonous research. With a draft version of the rice genome in hand we aim to structure and analyse the data in a similar standard as the Arabidopsis genome data. Large scale integration and analysis has been launched and the MIPS rice database has been released in late summer 2002. The functionality circumvents various search and browse options, exhaustive PEDANT analysis and functional categorisation. Beside the application for rice genomics this will lay an important foundation for comparative genomics and will be an indispensable resource for any genomic research in grass genomics.

○ European Union Chromosome 3 Arabidopsis Sequencing Consortium [M. Salanoubat, K. Lemcke,..., H.-W. Mewes and K.F.X. Mayer], The Institute for Genomic Research and Kazusa DNA Research Institute **Sequence and Analysis of chromosome 3 of the plant *Arabidopsis thaliana*.** Nature (2000) 408, 820-823. ○ The Kazusa DNA Research Institute, The Cold Spring Harbor and Washinghton University St Louis Sequencing Consortium & The European Union Arabidopsis Genome Sequencing Consortium [G. Murphy,....,K.F.X. Mayer,..., H.-W. Mewes and M. Bevan] **Sequence and Analysis of chromosome 5 of the plant *Arabidopsis thaliana*.** Nature (2000) 408, 823-826. ○ The Arabidopsis Genome Initiative [Genome Sequencing Groups; Genome Analysis Group: K.F.X. Mayer, O. White *et al.*; Contributing Authors] **Sequence and analysis of the flowering plant *Arabidopsis thaliana*.** Nature (2000) 408, 796-815. ○ The Arabidopsis Genome Initiative [Concept and Production: MIPS; technical Realisation: Biomax Informatics] ***Arabidopsis thaliana* Genome.** CD, Nature (2000). ○ M. Bevan, K.F.X. Mayer, O. White, J. A. Eisen, D. Preuss, T. Bureau, S. L. Salzberg and H.-W. Mewes **Sequence and Analysis of the Arabidopsis genome.** Current Opinion in Plant Biology (2001) 4, 105-110. ○ K.F.X. Mayer **Das Arabidopsis Genom, eine Blaupause für die Genome höherer Pflanzen.** GenomXPress (2001) 1/01, 3-5. ○ K.F.X. Mayer *et al.* **Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*.** Genome Research (2001) 11, 1167-1174. ○ B. Heidenreich, K. Mayer, H. Sandermann and D. Ernst **Mercury induced genes in *Arabidopsis thaliana*: identification of induced genes upon long-term mercuric-ion exposure.** Plant, Cell and Environment (2001) 24 (11), 1227-1234. ○ K.F.X. Mayer **Bioinformatische Analyse des Arabidopsis Genoms.** Vorträge für die Pflanzenzüchtung (2001) 52, 51-60. ○ H. Schoof, P. Zaccaria, H. Gundlach, K. Lemcke, S. Rudd, G. Kolesov, R. Arnold, H.W. Mewes and K.F.X. Mayer **MIPS *Arabidopsis thaliana* Database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome.** Nucleic Acids Research (2002) 30 (1), 91-93. ○ K.F.X. Mayer and H.-W. Mewes **How can we deliver the large plant genomes? Strategies and perspectives.** Current Opinion in Plant Biology (2002) 5, 173-177. ○ B. Morgenstern, O. Rinner, S. Abdeddaim, D. Haase, K.F.X. Mayer, A. Dress and H.-W. Mewes **Exon Discovery by Genomic Sequence Alignment.** Bioinformatics (2002) 18(6):777-87.

Fig 1: *Various MIPS plant genomics databases. MIPS plant genomics databases are both aimed as an information research for genome based, systematic research as well as a platform for comparative plant genomics.*

Synteny driven hierachical shotgun sequencing of the maize genome

As mentioned above the rice genome is an indispensable resource for any grass genomics in the future. Along with wheat and rice, maize is the most important crop genome worldwide. Beside its importance for agriculture it is long established and attractive model system. The genome contains 2500 Mb of sequence, the 20-fold size of the Arabidopsis genome and in the size range of the human genome. International efforts are now undertaken to sequence and analyse the maize genome. MIPS is serving as bioinformatic and genome analysis partner within a consortium aiming to develop a complete physical map of the maize genome, a high density marker map of the maize genome and to initially sequence 20 Mb distributed over all 10 Chromosomes. 450,000 BAC ends will be sequenced and 225,000 BACs will be fingerprinted. The BAC end sequences will be compared against the rice genome backbone and low copy matches will be used to exploit the maize-rice synteny for sequencing map construction. The maize sequences will be analysed in put into relation to other plant genomes not only for genes contained but also for repetitive and transposable elements which are thought to account for up 80% of the maize genome.

SPUTNIK: Reconstitution and comprehensive analysis of transcriptomes

While Arabidopsis and rice may form the core plant genomics spefic aspects and applied aspects mostly can only be studied or applied to crop plants such as leguminous plants, solanaceous or crop grasses.

There is a large short cut to access partial transcriptomic sequences. In total there are in excess of 2 million plant-derived ESTs within the public domain. While ESTs are beset with issues of redundancy, questionable quality and short read-length, they are a solid and computable form of data that stems directly from the plant transcriptome.

SPUTNIK has been implemented as a largely automated pipeline for the processing, clustering and annotation of large numbers of EST sequences. From the non-redundant cluster set, peptide sequences can be derived and with this, large scale genomic analyses can be performed with the partial reconstructed proteome. Based on them analysis for domains, structural and functional characteristics can be carried out. The resulting data is suitable for comparative plant genomics.

Initially all EST sequences from a particular plant species are assembled within a relational-database schema. A suite of advanced sequence cleaning and clustering algorithms (HarvESTer, Biomax informatics) is applied to the EST sequences and a very high-quality set of non-redundant consensus-sequences is assembled. The derivation of the correct coding sequence from EST singletons and tentative consensus sequences is essential for further analysis. To overcome limitations associated with standard modeling tools high stringency BLASTX analyses are used to define likely coding sequences in an organism specific manner, and from the results hexanucleotide residue (di-codon) usage tables are prepared for the individual species. The usage of a Hidden Markov Model and the resultant hexanucleotide probability tables allows the deduction of the correct ORF in a frameshift tolerant manner.

Annotation is anchored to the clusters and derived peptides using a variety of tools. Sequences are assigned to the MIPS functional catalogue. The implementation of structural methods (Known3D and SCOP) allows the estimation of conserved structural folds. The screening of sequences for Interpro domains provides another layer for annotation based on the widest curated collection of diagnostic domains and motifs.

The pertinent items of data are indexed and are tightly woven into the Sputnik relational database schema. Access to the data is provided through the world wide web, and a large number of graphical user interfaces have been implemented for both the search, retrieval and simplified display of resulting data. The framework has been implemented in a fashion, that any scientist should be easily able to zoom into suitable data with the minimum of time spent learning the nuances of the system.

pHOGs: probing for homologous groups of plant proteins

Initial analysis of the plant protein space showed pronounced differences to what can be observed within other kingdoms. Beside in part drastically overrepresented families, caused by for example a polyploid phase during evolution and by expansions of the respective families in plants in general, numerous families are underrepresented or not present at all within Arabidopsis. Thus, plant specific pecularities can already be observed on a comparably coarse level. In addition transfer of knowledge gained within model organisms such as Arabidopsis and rice to crop plants requires comprehensive analysis and knowledge about orthologous and paralogous relationships within and between species beyond the coarse BLAST similarity measure level.

○ P. Kosarev, K.F.X. Mayer and C. S. Hardtke **Evaluation and classification of RING finger domains encoded by the Arabidopsis genome.** Genome Biology (2002) 3 (4), research 0016.1-0016.12. ○ W. M. Karlowski, H. Schoof, V. Janakiraman, V. Stümpflen and K.F.X. Mayer **MOsDB: an integrated information resource for rice genomics.** Nucleic Acids Research (2003) 31, 190-192. ○ S. Rudd, H.-W. Mewes and K.F.X. Mayer **SPUTNIK: A database platform for comparative plant genomics.** Nucleic Acids Research (2003) 31, 128-132. ○ R. Kota, S. Rudd, A. Facius, G. Kolesov, T. Thiel, H. Zhang, N. Stein, K.F.X. Mayer and A. Graner **SniPping polymorphisms from the large barley EST collections.** Molecular and General Genomics (2003) 270, 24-33. ○ E.D. Brenner, D.W. Stevenson, R.W. McCombie, M.S. Katari, S.A. Rudd, K.F.X Mayer, P.M. Palenchar, S.J. Runko, R.W. Twigg, G. Dai, R.A. Martienssen, P.N. Benfey and G.M. Coruzzi **Expressed sequence tag analysis in Cycas, the most primitive living seed plant.** Genome Biology (2003) 4, R78. ○ H. Schoof, R. Ernst, V. Nazarov, L. Pfeifer, H.-W. Mewes and K.F.X. Mayer **MIPS *Arabidopsis thaliana* Database (MAtDB): An integrated biological knowledge resource for plant genomics.** Nucleic Acids Research (2004), 32, Database issue D373-D376. ○ S.-H. Shiu, W. M. Karlowski, R. Pan, Y.-H. Tzeng, K.F.X. Mayer and W.-H. Li **Comparative analysis of the Receptor-Like Kinase Family in *Arabidopsis thaliana* and Rice.** Plant Cell (2004), 16, 1220-1234. ○ K.F.X. Mayer, S. Rudd and H. Schoof **Arabidopsis and its use in cereal genomics.** In: Cereal Genomics (2004) Kluwer Academic Publishers, Dordrecht, Netherlands. ○ S. Rudd, M. Frisch, K. Grote, B.C. Meyers, K. Mayer and T. Werner **Genome-wide in silico Mapping of S/MARs in *Arabidopsis thaliana* suggests correlation of intragenic S/MARs with gene expression.** Plant Physiology (2004), 135, 715-722. ○ K. Schrick, D. Nguyen, W. Karlowski and K.F.X. Mayer **START Lipid/Sterol Binding Domains are amplified in plants and are predominantely associated with homeodomain transcription factors.** Genome Biology (2004) 5, R41.

Thus we are aiming to structure the plant protein space by both using completed plant genomes and the associated annotation as well as the, although fragmentary, extremely rich plant proteome resource stemming from the clustering of ESTs (SPUTNIK, see above). The analysis circumvents several main modules including a similarity measure, domain signatures and association of proteins with similar/identical domains and domain architectures which lead to plant specific homologous groups (pHOGs). Future work will focus towards implementation of a web interface and even more important towards the analysis of the apparantly huge amount of plant specific homology domains using sensitive bioinformatic methods (SESAM, Hidden Markov Models, Super Paramagnetic Clustering). In an additional step association of novel domains with functional data (functional role, EC numbers, association with domains with established molecular role) will be used for assigning tentative functional implications for the respective domains.

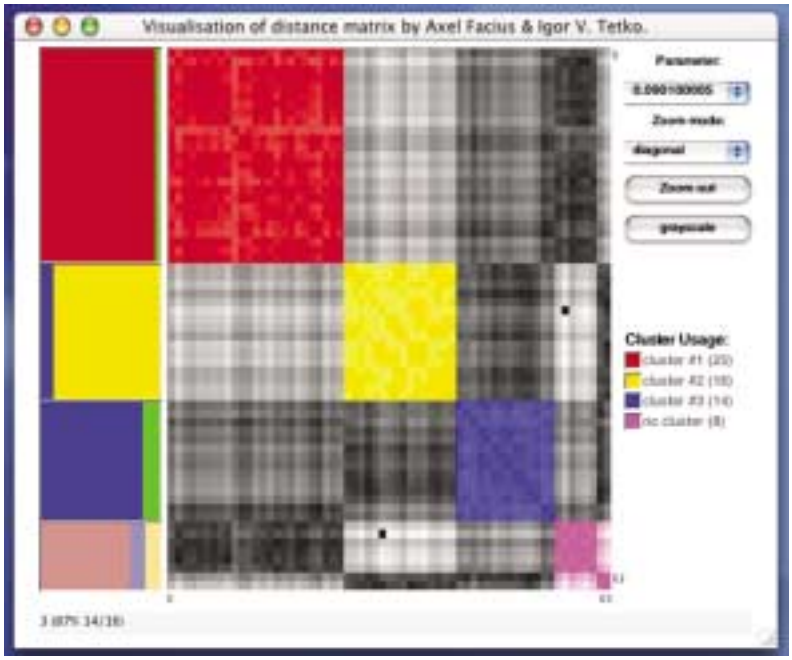


Fig 2: *Graphical representation of automatic protein family assignment by superparamagnetic clustering. More than 400 individual RING finger proteins from Arabidopsis thaliana have been automatically classified. In the right rectangle individual proteins constituting a specific subfamily are highlighted in coloured rectangles. The left rectangle gives the selectivity of the procedure by depicting the ratio of correctly assigned as well as false positives in the respective colour codes.*

GABI Primary Database

Bioinformatics
GABI-PD

Svenja Meyer, Axel Nagel, Steffen Schulze-Kremer
German Resource Center for Genome Research (RZPD), Berlin

Sequence- and SNP/InDel-information, data from mapping experiments and gene expression-, proteome- and metabolite analysis from diverse GABI groups are joined in a single database, the GABI Primary Database. Data are integrated with each other and with information from public databases and are associated with biological material, e.g. clones and mutants. User-friendly graphical interfaces allow effective data searches via the internet (https://gabi.rzpd.de).

GabiPD also supports GABI participants in bioinformatics since there is a great need from individual users.

Data integration and graphical user interfaces

The database is publicly accessible via the GabiPD webpage (https://gabi.rzpd.de). Registered users (GABI- and WPG-participants) can exclusively access non-public data sets via password authorisation. Table 1 shows which data are currently integrated in GabiPD and which user-group has access to the data.

The web-statistics show the database being used regularly for data searches and sequence similarity searches. GabiPD is not only accessed by German GABI- or WPG-participants, but also by foreign users from European and Asian countries, USA and Australia.

New data are integrated in GabiPD regularly. The flexible database structure and developed program modules, which allow a semi-automated data handling, ensure that data can be integrated very fast in GabiPD.

Different graphical user interfaces can be used for data searches.

GreenCards

GreenCards is the main search- and visualisation-interface which allows a text-based data search. GabiPD can be queried for clones, sequences, insertion mutants, SNPs, proteins, mapping information or BLAST results by entering, for example, sequence accessions, clone, gene, marker, or genotype names, gene functions or simply keywords.

Firstly all query hits are shown in an overview and can be selected for a detailed view to get numerous information on each search result, e.g. genotypic information, sequence data, GenBank accessions, SNP and InDel positions or chromosomal positions. Links to other GABI databases like GABI-MASC, GABI-Kat, SPUTNIK or ARAMEMNON are integrated. Links to plant material providers (GABI-Kat, ZIGIA and RZPD) are also included in the GABI Primary Database. This way GabiPD users are informed about available plant material (e.g. clones and insertion mutants) via GreenCards and all plant clones, existing at RZPD, can be directly ordered from the GreenCards interface.

Complex data, like sequence variations (SNPs and InDels) from diverse tetraploid potato cultivars (Rickert *et al.* 2003) are also displayed using GreenCards (fig 1).

A small "statistical function" is included in GreenCards, i.e. all search hits are shown in a hierarchical order grouped by cultivar.

Visualisation of mapping data

Mapping data are visualised by the Java servlet tool called YAMB ("Yet Another Map Browser"). Chromosomal positions of the mapped objects are read directly from the database and the maps are drawn dynamically. This way it is easy to add any new elements and view the actual map at once. Maps of different strains can be displayed in parallel, with same elements on homologous chromosomes linked with each other.

All maps can be zoomed out to overview all mapped elements or can be zoomed in for a more detailed view. The maps are interactive, that is all elements displayed on the maps are clickable. A click calls up GreenCards with detailed information regarding the selected object, e.g. sequence information, positions on the map, accession numbers, BLAST hits, references and SNP and InDel positions (fig 1).

BLAST

An web-interface is available to allow secure BLAST searches (BlastN, BlastP, BlastX, tBlastX oder tBlastN). Secure BLAST search means that no records of the search results are kept at GabiPD. The results are securely transferred and not made accessible for anyone other than the submitter of the query. The BLAST results are visualised graphically (display of the alignments) and can be downloaded. All non-public sequence data stored in GabiPD are made available for BLAST searches exclusively for GABI or WPG participants, while public sequence databases, which are updated from NCBI or TIGR once per week, are freely accessible.

On request public sequences, which are currently not available via GabiPD, can be made accessible for BLAST analysis.

Data from gene expression-, proteome- and metabolite-analysis; compliance of international standards

Data from Affymetrix™ gene expression experiments, proteome- and metabolite analysis from GABI projects are integrated in GabiPD. For integration of gene expresssion data, a database schema compliant to the schema proposed by the Microarray Gene Expression Data group (MGED, www.mged.org) was developed. This way it is ensured that gene expression data integrated in GabiPD are internationally comparable. First Arabidopsis Affymetrix gene expression data are displayed via internet using the web-version of the MapMan tool (see below).

Datatype	Species	Number	Status	Project
Sequences	Arabidopsis	~ 4,600	3	GABI-LAPP
		~ 1,000	5	GABI-LAPP
		~ 13,000	2	GABI *Arabidopsis* II
	barley	~ 140,000	1	GABI PLANT
		~ 10,800	2	
		~ 33,800	5	
		~ 4,000	2	GABI-LAPP/IPK
		~ 7,400	2	GABI-AGROTEC
	potato	~ 2,000	5	GABI CONQUEST
	sugar beet	~ 13,500	1	GABI Beet
T-DNA flanking sequences	Arabidopsis	~ 14,000	1	GABI-Kat
	poplar	~ 130	1	GABI POPLAR
mapping information	potato	12 genetic maps	5	GABI CONQUEST
		878 mapped elemente	5	
SNPs/InDels	potato	1,625 SNP/INDEL positions	5	GABI CONQUEST
	Arabidopsis		2	GABI *Arabidopsis* II
gene expression data	Arabidopsis	3 Affymetrix data sets	1	GABI *Arabidopsis* III
		2 Affymetrix data sets	5	GABI *Arabidopsis* III
proteomics data	Arabidopsis	1 2D-gel image, 32 massspectromic data sets	4	GABI *Arabidopsis* III
		2 2D-gel images, 390 identified spots	1	GABI-LAPP
		6 2D-gel images, ~ 3,300 identified spots	5	GABI-LAPP
metabolite data	Arabidopsis	~ 3,000	5	GABI *Arabidopsis* III
enzymatic activities	Arabidopsis	~ 1,000	5	GABI *Arabidopsis* III
Hybridisations	Arabidopsis	~2,700	1	ZIGIA
Links to plant material providers	Arabidopsis cDNA clones	~10,500	2	GABI *Arabidopsis* II, Klone verfügbar über RZPD
	Arabidopsis insertion mutants	~ 25,000	1	ZIGIA
		~ 9,000	1	GABI-Kat
BLAST results		~ 590,000	1-5	GabiPD

Tab 1: *Integrated data sets from GABI projects: type of data, species and status of data (1: public, 2: accessible for GABI and WPG, 3: accessible for WPG members, 4: accessible for GABI participants, 5: accessible for owner defined group) and project name (Status: 26. January 2004).*

Also Arabidopsis proteomics data have been entered in GabiPD. The 2-D gel images are visualised interactively and all information, which is available for characterized protein spots (e.g. mass spectrometric data, putative protein function, molecular weight and isoelectric point) are accessible via GreenCards.

Data from metabolite analysis and measurements of enzymatic activities are also stored in GabiPD. The graphical visualisation of these data is under way.

Project web pages

GabiPD has set-up web pages for certain GABI projects. The web-links are used as reference for publications. Via the "PoMaMo-web page" (Potato, Maps and More) (https://gabi.rzpd.de/projects/pomamo.shtml) potato data integrated in GabiPD, like mapping and sequence information or SNP/InDel data (Rickert *et al.,* 2003 Mayer *et al.*) are accessible and information on "potato-publications" as well as links to other Solanaceae databases are available.

The "*Arabidopsis thaliana* Proteomics Data" web page (https://gabi.rzpd.de/projects/Arabidopsis_Proteomics/) serves as publication reference (Giavalisco *et al.* accepted) and includes images of 2D-gels of different Arabidopsis tissues and complete lists of identified proteins on these gels. Images and data are also accessible for download.

GabiPD bioinformatics services (gabi@rzpd.de)

There is a great demand for customised bioinformatics services within the GABI community and GabiPD tries to meet these requirements. Sequence trace file conversion including clipping as well as submission of sequence data to public databases like GenBank, EST sequence clustering, sequence comparisons via BLAST analysis and subsequent examination of BLAST reports is covered by GabiPD.

In addition GabiPD supports GABI groups in data handling and data analysis, e.g. evaluation of gene expression data. On request, GabiPD assists also in statistical analysis. Genes represented on the Arabidopsis Affymetrix™ chip were assigned to AGI gene codes and to genes as annotated by the CATMA consortium. These information is currently available for download from our web pages (https://gabi.rzpd.de/services/Affymetrix.shtml).

Last but not least GabiPD develops in cooperation with GABI-partners specific user interfaces, adapted to the special requirements of the cooperation partners. This way the MapMan tool was developed in collaboration with the group of Prof. Mark Stitt, Max-Planck-Institute of Molecular Plant Physiology, Golm

More bioinformatics services are offered, please contact gabi@rzpd.de.

MapMan

The Java program MapMan is a tool to graphically map complex genomic data, e.g. Affymetrix™ gene expression data, proteomics or metabolite data, on metabolic pathways or regulatory networks for instance (O. Thimm, 2004).

Each gene, protein or metabolite is represented by a single box and the gene expression level or metabolite concentration is colour-encoded (fig 2). Genes or metabolites are grouped together according to their roles in metabolism or regulation for example. Advantage of MapMan is its high flexibility, i.e. diverse data types can be visualised and user can use their own biological expertise to adapt the tool to their special demands.

The program is available for download from the MapMan web page (https://gabi.rzpd.de/projects/MapMan/). MapMan is also used to display gene expression data, which are integrated into GabiPD, via the internet. The user can select a pathway, e.g. glycolysis or TCA-cycle, and an experiment from a menu and view up- and down-regulated genes in the context of the selected pathway.

○ O. Thimm, O. Blaesing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L. A. Müller, SY. Rhee and M. Stitt **MapMan: A User-Driven Tool to Display Genomics Data Sets onto Diagrams of Metabolic Pathways and other biological Processes.** The Plant Journal (in press) ○ A. M. Rickert, J. H. Kim, S. Meyer, A. Nagel, A. Ballvora, P. J. Oefner, C. Gebhardt **First generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome.** Plant Biotechnology Journal (2003) 1,399-410. ○ O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L.A. Müller, S.Y. Rhee and M. Stitt **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** The Plant Journal (2004) 37(6), 914-939. ○ S. Meyer, A. Nagel and C. Gebhardt **PoMaMo - a comprehensive database for potato genome data.** Nucleic Acids Research, Database issue 2005, in press. ○ P. Giavalisco, E. Nordhoff, T. Kreitler, K.-D. Kloeppel, H. Lehrach, J. Klose and J. Gobom **Proteome Analysis of *Arabidopsis thaliana* by 2-D Electrophoresis and Matrix Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry.** Proteomics, accepted.

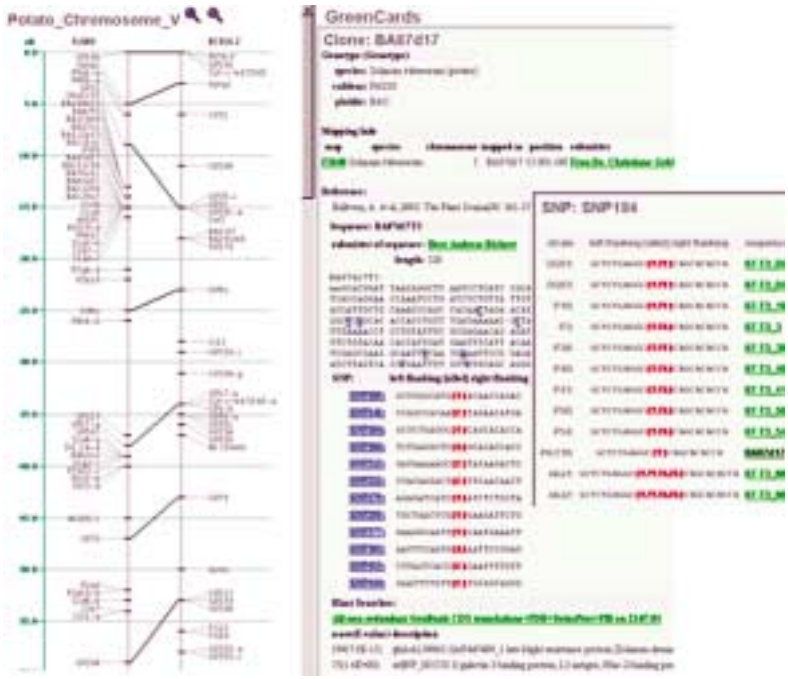


Fig 1: *Mapping data are visualised via a Java Servlet tool. Detailed information for each mapped element, e.g. sequence data and primer information are shown by GreenCards. Variable postions (SNPs and InDels) are highlighted in blue.*

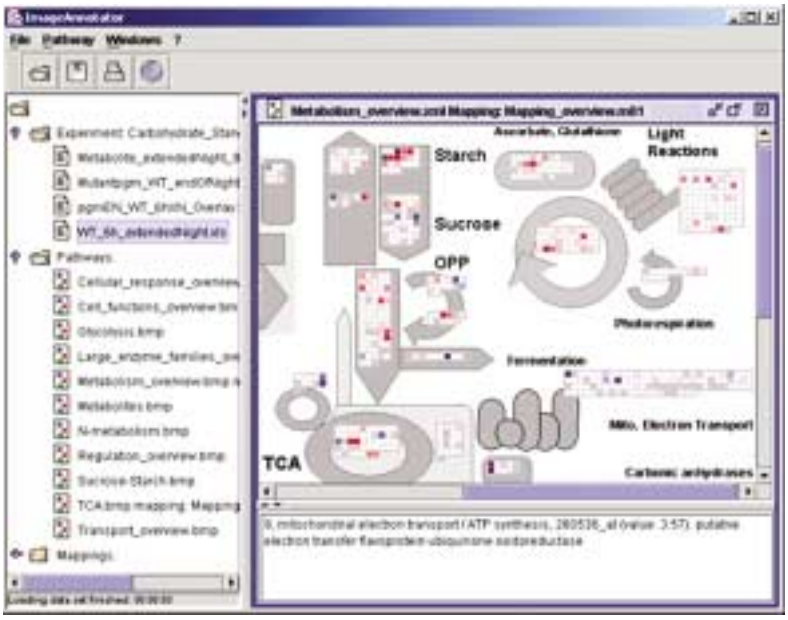


Fig 2: *Affymetrix™ gene expression data mapped on metabolic pathways using MapMan. The gene expression levels of genes are colour-encoded. Increased expression is indicated by blue, decreased expression by red colour.*

Arabidopsis

Overview

In GABI, *Arabidopsis thaliana* serves as one of the two major model organisms for the elucidation of plant gene functions. This small weed has been chosen by the international plant science community as plant molecular-genetic model system due to its advantageous characteristics such as its small size, short generation time, large number of seeds, efficient mutagenesis, extremely easy genetic transformation suitable to conductance in high throughput, and, most importantly, its compact genome structure. The latter character resulted in selection of Arabidopsis as the first plant species for complete genome sequencing. An international effort that resulted in the assembly of the highest quality and best-annotated genome sequence of any multicellular organism hitherto analyzed, was completed in the early phase of the GABI program in the year 2000. Arabidopsis has thus been further pushed forward in its role as the front-runner of plant genomics research world-wide, a role which it will continue to play over the next decade. The work conducted in the GABI program could therefore be built on an excellent technological and informational basis, unparalleled by any other plant species. While the available genome sequence allows researchers to design and conduct experimental work at an entirely new level and with novel research tools, it also very clearly showed the dramatic limitations of our current knowledge on plant gene function, since less than 10% of the estimated 25,000 genes had been studied experimentally at the time the genome sequence was uncovered.

The GABI projects devoted to the analysis of gene functions in Arabidopsis all follow the very general principle of linking genetic variation to phenotypic trait expression. They differ, however, in the way this principle is applied and thus they grouped themselves into three consortia, which together formed the `Arabidopsis-Superconsortium´. Its three pillars are: Consortium 1: `Genetic Diversity´, Consortium 2 `Protein Communities´, and Consortium 3 `Gauntlets: Genes, Phenotypes and Environment´.

The activites in Consortium 1 are directed towards the development of techniques, research tools, and resources that enable and support the analysis of natural diversity, a rich source of genes and gene-variants responsible for adaptation to diverse environmental conditions. The developed resources include defined seed stocks derived from natural accessions (seeds sampled at various sites world-wide) and crosses thereof and high-efficiency molecular genetic marker systems suitable for detection of the genetic diversity represented in these lines. The work in consortium 2 follows the general strategy of `reversed genetics´ that includes the use of sequence information to identify genes of (potential) interest and the application of creation or directed identification of genetic variants with altered activity of these genes. Together with molecular biological and biochemical analyses of the gene products, the phenotypic characterization of the genetic variants yield detailed information of the functions of the selected genes. These principles are here extended to a genome-wide approach with a focus on certain groups (communities) of proteins and their corresponding genes that carry out very important functions. In the third consortium, the `forward genetics´ principle is followed, which starts out with phenotypic selection of variants (usually from large populations) according to particular characteristics (defects or modifications) indicative of alterations in processes of interest caused by genetic changes. The main activities in the projects of this consortium thus consist of i) the establishment of growth conditions or treatments (`gauntlets´) suitable to uncover specific changes in characteristics and the development of analytical tools for assessment of the characteristics, ii) the application of these procedures to identify genetic variants with corresponding altered characteristics, and iii) the identification of the genes affected by the genetic modification. While formally assigned to individual consortia, the projects were right from their initiation designed to be cross-linked with the activities in the other consortia by using the genetic variants created in consortium 1, or the procedures for phenotyping established in consortium 3, or the genetically modified lines generated in consortium 2. Moreover, all groups active in the three consortia are very strongly supported by the materials and the information developed/collected by the resource centers and the bioinformatics centers.

Genetic Diversity Overview

The major goals of the Arabidopsis Consortium 1 `Genetic Diversity´ are the assessment of the genetic diversity represented by a large collection of natural accessions (seed stocks collected from a wide range of locations woldwide) and to develop seed stocks and plants lines that provide the means to assess the phenotypic diversity associated with the genetic variation.

Arabidopsis thaliana, which probably originated in central asia, is naturally occurring in Asia, Europe and Africa at a wide range of geographical locations, from northern Scandinavia (at 68°) to the Cape Verde Islands (at 16°) and from sea-level to high alpine locations. It has also been introduced to North America, Australia and Japan, most probably. Its broad geographic distribution coincides with adaptation of the local populations to a wide range of different growth environments. Accordingly, these natural populations show very considerable diversity in the expression of a large set of adaptive traits such as resistance to biotic stresses (interactions with various pathogens) or abiotic parameters such as high or low temperatures, drought or salt conditions and different day length regimes. As such kind of environmental conditions are of central relevance to agricultural productivity understanding of the genetic determinants underlying the different expression of the adaptive traits is highly desirable.

To support an efficient use of the natural diversity represented in the available accessions, the in project `Creation of novel genetic variants of Arabidopsis´, homogenous lines (varieties) are created through propagation via single seed descent and an extensive crossing program is carried out. In this way F1 and F2 progenies are established from crosses beween the accession-derived varieties and a set of standard reference varieties. Selected crosses are furthermore used to generate recombinant inbred lines (RILs) and genetic substitution lines (near isogenic lines, NILs). In the second project `Establishment of high-efficiency SNP-based mapping tools´ performed jointly by four partners constituting the `Max-Planck-Arabidopsis-SNP-Consortium´ (MASC) two different strategies were followed to detect DNA-polymorphisms in a set of 12 / 6 accessions. On the one hand, a total of 606 `sequence tagged sites´ (STSs) were analysed for 12 different accessions via PCR amplification and comparative sequencing and on the other hand a total of 9,400 `expressed sequence tags´ (ESTs) were generated through sequencing of cDNAs from 6 different accessions. In total 8,095 `single nucleotide polymorphisms´ (SNPs) were hitherto detected and collated together with relevant information in a database for convenient access and use. These polymorphisms have been used to develop a framework marker set consisting of 111 evenly spaced SNP-markers. These markers have been used to genotype 351 different accessions resulting in the hitherto most extensive and detailed information about genetic relationships between these accessions. This mapping tool has also been used to genotype a population of 423 RILs and to select a set of 93 NILs. Finally, a novel highly user-friendly and cost-effective SNP-detection technology has been developed and is being used to set up a large marker set complementary to the aforementioned framework marker set.

The experimental tools (SNP-based marker sets) and the resources (seeds stocks) developed within this consortium are now being used by numerous groups to identify genetic determinants of adaptive traits and form one a central basis for the joint Génoplante-GABI project called `Evaluation of natural diversity in Arabidopsis accessions for traits of agronomic or basic importance´ to be conducted through a collaborative effort of 15 French and 4 German groups.

Protein Community Overview

The primary goals of the Arabidopsis Consortium 2 'Protein Communities' are the classification and functional analysis at the genome level of two distinct protein classes, transcription factors and membrane transporters, that are critically involved in metabolic processes and the control of plant development, and to devise strategies for applying the gained knowledge in agriculturally important plant species.

Unlike prokaryotic cells, plant cells are highly compartmentalized and metabolic pathways localized in more than one compartment have to be connected by transport processes. The cellular components responsible for transport facilitation are integral membrane proteins. Transport proteins play a central role in plant nutrient acquisition and photosynthate allocation to different storage organs. Accordingly, transport processes are also important for biotechnological applications. Modification or modulation of transport processes in crop plants can be applied to improve important traits and to tackle agricultural problems.

The Genomics, Expression & Membranes project aims to identify all putative transporters and other membrane proteins and to select a subset of transporters for functional characterization. The two participating groups pursue an integrated concept consisting of bioinformatical and experimental modules. Novel transporters are identified by genome data mining using a newly developed Arabidopsis membrane protein database (ARAMEMNON) that offers superior features and functionalities, by expression profiling of putative transporters responding to changing conditions and by yeast suppression cloning. Functional characterization of novel transporters involves electrophysiology and knock-out (KO) mutants.

Transcription factors (TFs) are actors in regulatory networks that control the development of a plant, its metabolism and responses to environmental conditions. Arabidopsis and other plants contain approximately 2000 TFs, of which only very few are characterized in some detail. A deeper understanding of the roles of individual transcription factors and whole TF families will be crucially important for developing biotechnological applications that aim to improve agronomically important traits.

The Transcription Factor project aims to generate a wealth of in-depth information on various TF classes and the corresponding target genes. From six TF families, earmarked by different DNA binding domains, approximately 300 representative members are selected for analyses. Hallmarks are the phenotypic analysis of KO lines and transgenic lines that ectopically overexpress TF full length cDNAs, expression profiling of KO and transgenic plants to unravel regulatory networks, and yeast two-hybrid screens to identify interaction partners.

Gauntlets Overview

The primary goal of the Arabidopsis Consortium 2 'Gauntlets' is to identify genes that play a crucial role in determining how the model plant *Arabidopsis thaliana* responds to a wide range of abiotic and biotic factors, and analyse how they interact to determine fitness and productivity. Such responses play a major role in the ecological fitness and the agronomic value of plants, which are unavoidably exposed to large fluctuations in the surrounding world that they have to cope with by metabolic and developmental responses, rather than the behavioural responses that are so important in animals. To do this, a wide range of defined growth conditions ('gauntlets') are being established in which the response to different carbon balances, nutrient limitation, anaerobiosis, temperature, pH, salt, water and heavy metal stress, and biotic stress are clearly, reproducibly and controllably displayed. These provide experimental systems in which screens can be carried out to identify mutants, and in which the phenotypes of transgenic plants and ecotypes (natural diversity) can be evaluated. They also provide systems in which the physiological and molecular response can be dissected. High-throughput techniques will be exploited to analyse transcript-, protein/enzyme activity-, peptide-, oligosaccharide-, and metabolite-profiles. Interpretation of transcript and metabolite profiles is being supported by the development of novel software applications that allow user-driven display of the results in the context of state-of-the-art biological knowledge.

Working against this growing background of information, genes, proteins, peptides and metabolites involved in sensing and/or responding to environmental challenges will be isolated using complementary approaches, including non-biased horizontal analysis of transcripts and metabolites, and mutant screening based on novel phenotypes or expression patterns. Groups of genes, proteins/enzyme activities, cell wall structures and metabolites will be identified that are diagnostic for different states. Transformants with altered expression of candidate genes will be subjected to systematic analysis to provide a detailed analysis of gene function. Different members of the Consortium are focusing on nutrient responses, stress and particularly heavy metal responses, the analysis of cell wall structure and its role, and the extension of analytic techniques to allow them to be applied at the single cell level.

Creation of Novel Genetic Variants of Arabidopsis

Arabidopsis
Genetic Diversity

1 Rhonda Meyer, 2 Carsten Müssig, 2 Thomas Altmann
1 Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm
2 University of Potsdam, Departement of Biochemistry and Biology-Genetics,
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

The large collection of seed stocks sampled from natural populations adapted to a wide range of different growth environments represents a highly valuable resource for genes and gene-variants responsible for the expression of various agronomically highly important adaptive traits. These include resistance to biotic stresses (interactions with various pathogens) or abiotic parameters such as high or low temperatures, drought or salt conditions and different day length regimes. To make this gene pool accessible for efficient experimental analysis, homogenous *Arabidopsis thaliana* varieties are isolated and used in this project to generate F1s and F2s after reciprocal crosses to the two `reference varieties´ Col and C24. Such crosses have allowed us to uncover much more pronounced genetic effects on traits such as phosphate efficiency as compared to analysis restricted only to the evaluation of the natural accessions themselves (1, 2). The seeds of the homogenised natural accession-derived varieties and the F1 and F2 seeds are provided to the other GABI-Arabidopsis partners for evaluation of specific characteristics in order to detect novel genetic determinants of important traits. For a subset of (particularly informative) variety combinations thus identified, recombinant inbred lines (RILs) and genetic substitution lines (near isogenic lines, NILs) are generated and genotyped. The latter task will be supported by the results gained and materials produced in the collaborative GABI project `Establishment of high-efficiency SNP-based mapping tools´ (see 0312275A/9 TP4). Seeds of the genotyped RILs and NILs are provided to the other GABI-Arabidopsis partners for specific evaluation.

Establishment of homogenous varieties

Seeds of various Arabidopsis accessions have been sown and seeds of a single representative individual harvested (fig 1). From these seed stocks, plants were again grown for seed amplification and the resulting seeds bulked and stored. In total 175 Arabidopsis accessions (137 from different geographical origin) different have hitherto been processed and in this way converted into homogenous varieties.

Creation of variety x `reference variety´ F1s and F2s

The established homogenous varieties were used for reciprocal crosses to the `reference varieties´ Col, C24 and Nd F1 seeds plus the corresponding seeds from manual self-fertilisation (necessary control seed for physiological analyses) were derived from 400 crosses (74/52 different varieties crossed reciprocally to the three reference varieties). 348 of these crosses have hitherto been propagated to yield F2 seeds, which were harvested and stored (fig 2).

During growth and propagation of these lines surprisingly large differences in vigour of the F1s in comparison to their parents were observed, which resulted in up to 151% mid-parent-heterosis for dry biomass (fig 3 A, B). These observations highlight the importance of the unbiased crossing program conducted in this project and were used to select variety combinations for the generation of RILs.

Establishment of RIL populations (fig 4)

According to the observation of multiple traits showing different expression in the F1 as compared to the parents (biomass, phosphate efficiency, root morphology, gene expression profiles, metabolite profiles, flowering time) the combination of the varieties C24 and Col has been selected for the first RIL population to be generated. From each of the reciprocal crosses (C24 x Col and Col x C24) at least 212 and 233 lines have been propagated into the F9 generation. In these lines unexpectedly strong variation in flowering time that occurred. The delays in flowering time (and in seed set) became more and more enhanced in subsequent generations. Despite these delays, the C24 x Col / Col x C24 RIL population has been established and genotyped within the frame of the project: Genotyping (using the SNP-based mapping tools created in TP4) has been performed on F7 plants-derived DNA and F8/F9 seed stocks have been made available.

The generation of three further RIL populations has been achieved. Thus, a set of 485 C24/Nd lines, a set of 460 C24/Ak lines, and a set of 497 C24/Bch lines (equally derived from the reciprocal crosses each) have been propagated to the F8 generation. All three combinations were selected according to the observation of large trait variation (including vigour) among the parents and the F1s and F2s.

Establishment of a NIL population (fig 4)

For the generation of NILs the combination of the varieties C24 and Col has also been selected and subjected to a backcrossing program. Within this program, the BC3F2 and BC4F1 generations have been established. For a total of 178 independent BC3F1 plants DNAs have been isolated and the genotyping has been completed using the SNP-based mapping tools created in TP4 (see corresponding report). Thus, genotype data are available at 116 marker loci for the 178 BC3F1 plants. These data have been used to select individuals from the BC3F2 or BC4F1 generations for further analysis. 93 single or double segment substitution lines have been established constituting the NIL population.

This project is tightly linked to the GABI project `Establishment of high-efficiency SNP-based mapping tools´ (see: 0312275A/9 TP4), which provides efficient marker systems for the genotyping of the RILs and NILs. The two projects have been co-ordinated such that the C24 x Col combination has been prioritised in both and the integration will be extended to the other RILs to be genotyped (C24 x Nd and C24 x Ak). For information of the other GABI partners, a web site has been established (www.mpimp-golm.mpg.de/arab-diversity/index-e.html), which displays the available seed stocks plus further information.

Several GABI-Arabidopsis groups are using the material generated here for a diverse set of analyses including responses to various environmental parameters and different plant characteristics (see projects in the Arabidopsis consortium 3 `Gauntlets: Genes, Phenotypes and Environment´). The upcoming GABI-Genoplante joint project `Evaluation of natural diversity

○ R. A. Narang, A. Brüne and T. Altmann **Analysis of phosphate acquisition efficiency (PAE) in different *Arabidopsis thaliana* accessions.** Plant Physiol. (2000) 124, 1786-1799. ○ R.A. Narang and T. Altmann **Phosphate accquisition heterosis in *Arabidopsis thaliana*: a morphological and physiological analysis.** Plant Soil (2001) 234, 91-97. ○ K.J. Schmid, T. Rosleff-Sörensen, R. Stracke, O. Törjek, T. Altmann, T. Mitchell-Olds and B. Weisshaar **Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*.** Genome Res. (2003) 13, 1250-1247. ○ O. Törjek, D. Berger, R.C. Meyer, C. Müssig, K.J. Schmid, T. Rosleff-Sörensen, B. Weisshaar, T. Mitchell-Olds and T. Altmann **Establishement of a high-efficiency SNP-based framework marker set of Arabidopsis.** Plant J. (2003) 36, 122-140. ○ R.C. Meyer, O. Törjek, M. Becher and T. Altmann **Heterosis of biomass production in *Arabidopsis thaliana*; Establishment during early development.** Plant Physiology (2004) 134, 1813-1823.

in Arabidopsis accessions for traits of agronomic or basic importance´, is built upon the stocks developed within this project and will use this material to evaluate phenotypic variation for a wide range of important traits. The newly established priority program (`Schwerpunktprogramm´) of the German research Foundation (DFG) on the molecular analysis of heterosis in plants (SPP 1149, `Heterosis bei Pflanzen´) will have a central focus point on the use of Arabidopsis. The strong heterosis observed in the C24 x Col cross are a major topic of analysis in this program and the genotyped RILs and NILs created within this GABI project are essential for this program. The further work in the SPP 1149 will be strongly supported by the availability of further crosses showing heterosis (see fig 3B).

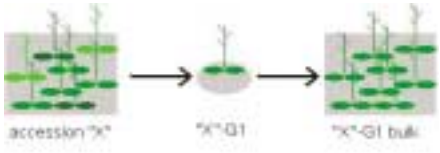


Fig 1: *Homogenous accessions are derived from a single representative individual selected from the different accession populations. 86 homogenous varieties have been established via a single representative individual of the following accessions:*
Aa-0; Ag-0; Ak-1; An-2; Ang-0; Ang-1; Ba-1; Bay-0; Bch-1; Bd-0; Be-0; Be-1; Bl-1; Bla-1; Bla-3; Br-0; Bs-1; Bs-2; Bs-5; Bsch-0; Bsch-2; Bu-0; Bu-2; Bur ; C-24; Ca-0; Chi-0; Cl-0; Co; Co-2; Col-0; Col-2; Col-3; Ct-1; Cvi; Da(1-12); Da-0; Db-0; Dijon-G; Dijon-M; Dr-0; Dra-0; Edi-0; Ei; Ei-2; Eil-0; El-0; En-2; Enkheim-D; Enkheim-T; Ep-0; Er-0; Est; Et-0; Fi-0; Gr; H55; Hi; Hodia-Obi-Garm; Je 54; La er; Ler-1; Lip-0; Lm; Lu; Nd; Ob-0; Old; Old-1; Oy; Petergof; RLD 1; Rsch; RU-1; S 96; Sha; Sn(5)-1; Sol-0; Sorbo; Sue; Ta; Te; Wil; Ws; Yo; Ze-0.

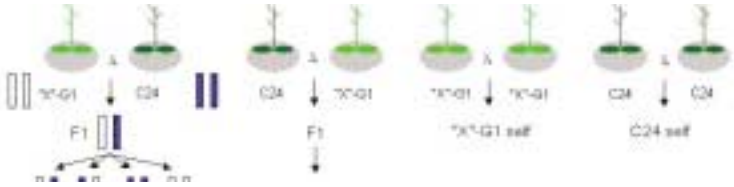


Fig 2: *Production of F1, F2 and selfed progeny from homogenous accessions,shown for crosses with reference variety C24 144 reciprocal F1s ("X"-G1 x C24, "X"-G1 x Col) have been established for the following accession-derived varieties:*
Aa-0; Ak-1; An-2; Ang-0; Ang-1; Bay-0; Bch-1; Bd-0; Be-0; Be-1; Bs-1; Bs-2; Bs-5; Bsch-0; Bsch-2; Bu-2; Bur; Aa-0; Ak-1; An-2; Ang-0; Ang-1; Bay-0; Bch-1; Bd-0; Be-0; Be-1; Bla-3; Bs-1; Bs-2; Bs-5; Bsch-0; Bsch-2; Bu-2; Bur; Ca-0; Chi-0; Cl-0; Co-2; Col; Col-2; Col-3; Ct-1; Cvi; Da-0; Da(1-12); Db-0; Di-G; Dijon-M; Dr-0; Dra-0; Edi-0; Ei-2; Eil-0; El-0; En-2; Enkh.-D; Enkh.-T; Ep-0; Er-0; Est; Et-0; Fi-0; Fr-4; Gr; H55; Hi; H-O-G; Je54; Laer; Lip-0er; Lm; Lu; Nd; Ob-0; Old-1; Oy; Petergof; RLD-1; Rsch; Ru-1; S96; Sha; Sn(5-1); Sorbo; Ta; Te; Wil; Ws; Yo.

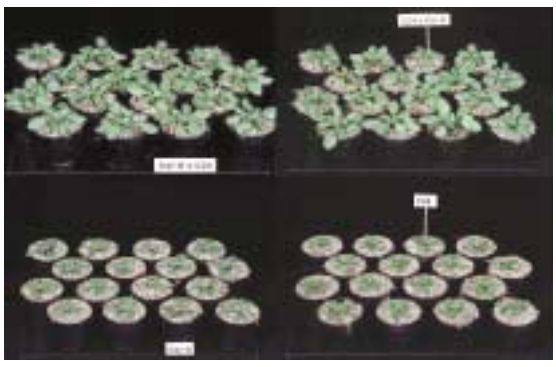


Fig 3A: *In the combination Col / C24 the reciprocal F1s (top) shows increased vigour compared to the parents (bottom).*

Fig 3B: *Mid-Parent-Heterosis for dry weight in F1's of "X"-G1 x Col, x C24 and x Nd. MPH, calculated as (mean F1-mean P)/(mean P), ranges from –34 to 151%.*

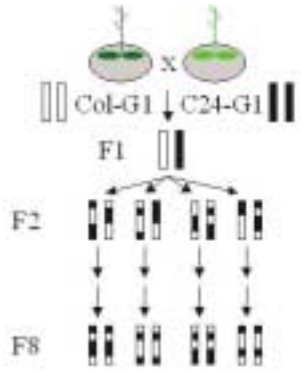


Fig 4: *Establishment of RILs shown for one of the reciprocal crosses of Col / C24.*

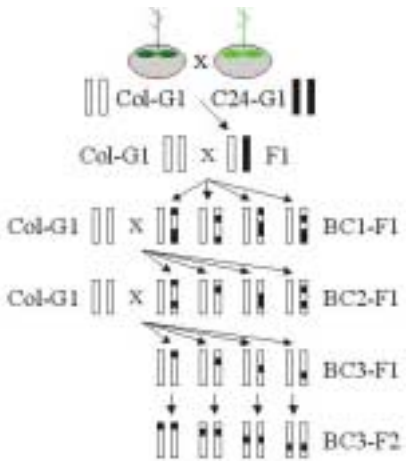


Fig 5: *Establishment of NILs shown for one of the reciprocal crosses of Col / C24.*

Large-scale SNP identification and analysis in the Arabidopsis genome

Arabidopsis
Genetic Diversity

K. Schmid, Thomas Mitchell-Olds
Max Planck Institute for Chemical Ecology (MPI-CE), Jena

The main project goal was to identify a genome-wide set of single nucleotide polymorphism (SNP) markers for high-throughput mapping in Arabidopsis in order to accelerate positional cloning and QTL mapping to find genes influencing important phenotypic traits. At the time of the project start, no such set of genome-wide SNPs was available. Additionally, the project was used to develop SNP technology for the application to crop plants. The markers and experimental approaches developed in this project were aimed to be available to other research groups working with genetic markers and were expected to have an important impact on future mapping projects in Arabidopsis and other plant species.

To identify SNPs, we PCR- amplified and sequenced about 600 evenly spaced regions in the Arabidopsis genome from twelve different varieties (Col-0, Ler, Cvi, C24 and Nd-1, Ei-2, CS22491, Gü-0, Lz-0, Wei-0, Ws-0, Yo-0). These sequences were then compared against the corresponding regions in the Arabidopsis genome and SNPs detected as sequence differences to the genome sequence.

We designed 8 different sets of STS primers to amplify genomic segments of about 600 bp length (Table 1). Two series of primers were designed to provide a set of about 100 markers that are spaced about 1 Mb from each other and optimized to amplify noncoding regions because they were expected to be highly polymorphic (Je_1MB set, Go-1MB). Additional sets of primers were designed to amplify regions that are spaced on average 100 kb. Three STS sets were known to contain candidate SNPs (TIGR, CAPS, ATHA_EST primer sets). The other sets amplify randomly chosen regions of the genome without prior information on SNPs.

In addition to the six accessions selected for SNP generation by EST sequencing by the GABI-MASC project partner at the MPI for Plant Breeding Research, Cologne, five other accessions and Col-0 as control were selected for the STS approach. We tested a total of 606 primer pairs for PCR amplification and with 595 (98%) of these primer pairs it was possible to obtain a PCR product from at least one accession (including Col-0). We obtained 4,995 STS sequences with >100 high-quality basecalls that were used for SNP detection.

A total of 2,236,318 non-redundant basecalls were compared with the Col-0 genome sequence. The comparison led to the identification of 3,773 SNPs and 619 InDel polymorphisms. A total of 620 (16%) of the SNPs were genotyped from all 12 accessions and 2,687 (71%) from at least eight accessions. Among all SNPs, 2922 (77%) are located in regions of the genome annotated as noncoding and 869 (23%) in coding regions. Among InDels, 617 were noncoding and only 2 coding. We were able to determine the coding status of 857 SNPs in coding regions and found 410 (48 %) replacement SNPs and 447 (52%) synonymous SNPs. Polymorphisms that were directly sequenced from both strands of the PCR products can be considered to be confirmed polymorphisms. Using this criterion, 2,331 (62%) of the SNPs and 343 (55%) of all InDels are confirmed polymorphisms in at least one accession. The proportion of false positive SNPs was estimated to be 0.043%

Frequency distribution of SNPs

To estimate the proportion SNPs that are rare polymorphisms, we calculated the relative frequencies of all SNPs in the STS sample (Figure 1). We calculated the frequency of all SNPs whose allelic states have been determined from at least 8 out of 12 accessions (n=2,681). Most SNPs are rare and segregate at low frequencies. A total of n=1,344 (51%) SNPs with a sampling depth of at least eight accessions occur as singletons. We also compared relative frequencies of noncoding, silent and replacement SNPs. Silent SNPs have the highest average frequency and lowest number of singlets whereas replacement SNPs show the opposite pattern, namely the lowest average frequency and the largest proportion of singletons among the three different SNP types.

A comparison of the average sequence divergence to Col-0 among accessions indicates that the Cvi-0 accession is substantially more divergent from Col-0 than the other accessions (Table 1). The number of base pairs that have to be analyzed to detect a SNP varies more than 1.5 fold between the most (Cvi-0) and the least divergent (Gü-0) accession.

To provide public access to the data generated in this study a web interface was created (https://www.mpiz-koeln.mpg.de/~GABI-Kat/MASC). It is possible to retrieve SNP markers based on information on genomic location, hits to protein coding genes or differences between two accessions that were surveyed in this study. The web interface provides information about primers for various genotyping methods (e.g., CAPS, primer extension, Pyrosequencing) which allows the design of large-scale experiments for mapping.

The data generated in this project were also used to investigate the evolutionary forces that have shaped patterns of genetic variation in the genome of *A. thaliana* (Schmid *et al.* 2005). Using coalescent simulations and maximum likelihood simulations, we found that the patterns of variation are not consistent with a simple neutral model that assumes no population structure, a constant population size and no selection. By testing several alternative models we found that selection may be responsible for unusual patterns at a few loci but that the genome-wide pattern of variation is mainly determined by demographic effects such as changes in population size and structure due to repeated glaciation events in Europe. These results are important for the interpretation of future studies of genetic and phenotypic variation in *A. thaliana*.

○ K.J. Schmid, T.Rosleff-Sorensen, R. Stracke, O.Torjek, T. Altmann, T. Mitchel-Olds und B. Weisshaar **Large-Scale Identification and Analysis of Genome-Wide Single-Nucleotide Polymorphisms for Mapping in *Arabidopsis thaliana*.** Genome Research (2003) 13, 1250-1257. ○ O. Törjek, D. Berger, R.C. Meyer, C. Müsig, K. Schmid, T. Rosleff-Sörensen, B. Weisshaar, T. Mitchell-Olds and T. Altmann **Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis.** Plant J. (2003) 36, 122-140. ○ K. J. Schmid, S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar and T. Mitchell-Olds **A multilocus analysis of genetic variation in Arabidopsis thaliana reveals a genome-wide deviation from a neutral model of nucleotide polymorphism.** Genetics (2005), in press.

Primer set	N	Percent successful amplifications	High quality base pairs	SNPs detected	SNP frequencies
Nonrandom Set					
CAPS	122	85	427,841	653	1:655
TIGR	89	96	320,807	609	1:527
ATHA_EST	48	97	208,490	309	1:675
Random Set					
CYS	56	85	185,588	254	1:731
CHR4	48	90	177,132	449	1:395
Je_1MB	100	88	414,536	637	1:651
Go-1MB	111	n/a	477,476	829	1:576
ADRUM_EST	32	47	18,080	33	1:548
TOTAL	**606**	**88**	**2,236,318**	**3,773**	**1:595**

Tab 1: *Summary of sequence-tagged-sequence (STS)-derived SNPs*

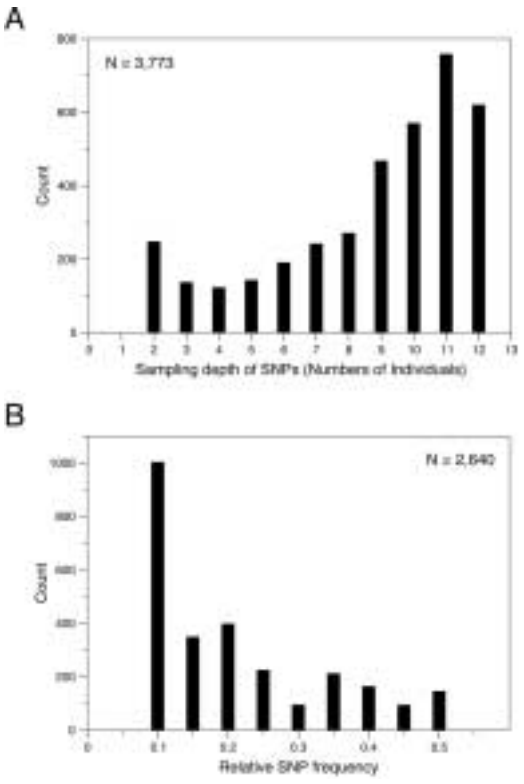


Fig 1: *(A) Sampling depth of SNPs derived from STS sequences. Sampling depth is defined as the number of different accessions from which the SNP was sequenced. (B) Frequency distribution of SNPs with a sampling depth of eight or more accessions.*

Establishment of high-efficiency mapping tools

Arabidopsis
Genetic Diversity

1 Dieter Berger†, 2 Otto Törjek, 2 Thomas Altmann
1 Max Planck Institute of Molecular Plant Physiology (MPI-MP), Golm
2 University of Potsdam, Departement of Biochemistry and Biology-Genetics,
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

The natural *Arabidopsis thaliana* accessions sampled from various different growth environments represent extensive phenotypic diversity related to adaptive traits. In order to make this diversity accessible to genetic analysis a highly efficient genome-wide marker system is required. The establishment of such a genetic marker system is the major aim of this project, which is integrated into a joint effort by a total of four groups (MPI-Z, MPI-CE, MPI-MG, MPI-MP; the MPG Arabidopsis SNP Consortium, MASC). To achieve this single nucleotide polymorphisms (SNPs) are identified among a set of *Arabidopsis thaliana* varieties (initially selected: Col-0, Laer, Cvi, C24 and Nd). The SNP information is than used to establish of a framework marker set of ca. 100 evenly spaced SNP-based markers for use in rapid mapping of new mutations and for genotyping of a widened set of accessions and of recombinant inbred lines (RILs) or genetic substitution lines (near isogenic lines, NILs) (see 0312275A/9 TP1). For this purpose, sites spaced by ca. 1.3 Mbp are selected from the Arabidopsis genome (initially according to the BAC-based physical map data, lateron according to the complete genomic sequence) and subjected to comparative sequencing throughout the different varieties. This is done by PCR amplification of ca. 600 bp genomic DNA fragments of the different varieties using primers designed according to the Col-0 A. thaliana genomic sequence. Comparative sequence analysis is thereafter used to identify polymorphic nucleotides. The information gained in this way is used to set up efficient SNP detection procedures (including novel techniques developed by the MPI-MG group). Upon validation through the MASC groups, the data and the mapping tools (detection procedures) will be provided to the other GABI partners. These tools will support QTL mapping and positional gene cloning, two genetic approaches central to the identification of (lead) genes of high economic value.

Selection of ca. 100 evenly spaced sites, primer design and amplification of genomic DNA fragments

In order to establish a marker set covering the entire Arabidopsis genome with even spacing of ca. 1.3 Mbp, the initial set of 100 sites (regions of DNA sequence) has been extended to 111. Primer design and PCR amplification of one fragment per region has been performed for the five initially selected varieties for a set of 118 sites and could also be achieved for a seven additional varieties, CS22491, Ei-2, Gü-0, Lz-0, Wei-0, Ws-0, Yo-0 through a tightened collaboration with the MPI-CE and MPI-Z groups. Since about 60% of the fragments analysed yielded polymorphic sequences, a total set of 155 fragments was amplified for the C24 and Col varieties (and sequenced, see below).

Sequencing of at least 100 genomic fragments for each of the different varieties

For all five initially selected varieties, at least one genomic DNA fragment was sequenced for at least 100 sites. All sequencing has been finished and it has been extended to and completed for a the additional seven varieties through the aforementioned tight collaboration with the MPI-CE and MPI-Z groups. For C24 and Col, a total of 155 DNA fragments have been amplified and sequenced (a total of 75 kb) in order to increase the number of sites for which SNPs are detected.

Sequence analysis and identification of polymorphisms

The obtained sequences have been subjected to multiple alignment analysis. Pairwise comparisons among all varieties resulted in the identification of 58% to 76% polymorphic sites (regions with at least one SNP). Within the 74,964 nt compared between C24 and Col 598 polymorphic nucleotides were detected. In addition, 85 sequence insertions/deletions were observed. The sequences generated in this project have been submitted for inclusion in the central MASC SNP-database established at MPI-Z (https://www.mpiz-koeln.mpg.de/~GABI-Kat/MASC/). To integrate all sequences data hitherto accumulated by MASC and to perform an analysis on a uniform data set throughout all MASC data, the data produced in this project were transferred to the MPI-CE group in raw format. Analysis of the integrated total MASC data set uncovered important information about the Arabidopsis genome evolution (Schmid *et al.* 2003).

Assembly of framework marker sets

Primary focus of the assembly of a (complete) framework marker set has been on the C24/Col combination of varieties, due to the immediate demand in the GABI project `Creation of Novel Genetic Variants of Arabidopsis´ (see 0312275A/9 TP1). From the sequence data gained here, a total of 111 polymorphic sites could be identified for this combination. Due to the availability of the sequence data generated by the collaborating MPI-CE and MPI-Z groups, a complete set of 111 polymorphic sites with optimised spacing (between 1 and 1.5 Mb) could be assembled (in the final selection the subset of sites which are also polymorphic between Col and Ler was maximised): 68 polymorphic sites derived from MPI-MP sequences, 38 polymorphic sites derived from MPI-CE sequences and 6 polymorphic sites derived from MPI-Z sequences (Fig 1). Framework SNP-sets were also assembled for the variety combinations Col/Ler (112 polymorphic sites), Col/Cvi (112 polymorphic sites), Col/Nd (107 polymorphic sites), Ler/Cvi (101 polymorphic sites) and Ler/Nd (102 polymorphic sites).

Development of SNP-based marker sets

With the availability of the completely assembled set of 111 polymorphic sites (C24 vs. Col), SNP-assays for marker analysis could set up (Törjek *et al.* 2003). A dual strategy has been followed here:

(1) Setup of arrangements for custom SNP analysis by service providers such as Qiagen, Sequenom, and GAG BioScience for the performance of large scale SNP analyses (e.g. for genotyping of entire RIL or NIL populations).

All three service providers were supplied with a test set of 15 DNAs derived from individuals of known genotype (C24 homozygotes; C24/Col heterozygotes; Col homozygotes) and were asked to perform SNP assays for 10 different marker loci. They all delivered the results (in each case 150 data

○ K.J. Schmid, T.Rosleff-Sorensen, R. Stracke, O.Torjek, T. Altmann, T. Mitchel-Olds and B. Weisshaar **Large-Scale Identification and Analysis of Genome-Wide Single-Nucleotide Polymorphisms for Mapping in *Arabidopsis thaliana*.** Genome Research (2003) 13, 1250-1257. ○ O. Törjek, D. Berger, R.C. Meyer, C. Müssig, K.J. Schmid, T. Rosleff-Sörensen, B. Weisshaar, T. Mitchell-Olds and T. Altmann **Establishement of a high-efficiency SNP-based framework marker set of Arabidopsis.** Plant J. (2003) 36, 122-140.

points), which turned out to be 100% correct. For reasons of user-friendlyness and lowest costs, the GAG BioScience company (Bremen) was selected for a further extended test analysis: DNAs isolated from 48 C24 x Col BC3F1 and Col x C24 BC3F1 plants was sent to this company asked to analyse the 10 polymorphic sites. Of the total of 480 data points to be analysed 478 were delivered and the results were consistent with the analyses of the same marker loci in samples of the identical DNAs at MPI-MP (see below) in 476 of the 478 cases. According to this high fidelity, GAG-Bioscience was commissioned to set up the SNP assays for the 111 polymorphic site (all of which are now established there).

(2) Implementation of efficient SNP assays locally for small to moderate scale SNP analyses (e.g. for (re-) evaluation of individual lines as in the frame of specific NIL creation or for mapping of individual mutations):

To this end, multiplexed SNaPshot assays have been established. The 111 marker assays have been assembled into 5- to 8-plex reactions (Fig 2). To confirm the viability of these assays the aforementioned DNAs of the 47 BC3F1 plants have been analysed for these 111 markers and all the 5040 data points yielded bonafied results (Fig 3). This part of the project will furthermore be closely coordinated with the MPI-MG group, who has set up a novel SNP assay system based on LDR (ligase detection reaction).

Of the 111 polymorphisms identified for C24 vs. Col 62 are also polymorphic in Ler/Col, 61 in Cvi/Col, 49 in Nd/Col, 20 in C24/Ler, 23 in C24/Cvi, 36 in C24/Nd, 10 in Ler/Cvi, 17 in Ler/Nd and 18 in Cvi/Nd. These subsets of markers can thus be used for mapping in crosses involving these varieties.

Most recently, a set of 351 accessions was subjected to genotyping using the framework marker set (Fig4). The data thus obtained provide the hitherto most extensive and detailed information about genetic relationships between these accessions. This marker system has also been used to genotype 423 (Col/C24) RILs and to select 93 (Col/C24) NILs.

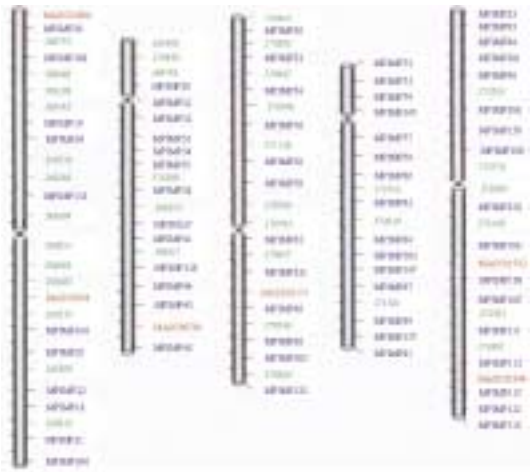


Fig 1: *SNP-based framework marker set assembled for the C24/Col-0 combination of varieties (SNP data collected by the three collaborating groups: MPI-MP: 68, MPI-CE: 38, MPI-Z: 6).*

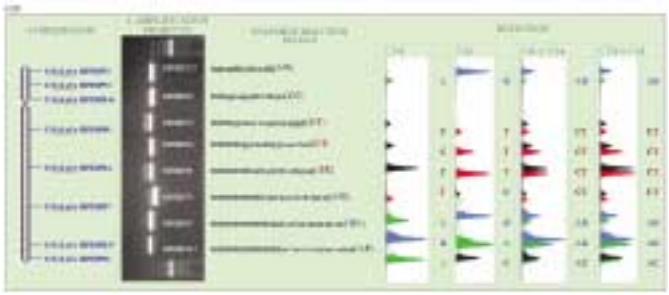


Fig 2: *Multiplex SNaPshot assay. Eight-fold multiplex reaction with SNP markers selected from the SNP-based framework marker set for C24 vs. Col-0.*

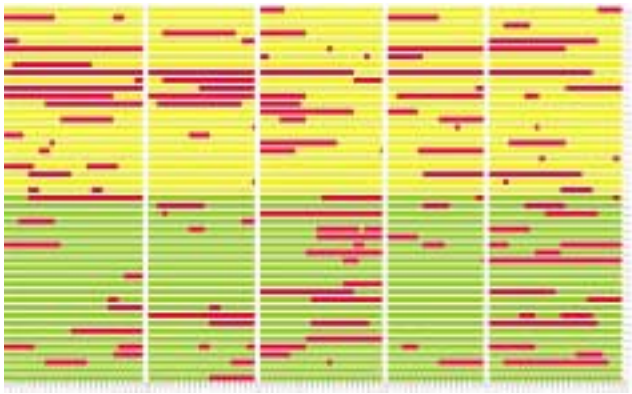


Fig 3: *Results of SNaPshot analysis for genotypisation of 48 reciprocal BC3F1 plants. The 24 Col x C24 BC3F1 and 24 C24 Col BC3F1 plants have been genotyped for 112 markers (yellow: homozygous Col; red: heterozygous; green: homozygous C24).*

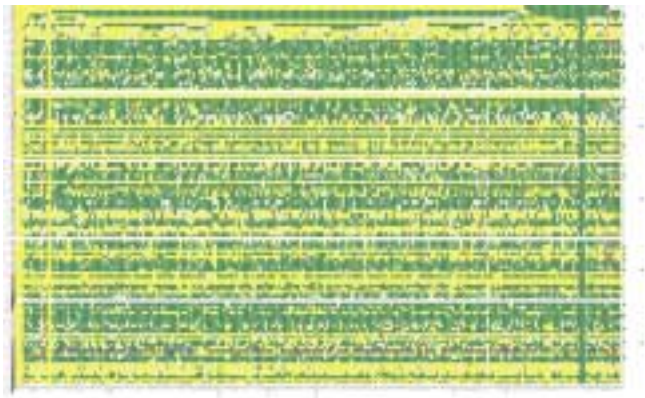


Fig 4: *Genotypisation of 351 Arabidopsis thaliana accessions using the established framework marker set (yellow: Col allele; green: C24 allele; red: heterozygous; blue: third allele; grey: n.d.).*

Establishment of high-efficiency SNP-based mapping tools and development of methods for genome-wide mutation detection

Arabidopsis
Genetic Diversity

1 Ralf Stracke, 2 Thomas Rosleff Sörensen, 3 Bernd Weisshaar
Max Planck Institute for Plant Breeding Research (MPIZ), Cologne

Project report

The goal of the project was to develop SNP (single nucleotide polymorphism) markers for high-throughput mapping in the model plant *Arabidopsis thaliana*, and to implement and validate the required technology for easy transfer to other plants as well as to other laboratories. SNP-based high-throughput mapping will accelerate positional cloning and QTL mapping, methodologies of central importance for the identification of functionally and economically important genes. This technology will be highly valuable for both academia and industry. Our task was to provide data for a high-density *Arabidopsis thaliana* SNP-based map, and to verify the newly established mapping tools in "real world" examples. We should contribute at least 750 SNPs (in comparison to Col-0 accession as standard) to reach the goal of about one SNP marker every 100 kbps on average in the high-density map. SNPs should be deduced from high-quality cDNA sequences (derived from selected accessions) by comparison of the data to Col-0 genomic sequence. Genetic variation indicates that two divergent genotypes may contain at a given SNP locus different alles; therefore, some of the detected SNPs in any of the selected genotypes might also be informative in other genotypes.

Ordered cDNA libraries of six accessions, namely Ler, C24, Cvi, Nd-1, Ei-2 and Ak-2 were prepared. These accessions were selected because they are most frequently used as standard laboratory accessions for mutagenesis experiments, genomic sequencing, set-up of mapping populations and/or on the basis of particularly great phenotypic or genotypic differences. The Ler, Cvi, Nd-1 and Ei-2 cDNA libraries were generated using RNA prepared from whole mature, six week old *A. thaliana* plants treated for 24 hours with different stresses (heat, cold, drought, wounding, darkness, UV-light, salt) to enrich for transcripts from genes related to interactions with the environment. The Ak-2 cDNA library was produced on the basis of RNA from inflorescences. Selection of this accession was based on the generation of a recombinant inbred line (RIL) set between Ak-2 and C24 within the GABI Arabidopsis Verbund I: Creation of novel genetic variants of Arabidopsis (Altmann, Förderkennzeichen 0312275 A). After transformation in *E. coli* the resulting bacterial colonies were picked and stored in 384-well format microtiter plates and copies of these libraries were given to the RZPD (Resource Center/Primary Database).

cDNA-inserts of the bacterial clones were amplified by PCR using flanking vector-primers, and the resulting PCR-fragments were used as templates in sequencing reactions using BigDye-terminator chemistry. A total of 10,706 ESTs were produced (Ler: 1,248; Cvi: 1,774; Nd-1: 1,629; Ei-2: 2,088; Ak-2: 1,248; C24: 2,719). These sequences were subjected to stringent quality filtering including vector clipping and quality trimming. This resulted in a total of 1,941,596 high-quality bps representing 7,465 high-quality reads (Table 1). These represent 5,289 distinct clusters with 79% singlet reads, showing that the libraries are of low redundancy. With the SNP detection tools (see below) a total of 4,327 candidate SNPs and 18 candidate InDels were detected in the sequences (InDels in noncoding regions were not included). The distribution on the five chromosomes shows that the whole genome is well covered by EST-derived SNPs with the exception of the centromeric regions. More detailed analysis showed that the majority of SNPs are located in coding regions (3,432; 79%), and among coding SNPs a significant proportion (1,101 of 3,432; 32%) did cause an amino acid replacement. Two SNPs lead to a nonsense codon. Among InDel polymorphisms, ten are in-frame and eight out-of-frame. Furthermore our 10,706 ESTs hit 3,229 annotated genes; 479 clustered ESTs did not match any of the currently annotated genes, and 201 sequences did match a total of 177 genes which had no EST hit before, making this EST data set useful for gene annotation in *A. thaliana*. – All sequence trace files have been transfered to RZPD.

A cDNA sequence analysis pipeline with software tools for the identification of high-quality EST sequences, vector clipping, comparison of sequence data to the Col-0 standard and SNP identification was developed and installed in close cooperation with Karl Schmid (MPI-CE). To provide public access to markers generated in this study, a "MASC SNP Database" with a web interface (http://www.mpiz-koeln.mpg.de/masc/; available upon publication of the data) was created and implemented which covers all SNPs generated by GABI project partners. This database is searchable for polymorphic sites on selected accessions, genomic location, hits to protein coding genes and includes "virtual crosses" (i.e. display of polymorphisms between a chosen pair of accessions other than Col-0; figure 1). As output "SNP Cards" are displayed with an unique identifier for the SNP and additional information such as upstream and downstream sequences, the SNP-type (i.e. noncoding/synonymous, coding-replacement, coding-silent), and primer sequences that will aid the application of these markers in large-scale mapping experiments using various genotyping methods (e.g. CAPS, primer extension, Pyrosequencing).

Since most EST-derived SNPs are derived from only a single sequencing reaction, they need to be considered as hypothetical. To estimate the proportion of false positives among these SNPs, we designed primers for 96 amplicons covering genomic regions with EST-derived SNPs and used them to generate and sequence PCR products. From 96 polymorphic sites that were analysed, 92 were confirmed to match the expected Col-0 sequence from MAtDB, two did display a difference, and two analyses failed due to PCR or sequencing problems. The PCR failure rate for the other accessions was higher than for Col-0 (on which the primers were designed), but in 81 cases data for both the targeted SNP and the Col-0 sequence were available. Of these, only three turned out to be incorrect, indicating a confirmation rate of 96%. In addition, among 1,858 SNPs that are located in EST clusters of at least two sequence reads, only eight differ between the individual sequence reads and appear to be sequencing errors or reverse transcriptase-induced mutations. We therefore conclude that due to our stringent quality criteria, a very high proportion of EST-derived SNPs are true polymorphisms.

A strategy basing on LCR (Ligase Chain Reaction) is followd by the GABI partner Aleksey Soldatov (MPI-MG) to provide an efficient SNP-detection tool. To ensure that these methodology can later be established in other laboratories without much trouble we will map mutant allels by using the developed SNP-based technique under realistic conditions. For mapping experiments we have chosen the *transparent-testa 1(tt1)* and *transparent-testa 10 (tt10)* mutant (EMS mutants in Ler background).

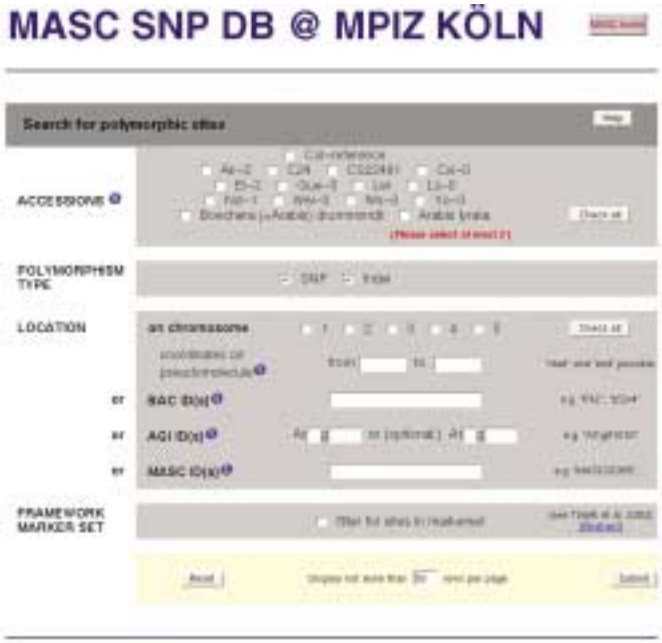


Fig 1: *Screenshot of the GABI "MASC SNP database".*

Accession	ESTs	ClusteredESTs	Mapped ESTs[a]	High-quality bp	SNPs	SNP/bp sequenced	InDels[b]
Ak-2	1,248	746	620	311,040	878	1:354	8
C24	2,719	1,193	916	398,384	1,137	1:350	5
Cvi-0	1,774	877	685	322,322	1,345	1:240	4
Ei-2	2,088	944	717	322,698	861	1:386	2
Ler	1,248	712	590	285,822	840	1:340	5
Nd-1	1,629	817	648	291,330	839	1:347	1
total	**10,706**	**5,289**	**4,176**	**1,941,596**	**4,327[c]**	**1:336**	**18[b]**

Tab 1: *Summary of EST-derived SNPs. [a]Mapped ESTs refer to ESTs which could be aligned with annotated genes. [b] Only InDels in open reading frames were counted. [c]Due to redundancy (e.g. intermediate frequency polymorphisms), the overall number of SNPs und InDels is lower than the sum of all accessions.*

Ligation detection reaction – TaqMan procedure for SNP detection on genomic DNA

Arabidopsis
Genetic Diversity

1 Aleksey Soldatov, 1 Tatiana Borodina, 2 Thomas Altmann, 3 Bernd Weißhaar, 4 Thomas Mitchell-Olds
1 Max Planck Institute of Molecular Genetics (MPI-MG), Berlin
2 Max Planck Institute of Molecular Plant Physiology (MPI-MP), Golm
3 Max Planck Institute for Plant Breeding Research (MPIZ), Cologne
4 Max Planck Institute for Chemical Ecology (MPI-CE), Jena

The goal of the project was to develop an effective SNP (single nucleotide polymorphism) detection method for genotyping a model plant *Arabidopsis thaliana* and transfer the technology to GABI scientific community. High-throughput SNP genotyping is required for effective positional cloning and is necessary for studying complex genetic traits (plant size, blooming time, fruitfulness, etc.). Reliable SNP detection methodology is highly demanded for scientific research and industrial applications.

We have elaborated a new high-throughput and cost-effective ligation detection reaction (LDR) – TaqMan SNP genotyping method (patent EP03019521). User-friendly experimental procedure allows to hope that it will become the method of choice for large-scale genotyping projects.

By now we have prepared a genotyping kit (for 140 SNP loci) for A. thaliana accessions Columbia and C24. The kit is optimized for end-point fluorescent closed-tube detection. Allelic discrimination for 30 loci of this kit is shown on figure 1.

The method is available in two formats: (i) "one tube – one locus" and (ii) "one tube – many loci". First is a two-step closed tube procedure (adding of two kit premixtures to the genomic DNA; no purifications and hybridizations), which takes about 3 hours and requires 0.5-50ng of genomic DNA (5ng for routine genotyping) per reaction. Price of genotyping is 1.5 Euro on commercial enzymes and 0.3 Euro on home-made. Second is a three-step procedure (for about 100 loci), which takes about 4 hours and requires 100-500ng of genomic DNA per one tube. Price per locus is ~0.6 Euro on commercial enzymes and ~0.1 Euro on home-made.

SNP discrimination is based on high fidelity of thermostable ligase. Ligase-discrimination allows analyzing a large number of loci simultaneously in one tube. This makes the method high-throughput and decreases the amount of genomic DNA required for one analysis. During the ligation the biallelic state of the SNP locus is converted into a bimarker state of ligated detector oligonucleotides. The state of the markers is then determined by a 5'-nuclease assay (TaqMan) with universal fluorescent probes.

The elaborated method doesn't need any SNP-specific optimization. The procedure involves specially designed oligonucleotides, which provide equal amplification of all ligated products during amplification step and allow using different detection schemes (closed-tube end-point detection; microarray-based analysis; real-time quantitative assay). The acceptable setup price (50$ per one locus ‡ enough for ~100,000 reactions) was achieved due to a new cost-effective method of synthesis of oligonucleotides with block structure (patent applied).

To select a basis for technology well suited for any plant laboratory we have tested a number of different approaches: minisequencing and arrested primer extension on microarrays; combination of allele-specific PCR with microarray hybridization; arrested primer extension on genomic DNA and detection of extended products by hybridization on microarrays. However, we had to stop our work in these directions while all the schemes mentioned had fundamental restrictions making them inadequate for routine genotyping.

The LDR-TaqMan genotyping approach was successfully applied also for genotyping of human DNA samples. A kit for 9 clinically important human SNPs was prepared and used for determination of allele frequencies of these SNPs in two East European populations.

Now we are working on miniaturisation and automation of the procedure and on increasing the number of loci analysed in a single reaction.

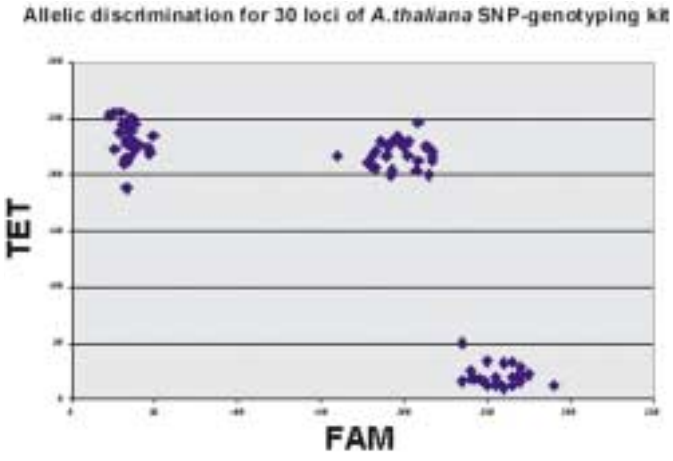


Fig 1: *Allelic discrimination in LDR-PCR method. The figure shows clear discrimination between homozygous and heterozygous loci.*

○ T.A. Borodina, H. Lehrach and A.V. Soldatov **Ligation detection reaction – TaqMan procedure for single nucleotide polymorphism detection on genomic DNA.** Anal Biochem. (2004) 333(2), 309-319. ○ A.M. Rickert, T.A. Borodina, E.J. Kuhn, H. Lehrach and S. Sperling **Refinement of single-nucleotide polymorphism genotyping methods on human genomic DNA: amplifluor allele-specific polymerase chain reaction versus ligation detection reaction-TaqMan.** Anal Biochem. (2004) 330(2), 288-97. ○ T.A. Borodina, H. Lehrach and A.V. Soldatov **Ligation-based synthesis of oligonucleotides with block structure.** Anal Biochem. (2003) 318(2):309-13.

Patents Ligation-based synthesis of oligonucleotides with block structure. PCT/EP 2004/003921 A.V. Soldatov, T.A. Borodina, H. Lehrach.
Ligation-based method of analysis of single nucleotide polymorphisms on genomic DNA. EP03019521 A.V. Soldatov, T.A. Borodina, H. Lehrach.

Genomics, Expression and Membranes: Identification and functional analysis of transport proteins

Arabidopsis
Protein Community

1 Reinhard Kunze, 2 Marcelo Desimone, 1 Anja Schneider, 1 Rainer Schwacke, 2 Anja Schmidt
2 Giorgio Gonella, 1 Eric van der Graaff, 1 Ulf-Ingo Flügge, 2 Wolf B. Frommer
1 University of Cologne, Department of Botany II
2 University of Tübingen, Department of General Genetics

As plant cells are highly compartmentalized, solute transport across the plasma membrane or intracellular membranes (e.g., mitochondria, plastids, vacuole) constitutes an important step in many biochemical pathways. In addition to transport, the membrane proteins involved often exert critical control and regulatory functions, that determine the metabolic state and influence the development of the plant. Thus, transport processes also play an important role in the expression of quality and the productivity of crop plants.

This project aims to identify and classify all transporters and membrane anchored sensor proteins of the model plant *Arabidopsis thaliana*, to functionally characterize a subset of these proteins, and to determine their expression regulation. As Arabidopsis belongs to the Brassicaceae family, novel insights gained in this project can be directly extended to crop plants from the same family (e.g. rapeseed, cabbages, mustard) and at least partially also to other plant families.

The project is performed cooperatively by the groups of U.-I. Flügge and W. B. Frommer. By parallel application of different strategies, distributing the experimental work between the two groups and combining the results, strong synergistic effects are achieved. Figure 1 gives an overview of the approaches used in this project. More detailed information are provided on our webpages: www.uni-koeln.de/math-nat-fak/botanik/bot2/agflue/HOME/projects/GABI_rkunze/index.html

Gene isolation by yeast complementation (Fig 1A)

By disruption of characterized transport proteins in yeast, strains are developed that are deficient in uptake of particular and important metabolites. Subsequently, selective growth conditions are established. Yeast screening systems for more than 30 different compounds (amino acids, nucleobases, vitamins, and hormons) have been established and optimized. The mutant strains are transformed with a Arabidopsis cDNA library in a yeast expression vector.

Several new Arabidopsis genes were identified by complementation, including three novel gene families: the ureide permeases (UPS) (Desimone *et al.* 2002), the mitochondrial basic amino acid carriers (mBAC) and the mitochondrial succinate-fumarate carriers (mSFC) (Catoni *et al.* 2003).

AtUPS1 was identified by complementation of an allantoin-uptake deficient yeast strain (Figure 2A). AtUPS1 is a highly hydrophobic protein with ten membrane spaning domains and a consensus sequence for a "Walker A" motif (Figure 2B) and the first member of the UPS family (Figure 2C).

RNA expression profiling (Fig 1B)

Novel transporters and other membrane proteins, that are up- or down-regulated in response to challenging growth conditions, or as a consequence of mutations in other transport processes, are traced by expression profiling. Specialized membrane protein macroarrays and full genome microarrays are used for screenings and more detailed analyses. We developed macroarrays by identifying all approximately 2000 Arabidopsis proteins with more than three membrane spans using the ARAMEMNON database (see below). Gene-specific cDNA fragments were cloned, amplified and printed on nylon membranes.

To this end, expression profiling experiments have been performed with knock out (KO) mutants of seven transporters and three growth condition gauntlets. The results show that in the KO mutants only very few (if any) other membrane proteins are deregulated, whereas changes in growth conditions can result in a significant number of deregulated transporters. For example, depletion of iron in the germination medium results in 47 down- and 16 upregulated (putative) transporters.

Transport activities and specificity measurements by electrophysiology (Fig 1C)

Electrophysiology of oocytes expressing plant transporters can be used to determine transport mechanisms and substrate specificities. More than 30 selected transporters from different protein families were expressed in *Xenopus laevis* oocytes. To confirm the membrane association, five of these proteins were additionally expressed as GFP-fusions. Approximately one third of the expressed proteins induce measurable currents. Detailed studies were performed on selected transporters.

Assays with six members of the AAP family revealed that none of the proteins has a strict specificity for a single amino acid. Each protein transports a variety of amino acids, however, the relative transport rates for individual amino acids differ considerably (Fischer *et al.* 2002).

For the allantoin transporter UPS1, the transport mechanism was determined. UPS1 co-transports allantoin and a proton across the membrane (Desimone *et al.* 2002).

For the AMT1 ammonium transporters it was demonstrated that they function as uniporters (Ludewig *et al.* 2002).

Bioinformatics (Fig 1D)

We have developed a novel database (DB) for *Arabidopsis thaliana* membrane proteins, ARAMEMNON, which integrates features that are presently only available from separate sources, and thus greatly facilitates the interpretation of gene/protein sequence data (Schwacke *et al.* 2003).

Several specialized membrane protein databases are accessible on the web, that use one or two algorithms for transmembrane span (TM) prediction. As predictions by different programs frequently disagree, a major objective of the ARAMEMNON DB is to provide TM predictions generated by different computation programs and a comparative graphical display.

ARAMEMNON also provides subcellular localization predictions by signaling peptide recognition programs. Another objective is to identify related proteins and determine the structure of the corresponding protein families. The major features of the ARAMEMNON database are:

- Complete membrane protein collections from *Arabidopsis thaliana* and Synechocystis (additional membrane protein collections from rice, yeast and mammals are in progress)

○ W. Fischer *et al.* **Low and high affinity amino acid H+ cotransporters for cellular import of neutral and charged amino acids.** The Plant Journal (2002) 29, 717-732. ○ Ludewig *et al.* **Uniport of NH4+ by root hair plasma membrane ammonium transporter LEAMT1.** The Journal of Biological Chemistry (2002) 277, 13548-13555. ○ M. Desimone *et al.* **A novel superfamily of transporters for allantoin and other oxo derivatives of nitrogen heterocyclic compounds in Arabidopsis.** The Plant Cell (2002) 14, 847-856. ○ R. Schwacke *et al.* **ARAMEMNON: a novel database for *Arabidopsis thaliana* integral membrane proteins.** Plant Physiology (2003) 131, 16-26. ○ E. Catoni *et al.* **A potential role of the mitochondrial succinate-fumarate translocator in lipid mobilization of seedlings and ethanolic fermentation in pollen of Arabidopsis.** FEBS Letters (2003) 534, 87-92.

- Transmembrane span predictions by seven programs and comparative graphical representation (additional predictions in progress)
- Subcellular localization predictions by eight programs (additional predictions in progress) and experimental data (if available)
- Display of related paralogs and orthologs
- Display of all genes in the chromosomal neighborhood
- Display of gene families ("clusters", very similar to neighbor joining trees)
- Display of sequences (protein, predicted ORF, experimental full length cDNA, genomic DNA, 5' and 3' UTRs)
- Bibliography with links to PubMed (including papers not related to sequence data)
- Improved annotation (including protein names from publications)

The ARAMEMNON database is accessible at http://aramemnon.botanik.uni-koeln.de

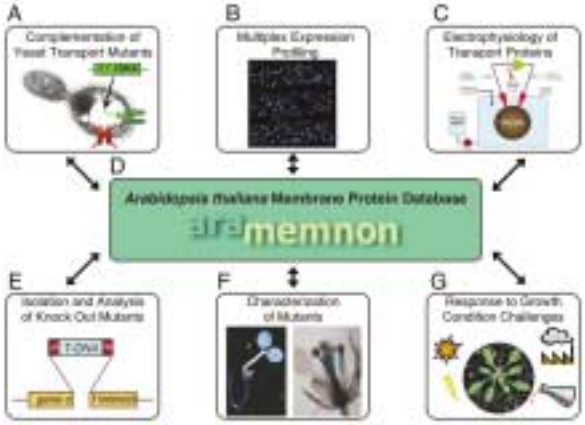


Fig 1: *Overview of the project. Novel putative transporters are identified by sequence analysis (D) and complementation of yeast mutants (A), and characterized by mutant analysis (E-F), electrophysiology (C) and expression (B) in response to growth conditions (G).*

Isolation of knock-out mutants (Fig 1E)

The phenotypic and physiological characterization of loss-of-function mutants is a crucial step in determining the role of the respective proteins. We are primarily interested in isolating mutants of orphan transporters that were identified

- by bioinformatical analysis using the ARAMEMNON database as a tool,
- by expression profiling to be deregulated in response to challenging growth conditions like depriviation of nutrients or different kinds of abiotic stress,
- to have pronounced sequence similarity to known transporters in other organisms.

Mutant lines are obtained from GABI-Kat, the Salk Institute Genome Analysis Laboratory (SIGnAL), or by screening our local resources. We have built up an in-house resource for conventional PCR-screenings of 57.200 T-DNA insertion mutants. If these resources do not recover a mutant for the gene under investigation, we also use the screening service offered by the Arabidopsis Knockout Facility (AKF) at the University of Wisconsin.

The mutant lines are being subjected to phenotypic analyses under different growth conditions. In summary, to date

- 224 T-DNA insertion lines for (putative) transporters and other membrane proteins have been obtained,
- 56 lines are genotypically verified,
- 12 lines are homozygous lethal,
- 44 lines are phenotypically analyzed and subjected to growth condition gauntlets.

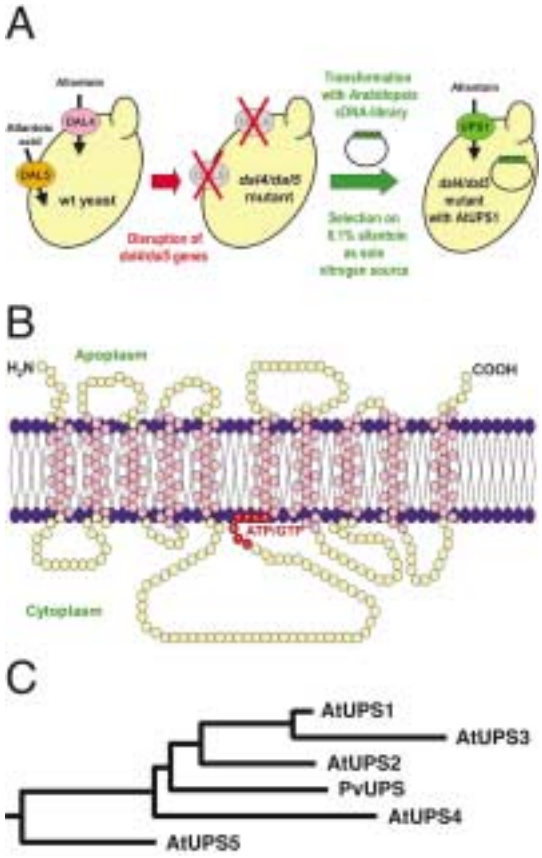


Fig 2: *(A) Strategy to isolate Arabidopsis ureide permeases by yeast complementation. (B) Structure of the AtUPS1 protein. The "Walker A" motif residues are marked in red. (C) The AtUPS family consists of five proteins. Close relatives are found in many other plants.*

Transcription factor function search: How do individual factors regulate agronomical important processes in plants? (Subproject A)

Arabidopsis Protein Community

Isabell Witt, Maria Ines Zanor, Bernd Müller-Röber
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

Goal

Transcription factors (TFs) regulate the expression of downstream target genes and thereby contribute to the establishment of complex traits in higher plants, including developmental features, cell differentiation, biosynthetic pathways and adaptation to environmental stresses. The evolution of higher plants is tightly coupled to TF evolution and the high TF number in plants places them on the same level of complexity with e.g. Drosophila. Several examples demonstrate already the large biotechnological potential of plant TFs for the modification of traits. The aim of this GABI sub-project is to establish a knowledge basis on *Arabidopsis thaliana* TFs for future use in agriculture and plant biotechnology. In this part of the project the function of a subset of 200 of the about 1700 plant TF genes will be analysed. This subset is defined by the type of the DNA-binding domain of a given TF. To assign functions to single members of the large TF families, over-expression (OX) lines will be generated and knock out (KO) lines will be identified. TF function will be deduced from visible phenotypes, the identification of target genes, and interacting factors. The development and application of suitable bio-informatics tools will allow integration of and comparisons with the data obtained by Gauntlets groups.

Project Status

The sub-project started 01.08.2000 and since then TFs were selected based on different criteria:

- The plant specific Dof type TFs (37 members) were chosen for functional analysis, because there was already research done in the lab of applicant on some of these factors before this project started. 32 selected.
- TFs of the plant specific AP2/ERF family (> 100 members) are expressed in a variety of tissues and are involved in stress responses and various developmental processes. 54 selected.
- NAC TFs (>100 members) represent a plant specific family, that has not been described extensively in the literature. NAC TFs have been shown to be involved in pattern formation, organ separation and other developmental processes. 25 selected.
- The entire SCARECROW family (26 members) was selected, because several members of this family were altered in their expression level in + and – sulphur experiments.
- 15 TFs from different families were differentially expressed in + and – sulphur experiments. 38 TFs were significantly induced or reduced after treatment with different chemicals. 10 additional factors harbour C2H2 motifs.

Coding regions of TF genes were isolated from first-strand cDNA pools from different tissues and cloned behind the 35 S promoter into pGreen0229. Transgenic A. thaliana lines were identified via growing T1 seeds on soil followed by BASTA™ spraying for selection on seedling stage plantlets. Between zero and > 100 resistant plants were recovered after BASTA™selection, depending on the expression of a particular TF. During the 29 months since the project started 123 TF OX plants were created and T1 lines selected after BASTA™ spraying. In addition 37 constructs with the inducible Top10 promoter (Love *et al.* 2000) are in planta. Transgenic lines expressing elevated levels of transcription factors were identified by Northern blot analysis. More than 45 different transgenics were tested for over expressed TFs by this method. Whenever the level of the over expressed TF is well detectable on the Northern blot, the corresponding visible phenotype is usually clearly deviating from control plants carrying the empty vector.

To identify changes in transcript levels in transgenic plants over expressing an individual TF, Affymetrix GeneChips™ were used. The Dof TF family was investigated first. In Arabidopsis it has 37 members sharing a highly conserved cysteine-rich motif for a single zinc finger. Members of this family have been analysed in maize, barley, tobacco and potato, but only few Dof TFs such as OBP1-3, DAG1 and DAG2, were investigated in Arabidopsis yet. The aim of this project is to identify the downstream genes of selected Dof transcription factors in Arabidopsis and to integrate them with the visible phenotypes including altered leaf and flower morphology. Based on protein similarities the Dof family was subdivided into eight subfamilies. Arabidopsis lines were created over expressing 31 different Dof TFs representing these subfamilies. Considering their visible phenotype and their position on a distance matrix tree, different lines were selected representing nine Dof TFs that were investigated on GeneChipsTM. Using the pair wise comparison tool of the Affymetrix software Microarray Suite 5.0, we were able to detect significant differences in the expression level between the transgenic probes and wild type plants. Preliminary evaluation of the data showed that some of the Dof TFs share common downstream genes suggesting partially overlapping functions, but also unique downstream genes that seem to be uniquely changed for the corresponding Dof TF. A summary for uniquely altered genes observed in the frame of our experiments is presented in table 1. In a specific example, a transgenic line overexpressing a TF developed leaves with small blades and extended petioles, elongated internodes, and an increased number of side branches when compared with the wild-type. Expression profiles revealed differential expression of many genes primarily related to abiotic stress, cell wall, cell scaffold and hormones that could be linked to enhanced abiotic stress tolerance (drought, oxidative stress) and to the altered leaf morphology. Various stress treatments revealed that these transgenic lines indeed have an elevated stress tolerance when compared to wild type plants.

The detection of "true" knock-out alleles in which the allele can surely

be predicted to be "NULL" is crucial to analyse gene function. The GABI-KAT resource has implemented sequence-based selection of knock-out lines and has proven to yield large numbers of T-DNA tagged lines which are immediately available without additional strings attached. 24 KO lines for TF genes have been ordered from the GABI-KAT collection. For 21 KO lines the corresponding OX plants are available within the project. Depending on the needs of the genetic material, lines will be backcrossed, co-segregation will be confirmed and they will then be analysed in detail. 18 lines are currently under investigation by Southern blot analysis. As access to KO lines is not limiting any more, the RNAi approach will only be applied for genes of outstanding interest.

Collaborative work

There is close scientific exchange with the group of Bernd Weißhaar that is working on one of the two GABI TF sub-projects. In several seminars in Golm and Cologne the progress of the two sub-projects was discussed co-ordinated and information and methods were exchanged. Expression profiling data were sent to the GABI Primary Database for later use by other GABI members. Data processing is done in collaboration with GABIinfo. To develop an effective and bio-informatically robust pipeline for data analysis, the expression profiling data were sent to Sabine Tornow (AG Mayer at MIPS). Comparison of data mining software for expression profiling data was started. Our data were processed with a program called Coupled Two Way Clustering. Cluster analysis revealed that e.g. in TF 88 many photosynthetic genes are down regulated which is in accordance with the stress phenotype. 18 Dof cDNA clones have been delivered to GABI-LAPP (B. Kersten, D. Cahill) to include the encoded proteins on the protein chip after expression in E. coli. An antibody against TF88 was produced and tested positively on the protein chip

In order to obtain as much information as possible from the transgenic plants, selected lines of special interest are analysed on the metabolite level. Dr. Ni Di-an (DAAD) in collaboration with Joachim Kopka at the MPI-MP gained preliminary results and found about 20 significantly altered metabolites including an increase in proline, that is known to be elevated in stressed plants.

Joachim Selbig and co-workers are working on possibilities to integrate Affymetrix data into the in-house expression profiling database at the MPI-MP in Golm.

Outlook

- Cloning of the remaining TF genes.
- Detailed analysis of the expression profiling data.
- Further proof of biological relevance of the data derived from GeneChipsTM experiments.
- After proof of biological relevance TF over expression in different crop plants.
- For selected lines of major interest integration with metabolite profiling data and promoter studies.

Fig 1: Plants overexpressing Dof TF88 are most resistant to oxidative stress *Plants treated with MV (paraquat) + high light.*

Tab 1: *shows the uniquely regulated genes of nine Dof TFs over expressing transgenics after evaluation of Affymetrix GeneChip results. Genes highlighted in red are up-regulated, genes highlighted in blue are down-regulated genes. Highlighted boxes represent functional categories that have significantly higher numbers of genes than found in other lines and thus seem to be characteristic for the corresponding Dof TF.*

Transcription factor function search: How do individual factors regulate agronomically important processes in plants? (Subproject B)

Arabidopsis
Protein Community

Gieta Dewal, Martin Sagasser, Illona Zimmermann, Bernd Weisshaar
Max Planck Institute for Plant Breeding Research (MPIZ), Cologne

Goal

Transcription factors (TFs) regulate the expression of downstream genes and thereby contribute to the establishment of complex traits in higher plants, including developmental features, cell differentiation, biosynthetic pathways and adaptation to environmental stresses. The evolution of higher plants is tightly coupled to TF genes, and several examples already demonstrate the large biotechnological potential of plant TFs for the modification of traits. The aim of this GABI subproject was to establish a knowledge basis on *Arabidopsis thaliana* TFs for future use in agriculture and plant biotechnology. In the cooperation project between Müller-Röber and Weisshaar, the function of 300 of the about 1700 plant TF genes was analysed. 100 of these genes are investigated in Köln (Weisshaar). The subset is defined by the type of the DNA-binding domain of a given TF. To assign functions to single members of the large TF families, overexpression (OX) lines and knock-out (KO) lines were used. TF function will be deduced from visible phenotypes of the OX and KO lines, the identification of target genes, and interacting factors. The ultimate goal of the project is to use the generated data to identify TF genes which have an impact on agronomically important traits.

Selection of TF genes and isolation of TF cDNAs

To select genes for the "MPI-Z" 100-gene subset, *A. thaliana* genomic sequence and EST data were used to identify "new" TF genes on the basis of the existence of MYB, bZIP, bHLH, and WIP (zinc finger type) DNA binding domains. Cluster analyses were performed for the entire gene families of interest. These analyses were based on previous results on MYB genes (Stracke *et al.* 2001) and parallel EC-funded work in the group on bZIP genes (Jakoby *et al.* 2002), and benefited from our experience with these TF gene families. Results on bHLH gene family were published recently (Heim *et al.* 2003). Based on the clustering, genes were selected for further characterization. Criteria applied were e.g. to include all members of small subgroups, or putative homologues to factors of known function from other species. In addition, information on genes under investigation in other international laboratories was considered in an attempt to concentrate on genes offering good chances to generate new information. Decisions about TF genes to be included into the 100-TF-gene subset were made in agreement with the group of B. Müller-Röber at MPI-MP. Table 1 lists the selected TF genes.

The 100 full-length coding sequences (cDNAs with complete open reading frames) for these genes were produced by PCR techniques (5'- and 3'-RACE). Basic information on the expression pattern of any given gene was obtained by the parallel use of RT-PCR templates from 14 different sources (root, leaf, stem, flower, silique, cell culture, and cell culture treated with cycloheximide, UV, flagellin elicitor, salicylate, 1-aminocyclopropan-1-carboxylic acid, methyljasmonate, 4°C or 37°C respectively). The TF cDNAs were inserted into the pDONR plasmid (Invitrogen) to yield ENTRY clones suitable for the GATEWAY system to be applied in subsequent project phases. All of the clones were confirmed by sequencing.

Generation of constructs for systematic Yeast-two-hybrid screens

A total of 82 TFs of 4 different TF families have been tested with the Y2H-system. These experiments showed that about 74% of these TFs can act as transcriptional activators of the reporter gene when fused to the BD. 26% can be used for Y2H-screens. Usage of the low copy Cen based plasmids did not prevent autoactivation. The SOS recruitment system ("Cyto-Trap") was not suitable for our purposes. The only system which did work in principle was the split ubiquitin system (Johnson *et al.* 1994), however, this system still needs to be adopted for higher throughput.

We performed 38 Y2H screens resulting in 484 different candidates. In this screen 1-43 million zygotes were tested per screen and three libraries were examined. Blast results of all andidate genes showed that 65% are known proteins. 10% of the candidates are RNA/DNA binding proteins eg. MYB-, bHLH-, bZIP- and MADS BOX proteins. In a matrix approach interesting candidates were tested against activation domain fused TFs which could not be used as a bait in the Y2H. Interesting candidates will be studied in future.

Overexpression lines and KO Lines

The system used for the construction of overexpression lines was initially based on pGPTV binary vectors as described in Sagasser *et al.* (2002). To overcome the restrictions of this low copy vector and to also be able to apply the GATEWAY technology to this part of the project, a new vector had to be chosen. This new plasmid, designated pJAN33, is based on the high copy vector pPAM (GenBank acc. no. AY027531). It confers kanamycin resistance and allows GATEWAY based cloning of a given cDNA (ORF). The cDNA is expressed under control of a double CaMV 35S promoter and there is an intron between transcription start and the cloned cDNA. KO Lines were mainly isolated from GABI-KAT but also from other resources for insertional mutagenesis. An overview on all lines is given in table 1.

Phenotypic analysis of KO lines and overexpressors

Results from the analysis of MYB KO and overexpressor lines show that obvious phenotypes are the (rare) exception. Additional information e.g. on the expression pattern, interacting partners, function of homologues in other species or other members of the same subgroup of the TFs, will be used to come up with an "educated guess" on the function of a given TF. Based on this assumption, specific features of the mutants / overexpressors will be analysed in comparison to wild type plants. The analysis of the 100 + 32 = 132 lines generated in the first phase of the project has just started.

○ R. Stracke, M. Werber and B. Weisshaar **The R2R3-MYB gene family in *Arabidopsis thaliana*.** Curr. Opin. Plant Biol. (2001) 4, 447-456. ○ M. Jakoby *et al.* **The bZIP family of transcription factors in *Arabidopsis thaliana*.** Trends Plant Sci. (2002) 7, 106-111. ○ M. A. Heim, M. Jakoby, M. Werber, C. Martin, B. Weisshaar and P. C. Bailey **The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-wide Study of Protein Structure and Functional Diversity.** Molecular Biology and Evolution (2003) 20, 735-747. ○ M. Sagasser *et al.* **A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins.** Genes & Dev. (2002) 16, 138-149. ○ I.M. Zimmermann, M.A. Heim, B. Weisshaar and J.F. Uhrig **Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins.** Plant Journal (2004) 40, 22-34. ○ P.C. Bailey, C. Martin, G. Toledo-Ortiz, P.H. Quail, E. Huq, M.A. Heim, M. Jakoby, M. Werber and B. Weisshaar **Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana*.** Plant Cell (2003) 15, 2497-502.

Summary and Outlook

The first half of the funding period of this sub-project was used successfully to establish a profound basis to generate the desired data until the end of 2003. Great efforts have been made to obtain error-free cDNA clones comprising 100 full-length ORFs of the mostly low-expressed TFs. First of all, this work allowed to update gene models and to learn more about relationships among genes in the families under study. The results also showed, as discussed by the MIPS/GABI-Info annotation group, that gene models not based on cDNA sequence (or EST) data are especially error-prone. Second, this effort was justified since the GATEWAY technology established within the project now allows application of the established cDNA constructs in many "downstream" experiments, which all benefit from the high quality of the clone set. This includes easy construction of OX lines and Y2H constructs without the necessity for re-sequencing. Third, the expression data generated by the RT-PCR approach taken will serve as the basis to select experimental conditions for the analysis of OX and KO lines.

At present, the first results in terms of interesting OX and KO phenotypes are upcoming. These data will need confirmation, and supplementary data will be generated during the year 2003 for the genes currently under study. In addition, phenotypic differences to wildtype in the agronomically most interesting areas (response to pathogens, stresses, nutrient availability, yield etc.) are probably more subtle and are only detectable under specialised experimental conditions.

Gene family	number	FL cDNA clones	OX Lines	KO Lines
R2R3-MYB	39	39	39	9
MYB-3R	5	5	5	2
MYB-1R	3	3	3	
bZIP	16	16	16	5
bHLH	31	31	31	14
WIP (ZF)	6	6	6	2
total	**100**	**100**	**100**	**32**

Tab 1: *The table shows the TF genes selected for the analysis in this project, the full length (FL) cDNA clones, the overexpression (OX) lines and knock-out (KO) lines produced or identified.*

ectopic overexpression of AtMYB23, glabrous phenotype *wildtype plant*

Fig 1: *Example for first upcoming overexpression phenotypes. The shown line is a MYB23 overexpression line.*

Carbon and Nutrient Signalling: Test Systems, and Metabolite and Transcript Profiles

Arabidopsis
Gauntlets

Mark Stitt
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

This project investigates the molecular mechanisms that regulate metabolism, growth and development of plants in response to the availability of carbon, nutrients, oxygen and low temperature. The general strategy is to develop a series of defined growth conditions, or 'gauntlets' (German: 'Spießrutenlauf'). These are used to screen genetic diversity and discover genes by forward genetics. They also provide experimental systems in which multilevel analysis is performed, to characterize system responses and identify key genes that can subsequently be subjected to functional analysis. This report, which was written 34 months into a 48 month program, outlines the technology platforms that have been developed and gives selected examples from the results obtained so far.

Growth in 'gauntlets' to make the plant talk

'Gauntlets' are standardised high-throughput growth conditions, which display phenotypes that are generated by defined challenges. The phenotype, which should where possible be quantitative, may be visual or involve changes of metabolites that can be detected by a high throughput platform. 'Gauntlets' are used to screen for mutants, to identify natural variation in ecotypes (see Arabidopsis Verbund I), to characterize mutants provided by other members of the Arabidopsis GABI community, and to define suitable conditions for detailed metabolite, transcript and enzyme activity profiling. 'Gauntlets' have been developed for responses to the C, N and P supply, low temperature and hypoxia. Typically, a suite of complementary 'gauntlets' is developed. Some involve simplified treatments that rapidly affect the entire plant (e.g. the effect of an abrupt change on seedlings in liquid culture). Others involve more complex system like growth in defined conditions on sterile agar plates, often orientated vertically to score root growth rates and root architecture. An example is shown in Figure 1, which schematically depicts a gauntlet to detect genotypes that are altered in the ability to store resources to cope with a long night. Other 'gauntlets' address the impact on reproductive growth. Where possible, complementary 'gauntlets' are developed that allow a complex response (e.g., the response to low nitrogen) to be dissected (e.g. to separate the specific effects of nitrate as a signal from the general effects of a change in the nitrogen supply).

Listening for the answer: multilevel profiling of transcripts, enzyme activities and metabolites.

Transcript profiling is carried out with commercial near-full genome 22K Affymetrix arrays. Depending on the tissue used for the experiment, 14-18,000 genes are called positive. This approach is complemented with a real time RT-PCR platform, which has been developed in house by M. Udvardi and W.-R. Scheible and detects 1500 of the ca. annotated 1600 transcription factors in Arabidopsis. Less than 600 are detected in the same extracts using 22K Affymetrix arrays.

As enzyme activities can be measured with high precision, they provide an interim alternative to quantitative proteomics. About 40 activity tests for enzymes involved in photosynthesis, sucrose and starch synthesis, sucrose degradation, primary nitrogen metabolism, amino acid synthesis and degradation, nucleotide synthesis and lipid metabolism have been developed and to ELISA format and a pipetting robot. Enzyme activity profiling is used as a strategy to analyse genetic diversity in ecotypes, F1 hybrids, RIL's and NIL's (see Arabidopsis Verbund I), and in combination with transcript arrays for multilevel profiling. The analytic platform is illustrated in figure 2. One of the advantages of this approach is that it can be used for crops when sequence information is not available to support proteomics.

Metabolite profiling is an essential component of our phenotyping platform as well as being an important role in experimental design (1). There are several components in the platform: Assays for sugars, starch, protein, chlorophyll, nitrate, key phosphorylated intermediates and amino acids are performed on ELISA plates using a pipetting robot to allow large numbers of samples to be processed, e.g., during the development of new 'gauntlets'. Automated HPLC analysis of nucleotides, amino acids and phenylpropanoids also allow large numbers of samples to be analysed in parallel. GC/MS is used for unbiased analysis of a wide range of metabolites. Highly sensitive enzymecycling assays allow measurement of every important metabolite between sucrose, starch and lipid biosynthesis in small samples like growing Arabidopsis seeds. They are based on new cycling assays for ATP, ADP, ADPGlc, UDPGlc, PPi, and many glycolytic metabolites, as well as optimised protocols for NAD(H), NADP(H) and acetyl-CoA (2). As metabolite analysis is essentially genotype-independent, this complete platform can be used in combination with crops.

To support analysis of transcript profiles, a tool has been developed in collaboration with the RZPD to allow automatic display of expression profile data onto customer-made sketches of pathways and cellular functions (see note added in press). This tool is being extended to display enzyme and metabolite data. Figure 3 shows an example of the interface seen by the user, displaying in this case 1000's of genes involved in metabolism on to a simple sketch that divides metabolism into over 100 different hierachicial functional areas. Many other diagrams are also available, that that resolve metabolism to the level of the individual enzyme in pathways, as well as high level schemes that depict the response of genes assigned to general areas of cellular function (fig 4).

○ 1. M. Stitt and A. R. Fernie **From Measurements of Metabolites to Metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions Current Opinion** in Biotechnology (2003) In Press. ○ 2. Y. Gibon *et al.* **Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system.** The Plant Journal (2002) 30, 221-235. ○ 3. M. Stitt *et al.* **Steps towards an integrated view of nitrogen metabolism. Journal of Experimental.** Botany (2002) 370, 959-970. ○ 4. V. Hurry, R. Furbank and M. Stitt **The role of inorganic phosphate in the development of freezing tolerance and the acclimation of photosynthetic carbon metabolism to low growth temperature is revealed by studies of pho mutants of *Arabidopsis thaliana*.** The Plant Journal (2000) 24, 383-396. ○ V. Hurry, R. Furbank and M. Stitt **The role of inorganic phosphate in the development of freezing tolerance and the acclimation of photosynthetic carbon metabolism to low growth temperature is revealed by studies of pho mutants of *Arabidopsis thaliana*.** Plant J. (2000) 24, 383-396. ○ Y. Gibon, H. Vigeolas, T. Tiessen, P. Geigenberger and M. Stitt **Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system.** Plant J. (2002) 30, 221-235. ○ M. Stitt, C. Müller, P. Matt, Y. Gibon, P. Carillo, R. Morcuende, W.-R. Scheible and A. Krapp **Steps towards an integrated view of nitrogen metabolism.** J. Exp. Bot. (2002) 370, 959-970. ○ M. Stitt and V. Hurry **A plant for all seasons: Arabidopsis as a model system to investigate the cold acclimation of photosynthetic carbon metabolism.** Curr. Opinion Plant Biol. (2002) 5, 199-206.○ M. Stitt and A.R. Fernie **From Measurements of Metabolites to Metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions.** Curr. Opinion in Biotech (2003) 14, 136-145. ○ J.H.M. Hendriks, A. Kolbe, Y. Gibon, M. Stitt and P. Geigenberger **ADP-glucose pyrophosphorylase is activated by post-translational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species.** Plant Physiol. (2003) 133, 838-849. >

Unravelling what plants do when carbohydrates are exhausted

Many metabolic and developmental processes in plants are regulated by sugars. The signalling pathways remain controversial, however, as most previous studies have concentrated on the effect of high exogenous sugar on germinating seeds. Our aim is to learn how physiological changes modify agronomically-relevant traits.

We have established 'gauntlets' that display phenotypes that are generated by changes in the sugar supply. One example is the transfer of plants from a long day to a short day light regime (fig 1). Carbohydrates especially starch accumulate in leaves during the day and are remobilised to support continued metabolism during the night. Most of the starch is consumed at the end of the 'normal' night. When the night is lengthened, the residual starch is exhausted within 2-4h and sugars fall to low levels. Seedlings growing on vertical plates and root growth were switched from long to short day conditions, and root growth was scored to uncover variation between ecotypes in their ability to adjust to this stress. We have also used this experimental system to identify genes involved in sensing and adjusting to a shortfall of carbohydrate. There is a diagnostic decrease of glycolytic intermediates during the first 4h of the extended night, which is then reversed as alternative sources of carbon are tapped (e.g., remobilisation of amino acids from proteins). This temporary imbalance between carbon and nitrogen metabolism is corrected within two diurnal cycles, because starch synthesis is increased to store more carbon store for use during the long night. Transcript profiling at five time points spanning the transient decrease and recovery of glycolytic intermediates revealed that expression of about 800 genes is changed. Transcripts for enzymes involved in nitrate and sulphate assimilation decreased, and transcripts for some key enzymes in amino acid catabolism rose (fig 3). These were accompanied by changes in the activities of the encoded enzymes. Most importantly, we found changes in transcripts for many genes encoding transcription factors, protein kinases and signalling components very early in the transition, before glycolytic intermediates started to recover. These included changes of transcripts for specific members of gene families for trehalose phosphate synthase and trehalose phosphatase and for ethylene synthesis and ethylene responses. We are using knockout mutants to define their role in low sugar signalling.

Plants use nitrate as a signal when nitrogen runs out

Plants differ from most other organisms in that that nitrate is their major nitrogen source (3). We are interested in learning how metabolism, growth and development respond to changes in the nitrogen supply, and discovering which of the responses are triggered by signals derived from nitrate.

In one approach, plants are grown in 'gauntlets' with glutamine as a constitutive nitrogen source, in the presence of low or high nitrate. Metabolite analyses confirmed that in this system nitrate des not affect amino acid and protein levels. Low nitrate stimulates lateral root formation and growth. This is an adaptive response, which presumably promotes foraging and nutrient acquisition. Screening of T-DNA insertion lines identified

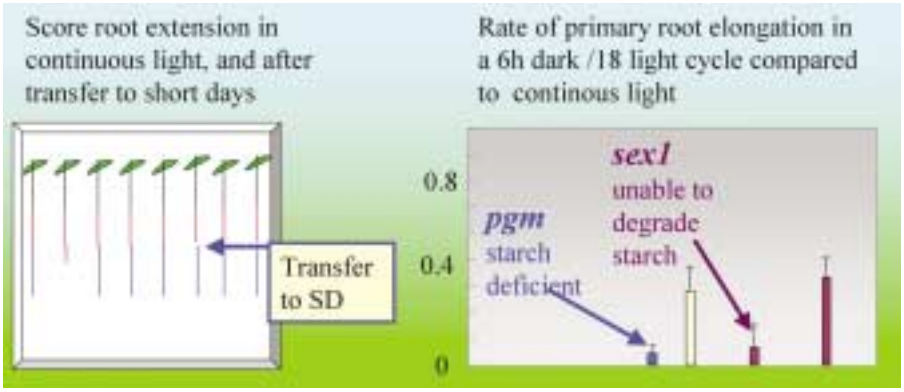


Fig 1: *‚Gauntlet' to identify genotypes that are impaired in their ability to store appropriate resources to survide a long night. This screen will identify mutants in biochmeical pathways needed to synthesise or mobilise important storage compounds (e.g., starch), as well as mutants imparied in sensing photoperiod or transient lack of resources.*

13 genes putatively involved in the regulation of root architecture, including several signalling and regulatory proteins. Low nitrate also accelerates the transition from vegetative to floral growth. This 'escape' strategy has obvious ecological and agronomic consequences. Mutants impaired in photoperiod, vernalisation and 'autonomous' pathways for floral induction still respond to nitrate. Experiments with mutants in the giberellic acid-dependent pathway are in progress to determine if nitrate acts via this or a completely novel pathway.

In a complementary approach we investigated changes of metabolites and transcripts after supplying nitrate to deficient plants. Except for an increase of nitrate there were no substantial changes of metabolites in primary carbon and nitrogen metabolism during the first 30min. About 500 transcripts changed within 30min, including genes for transporters and enzymes required for the uptake and assimilation of nitrate, and a significant number of genes that encode transcription factors, protein kinases and phosphatases, and signalling components. After 3h, there were marked changes in metabolite levels showing that nitrate was being utilised. The levels of minor amino acids did not rise, however, indicating that their use is stimulated by more direct signals. Between 30min and 3h, there was a general increase in the transcripts for many of the genes encoding enzymes in pathways that utilise nitrogen like amino acid synthesis, nucleotide synthesis, chlorophyll synthesis. Transcripts also rose for many genes required for RNA synthesis and processing, and especially protein synthesis where over 70% of all genes assigned to this functional class were induced. Knockout mutants in genes that rise after 30min are being used to define their role in this reprogramming of cellular function.

We are also analysing responses to low phosphate and sulphate. The aim is to develop a data base displaying the response of genes and specific gene family members to depletion and re-supply of each of these nutrients in a simple system in young plants, and to then extend it to a later stage in the life history when massive remobilisation of resources is occurring during seed filling.

Plants slow metabolism down to avoid driving themselves into anoxia

In contrast to large multicellular animals, plants lack a circulatory system and specialised pigments to transport oxygen. As a result, metabolically active or bulky plant tissues can become seriously hypoxic: for example oxygen levels fall to 2% in the centre of a growing potato tuber. As this problem has been not appreciated in the past, research into plant responses to

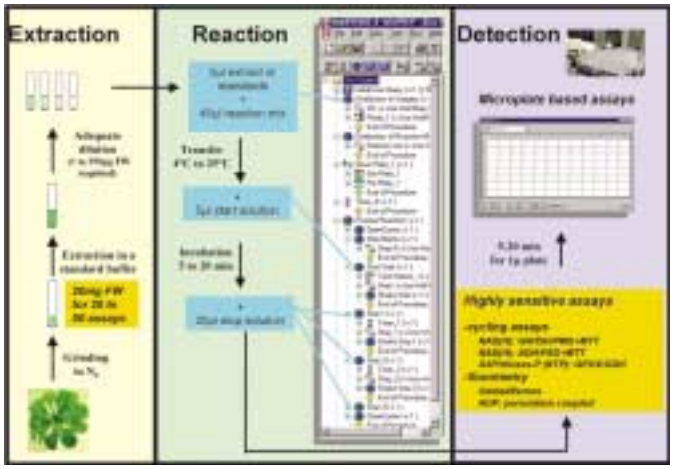


Fig 2: *Robot-based enzyme tests.*

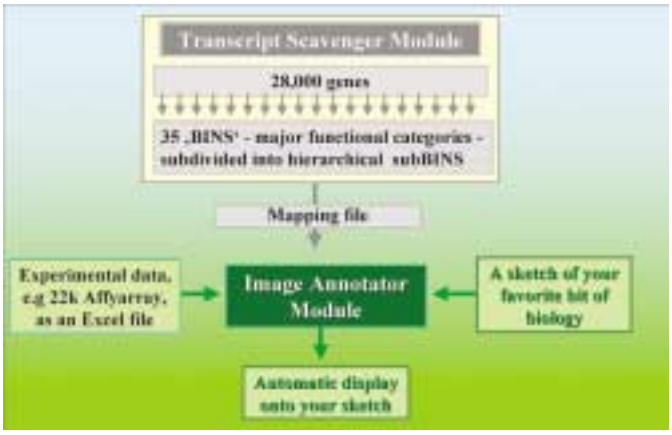


Fig 3: *MapMan.*

○ O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L.A. Müller, S.Y. Rhee and Mark Stitt **MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** Plant J. (2004) 37, 914-939. ○ T. Czechowski, R.P. Bari, M. Stitt, W.-R. Scheible and M.K. Udvardi **Real-Time RT-PCR profiling of 1500 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** Plant J. (2004) 38, 366-379. ○ J. Kopka, A.F. Fernie, W. Weckwerth, Y. Gibon and M. Stitt **Metabolite profiling in Plant Biology: Platforms and Destinations.** Genome Biology (2004) 5, 109. ○ Y. Gibon, O. Bläsing, N. Palacios-Rojas, D. Pankovic, J.H.M. Hendriks, J. Fisahn, M. Höhne, M. Günther and M. Stitt **Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars, and posttranslational activation of ADP glucose pyrophosphorylase in the following light period.** Plant J (2004a) 39:847-862. ○ W.-R. Scheible, R. Morcuende, T. Czechowski, D. Osuna, C. Fritz, N. Palacios-Rojas, O. Thimm, M.K. Udvardi and M. Stitt **Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes and signalling infrastructure by nitrogen.** Plant Physiology (2004) 136, 2483-2499. ○ Y. Gibon, O. Bläsing, J. Hannemann, P. Carillo, M. Höhne, J. Cross, J. Selbig and M. Stitt **A robot-based platform to measure multiple enzyme activities using a set of cycling assays: comparison of changes of enzyme activities and transcript levels in Arabidopsis during diurnal cycles and in prolonged darkness.** Plant Cell (2004b), in press. ○ E. Zuther, K. Buchel, M. Hundertmark, M. Stitt, D.K. Hincha and A.G. Heyer **The role of raffinose in the cold acclimation of Arabidopsis.** FEBS Letters: (2004) in press.

low oxygen has concentrated on the response to externally-imposed anoxia. We have characterized the responses of Arabidopsis root growth and seed metabolism (2) to falling oxygen. Growth and metabolism are slowed down in response to small changes in external oxygen, even in these very small organs. This represents an adaptive response, which decreases oxygen consumption and prevents internal oxygen levels from falling further. It occurs at much higher oxygen levels than those that induce fermentation. Current experiments are investigating the responses of transcripts to small changes in oxygen tensions to discover genes involved in the perception and transduction of signals that lead to slowing of metabolism and growth when internal oxygen levels fall in plants, and using reverse genetics to functionally analyse candidate genes prospectively involved in oxygen sensing and signalling in plants.

Cross talk between phosphate-signalling and low temperature acclimation

Exposure of plants to low but non-freezing temperatures induces an adaptive response that allows them to survive at subzero temperatures. Acclimation of photosynthetic metabolism to low temperature involves post-translational regulation of enzymes, changes in expression of genes for specific pathways and change in the overall protein concentration due to changes in the proportion of the cell occupied by the vacuole, allowing maintenance of high rates of photosynthesis and sucrose synthesis. Gauntlets' were established to identify genotypes blocked in these responses. These investigations uncovered an unexpected interaction with phosphate nutrition and signalling. Two mutants were identified that are blocked in post-translational activation of sucrose phosphate synthase at low temperatures. Unexpectedly, they were also defective in post-translational activation in low phosphate. Even more intriguingly, several aspects of cold acclimation were enhanced in pho1 and impaired in pho2 mutants, which have low and high shoot phosphate, respectively (4). We recently succeeded in molecular cloning of PHO2 by fine mapping, use of 22K Affymetrix arrays to identify a candidate whose transcript is strongly reduced in pho2, sequencing of the identified gene in wild-type plants and pho2, and complementation of pho2 by the wild-type gene. The sequence predicts that PHO2 is a component in a regulatory pathway, opening a new route into the functional characterization of phosphate signalling and low temperature responses in plants.

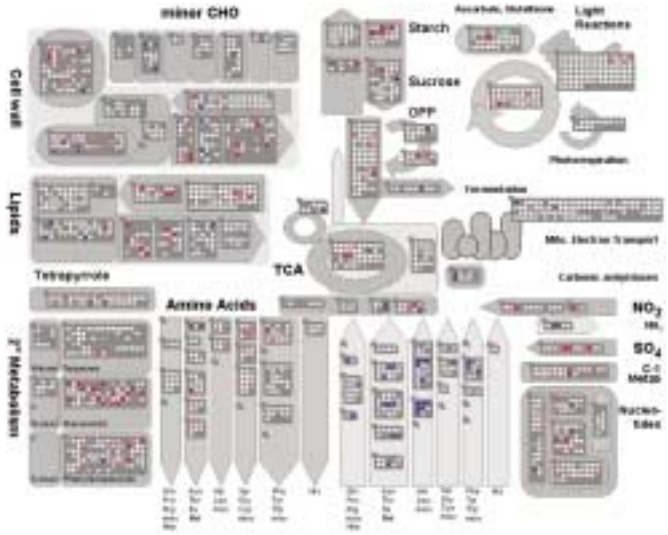


Fig 4: *Changes of genes involved in metabolism after extending the night for 6 h. Each gene assigned to a particular functional area is shown as a small square: blue denotes an increase, white no change and red a decrease. Genes called ‚not present' are shown grey, and cannot be distinguished from the background.*

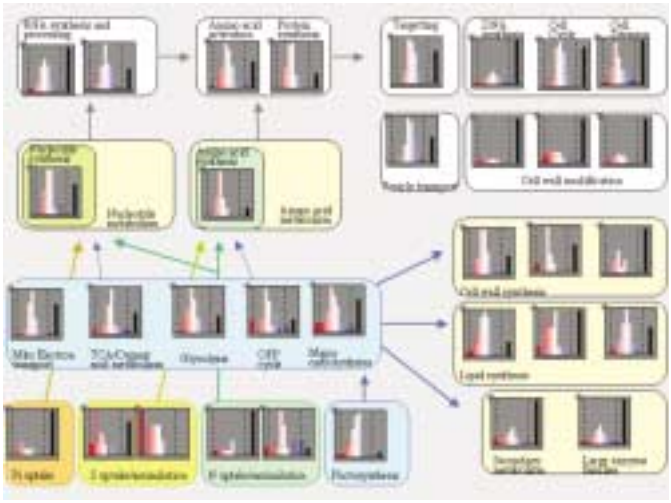


Fig 5: *Global overview of changes in expression of genes involded in biosynthesis and growth after 6 h of an extended night. All of the genes assigned to a given functional area are treated as a population and the change plotted as a frequency histogram. Blue denotes an increase and red a decrease. Genes called ‚not present' are shown as a grey bar.*

Analysing cell wall biosynthesis to study its role in biotic and abiotic stress reactions

Arabidopsis
Gauntlets

Nicolai Obel, Björn Usadel, Tze Siang Choo, Markus Pauly
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

Aim and objective

The cells of all higher plants are surrounded by a wall responsible for the structural integrity and shape of the cells, the transport of metabolites, and defence reactions against pathogens. Plant cell walls are not static but dynamic entities that undergo metabolism particularly during the processes of cell elongation and cell differentiation (Carpita & Gibeaut 1993). Plant cell walls also represent a major renewable resource that is used extensively by human mankind in e.g. the food, feed, textile and paper industry.

Cell walls consist of numerous co-extensive networks. The main constituents are cellulose microfibrils embedded in a hydrated matrix composed of various hemicelluloses and pectic polysaccharides. Although much is known about the complex structures of the cell wall constituents, very little knowledge exists about its biosynthesis and the relationship of the structural features to their functional properties. The identification of genes responsible for wall polysaccharide synthesis will allow engineering of wall polymers opening an unprecedented opportunity to create a novel structural diversity of wall polysaccharides with different functional properties.

The overall aim of this project is hence the identification of genes relevant for cell wall biosynthesis using a functional genomic approach. The function of these genes is tested in vivo by studying the effects in mutants subjected to different biotic and abiotic stress responses. The project plan consists of three main subjects: 1) the development of a plant cell wall derived oligosaccharide profiling method 2) the use of this assay to screen for mutants with structurally altered cell walls 3) the investigation of the responses of the identified mutants to "Gauntlets", biotic and abiotic stresses.

Oligosaccharide profiling – a rapid and sensitive method to identify wall mutants

The identification of cell wall mutants is a daunting task, mainly due to the time-consuming and labor-intensive methods used to analyze wall polysaccharides in detail (Darvill *et al.* 1980). Therefore, an oligosaccharide profiling method was developed that allows the rapid identification of potential wall mutants (Fig 1). The method facilitates the use of specific glycosylhydrolases that solubilize a particular wall polysaccharide. The released oligosaccharides are then subjected to matrix assisted laser ionization time of flight (MALDI-TOF) mass spectrometry (MS). Spectra are obtained within 1-2 minutes, and give an account of the various oligosaccharide structures present in the tested wall material (Fig 1, example right corner). In addition, the relative abundance of the various oligosaccharides can be obtained by integration of the ion signal areas. Taken together a qualitative and quantitative oligosaccharide profile can be established. In Figure 1 an oligosaccharide profil of the major hemicellulose present in plants, xyloglucan, is presented. Comparisons of the profile obtained from wildtype leaf material with those of a variety of wall mutants clearly demonstrates the usefulness of the method as those mutants could have been clearly identified using the oligosaccharide profiling method (Lerouxel et al, 2002). However, a number of limitations to the method exist. Structural isomers cannot be distinguished as they have the same molecular mass. The oligosaccharide profile only reflects structures released by the enzyme. But enzymes are usually not able to solubilize a wall polysaccharide in its entirety. Therefore, once a plant mutant has been identified based on an altered oligosaccharide profile, a detailed structural analysis of the wall has to be performed to confirm the polysaccharide change, and to uncover potential additional wall alterations. Nevertheless, the profiling method developed here is very useful as it can be applied to a variety of wall polymers (xylans, pectins), to a variety of species (so far tobacco, potato, tomato, rice and barley were tested), and to a variety of tissues (leaf, stem, root, flower, seeds). Another practicable use of this method could be quality control of industrial wall polysaccharides and wall derived materials.

The method has been optimized in terms of speed and sensitivity to accommodate a medium to high throughput sample analysis (Fig 2). This improved protocol allows the analysis of plant material generated in Arabidopsis Verbund I "Genetic Diversity". For example a chemically mutagenized plant population is screened to identify novel wall mutants. Furthermore, insertional knockout lines of genes putatively involved in cell wall biosynthesis obtained from the GABI-KAT consortium can be tested, if they are indeed involved in wall build-up, and if yes, which polysaccharide structure is affected. Once wall mutants have been identified, the affected genes will be mapped and cloned. In addition, isolated mutants will be subjected to various biotic and abiotic stresses to assess the function of the gene. In conclusion, the oligosaccharide profiling method developed in the GABI program will give vital insights into polysaccharide biosynthesis and its function.

○ T. Choo *et al.* **Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting** Plant Physiology (2002) 130, 1754-1763. ○ A. Darvill *et al.* **The primary cell walls of flowering plants, in Tolbert, NE editor,** The Biochemistry of Plants (1980) Vol. 1 Academic Press, New York, 91-162. ○ N. Carpita and D. Gibeaut **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth,** Plant Journal (1993) 3, 1-30. ○ E. Zablackis *et al.* **Substitution of L-fucose by L-galactose in cell walls of mur1,** Science (1996) 272, 1808-1810. ○ Vanzin *et al.* **The mur2 mutant of *Arabidopsis thaliana* lacks fucosylated xyloglucan because of a lesion in fucosyltransferase AtFUT1,** Proc. Nat. Academy of USA (2002) 99 (5), 3340-3345.

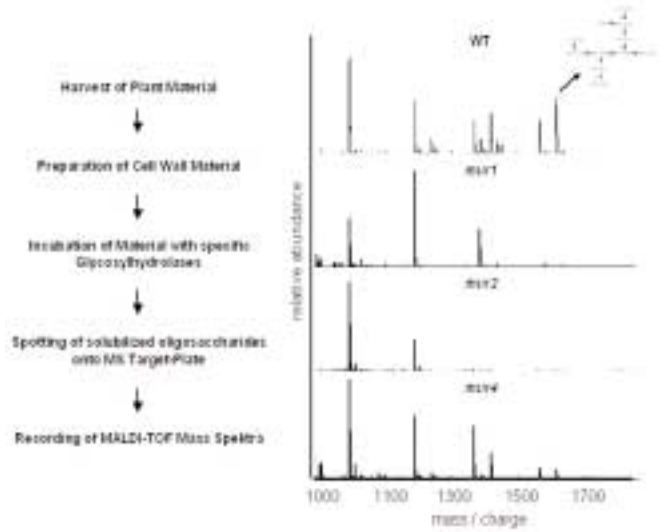


Fig 1: Left: *Establishment of an Oligosaccharide Profil using specific glycosyl-hydrolases and MALDI-TOF mass spectrometry (Choo et al. Plant Physiology (2002) 130, 1754-1763).*
Right: *MALDI-TOF spectra of xyloglucan oligosaccharides derived from leaf material of Arabidopsis thaliana wildtype (WT) and known xyloglucan mutants (mur1 - mur4) treated with a xyloglucanase, specifically solubilizing this hemicellulose. Ion signals represent oligosaccharides of known structure. An example is shown for m/z 1597 presented in the upper right corner.*
mur1- Inactivation of the de novo synthesis of GDP-fucose (Zablackis et al. 1996).
mur2 – Inactivation of a fucosyltransferase (Vanzin et al. 2002)
mur4 – Inactivation of an epimerase converting UDP-xylose to UDP-arabinose

Fig 2: *Optimisation of the oligosaccharid-profiling method for middle to high throughput analysis of wall materials. (A) The analysis can be performed on a single 4 day old etiolated hypocotyl. (B) Spotting of the sample onto a target plate is carried out by a pipeting robotor. (C) Automatic annotation and comparison of recorded spectra via the developed AMAS software [www.mpimp-golm.mpg.de/pauly "MS-based tools"). The program highlights MS-files in color for easy identification of significantly altered profiles.*

Spatio/temporal analysis of gene expression profiles within single cells and specific tissue types of developing Arabidopsis plants

Arabidopsis
Gauntlets

Elke Lieckfeldt†, Ulrike Simon Rosin, Daniela Zöller, Berit Ebert, Julia Kehr, Joachim Fisahn
Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam-Golm

Trichomes are hair like structures that develop on the surface of plant leaves. In *Arabidopsis thaliana* trichomes consist of single cells that are located in the outer epidermal cell layer. Due to their exposed location trichomes are model systems to study cell differentiation and cell cycle control. During leaf development trichomes differentiate from protodermal cells. The initial stages of trichome formation are associated with repeated endoreduplication. Several mutants in trichome formation provided information on genes involved in the control of trichome initiation and patterning. However, a more detailed understanding of the processes controlling trichome formation can be obtained by single cell sampling techniques (fig 1). This technique uses glass micro capillaries with an opening diameter of 1um that are inserted into single epidermis or trichome cells. Due to the turgor pressure the contents of the cell enters the micro capillary. Subsequently the capillaries are withdrawn from the cells and the single cell content is analysed by RT-PCR for gene expression profiling, GCMS, CE for metabolites and LCMS/MS, 2Dgel electrophoresis and SELDI for protein composition.

Routine collection of single cell contents has been successfully established in the first year of the project. In particular, single trichome, epidermis and basal cells of young and mature Arabidopsis leaves were harvested by glass micro capillaries. More recently, we developed a protocol for collection of cell sap from early stages of trichomes of developing primary leaves of 6 to 9 day old Arabidopsis seedlings. Up to 1000 single cells can be harvested per day.

Trichome basal cells putatively interact by cell-cell-communication in the epidermal layer during the initial phase of trichome formation. Therefore, we included samples of these basal cells in our gene expression studies. Collection of cell sap from basal cells is comparable to epidermal pavement cells, and thus straight forward.

RNA transcript analysis

To learn more about genes differentially expressed during primary leaf development in general and, moreover, to identify possible candidate genes which are involved in trichome formation, transcript analysis was also performed on primary leaf tissue of 6 to 9 day old Arabidopsis seedlings (fig 2). Total RNA was isolated, labelled and directly used for hybridisation of filter arrays and Affymetrix chips.

DNA glass arrays which were initially provided by GABI members are not available anymore. Due to this discontinuation of gene chip supply we had to switch to an alternative source of Arabidopsis chips. Therefore, we are currently using the Affymetrix system which is provided by the GABI resource center. Total RNA isolated from primary leaves of 6 to 9 day old seedlings has been used for hybridisation of 20 Affymetrix chips. Results of Affymetrix chip hybridisations are preliminary processed by the GABI resource center. Hybridisation data are then returned for further processing to the individual experimenter. Comparative analysis, in which expression data of two chips are compared, was performed by the software package Microarray SuiteTM(Affymetrix).

Simultaneously, we have continued filter array hybridisations on our in-house-made Arabidopsis filters. Overall 40 filters have been hybridised with 1-2µg of double stranded cDNA from pooled single cells (epidermal, trichomes at different developmental stages, basal), and with 10-15µg of total RNA from primary leaves of 6 to 9 day old Arabidopsis seedlings, respectively. Pairwise comparison based on a t-test was applied to our single cell expression data. Therefore, we used the in-house developed software program HARUSPEX (S. Kloska, bioinformatics group MPI-MP). Radioactive images of the filters (scanned with a phosphor imager) were analysed by the Array VisionTM software. Expression levels of transcripts derived from this raw data analysis were saved in data files which subsequently were transferred to the HARUSPEX program for normalisation and further analysis.

In order to analyse expression data coming from filter hybridisation, we developed a mathematical model. Based on the calculation using average and standard deviation of filter array hybridisation data, it was possible to determine the "real expression value" of genes in specific cell types. The calculation allowed statements about the expression profiles of specific genes in epidermal-, basal- and trichome cells. The prediction of expression data could be subdivided into 125 categories. These categories were separated in "allowed" and "non-allowed", or in other words, possible and impossible categories. 67 allowed categories could be defined, which include the information about gene expression in specific cell types. In near future the expression data from genes of interest in single cell types will be verified by real-time RT-PCR experiments.

In parallel, we are currently working to establish an additional model with data coming from filter array hybridisation and Affymetrix chip hybridisations with RNA derived from primary leaves. This model allowed the predication of specific gene expression in primary leaves at different developmental stages (6d-9d). Comparison and cluster analysis of primary leaf gene expression profiles will answer many questions related to primary leaf development of Arabidopsis.

Metabolite analysis on the single cell level

Capillary electrophoresis was used to analyse metabolites (sugars, organic acids, amino acids and ions) from single trichome and epidermal cells of mature leaves. Because of the very small sample volume of trichome cell sap, the loading procedure into the capillary was modified, and is done electro kinetically now. To produce larger signals (often values are close to the detection limit) and to improve the reproducibilitiy of the measurements, cell sap from 100s of trichomes was also used for CE. A number of experiments had to be performed concerning the internal standardisation of the measure-

ments. With the optimised CE protocol we have been able to detect several ions (K, Na, Ca), sucrose and glucose in single trichome and epidermal cells. To gain as much as possible information about cell specific metabolite profiles, we are using the pooled sap from 100-200 trichomes and epidermal cells for GCMS/LCMS analyses. This GCMS analysis is a joint effort with AG Fiehn (MPI Golm). Several hundred metabolites could be identified within single cell extracts and this analysis is still continued.

Protein analysis on the single cell level

For comprehensive protein profiling we follow three strategies: (1) the SELDI protein chip technique (Ciphergen) from pooled single cells (the same material as used for metabolite analysis), (2) 2D PAGE from primary leaves of 6 to 9 day old Arabidopsis seedlings, from several ecotypes differing in trichome phenotype (Col-0 – many trichomes, C24 – limited number of trichomes), and from mutants showing interesting trichome phenotypes (trichomes absent: gl1, ttg1; very hairy leaves: N1139, N318, N1093 from the NASC seed stock center). (3) LCMS/MS of single trichome and epidermal cells.

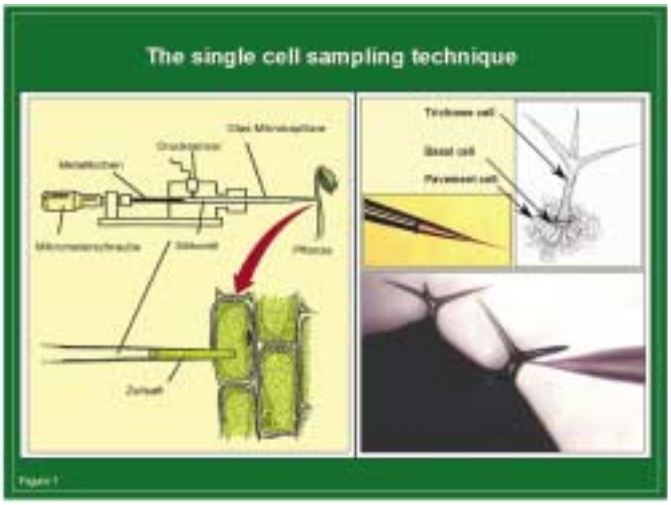
SELDI experiments have been performed in cooperation with Dr. v. Eggeling (University Jena) and started in January 2002. The cooperation with this group has turned out to be very fruitful and good results have been obtained. Mathematical analysis of SELDI expression profiles is in collaboration with the bioinformatics group of Joachim Selbig (MPI Golm) and significant differences have been detected between trichome and epidermis protein compositions.

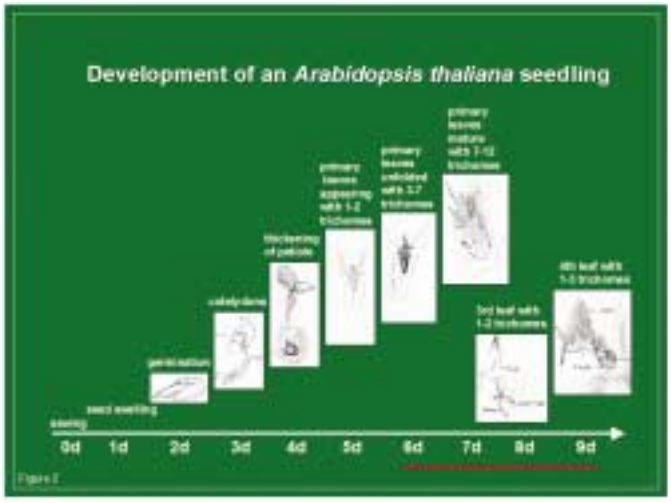
Proteins of 5 week old Arabidopsis leaves from various ecotypes (C24, Col0, LE, Enkheim Frankfurt) were extracted and separated by 2D gel electrophoresis. Furthermore, the cytosolic protein fractions of leaves from LE were compared to two trichome less mutants (N64, N8; obtained from ABRC, Columbia, Ohio, NASC, U.K.). Similarly, proteins derived from N1139 were compared to a Arabidopsis mutant with a highly increased number of trichomes. Larger gels were run by P. Giavalisco with material extracted from LE, N64 and N89. Comparison of gels was achieved by the software Melanie (genebio). From this mathematical analysis a number of spots were eliminated that will be further analysed and sequenced by QTOF techniques (J. Kehr).

LCMS/MS is presently the most promising method to identify proteins from single trichome cells of Arabidopsis. In collaboration with the group of W. Weckwerth (MPI Golm) 50 proteins extracted from single trichome cells have been identified. This list contains a number of known and unknown proteins. The localisation of these unknown proteins to trichome cells provides the first level of functional identification of these novel proteins. Some of the proteins identified by LCMS/MS have been associated previously with trichome specificity. Therefore, providing a proof of our present concept.

Present achievements in relation to the outlined project

Several novel candidate genes have been identified by chip and filter hybridisation on the single cell level that are controlling trichome development. The identified candidates are presently evaluated by real time PCR. Metabolite profiles of single trichome and epidermis cells have been successfully obtained. Protein composition of single trichomes is presently on the way and 50 proteins have been identified.





Searching for signals: stress-induced changes in Arabidopsis metabolite, peptide and protein patterns

Arabidopsis
Gauntlets

1 Stephan Clemens, 2 Jürgen Schmidt, 2 Ludger Wessjohann, 1 Dierk Scheel
Leibniz Institute for Plant Biochemistry, Halle/Saale
1 Department of Stress and Development Biology
2 Department of Bioorganic Chemistry

Aims and Objectives

Plants have to cope with a variety of constantly occurring and potentially stressful changes in their environment, such as pathogen attack, cold, shade, high light levels, ion imbalance, drought, exposure to toxic compounds or high salinity. Most, if not all of the plant responses to such changes involve the perception of signals and the synthesis of proteins, peptides or metabolites that counteract the respective stress. In general, the perception of stimuli by plant cells is poorly understood. Most of the signaling molecules and many of the physiological changes elicited by stress remain unknown. Towards the end of an unbiased and comprehensive biochemical phenotyping of plants we initiated a project aiming at the use of state-of-the-art analytical tools for the profiling of soluble proteins, peptides and (secondary) metabolites in *Arabidopsis thaliana* roots and leaves (fig 1). This extensive profiling is applied to the monitoring of stress-induced changes as well as to the identification of novel signaling molecules and of secondary metabolites, peptides and proteins directly constituting stress responses. Furthermore, this approach is being developed into a tool for the elucidation of various plant processes that is generally applicable to any plant species. In the following we report on the progress made with special emphasis on the metabolite aspect.

LC-MS-based metabolite profiling

A characteristic of plant life is the production of an enormous number of natural compounds (secondary metabolites). Secondary metabolites are known to have crucial roles in plant development as well as in the interaction of a plant with its biotic and abiotic environment. Ample evidence has been obtained in the past decades for a wide range of functions. Arrays of toxic compounds help defending against phytopathogenic bacteria and fungi or deterring herbivores. Secondary metabolites can function as signals both internally or in communication with symbionts such as rhizobacteria or mycorrhizal fungi. Also, they provide protection against a variety of abiotic stresses such as UV light, drought or high salt concentrations.

The majority of biosynthetic reactions in plant secondary metabolism is catalyzed by specific enzymes. Appropriately, completion of the *Arabidopsis thaliana* genome sequence revealed that a significant proportion of the ca. 25000 predicted genes encode proteins assumed to function in secondary metabolism (The Arabidopsis Genome Initiative 2000). This is illustrated by the fact that the *A. thaliana* genome contains, for instance, >250 cytochrome P450 genes, >100 acyl transferase genes and >300 glycosyl transferase genes. There is a huge discrepancy between the number of these genes and the number of known reactions catalyzed by these types of enzymes in Arabidopsis leading to the conclusion that a large number of metabolites have yet to be identified. Thus, understanding a significant part of Arabidopsis biology requires methods allowing both the sensitive detection and quantification as well as the identification of secondary metabolites. Applying such techniques to various genetic backgrounds and environmental as well as developmental conditions then would help elucidating the function of such compounds and of the genes involved in their biosynthesis. All of the above also applies to any other given plant species of scientific and/or economic importance.

Metabolomics has emerged as the third major path of functional genomics besides mRNA profiling (transcriptomics) and proteomics (Sumner *et al.* 2003). Metabolomic approaches seek to profile metabolites in a non-targeted way, i.e. to reliably separate and detect as many metabolites as possible in a single analysis. Profiling schemes for Arabidopsis and other plants have been developed in recent years (Roessner *et al.* 2000). Main focus of these mostly GC-MS-based approaches have been metabolites of the primary metabolism such as sugars, amino acids, organic acids or sugar alcohols. Several hundred compounds can be robustly and reliably detected. However, already these first pioneering reports emphasized the need for complementary LC-MS based approaches to allow a more comprehensive profiling (Roessner *et al.* 2000) as the metabolome is chemically highly complex and every analytical procedure is necessarily limited in scope.

We developed a profiling scheme using the robust and well-established separation of extracts by capillary LC on reversed phase material coupled to state-of-the-art mass spectrometry. The basic assumption was that such an approach would extend the range of metabolome analysis to those compound classes not amenable to GC analysis. These would include a significant fraction of the plant secondary metabolism. A few years ago quadrupole-time-of-flight mass spectrometers (Q-TOF-MS) were introduced. They combine time-of-flight mass analysis with electrospray ionization (ESI) resulting in high sensitivity, high mass resolution and high mass accuracy (Chernushevich *et al.* 2001) and therefore are principally well-suited for comprehensive metabolite profiling. To our knowledge, no reports of such an application of hybrid mass spectrometers have yet been published. It has been a major undertaking within this project to establish the use of such hybrid mass spectrometers for plant metabolome analysis (von Roepenack-Lahaye *et al.* submitted).

Following the setup of the plant growth system, the mass spectrometers and the analytical scheme, a thorough evaluation of variability was undertaken for the generation of plant material and the analytical procedures. We determined that the linear range of the MS analysis allowed to obtain quantitative data. The analytical and biological variance were found to be in a range accepted for other types of large-scale global analyses. Data analysis is a major bottleneck of metabolite profiling by Cap-LC-ESI-Qq-TOF-MS. Typically, the total ion chromatogram does not resolve distinct peaks (fig 2). Extensive data convolution is needed to extract mass spectra. We tested and optimized MetaboliteID, a metabolite processing software (Applied Biosystems). MetaboliteID allows to automatically extract mass spectra from the total ion chromatogram and to generate peak lists displaying retention time, mass and intensity of a peak. For the processing of the large data sets we developed macros and simple programs that allow, for instance, to generate master lists using retention time and mass as

○ The Arabidopsis Genome Initiative **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** Nature (2000) 408, 796-815. ○ I. V. Chernushevich, A. V. Loboda and B. A. Thomson **An introduction to quadrupole-time-of-flight mass spectrometry**. J. Mass Spectrom. 36 (2001), 849-865. ○ U. Roessner, C. Wagner, J.Kopka, R. N. Trethewey and L. Willmitzer **Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry.** Plant J. (2000) 23, 131-142. ○ L. W. Sumner, P. Mendes and R. A. Dixon **Plant metabolomics: large-scale phytochemistry in the functional genomics era.** Phytochem. (2003) 62, 817-836. ○ E. von Roepenack-Lahaye, T. Degenkolb, M. Zerjeski, M. Franz, U. Roth, L. Wessjohann, J. Schmidt, D. Scheel and S. Clemens **Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry.** Plant Physiology (2004) 134, 548-559.

identifiers, to adjust retention times and to average and compare experiments. State-of-the-art is now the robust detection of about 1400 distinct mass signals in methanolic leaf extracts of control plants and about 800 mass signals in respective root extracts. Subtle metabolic differences can be detected as demonstrated by the analysis of several Arabidopsis mutants.

A strength of Cap-LC-ESI-Q-qTOF-MS-based profiling, in addition to high sensitivity and resolution, is the possibility of obtaining structural information on metabolites of interest via tandem mass spectrometry. The majority of secondary metabolites in *Arabidopsis thaliana* – or any other plant - are as yet unknown. Consequently, few reference spectra or reference compounds are available. It is therefore very important to be able to gain structural information on metabolites showing interesting changes correlated with a certain condition or a certain genetic background. In many cases we are able to tentatively assign a structure to a given compound, demonstrating that the use of Cap-LC-ESI-Q-qTOF-MS exceeds the profiling and can be used for the identification of interesting metabolites in any kind of biological sample.

Peptides

Peptide signals have long been known in animals and bacteria. Hundreds of them have been found in non-plant eukaryotes and have been shown to play a pivotal role in cell-to-cell signaling. In plants, very few peptide signals have been identified to date. However, there is growing indirect evidence that peptides could be as important for signaling in plants as they are in animals. Peptides often are ligands of receptor-like kinases (RLKs) and hundreds of sequences encoding receptor-like kinases have been found in the Arabidopsis genome. Ligands have not been identified yet for most of these potential receptors. The extracellular domain of most RLKs contains leucine-rich repeat sequences which are implicated in protein-protein interactions. Thus, receptor-like kinases from plants may mostly recognize peptide ligands.

The intercellular washing fluid presumably contains many of the long-distance signaling molecules traveling between roots and shoots as well as many of the molecules involved in cell-to-cell signaling. It has been a major goal of this project to establish analytical procedures of maximum sensitivity in order to make the direct biochemical identification of novel signals feasible. We developed the nanospray-Qq-TOF-MS analysis of intercellular washing fluid. This fluid is harvested by infiltration and assays with marker enzymes show that there is only minimal contamination with cellular components. The sensitivity of the Q-STAR allows to directly visualize molecules in the metabolite and peptide mass range. Pilot experiments have also been carried out with an FT-ICR mass spectrometer.

Proteomics

The monitoring of changes in Arabidopsis protein patterns is focused on the separate analysis of soluble root, leaf and intercellular washing fluid proteins. Proteins are separated in large-format high-resolution two-dimensional gels. Samples are run in different pH gradients: 3-10, 4-7, and 5-6. Protein spots are cut out, trypsin-digested and analyzed by MALDI-TOF-MS. A large number of Arabidopsis proteins that change in abundance upon biotic or abiotic stress treatments have been identified in this way.

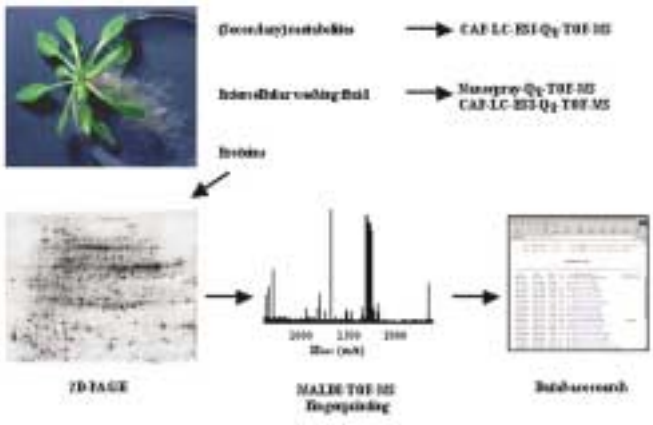


Fig 1: *Arabidopsis roots and leaves are analyzed separately. Soluble proteins are separated in large-format two-dimensional gels. Proteins of interest are cut out, trypsin-digested, analyzed by MALDI-TOF-MS and identified based on the peptide fingerprint. Methanolic extracts are analyzed by CapLC-ESI-Qq-TOF-MS. Intercellular washing fluid is analyzed by Nano-ESI-Qq-TOF-MS.*

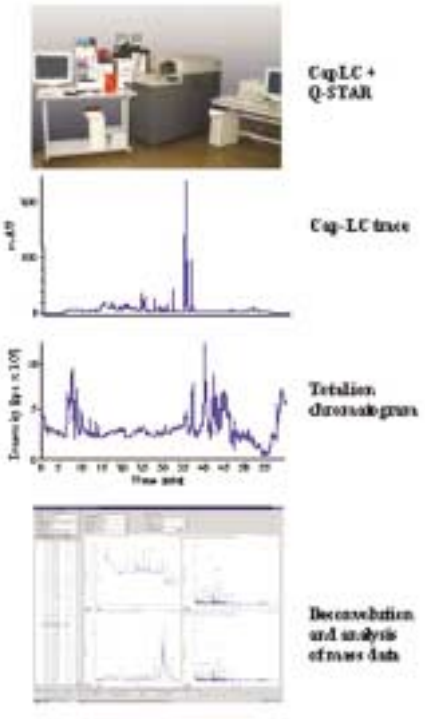


Fig 2: *Profiling of secondary metabolites using CapLC coupled to ESI-Qq-TOF-MS. The total ion chromatogram requires data deconvolution to extract mass spectra. Lists of mass signals providing information on mass, retention time and intensity are generated and processed.*

Genome Analysis in Rapeseed – Genetic, physical and functional characterization using Arabidopsis resources

GABI-GARS Overview

Oilseed rape (*Brassica napus*) is one of the leading oil crops world-wide. The cultivation of winter oilseed rape as the major European oil crop has increased tremendously during recent decades due to great progress in rapeseed breeding. Comparative crop improvement can probably not be achieved in future by conventional breeding methods in a short-term frame, particularly regarding quantitative characters such as grain and oil yield that are controlled by several or numerous genes scattered throughout the entire genome. Therefore, an improved understanding of the regulation and function of the responsible genes is beneficial in order to accelerate breeding success. Chromosome and gene duplications have played an especially important role in genome evolution in the genus Brassica. As the most important close relative of Arabidopsis, oilseed rape is likely to be one of the major crops to benefit most from the new information and technologies becoming available in the field of plant molecular genetics. Thus it can be expected that practical rapeseed breeding will in the next decade reap considerable benefits from molecular genetic research, technology and marker development.

The GABI-GARS research consortium is investigating key aspects of *B. napus* genetics and development in a co-operation involving many of the major scientific institutions and commercial breeders involved in oilseed rape breeding and research in Germany. The overall aim of the joint project is to exploit the possibilities offered by molecular biology – including genetic maps, physical maps and synteny studies using Arabidopsis genomic information – to create a basis for the efficient use of modern techniques to improve oilseed rape as a crop. The main focus is on identification of genes involved in the regulation of metabolic pathways influencing seed development. The generation of resources including mapping populations segregating for important traits, a series of intervarietal substitution lines, a physical (BAC) library and seed EST libraries, together with information on synteny between Arabidopsis and rapeseed, have laid a foundation for rapid identification, characterization and cloning of agronomically important genes. The direct involvement of commercial rapeseed breeding companies in the GABI-GARS consortium, in close cooperation with the participating scientific institutions, has created a clear pathway linking the basic scientific research on the *B. napus* genome, based on resources and knowledge from Arabidopsis, to the exploitation of the results in practical oilseed rape breeding.

Biochemical investigations of seed development in rapeseed for functional gene analysis *(Brassica napus)*

Rapeseed
GABI-GARS

Roland Baetzel, Wilfried Lühs, Wolfgang Friedt
Justus-Liebig-University Giessen, Department of Crop Science and Plant Breeding

During maturation of the rape seed, four phases have generally been recognised in terms of the accumulation of major storage products, which are used as biochemical marker points during rapeseed development: (i) a cell division phase, (ii) onset of rapid synthesis of storage oil, (iii) primarily storage protein synthesis, and (iv) primarily apo-lipoprotein (oleosin) synthesis during the dehydration phase. Using the available information which describes the kinetics of appearance of plant compounds during the development and ripening of oilseeds, a specific metabolic profiling of a winter rapeseed cultivar with high oil yield has been performed in this GABI-GARS subproject. Following the cultivation of a specific winter rapeseed genotype under standardised conditions in growth chambers this general temporal scheme was biochemically monitored with sample collections over the 3-month period of seed development in order to separate the fruit tissue into pod material and seeds with a liquid nitrogen-cooled apparatus. Morphological characteristics and detailed biochemical studies were performed in order to elucidate temporal and spatial aspects of the synthesis and metabolism during rapeseed embryogenesis leading to a stage-specific gross seed and silique composition including typical compounds such as triacylglycerols and fatty acids, protein species, and important secondary plant compounds.

The following primary aims of the project were achieved:

1. Development and collection of stage-specific plant material under highly standardised conditions
2. Recording of morphological characteristics during pod and seed development
3. Evaluation of biochemical and physiological parameters of clearly described stages of seed development in rapeseed with particular focus on lipids and proteins

Development of plant material

Based on analyses carried out on field material that showed the early developmental stages of most interest in terms of metabolic development, it was decided to sample more regularly in the early seed stages than originally planned. Harvests were thus carried out at 14, 17, 20, 23, 28, 35, 42, 49, 63 and 70 days after pollination (DAP) instead of weekly intervals.

Plants sown at 5 time intervals were transferred to growth chambers after vernalisation and grown under controlled conditions. Out of 225 available plants 150 were used for controlled pollinations. Male sterile plants ('Express-MSL') were hand-pollinated on 3 days each week with pollen from the fertile line 'Express 617'. Because very young material specific for the first embryonic cell division (DAP 7) is likely to be of great interest in terms of gene expression, appropriate material was harvested in addition to the above mentioned DAP stages. Beginning 7 days after pollination (DAP 7), immature pods were harvested three times weekly from August 2001 until the end of 2001. The harvested pods representing different developmental stages were stored at –80°C for molecular and biochemical analyses. The harvested fruit tissue was separated into pod material and seeds with a special liquid nitrogen-cooled apparatus. The whole material of each DAP stage was pooled and divided into four homogeneous sub pools. One part of the pooled material was transmitted to Partner B (SunGene) for further analyses.

Morphological characteristics

The different seed and embryo maturation stages were recorded and measured photographically. Developing seeds were visible in the silique within DAP 7, the embryos were microscopically small until DAP 14. Up to DAP 49 a gradual increase in seed size was observed, starting with 1.68mm length and 1.21mm width until DAP 49 (2.85mm length, 2.64mm width). The average pod length showed a rapid increase from DAP 7 until DAP 23 (23.7mm to 63.7mm, respectively), thereafter pod length remained nearly stable. Also the number of seeds per pod was determined, but there was no significant change from DAP 14 onwards, the average number of seeds per pod for all DAP was 24. To determine the temporal development of fresh weight, dry weight and water content during the harvesting of samples, corresponding measurements were taken 3 times from 10 pods for each maturation stage (fig 1). The water content of the seeds showed a maximum at DAP 28 and decreased sharply until maturity. Dry weight starts to increase from DAP 20 until maturity of seeds. The final average seed weight was 5.6mg at maturity.

Storage product formation

The biosynthesis of seed storage product occurred in three distinct phases as described earlier. For total lipids a stable content was recorded until DAP 20 with a gradual increase until DAP 28. The strongest increase of total lipids was observed between DAP 28 and DAP 70. Individual lipid classes were separated by solid phase extraction using a cyanopropyl-bonded silica phase. The changes in different lipid classes of developing rapeseed is shown in fig 2. For glycolipid content there was no significant change in the whole seed until maturity: Starting with 2.7% at DAP 14 there was a very slow increase from 3.7% (DAP 28) to 7.3% (DAP 49) with a rapid final decrease to 3.8% at DAP 70. For phospholipid content a gradual increase from DAP 14 (12.5%) to DAP 28 (17.4%) with finally a fast decrease till DAP 63 (5.0%) was observed. The neutral lipids as the main lipid class showed a slow accumulation during DAP 14-28 followed by a very rapid increase up to 52.2% at DAP 63. Afterwards, there was a gradual decrease until the seed attained maturity. Separation of lipid classes and detailled analysis of their fatty acid composition yielded the following results: At DAP

14 the neutral lipids showed a fatty acid composition with approximately equal amounts of oleate (18:1, 35.4%) and linoleate (18:2, 30.2%) plus minor amounts of linolenate (18:3, 9.6%), palmitate (16:0, 8.1%) and stearate (18:0, 5.8%). The fatty acid composition changed rapidly after the onset of storage lipid synthesis at DAP 23 until DAP 42. After this stage the fatty acid composition was similar to that of mature seeds (Fig 3). In contrast the fatty acid composition of polar lipids at DAP 14 was characterized by large amounts of 16:0 (15-20%), 18:2 (40-50%) and a lower level of 18:1 (3.0% phospholipids, 17.8% glycolipids). The fatty acid composition of the polar lipids mature seeds was approximately as follows: 16:0 (6%), 18:0 (1%), 18:1 (60%) and 18:2 (20%).

Outlook

The data of this project and the results obtained by Partner B (SunGene) will give a detailed overview of biochemical parameters during seed development of rapeseed and enable the possibility to identify key regulatory genes involved in biosynthetic pathways of metabolites using seed-specific microarrays.

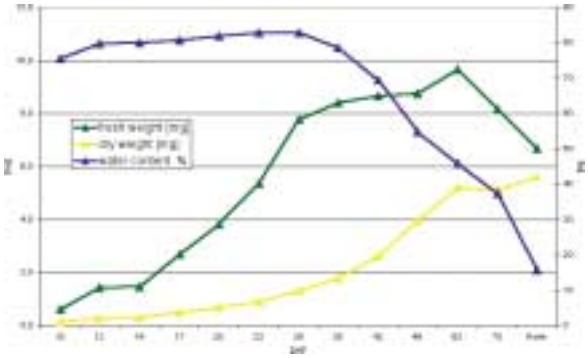


Fig 1: *Dynamics of fresh weight, dry weight and water content of developing seeds of B. napus cv. 'Express' (means of pooled seed material, 3 x 10 pods).*

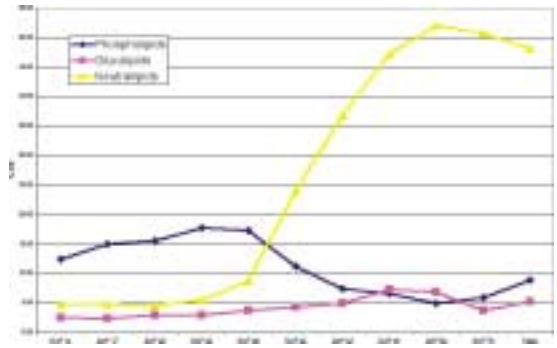


Fig 2: *Development of polar lipid and neutral lipid contents of developing seeds of B. napus cv. 'Express' (means of 8 replication per DAP).*

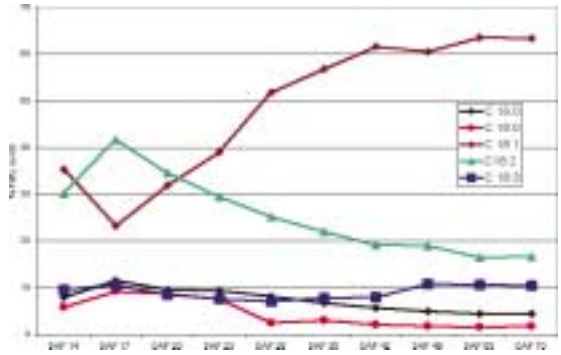


Fig 3: *Fatty acid composition (neutral lipids) in B. napus cv. 'Express' at different developmental stages (means of 8 replications per DAP).*

A strategy to identify regulatory genes important for the regulation of metabolic pathways in seeds during rapeseed development (Subproject B)

Rapeseed
GABI-GARS

Ute Heim, Michael Geiger, Karin Herbers
SunGene GmbH & Co KGaA, Gatersleben

Seed development of *Brassica napus* is associated with the import of N and C compounds thought to be largely produced in photosynthetically active leaves and to a lesser degree in the siliques during early phases of seed development. Reduced N and C are converted into former principal classes of storage compounds, i.e. starch, proteins and oil, they also serve as source for the synthesis of membrane lipids, non-storage proteins with enzymatic functions, different isoprenoid compounds, phenolic compounds and cell wall constituents. There is little knowledge about the factors governing the intricate network of seed metabolism. To identify influential regulatory genes for seed development and regulation it is indispensable to know the biochemical processes accompanying and/or determining seed development from immature to mature seed. This project focused on *Brassica napus* in order to facilitate a systematic biochemical and molecular biological analysis of development of an oil-rich seed. In order to analyse biochemical events and compare them to gene expression *Brassica napus* seeds cv. "Express" (winter rapeseed) was grown under highly standardised conditions in growth chambers of the university Gießen (Subproject A, Wolfgang Friedt).

Metabolic profiling of rapeseed siliques and seeds

To analyse the course of metabolic variation during development levels of different primary and secondary metabolites were determined in seeds and siliques. Carotenoids, chlorophylls, tocopherols, tocotrienols, starch, glucose, fructose, sucrose and amino acids were measured. Seeds and siliques were harvested 14, 17, 20, 23, 28, 35, 42, 49, 63 and 70 days after pollination (DAP).

In siliques the soluble sugars glucose and fructose showed a similar behaviour. At the end of drying stage (70 DAP, data not shown) a significant decrease were observed. Highest levels were found after 20 days of silique development. In seeds both soluble sugars were present in high amounts during the beginning of seed development followed by a 80% decrease measured at 28 DAP (data not shown).

The content of sucrose showed two peaks during silique development. The first was measured at 20 DAP where also the reducing sugars glucose and fructose were highest. The second peak was found at the end of silique development (63 DAP, data not shown). In seeds highest levels of sucrose were reached 40 DAP (data not shown). In contrast to the respective data on siliques the ratio between hexose and sucrose was very high during the cell division phase and dropped between 30 and 40 DAP when the storage phase is initiated (Fig1a,b).

In siliques the highest content of starch, shown as glucose equivalents/g fw, were found 35 DAP followed by a continuous decrease. 70 DAP starch was no longer detectable in the siliques. In seeds highest levels of starch were found at the beginning of seed development, followed by a strong decrease during later development. 49 DAP no starch could be detected (Fig1 b).

In siliques chlorophyll decreased during development to 90% (Fig1 c). In siliques higher values were present during the first stages of development, while in seeds highest amounts were found during middle stage (Fig 1d).

In siliques more than 90% of total tocopherol content was a-Tocopherol. g-Tocopherol was only present in very limited amounts (data not shown). An increase of total tocopherol was observed during overall silique development. In rapeseed seeds both tocopherol forms showed a significant increase during seed development (Fig 1f). d-Tocopherol was also detectable, but the percentage was below 1% of total tocopherol (data not shown). a- Tocopherol levels were highest during early development, whereas more g-Tocopherol accumulated during later stages of development (Fig 1f). Tocotrienols were not detectable in siliques and seeds (data not shown).

The data of biochemical profiling of seed development are expected to form a valid basis for activities aiming at the identification of relevant gene functions for the manipulation of seed development.

Expression profiling

In cooperation with the MPI for Molecular Plant Physiology (Thomas Altmann, Otto Toerjek) and RZPD (Johannes Maurer) an experiment was performed to prove the suitability of the "GeneChip Arabidopsis Genome Array" (Affymetrix) for transcriptional profiling of seed development of rapeseed. Two different genomic DNAs of rapeseed cultivar "Express" (BNI and BNII) and genomic DNAs of Arabidopsis cultivars "Columbia" (Col) and "C24" (C24) were isolated, labelled and hybridised with the Affymetrix chip. The results demonstrated that the hybridisation pattern between the homologues probes (BNI and BNII or Col and C24 respectively) were very similar. This fact and the low background have demonstrated the high quality of hybridisation experiments. Evaluation of this experiment has shown that from Arabidopsis 'Col' and 'C24' 205 and 291 genes, respectively, were not detected on the Affymetrix chip (8200 genes). On the other hand, from *Brassica napus* 7621 genes and 7772 genes, respectively, could not be detected. This means that approximately 95% of the oligonucleotides derived from Arabidopsis genes do not have sufficient homology to hybridise with the rapeseed probes under the conditions used. The high stringency of the evaluation software and/or lacking homologies between sequences of Arabidopsis oligos and rapeseed may explain these results. To solve the question how many oligonucleotides of the Affymetrix chip have sufficient homology to known Brassica sequences a bioinformatics approach was initiated and performed with the MIPS (Münchner Informa-

tionszentrum für Proteinsequenzen, Klaus FX Meyer, Heidrun Gundlach). 11 oligos from each 5000 genes of the Affymetrix chip were blasted against 300,000 sequences of a Brassica oleracea database. The data confirmed the first experiment in which only approximately 5% of the Arabidopsis genes could be allocated to rapeseed genes. These results demonstrated that the homologies between the oligonucleotides representing 5000 Arabidopsis genes were not sufficiently high to use the Affymetrix chip as a tool for transcriptional profiling for seed developing of rapeseed.

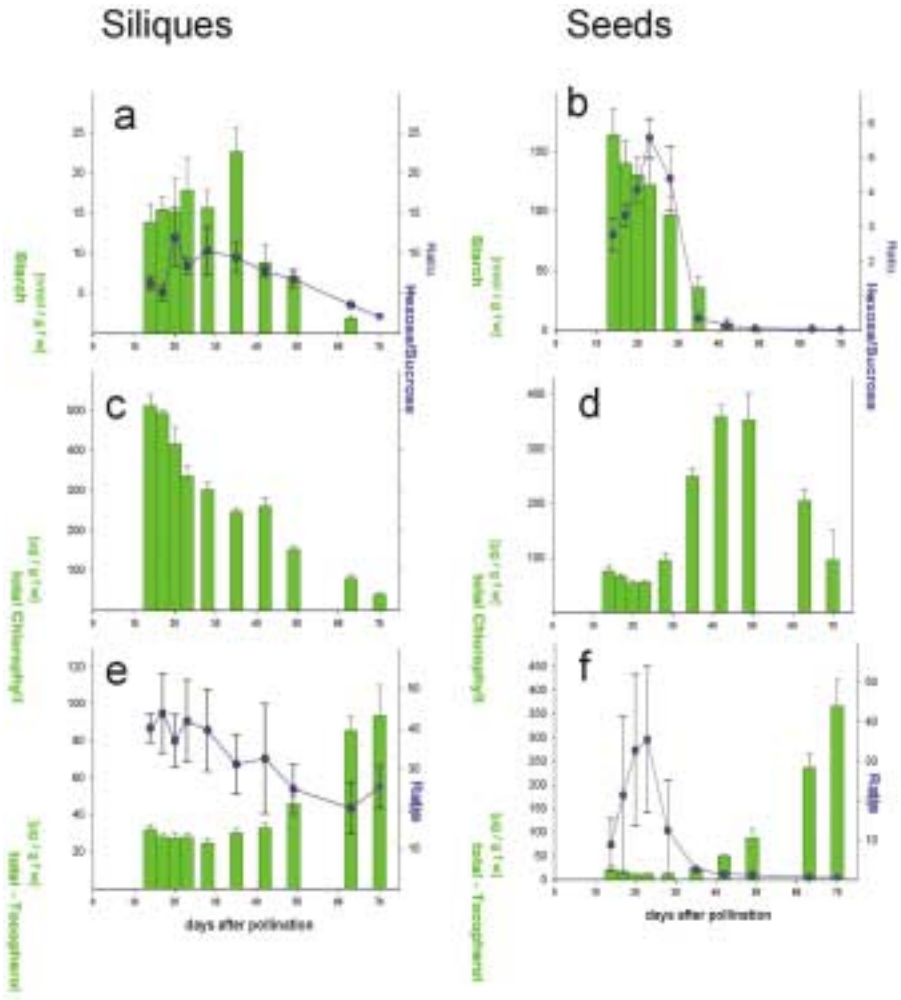


Fig 1: Results of metabolic profiling. *The figure shows the amounts of different metabolites (a, b: starch shown as glucose equivalents; c, d: chlorophyll; e, f: Tocopherol) and ratios between metabolites (a, b: hexose/sucrose; e, f: a-Tocopherol/g-Tocopherol) in siliques (a,c,e) and seeds (b,d,f) of Brassica napus, cv. "Express" at different stages of their development. Seed material was harvested 14, 17, 20, 23, 28, 35, 42, 49, 63, 70 days after pollination (DAP) from plants grown under highly standardised conditions in grow chambers. The values represent means ± SD of 5 separate seed pools.*

Development of intervarietal substitution lines, QTL mapping, and synteny analysis

Rapeseed
GABI-GARS

Wolfgang Ecke, Rubens Marschalek, Heiko Becker
Georg-August University Göttingen, Department of Crop Science and Plant Breeding

Introduction

Model organisms, like Arabidopsis or yeast, are very suitable to study the basic functions of genes. In crop plants, however, it will be more important to analyse and functionally characterize the variation between different alleles of the same gene since this is the basis for the optimisation of crop plants with respect to human needs. At the moment no comprehensive conclusions about the effects of allelic variation on the phenotype of a plant can be drawn from a purely molecular description. Therefore, an analysis of allelic variation should start on the phenotypic level.

The dissection of phenotypic variation and the assignment of phenotypic effects to allelic variation at individual genetic loci, e.g. quantitative trait loci (QTL) in the case of complex traits, can be achieved by genetic mapping. One approach to QTL mapping that has the added advantage of also allowing a fine mapping of QTL is the utilization of intervarietal substitution lines (Howell *et al.* 1996). Such lines are produced by recurrent marker assisted backcrossing (Fig 1) and contain only one donor segment in the genetic background of the recurrent parent. In a complete set the distinct donor segments of the different lines cover all of the donor genome. By comparing trait expressions across these lines QTL can be detected with a high power and can be assigned to small segments of the genome. When QTL have been fine mapped it becomes possible to utilize the information from model organisms that have been thoroughly characterized with respect to basic gene functions to identify candidate genes for the mapped QTL by analysing the synteny between the model organism and the crop plant.

Project aims

The objectives of the project can be summarized as follows:

1. The development of two series of intervarietal substitution lines for rapeseed. The substitution lines will be developed to provide a tool for (a) an efficient detection of QTL, (b) a precise phenotypic characterization of QTL effects, and (c) a fine mapping of selected QTL that have shown major effects on agronomically important traits with the objective to identify candidate genes for these QTL that can then be further characterized at the molecular level. The first series of intervarietal substitution lines will be developed from a cross between the two winter rapeseed varieties 'Mansholt´s Hamburger Raps' and 'Samourai' with the old Dutch landrace 'Mansholt' as donor and the current French variety 'Samourai' as recurrent parent. A segregating doubled haploid population of this cross has already been used to map a large number of QTL for agronomically important traits like oil and protein content, grain weight and harvest index by interval mapping. On the other hand, in a cross between two winter rapeseed varieties the allelic diversity of loci controlling basic properties of winter rapeseed as a crop plant and its specific ecological adaptations like biannual growing habit or winter hardiness is likely to be limited. Therefore, the second series of intervarietal substitution lines will be developed from a cross between 'RS239', a resynthesized rapeseed genotype with annual growing habit, as donor parent and the current high yielding winter rapeseed variety 'Express' as recurrent parent.
2. The mapping of QTL for agronomically important traits in rapeseed. The intervarietal substitution lines will be used to identify, map and characterize QTL for the major phenological and morphological traits as well as yield, yield related, and quality traits of rapeseed.
3. A global analysis of the synteny between Arabidopsis and rapeseed. Based on an RFLP map of the rapeseed genome the synteny between Arabidopsis and rapeseed will be analysed. This shall provide a means to utilize the information available from Arabidopsis genome research to identify candidate genes for QTL that have been genetically mapped in rapeseed.

Results

Development of intervarietal substitution lines from the cross 'Mansholt' x 'Samourai'

Utilizing AFLP markers from a genetic map developed in the segregating doubled haploid population of this cross for marker-assisted selection the development of the substitution lines has progressed into the BC3 generation. A total of 54 BC3 plants were selected and again backcrossed with 'Samourai' to produce the BC4 generation. The selected BC3 plants contain between 3 and 8 donor segments with an average of 5 segments, indicating that it will be possible to recover the first substitution lines with only one donor segment in the BC4 generation.

Development of intervarietal substitution lines from the cross 'Express' x 'RS239'

In a first step a genetic map was developed in a BC1 population of 90 plants. The map comprises 216 donor specific AFLP markers derived from 23 primer combinations that define 176 genetic loci distributed on 22 linkage groups with a total length of 1327 cM (Kosambi). All BC1 plants were again backcrossed to 'Express' and based on the marker analysis in the BC1 population 9 BC2 families were selected. These families are the progeny of BC1 plants that contain between five and ten large donor segments encompassing either complete linkage groups or at least half of a linkage group. Together these segments cover all of the mapped genome. Marker analyses are conducted on about 30 plants per BC2 family using the markers on the genetic map to identify a new set of genotypes with donor segments covering the mapped genome. In addition, all BC2 plants have been backcrossed with 'Express' to produce the BC3 generation.

○ P.M. Howell, D.F. Marshall, D.J. Lydiate **Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection.** Genome (1996) 39, 348-358. ○ M. Uzunova *et al.* **Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content**. Theor. Appl. Genet. (1995) 90, 194-204.

Synteny analysis

A synteny map of Arabidopsis and rapeseed was developed based on an RFLP map previously developed using the doubled haploid population from the cross 'Mansholt' x 'Samourai' (Uzunova *et al.* 1995). The map comprises 207 RFLP markers derived from 166 probes. Of these probes 155 could be successfully sequenced. Using DNA:DNA sequence comparisons with an Arabidopsis DNA sequence databank 162 loci in the Arabidopsis genome homologous to 139 of the 155 sequenced probes could be identified. The fact that some probes showed homology to more than one locus in Arabidopsis probably reflects the duplicated nature of the Arabidopsis genome. For the remaining 16 probes no homologous loci could be identified in Arabidopsis. Since a number of probes had shown more than one marker in rapeseed, the total number of RFLP markers on the genetic map for which homologous loci in the Arabidopsis genome could be identified was 175.

The size of regions showing collinearity between Arabidopsis and rapeseed varies greatly. On linkage group 8 for example no collinearity is apparent, adjacent markers show homology to different chromosomes of Arabidopsis, indicating a small-scale synteny (Fig 2). On the other hand, with one exception all markers in the lower part of linkage group 17 show homology to a continuous region of chromosome II of Arabidopsis, indicating a collinearity extending over more than 50 cM in rapeseed and nearly 16,000 kb in Arabidopsis (Fig 2). The results so far indicate that the synteny relationship of Arabidopsis and rapeseed is characterized by a high number of chromosomal rearrangements. Nevertheless, there are many segments where collinearity extends over several cM up to tens of cM.

Outlook

The development of intervarietal substitution lines will be continued for both crosses. After recovering lines with only one donor segment field trials will be conducted to map QTL for agronomically important traits. Based on the synteny study it will then be possible to identify candidate genes for QTL located in regions with extended collinearity between rapeseed and Arabidopsis.

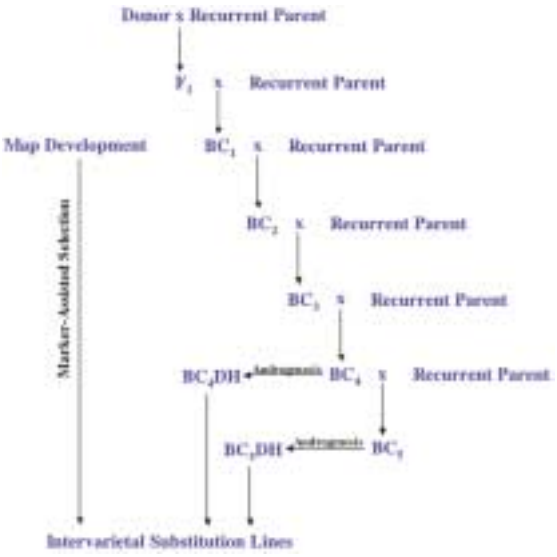


Fig 1: Development of intervarietal substitution lines by recurrent marker assisted backcrossing. *In each backcross generation a set of genotypes is selected which as small a number of donor segments as possible but with donor segments that together cover all of the donor genome.*

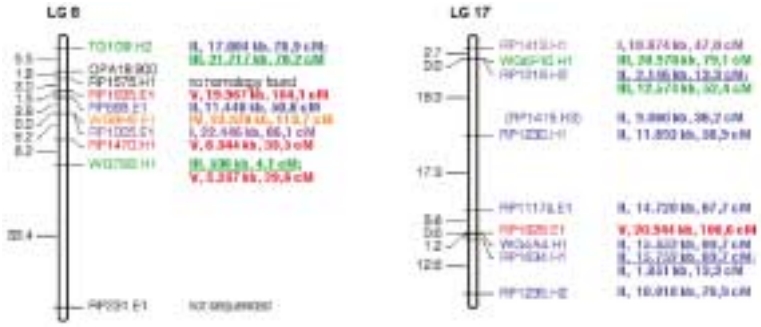


Fig 2: Synteny maps of linkage groups 8 and 17 of the rapeseed genetic map. *The roman numerals indicate the Arabidopsis chromosomes, the following figures the positions of the loci homologous to the corresponding RFLP markers on the sequence maps of the chromosomes.*

Seed colour in oilseed rape *(Brassica napus)*: Classical genetics, QTL analysis and candidate genes for map-based cloning and marker development

Rapeseed
GABI-GARS

1 Ana Gloria Badani, 1 Rod Snowdon, 1 Roland Baetzel, 1 Tina Lotz, 2 Renate Horn, 1 Wilfried Lühs, 1 Wolfgang Friedt
1 Institute for Crop Science and Plant Breeding I, Justus Liebig University, Giessen
2 Institute of Biological Sciences, Division Plant Genetics, University of Rostock

The yellow seed colour trait is of particular interest for oilseed rape breeding because the associated reduction in testa thickness (fig 1) results in reduced dietary fibre content after oil extraction, considerably improving the nutritional quality of rapeseed meal. Selection for yellow seed colour is difficult due to significant environmental effects, however, therefore a better understanding of the genetic control of the trait, and development of molecular markers for use in oilseed rape breeding, were two primary aims of the GABI-GARS project. For *Arabidopsis thaliana* a large number of mutants have been described which show an analogous transparent testa (tt) phenotype, and the corresponding genes have been identified and cloned in the model plant. In this project comparative genetic and physical mapping are being used to identify and characterize gene loci controlling seed colour in *Brassica napus*. We are particularly interested in the detection of associations between known Arabidopsis transparent testa genes and quantitative *B. napus* loci for yellow seed colour. Using a combination of classical genetics (segregation analyses in cross progenies), genetic mapping, QTL analysis and candidate gene studies our ultimate aim is the map-based cloning of important B. napus genes involved in the expression of seed colour and dietary fibre content.

In order to identify gene loci contributing to yellow seed colour in B. napus, segregating mapping populations (designated 'YE1' and 'YE2') were developed using lines derived from two distinct sources of true-breeding yellow-seeded oilseed rape crossed with different black-seeded genotypes. Genome maps were constructed using AFLP and SSR markers and aligned using reference markers that allowed identification of all linkage groups. Seed colour data was obtained for the segregating cross progeny using a digital imaging system, and this trait was found to be correlated with the content of anti-nutritive acid detergent fibre (ADF). Two corresponding quantitative trait loci (QTL) for seed colour were localised in the respective genetic maps, along with a third minor QTL in 'YE2'. One major locus explaining more than 60% of the phenotypic variance in seed colour was localised at the same position on linkage group 1 in the respective populations. Markers associated with yellow seed colour at this locus were verified in a third, independent population. A second QTL for seed colour co-localised on linkage group 19 in both crosses. A further minor QTL was found in 'YE2' but not in 'YE1'. Interestingly, the major locus for seed colour co-localised with a significant QTL for ADF (fig 2).

Combined QTL and segregation data for seed colour and content of ADF in doubled haploid and/or F2 offspring from the two crosses suggest that seed colour in B. napus is controlled by a dominant gene that contributes to a reduction in anti-nutritive fibre content, accompanied by one ('YE1') or two ('YE2') epistatic genes that may influence pigment biosynthesis. Genetic markers linked to the major gene have been identified that can now be used to assist in oilseed rape breeding for yellow seed colour and reduced crude fibre.

The most promising candidate genes to fit the genetic model suggested by the data we have obtained are key *A. thaliana* TT-genes. A number of relevant TT-genes code for specific enzymes involved in the flavonoid biosynthesis pathway, whereas the seed-specific gene TT1 encodes a WIP zinc finger-containing protein domain involved in development of the seed endothelium, the tissue layer in which tannin pigments accumulate. Because of its association with the endothelium TT1, which is implicated as a potential regulatory factor, represents a potential candidate for the putative major gene controlling ADF crude fibre accumulation in B. napus. Localisation of genomic sequence tags for TT1 and other relevant TT-gene candidates in the respective genome maps should ultimately assist in the confirmation of candidate gene loci associated with the QTL for seed colour. Together with the genetic markers linked to seed colour loci, these genes are now being used in an attempt to clone major oilseed rape genes contributing to yellow seed colour and antinutritive fibre content in the crop species. A large-insert B. napus BAC-library (8x genome coverage), also developed during the GABI-GARS project, will assist in this work. Ultimately we expect to develop new molecular tools for improved marker-assisted breeding of these important traits.

○ A.G Badani, R.J. Snowdon, R. Baetzel, F.D. Lipsa, B. Wittkop, R. Horn, A. De Haro, R. Font, W. Lühs and W. Friedt **A model for the inheritance of seed colour in oilseed rape (*Brassica napus* L.) based on analyses of segregation data, QTL and associated quality traits in two genetically distinct crosses.** Genetics (2005), submitted. ○ A.G. Badani, R. Snowdon, R. Baetzel, W. Lühs, R. Horn and W. Friedt **QTL mapping for yellow seed colour in oilseed rape (*Brassica napus*).** Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 85-87. ○ R. Baetzel **Stabilität und Genetik von wertbestimmenden Inhaltsstoffen gelbsamiger Rapslinien (*Brassica napus*) als Basis für die Züchtung von qualitativ hochwertigen Körneröl-raps-Sorten (Dissertation Theses).** Agrarwissenschaften, Giessen, 2004. ○ R. Baetzel, W. Lühs, A.G. Badani and W. Friedt **Development of segregating populations in the breeding of yellow-seeded winter rapeseed (*Brassica napus* L.).** Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 238-242. ○ R. Baetzel, W. Friedt and W. Lühs **Seed colour assessment in rapeseed (Brassica napus L.) by near-infrared reflectance spectroscopy.** Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 243-246. ○ T. Lotz, R. Snowdon, R. Horn, G. Dewal, B. Weisshaar, W. Friedt, H. Belcram, M. Caboche and B. Chalhoub **Molecular analysis of *Arabidopsis thaliana* transparent testa genes (tt) in *Brassica napus*.** Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 109-111. ○ R.J. Snowdon and W. Friedt **Molecular markers in Brassica oilseed breeding: current status and future possibilities.** Plant Breeding (2004) 123, 1-8.

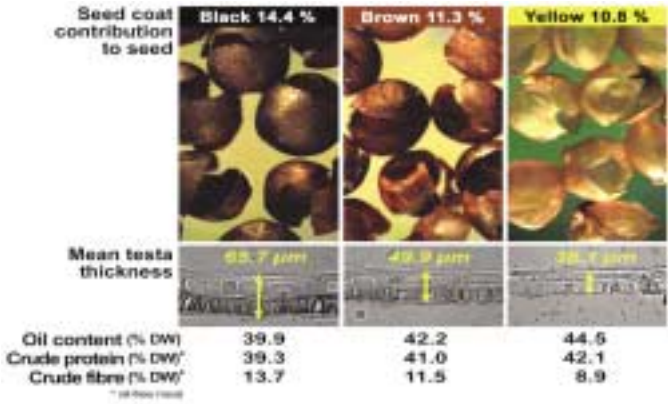


Fig 1: *The yellow seed colour trait is of particular interest for oilseed rape breeding due to the significant correlation between seed colour and anti-nutritive dietary fibre. Yellow seediness is caused by a reduction in thickness of the seed coat, which is associated with a significant decrease in fibre content in the seed meal after oil extraction. The proportion of oil and protein per seed dry weight is often also higher in yellow seeds than in black seeds, due to the lower overall contribution of the testa to the seed.*

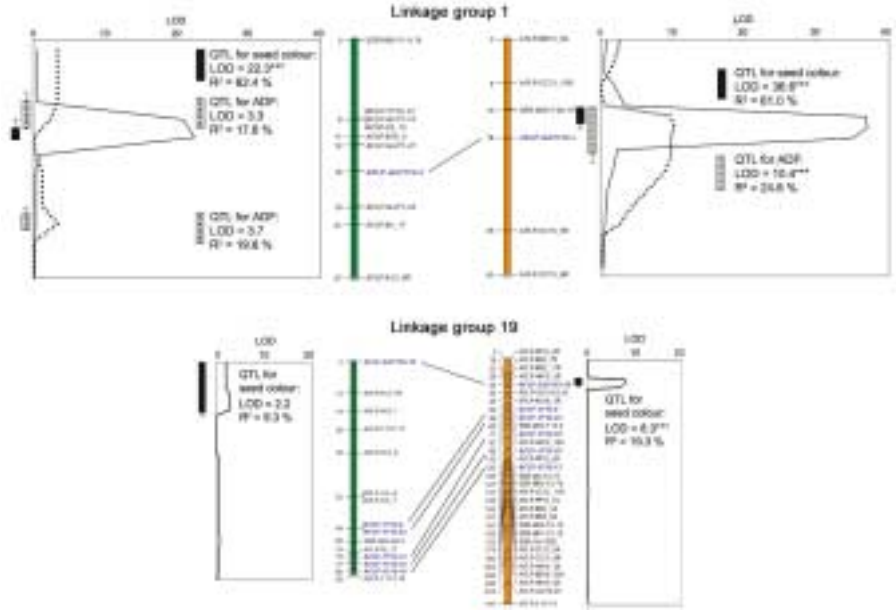


Fig 2: *Seed colour QTL (solid lines) were found to co-localise on two aligned linkage groups in two distinct segregating crosses 'YE1' (green linkage groups) and 'YE2' (orange linkage groups) between different yellow- and black-seeded oilseed rape genotypes. Furthermore, the major locus (contributing more than 60% of the genetic variation for seed colour) was found to co-localise with a QTL for acid detergent fibre (ADF; dashed lines). Alignment of the linkage groups was possible due to the presence of common reference markers in the two maps (joined by lines). A third QTL was found in 'YE2' but not 'YE1': This is in agreement with segregation data that suggest a digenic segregation between black- and non-black seeds in 'YE1' and a trigenic segregation in 'YE2'. The same dominant major gene appears to be involved in expression of seed colour in both crosses, with one ('YE1') or two ('YE2') further genes acting epistatically.*

Oilseed rape *(Brassica napus* L.*)* – Establishment of Segregating Progenies and Field Evaluation

Rapeseed
GABI-GARS

Christoph Lüdecke, Heinrich Busch, Dieter Stelling, Konstanze Richter
Deutsche Saatveredelung Lippstadt-Bremen GmbH (DSV), Lippstadt

Introduction

The overall topic of the GABI-GARS project was (1) to analyse the genome of oilseed rape (*Brassica napus*) with respect to genes for seed colour and (2) to create the basis for the identification and functional analyses of regulatory genes involved in the development of its seeds. The role of Deutsche Saatveredelung Lippstadt-Bremen GmbH (DSV) and the other breeding companies involved has been to support the project with suitable rapeseed materials, breeding knowledge and propagation capacities to develop the required plant materials for the project.

Primarily this involved extensive multi-year and multi-location field testing and the evaluation of diverse agronomical and oil quality characters of oilseed rape breeding populations that were developed for subsequent genetic analysis by the scientific partners.

Project aims & results

The task of DSV was to develop and to propagate rapeseed materials and to characterize these in the field. The plant material was then taken over by the scientific project partners (Universities of Giessen and of Göttingen) for further analyses.

The development and propagation of rapeseed materials was done for two sub-projects within GABI-GARS, namely the subprojects "seed colour" and "substitution lines". For the first subproject, the aim of DSV was to develop segregating progenies and to deliver DNA, leaf and seed samples along with seed quality data to the scientific partners. For the subproject "substitution lines", the task of DSV was to multiply seed of the substitution lines and to test them in field trials.

Subproject "seed colour"

For the subproject dealing with the colour of B. napus seeds, DSV received a segregating F2 population from project partner University of Giessen. In autumn 1999, the F2 offspring was sown in field trials and evaluated with respect to their agronomic characteristics. A total of approximately 1000 F2 individuals were evaluated in the field. In spring 2000, around 800 of the plants that survived the winter period were selfed by bagging to produce F3 seeds. From each of the appr. 800 F2 plants, two leaf samples were taken and either frozen or used for DNA extractions. Seeds from appr. 750 selfed F2 plants as well as from some open-pollinated plants could be harvested. Both types of F3 seeds derived from selfing and open pollination were analysed with near-infrared spectroscopy (NIRS) for their seed quality, i.e., content of oil, glucosinolate and protein, oleic and linolenic acid. All the seed samples together with the quality data and spectra were given to the project partner University of Giessen who screened the material for the seed colour trait.

Based on the seed colour analyses, selected F3 seeds derived from 361 F2 plants were sown in autumn 2000. Again, the plants were evaluated with respect to their agronomic characteristics. About 1600 plants were selfed in spring 2001 to produce F4(3) seeds. Leaf samples were taken from the F3 plants and deep-frozen. F4 seeds derived from selfing as well as from open pollination were analysed with NIRS for their seed quality. All the F4(3) seeds together with the quality data and spectra were delivered to project partner University of Giessen who also screened this material for the seed colour trait.

Subproject "substitution lines"

For the subproject "substitution lines", DSV has received preliminary substitution lines, derived from the cross 'Mansholts' x 'Samourai', from the project partner University of Göttingen. These lines will be selfed for plant propagation. Harvested seeds will be used for field trials and scoring of the respective lines for different agronomic and quality parameters. Development of substitution lines in a winter-annual crop species like oilseed rape is a considerably time-consuming process, therefore their completion was planned from the outset of the project over a time-frame of six years. Under favourable conditions the development could have already been finished in 2002, but this was not achievedby the University of Göttingen due to some technical limitations and the presence of still numerous donor segments in the substitution lines in generations BC2 and BC3. Once completed, however, these lines will represent an extremely powerful tool for the genetic analysis of important traits in oilseed rape. Because of the potential benefit of such materials, the breeding companies involved in the GABI-GARS project decided to take over the costs for a prolongation of this subproject for one further year, in order to support the further development of the substitution lines and to achieve the most important aims and milestones of this sub-project.

Outlook

The already and probably in the near future identified and characterized markers of the two sub-projects "seed color" and "substitution lines" can and will be used directly in the oilseed rape breeding programs of the breeding companies involved in the GABI-GARS project.

Especially for traits of oligogenic inheritance (like seed colour or some other quality parameters) which, in addition, are also influenced by some environmental factors, the selection should become much easier with the help of these molecular markers. Therefore, the costs of breeding and seed production with respect to these traits can be expected to be reduced considerably.

Substitution lines represent an extremely valuable tool for a targeted genetic analysis of traits. In particular for quantitative traits, it is possible using substitution lines to gain informations regarding the specific contribution(s) of different genome segments to the phenotypic expression of each single trait and/or its components. Prerequisites are (I) a full set of oilseed rape substitution lines which is not available yet and (II) an extensive evaluation of the substitution lines in the field, greenhouse and lab which the involved breeding companies intend to do in following projects. By the utilisation and optimisation of known statistical and quantitative-genetical methods it then should be possible to detect a high number of OTLs and associated markers which, for example, are suited to increase the selection efficiency of the rapeseed breeder in several cases significantly.

The entire GABI-GARS project has suffered from significant cuts in the original project application that, for example, have led to the unavailability of an oilseed rape BAC library and a cDNA library for developing *Brassica napus* seeds. Without these genomics tools the development of molecular markers that can be directly and usefully applied in breeding programmes has been less feasible. The outlook for successful economic utilisation of the results of the GABI-GARS project, therefore, is smaller than in the original forecast. Parts of the initial project's goals, however, may be achieved in following projects.

Development and phenotypic description of basic material for genome analysis of oilseed rape

Rapeseed
GABI-GARS

Lutz Brahm, Jutta Böhm
Saatzucht Hadmersleben GmbH, Hadmersleben

Introduction

Winter oilseed rape (*Brassica napus*) plays a major role within the domestic crop plants through its acreage of more than 1 million ha. The oil of 00-oilseed rape being present in content of about 45% in the seed is of excellent quality for nutritional purposes and useful for industrial applications as well. After crushing the resulting meal contains about 40% protein with a well-balanced amino acid composition. Nevertheless the high content of crude fibre (ca. 12%) of rapeseed meal induces a worse evaluation than soybean meal (ca. 7%) from an energetic point of view.

Therefore, further improvement of the seed quality with regard to a reduced crude fibre content is of major interest in practical plant breeding. The aim is to increase the oil and protein content as well as digestibility and energy concentration of the meal.

Arabidopsis thaliana has been widely used to discover the basic functions of genes. The knowledge that has been gained in that model crop should be useful also in investigating traits of agronomic importance in other crop species, especially in a related species like oilseed rape with both crops belonging to the same family of Brassicaceae.

Project aims

The aim of the project is to provide and characterize basic plant material for synteny studies using Arabidopsis resources.

In that respect a large DH population segregating for seed colour and crude fibre content should be developed in collaboration with Norddeutsche Pflanzenzucht Hans-Georg-Lembke KG (GABI-GARS: breeding and biotechnology development of genetic Rapeseed material, 0312286 H). This population is used to study the synteny of Arabidopsis and Rapeseed at respective loci inheriting seed colour in both species, and to investigate crude fibre content and oil content at the University of Giessen (GABI-GARS: exploiting microsynteny and the Arabidopsis genome resources for marker development and novel strategies of gene cloning in oilseed rape, 0312286 A).

In addition, a set of intervarietal substitution lines should be provided by the University of Göttingen (GABI-GARS: development of intervarietal substitution lines, QTL mapping, and synteny analysis, 0312286 C) for field evaluation of important agronomic traits (yield, yield related, and quality characters). These lines are produced by recurrent marker assisted backcrossing, and contain only one donor segment in the genetic background of the recurrent parent. The distinct donor segments of the substitution lines should cover the whole donor genome. These lines provide a tool for the efficient detection of QTL for important agronomic traits and the precise phenotypic characterization of QTL effects. Subsequently, QTL that have shown major effects on important traits can be fine mapped with respect to the identification of candidate genes by exploiting synteny between Rapeseed and Arabidopsis.

Results

DH population development

Starting from 10 clones of a F1 plant of the cross 'Express 617 (00)' x 'Yellow II (00)' provided by Norddeutsche Pflanzenzucht Hans-Georg-Lembke KG, finally, 330 DH lines were developed.

'Express 617' is a inbred line derived from the high yielding, black seeded winter rapeseed variety 'Express', with a high oil, low erucic acid and low glucosinolate content. 'Yellow II (00)' is a yellow seeded inbred with low erucic acid and low glucosinolate content.

Following microspore isolation 3680 embryos were transferred to solid MS-media. After development of shoots and roots, 532 plantlets were planted in soil. 502 DH plants were analysed for their ploidy levels using a flow cytometer. 299 showed spontaneous chromosome doubling, 158 plants were haploid and treated with colchicines. 21 plants were triploid and 24 tetraploid. Finally, 457 were adapted to the greenhouse. 127 of those showed no pot set. At the end, seed of 330 DH plants was harvested.

A second generation was cultivated in the field and greenhouse for seed multiplication. 276 DH lines were planted to the field and 54 were grown in the greenhouse. 254 DH lines derived from selfings in the field were transferred to the University of Giessen for further experiments with respect to seed colour. The DH lines have been analysed for quality traits using Near Infra Red Spectroscopy and seed colour.

Field evaluation of substitution lines

Two crosses are used to develop intervarietal substitution lines at the University of Göttingen

a) 'Mansholt´s Hamburger Raps' x 'Samourai'
   In the first cross the old Dutch landrace 'Mansholts' is used as the donor and the current French variety 'Samourai' is used as the recurrent parent. Both are winter types. In a marker assisted selection procedure BC3 plants were selected for further backcrossing with an average of five donor segments. First substitution lines with only one donor segment are expected to be discovered in BC4.
b) 'Express' x 'RS239'
   'RS239' is a resynthesised rapeseed with annual growing habit, and serves as the donor parent in the second cross. The current high yielding variety 'Express' is used as the recurrent parent. First a genetic map was developed followed by marker assisted selection in BC1 and BC2. BC2 plants have been backcrossed to 'Express' to produce the BC3 generation.

Up to now no substitution lines containing only one donor segment could be discovered in either of the crosses. Hence, no field evaluation could be conducted to characterize important agronomic traits in the substitution lines. The development of intervarietal substitution lines will be continued in both crosses. Following the discovery of lines with one donor segment, the field evaluation those lines will enable mapping of QTL for agronomically important characters. Finally, candidate genes can be identified based on synteny studies exploiting Arabidopsis resources.

Genome Analysis of Sugar Beet

Sugar Beet Overview

Sugar beet is the only sucrose storing crop of moderate climates. It has the highest product profitability of all agricultural crops in central Europe. Sugar beet is a biennial species forming a leaf rosette and a storage root in the first year. It is a diploid species with 18 chromosomes and a fairly small haploid genome size of 758 Mb. There are only three beet breeders of worldwide importance, two of them are located in Germany. Both are partners of the different GABI sugar beet projects together with one sugar producing company, one start up company and six academic partners.

There are three GABI projects dealing with sugar beet:

- GABI-Beet will provide new technologies for molecular breeding of sugar beet. During this project a large collection of ESTs will be sequenced and fully annotated, high density marker maps largely based on SNPs will be established together with a representative BAC library of beet and the genome organization of cultivated and related wild beet species will be investigated.
- The aim of GABI-BOLT is to clone the gene for early bolting which is important for breeding and cultivation of beet.
- Sweet-GABI has the aim to combine structural and functional genomics to analyze sugar metabolism and root development in sugar beet.

It is expected, that the technology to be developed will also have major impact on breeding of other crop species. GABI-Beet aims to concentrate on regions of the genome where genes characteriztic for a root crop such as root diseases, sucrose storing and early bolting are located. Sugar beet will thus be a model species for root crops as well as for crop species with small genomes.

The molecular marker map to be established during GABI-Beet will be enriched with EST derived SNP markers. The ESTs will be selected due to the predicted function of the derived proteins. Since the economic value of sugar beet is essentially depending on its quality, genes involved in carbohydrate and nitrogen metabolism, disease and stress resistance will be mainly selected. Thus, a functional map will be accomplished with marker genes which can serve as candidate genes for later cloning of genes which contribute to the genetic variation of traits listed above. Therefore, the two projects GABI-Bolt and Sweet-GABI use resources established within the project GABI-Beet like large insert libraries and EST collections. The results will help us to understand developmental processes like storage organ formation and shoot induction and responses to biotic and abiotic stresses. Molecular markers will be valuable tools for selecting elite breeding material and for map based cloning of genes. Finally, the results of these projects will help to select adapted breeding material with superior quality characters.

GABI-Beet Overview

GABI-Beet will provide new technologies for the molecular breeding of sugar beet. Marker technology (RFLPs, RAPDs, SSRs, AFLPs) has already been employed successfully in practical beet breeding. Here, we aim to introduce SNPs based on expressed genes and anonymous markers in combination with a high throughput system for marker analysis as a new tool for molecular breeding suitable for research area 2. GABI-Beet is devided into four major projects: project A (sequencing of ESTs and bioinformatics) which is the requirement for project B (high density mapping and allelic variation detected by SNP technologies) with the integration of conventional and functional markers (functional map). This approach is complemented by the comparative mapping and genome analysis by in situ hybridization (project C) and the construction of representative large insert libraries (project D). The GABI-Beet project is focussing the following milestones:

A. Supply of extensive sequence information of transcribed regions of the sugar beet genome,
B. Establishment of an integrated data base for molecular and genetic information and the development of new marker systems based on single nucleotide polymorphism (SNP),
C. Comparison of the genome structure of wild and cultivated beet with respect to repetitive and mapped low-copy sequences,
D. Construction of a BAC library from sugar beet genome

Milestones and results

A representative BAC library of beet has been constructed. All clones have been spotted on nylon membranes in a high density format. DNA pool are presently been established for screening the library by PCR. 12,639 cDNA have been sequenced from a librray normalized by oligofingerprinting. 9,730 different ESTs have been identified and 4,670 sequences have been annotated and assigned to functional categories. A core collection of previously mapped RFLP markers has been sequenced. A database (Beetbase) has been established containing sequence information of all clones mapped so far. SNPs can be easily identified by comparison of sequences from a panel of 16 different beet genotypes. As a result, a high frequency of SNPs in sugar beet became obvious. The present status with respect to EST mapping over all groups of GABI-Beet is shown in the following table (tab 1). The PCR products have been obtained using EST derived primers.

For mapping of ESTs four different beet populations have been selected. Taken together, DNA has been isolated from more than 380 plants. A sugar beet Ty1-copia element has been sequenced and its distribution around the beet genome has been determined by means of fluorescence in situ hybridization. For comparative mapping of cultivated and wild beet genomes segregating pouplations have been established by crossing different wild beet species. Production of wild beet-sugar beet addition lines failed so far due to non-viability of hybrid seeds. A Rhizoctonia resistance tests has been worked out and a sugar beet population has been phenotypicall characterized in terms of resistance to this economically important pathogen. As a next step, genes involved in Rhizoctonia resistance will be mapped preferentially using ESTs as markers. More detailed information is available from the individual short reports listed below.

Partners of GABI-Beet

GABI-Beet is a joint project between 11 partners from academic research institutions and industry coordinated by partner Kiel.

- Institute of Crop Science and Plant Breeding, Kiel
- Federal Center of Breeding Research of Cultivated Plants, Braunschweig
- Institute for Sugar Beet Research, Göttingen
- GSF, Neuherberg
- MPI for Breeding Research, Plant Breeding and Yield Physiology Department, Cologne
- ADIS, Cologne
- TraitGenetics, Gatersleben
- Institute of Plant Breeding, Halle
- Kleinwanzlebener Saatzucht AG, Einbeck
- A. Dieckmann-Heimburg, Nienstädt
- Nordzucker AG, Braunschweig

	No. of amplicons generated	No. of amplicons analyzed in panel	No. of amplicons mapped in one (two) population(s)
Total (done)	981	718	601 (295)
Total (planned)	>1000*	>1000*	1000 (100)

**: depending on efficiency of amplification in different Beta genotypes*

Tab 1: *Number of amplicons analyzed for SNPs and mapped loci.*

GABI Bolt Overview

Sugar beet is a biannual species. In the first year it forms a storage root and a leaf rosette. Shoot elongation (bolting) and flowering takes place in the second year or after a cold treatment of >10 weeks. For beet production bolting is an unfavourable character because it results in severe sucrose loss and low quality. For breeding, however, early flowering is preferred because the breeding process can be shortened. In sugar beet, the locus B was found to be responsible for early bolting with B being completely dominant over b. Beets harbouring the B allele make shoot elongation without vernalization. This gene can be used in beet breeding to shortcut the breeding process for example for rapid selection of maintainer plants (O-types). For seed production, however, the B allele has to be replaced by the recessive allele to insure bolting resistance in the field. Cloning of the B gene offers the possibility to manipulate the annual behaviour by transgenic technology. This has benefits for both breeders and growers. First, the introduction of the early flowering trait into breeding material could save time during the breeding process. Also, unwanted crosspollination with annual wild beets which can take place during seed multiplication can be easily monitored by marker technology using the B sequence as a probe. Second, full control of bolting character is a prerequisite for breeding of cold resistant beets which can be sown before winter. The yielding potential of winter beets is much higher compared to spring beets. However, sowing before winter requires cold tolerance and extreme bolting resistance. Introducing the bolting gene into non-bolting genotypes is a prerequisite for breeding because otherwise these plants would never produce flowers. Inactivation by antisense technology is a means to restore the non-bolting phenotype in the F1 hybrid. It can be expected that cloning of the B gene will have a major impact on beet breeding and breeding of other root crops with biannual behaviour.

The aim of the project is to clone the bolting gene B. The sequence will be analyzed and the gene will be transferred to non-bolting beets. Antisense transformation of bolting beets carrying the B allele will be made to restore the biannual character. An easy to perform marker test relying on the B sequence will be designed for routinous screening of seed lots.

Sweet GABI Overview

Obviously, sugar content is one of the most important traits in sugar beet. Sugar content is also a highly heritable trait and is therefore well amenable to classical breeding procedures. However, a negative correlation exists with yield, making the simultaneous improvement of both traits more difficult.

The aim of the Sweet GABI project is to identify genes, that can be considered candidate genes for sugar content, and can be used in marker assisted sugar beet breeding. In addition, we hope to get a better understanding of the reasons for the negative correlation of sugar content and yield.

QTL analysis for sugar content has been performed and revealed several regions of the genome which explain the genetic variation of this trait. To widen our understanding of the inheritance of sugar content, a wide cross between sugar beet and red table beet was made and the progeny will be tested in field trials in the frame of the Sweet GABI project. Besides sugar content, quality, and morphological data will be evaluated.

At the same time, developing tap roots are analyzed during the growing period, samples are taken for microscopic analysis of the expansion process and the development of cambium rings.

Expression analyses is applied to monitor gene expression during the sugar accumulating period and to compare gene expression between outer and inner cambium rings, the latter having a higher sugar content.

From these experiments, candidate genes will be selected and put on the genetic map. In case, they map into QTL regions already identified, they enter a next step of verification by being analyzed in association studies in a wide panel of breeding lines.

By this approach we combine genetics with structural and functional genomics and we integrate all the information, sugar beet breeders have accumulated for their breeding material.

Genome analysis of sugar beet, a model species for root crops
(Subproject A)

Sugar beet
GABI-Beet

1 Christian Jung, 1 Uwe Hohmann, 1 Maike Niemann, 1 Thomas Schmidt, 1 Tim Thurau,
2 Hans-Werner Mewes, 2 Stephen Rudd, 3 Georg Büttner, 3 Bernward Märländer, 4 Lothar Frese
1 Christian-Albrecht University of Kiel, Department of Crop Science and Plant Breeding, Kiel
2 National Research Center for Environment and Health (GSF), Institute for Bioinformatics, Neuherberg
3 Institute for Sugar Beet Research (IFZ), Göttingen
4 Federal Center for Breeding Research on Cultivated Plants (BAZ), Braunschweig

Objectives of GABI-Beet

GABI-Beet will provide new technologies for molecular breeding of sugar beet. Marker technology (RFLPs, RAPDs, SSRs, AFLPs) has already been employed successfully in practical beet breeding. Introducing SNPs based on expressed genes and anonymous markers create high density functional maps and add a new quality to marker assisted breeding. GABI-Beet aims to concentrate on regions of the genome with genes characteriztic for a root crop such as root diseases, sucrose storing and bolting. Sugar beet will thus be a model species for root crops as well as for crop species with small genomes. Those maps will give direct access to genes of agronomical importance. Wild beet species of Beta sections II and IV harbor valuable genes but due to strong crossing barriers, these species have not been used by beet breeders in the past. GABI-Beet aims to reveal relationships between cultivated and wild species of this genus. The main goals of GABI-Beet can be summarized as follows:

Milestones of the GABI-Beet project Kiel

- Construction of a BAC library of the B. vulgaris genome
- Sequencing allelic variants from ESTs and RFLPs
- High density mapping (ESTs, SNPs, STS)
- Comparative mapping (*B. procumbens*, *B. lomatogona*)
- Genome organization of repetitive DNA classes (LINE, Ty1-copia elements, Ty3-gypsy retrotransposons)
- Annotation of beet ESTs (subcontractor Munich)
- Wild beet populations and addition lines (subcontractor Braunschweig)
- Rhizoctonia resistance tests with one of the mapping populations (subcontractor Göttingen)

Results

Annotation und cluster analysis of ESTs (Munich)

From the 12,639 EST sequences available from the MPIZ Cologne 12,003 sequences (representing 6.3 Mbp of sequence) were clustered into a unigene set of 9,730 sequences. 8,004 sequences were singletons and 1,726 were multi-member clusters. The average of 1.23 ESTs per unigene is an extraordinary low value for an EST collection and demonstrates the value of the oligofingerprinting procedure used in the initial selection of unique cDNA clones.

The annotation of the unigene collection has again been performed using the SPUTNIK system. BLASTX comparison of the sugar beet sequences against the most recent update from the *Arabidopsis thaliana* genome reveals that only 4,670 sequences are homologous to Arabidopsis genes using the expectation value of 10e-30. 503 sequences can be matched to a non-redundant protein database of all known and predicted peptides using the expectation value of 10e-30. The genome plasticity between Arabidopsis and sugar beet is > 40% (3,506 out of 9,730 sequences). This difference could be explained by (a) quite different expressed and annotated gene sets in Arabidopsis and in sugar beet, (b) large number of contaminants, (c) selection for rarer sequences within the cDNA collection by oligonucleotide fingerprinting technique or (d) enrichment of sequences (i.e. small RNAs, small peptides) that have been missed in the annotation of the Arabidopsis genome, where gene modeling algorithms were trained to select for long sequences.

Development of high-density marker maps

For SNP and RFLP marker analysis, a representative set of 16 Beta-genotypes that contains conventional breeding material as well as 3 accessions of leaf beet, red beet and *B. maritima* was established and analysed for the abundance and distribution of SNPs by sequencing RFLP-derived PCR-fragments. For mapping, parents for four mapping populations were selected.

Selection of disease resistance-related ESTs

From the EST-database „SPUTNIK" those ESTs have been selected that show homology to disease resistance related sequences on nucleotide and amino acid level. This selection is focused on resistance gene analogs (RGA) as well as on sequences with homology to genes which are involved in general pathogen defence (defence related genes - DR). In addition, protein domains were used to identify relevant ESTs. Short conserved amino acid motifs were used in a BLAST-search to detect putative resistance related genes. So far, 187 ESTs have been selected to be processed in Kiel. Those ESTs will be mapped as RFLP markers and as SNP markers.

Mapping of resistance related ESTs as RFLPs and SNPs

Approximately 350 ESTs will be mapped which will be identified on their sequence homology to disease related genes. For RFLP mapping 72 EST-derived primer combinations were tested. Fourty-six PCR-fragments (64%) were generated from genomic DNA of sugar beet line KWS-2320. 15 polymorphic loci were identified and will be mapped in the mapping populations. Nine probes showed a highly repetitive hybridization pattern which makes identification of single polymorphic bands impossible. The mapping data will be send to the university of Halle for construction of an integrated genetic map.

For SNP marker development 37 primer combinations have been tested with DNA from 8 plants from the beet panel. With 19 primer combinations (51%) a specific PCR fragment could be generated with all the 8 genotypes. In parallel, all 187 primer combinations will be used for genotyping of the mapping populations. SNP genotyping will be done partly within the Beet project by spin off company TraitGenetics and another company specialized for SNP analysis using MALDI-TOF technology (GAG BioScience, Bremen). Together with the the MPIZ Köln and TraitGenetics Gatersleben we aim at mapping of 180 SNP-marker in one population and approximately 90 SNPs in a second population. We have started SNP detection by cleaving PCR fragments followed by agarose gel electrophoresis (CAPS,

○ U. Hohmann *et al.* **A bacterial artificial chromosome (BAC) library of sugar beet and a physical map encompassing the bolting gene B.** Mol Gen Genet (2003), 296, 126-136 ○ R. Herwig *et al.* **Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes.** Plant J. (2002) 32, 845-857. ○ S. Hunger *et al.* **Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (Beta vulgaris L.).** Genome (2003) 46, 70-82

cleaved amplified polymorphic sequences). Since a capillary sequencer became available a short time ago we will alternatively sequence alleles in the future in order to determine SNPs.

Comparison of the genome structure of wild and cultivated beets

The genome of B. vulgaris consists of some 60% repeated DNA. The majority of this DNA sequences can be assigned to satellite DNA and transposable elements. The aim of this research area within GABI-Beet is the identification, molecular characterization and chromosomal localization of repetitive sequences. Experimental work was focused on (a) isolation and molecular characterization of LTR retrotransposons and (b) identification and diversity analysis of class II transposable elements.

A new sugar beet Ty1-copia element has been isolated and designated Tbv1. Tbv1 is 5.2kb long and flanked by LTRs. The variability of the LTR indicates a considerable divergence of Tbv1 within the B. vulgaris genome which is in line with the wide-spread distribution within Beta species. The chromosomal localization of Tbv1 using fluorescent in situ hybridization (FISH) displayed that Tbv1 is present on all B. vulgaris chromosomes and scattered throughout the genome. There are no regions of particular exclusion or amplification. The clustering of hybridization sites is consistent with the nested organization of retrotransposons in plant genomes. A new Ty3-gypsy retrotransposon has been isolated and designated Beetle. So far, we have sequenced 7,698bp and analyzed the internal domain corresponding to the polyprotein of the retroelement. Transposons of ClassII are subdivided into superfamilies En/Spm, Ac/Ds, Mutator and Tc1/Mariner. Southern experiments showed that these transposons are highly amplified in species of the genus Beta. A recent PCR-based study has shown that Mariner transposons are also widely distributed in higher plant genomes. The comparison of Mariner transposons from B. vulgaris, B. corolliflora, B. nana, and B. procumbens revealed that these transposons exist in subfamilies within Beta genomes.

Comparative mapping of sugar beet and wild beet species from sections II and IV and development of addition lines (Braunschweig)

Two different crosses between four wild species have been produced. In the first cross, B. macrorhiza x B. lomatogona, two morphologically distinct accessions of B. macrorhiza are being used. In total, five pair crosses within the B. macrorhiza species using accessions from Turkey (BGRC63023) and Daghestan (BGRC58253) were conducted. In the second cross, B. procumbens x B. webbiana, two of the pair crosses currently flowering in spatial isolation to produce more seeds. Sufficient material will be available to construct the wild beet mapping populations. For the production of monosomic addition lines hybrids between B. patellaris x B. vulgaris subsp. vulgaris three accessions of B. patellaris (BGRC35322, 54753, 57667) of very different geographic origin were selected as female parents. In total, 126 pairs of genotypes (B. patellaris x B. vulgaris) were crossed yielding an average set of 3.2 seeds. All seeds (about 400) were analysed by a squash dot test with a B. vulgaris-specific repetitive probe. As a result, none of the plants turned out to be a hybrid between both species. This may indicate that such crosses using the wild beet as seed parent are not feasable.

Rhizoctonia resistance tests (Göttingen)

Significant technical improvement was achieved at the IfZ (Göttingen) to characterize and quantify the resistance to Rhizoctonia in sugar beet (fig 1). The evaluation of an ELISA test for the quantification of Rhizoctonia solani was continued and optimized. The method was now applied in addition to the visual monitoring of the fungal infection. The biotest on 104 S2 families provided by the Dieckmann breeding company revealed that the breeding material is segregating over a larger range as expected. First segregation analysis of monitored parameters showed that the Rhizoctonia resistance is segregating as an oligogenic character. To obtain an estimation of the number of genes involved in the resistance a QTL analysis will be performed using selected R-ESTs for mapping.

Construction of a large insert library from sugar beet

For positional cloning of candidate genes in GABI-BOLT and Sweet-GABI projects a representative BAC library with 57,600 clones and a 8.5x genome coverage was constructed. This achievement fulfills one of the milestones in GABI-Beet. At present, the BAC library is used in GABI-BOLT to clone the bolting gene (Hohmann *et al.* 2003). With the financial support from the Feuerwehrfonds of GABI a working copy of the library can be distributed to all GABI-Beet members. The fabrication of macroarrays (nylon membranes) and three dimensional pools of bacteria and DNA will allow the screening of the library with Southern hybridization and PCR based technology with the option of multiplexing.

Fig 1: *Sugar beet roots resistant and susceptible to Rhizoctonia solani.*

Genome analysis of sugar beet: a model species for root crops
(Subproject B)

Sugar beet
GABI-Beet

Silke Möhring, Ralf-Schäfer-Pregl, Jörg Wunder, Thomas Rosleff-Sörensen, Charlotte Bulich
ADIS (DNA core facility headed by Bernd Weisshaar), Katharina Schneider
Max Planck Institute for Plant Breeding Research (MPIZ), Cologne

Aims of the GABI-Beet project at MPIZ in Köln

The project GABI-Beet is focused on genome analysis of sugar beet, an important plant for sucrose production in regions of moderate climate. It is the aim to create state-of-the-art tools for structural and functional genome analysis and to make them available to more applied projects related to the breeding process.

In the group at MPIZ the tasks are threefold:

1. In the first part EST sequences of 10,000 unique cDNA clones are generated representing a source of putative candidate genes which influence agronomic traits. These cDNA clones and EST sequences will also be used for expression profiling using macroarray analysis in the Sweet-GABI project.
2. It is a further aim at MPIZ to establish and maintain an interactive database called „Beetbase" to collect all primary data generated in GABI-Beet. Beetbase is intended to serve as a bioinformatics tool for the applied needs of sugar beet geneticists supporting and complementing the resource centre GABI-Info at MIPS (Munich Information Center for Protein Sequences).
3. In the field of molecular markers, it is the aim to generate 350 single nucleotide polymorphism or „SNP" markers. Therefore SNPs are first detected in a panel of selected breeding lines by comparative sequencing. Genetic mapping of these markers is expected to improve the quality of the existing genetic maps of sugar beet and to identify genes and genomic regions with impact on agronomic traits. The analysis shall also include an evaluation of different SNP mapping techniques to select a method with high scientific and economic efficiency.

Results

EST sequencing

As a resource for expressed sequences from sugar beet, a set of 10.000 unique cDNAs from four different tissues (Herwig *et al.* 2002) was sequenced within GABI-Beet. In a first run 5' end sequences were determined by the DNA core facility ADIS on an Applied Biosystems (Weiterstadt, Germany) Abi Prism 3700 DNA analyzer based on 96 capillaries for high-throughput sequencing and BigDye-terminator chemistry. In the initial two rounds of sequencing high quality sequences of 432 bp length on average were generated for 97.3% of all clones and transmitted to Beetbase. The results of resequencing 96 randomly selected clones from all of the 28 384-plates confirmed previous data in 93.7% of the cases. All sequence data were transferred to the GABI primary database in Berlin on Feb. 15th, 2002.

The Beetbase

The Beetbase is intended to provide an exchange platform for the GABI-Beet partners who want to generate and use data on EST sequences, SNP analysis and marker development. The DNA sequences of 200 known RFLP markers including primer sequences for specific PCR-based STS assays are available from the Beetbase. For 63 converted RFLP markers sequence alignments generated from test panel genotypes (see below) and tables with information on position and type of nearly 1500 SNPs were integrated as well. EST sequence data were analysed after removal of vector and adaptor sequences. A data set of 14779 EST sequences was passed on to MIPS for detailed functional annotation. A direct crosslink between the two databases was established to enable the use of sequence annotation tools from MIPS. The access to Beetbase is possible through a password protected https-page via the www. The graphical-user-interface (GUI) was created using a combination of HTML- and PHP-code as well as 'Javascript' und 'Java - applets' for some special applications like the interactive alignment editor 'Jalview'. Currently the EST-derived SNPs of the three GABI-Beet partners developing markers are being integrated, a new enhanced version of the GUI is under way and the Beetbase will be improved with respect to data accessibility and security.

Development of SNP markers

The targets for SNP marker development are EST sequences with homology to genes involved in carbohydrate metabolism and beet development. Secondly, known RFLP markers were converted to SNP markers as references in the genetic map. The first step in the development of SNP markers was the establishment of STS (sequence tagged sites) assays to generate the targets for SNP analysis. For 111 RFLP markers specific STS assays were developed. With respect to the ESTs, in total 276 were worked on.

To analyse polymorphisms, the markers were tested on the sugar beet screening panel which consists of 13 parental and F1 genotypes of the six different mapping populations, namely DIE1, DIE2, DIE3, KWS1, KWS2, KWS3. Additionally one genotype each of red beet, Swiss chard and Beta vulgaris ssp. maritima are included. The PCR fragments of 248 EST- and 89 RFLP-related sequences were prescreened for polymorphisms by SSCP (single strand conformation polymorphism) in non-denaturing polyacrylamide gels (example in Fig 1). This method is known to detect polymorphisms at the SNP level. The average degree of polymorphism in the six mapping populations varies from 36 to 72%. On average 3.3 different alleles were detected per marker in the EST- and 6.7 alleles in the RFLP-related sequences among the screening panel genotypes.

The PCR products of 206 STS markers were sequenced in 8 to 16 screening panel genotypes by the ADIS unit at MPIZ. Sequence analysis and alignments were performed using the software of Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisconsin. The sequences of 37 STS markers generated from at least 14 genotypes were analysed to determine the SNP distribution within the screening panel. The analysis was performed separately for RFLP-related sequences and coding as well as non-coding regions of ESTs. A summary of averaged data is given

○ S. Hunger *et al.* **Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris L.*).** Genome (2003) 46: 70-82. ○ R. Herwig *et al.* **Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes.** Plant J. (2002) 32, 845-857. ○ S. Möhring *et al.* **Multiplexed, linkage group-specific SNP marker sets for rapid genetic mapping and fingerprinting of sugar beet (*Beta vulgaris* L.).** Molecular Breeding (2004) 14:475-488.

in table 1. In total 416 SNPs were detected. Concerning the type of nucleotide substitution, about 61% transitions and 38% transversions were identified whereby the transition/transversion ratio was higher in coding regions than in RFLP-derived or non-coding regions. Triallelic SNPs were found at a frequency of approximately 0.33%. The frequency of the less abundant allele of the SNPs was 25%.

The average nucleotide diversity π_a was computed as the arithmetic mean of all pairwise sequence comparisons in which the number of differences per nucleotide site was divided by the sequence length in base pairs. On average the nucleotide diversity was π_a = 8.49 x 10-3. The average nucleotide diversity showed a higher value in the RFLP-related sequences and the non-coding regions compared to the coding region which can be explained by the higher selection pressure on coding sequences.

Genetic mapping

In total 221 STS markers based on RFLP- and EST-derived sequences were so far mapped in the three populations DIE2, KWS1 and KWS2 with the main focus on the two populations DIE2 and KWS1. Ninety-eight EST markers were mapped in DIE2 and 141 KWS1. Apart from 8 dominant markers, all markers were scored codominantly. The mapping data for 24 converted RFLP markers in the KWS2 population were part of a publication (Hunger *et al.* 2003). For 155 STS markers segregation analysis was performed by SSCP analysis. Four STS markers were analysed by the CAPS (cleaved amplified polymorphic sequence) method.

The SNaPshotTM multiplex reaction, a primer extension-based SNP scoring method developed by Applied Biosystems, was established as a method for increased throughput. A protocol was developed to multiplex up to six primer pairs for PCR amplification. For SNP genotyping, up to six loci were successfully analysed in parallel in linkage group-specific multiplexes (Möhring *et al.* in press). One hundred and twelve SNaPshot markers were mapped in the DIE2, KWS1 and KWS2 population, respectively. Twenty-five SNaPshot markers were mapped in the DIE2, KWS1 and KWS2 population, respectively. The whole population or random samples were retested by SSCP and CAPS analysis to confirm the SNP scoring results.

To test a further SNP scoring method based on MALDI-TOF analysis (Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry) and to evaluate the option of an external service, a pilot project with five markers was carried out together with the company GAG BioScience (Bremen) in 2002. For four markers the results obtained by MALDI-TOF and SSCP were consistent in 95% of the samples confirming the suitability of the MALDI-TOF analysis for mapping purposes in sugar beet. In total 68 markers were successfully analysed in this external facility.

	Sequence length [bp]	Number of SNPs	Average nucleotide diversity [π_a]	Frequency of less abundant allele [%]	Transitions [%]	Transversions [%]	Triallelic SNP [%]
data of 37 marker loci	317007	416	8.49 x 10-3	25	61.33	38.33	0.33

Tab 1: *Characterization of SNPs in PCR fragments amplified from the screening panel genotypes based on 37 different STS markers.*

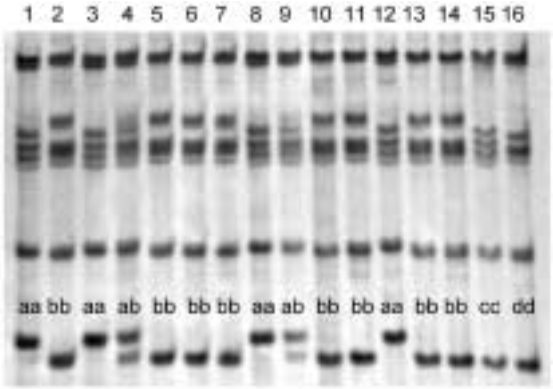


Fig 1: *SSCP analysis of the marker MP124 after Mse I digest in the 16 screening panel genotypes.*

Genome analysis of sugar beet, a model species for root crops
(Subproject C)

Sugar beet
GABI-Beet

Dagmar Kulosa, Andreas Polley, Susanne Reich, Carina Kehling and Martin Ganal
TraitGenetics GmbH, Gatersleben

Molecular markers are important breeding tools that are used for the identification and transfer of agronomically important traits into elite breeding lines and new varieties. At present various molecular marker techniques such as microsatellites (SSR, simple sequence repeats) or AFLPs (amplified fragment length polymorphisms) are being used in breeding laboratories. Their main disadvantages are limited availability, high costs per datapoint and other factors. Single nucleotide polymorphisms (SNPs) are a molecular marker system that is based on the difference of single bases between two individuals. SNPs are markers that can be found anywhere in the genome in nearly unlimited numbers since every nucleotide difference between individuals can be exploited as a molecular marker. In the last years, a variety of technologies have been developed, predominantly in human genome analysis, that can be used for the detection, mapping and analysis of SNPs in large numbers. For practical plant breeding, the potential of SNP markers is that they can be found in candidate genes for important traits and used for association studies. Furthermore, multiparallel SNP analysis offers a possibility to reduce costs in molecular marker analysis for plant breeding in marker-assisted backcross or selection projects or QTL (quantitative trait loci) analyses.

The project part currently performed at Gatersleben is predominantly part of an concentrated effort of the GABI-Beet partners at Kiel, Cologne and Gatersleben to identify SNPs in approximately 1,000 sugar beet genes and map the majority of them in segregating populations. Each partner shall identify SNPs in approximately 330 ESTs. Starting point for this project was the available set of more than 10,000 ESTs that has been sequenced by the project partner in Cologne and provided to the project partners.

In a first step, the entire EST set was used to design primers for amplification of each EST using bioinformatic tools. This included automated quality clipping, vector removal and primer design for amplification products in the size range of approximately 500 base pairs. At Gatersleben, ESTs from the functional categories stress, transcription, protein destination and protein synthesis were used for SNP detection. Since this was not sufficient to achieve the specified number of ESTs, additional randomly chosen ESTs were selected for SNP analysis.

Under highly standardized conditions, primers were tested for amplification in a small panel of sugarbeet lines in a first step. Functional primer pairs defined by one specific amplification product were subsequently used for the amplification of a panel consisting of the parents and F1 individuals of a number of sugar beet mapping populations provided by the breeding companies and partners in Sülbeck (Fa. A Strube-Dieckmann) and Einbeck (Fa. KWS SAAT AG), as well as Beta maritima, root beet and Swiss chard as close relatives of sugar beet.

For SNP identification, amplification products of those primers that yielded single fragments in most or all of the panel lines were sequenced from both sides and analysed for SNPs by sequence alignment (fig 1). At present, primers for more than 1,000 ESTs have been tested in this way at TraitGenetics. Approximately 400 ESTs have yielded single fragments in the sugar beet panel lines under highly standardized conditions. Those 400 functional primer pairs have been sequenced in the sugar beet panel lines. 75% of the investigated amplicons have shown one or more SNPs (fig. 2). 20% of the amplicons have shown one or more InDels (insertion/deletion). Taking together the sequence data of more than 200 kb of investigated genomic sequences, one SNP was found on average every 83 base pairs and one InDel every 736 base pairs resulting in one sequence polymorphism (SNP or InDel) every 75 base pairs reflecting the high level of polymorphism in this outbreeding plant species. More than 50% of all investigated amplicons displayed more than 2 SNPs so that the haplotype structure within that gene could be investigated (Fig 3). More than 5 different haplotypes could be identified in some genes confirming the high level of genetic variation in the sugar beet gene pool.

Currently, the SNP data are being collected in a database that comprises all relevant information for individual ESTs and SNPs including sequence polymorphism information and allele/haplotype tables for each SNP as well as quality control data. Additionally, a database is under construction that permits the selection of individual SNPs based on their polymorphism in individual mapping populations for further assay development.

For genetic mapping of SNPs, DNA has been extracted from the individuals of a set of populations. SNP mapping has been started based on the available technology at TraitGenetics (capillary electrophoresis on ABI31000) using primer extension methods. At present, this has the advantage that SNP analysis can be performed on the same equipment as microsatellite analysis and thus the developed SNP markers can complement SSR analysis in order to achieve a higher marker density in mapping projects of agriculturally important traits. First results on a number of SNPs indicate that by multiplexing the SNP detection procedure, it should be possible to reduce the costs per datapoint to less than 1 Euro compared to a microsatellite datapoint at 2-3 Euro. In parallel, work is in progress to investigate the quality and costs of other SNP detection methods such as mass spectroscopy and other fluorogenic assays for sugar beet genome analysis and mapping.

TraitGenetics will have a set of more than 300 mapped genes available for the sugar beet genome at the end of the project. Together with the SNPs identified by the other project partners, this set can be used in association genetics and for the complementation of microsatellite mapping in breeding projects. The project serves as a starting point to identify SNPs in many more genes so that in the long term SNP markers could replace other marker systems in sugar beet breeding.

○ R. Herwig *et al.* **Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25000 potential sugar beet genes**. The Plant Journal (2002) 32, 845-857.

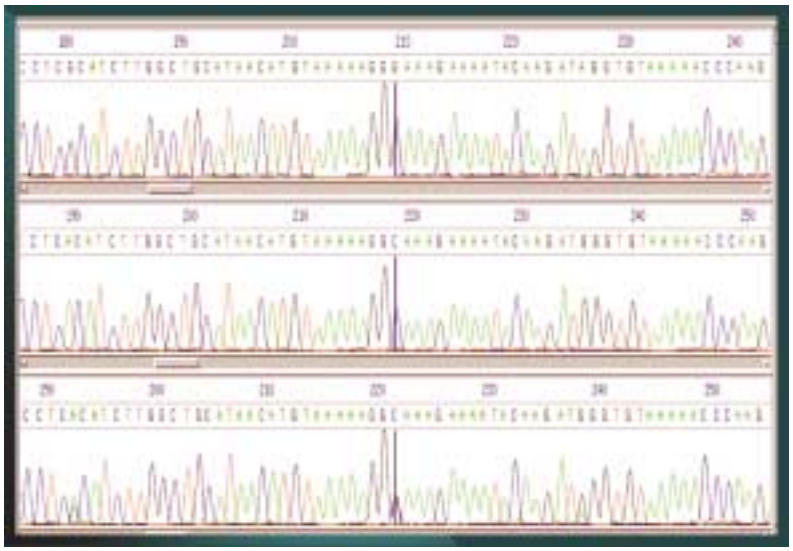
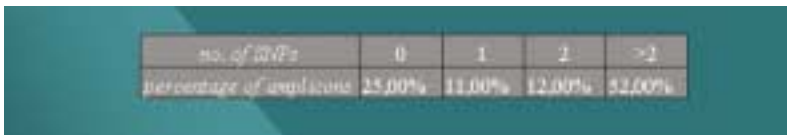
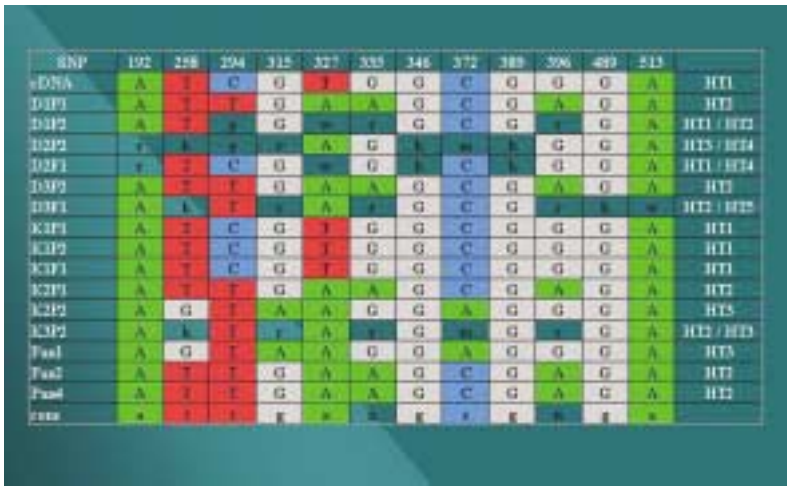


Fig 1: *SNP identification in a sugar beet EST. The vertical line shows an SNP in the two parents and the F1 of a mapping population.*



no. of SNPs	0	1	2	>2
percentage of amplicons	25.00%	11.00%	12.00%	52.00%

Fig 2: *Level of polymorphism in the investigated amplicons.*



SNP	192	258	294	315	327	335	346	372	389	396	489	513	
cDNA	A	T	C	G	T	G	G	C	G	G	G	A	HT1
D1P1	A	T	T	G	A	A	G	C	G	A	G	A	HT3
D1P2	A	T	T	G	W	T	G	C	G	T	G	A	HT1 / HT2
D2P2	r	k	s	s	A	G	k	m	k	G	G	A	HT3 / HT4
D2F1	r	T	C	G	W	G	T	C	k	G	G	A	HT1 / HT4
D3P2	A	T	T	G	A	A	G	C	G	A	G	A	HT3
D3F1	A	k	T	s	A	s	G	C	G	s	k	w	HT2 / HT5
K1P1	A	T	C	G	T	G	G	C	G	G	G	A	HT1
K1P2	A	T	C	G	T	G	G	C	G	G	G	A	HT1
K1F1	A	T	C	G	T	G	G	C	G	G	G	A	HT1
K2P1	A	T	T	G	A	A	G	C	G	A	G	A	HT2
K2P2	A	G	T	A	A	G	G	A	G	G	G	A	HT5
K3P1	A	k	T	r	A	s	G	m	G	s	G	A	HT2 / HT3
Fus1	A	G	T	A	A	G	G	A	G	G	G	A	HT5
Fus2	A	T	T	G	A	A	G	C	G	A	G	A	HT1
Fus3	A	T	T	G	A	A	G	C	G	A	G	A	HT2
cons	a	t	t	g	w	m	g	c	g	w	g	a	

Fig 3: *SNPs within an amplicon derived from a sugar beet EST. Deduced haplotypes are displayed on the right side.*

Genome analysis of sugar beet, a model species for root crops
(Subproject D)

Sugar beet
GABI-Beet

Georg Koch
A. Strube-Dieckmann, Nienstädt

Objectives

- Long-term storage of reference populations, segregating populations and panel plants.
- Development of populations segregating for Rhizoctonia resistance.
- Clonal propagation of plants and leaf sampling as well as large scale DNA extractions.
- Adaptation of SNP methods and high-throughput screening technologies at the company site adapted to the needs of plant breeding.
- Contributing typical breeding and genetic knowledge to the project.

Results

Long-term storage of reference populations, segregating populations and panel plants

Panel plants, reference population plants and the plants from the segregating population are in long-term tissue culture storage (gene bank), in total approx. 350 genotypes with each more than 5 clones.

Clonal propagation of plants and leaf sampling as well as large scale DNA extractions

All plants were in vitro multiplied, rooted and transplanted to achieve sufficient plant material for the whole project and all project partners. In most cases DNA was extracted on site, also in large quantities (> 1mg DNA per genotype).

Development of a segregating population with resistance to *Rhizoctonia solani* AG 2-2 IIIB

Population development were performed according the scheme shown in Fig 1. A resistant single plant was hand crossed with a susceptible elite single plant, a plant of the progenies (F1) selfed and 250 plants were selected out of those F2 progenies and again selfed to produce F3 families. All selfing steps were performed under strict isolation to prevent foreign pollen from cross fertilisation. The success of the hand crosses as well as the confirmation of the expected segregation of the F2 selfing population was proved with molecular markers.

In parallel several hand crosses from two different resistance sources were produced and the resulting hybrids by means of STS or SSR markers confirmed (Fig 2) and under strict isolation selfed. Several of those F2 populations were chosen for further analysis and the expected 1:2:1 segregation corresponding a Hardy-Weinberg-population proved by PCR markers. These results were used as supportive confirmation that the strict selfing was successful.

Prove of the resistance to Rhizoctonia solani AG 2-2 IIIB in the segregating F2 populations with a rapid bioassay

After completion of the F2 seed increase six F2 populations were chosen. These six F2 populations were subjected to a Rhizoctonia bioassay in the greenhouse. This rapid bioassay method should confirm segregation of Rhizoctonia solani AG 2-2 IIIB resistance and allowed us to judge on the relative resistance strength segregating in the different populations. Test method are adopted from several known protocols: (1) Pre-culture of single plants in multi-pot plates or 7cm or 9cm pots and Rhizoctonia-free soil substrate for ca. 2 months; (2) afterwards inoculation of the petioles with ca. 12 under appropriate conditions prepared infected barley grains and (3) scoring of the percentage of the diseased black root area ca. 3-4 weeks after inoculation. These assays are not intended for high quality disease ratings which are necessary for the mapping effort. However, results are appropriate for an estimation on the segregation of resistance/susceptibility and relative resistance strength in the various populations. Population DIE-4 F2 was finally chosen according their strong resistance. Population DIE-3 F2 was also advanced to F3 as backup population in case of any problem during seed increase of DIE-4 F2, but later not used.

Available material of the chosen segregating population DIE-4 F2 for mapping resistance to Rhizoctonia solani AG 2-2 IIIB

Seed was harvested and transferred to partner IfZ for resistance testing. S2 single plants leaf samples were taken from the plants out of the bioassay at IfZ, freeze-dried and stored until further need.

In parallel to the seed production 5g leaf samples of each F2 single plant were harvested for marker analysis. In limited amounts DNA of the parental plants was available and transferred to partner CAU.

The F2 single plants are initiated in tissue culture and will be available throughout the project for regeneration of the identical genotypes.

Sequencing of the RFLP clones from chromosome 3 and 4

To fuel the SNP discovery at the sites of the university partners which started a few weeks later with the project, we provided partner Kiel with initial sequence information. Plasmid DNA of the clones from Chromosome 3 (28 RFLPs) and 4 (15 RFLPs) were produced, sequenced for the RFLP clones of chromosome 3 and the sequences delivered to partner Kiel.

Sequences will be put into publicly available data bases by our project partner Kiel.

Set-up of a SNP detection method

Initial idea was to establish SSCP methods with self-designed cooling plates optimised for higher throughput. However, during the first year of the project sequencing as a method to uncover SNPs as well as several standard methods to detect known SNPs in samples were successfully established and are now used widely and routinely. Commercial laboratories offering services which are very attractive with respect to costs and throughput. Further improvements in the price per data point are expected. This fast and partially unexpected market and technology progress makes a gel based SNP detection method more and more unattractive and less competitive. Simultaneously partner MPIZ could show that the advantages of SSCPs are linked to several drawbacks (laborious and limited in their potential to detect all available SNPs under restrictive but practical conditions). Therefore, we have not continued this approach and switched over to an evaluation of commercial service providers and to stronger efforts to convert the collected SNP information provided by the partners in the project to CAPS or InDel markers which could be separated very time- and cost efficient in simple agarose gels. Additionally, the generated SNP information need to be integrated and tested for robustness, reliability and polymorphism rate in our germplasm. This work can only be in part done by the project partners and is a very vital part of final success of the SNP development.

Set-up of SNP identification and SNP detection in practical breeding processes

We started to prepare and to collect the material for a model SNP detection. In addition and completion to the work done by the project partners we included specific material groups which are important in the breeding process like diploid and tetraploid accessions. For that material we have sequenced exemplified loci to detect uncovered SNPs with the available standard methods. The goal is to establish methods and procedures which do fit in the practical breeding work.

Together with Pyrosequencing Inc. assays were developed to differentiate the allele doses at SNP loci. A typical successful application is shown in Figure 3. Peak heights revealing the allele doses differences of Simplex, Duplex and Nulliplex plants (from bottom to top, C peak).

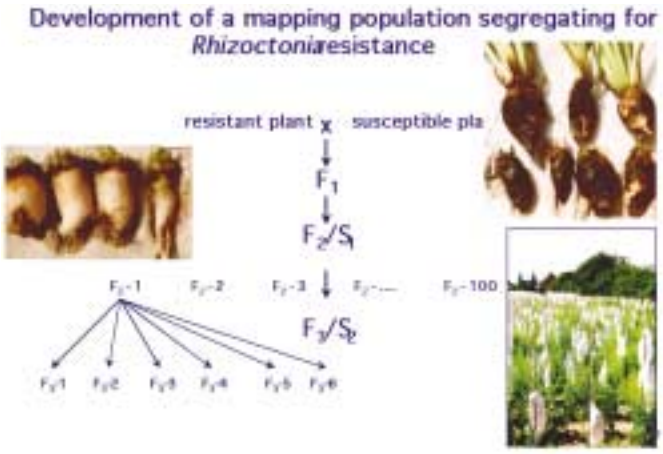


Fig 1: *Path of the population development.*

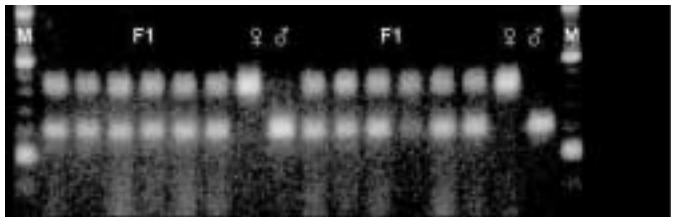


Fig 2: *Prove of the hand crosses for a successful crossing; confirmation of hybridity.*

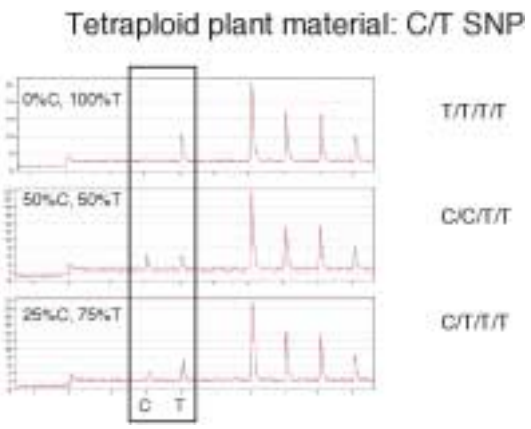


Fig 3: *Pyrosequencing SNP assay differentiating allele doses at SNP loci.*

Cloning of the bolting gene B from sugar beet (GABI-BOLT)

Sugar beet
GABI-BOLT

1 Christian Jung, Uwe Hohmann, 2 Britta Schulz, 3 Georg Koch
1 Christian-Albrecht University of Kiel, Department of Crop Science and Plant Breeding, Kiel
2 KWS SAAT AG, Einbeck
3 Strube-Diekmann, Nienstädt

Objectives of GABI-BOLT

The aim of the GABI-BOLT project is to clone the bolting gene B and to design an easy to perform marker test relying on the B sequence for routine screening of seed lots. The genetic manipulation of the bolting behaviour of sugar beet in combination with marker-assisted selection will have major impact on sugar beet breeding. It will be a scientific and commercial challenge of the following years after the cloning of the B gene to study the feasibility and to exploit the B gene by utilisation of annual lines during breeding and complete bolting resistant hybrids for root production in commercial fields. Sugar beet can serve as a model for root crops. Cloning of the B gene will allow further analysis of the flower induction including the vernalisation requirement. The gene sequences could be used as a heterologous probe to identify similar genes in other biennial crops.

Results

High-density genetic map of the gene B-region

Prior to this project, the B locus had been mapped with RFLP markers to chromosome 2 of beet. Fine mapping had been done with AFLP markers in combination with bulked segregant analysis (El-Mezawy *et al.* 2002). Altogether 10 markers had been located within a 2.6 cM region around the gene using a mapping population of 232 F2 plants. Later, a map based on 775 plants and more than 50 markers linked to the gene in a 25 cM interval was established. During GABI-BOLT, a high resolution genetic map based on 2134 plants of a segregating F2 population and four AFLP markers in a 0.64 cM region around gene B was constructed. This genetic map is the basis for physical mapping of the gene B region.

Large insert library screening

Before the project started, 8 YACs had been isolated with the help of AFLP-markers 17 linked to the B gene and a first physical map had been constructed. The YAC termini and the B linked AFLP markers were used for screening a BAC library with higher genome coverage. A BAC library of a biennial (bb) genotype with 57,000. clones (8.5 x genome equivalents) became available by the GABI-Beet project by the end of 2000. 12 YAC termini and four AFLP marker fragments that had been used to identify 56 different BAC clones from the gene B region. Of these, 10 BACs gave signals with more than one probe indicative for a tight relationship between genetic and physical distances.

Identification of overlapping BACs

The end sequences of 49 BACs were determined and end specific sequences of 42 BAC ends were amplified by PCR using BAC end specific primers and the BACs as templates. For identifying overlapping BACs, a two-step-PCR based screening strategy was followed. On the basis of PCR results two contigs were constructed from the gene B region. The BAC contigs are primarily based on the BACs that have been identified with the AFLP primer combination P05 and the b-linked marker allele. Hence, the most closely linked b-allele marker P05_b162 is present in one of the two contigs. One contig contains 8 overlapping BACs. The other consists of 14 BACs that are overlapping by at least one BAC end. However, the orientation of individual BACs can not be determined unequivocally unless both ends have been mapped and overlapping BACs have been identified. The contig analysis using PCR-fingerprinting was complemented with RFLP-fingerprinting und AFLP-fingerprinting techniques and revealed 3 contigs with minimal tiling path lengths of 271,3 kb, 247,5 kb and 404,2 kb, respectively. In total, the three contigs from a 3.2 cM region are spanning 923 kb.

Gene B contig

Four BACs (B1, B10, B22 and B70) spanning 360 kb cover a genetic distance of 1.8 cM (between marker loci p14_b118 and p05_162). The ratio between physical and genetic distance in the gene B region was calculated to be 200kb/cM which is substantially lower than the estimated average ratio for the whole genome which is 1.1 Mbp of DNA/cM. The ratio between genetic and physical distances are in correspondence with previous findings from positional cloning experiments with other plant species. The ratios between physical and genetic distances in the given genome region is one important parameter for the success of a positional cloning experiment. The genetic distance between AFLPs p05_b162 and p04 B159 was calculated to be 0.14 cM. The two BACs B10 and B22, overlap by approximately 50 kb, thus both together span a region of about 200 kb (fig 1).

Mapping with BAC-derived probes

For the integration of physical and genetic maps we focused on anchoring the BACs from the physical map to the genetic map. The strategy for genetic mapping of BAC ends was performed by a) determination of the genomic copy number of the termini, b) converting BAC termini into markers and mapping these markers with plants from the high-density mapping population of 2134 F2 individuals. So far, from 32 BAC end sequences 10 have been converted into STS- or CAPS marker and then mapped in those plants to fine characterize the recombination breakpoint. Two dominant and three co-dominant CAPS marker were developed and used for mapping. An example is given in figure 1. CAPS marker 29tAluI yielded three bands visible with the BB and bb genotype. The largest band is common for both genotypes; the polymorphic bands are indicated by arrows (fig 2). In analyzing non-bolting recombinants the five markers co-segregated perfectly with the phenotype. Genetic mapping with the new developed BAC end-derived markers will refine the position of gene B (fig 2).

○ U. Hohmann *et al.* **A bacterial artificial chromosome (BAC) library of sugar beet and a physical map encompassing the bolting gene B.** Mol Gen Genet (2003), 269, 126-136 ○ A. El-Mezawy *et al.* **High resolution mapping of the bolting gene B of sugar beet.** Theor Appl Genet (2002) 105,100-105.

Marker test in breeding material

The seed companies Strube-Heimburg and KWS SAAT AG tested three markers (102R, 29t und 131s) in their breeing material. So far, none of them was suitable for all genotypes to differentiate the B and b alleles. Therefore, additional markers from the gene B contig will be developed.

Shot gun libraries from candidate BACs

Three BACs (BAC1, BAC70 and BAC131) were selected to construct shot-gun libraries. Single sequence reads of 96 clones from each BAC revealed more than 150 kb non-redundant DNA sequence information and more than 50 potential ORFs. At present, some of these ORFs will be converted into co-dominant SNP markers for marker-assisted selection and determination of candidates for bolting gene by analysis EMS mutants.

EMS mutant analysis

To increase genetic variation at the B locus by producing loss of function mutants approximately 10,000 seeds of BB plants were EMS treated. The EMS treatment was optimized and performed in different series with varying concentrations of EMS (0,5 – 1%) and different exposure times to EMS. The effect of the mutant treatment was proven by reduced or increased growth, necrotic lesions, modified cotyledons (number and form). More than 66,000 M2 plants obtained from selfing 3,300 M1 parents were investigated in the field. Among these, nine non-bolting plants have been identified from nine different M2 families. Two additional progenies have been selected with single plants showing reduced height. Moreover, 50 progenies were selected for phenotypes with reduced growth and delayed bolting tendency. These plants will be analyzed with candidate ORFs for the detection of mutant alles.

cDNA library construction

Because the B allele will be absent from the BAC library (bb), it is planned either screen two cDNA libraries from a bolting genotype, one in the time range of four weeks before bolting and one four weeks after the onset of bolting. RNA from two different genotypes (930190: bolting and KWS2320: non-bolting) and different tissues (leaf, stem, flowers, and roots) will be isolated to determine the spatial and temporal expression of the gene. Construction of cDNA libraries by the RZPD in Heidelberg (cDNA cloning) and Berlin (picking and arraying) has been approved by the SCC of GABI.

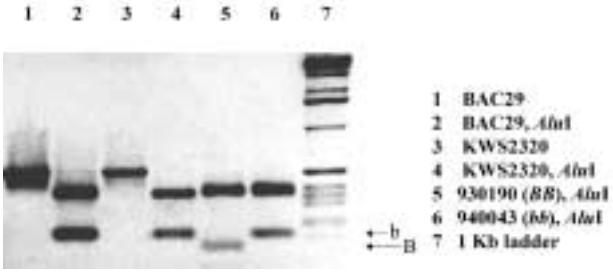


Fig 1: *Minimal tiling path for the central gene B region.*

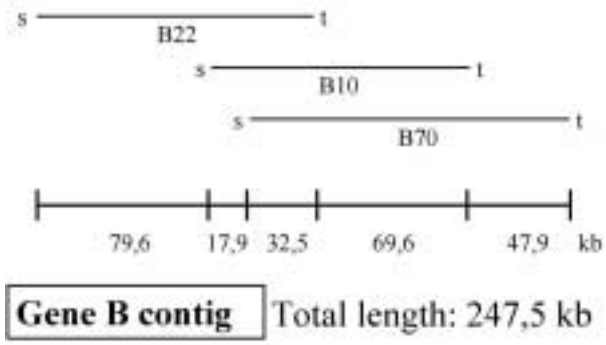


Fig 2: *CAPS marker 29t. DNA of BAC29 and of genomic sugar beet DNA from DH line KWS2320 (BAC donor plant, bb), 930190 (BB) and 940043 (bb) was amplified with primer pair 29t. Some of the PCR products were digested with restriction enzyme AluI and the products separated on a 1.5% agarose gel.*

Fig 3: *EMS mutants in sugar beet. Non-bolting (left) and late bolting (right) plants were identified.*

Sweet-GABI: Crosslinking Structural and Functional Genomics with Regard to Metabolism and Development in Sugar Beet

Sugar beet
Sweet-GABI

1 Britta Schulz, 2 Katharina Schneider
1 KWS SAAT AG, Einbeck
2 National Research Center for Environment and Health (GSF), Neuherberg

General outline and aims of the project

In Sweet-GABI it is the aim to combine structural and functional genomics to analyze sugar metabolism and root development in sugar beet.

Genomic information already available and generated within GABI-Beet is the basis for this project. In particular, Sweet-GABI uses the sequence information from ESTs. Differential gene expression is analyzed using new technologies that enable transcription profiling. This way new genes will be identified which play a key role in metabolic and developmental processes. To relate the molecular information to phenotypic effects, a QTL analysis will be performed and all new candidate genes are tested for their influence on the traits investigated. This approach is intended to provide new molecular markers for the selection of sugar beet plants with improved quality traits in commercial breeding programs.

Task of project partner KWS

KWS provides plant material and phenotypic data for the project. A segregating F_2 population has been established, resulting from the cross Beta vulgaris var. altissima (sugar beet) x Beta vulgaris var. conditiva (red beet). The individual F_2 plants were selfed to generate F_3 lines for phenotypic evaluation of traits relevant for tap root development and sucrose content. QTL analyses will be performed to identify genetic factors influencing beet anatomy as well as sugar yield and physiology.

To assess the extent of the phenotypic variation within this cross, plots of the F_2 population have been included. In 2003 the F_2 lines of the whole population will be grown and tested in replicated field trials to generate the data for QTL analysis.

Task of project partner MPIZ

A combination of new methods of functional genomics and genetic analysis will be applied at the MPIZ to address the questions of the Sweet-GABI project. The work plan comprises the identification, characterization and verification of candidate genes relevant for tap root development, traits related to processing quality as well as carbohydrate metabolism and sucrose content

Characterization of plant material to define stages of tap root development

A study of the basic anatomy of tap root development was performed to define developmental stages and tissues for RNA extraction, cDNA synthesis and differential analysis. For this purpose two different genotypes were grown in the field to study the development of the sugar beet plant. Plant samples were taken in regular intervals. Plant size, weight and the number of cambium rings in the tap root were recorded. Microscopic sections of the early stages were documented to follow cell division and expansion of the beet. In a comparison of the different samples, the inner rings showed the highest sucrose content. These phenotypic data will be the basis for selecting tissues to be analyzed at the molecular genetic level.

Molecular genetic analysis

One major goal of the project is to identify transcripts which are specifically expressed at defined developmental stages and tissues which can be considered candidate genes for quality traits of sugar beet based on their expression pattern. To perform expression profiling, we established two techniques: cDNA AFLP and cDNA macroarrays.

In a first step, transcripts were classified with respect to their expression in tap root compared to leaf, inflorescence and seed tissue. Genes, which showed predominant expression in the tap root, were selected.

Results obtained by the cDNA AFLP approach

Using 125 primer combinations, cDNAs from 9 different tap root samples, 4 leaf samples, an inflorescence and a seed sample, were analyzed, respectively. A total of 50 transcript-derived fragments, which mainly occurred in tap roots, were identified. For 32 fragments, specific PCR-based assays could be developed, and the expression pattern could be verified by RT-PCR experiments. According to DNA sequence analysis, they specify 19 different genes. To obtain more sequence information, 3' and 5' RACE (rapid amplification of cDNA ends) experiments were established and allowed the extension of 13 fragments. In 10 cases, additional sequence information could be retrieved from matching ESTs. For 7 candidates, there were no homologous genes with known function in the databases, the others showed homologies to gene products mainly involved in signal transduction and cell wall metabolism.

Results of the macroarray analysis

An integrated approach involving EST sequencing, data mining and multiparallel expression profiling by macroarray analysis was established to identify tap root-expressed genes. The DNA sequences of 3250 ESTs derived from young sugar beet plants specified 2049 unique gene products with putative functions in primary and secondary metabolism (651), transport processes (136), signal transduction (78) and cellular organization (39).

Expression profiles from roots, leaves and inflorescences of field-grown plant material were generated. Two different samples of each organ were analyzed to reduce sampling effects, which accounted on average for 30.3% of spots with at least two-fold deviation. For 76 unique cDNAs, the amount of detected transcript in roots was at least twice as high as in other organs. Macroarray expression data were confirmed by Northern blot analysis and quantitative RT-PCR experiments concerning eleven cDNAs.

The experiments are now extended to the GABI-Beet set of 10,000 cDNA clones. Further differential analysis will be performed with these filters to compare samples from the time course and tissues from the inner and outer cambium rings.

Verification of candidate genes

A functional genetic map of sugar beet, containing the map positions of 75 genes selected from the metabolic pathways related to carbohydrate metabolism was available at the beginning of the project. This map has been extended by adding genes with preferential transcription in the tap root as identified by cDNA-AFLP and by macroarray analysis. So far 44 new genes have been placed on the genetic map.

The mapping activities in the new cross (sugar beet x red table beet) have started. The degree of polymorphism in this population is rather high, as expected for a cross between subspecies.

One approach to validate candidate genes at the genetic level involves association studies, exploiting linkage disequilibrium. Candidate genes were chosen based on their linkage to QTLs for sucrose content and quality traits as well as based on the results of the cDNA-AFLP analysis. The allelic configuration of these genes was determined in breeding lines by SSCP (single strand conformation polymorphism) analysis. On average 3 to 6 haplotypes could be identified. A random selection of the SSCP alleles have been verified by comparative sequencing of the PCR fragments amplified from the breeding lines.

The statistical methods to detect a possible association between a specific allele and a phenotype may still have to be adjusted to the experimental design. DNA sequence analysis will then reveal the molecular origin of the polymorphisms and contribute to understanding the associated phenotypic differences for validated candidate genes.

Sweet-GABI has already yielded a vast amount of data, especially with respect to the identification of potential candidate genes. Genetic mapping and association studies have also been established as methods to crosslink functional to structural genetics. The project has generated extremely valuable tools like macroarray technology that can be applied to compare expression patterns of any given set of tissues, developmental stages and genotypes. The new segregating population of sugar beet x red table beet is a further resource that can be used to for the genetic analysis of additional quantitative traits.

In the future,we want to go into more detail with promising candidate genes. This will include additional molecular, biochemical and physiological experiments to characterize their functions and to relate allelic differences in the functions of the respective genes. Such analyses will reveal the end products of gene activity and contribute to the verification of candidate genes.

Fig 1: *Segregation for shape and colour of the tap roots, derived from a cross between sugar beet (Beta vulgaris var. altissima) and red table beet (Beta vulgaris var. conditiva).*

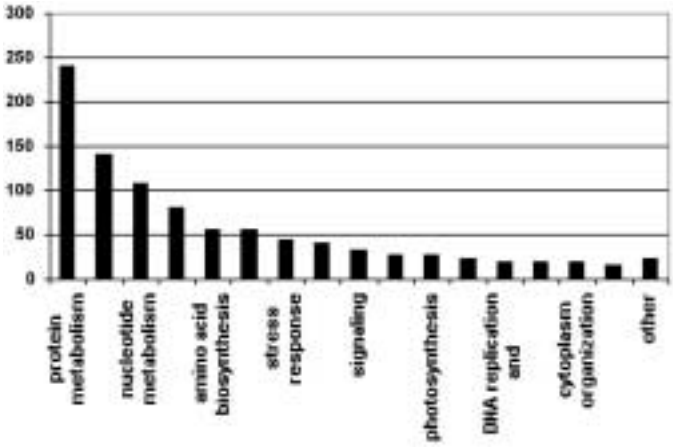


Fig 2: *Functional categories of sugar beet gene products with predicted functions deduced from 973 non-redundant EST sequences. Column height corresponds to the number of members for each category as indicated on the y-axis.*

Genome Analysis of Potato

GABI-CONQUEST Overview

The CONQUEST project (Genes CONtrolling QUantitativE traits of Solanum Tuberosum) is a collaborative project between the Max-Planck-Institute for Plant Breeding Research (MPIZ) and the potato breeding companies Böhm-Nordkartoffel Agrarproduktion (BNA) and Saka-Ragis Pflanzenzucht GbR (SARA).

Project goals are the elucidation of the molecular basis of quantitative trait loci (QTL) for pathogen resistance in potato and the development of diagnostic markers for the targeted resistance traits. Quantitative resistance to late blight, caused by the oomycete Phytophthora infestans, and resistance to the root cyst nematode Globodera pallida are important traits for the potato breeding industry which are difficult and time consuming to evaluate and, therefore, difficult to select based on the phenotype. Knowledge on number, genomic position, coding function, naturally existing molecular and functional variants of genes affecting these characters will make possible the precise genotypic selection of plants with superior alleles early in the breeding process.

A number of resistance QTL and candidate genes have been previously identified on the potato molecular map. Several genes that control resistance to late blight or root cyst nematodes map to a specific genome segment on potato chromosome V (resistance hot spot). This segment is characterized by physical mapping and genomic sequencing, in order to identify candidates for the resistance genes. The genomic sequence is the prerequisite for further functional analysis of the candidate genes.

SNP (Single Nucleotide Polymorphism) markers were identified, which tag, besides the resistance hot spot on chromosome V, other regions of the potato genome known to harbor loci for resistance to various pathogens. The linkage of SNP markers with resistance to late blight or nematodes was tested in segregating tetraploid breeding materials developed and evaluated by the collaborating breeding companies. SNPs closely linked to superior resistance alleles were detected and developed into specific diagnostic markers for marker assisted selection.

Identification and characterization of genes controlling quantitative agronomic characters in potato by a candidate gene approach

CONQUEST (Genes **CON**trolling **QU**antitativ**E** traits of **S**olanum **T**uberosum)

Potato
GABI-CONQUEST

1 Christiane Gebhardt, 2 Hans-Reinhard Hofferbert, Eckart Tacke, 3 Jens Lübeck
1 Max Planck Institute for Plant Breeding Research (MPIZ), Cologne
2 Böhm-Nordkartoffel Agrarproduktion mbH (BNA), Ebstorf
3 SaKa-Ragis Pflanzenzucht GbR (SARA), Windeby

Goals

- To identify at the molecular level candidate genes for controlling quantitative resistance characters which are highly relevant for potato production. The characters targeted are field resistance to Phytophthora infestans causing the late blight disease and resistance to the root cyst nematode Globodera pallida.
- To identify and characterize at the molecular level natural candidate gene alleles affecting the target characters in a positive direction (superior alleles).
- To develop specific diagnostic assays for superior candidate gene alleles to be used in marker assisted selection.
- To isolate superior candidate gene alleles for the purposes of functional analysis, quantitative complementation assays and genetic engineering.

Background

The potato function map for genes controlling qualitative and quantitative resistance to different types of pathogens (Gebhardt and Valkonen 2001) reveals the existence of "hot spots" for resistance in the potato genome. Resistance hot spots are genome segments where genes conferring resistance to different pathogens are clustered. A prominent resistance hot spot is located on potato chromosome V and includes, besides R-genes conferring qualitative resistance to Phytophthora infestans (late blight) and Potato Virus X (PVX), major QTL (quantitative trait loci) for resistance to P. infestans and the root cyst nematode Globodera pallida, the two resistance traits targeted in CONQUEST. A high resolution map of the genome segment has been constructed (Meksem *et al.* 1995, Ballvora *et al.* 2002). A BAC (bacterial artificial chromosome) library of circa 100 000 clones is available for the project.

Based on the potato function map for resistance, we hypothesized that the observed clustering of resistance loci for qualitative and quantitative resistance to different pathogens is the result of gene duplication and divergence of ancestral genes. This suggested that genes controlling qualitative (R genes) and quantitative resistance (QTL) are related. Molecular cloning and sequencing has unraveled the structure of major classes of R genes. Genes with sequence similarity to R genes are considered, therefore, as candidates for controlling quantitative resistance. This was also supported by similar map locations of resistance-gene-like (RGL) loci and resistance QTL (Leister *et al.* 1996). Isolating and sequencing of BAC clones containing genes with sequence similarity to plant R genes that are located in map segments with QTL for resistance to late blight and/or root cyst nematodes provides, therefore, the entry point for the identification at the molecular level of candidate genes for controlling these quantitative resistance traits. Developing SNP (Single Nucleotide Polymorphism) markers from BACs containing RGL sequences provides the molecular tools to identify natural candidate gene alleles affecting the target characters in a positive direction (superior alleles). Generating and evaluating plant populations that segregate for superior resistance alleles provides the genetic material for the identification and molecular characterization of superior resistance alleles. As a minor deviation from the original proposal, we also considered some PR genes and anonymous RFLP markers for SNP marker development in order to tag QTL in the potato genome that are, as known so far, not linked to RGLs.

Results

MPIZ

- Construction of a physical map based on BAC clones around the resistance hot spot on potato chromosome V. Genomic sequencing of BACs in the contig.

Fourteen BAC clones have been ordered in two partial contigs in the 3 cM interval GP21 – GP179 (Figure 1). The BACs cover ca. 400 kb genomic sequence. Six BACs were fully sequenced, corresponding to ca. 300 kb genomic sequence of potato. Thirty six putative genes were identified within the sequenced genomic region, among others RGLs, one of which encodes R1, a major gene for resistance to late blight (Ballvora *et al.* 2002).

- Isolation and partial sequencing of BACs carrying resistance-gene-like (RGL) candidate genes and mapping to regions in the potato genome where QTL for resistance to the target traits have been located.

Using 5 different RGL markers as probes, 230 BAC clones were selected. The BACs were grouped by RFLP fingerprinting. Fourty seven BACs having different fingerprints were mapped to 13 different regions on 10 potato chromosomes. Sequencing 117 ends of BAC insertions resulted in ca. 40 kb genomic sequence. Twenty-two BAC end sequences were similar to known resistance genes.

- Identification of SNP markers at candidate loci based on partial BAC sequences.

PCR products were generated from 17 tetraploid and 11 diploid potato genotypes based on 65 BAC end sequences and, in addition, 21 sequences of pathogenesis-related genes or RFLP markers mapping to candidate regions. Comparative sequence analysis of the PCR products (in total ca. 41 kb genomic sequence) identified 1539 SNPs and 134 InDels (Insertion-Deletion polymorphisms). The sequence information, SNP and InDel markers were deposited in a potato genome database that is under construction at RZPD, Berlin, in collaboration with Dr. Svenja Meyer and Dr. Axel Nagel.

- Genotyping with SNP markers the phenotypic extremes (25 cases and 25 controls per family) of four different F1 families developed and evaluated for the target traits by BNA and SARA. Detection of linkage between SNP markers and quantitative resistance.

Technologies for quantitative genotyping of SNP markers in tetraploid potato were evaluated and compared in collaboration with Dr. Peter Oefner at SGTC, Stanford, USA. The methods of pyrosequencing, primer extension followed by DHPLC and Dideoxy-sequencing were all found to be feasible (Rickert *et al.* 2002). Due to practical reasons, Dideoxy-sequencing was

○ A. Ballvora *et al.* **The R1 gene for potato resistance to late blight (Phytophthora infestans) belongs to the leucine zipper/NBS/LRR class of plant resistance genes.** The Plant Journal (2002) 30, 361-371. ○ C. Gebhardt and J. P. T. Valkonen **Organization of genes controlling disease resistance in the potato genome.** Annu. Rev. Phytopathol. (2001) 39, 79-102. ○ D. Leister *et al.* **A PCR based approach for isolating pathogen resistance genes from potato with potential for wide application in plants.** Nature Genetics (1996) 14, 421-429. ○ K. Meksem *et al.* **A high-resolution map of the vicinity of the R1 locus on chromosome V of potato based on RFLP and AFLP markers.** Mol. Gen. Genet (1995) 249, 74-81. ○ A. M. Rickert *et al.* **Genotyping of SNPs in a polyploid genome by pyrosequencing.** Biotechniques (2002) 32(3), 592-593, 596-598, 600. ○ A.M. Rickert, J.H. Kim, S. Meyer, A. Nagel, A. Ballvora, P.J. Oefner and C. Gebhardt **First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome.** Plant Biotechnology Journal (2003) 1, 399-410. ○ A.M. Rickert **Entwicklung und Analyse von SNP-Markern (single nucleotide polymorphisms) in Pathogenresistenz-vermittelnden Regionen des Kartoffelgenoms (Dissertation, Theses).** Universität zu Köln, 2002. ○ C. Gebhardt, A. Ballvora and A. Rickert **SaKa-Ragis Pflanzenzucht GbR, Windeby, BNA Zuchtgesellschaft mbH, Ebstorf; The German potato genome project GABI-CONQUEST: Results and Perspectives.** Vorträge zu Pflanzenzüchtung (2002) Supplement 1: 109. ○ C. Gebhardt **Rein in die Kartoffeln! Das GABI Projekt CONQUEST.** GenomXPress (2002) 2, 5-6. ○ M.A. Sattarzadeh **SNPs (Single nucleotide polymorphisms) and haplotypes linked to QTLs for resistance to *Globodera pallida* in tetraploid potato (Diplomarbeit, Theses).** Rheinische Friedrich-Wilhelms-Universität zu Bonn, 2003.

chosen for the purpose of SNP genotyping the material provided by BNA and SARA within the ongoing project. Two populations selected for high resistance or high susceptibility to the root cyst nematode G. pallida were genotyped for 262 SNPs and 17 indels tagging 11 different genomic regions. Markers in 3 regions were linked in both populations to resistance QTL at $P < 0.001$.

- Development of allele specific marker assays for application in breeding.

A haplotype specific PCR assay was developed for SNP markers linked to a major QTL for resistance to G. pallida in the population of SARA.

BNA

- Generation, propagation, and phenotypic resistance evaluation of three F1 families, one family for resistance to G. pallida and two families for resistance to late blight.

Three hybrid families were generated (ca. 1200 clones). 367 clones of the family that segregates for nematode resistance were evaluated for resistance. Cases and control plants were selected based on this test and provided to MPIZ for genotyping. The two remaining families were evaluated in the field for late blight resistance and earliness in two consecutive years.

- Selection of 25 highly resistant (cases) and 25 susceptible plants (controls) each from two families, one for nematode resistance and one for late blight resistance.

Completed for the nematode resistance and late blight resistance. Crosses were performed with some clones of the nematode resistant cases.

SARA

- Generation, propagation, and phenotypic resistance evaluation of three F1 families, one family for resistance to G. pallida and two families for resistance to late blight.

Three hybrid families were generated (ca. 1200 clones, ca. 400 clones per family). The family that segregates for nematode resistance was evaluated for resistance. Based on the evaluation, cases and control plants were selected and provided to MPIZ for genotyping. The two remaining families were evaluated in the field for late blight resistance in two consecutive years.

- Selection of 25 highly resistant (cases) and 25 susceptible plants (controls) each from two families, one for nematode resistance and one for late blight resistance.

Completed for the nematode resistance and late blight resistance.

Perspectives

SNP markers have been identified during the ongoing project that are physically tightly linked to candidate genes for quantitative resistance. These markers are a valuable resource for association studies and genetic fingerprinting. Finding association between SNP markers and quantitative resistance traits in breeding materials will increase the value of such markers for marker-assisted resistance breeding. For application on a wide scale, however, more cost-effective and robust technologies for parallel SNP genotyping are required.

For few SNP markers that are linked to superior resistance alleles present in the genetic material used in the ongoing project, specific PCR assays are being developed. These marker assays can be used in a verification phase and, after that, for marker-assisted selection.

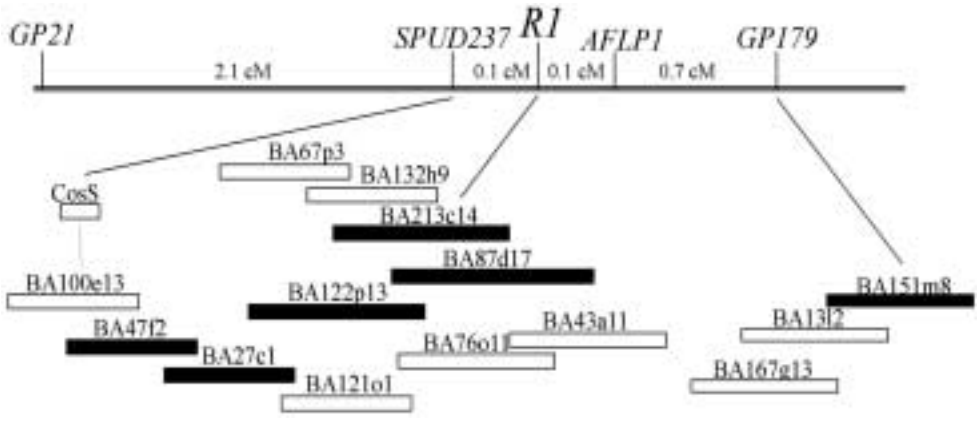


Fig 1: *Current status of the physical map for the resistance hot spot on potato chromosome V in the interval GP21-GP179. BACs in the contig are shown below the genetic map. Solid black bars indicate the BACs that have been sequenced.*

Genome Analysis in Poplar

GABI-POP Overview

The sufficient production of wood without disturbing primary forests will be one of the most important issues for the near future. It was calculated that without improving tree growth or increasing forest area the availability of wood becomes a limiting factor in the next 50 years. In the last 15 years many techniques have been developed such as tissue culture, transformation (gene technology), and structural and functional genomics for various tree species. The basic function of many genes in trees have been unraveled by using the overall plant model species Arabidopsis but few characteristics like longevity, extensive wood production, and long-term adaptation processes can not be studied. Populus has become a model woody species because of its small genome, high propagation capacity, relatively short vegetation periods, and unlike Arabidopsis commercial and ecological significance.

The objectives of the work carried out are the tagging and sequencing of genes and promoters from the model tree species Populus spp. by insertion mutagenesis. A high number of transgenic lines harboring different gene constructs (Fladung *et al.* 1997) were analyzed in respect to T-DNA integration and T-DNA flanking genomic regions. A part of these lines carry the Ac transposon from maize in combination with the phenotypically visible marker genes rolC or iaaL (Fladung *et al.* 1997, Fladung and Ahuja 1997). The excision of Ac during leaf development causes restoration of the marker gene, for example, rolC expression was indicated in the 35S-Ac-rolC transgenic aspen plants by light-green sectors in green leaf background.

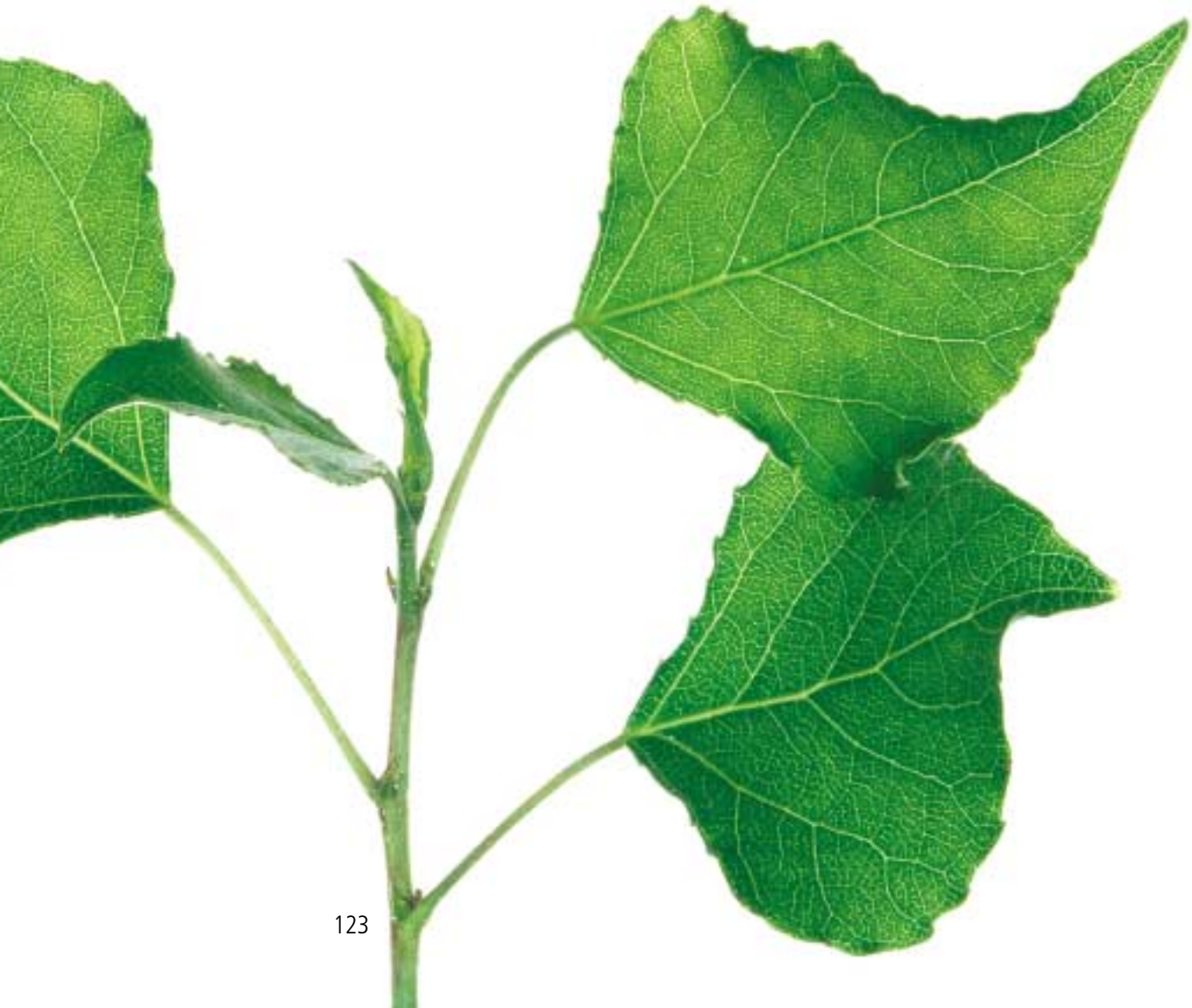
The new positions of Ac in the aspen genome were identified and flanking genomic regions partly sequenced. Interestingly, about one third of the sequences obtained show high similarity to known genes from databanks. This frequency is much higher than the one obtained in the similar study from the genomic sequences flanking the T-DNAs. The technique of transposon tagging is, therefore, suitable for inducing knock-out or knock-in mutants in Populus.

Transposon tagging is a powerful tool for the identification of genes in many plant species including trees but in general its use is limited by low transformation efficiency and low regeneration capacity during tissue culture. Therefore, in trees the practical application of transposons for gene tagging is principally limited due to the long vegetative cycles and, thus, the hemizygous status of the element in the genome. In addition, the species of the genus Populus are dioecious and, therefore, no selves are possible.

In order to overcome these difficulties haploid plants can be used for gene tagging in Populus. In annual plant species haploids play an important role in modern plant breeding. The induction of haploids in woody plants has also been reported (Chen 1987) but, unfortunately, no haploid line was made available upon request.

The activities of the work carried out can be summarized as following:

1. Detailed molecular analysis of variant phenotypes from T-DNA tagged transgenic lines.
2. Sequencing of genomic regions flanking the new Ac-insertion sites.
3. Induction and establishment of haploid P. nigra hybrid lines.
4. Transformation of haploid lines obtained with Ac.



Isolation of tree-specific genes and promoters by a transposon tagging approach

Poplar
GABI-POP

Matthias Fladung, Frank Deutsch, Sandeep Kumar
Institute for Forest Genetics and Forest Tree Breeding (BFH), Großhansdorf

Detailed molecular analysis of variant phenotypes from T-DNA tagged transgenic lines

Transgene flanking genomic regions of in total 30 transgenic lines including the lines showing variant phenotypes were analyzed using Inverse-PCR strategy. In total, 27 right T-DNA/plant junctions, 20 left T-DNA/plant junctions, and ten target insertions from control plants were obtained. The right end of the T-DNA in 18 transgenic lines was conserved up to the cleavage site and the right border repeat was deleted in nine junctions. Nucleotides from the left border repeat were present in 19 transgenic lines out of 20 cases analyzed (Fladung and Kumar 2002, Kumar and Fladung 2001, 2002).

Comparison of the genomic target sites prior to integration to the T-DNA revealed that in three transgenic lines T-DNA inserted into the plant genome without any notable deletion of genomic sequence. However, deletions of a few nucleotides to more than 500 bp were observed in other transgenic lines. Filler DNA of up to 235 bp were observed on left and/or right junctions of six transgenic lines which in most of the cases originated from the nearby host genomic sequence or from T-DNA. Short sequence similarities between recombining strands near break points without any bias to the left or the right T-DNA end were obtained in most of the lines analyzed suggesting that both T-DNA ends may interact during the integration process. Small sequence similarity occurring at the T-DNA 3´ end could act as a primer for DNA fill-in synthesis, whereas at the 5´ end it could adjoin the recessed strand of the partner terminus and be fixed by ligation. The integration model is reminiscent of the model for the repair of genomic DSB in somatic plant cells based on synthesis-dependent strand-annealing (SDSA, Kumar and Fladung 2000).

The "Flanking Sequence Tags" (FSTs) sequences obtained for the T-DNA genomic regions were compared to public databases for similarity to known proteins and genomic sequences (Kumar and Fladung 2002). Only 16% of the T-DNA insertions landed into or near known or predicted genes.

Sequencing of genomic regions flanking the new Ac-insertion sites

Three independent transgenic lines (35S-Ac-rolC #2, #3 and #10) containing a single copy of 35S-Ac-rolC transgene as already demonstrated by Southern analysis and showing light-green sectors on leaves indicating active excision of Ac were selected for the determinations of the FSTs. The light-green sectors were harvested for further molecular analyses (Kumar and Fladung 2003). DNA from several independent light green leaf sectors was used in Inverse-PCR or TAIL-PCR reactions to amplify aspen genomic DNA flanking the Ac insertions. The sequences obtained were compared to public databases for similarity to known proteins and genomic sequences. The sequences analyzed in three independent transgenic lines and the BLASTx search results are summarized in Table 1. In total, 75 FSTs have been submitted to the GABI database so far.

The data show that out of the 75 sequences analyzed 22 sequences gave significant BLASTx hits. This might be a minimal estimate keeping in view the small amount of informations from the aspen genome available in the public databases. However, we expect that more DNA sequences disrupted by these insertions become recognizable as putative genes as the poplar genome is fully sequenced in the end of year 2003.

Such a catalogue of Ac flanking DNA sequences could be used to make a FST database, suitable for aspen reverse genetics. The total frequency of FSTs in predicted genes (Table 1) was calculated ranging from 25-33% (mean 29%) on the basis of homology to the public databases. Details of the Ac-tagged hits are given in Kumar and Fladung 2003. FSTs in predicted genes like DNA methyltransferase can provide informations on the unique function of these genes in a long-lived tree system. The preferential insertion of the Ac element in the coding regions of the aspen genome as indicated by our results suggests that transposon mutagenesis will be a very important tool to discover gene functions by reverse genetics or forward genetics strategies.

The frequency of hitting coding sequences in the same aspen system is for Ac insertion double as that for T-DNA integration. Hence, the Ac system seems to be more effective for gene tagging and promoter trapping studies in aspen.

The T-DNA/Ac FSTs obtained from aspen genome were searched for T-DNA tagged Arabidopsis lines in GABI-KAT database (http://www.mpiz-koeln.mpg.de/GABI-Kat/). We obtained five Arabidopsis lines where T-DNA appears to be inserted into coding regions that are similar to aspen T-DNA/Ac FSTs (Table 2). Seeds were ordered to study Arabidopsis lines with a T-DNA in similar genes.

Induction and establishment of haploid P. nigra hybrid lines

Microspore culture, callus formation, and regeneration of plants was established for P. nigra hybrid lines (Deutsch *et al.* submitted). From the 113 calli obtained in 2001 one third turned green in light and some embryos formed green cotyledons and hairy roots. Out of these only one line (M22-1c5 regenerative callus) was found to be haploid in flow cytometer measurements analysis. Regenerated plantlets of this line are dwarfish and slow growing, and unfortunately, double-haploid. However, the regenerative callus line M22-1c5 is still haploid in continuous in vitro culture for 21 months up to date.

In 2002, one third of the 1375 calli obtained turned green and one sixth initiated organogenesis. Most of the transparent calli turned brownish but

○ Z. Chen **Induction of androgenesis in hardwood trees.** In: Bonga JM, Durzan DJ (eds) Cell and tissue culture in forest trees. Martinus Nijhoff Publishers, Dordrecht, (1987) vol 2, pp. 247-268. ○ M. Fladung and M. R. Ahuja **Excision of the maize transposable element Ac in periclinal chimeric leaves of 35S-Ac-rolC transgenic aspen-Populus.** Plant Mol. Biol. (1997) 33, 1097-1103. ○ M. Fladung, S. Kumar and M. R. Ahuja **Genetic transformation of Populus genotypes with different chimeric gene constructs: Transformation efficiency and molecular analysis.** Trans. Res.(1997) 6, 111-121. ○ F. Deutsch *et al.*: **Regeneration of stable haploid poplar callus lines from immature pollen culture.** Paper submitted. ○ M. Fladung and S. Kumar **Gene stability in transgenic aspen-Populus – III. T-DNA repeats influence transgene expression differentially among different transgenic lines.** Plant Biology (2002) 51, 329-338. ○ S. Kumar and M. Fladung **Transgene repeats in aspen: molecular characterisation suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome.** Molecular Genetics Genomics (2000) 264, 20-28. ○ S. Kumar and M. Fladung **Gene stability in transgenic aspen (Populus). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen.** Planta (2001) 213, 731-740. ○ S. Kumar and M. Fladung **Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration.** The Plant Journal (2002) 31, 543-551. ○ S. Kumar and M. Fladung **Somatic mobility of the maize element Ac and its usability for gene tagging in aspen.** Plant Molecular Biology (2003), in press.

some formed green cores of solid regenerative cells and shoots. Flow cytometer measurements of 66 selected regenerative calli revealed that five calli were haploid, 58 diploid (including two calli showing a second peak at n and two calli at 4n), two tetraploid, and one callus was aneuploid. All regenerative calli analyzed in flow cytometer measurements were investigated for their haploid origin using five microsatellites- (SSR-) markers. The microsatellites were heterozygous in the "father" trees 'Aue 1' and 'Aue 2'. All haploid and most diploid and tetraploid calli were homozygous for either alleles.

One haploid regenerative callus line (FD25) regenerated haploid plants. The line FD25 is well growing in in-vitro culture despite its haploid status. At present measurements are repeated to confirm long-term haploidy in the regenerative callus line, and regeneration experiments from leaves of regenerated haploid plants are underway. The haploid plantlets are currently multiplied for transformation experiments.

Transformation of haploid lines obtained with Ac

Transformation experiments using callus and leaf tissue from the haploid clone M22-1c5 with the 35S-Ac-rolC construct were performed. Two stable transgenic haploid Populus lines carrying Ac were obtained that are being tested for Ac-excision. One Ac-transgenic regenerative callus line is haploid, the second is double haploid. Regeneration of Ac-transgenic plantlets is in progress. The haploid M22-1c5#23-1 transgenic line was investigated for Ac transposition. Ac insertion in one case has been found in the npt-II gene of the T-DNA. Southern analysis are underway to confirm this result. Following transformation of the two additional haploid regenerative callus lines FD25 and FD4 so far twelve and four transgenic calli, respectively, putatively transgenic for the 35S-Ac-rolC gene construct were obtained. From three independent transgenic lines, transgenic haploid plantlets are being rooted.

Transgenic line	FSTs analyzed	Significant BLASTx hits
Esch5:35S-Ac-rolC#3	41	12 (29%)
Esch5:35S-Ac-rolC#10	24	6 (25%)
Esch5:35S-Ac-rolC#2	10	4 (40%)
Total	**75**	**22 (29%)**

Tab 1: *Frequency of Ac insertions into predicted genes (Kumar and Fladung 2003).*

Aspen sequence (GABI-PD code)	Transgenic line	T-DNA/Ac	Gene annotation	GABI-KAT Gene code	GK line ID
B2_2_RE_B	Brauna11:35S-rolC#2	T-DNA	Putative methyltransferase	At4g10760	353B10
E2_5RE_B	Esch5:35S-rolC#5	T-DNA	Receptor-like kinase	At5g38280	254G07
SK316CUP	Esch5:rbcS-rolC#11	T-DNA	potassium channel	At4g22200	152C05
SK467	Esch5:35S-Ac-rolC#2	Ac	putative phosphoribosylanthranilate transferase	At4g11610	004G08 147G08 234C05
SK510	Esch5:35S-Ac-rolC#3	Ac	Hypothetical protein	At4g03540	301G06

Tab 2: *GABI-KAT search results of T-DNA/Ac flanking genomic sequences obtained from aspen.*

Cereals

Cereals Overview

On a worldwide level more than 70% of the human diet stem directly or indirectly from cereals. While rice and maize dominate the production in Asia and North America, respectively, wheat and barley are the major cereal species grown in Europe.

Barley (*Hordeum vulgare*) represents both a model for genome analysis and an agriculturally important plant species. This is because the seven barley chromosomes represent the basic genome complement of all species within the *Triticeae* tribe, which includes with wheat, oats, Triticale and rye additional species of major agricultural importance.

The usefulness of the barley genome as a model on the one hand emerges from the idea to apply the resources available for this species to other cereals and, on the other hand, to use the barley genome as a platform for the transfer of genetic information from rice.

Despite being composed of 12 chromosomes, the rice genome is about 12 times smaller than the barley genome and 36 times smaller than the wheat genome.

Not least because of its small size, the final sequence of rice is expected to be available by the end of 2004.

To what extent the genomes of barley and rice have diverged during the past 60 million years of common evolution is studied in the **GABI Map** consortium by comparative sequencing of selected regions in both species. The outcome of this project provides not only insight into evolution of the two genomes but also information on the usefulness of the barley genome as a platform for the development of strategies and approaches to exploit the wealth of information now available from the extensively analyzed rice genome.

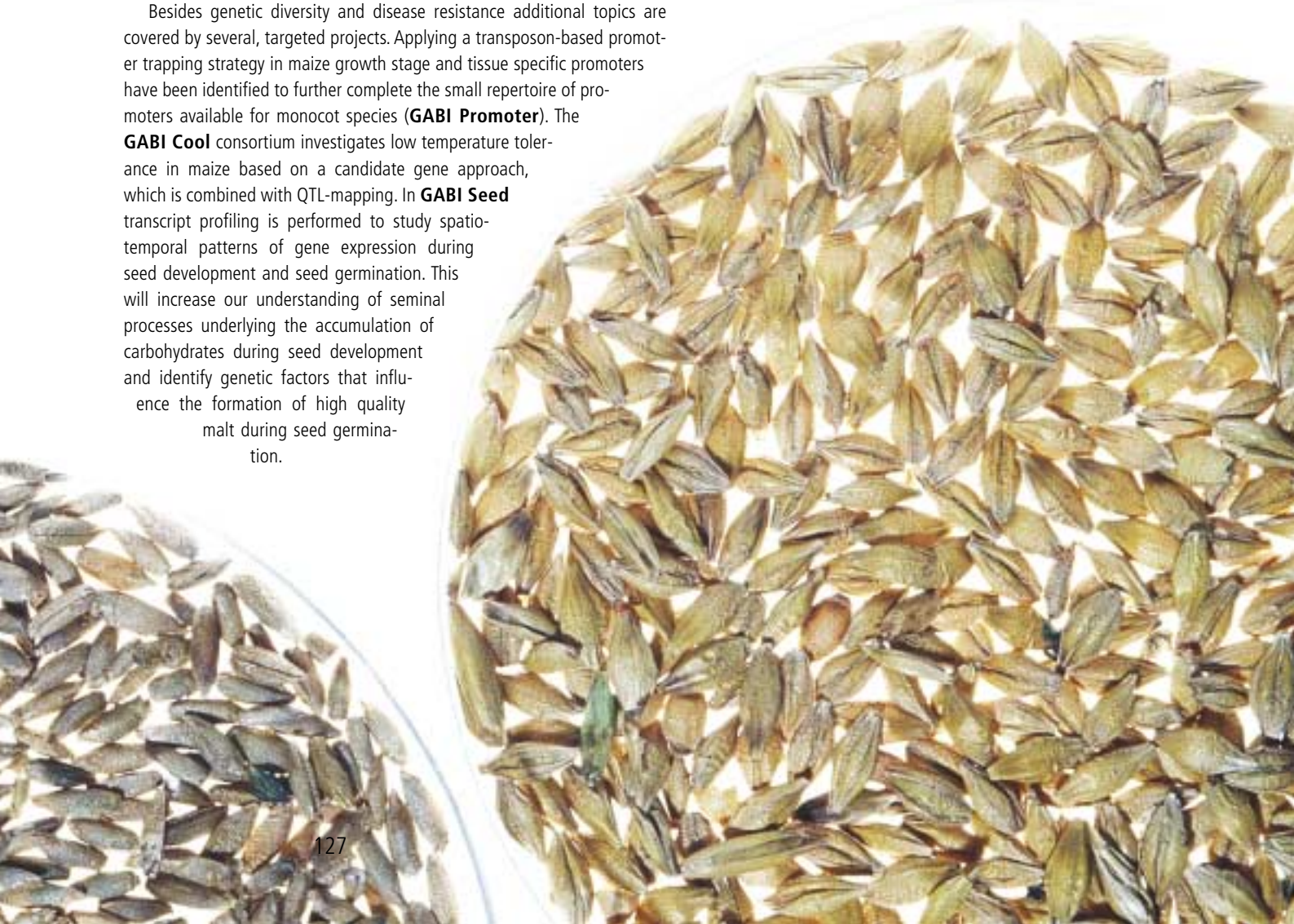
The majority of GABI-research into cereals focuses on two major areas that are of specific relevance to plant breeding: analysis and exploitation of genetic diversity from wild species, and disease resistance. The exploitation of the genetic diversity present in the wild relatives represents an enormous, yet untapped, potential for genetic improvement of cultivated plants. One approach to make better use of genetic diversity rests on the marker-assisted introgression of defined chromosome fragments from a wild relative into a cultivar as it is performed in barley (**GABI Barley Diversity**), wheat (**GABI-AB-QTL**) and rye (**GABI Rye**). The corresponding introgression lines are ideally characterized by carrying one defined chromosome segment from a wild species in the genetic background of the cultivar. Their generation critically depends on the availability of markers suitable for high throughput screening. Therefore, marker development has also been included in several projects.

As with other plant species disease resistance represents a major breeding goal in cereals. There is evidence that a large number of resistance genes against viral, fungal and animal pathogens belong to the NBS/LRR class of genes, which is characterized by distinct DNA signatures that are conserved across species.

The project **GABI Rye** aims at exploiting the sequence information of such known resistance genes to identify genetic markers and candidate genes for various kinds of fungal and viral disease resistance.

The cultivation of resistant cultivars increases yield stability and reduces the need for agrochemicals. However, infection by the pathogenic fungus *Fusarium graminearum* not only reduces yield, but dramatically increases health risks due to the production of harmful toxins, which persist in the grain. The identification of differentially expressed genes involved in the plant response to this pathogen may facilitate the development of efficient strategies to breed for resistance (**GABI Agrotech**). In contrast to non-obligate biotrophic fungi like *Fusarium*, which are able to infect a broad host range, obligate biotrophs such as powdery mildew (*Blumeria graminis*) usually are restricted to a narrow host range, in most cases comprising only one species. Here, susceptibility of a plant species to a given pathogen is the exception, while (non-host) resistance is the rule. The **GABI Nonhost** consortium aims at unraveling the molecular basis of non-host resistance applying a functional genomics approach in both barley and Arabidopsis.

Besides genetic diversity and disease resistance additional topics are covered by several, targeted projects. Applying a transposon-based promoter trapping strategy in maize growth stage and tissue specific promoters have been identified to further complete the small repertoire of promoters available for monocot species (**GABI Promoter**). The **GABI Cool** consortium investigates low temperature tolerance in maize based on a candidate gene approach, which is combined with QTL-mapping. In **GABI Seed** transcript profiling is performed to study spatio-temporal patterns of gene expression during seed development and seed germination. This will increase our understanding of seminal processes underlying the accumulation of carbohydrates during seed development and identify genetic factors that influence the formation of high quality malt during seed germination.



Development of new DNA marker systems and utilization of genetic resources for Barley

Barley Diversity Overview

Participating subprojects:

I. Development of single nucleotide polymorphism (SNP) markers for barley. (A. Graner, IPK-Gatersleben, F. Ordon, University of Giessen, J. Weyen Saaten-Union Resistenzlabor GmbH, Hovedissen)

II. Assessment of genetic diversity in barley germplasm and varieties using microsatellite markers. (M. Röder, IPK-Gatersleben)

III. AB-QTL analysis in barley: Detection of favorable genes for quality and yield traits from barley wild species and their introgression into the elite germplasm of cultivated barley by means of DNA markers. (K. Pillen and J. Léon, University of Bonn)

The three subprojects of the BARLEY Diversity co-operation are focused on the generation of new DNA markers in barley and on the assessment and use of the genetic diversity present in the genus Hordeum.

PCR-based markers like microsatellites are very useful in regard to the assessment of the genetic diversity. However, a further reduction of the diagnostic costs would be beneficial in order to utilize DNA markers on a large scale in plant breeding where populations of several thousand individuals have to be screened routinely. High-throughput marker technologies might be the key for cost reduction. Very promising candidates in this respect are single nucleotide polymorphisms (SNPs). SNPs identify base substitutions between short, PCR-amplified DNA fragments from different plants. Because of their unparalleled abundance they can be used in whole genome scans to identify genes regulating simple and complex traits. It is expected that the genotyping of plants with SNPs will reduce the current costs per genetic data point dramatically. Thus, the aim of subproject I was to establish SNPs as a new class of DNA markers in barley. It is planned to apply the resulting barley SNPs to all barley co-operative projects where DNA polymorphisms are a pre-requisite for genetic mapping or molecular breeding.

Since the beginning of the century, much effort has been made towards the collecting and preservation of unadapted germplasm. So far, millions of accessions of wild and cultivated plant species have been deposited in seed banks worldwide. However, only a limited number of these accessions are used as a genetic resource for crop improvement. The situation might change considerably in future due to the application of DNA marker technologies. For instance, the characterization of accessions by means of DNA fingerprinting could foster the selection of exotic germplasm as donors for the improvement of elite varieties. We implemented subproject II in order to characterize the genetic diversity which is present among barley varieties and wild species. Here, highly polymorphic microsatellites are used in order to assess approximately 1,000 barley accessions. Later, the marker-characterized germplasm will be used as new genetic donors for the improvement of monogenic and polygenic traits in plant breeding and in basic research projects.

After molecular markers became available for breaking up quantitative agronomic traits into its underlying polygenes by means of the quantitative trait locus (QTL) analysis, it also became feasible to identify favorable exotic QTL alleles from wild species which can improve agronomic traits. We currently carry out subproject III to achieve this goal in barley. Here, we apply the advanced backcross QTL strategy in order to generate new, improved breeding material that contains previously inaccessible favorable exotic QTL alleles in a relatively short period of time. In addition, the subproject will result in the production of nearly isogenic lines (NILs) containing single favorable QTL segments from the exotic donor in an otherwise complete elite background. The NILs will be the ideal basis for both breeding of new barley varieties as well as for basic research, for instance, on map-based cloning of the localized exotic QTL genes.

AB-QTL analysis in barley: Detection of favorable genes for quality and yield traits from barley wild species and their introgression into the elite germplasm of cultivated barley by means of DNA markers

Barley
Barley Diversity

Klaus Pillen, Maria v. Korff Schmising, Huajun Wang, Jens Léon,
University of Bonn, Department of Crop Science and Plant Breeding, Bonn

Introduction

In this study, the AB-QTL (Advanced Backcross Quantitative Trait Locus) strategy developed by Tanksley & Nelson (1996) is used to detect genes affecting quantitative traits of agronomic importance from interspecific crosses between cultivated barley (*Hordeum vulgare ssp. vulgare*) and wild barley (H. vulgare ssp. spontaneum). The particular aim of the project is to localize wild barley alleles which significantly improve a quantitative trait, as shown by Pillen *et al.* (2003), Pillen *et al.* (2004) and Piepho & Pillen (2004). The traits under investigation are parameters of agronomic performance, nitrogen efficiency, malting quality and disease resistance. At the end of the project, favorable wild barley QTL alleles will be transferred into the elite barley gene pool by means of the production of QTL-NILs (near isogenic lines). This project is similar in regard to, both, the strategy and the goals of wheat project 0312862 which is also currently conducted in our laboratory.

Materials and methods

Population development

Four advanced backcross (AB) populations, in total including 800 BC2DH lines, were developed from crosses of two spring and two winter elite varieties with two wild barley accessions (fig 1). For this, the spring barley cultivars Scarlett and Thuringia were crossed with the wild barley accession ISR42-8 (S42 and T42, respectively) and the winter barley varieties Carola and Theresa were crossed with the wild barley accession ISR101-23 (C101 and T101, respectively). The BC2 individuals were used for doubled haploid (DH) production through anther culture.

SSR marker analysis

A total of 220 SSR markers were screened to reveal DNA polymorphisms between the parents of the four crosses. Approximately 100 polymorphic SSR markers were selected for consecutive genotyping. These SSRs yielded consistently good PCR products, produced minimal stutter bands and were evenly distributed across the chromosomes. The SSR markers were assigned to the chromosomes and their chromosomal position was estimated with the aid of the two barley consensus maps of Kleinhofs *et al.* (1993) and Ramsey *et al.* (2000). However, 40 SSR markers are not yet mapped in any reference population. These markers were assigned to chromosomes by means of a set of wheat barley addition lines. In order to genetically locate these SSR markers, we now map these SSR markers in the F2 population Thuringia*42-8 and in the Oregon-Wolfe-Barley (OWB) reference population (von Korff *et al.* in press).

Field evaluations of quantitative traits

The BC2DH lines will be field investigated during the years 2003 and 2004 at four locations in Germany. These are the experimental station Dikopshof (University of Bonn) and the breeding stations Gudow (Nordsaat, Schleswig-Holstein) and Irlbach (Ackermann, Bavaria) plus Leutewitz (DSV, Saxony) only for winter barley and Morgenrot (Breun, Saxony) only for spring barley. Each BC2DH line will be tested under three treatments per location. A standard agronomic treatment will be used for evaluating yield and quality related parameters. A low nitrogen fertilization regime will be used again for evaluation of the agronomic performance. We expect to localize QTL alleles for high nitrogen efficiency by comparing the standard and the low nitrogen fertilization levels. These QTLs can later be used for breeding low input barley varieties which require less nitrogen.

The analysis of agronomic performance will include the quantitative traits plant height, time of flowering, lodging after flowering, lodging before harvest, bending of ears, bending of stems, thousand grain weight and total yield. The following measurements will only be carried out at the experimental station Dikopshof: spikes/sqm, kernels per spike and harvest index. Malting quality parameters will be examined only from grain samples taken from the Dikopshof. The malting analyses will include the parameters total and soluble protein content, germination rate, water absorption, fermentability, a and b amylase activity, b glucan content, diastatic power and malt extract.

In a third block, we will evaluate the level of disease resistance present in the BC2DH lines. For this, the AB lines are cultivated without any fungicide treatment. We will record the natural occurrence of leaf and spike related diseases, such as BaYMV (Barley yellow mosaic virus), powdery mildew (Blumeria graminis), net blotch (Drechslera teres), leaf scald (Rhynchosporium secalis), brown rust (Puccinia hordei), Fusarium head blight (Fusarium ssp.) as well as the occurrence of non-parasitic browning.

Results and discussion

Population development

The number of DH lines and corresponding BC1 families varied considerably between the four barley populations. It was mainly determined by the survival rate during the tissue culture. In the winter barley populations the two-row barley lines were discarded in order to obtain a homogeneous population of six-row DH lines. No further selection was exercised during the generation and propagation of the BC2DH lines. The Scarlett and Thuringia BC2DH populations count 310 and 84 DH lines and were derived from 12 BC1 parents each. The generation of 284 Carola and 114 Theresa BC2DH lines can be traced back to 8 and 7 BC1 parents, respectively.

○ A. Kleinhofs *et al.* **A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome.** Theor. Appl. Genet. (1993) 86, 705-712. ○ L. Ramsay *et al.* **A simple sequence repeat-based linkage map of barley.** Genetics (2000) 156, 1997-2005. ○ S. D. Tanksley & J. C. Nelson. **Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines.** Theor. Appl. Genet. (1996) 92, 191-203. ○ R. van Berloo **GGT: Software for the display of graphical genotypes.** The Journal of Heredity (1999) 90, 328-329. ○ K. Pillen, A. Zacharias and J. Léon **Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.).** Theor Appl Genet (2003) 107, 340-352. ○ K. Pillen, A. Zacharias and J. Léon **Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare ssp. spontaneum*.** Theor Appl Genet (2004) 108, 1591-1601. ○ H.P. Piepho and K. Pillen **Mixed modelling for QTL x environment interaction analysis.** Euphytica (2004) 137, 147–153. ○ M. von Korff, H. Wang, J. Léon and K. Pillen **Development of candidate introgression lines using an exotic barley accession (*H. vulgare ssp. spontaneum*) as donor.** Theor Appl Genet (2004) (DOI: 10.1007/s00122-004-1818-2). ○ M. von Korff, J. Plümpe, W. Michalek, J. Léon and K. Pillen **Insertion of 18 new SSR markers into the Oregon Wolfe Barley map.** Barley Genetics Newsletter (2004) (in press).

Genotyping

The polymorphism survey revealed a polymorphism rate between 80 and 84% in the respective crosses. So far, the spring barley populations S42 and T42 have been genotyped with 87 and 69 SSR markers. In addition, genotype data from 72 and 49 SSR markers have been developed for the winter barley populations C101 and T101. In the two winter barley populations 72 and 49 genotyped SSR markers span 969 and 836 cM, respectively, with an average marker density of 15 and 17 cM. In the spring barley population 87 and 69 genotyped SSR markers cover 1042 and 836 cM, respectively, with an average marker density of 13 and 12 cM.

The genetic constitution of the BC2DH plants was visualized using the GGT software (van Berloo 1999). The graphical genotypes generated by the software provide an estimation of the wild barley genome portion and visualize the positions of chromosomal donor segments present in each BC2DH plant. In Fig 2 the distribution of the donor segments across all seven barley chromosomes is shown for 254 BC2DH lines of the cross C101. The GGT software can also be used to assist in the selection of the BC2DH plants for the production of NILs. In this case, DH lines are selected based on (1) the presence of the target QTL donor interval and (2) a low remaining portion of the donor genome, especially on the chromosome carrying the target donor segment. These QTL-NILs provide a valuable resource for unraveling gene functions, both, in regard to transcription profiling and to map-based cloning of a QTL. Two sets of NIL candidates from the spring barley populations S42 and T42 have already been selected (von Korff *et al.* 2004).

Calculated from the genotyped SSR markers, the mean percentage of wild alleles present in the DH lines was estimated with 14% in S42, 12% in T42 and 12% in the two winter barley populations C101 and T101. These values are very close to 12.5% which is the theoretical expectation of the donor genome present in a BC2DH population.

Future work

In order to fill in some remaining gaps in the barley map, selected SSR and SNP markers from Prof. Graner, Gatersleben, will be genotyped. Ultimately, the genotype data and the phenotype data will be used in order to locate favorable QTL alleles of wild barley which improve quantitative agronomic traits.

Co-operating breeding companies

Deutsche Saatveredelung Lippstadt-Bremen GmbH, Lippstadt
J. Ackermann Saatzucht, Irlbach
Nordsaat Saatzucht, Böhnshausen
Saatzucht Josef Breun GdbR, Herzogenaurach

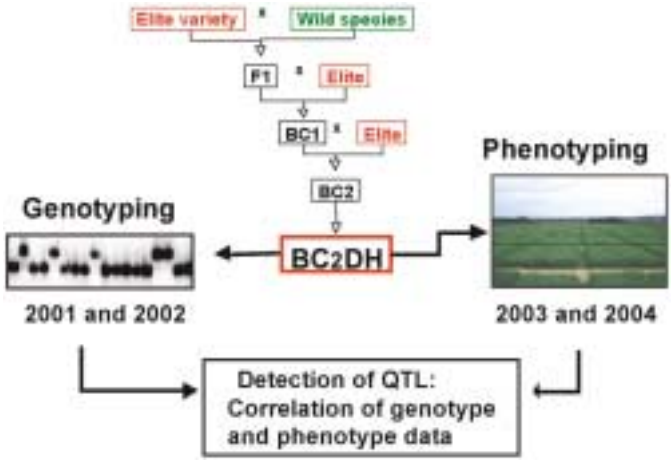


Fig 1: *The Strategy of the AB-QTL analysis in barley. Elite barley varieties are crossed with wild barley accessions. After two rounds of backcrossing, doubled haploids (BC2DH) are produced. The wild barley alleles are then used as donors in order to detect QTLs and to improve the elite barley varieties through the introgression of favorable, trait improving wild QTL alleles.*

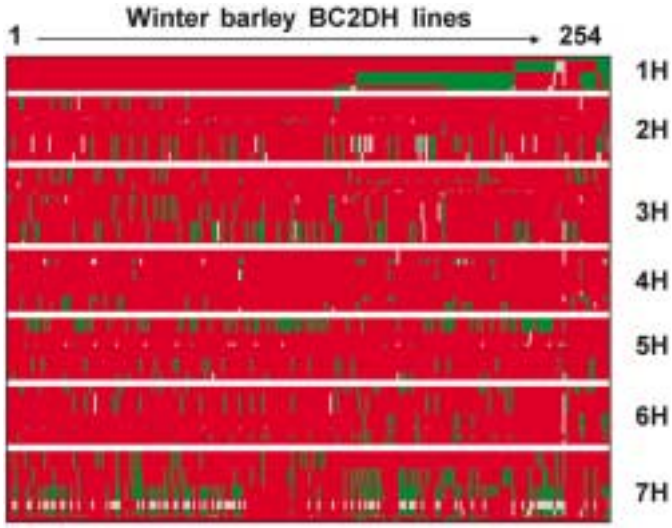


Fig 2: *Distribution of wild barley donor segments (green) for 254 winter barley BC2DH lines of the cross C101 by means of graphical genotypes. The recipient genome is labeled in red. Grey segments indicate missing genotype data. The BC2DH lines are sorted for presence of the donor segments on the first barley chromosome (1H).*

Development of single nucleotide polymorphism (SNP) markers for Barley *(Hordeum vulgare* L.*)* (Subproject A)

1 Frank Ordon*, 1 Gisela Neuhaus, 2 Jörg Schondelmaier, 2 Jens Weyen
1 Justus-Liebig-University, Institute of Crop Science and Plant Breeding I, Giessen
2 Saaten-Union Resistenzlabor GmbH, Leopoldshöhe
* present address: Institute of Epidemiology and Resistance, Federal Centre for Breeding Research on Cultivated Plants, Aschersleben

Introduction

The subproject "Development of trait specific SNPs" focuses on the identification of SNPs in STSs linked to resistance genes against diseases of agronomical importance, i.e. barley yellow mosaic virus disease (BaMMV, BaYMV, BaYMV-2), barley yellow dwarf virus (BYDV) and *Rhynchosporium secalis*. Based on these SNPs a set of specific markers suitable for marker assisted selection procedures and genotyping is developed. Because of the economical importance of these genes in barley breeding much emphasis is given to detect as many SNPs in closely linked fragments as possible. For this purpose about 300 barley gemplasms covering a wide range of genetic diversity including *H. vulgare* ssp. *spontaneum* are analysed for respective SNPs. At the beginning of the project different types of markers (RAPDs, AFLPs) had been known for resistance genes *rym4*, *rym5*, *rym9*, *rym11*, *Rh* and *Ryd2* in barley already, but only very few had been converted into STSs suitable for SNP-screening. Besides this, mere DNA-sequencing for SNP-detection has been rather expensive in this time, but several gel-based systems like TGGE, SSCP-analysis and heteroduplex analysis had been known for SNP-screening and new techniques like Bess-T-Sequencing™ had just been developed. In order to identify SNPs in STSs linked to resistance genes, the work was briefly divided into the following sub-tasks: (i) development of additional STSs out of RAPDs and AFLPs, (ii) establishment of a SNP-detection system, (iii) SNP-screening of trait specific STSs, (iv) sequencing of trait specific SNP carrying fragments, (v) development of allele specific primers and marker systems for genotyping.

Results

Due to the fact that flanking STSs were not available for all loci at the beginning of the project, closely linked RAPDs and AFLPs have been converted into STSs. In this respect one RAPD located 0.8 cM distal of *rym5* - which is of special importance in barley breeding – and an AFLP located 0.05 cM proximal of this locus (Pellio *et al.* 2004) have been converted into STSs. Besides this, additional STSs suitable for SNP-screening have been developed out of RAPDs and AFLPs linked to genes under investigation.

SNP-screening – in a first step based on SSCP-analysis on MDE-Gels (Mutation Detection Enhancement) followed by silver staining - was conducted on a sub-sample of germplasms selected due to prior results on genetic diversity out of a collection of about 300 barley accessions comprising germplasms derived from different geographic origin, released German winter barley cultivars, and *H. vulgare* ssp. *spontaneum* accessions from Turkey and Israel. By this approach between one and four alleles were detected per STS. However, when different alleles were sequenced, in most cases no SNPs were detected leading to the conclusion that SSCP-analysis is no reliable tool in our hands. Therefore, in a next step fluorescence labelled Bess-T-Sequencing™ (Base Excision Sequence Scanning) on an automated DNA-sequencer (LiCOR 4200, Fig1) was established. Bess-T-Sequencing detects all SNPs involving thymidine (T) resulting in a detecting rate of 83% provided both strands are analysed. Employing Bess-T-Sequencing on STS AF18H533 closely linked to *rym5* leads to the detection of 1 SNP/533 bp. Out of 285 genotypes analysed a T was detected in 76 accessions corresponding to about 27%. In STS Y57c10 which is also linked to *rym5* SNPs were detected in a much higher frequency, i.e. 1SNP/69 bp. Figure 1 shows the result of a Bess-T-Sequencing on different genotypes using the fluorescence labelled forward primer of Y57c10 (Neuhaus *et al.* 2003). Polymorphic fragments were detected at about 220bp, 300bp, and 350bp (Fig 1) and additional SNPs have been detected by analysing the reverse strand. It turned out that respective SNPs are combined to eight different haplotypes which in a next step have been sequenced (Fig 1). As can be seen in Figure 1 all the SNPs detected by Bess-T-Sequencing were verified by sequencing respective genotypes and all of them involve T and C. Out of these SNPs, four of them involve a restriction site and those affecting a 4 or 6 bp restriction enzyme (Table 1) were used for genotyping. Out of 209 genotypes analysed up to now 188 (89.9%) were digested by *HaeII*, 66 (31.6%) by *EcoRV* and 63 (30.1%) by *Age*I. By this procedure eight different haplotypes were identified with frequencies between 65% and 0.5% (Tab 1, Ordon *et al.* 2003). In case genotyping was not facilitated by restriction digestion, allele specific primers were designed and used for genotyping. Out of 313 genotypes tested with a primer pair specific for a G/C exchange at 428 bp in Y57c10, 251 (80%) showed amplification indicative for the presence of C. Adding these results to the CAPs analysis of Y57c10 (Tab 1) 13 different haplotypes can be distinguished.

Within STS-G06H532 – developed out of the RAPD-marker OP-G06 (Bauer *et al.* 1997) and closely linked to *rym11* – no unequivocal results were obtained by Bess-T-Sequencing. The same holds true for Ylp closely linked to *Ryd2* (Ford *et al.* 1998). Therefore, a subset of 10 genotypes representing a broad spectrum of genetic diversity and respective donors of resistance were directly sequenced. In case of Ylp no additional SNPs were detected except the A to C exchange affecting the restriction site of *HSP92*II (Ford *et al.* 1998). Out of 327 genotypes analysed Ylp was digested by *HSP92*II in 319 cases (97.5%) indicative for A and the susceptibility encoding allele and only 8 genotypes, i.e. 2.5%, were not digested indicative for the presence of *Ryd2*. Within STS-G06H532 2 SNPs were detected by sequencing resulting in a SNP-frequency of 1 SNP/266 bp. Both SNPs do not affect restriction sites. Therefore, as the development of allele specific primers is often difficult and in case of a dominant marker a failure of PCR leading to a misinterpretation of results has to be taken into account, additionally the READIT®-Genotyping-System (Promega, Madison) was tested on a subset of genotypes. As can be seen in Table 2 this system which is in principle based on the SNP specific detection of the amount of light produced in a luciferin/luciferase reaction facilitates an unequivocal identification of respective SNPs.

Conclusion and future prospects

By the approach described above, i.e. development of STS closely linked to resistance genes followed by SNP-screening on a set of genotypes of known

○ G. Neuhaus *et al.* **First results on SNP-scanning in fragments linked to resistance genes against the barley mosaic virus complex.** J. Plant Diseases and Protection (2003) 110, 296-303. ○ F. Ordon *et al.* **Marker based strategies for resistance breeding to viruses (BaMMV, BaYMV, BaYMV-2) in barley (*H. vulgare L.*).** Proc. EUCARPIA Cereal Section Meeting "From bodiversity to genomics", 21-25 November, 2002, Salsomaggiore, Italy, (2003) 243-247. ○ B. Pellio *et al.* **Development of PCR-based markers closely linked to *rym5*.** J. Plant Diseases and Protection (2004)111, 30-38. ○ E. Bauer *et al.* **Molecular mapping of novel resistance genes against barley mild mosaic virus (BaMMV).** Theor. Appl. Genet. (1997) 95, 1263-1269. ○ C. M. Ford *et al.* **Rapid and informative assays for Yd2, the barley yellow dwarf virus resistance gene, based on the nucleotide sequence of a closely linked gene.** Molecular Breeding (1998) 4, 23-31.

genetic diversity and SNP-scoring on a large set of genotypes, SNP frequency was estimated in the range of 1SNP/69 bp and 1SNP/533bp and the frequency of the more abundant allele varied between 68.4 and 97.5%. The analyses of additional STSs is in progress. By this approach detailed information on SNP and haplotype frequencies in STSs linked to resistance genes is gained on a large set of genotypes facilitating the estimation of genetic relatedness by cluster and principal coordinate analysis (PCoA) and the use of respective SNPs in marker based selection procedures by conventional methods (e.g. CAPs, allele specific primers) or by effective high throughput systems in the future. Besides this, by screening these genotypes with a set of genome covering SNPs more detailed information on SNP frequencies and haplotypes within the barley gene pool may be gained in the future facilitating – together with phenotypic data – linkage disequilibrium mapping.

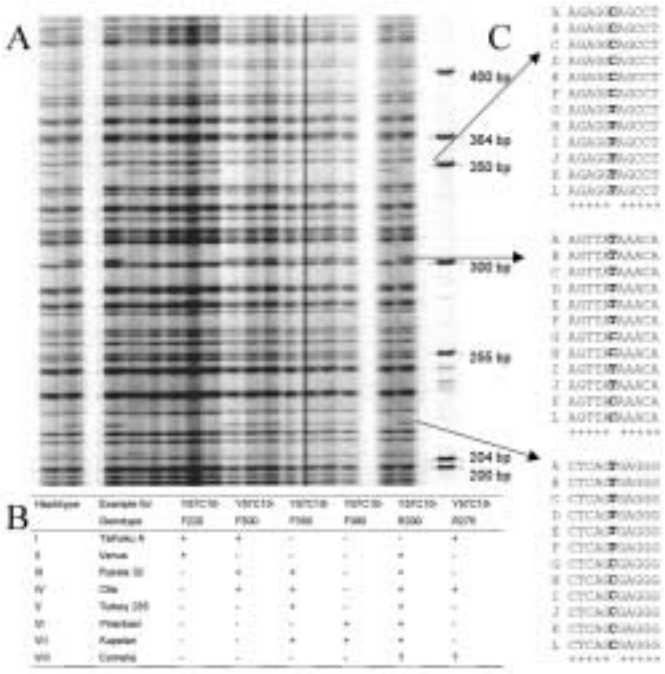


Fig 1: *Results of Bess-T-sequencing of Y57c10 on a subsample of genotypes revealing SNPs at 220bp, 300bp and 350bp (A), haplotypes detected in Y57c10 (B), and results of sequencing of haplotypes for SNPs detected at 220bp, 300bp and 350 bp (C).*

Haplotype	*Hae*II	*Eco*RV	*Age*I	Sum	%
I	+	+	+	40	19.1
II	+	+	-	7	3.4
III	-	+	+	17	8.1
IV	+	-	+	5	2.4
V	+	-	-	136	65.1
VI	-	+	-	2	1.0
VII	-	-	+	1	0.5
VIII	-	-	-	1	0.5

+=digested, -=non digested

Tab 1: *Combination of restriction-patterns within Y57c10 based on the analysis of 209 genotypes.*

Genotype	Origin	[1]RLU Allele 1 (G)	RLU Allele 2 (C)	[2]BKG RLU	[3]Adjusted Allele 1	Adjusted Allele2	[4]RRR	Allele
Carola	Germany	411.4	150.7	123.8	287.6	26.9	0.91	1
Tokyo	Germany	312.9	138.9	124.1	188.8	14.8	0.93	1
Alraune	Germany	170.3	738.3	179.8	-9.5	558.5	-0.02	2
Angora	Germany	174.3	1121	190.9	-16.6	930.1	-0.02	2
Chukurin Ibaraki 1	Japan	168.3	944.5	156.9	11.4	787.6	0.01	2
Muju covered 2	Japan	159.9	674.9	180.4	-23.5	494.5	-0.05	2
Pinarbasi	Turkey	179.2	1237	182.3	-3.1	1054.7	0.00	2
Russia 57	Russia	231.3	727.5	128.3	103	599.2	0.15	2
Turkey 235	Turkey	149.5	885.4	127.4	22.1	758	0.03	2
09-09	Israel	165.5	831.4	137.3	28.2	694.1	0.04	2

[1]RLU=relative light units, [2]BKGRLU=background relative light units, [3]=RLU for allele – background RLU, [4]RRR=adjusted allele 1 RLU/(adjusted allele 1 RLU + adjusted allele 2 RLU)

Tab 2: *Results of the analysis of a SNP in STS-G06532 by the READIT® SNP Genotyping system.*

Development of single nucleotide polymorphism (SNP) markers for Barley (*Hordeum vulgare* L.) (Subproject B)

Barley
Barley Diversity

Raja Kota, Nils Stein and Andreas Graner
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

Project goals

Recent advances in DNA sequence analysis and the establishment of chip-based assays have provided the framework for large-scale discovery and analysis of DNA variation. In this context, single nucleotide polymorphisms (SNPs) are of particular interest. SNPs are the most abundant form of genetic variation and are estimated to occur with a frequency of one out of a few hundred nucleotides. The SNP approach has several advantages over other marker assays, of which a few are listed below

1 SNPs are biallelic in populations and give rise to clear data patterns.
2 SNPs in genes can be targeted using existing EST databases.
3 EST derived SNPs are associated with the coding regions of the genome
4 SNP detection is amenable to high-throughput detection

The above mentioned features make SNPs the appropriate marker system in barley genome research for large scale analysis of genetic diversity, haplotyping in collections of unrelated and related germplasm, genome scanning in backcross populations and the development of diagnostic markers to identify and monitor allelic diversity regarding agronomic traits. The project, therefore, was focused on 1) the establishment of SNP discovery technology, 2) the determination of SNP frequency in the barley genome and genetic mapping of a set comprising 250 SNPs, 3) the development of SNP assay systems for specific traits and 4) the construction of a SNP database.

Establishment of the technology for SNP discovery

To identify SNPs in barley, we have taken advantage of EST sequences that have been generated from H. vulgare cv. Barke cDNA libraries within the projects GABI-PLANT C and GABI-SEED. Based on the sequence information from cv. Barke cDNA sequences primer pairs were preferentially designed to target the 3' UTR (untranslated regions) of ESTs. For undertaking SNP identification the target sequence then has been determined from six barley varieties currently being used in mapping studies (GABI-PLANT A) - these are: Igri, Franka (mapping population 1), Steptoe, Morex (mapping population 2), Dom and Rec (mapping population 3). Following PCR, amplicons were excised from agarose gels, purified by gel extraction and sent for sequencing. The obtained sequences from the six varieties along with the sequence from Barke were aligned and screened for SNPs. Using this approach, 460 out of 709 primer pairs yielded a single band. So far, sequencing results have been obtained for 442 SNP primer pairs. Two-hundred and fifty-seven primer pairs (58%) identified SNPs in minimum one of the given mapping populations. Seventy-four primer pairs identified SNPs between Igri and Franka, whereas 157 identified SNPs between Steptoe and Morex and 192 between Dom and Rec, respectively. Moreover, in approximately 31% of the cases, the fragment size amplified was larger than expected which can be attributed to the presence of introns within the coding regions.

Determination of the frequency of SNPs in the barley genome and the identification of a genome wide set comprising 250 SNPs

With respect to the parental combination, so far, the frequency of SNP occurrence ranged from 1/291 bp (mapping population 3) to 1/600 bp (mapping population 1). When compared among 8 barley accessions including one variety of H. spontaneum it was approximately 1/135 bp with: 80 amplicons displaying 1 SNP, 52 amplicons displaying 2 SNPs and 142 amplicons displaying 3 or more SNPs, respectively. In addition, 96 (22%) of the ESTs sequenced contained INDELs ranging from 1 bp to 164 bp. For individual polymorphic ESTs up to 8 haplotypes could be observed.

The PIC (Polymorphism information content) and the π-value (diversity index) calculated for all SNP-markers ranged from 0 to 0.87 and 0 to 0.03, respectively. Markers with higher PIC values tend to be more informative and to be of more use in association and linkage studies because of their allelic variation. However, SNPs are biallelic marker assays and the varying PIC values are dependent on the number of haplotypes detected in each SNP marker. In our study, 46% of the markers have a PIC value of more than 0.5 making them quite useful for linkage studies.

For genetic mapping, SNPs were analysed in doubled haploid (DH) progeny lines of the above mentioned mapping populations using a semi-automated DHPLC (Denaturing high-performance liquid chromatography) (fig 1) (Kota *et al.* 2001a; 2001b). So far, 230 SNPs have been mapped to various linkage groups (fig 2 and tab 1).

Because of the considerable number of amplicons that could not be mapped due to the lack of polymorphism (44% of all candidate markers) an in silicio database mining approach was initiated in a joint effort together with GABI-INFO (Kota *et al.* 2004). It builds on the fact that ESTs are derived from a number of different genotypes and possible polymorphisms may be identified by the alignment of sequences originating from different cultivars. A preliminary list of 113 ESTs potentially displaying a SNP was prepared. Sixty of these ESTs, containing 118 candidate SNPs, were selected for molecular confirmation. Upon PCR amplification, 36 primer pairs containing 63 potential SNPs yielded single bands (64%), while 14 (23%) produced no PCR product, and 10 (17%) showed PCR products with either multiple bands or weak bands and thus were not further analysed. Fifty-four (86%) of the 63 candidate SNPs were verified using the direct sequencing approach. However, nine SNPs could not be confirmed and may be explained by the presence of "false SNPs" among the sequences obtained from public databases. In this study, 28 out of 36 EST-derived sequences identified via the in silicio approach displayed one or more polymorphisms within our mapping populations. This clearly indicated a significant potential for the efficiency improvement of the polymorphic SNP identification.

○ R Kota *et al.* **Generation and comparison of ESTs derived SSRs and SNPs in barley (*Hordeum vulgare* L.).** Hereditas (2001a) 135, 145-151. ○ R Kota *et al.* **Application of DHPLC for mapping of SNPs in Barley (*Hordeum vulgare* L.)**. Genome (2001b). 44, 523-528. ○ T. Thiel, R. Kota, I. Grosse, N. Stein and A. Graner **Snp2caps: A snp and indel analysis tool for caps marker development.** Nucleic Acids Research (2004) 32. ○ R. Kota, S. Rudd, A. Facius, G. Kolesov, T. Thiel, H. Zhang, N. Stein, K. Mayer and A. Graner ○ **Snipping polymorphisms from large est collections in barley (*Hordeum vulgare* L.).** Molecular Genetics and Genomics (2003) 270, 24-33.

Development of a SNP database

To coordinate and support the mapping efforts of the projects GABI-Diversity C and GABI-Plant A with the barley EST resources program (GABI-Plant C) a MS Access-based database has been created that allows for a facilitated selection of non redundant unigenes for the individual mapping projects. With regard to the SNP-markers it contains the following information: SNP ID for a selected EST, primer sequences, amplicon sizes and whether primer pairs yielded single or multiple amplicons, ESTs containing INDELs and introns, SNPs present in a given mapping population, the mapping population and its linkage group. So far, access to the database is restricted to the participants of the project. After porting the data to an ORACLE-database structure (in progress) it is planned to provide public access to the database (in close communication with the PLA of GABI) via the IPK-PGRC website.

Conclusions

About 300 genes selected from a barley unigene set will be integrated into a barley consensus transcript map as SNP markers. This includes genes that have been identified as being differentially expressed in developing or germinating seeds (GABI-SEED). The consensus map is based on three major DH-mapping populations "Igri" x "Franka" (IF), "Steptoe" x "Morex" (SM), and "Oregon Wolfe Dom" x "Oregon Wolfe Rec" (OWB) (fig 2). In addition, our results convincingly show that in silico mining of SNPs from public databases is a cost-effective strategy, as it takes advantage of existing EST sequences thus providing a set of useful EST-SNPs for mapping studies.

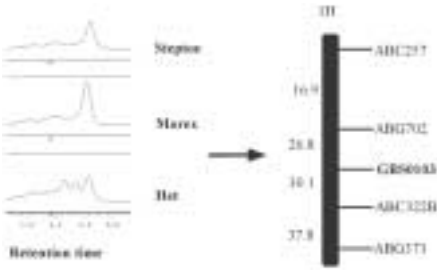


Fig 1: *DHPLC-elution profile of marker GBS0103 in the "Steptoe" x "Morex" barley DH-population and partial genetic map of barley chromosome 1H displaying the genetic position of the marker. "Het" represents a mixture of PCR products from "Steptoe" and "Morex" in equimolar ratio.*

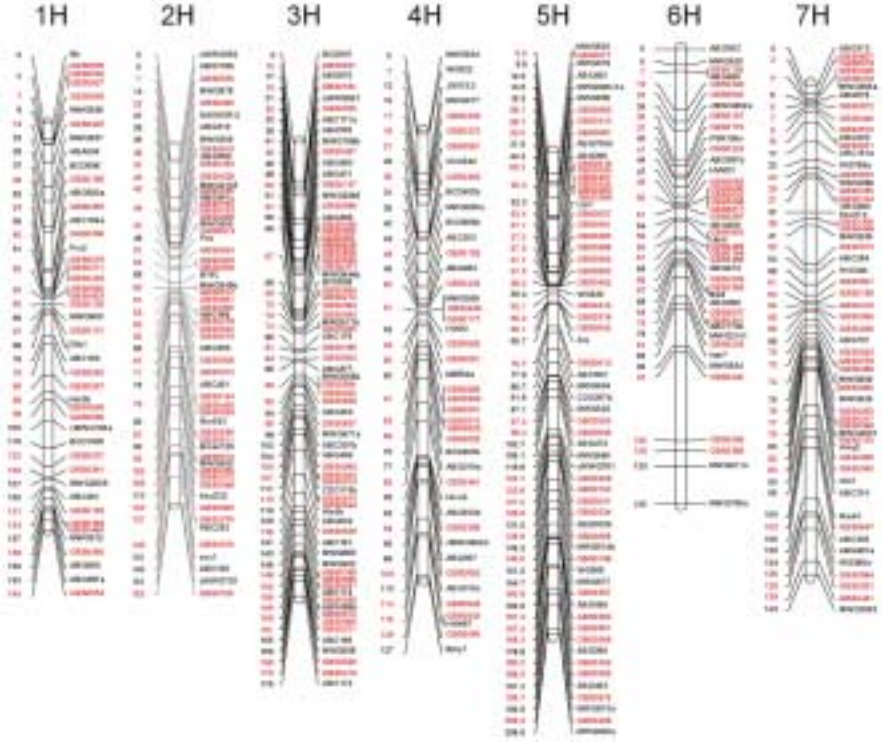


Fig 2: *A genetic consensus map of barley highlighting 216 EST-based SNPs mapped via the DHPLC assay. Markers displayed in red are SNP derived (GBS), while those in black are bin markers from genetic reference maps, respectively.*

Development of new DNA marker systems and the use of genetic resources for Barley

Barley
Barley Diversity

Ludmilla Malysheva, Martin Ganal, Marion Röder
Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben

Project goals

Barley is one of the most important large-genome cereals with a large number of varieties and accessions worldwide. Barley is being grown over a large geographic area and individual varieties are adapted to a wide range of environments. The systematic assessment and molecular characterization of barley genetic resources is the subject of this project. We use a set of mapped microsatellites distributed throughout the seven barley chromosomes to survey representative groups of the barley germplasm originating from different geographical regions. The resulting database of DNA profiles includes approximately 50,000 datapoints and will provide an overview of the genetic diversity found in barley varieties and genebank accessions.

Based on these marker data it will be possible to address the following questions:

- How related are the lines and varieties to each other?
- Is it possible to discriminate the current barley varieties for variety identification and protection purposes?
- How much genetic diversity is found in the current European barley varieties in relation to old varieties and to other geographical origins?
- Is it possible to use genebank material for the broadening of the genetic base in barley breeding?

The characterization of the population structure of the current barley genepool will provide the basis for exploiting the present genetic diversity in further studies and for associating the diversity detected with molecular markers with phenotypic characters.

Plant material

For the microsatellite analysis 948 barley lines were collected. They comprise of 332 modern spring and winter barley varieties from various breeders in Europe and a total 616 accessions from the barley core collection of the genebank in Gatersleben with origins from Europe (225), Asia (168), America (145), Africa (29) and the Near East (49). The core collection is a set of barley genebank lines which represents the entire genetic diversity within barley germplasm. DNA was extracted from pools of 5-10 plants per accession in order to detect possible heterogeneities within an accession.

Microsatellite markers

In total 522 barley microsatellites from various sources were tested and categorized for their amplification quality. Of these markers, 185 were classified as very good and 122 were classified as good. A total of 50 markers that are evenly distributed on all seven barley chromosomes were used for the analysis. The fragment analysis was performed on automated laser fluorescence sequencers (ALFexpress, Amersham Biosciences) (Fig 1).

Construction of the database

By the end of the project, the genotyping of all datapoints has been finished and the datapoints have been entered into the database (tab 1). For each marker an allele database was produced which contains the genotyping information for all barley accessions coded in a 1/0 format (fig 2).

For the analyzed barley genotypes (including two H. spontaneum accessions) we identified a total of 800 alleles with the fourty-eight microsatellite markers included in the statistical analysis. The average number of alleles per locus was 16.7. The number of alleles per locus ranged from 5 to 33 (tab 1). The polymorphism information content varied from 0.47 to 0.92. All accessions except two varieties originating from the same breeder could be distinguished on the basis of the marker data confirming the high discrimination power of the employed microsatellite markers.

The utility of the database for assessing first trait/marker relationships was tested for growth habit (spring or winter type). For a limited number of microsatellites, alleles were identified for which the allele frequencies were associated with the respective trait.

Conclusions

The genetic resources of cultivated barley were characterized with microsatellite markers evenly distributed through the barley genome. The resulting database is a prerequisite for a further exploitation of the genetic diversity present in the barley genepool with modern methods of genetics and genomics. As another product of this study, a set of microsatellite markers was identified that permits variety identification for such purposes as variety registration and variety protection.

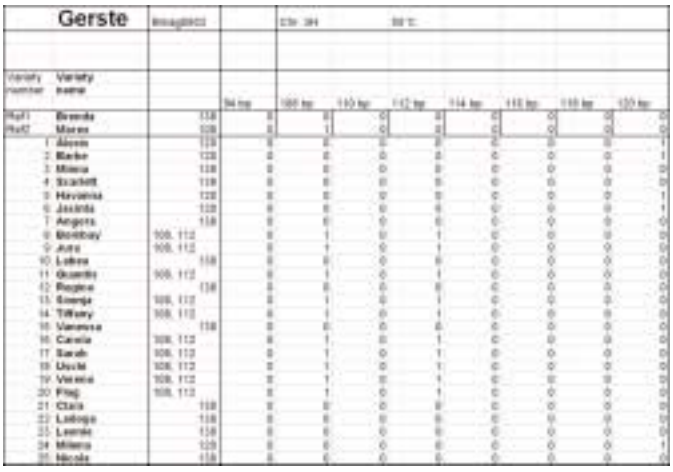
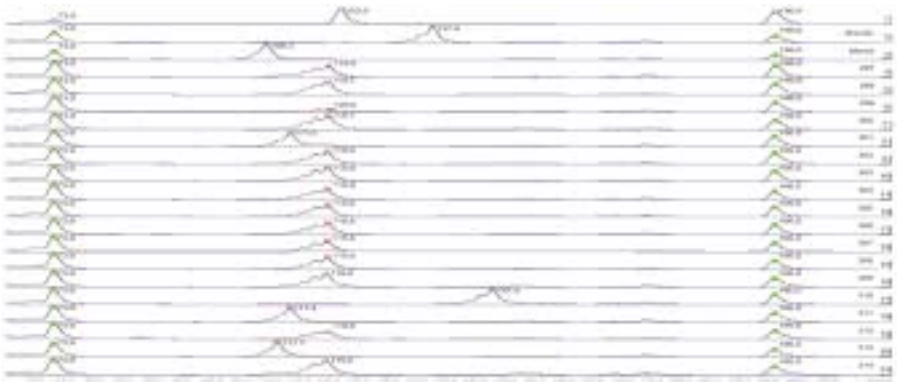


Fig 2: *Section of the preliminary database file for the marker Bmag0603.*

	Marker	Chromosome	No. of alleles detected
1	Bmac0032	1H	33
2	Bmag0211	1H	13
3	Bmag0382	1H	14
4	Bmag0579	1H	12
5	Bmag0718	1H	13
6	GBMS0184	1H	13
7	HVM20	1H	12
			∑=110
8	Bmag0518	2H	21
9	Bmag0749	2H	9
10	GBMS0160	2H	15
11	GBMS0229	2H	12
12	GBMS0247	2H	18
13	HVM36	2H	19
14	HVM54	2H	12
			∑=106
15	Bmag0013	3H	25
16	Bmag0225	3H	22
17	Bmag0603	3H	25
18	EBmag0705	3H	20
19	GBMS0046	3H	30
20	GBMS0189	3H	27
21	HVM60	3H	19
			∑=168
22	EBmac0701	4H	17
23	EBmac0788	4H	20
24	EBmac0906	4H	11
25	GBMS0087	4H	17
26	GBMS0133	4H	13
27	HVM40	4H	14
28	HVM67	4H	11
			∑=103
20	EBmac0684	5H	10
30	GBMS0032	5H	14
31	GBMS0119	5H	14
32	GMS001	5H	13
33	GMS027	5H	32
34	GMS061	5H	7
			∑=90
35	Bmac0040	6H	28
36	Bmac0316	6H	26
37	Bmag0613	6H	27
38	EBmac0602	6H	21
39	GBMS0083	6H	14
40	GBMS0125	6H	6
41	HVM65	6H	5
			∑=127
42	Bmag0135	7H	15
43	Bmag0507	7H	18
44	EBmac0755	7H	20
45	GBMS0035	7H	6
46	GBMS0111	7H	15
47	GBMS0183	7H	10
48	GBMS0192	7H	12
			∑=96

Tab 1: *List of markers analyzed for the project.*

Fig 1: *Amplification profiles of the microsatellite marker Bmag0603. Printout of the computer program Fragment Analyzer 1.02.*



Functional genomics of developing and germinating Barley seeds

GABI Seed Overview

Seeds harbour the embryo proper of the plant and a large amount of storage products as substrate for seedling growth during germination and early seedling development. Crop plant seeds provide the most important foodstuff for humans and domesticated animals. They are a concentrated source of carbohydrates, proteins and lipids but also of minerals, vitamins and fibers. But seeds have more features important to mankind. They survive long periods of unfavorable conditions, they can be shipped around the globe and they store the genetic makeup which has been and still is steadily improved by combined efforts of breeders and scientists with respect to human needs.

Among crop plants, cereals are by far the most important. Whereas maize, rice and wheat are economically leading, barley is the experimentally most amanable cereal of the Northern hemisphere and is mainly used for animal feeding and for brewing. It has been chosen within GABI as both, a crop plant and a cereal model plant to investigate in detail seed development and germination (Fig 1) to better understand storage product synthesis and accumulation as well as germination and its most important application, the malting process of the brewing industry. Other cereal characters, breeders would nowadays like to change or improve are, for example:

- yield (still one of the most important global goals);
- grain disease resistances;
- technical parameters with respect to harvest, storage, processing, milling- and baking quality;
- content of specific substances (starch, protein, specific fatty acids etc.);
- elimination/reduction of substances critical for all or certain groups of people like allergenes and substances as gluten (responsible for coeliaky) or phytin/phytase;
- improving the nutritional value for farm animal feeding (content of essential amino acids, elimination of critical substances, improving digestibility).

The listed parameters are usually determined by several/many genes (quantitative trait loci or QTLs). A better understanding of the molecular and physiological processes underlying the expression of the described characters will certainly improve our abilities to achieve the strived for goals by marker-assisted breeding, by genetic engineering or by a combination of the two approaches.

General projects goals

The major goals of the GABI-SEED project defined in the project proposal were:

- establishment (in a joint effort with GABI-PLANT) of a large collection of cDNAs from developing and germinating barley seeds;
- production and provision of high density "unigene" DNA arrays;
- global expression analyses by cDNA arrays (which also helps to identify interesting promoters, an important tool for biotechnology);
- medium-scale in situ hybridization with the final goal of establishing a 3D expression database;
- Development of nearly isogenic lines with respect to malting quality parameters as basis material for expression analysis;
- analysis of gene expression during the malting process;
- expression profiling of different varieties by means of cDNA-AFLPs;
- establishment of supportive bioinformatic tools and management data bases.

The project is divided into three subprojects investigated by three research groups:

1 Functional genomics of barley seed development;
2 Functional genomcis of germinating barley seeds;
3 Functional genomics of malting quality.

Functional Genomics of Barley Seed Development (Subproject A)

Barley
GABI Seed

Winfriede Weschke, Nese Sreenivasulu, Volodymyr Radchuk, Sabine Gubatz, Ulrich Wobus
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

ESTs and Macroarrays

During early grain development (0-7 days after flowering = DAF), caryopses were separated into maternal and mainly filial tissues (embryo and endosperm). Two cDNA libraries were constructed using RNA from the separated tissue material together with two additional libraries representing whole caryopses between 8 to 15 and 16 to 25 DAF. From these libraries, 47,066 expressed sequence tags (ESTs) were generated. Along with other EST collections available from other stages of plant development (GABI-PLANT) in silico expression analysis has been performed to identify genes preferentially expressed in seeds (Zhang *et al.* 2004). Already within 2001, two types of macro-arrays had been constructed and extensively used for expression analyses: a 711 cDNA array with at least 620 unigenes (Sreenivasulu *et al.* 2002) and a 1,400 cDNA array (Potokina *et al.* 2002; Sreenivasulu 2004). Recently, a high density array based on the EST collection described above, containing 11,787 tentative unique genes from developing seeds, has been fabricated by the RZPD/Berlin and successfully tested in our laboratory.

Transcriptom Analysis of Seed Development

Seed development is generally divided into 3 phases: (1) the cell division or pre-storage phase (embryogenesis sensu strictu), (2) the maturation or storage phase (synthesis and accumulation of storage products) and (3) the desiccation phase (water loss). Our cDNA libraries are representative of the pre-storage and storage phase (0-25 DAF, fig 2). Expression analyses were focussed on the pre-storage and early to mid storage phase (0-12 DAF, fig 2).

After an initial analysis of barley caryopses gene expression using the 711 cDNA macroarray (Sreenivasulu *et al.* 2002), we investigated the time span 0-12 DAF, representing the cell division and early to mid maturation phase, in 2-day steps in maternal (mainly pericarp) as well as filial (endosperm/embryo) tissue preparations using the 1,400 macro-array.

Principle component analysis (PCA) of all expression data revealed that the networks of gene expression especially in filial tissues define, besides of the pre-storage phase (0, 2 and 4 DAF) and the storage-phase (10 and 12 DAF), intermediate developmental steps (6 and 8 DAF) indicating dramatic reprogramming of the transcriptional machinery. We define this period as intermediate or transition phase (fig 3).

K-mean clustering of expression profiles resulted in 16 gene sets, which can be arranged into 6 cluster groups. 7% of the genes (=79) were preferentially expressed in maternal tissues, 22% (=258) showed higher expression in filial tissue preparations. Clusters are often composed of functionally related genes as schematically indicated in fig 2. Maternal tissue-specific clusters are characterized by up-regulation during DAF 6-12 of certain protease genes and genes involved in hormonal-induced processes and lipid mobilization pointing to degradation processes and programmed cell death during formation of the final protective husk of the barley grain. A pre-storage gene set with high expression during 0-4 DAF in filial tissue probes contains genes mainly involved in regulation of cell division and cell elongation. A group of genes characterized by a bell-shaped expression profile with a peak at 6-8 DAF is associated with photosynthesis and energy production. The storage phase is mainly represented by two sets of genes. Genes of carbohydrate metabolism (starch biosynthesis, glycolysis and citrate cycle) are highly expressed during 6-12 DAF and storage protein and protease inhibitor genes during 10-12 DAF. Thus, biosynthesis of these two major groups of storage compounds is not co-ordinately regulated. (http://pgrc.ipk-gatersleben.de/seeds/)

Spatial Patterns of Gene Expression: 3D-Models of Developing Seeds

Gene expression is highly regulated in time and space, but global analyses of gene expression neglects fine resolution of spatial expression patterns. *In situ*-hybridization is the most appropriate method but still provides only a 2-dimensional data set. We therefore aimed at building 3-dimensional computer models as frames to store 2-D patterns and to reconstruct 3D images of gene expression patterns. Three models have been finished visualizing the caryopsis at anthesis (0 DAF), with syncytial (3 DAF) and cellularized, closed endosperm (6 DAF). Model development is based on the software Amira. Ongoing adaptations are done in cooperation with the software developer, the Konrad-Zuse Zentrum (Berlin). Digital videos are now produced from the models and will be available through a website connected to the IPK home page.

Figure 4 demonstrates several important aspects of the model. First, using of 3µm transversal sections for modeling results in a virtual image of a high-resolution longitudinal section (left) nearly identical to the histological picture (down right). Second, from a set of *in situ* hybridized sections (lower middle pictures; arrows point to the outer endospermal cell row labeled for cell wall-bound invertase1, CWINV1), a 3D-expression pattern of CWINV1 can be reconstructed (upper middle part) showing labeling of two cell types in blue and green (see also Weschke *et al.* 2003). Clear pictures of gene function networks are expected from integration of 3D-mRNA patterns of functionally related genes. Consolidating our recent findings a review has been published focusing on specific aspects of molecular physiology of developing seeds and novel insights gained by genomic approaches (Wobus *et al.* 2004).

○ E. Potokina *et al.* **Differential gene expression during seed germination in barley (*Hordeum vulgare L.*).** Funct. Integr. Genomics (2002) 2, 28-39. ○ N. Sreenivasulu *et al.* **Identification of genes specifically expressed in maternal and filial tissues of the barley caryopsis: a cDNA array analysis.** Mol. Gen. Genomics (2002a) 266, 758-767. ○ W. Weschke *et al.* **The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development.** Plant Journal (2003) 33, 395-411. ○ W. Michalek *et al.* **EST analysis in barley defines a unigene set comprising 4,000 genes.** Theoretical Applied Gentics (2002) 104, 97-103. ○ E. Potokina *et al.* **Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare* L.)** Molecular Breeding (2004) 14, 153-170. ○ N. Sreenivasulu *et al.* **Mining functional information from cereal genomes – the utility of expressed sequence tags.** Current Science (2002) 83, 965-973. ○ N. Sreenivasulu *et al.* **Transcriptome changes in foxtail millet genotypes at high salinity: Identification and characterization of a PHGPX gene specifically upregulated by NaCl in a salt-tolerant line.** Journal of Plant Physiology (2004) 161, 467-477. ○ N. Sreenivasulu *et al.* **Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains.** Plant Journal (2004) 37,539-553. ○ H. Zhang *et al.* **Large-scale analysis of the barley transcriptome based on expressed sequence tags.** Plant Journal (2004) 40, 276-290. ○ U. Wobus *et al.* **Molecular physiology and genomics of developing barley grains.** Recent Research Developments in Plant Molecular Biology (2004) 37/661 (2): 1-29.

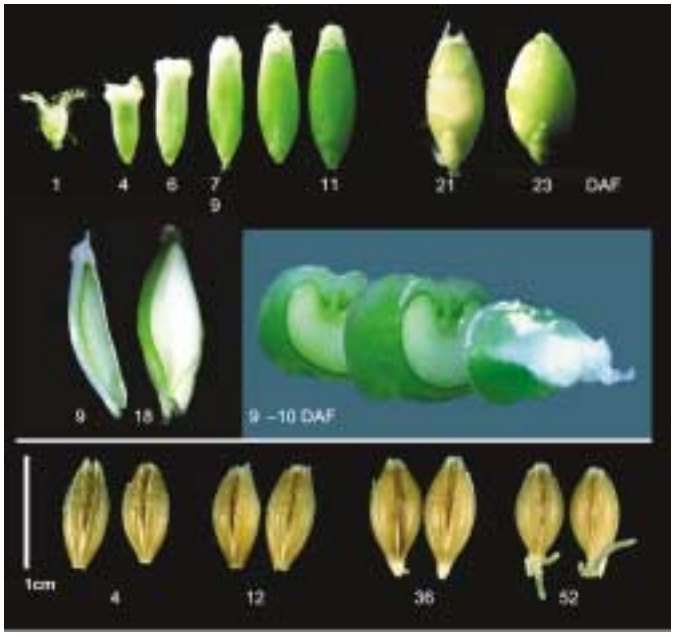


Fig 1: *Developing (upper panel) and germinating barley seeds (lower panel). To show the organization of the developing barley grain (lower part of the upper panel), parts of the maternal tissue were peeled off to show the surface of the embryo sac (left, 9 DAF). The separated filial grain part was longitudinally cut (left, 18 DAF) and the developing grain was sectioned transversally into three pieces (right, 9-10 DAF). DAF: days after flowering; germinating seeds are shown 4, 12, 36 and 52 hrs after imbibition.*

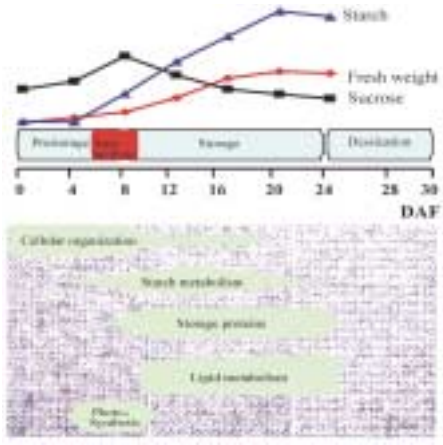


Fig 2: *Stages of barley grain development correlated to fresh weight increase and sucrose and starch profiles during development. The intermediate phase of barley grain development was identified by Principle Component Analysis (PCA) (see Fig 3). Metabolic pathways dominating different developmental phases (illustrated by green areas) were identified by data mining of ESTs from specific cDNA libraries. Background behind the green areas: pattern resulting after hybridisation of a labelled cDNA probe representing the transcriptom of the filial part of caryopses 4 DAF to an array filter containing about 12,000 caryopsis-specific ESTs. DAF: days after flowering.*

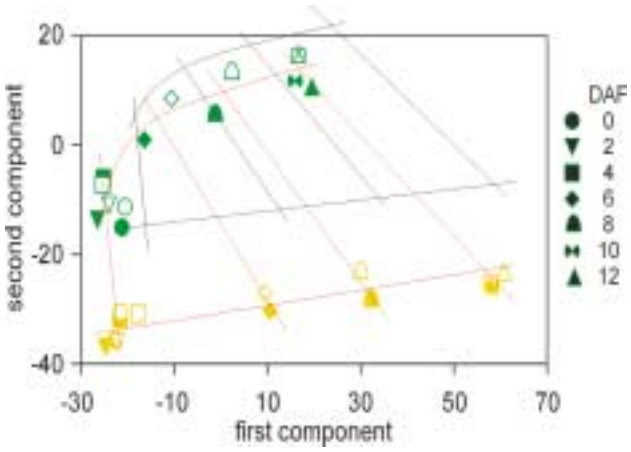


Fig 3: *Steps during barley grain development as identified by Principle Component Analysis (PCA) of expression data. Red lines in the figure connect PCA results obtained by using normalized, array- and gene centred hybridisation signals of about 1,400 cDNA fragments spottet on a cDNA macro-array filter. Black lines connect PCA results of only 337 highly regulated candidate genes (basis for K-mean clustering, see below). Green symbols: maternal tissues; yellow symbols: filial tissues; filled symbols: experiment 1; open symbols: experiment 2. (Sreenivasulu et al. 2004)*

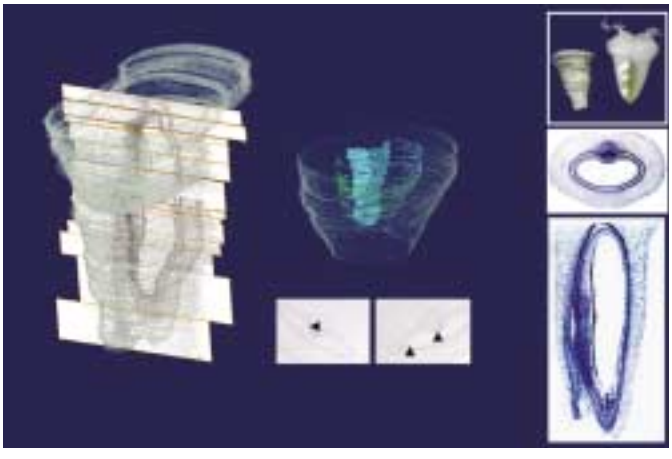


Fig 4: *3D-modelling of a caryopsis with syncytial endosperm and 3D-pattern of cell wall invertase-mRNA expression. See text for further explanation.*

Functional genomics of germinating Barley seeds (Subproject B)

Barley
GABI Seed

Elena Potokina, Nils Stein, Andreas Graner
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

The principal goal of the project is to reveal the spatio-temporal pattern of gene expression during germination and to develop an approach to relate differential gene expression as it is observed on a cDNA array to the expression of agronomic traits (functional association).

Differential gene expression during seed germination in barley

A barley cDNA macroarray comprising 1440 mostly unique genes was used to score the spatial and temporal expression patterns of gene expression in embryo, scutellum and endosperm tissue during different stages of germination (Potokina *et al.* 2002). Among the set of expressed genes 68 genes displayed the highest mRNA level in endosperm tissue, 58 genes were upregulated in both embryo and scutellum, 11 genes were specifically expressed in the embryo and 16 genes in scutellum tissue. Based on BlastX analyses 70% of the differentially expressed genes could be assigned a putative function. One set of genes expressed in both embryo and scutellum tissue included functions of the cell division cycle, protein translation, nucleotide metabolism, carbohydrate metabolism and some transporters. The other set of genes expressed in endosperm codes for several metabolic pathways including carbohydrate and amino acid metabolism as well as protease inhibitors and storage proteins. Northern blots with B3-hordein (HY06A05) and trypsin inhibitor (HY04J03) confirmed the specific presence of mRNA in endosperm tissue. However, only degraded transcripts could be detected (fig 1). Those genes belong to a cluster of 29 genes which are characterized by the decrease of their mRNA levels during germination. This observation provides experimental evidence that the endosperm of germinating barley grain contains a considerable amount of abundant mRNAs which probably were produced during seed development and which are degraded during the early stages of germination.

Functional association between complex traits and cDNA array based expression patterns

To relate quantitative differences in gene expression to the phenotypic variation of quantitatively inherited traits a generic strategy was devised (Potokina *et al.* 2004). This offers the possibility to identify candidate genes for agronomic traits that display quantitative variation. To prove the principle, the strategy is demonstrated in a pilot study using a cDNA array (1440 ESTs) to identify genes involved in the variation of the complex trait "malting quality" in barley. To this end, RNA expression was analysed in malt samples of a set of 10 barley accessions that were previously characterized for 6 quality associated trait parameters.

Candidate genes can not simply be identified by differential gene expression observed between two phenotypically most contrasting lines, since the differential gene expression in a random sample is not only restricted to the trait under consideration. Our strategy of candidate gene identification follows a two step procedure: 1) the relationship between a representative number of different well described lines (populations, cultivars) is determined by cluster analysis based on phenotypic parameters, 2) extraction of differentially expressed genes whose expression profiles support the phenotype-based relation between the lines. This approach allows to eliminate a background noise and to distil a set of candidate genes whose expression is related to the quantitative differences in the target trait.

Technically, the selection of the candidate genes was achieved via a correlation of dissimilarity matrices (Manhattan distance matrix) that were on one hand based on trait parameters and on the other hand on expression data (fig 2). As expected, a comparison based on the complete set of differentially expressed genes did not reveal any correlation between the matrices, since not all genes that show differential expression between the 10 cultivars are responsible for the observed differences in malting quality. However, by iteratively taking out one gene (with replacement) and recomputing the correlation, those genes positively contributing to the correlation could be identified. Using this procedure 19 genes (ESTs) were identified that yielded a dissimilarity matrix which significantly correlated with the malting dissimilarity matrix at the 5% level.

In case a differentially expressed gene has a direct impact on quality parameters (cis-acting factor) the genetic linkage of this gene to a quantitative trait locus (QTL) may be taken as additional evidence for the validation of a candidate gene. Therefore, selected candidates were mapped in a population segregating for malting quality traits (http://wheat.pw.usda.gov/ggpages/SxM/). To date, six out of seven mapped candidate genes displayed linkage to QTLs for malting parameters.

The approach developed in the project represents a generic strategy to relate variation in gene expression to the variation in a phenotypic trait. The procedure obviates the necessity to develop experimental populations (e.g. segregating populations), near isogenic lines or mutants, which are not available for most agronomic traits. The functional association strategy described here capitalizes on the genetic diversity already present in a set of breeding lines, cultivars or a population. Thus, it may provide an efficient link between functional genomics and plant breeding.

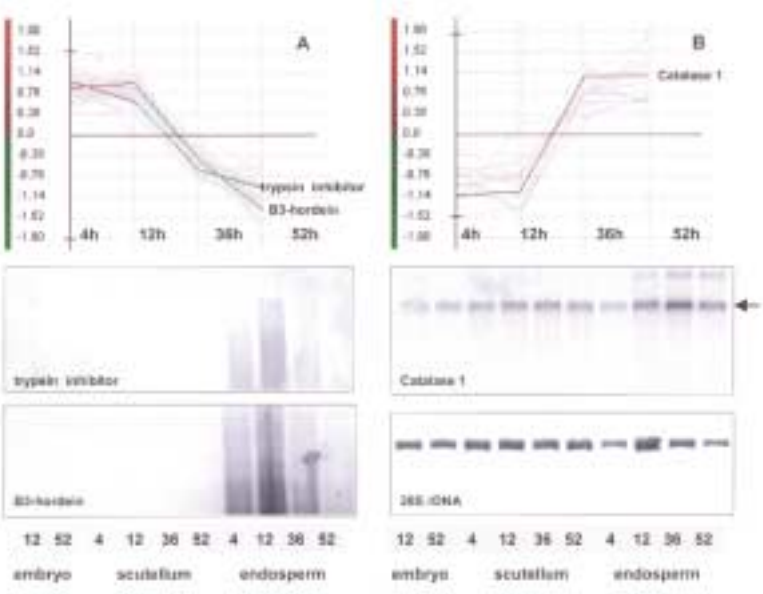


Fig 1: *Degraded transcripts in the endosperm of germinating barley grain. Two different sets of genes up-regulated in endosperm tissue display decreasing and increasing courses of mRNA expression along the time scale. Northern Blot analysis of temporal specific distribution of A: B3-hordein (HY06A05); trypsin inhibitor precursor (HY04J03) and B: catalase 1 (HY03C01) is shown. All lines contained 5mg of total RNA from embryo, scutellum and whole endosperm (aleurones plus starchy endosperm). RNA profiles for clones HY04J03 (A), HY03C01 (B) and 26S rDNA (control) were obtained from the same blot.*

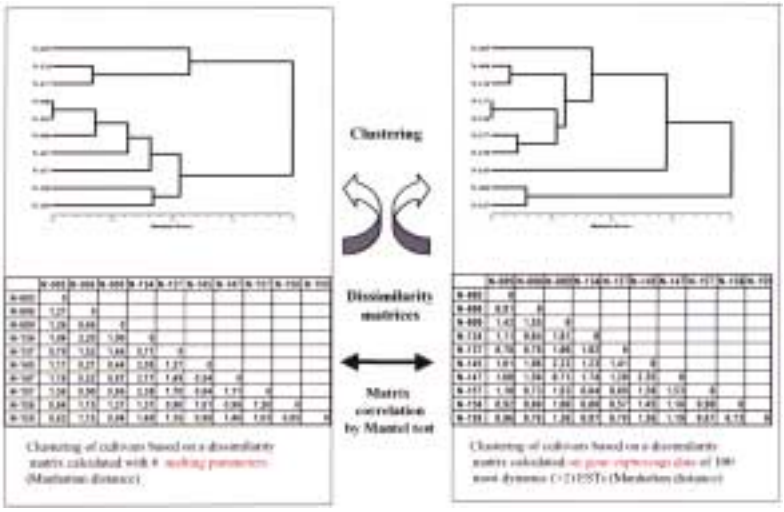


Fig 2: *Functional association analysis. Identification of candidate quality genes via matrix correlation by Mantel test. First, cultivars can be clustered based on a dissimilarity matrix calculated from either malt data (left) or the expression data (right). Next, the significance of the correlation of the malt- and expression-dissimilarity matrices were evaluated by a Mantel test. Using the complete dataset of the 100 most dynamic genes, the initial correlation of malt and expression data was not significant. However, by iteratively taking out one gene (with replacement) and repeating the Mantel test, those genes positively contributing to the correlation with the malting dissimilarity matrix could be identified.*

Functional Genomics of Malting Quality (Subproject C)

Barley
GABI Seed

Markus Herz, Sabine Mikolajewski, Günther Schweizer
Bavarian State Research Center for Agriculture, Freising

Development of nearly isogenic lines as basis material for expression analysis

Chromosomal regions contributing to malting quality have previously been identified at the LfL by the construction of a QTL map based on a double haploid (DH) population derived from the cross of two malting barley varieties "Alexis" x "Steina" (Hartl *et al.* 2000). These genomic regions influence different malting quality traits and maintain their strong effects in each of the six environments under investigation. In four particular genomic intervals on chromosomes 1H, 4H, 5H and 6H highly effective QTLs have been detected and therefore these regions have been termed as "QTL-Hot-Spots" for malting quality.

Eleven individual lines were selected from this phenotypically and genotypically well characterized population to establish the backcrossing program for the development of nearly isogenic lines (NILs) for malting quality QTLs.

Genetic markers flanking interesting genomic intervals on chromosomes 1H, 4H, 5H and 6H were used to select for the four genomic regions contributing to malting quality characters. Out of a set of 148 AFLP markers chosen for marker assisted selection, 25 AFLPs are located within the four QTL regions and 123 support the selection of the genetic background. To saturate the map 30 additional microsatellite markers have been included into linkage analysis during the project period. Twenty four of these markers have been chosen from a published SSR map (Ramsay *et al.* 2000) and six further SSRs have been selected from a set of markers developed from EST sequences by the IPK independently from GABI. Out of these, 10 SSRs from the SCRI collection and four SSRs from the IPK could be located in the bounds of the QTL intervals.

Four cycles of backcrossing and one generation of selfing resulted in a total of 8 BC4S1 lines. For each QTL-Hot-Spot two corresponding lines are available which display identical genetic constitution except the different introgression at the respective QTL interval. After further selection and seed multiplication, field trials in several environments are in progress to validate the effects of the QTL regions which have been introgressed into a different genetic background.

The corresponding NILs will be well suited to perform subtractive suppression hybridisation (SSH) experiments aiming at the identification of differential expressed genes correlated to malting quality traits.

In general, the QTL NILs represent a very well defined material as basis for the analysis of certain genomic regions which are characterized for their contribution to malting quality. Genes involved in malting quality traits can be identified and specific markers for the selection in highly adapted breeding accessions can be developed (fig 1).

Analysis of gene expression during the malting process

As illustrated by figure 2 the malting process in the micromalting facility at the LfL is a strictly controlled fully automated procedure taking place in a closed system. A number of 48 samples can be processed in a single micromalting machine. The entire micromalting facility consists of five machines which allow a throughput of 240 samples per week. Routinely 12 important parameters are analyzed in the quality laboratory operated by the LfL to determine the malting quality of the samples. Due to the special construction and operating mode of the machinery it is possible to take samples at any point of time during the malting process.

During the project period various methods for RNA analysis have been evaluated and adapted for the purpose of analyzing malting quality at the level of gene expression at the LfL. Sophisticated methods for RNA extraction from highly polysaccharide containing malt kernels, cDNA synthesis, RT-PCR, Real-Time-Quantitative PCR and cDNA-AFLP are now available analytical tools to allow a close look at the expression of genes contributing to malting quality.

Using gene specific RT-PCR, the changes of expression of two representative genes which are involved in cell wall degradation and starch metabilosm, Limit-Dextrinase and Xylane-Endohydrolase, could be visualized during the malting process. Limit Dextrinase shows an expression maximum after 22 hours, while the expression of Xylane-Endohydrolase starts after 56 hours. These results correspond to the measurements of the activity of theses enzymes described in literature (Narziß 1999) and could also be confirmed by a Real Time PCR assay for the purpose of relative quantification of the gene encoding Limit-Dextrinase. Gene expression was traced at six points of time during malting of the variety "Barke" between 2 hours and 56 hours of malting. Additionally an EST from the IPK-EST library with similarity to a Peroixdase gene (HY10B21v) was used as template for quantitative PCR. In contrast to Limit dextrinase, the expression of this gene rises continuously to a level of approximately 20-fold of the expression at 56 hours of the micromalting process.

On the basis of these results the time frame for taking samples was restricted to the first 76 hours of malting.

Expression profiling of different varieties by means of cDNA-AFLPs

Eleven spring barley varieties have been selected from a set of varieties grown for the Landessortenversuche (Bavarian regional variety trials). Three feeding varieties, "Orthega", "Baccara" and "Eunova" and eight malting varieties, "Alexis", "Barke", "Steffi", "Thuringia", "Scarlett", "Ria", "Anna-bell" and "Danuta" have been chosen from this set to evaluate the accuracy of the method and the degree of differential expression.

From the starting point of the micromalting process until the first 76 hours 18 samples were taken from each variety in various time intervals between 2 and 8 hours (fig 2).

In total 396 samples are available for diverse investigations. On order to identify even rarely expressed transcripts the cDNA-AFLP-method was chosen to generate expression profiles.

○ L. Hartl, G. Schweizer, M. Herz and M. Baumer **Molekulargenetische Lokalisierung von QTL für die Malzqualität der Gerste.** Bericht über die Arbeitstagung 2000 der Vereinigung österreichischer Pflanzenzüchter, 21. bis 23. November 2000, (2000) 117-122. ○ Mikolajewski *et al.* **Untersuchung differentieller Genexpression im Verlauf der Vermälzung von Gerstenkörnern – ein Vergleich von Brau- und Futtergerstensorten mittels cDNA-AFLP-Technik.** Vortr. Pflanzenzüchtung (2002), 54, 405-408. ○ L. Ramsay *et al.* **A simple sequence repeat-based linkage map of barley.** Genetics (2002) 156, 1997-2005. ○ L. Narziß **Technologie der Malzbereitung.** Ferdinand Enke Verlag Stuttgart (1999). ○ M. Herz and S. Mikolajewski **Nucleic acid molecules involved in quality traits and methods for their isolation.** European patent application (2003) 03 010 518.13. ○ M. Herz, S. Mikolajewsi, G. Schweizer and M. Baumer **Identification of candidate cDNAs correlated to malting quality of barley by means of cDNA-AFLP analysis and differential genotype pooling.** Proceedings of the 29th EBC congress (2003), Dublin 2003

CDNA AFLP analysis was performed based on the restriction enzymes PstI and MseI for digestion of the ds cDNA (Mikolajewski *et al.* 2001).

Until now expression patterns of all eleven barley varieties have been compared at 2 hours and 24 hours stages of the malting process using 25 AFLP primer combinations. An average number of 26 bands per primer combination was detected. In total 654 differential transcript derived fragments (TDF) could be identified.

Reproducible differences in expression patterns occurred both between cultivars and between different germination stages during the malting process. According to the pedigree of the varieties as well as their malting quality parameters, 80 fragments were identified which were most likely to be able to distinguish between varieties and feeding and malting quality, respectively. In addition, bands which displayed differences in expression between the two points of time were taken into account for selection. The accuracy of the selection was confirmed by the comparison of a similarity cluster of the 80 selected differential TDFs with clusters based on phenotypic quality data as well as information of genomic AFLP markers of all eleven varieties. Until now sequence information of eleven differential TDFs is available.

Using BLAST search the sequences of these eleven TDFs were compared to known sequences in public databases. BLAST searches resulted in similarity to sequences of different grass species. Several fragments show similarity to genes which are presumably involved in starch metabolism.

In addition sequence information of several differential TDFs gives evidence that genes known to be involved in stress defence reactions and protein protection may play a role in the context of malting quality.

A recently developed method named *MAGS* (**M**arker **A**ssisted **G**enotype **S**creening) (Herz & Mikolajewski, 2003, Herz *et al.* 2003) will allow the assignment of the differential TDFs to chromosomal regions which contribute to malting quality. This novel method is based on the marker assisted pooling of selected members of the QTL mapping population. Two contrasting pools each consisting of ten plants carrying the favorable marker alleles for a QTL interval and the respective phenotypic performance are compared to each other on the gene expression level.

First detailed *MAGS* analyses for a QTL on chromosome 6H identified three unique gene sequences which could be assigned to this QTL interval. These sequences will be used to develop molecular markers to facilitate the selection for malting quality in barley breeding programs.

During the project period of GABI 1 a solid scientific basis has been established to perform extensive expression analyses concerning malting quality in barley. Differential plant material is now available as well as the methodology to identify even rarely transcribed mRNAs by the application of cDNA-AFLPs and SSH. Confirmation of the correlation of identified sequences and genes to malting quality will be possible by mapping these loci in the malting quality QTL map developed by the LfL and by estimating the influence of these loci on important quality characters.

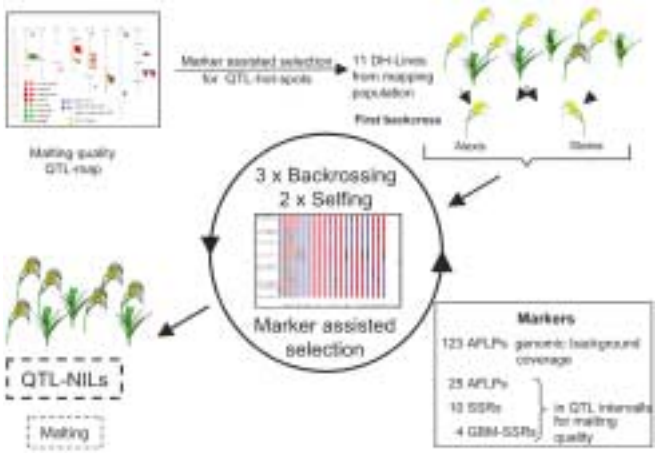


Fig 1: *Development of QTL-NILs for malting quality.*

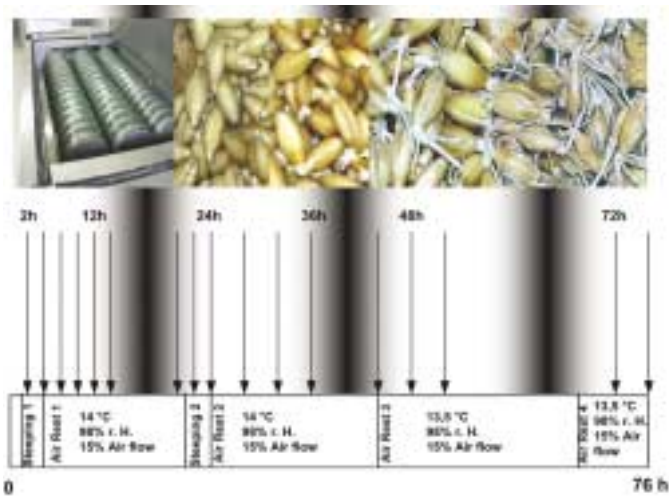


Fig 2: *Scheme of the first 76h of the micromalting process. Photos in the upper part show the micromalting machines and the development of the germinating seeds during the micromalting procedure. Arrows represent points of time for sampling.*

Comparative sequencing of two regions of the Barley and Rice genome

Barley
GABI Map

Roberto Cossu, Dragan Perovic, Nils Stein, Martin Ganal, Andreas Graner, Marion Röder
Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben

Project Goal

The goal of the project is the sequencing of two regions of the barley genome in comparison to their respective syntenic region in the rice genome. The two intervals comprise of the genomic region around the *Rh2*-resistance gene on chromosome 7H and the region of the rym4/rym5-resistance gene on chromosome 3H of barley. The results of this study will provide information concerning the level of synteny in regions containing disease resistance genes and analyze the possibility of utilizing the sequenced rice genome for map based cloning in barley.

Analysis of the *Rh2* region on chromosome 7H

Rice (*Oryza sativa*) and barley (*Hordeum vulgare*) are both members of the Gramineae. Whole genome mapping by means of RFLP markers has shown that large linkage blocks are conserved between both organisms. From previous research, it was known that the region on chromosome 1 (7H) of barley harbouring the resistance gene *Rh2* against scald (*Rhynchosporium secalis*) between the markers PSR119 and R2869 shows a high level of macrosynteny to a genomic region on rice chromosome 6. Based on these mapped markers, a BAC contig of approximately 250 kb was constructed by screening a barley library of the cultivar Morex. The markers hybridizing to the BACs are cosegregating with *Rh2* in a population of approximately 750 plants (1500 gametes). The respective BAC contig was delimited to the right side by one recombination event in relation to *Rh2*. An extension of the contig towards the left side was not yet successful due to a high amount of repetitive sequences in this region. Therefore the information of the known rice PAC sequence was exploited in order to find a new starting point for contig construction starting from the other side of the cosegregating region. RFLP probes were derived from the hypothetically predicted genes on the rice-PAC chromosome 6 and used for a further screen of the barley BAC library. This led to the isolation of a BAC which maps one recombination left from *Rh2*. The current work focuses on closing the gap between this BAC and the cosegregating contig.

An increase of the population size to approximately 2500 plants led to the discovery of seven new recombinants in the immediate vicinity of the *Rh2* gene. These new recombinant plants seperate the BAC contig by one recombination from the *Rh2* gene, while the BAC on the left side spans four recombinant plants and is now separated three recombinants from the *Rh2*-gene.

The DNA sequence of the three BACs comprising the 250 kb interval was determined by shotgun sequencing of the individual barley BACs. The complete sequence was assembled into one contig which still contains three small gaps (Fig 1). The genomic sequence of rice in the respective syntenic region on rice chromosome 6 has been published by the Japanese rice genome project.

First results from the sequence analysis are depicted in Fig 1. As in other organisms within the Gramineae that have large genomes, most of the sequences region of 250 kb are retrotransposons or other repetitive sequences. For three putative rice genes homologous putative genes were found in the sequenced barley contig confirming the microsynteny between rice and barley in the respective region. Of particular interest is the KAO1 gene which encodes ent-kaurenoic acid oxidase, an enzyme of the gibberellin biosynthetic pathway with homologues in maize and Arabidopsis. In barley and maize mutations in KAO give rice to gibberellin-responsive dwarf phenotypes. The second gene OSEYA1 has homology to the Drosophila mutant eyes absent, homologues were also described for other plant species such as Arabidopsis and maize. For the third gene no known homologues were found. The barley sequence of 250 kb corresponds to approximately 15 kb in rice and has thus undergone a large expansion by multiple insertions of retrotransposons. A more detailed analysis of the respective elements and genes is currently being performed.

By exploiting the so far complete synteny in gene order between rice and barley, further barley BACs have been isolated to expand the investigated region and will be sequenced in the near future.

Analysis of the genomic region around rym4/rym5 on chromosome 3HL

The *rym4/rym5* locus of barley confers resistance to barley mild mosaic virus (BaMMV) and barley yellow mosaic virus (BaYMV-1, -2) and has been mapped to the telomeric end of barley chromosome 3HL. Almost complete marker colinearity has been reported for barley chromosome 3H and rice chromosome 1 and the putatively syntenic region of the *rym4/5* region in rice was targeted by comparative and indirect mapping of the rice EST-based markers C112 and RZ783 (distal) as well as the rice EST-based marker R117 and a barley probe No777 (519H2) (proximal), respectively. This allowed to characterize a syntenic interval in barley and rice with markers C112 and R117 flanking the *rym4* gene in barley. The target interval in rice is covered by four overlapping BACs (Acc.-#: AP003263, AP004365, AP003448, AP003277, http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/status.pl) forming a contig of about 580 kb, with a physical distance of 490 kb between the two markers R117 and C112. In analogy to the *Rh2* project sequencing in rice is not further required.

High resolution mapping of the putative alleles *rym4* and *rym5* was performed in two different mapping populations (*rym4*: Igri x Franka, *rym5*: Alraune x W22, each about 1000 F2). Initial physical mapping led to the construction of two contigs starting from the closest flanking markers MWG838 and RZ783 of the *rym4* mapping population. However, efforts were not successful to bridge the gap and to connect the two contigs. Alternatively, physical mapping was started from a marker AFLP1 segre-

○ T. Wicker, W. Zimmermann, D. Perovic, A.H. Paterson, M. Ganal, A. Graner and N. Stein **A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley hv-eif4e locus: Recombination, re-arrangements and repeats.** The Plant Journal (2004), in press (doi: 10.1111/j.1365-313x.2004.02285.x).

gating 0.03 cM proximal to the *rym5* locus in the increased second mapping population (1500 F2, Fig2). Until now a 400 kb BAC contig was constructed towards *rym5* by three chromosomal walking steps. This contig carries the closest proximal marker and a 250 kb cosegregating region but no distally flanking marker. Therefore, the contig does not yet bridge the locus genetically. A recent fourth walking step has likely extended the contig towards this missing marker by another 90 kb.

Three barley BACs 519, 801 (Fig 2) and 204, cosegregating with the *rym4/rym5* locus, were completely sequenced leading to a total contiguous sequence of 250 kb. The sequence of BAC 519 still contains two small gaps of GC-rich repeats, impossible to bridge by sequencing in a reasonable effort. The sequence information of these minor gaps is not relevant for the further sequence analysis. Sequencing of a fourth BAC 778 which carries the proximal flanking marker is accomplished but still exhibits 7 gaps due to a high density of highly conserved repetitive elements.

Initial sequence analysis of the contig reveals a region with very low gene density. Based on sequence similarities only two open reading frames (ORF) could be identified within a 10 kb fragment (Fig 2). This equals a density of 1 gene per 185 kb. The contig consists mainly of repetitive DNA elements with about 50% belonging to the retroelement classes BARE-1 (Ty/copia-like) and gypsy. Their linear distribution along part of the contig is visualised in Fig 2. However, a further analysis of the repetitive DNA sequences is still ongoing to determine the detailed sequence organisation. A putative orthologue of barley ORF1 is present on the rice BAC AP003448 in the same orientation as in barley. This is a confirmation of the postulated synteny between the rice and the barley genome in the *rym4/5* region. However, a homologue to ORF2 is not present in a colinear position of the target interval of rice chromosome 1L but at another position of the short arm of rice chromosome 1. Interestingly, in rice the ORF1 homologue is only approximately 40 kb away from the rice EST-marker RZ783 – a marker that was mapped 1.42 cM distal to *rym4* in the population Igri x Franka (data not shown). Besides the colinearity of ORF1 between barley and rice, so far the comparative sequence analysis of the available rice and the reported barley BAC sequences did not reveal any further sequence similarities.

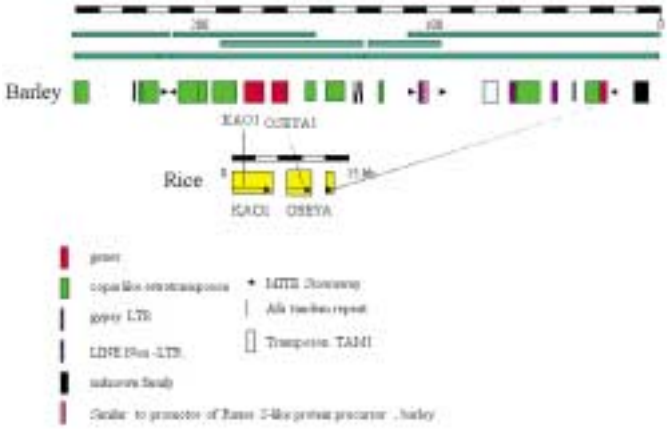


Fig 1: *Structure of orthologous regions of chromosome 7H in barley and chromosome 6 in rice. From top: Size scale in kb; display of the sequenced barley BAC contig; analysis of the barley sequence for genes and repetitive elements; display of the syntenic rice sequence with three predicted genes.*

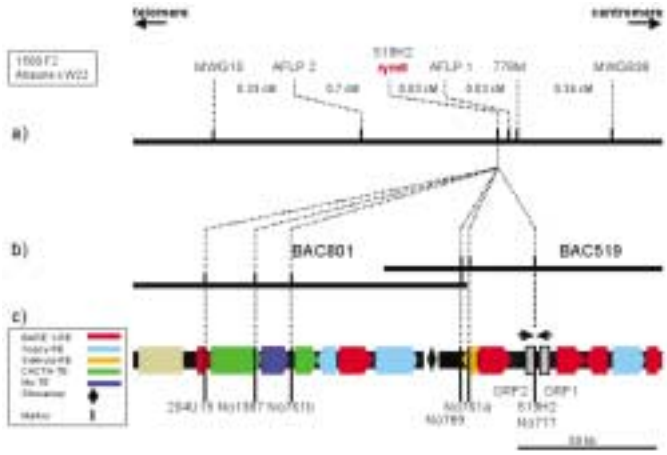


Fig 2: *Genetic and physical mapping of the rym4/5 locus. a) High resolution genetic map of rym5 in population Alraune x W122. b) Partial representation of the cosegregating barley BAC contig. Cosegregating markers are indicated by dotted lines. c) Sequence organisation of the displayed barley BACs.*

Conclusions

The results of this study demonstrate the use of synteny information for genome walking and the isolation of agronomically important genes such as disease resistance genes and QTLs exploiting the rice genome sequence for large-genome crops such as barley, wheat and rye within the Gramineae.

A Consortium-Based Functional Genomics Initiative on Plant Nonhost Disease Resistance

Barley/Arabidopsis
GABI-Nonhost

1 Jane Parker, 1 Paul Schulze-Lefert, 2 Patrick Schweizer, 3 Dierk Scheel, 3 Thorsten Nürnberger, 4 Markus Frank
1 Max-Planck Institute for Plant Breeding Research, Cologne
2 Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben
3 University of Tübingen, Center for Plant Molecular Biology, Tübingen
4 BASF Plant Science, Ludwigshafen

Nonhost resistance, one of the least understood areas in plant-microbe interactions, describes the fact that an entire plant species is normally immune to a specific parasite. Nonhost resistance most probably constitutes the major form of plant disease resistance in the field. Also, nonhost resistance is known to be considerably more durable in nature compared to strain-specific R gene dependent resistance. Therefore, understanding the molecular mechanisms of nonhost resistance may pave the way to introducing durable, race-nonspecific resistance into crop plants. The research consortium GABI-NONHOST utilizes the gene silencing technology RNA interference by double-stranded RNA (dsRNAi), together with conventional mutational tools, to overcome the problem of genetic redundancy of conventional mutant screens, i.e. to genetically disable steps in a pathway that are carried out by multiple sequence-related genes.

GABI-NONHOST focuses on the GABI model plants Arabidopsis and barley and consists of two major interconnected phases: A first phase of tool development and a second phase of parallel biological screenings and identification of gene functions. In Arabidopsis, phase 1 comprises the establishment of a deep leaf EST database, followed by the production of a transgenic dsRNAi mutant population carrying constructs that trigger dsRNAi in vivo. In barley, phase 1 comprises the establishment of a deep EST database from leaf epidermis, followed by the production of a collection of dsRNAi constructs for transient expression in single epidermal cells. Plant/construct collections will represent a large portion of the expressed unigene set of leaf (epidermal) tissues. In both systems, biological screens with foliar nonhost-pathogens will be used in phase 2 to identify silenced target genes required for nonhost resistance. In Arabidopsis and barley, whole-plant and microscopic screening (based on the interaction of single transformed epidermal cells with nonhost pathogens) systems will be employed, respectively. Proof of concept for the isolated nonhost resistance candidate genes will be achieved by their functional overexpression in model plant species Arabidopsis and barley or wheat in order to analyze their resistance behaviour to host fungi.

Gene silencing by transient single-cell dsRNAi in barley

In barley, a cDNA library derived from mRNA of peeled epidermal strips challenged with host and nonhost fungi (barley and wheat powdery mildew) was established and shotgun sequenced. The resulting set of approx. 3,600 unigenes is currently subjected to the gene function analysis by transient single-cell dsRNAi in barley epidermal cells (Schweizer *et al.* 2000). A plasmid vector (pUAMBN) for the expression of inverted repeat dsRNAi constructs in monocots has been generated containing a strong constitutive promoter upstream of two inversely localized GATEWAY® recombination cassettes. ESTs can be cloned in this vector by simple recombination in a library scale. Moreover, a high-throughput method for the simultaneous integration of 96 independent PCR products via GATEWAY® recombination in a microtiter scale was developed. The resulting dsRNAi constructs can be directly used in the transient single-cell dsRNAi approach in barley to analyze the penetration behaviour of host and nonhost fungi, as wheat and barley powdery mildew. Over the last three years, expertise has been accumulated in using single-cell transformation protocols to study host gene functions in interactions between barley and powdery mildew fungi. Recent technological improvements of gene delivery via the particle gun ('hepta adaptor', BIO-RAD) made it possible to modify the original protocols for high-throughput analysis (Schweizer *et al.* 1999), making it possible to phenotypically score within 30 working days a dsRNAi library comprising 500 genes for altered infection phenotypes. Functional analysis of barley dsRNAi constructs is carried out with the fungal host pathogen Blumeria graminis f. sp. hordei and the nonhost Blumeria graminis f. sp. tritici. We are currently trying to extend this transient gene function assay also to other nonhost taxa, such as Erysiphe pisi or Colletotrichum lagenarium.

More than 500 cDNAs expressed in the barley epidermal library have been tested yet. Interestingly, one candidate gene has recently been identified the silencing of which leads to a partial overcome of mlo-based race-unspecific powdery mildew resistance in barley. This gene is implicated in cytosceletal dynamics of the plant cell. This result provides a proof of concept for the combination of high-throughput cloning of dsRNAi with transient single-cell dsRNAi.

Heritable loss of function mutagenesis in Arabidopsis

Besides conventional EMS mutagenesis, we will create an Arabidopsis mutant population in which the vast majority of genes expressed in green leaves will be silenced. Since genes expressed in foliar tissue represent only a fraction of the entire Arabidopsis genome, this will lead to manageable numbers of dsRNAi lines that need to be generated by the consortium. EST libraries derived from mRNAs from pathogen-challenged foliar tissue have been established and are currently being screened.

We will utilize a two-step procedure for the generation of dsRNAi populations. In a first step, we will create hairpin dsRNA constructs that are driven by a constitutive promoter. In a second step, we will establish lines containing those genes under the control of an inducible promoter like the ethanol-inducible promoter for which no constitutive dsRNAi transgenic lines can be obtained (e.g. due to lethality resulting from loss of gene function). The cloning strategy will be adapted to 96-well microtiter plates to eliminate time-consuming steps like gel electrophoretic separations and restriction enzyme digestions. We will as well for this purpose modify the phage-site-specific recombination in vitro (GATEWAY®). The inverted repeat

○ A. Falk *et al.* **EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases.** Proc Natl Acad Sci U S A (1999) 96, 3292-7. ○ P. Schweizer *et al.* **A transient assay system for the functional assessment of defense-related genes in wheat.** Molecular Plant-Microbe Interactions (1999) 12, 647-654. ○ P. Schweizer *et al.* **Double-stranded RNA interferes with gene function at the single-cell level in cereals.** Plant Journal (2000) 24, 895-903.

constructs will be directly used for Agrobacterium-mediated Arabidopsis transformation. Transgenic Arabidopsis lines will be selected for each hairpin dsRNAi line. Seeds from the transgenic lines will be collected and subsequently distributed to the consortium partners for biological mutant screens.

We will screen for mutations compromising resistance to a range of nonhost downy mildew isolates (*Peronospora parasitica*) utilizing the dsRNAi mutant population generated in ecotype Col-0. Mutant plants will be tested microscopically and macroscopically for their effects on these downy mildew isolates. They will also be tested with other pathogen isolates, e.g. barley powdery mildew (*Blumeria graminis f. sp. hordei*), rice blast (*Pyricularia oryzae*) and late blight (*Phytophthora infestans*) or the NPP1 protein elicitor, used by other groups within the consortium. This will identify common components operating in nonhost resistance. Importantly, the proposed consortium-based strategy will allow cross-referencing mutant lines and establish whether genes required for nonhost resistance to one pathogen are co-utilized in resistance to other, unrelated pathogen types.

In Arabidopsis, the locus pen2 was identified that is required for efficient abortion of Blumeria graminis f. sp. hordei penetration attempts through leaf epidermal cell walls. Interestingly, this mutant also displays an increment in susceptibility to the oomycete Phytophthora infestans. The identified EMS-induced mutant gene has been isolated by a conventional map-based cloning protocol. PEN2 encodes a class family I-glucosidase. We are currently attempting to identify the substrate of this enzyme to get a better insight in the molecular basis of the penetration resistance of Arabidopsis to nonhosts.

Moreover, a null eds1 mutant line (Falk *et al.* 1999) implicated in nonhost resistance to oomycetes will be re-mutagenised. M2 plants from this population will be screened for further increases in plant susceptibility to the Brassica oleracea-infecting isolate (P-006) of Peronospora parasitica. In this way it should become possible to peel off additional layers of resistance to this pathogen. New loci will be characterized phenotypically, assessed for dominance/recessiveness and mapped by crossing to a polymorphic eds1 mutant line.

Proof of concept and exploitation

We will achieve proof of concept for the isolated nonhost resistance candidate genes by the functional overexpression of these genes in a dicot (Arabidopsis) and in monocot (barley, wheat) model plant species and the evaluation of increased resistance to host pathogens belonging to commercially important taxa. In the context of both GABI-NONHOST and in subsequent in-house activities, BASF Plant Science will work on a proof of concept for the isolated nonhost resistance candidate genes by their functional expression in a dicot (Arabidopsis) model plant and in monocot (barley, wheat) crop plant species followed by the evaluation of an increased resistance against biotrophic and necrotrophic host pathogens. If a proof of concept in these plant species can be achieved, BASF Plant Science will transfer the candidate nonhost resistance genes to economically important crop plant species for a further product development. This project has the potential to yield new crop plants carrying genes for durable resistance against host pathogens, besides providing a general resource for other research programs on green plant tissue.

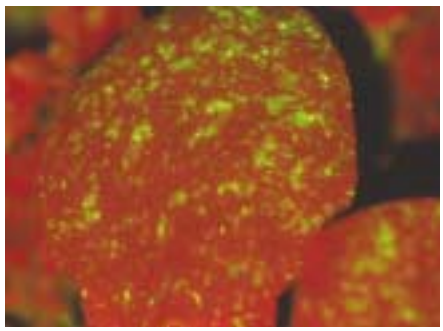


Fig 1a+b: The pen2 mutant in Arabidopsis displays an increment in susceptibility to the nonhost barley powdery mildew *(Blumeria graminis f. sp. hordei). The left picture shows a leaf of Col-0 inoculated with conidiospores of barley powdery mildew under fluorescence light. In contrast, the pen2 mutant (right picture) is characterized by single-cell hypersensitive reactions (HR) in response to fungal inoculation as indicated by the autofluorescence in the epidermal cells. The HR is the consequence of a successful penetration of the epidermis cells which does only very rarely occur in wildtype. (Fig Volker Lipka, MPIZ Cologne)*

Fig 2: Functional genomics of plant-microbe interactions on the single-cell level in barley leaves. *Barley powdery mildew (Blumeria graminis f. sp. hordei) penetrating a barley epidermis cell stained by a GUS reporter construct. Frequecny of the formation of finger-like fungal haustoria can easily be pursued in the microscope. The effect of dsRNAi constructs bombarded in the barley epidermal cells on the penetration efficiency of the fungus will be determined. (Fig Patrick Schweizer, IPK Gartersleben)*

Functional Analysis of the Barley Genome: Functional Characterization of Agronomically Relevant Genes and their Use to Improve Disease Resistance to the Genus Fusarium

Barley
GABI Agrotec

1 Carin Jansen, 1 Sibylle von Rüden, 1 Rebekko Fensch, 1 Karl-Heinz Kogel, 2 Sophia Biemelt, 2 Jochen Kumlehn, 2 Uwe Sonnewald, 3 Angelika Felk, 3 Simon Henning, 3 Joe-Gmem Ekanem, 3 Wilhelm Schäfer
1 Justus-Liebig-University of Gießen, Institute of Phytopathology and Applied Zoology
2 Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben
3 University Hamburg, Institute of General Botany (AMPIII)

Aim of the project

The project "Functional analysis of the barley genome: Functional characterization of agronomically relevant genes and their use to improve disease resistance to the genus Fusarium" aims at the identification, functional characterization and agronomical application of barley genes that strengthen the resistance against necrotrophic fungi of the genus Fusarium. These genes should be isolated from barley roots, leaves, and spikes.

Research plan

1) Infection of differentially susceptible barley accessions with Fusarium at roots and spikes.
2) Isolation of mRNA from infected plant material.
3) Generation of cDNA libraries and partially normalisation.
4) Sequencing of up to 10,000 ESTs.
5) Identification of Fusarium-induced genes in roots, leaves, and spikes of barley by generation of expression profiles via cDNA arrays.
6) Transient functional analysis of induced genes by RNAinterference or over-expression.
7) Generation and validation of stably transformed plants.

Current state

The success of this project depends mainly on the availability of a sufficient number of defence related candidate genes. For this purpose extensive epidermis specific cDNA libraries from barley plants that were treated with different biotic and abiotic stimuli (e.g. the chemical resistance inducer BION (Beßer *et al.* 2001), crude toxin preparation of Bipolaris sorokiniana (Kumar *et al.* 2002), and conidia of Blumeria graminis f. sp. tritici (non-host) were generated. Before sequencing two of the cDNA libraries were normalised to accumulate rare transcripts. The cDNA fragments were ligated subsequently into suitable vectors and ca. 10,000 clones were filed in 96-well plates. Approximately 2,000 clones were sequenced and the sequences were blasted against the HUSAR (Heidelberg Unix Sequence Analysis Resources) database (http://genius.embnet.dkfz-heidelberg.de/menu/w2h/w2hdkfz/index.shtml). The cDNA fragments of nearly 3,000 clones were amplified by PCR and spotted on nylon membranes. Currently hybridisations of these filters with probes from RNA of Fusarium infected and non-infected (control) barley leaves are performed. Figure 1 shows a small cut out from the first hybridisation experiment.

Besides this the expression profiles of these epidermal genes are also compiled after inoculation with Bipolaris sorokiniana and Blumeria graminis.

Additionally to the epidermis specific libraries, three root specific libraries were generated. For this more than 15 barley accessions were tested for their susceptibility against Fusarium culmorum in the root. An inoculation method that was established to infect barley roots with *Bipolaris sorokiniana* (1) was used to find differences in the severity of Fusarium-infection on the roots. The roots of 3-day-old barley seedlings were submerged in a Fusarium culmorum spore-suspension (10,000 spores/mL), wrapped in filter paper and incubated for 7 days (22°C, 16 h light period, 100% rel. humidity) in tubes filled with 10 ml of water. One week after inoculation the extent of disease symptoms, indicated by a brown discoloration, was measured as percentage of infected roots, and the reduction of root growth caused by Fusarium infection was determined. With this method it was possible to identify accessions which were more resistant to Fusarium culmorum than others. Finally the cultivars Nikel and Carola were chosen as more resistant and more susceptible, respectively. Roots from 6-day-old seedlings of both accessions were inoculated with a suspension of *Fusarium culmorum* spores (approx. 300,000 spores/mL) or 0.02% Tween-water as uninfected control and harvested 6, 24, 48, 72 and 96 hpi. One cDNA-library was established from uninfected roots of both cultivars, the two others from Fusarium-infected roots (6-96 hpi) of Nikel and Carola, respectively.

For the generation of the three cDNA libraries the Creator SMARTTM cDNA Library Construction Kit (BD Biosciences) was employed to obtain full length clones. The success of the cloning was determined by counting the plaque forming units (108 – 109 Pfu/mL). The mean size of the fragments assigned by PCR was approx. 1,500 bp. Nearly 14,000 clones were filed in 96-well plates, 2,500 of them were sequenced and the sequences were blasted against the HUSAR (Heidelberg Unix Sequence Analysis Resources) database (http://genius.embnet.dkfz-heidelberg.de/menu/w2h/ w2hdkfz/index.shtml). 60 cDNA arrays with 4,608 clones were spotted and are currently hybridised with different probes to identify Fusarium-induced genes in the roots.

For the generation of cDNA libraries from Fusarium-infected ears extensive tests including 14 accessions were conducted to find differences in the susceptibility to Fusarium graminearum in barley spikes. Besides Chevron and Hazera, 12 genotypes from the Institute of Agrobiotechnology (Tulln, Austria) and The International Maize and Wheat Improvement Center (CYMMIT, Mexico) were tested for their resistance status to Fusarium by inoculation of each spikeletts with a drop of conidia. 4 dpi (days post inoculation) the ears were harvested to isolate total RNA. The course and severity of Fusarium infection could be monitored via RT-PCR using a DNA oligo

○ E. Liljeroth, H.-B. Jansson, W. Schäfer **Transformation of Bipolaris sorokiniana with the GUS gene and use of studying fungal colonization of barley roots.** Phytopathology (1993) 83(12), 1484-1489. ○ K. Beßer, B. Jarosch, G. Langen, K.H. Kogel **Expression analysis of genes induced in barley after chemical activation reveals distinct disease resistance pathways.** Molecular Plant Pathology (2000) 1, 277-286. ○ J. Kumar, P. Schäfer, R. Hückelhoven, G. Langen, H. Baltruschat, E. Stein, S. Nagarajan, K. H. Kogel **Pathogen Profile: Bipolaris sorokiniana, a cereal pathogen of global concern. Cytological and molecular approaches towards better control.** Molecular Plant Pathology (2002) 3, 185-195.

primer set designed against the beta-tubulin gene of *F. graminearum* that gives no amplification product on the barley genome. Currently cDNA libraries representing the infected transcriptome of different barley cultivars are being established.

To facilitate the cytological analysis of the barley-Fusarium interaction on spikes, leaves and roots, a *Fusarium graminearum* isolate was transformed stably with the gfp gene that codes for the green fluorescent protein. The gene is driven by a truncated promoter of the iso-citrate ligase from *Neurospora crassa* and is transcribed in the Fusarium cells. Figure 2 shows micrographs from barley leaves 48 and 96 hpi, respectively, inoculated with the transformed *F. graminearum isolate*.

Because there were already candidate genes identified at the onset of the project that could probably serve as resistance factors to necrotrophic pathogens including Fusarium, the generation of stably transformed barley plants was set in immediately at the beginning of the project work. The Bci7 gene and the Bci9 gene coding for a proteinase inhibitor and an apyrase, respectively, were cloned between an ubiquitin promotor and a nos-terminator. Employing the established microspore transformation system the Ubi-Bci9 construct is currently transformed into barley plants.

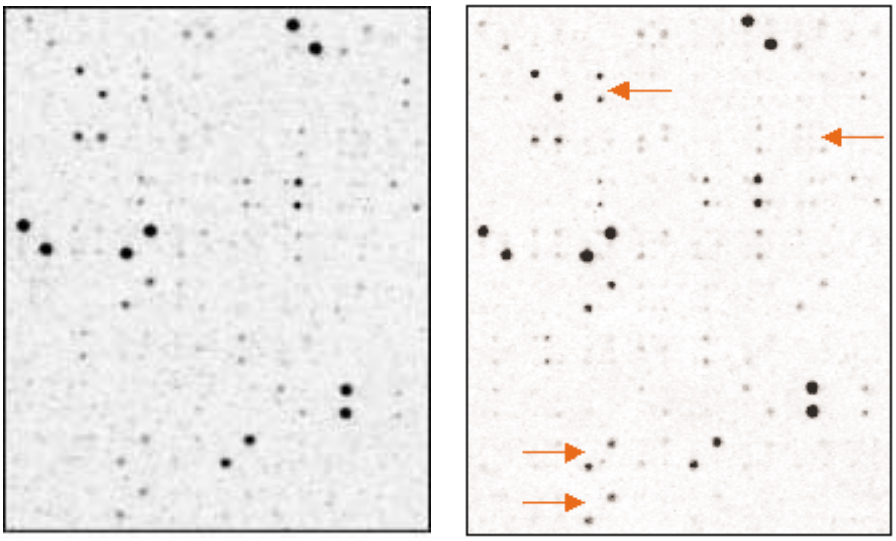


Figure 1: *Preliminary results of a cDNA-array hybridisation. Approx. 1,500 clones of cDNA library obtained from epidermal layer of barley leaves treated with a crude toxin preparation from culture filtrate of Bipolaris sorokiniana were spotted on filters. A: Cut out of a filter hybridised with a probe from non-infected barley leaves (control). B: Cut out of a filter hybridised with a probe from Fusarium-infected leaves (24 hours post inoculation, hpi). The red arrows indicate the double spotted clones which are seemingly differentially expressed in control and Fusarium-infected leaves.*

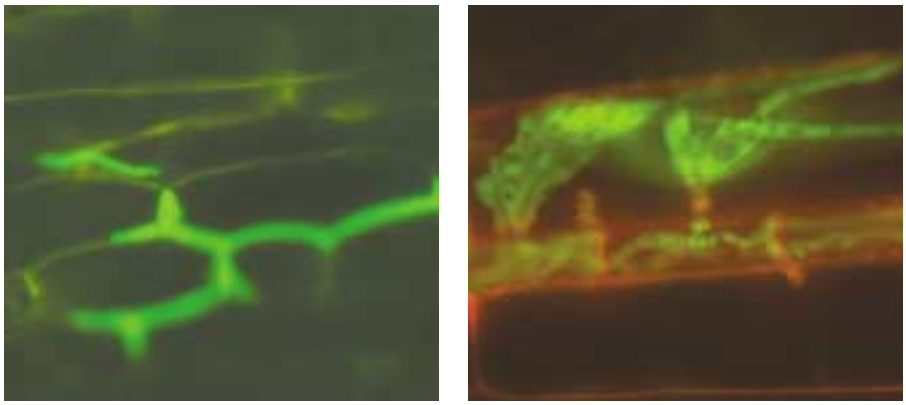


Figure 2: *Micrographs from barley leaves infected with a F. graminearum isolate transformed with gfp. A: Fusarium hyphae growing along the anticline cell walls of the epidermis (24 hpi). B: Fusarium hyphae growing inside dead epidermal cells.*

Assessment and Functional Analysis of Agronomically Important Genome Regions in Hybrid Rye

GABI Rye Overview

In hybrid rye breeding and research, Germany holds a leading position worldwide. In 1998, the German rye acreage amounted to 990,000 hectares. About 60% of this area were planted to hybrid varieties. Although acreages of barley and wheat are considerably higher, the hybrid rye seed business is similarly profitable due to the need of buying new seed every year and a more favourable license situation for hybrid than for line varieties. Rye considerably outyields barley and wheat on nutrient-poor, dry soils and in water resources areas, where nitrogen fertilizaton is limited by legislation.

Genome analysis is considered particularly rewarding for rye, because (1) it is an outbreeding diploid species with a high degree of polymorphism, (2) it is a parent species for triticale and (3) it exhibits specific traits, including biotic and abiotic stress tolerances, that cannot be assessed in related gramineous model organisms. Furthermore, rye represents the only hybrid crop among the small grains and could serve as a model for breeding hybrids in wheat and triticale in the future.

Disease resistance remains to be a major issue in rye breeding programes, especially in hybrid breeding with a narrower genetic basis. In addition, rye has been a valuable genetic resource in respect to major disease resistance genes for the breeding of other cereals, i.e, wheat and triticale. Genes for resistance to leaf rust, stem rust, yellow rust, powdery mildew as well as aphids were transferred from rye into wheat cultivars. In contrast to the large and highly complex allopolyploid genome of the latter species, genetic analysis and mapping of disease resistance genes is relatively straightforward for the diploid and highly polymorphic rye. The size and highly repetitive nature of the rye genome, however, renders it recalcitrant in respect to position-based gene isolation strategies. Thus, identification of resistance genes in rye and their characterization at the molecular level could greatly benefit from the map and sequence data for similar genes which were isolated from other species, e.g., barley, rice, and wheat. Analysis of the synteny relationships between resistance genes in rye and those found in other plant species (Research Area 1) will contribute to the question to what extent the anticipated potential of synteny-based gene discovery can be exploited in selected cereals.

The under-utilized weedy rye species of the Near East as well as European landraces and rye populations may contain alleles for agronomically highly valuable qualitative and quantitative traits that are missing in our adapted breeding materials (Research area 2). To design highly efficient introgression protocols two marker-based procedures will be studied and complemented by computer simulations (subproject 2a). In subproject 2b+2c two backcross breeding populations with exotic donor parents were developed by practical breeding aspects.

For the estimation of genetic effects of introgressed chromosomal segments from genetic resources as well as for efficient marker-assisted selection in breeding programmes highly polymorphic, codominantly inherited and sequence-characterized markers are the tools of choice. Thus, the development and mapping of a sufficient number of microsatellite markers in rye (Research area 2, subproject 1) will considerably improve the efficiency of introgression approaches exemplified in subproject 2a+2b+2c.

Structural and Functional Comparison of Disease Resistance Genes in Rye and Other Plants

GABI Rye

Thomas Bringezu, Bernd Hackauf, Bodo Linz, Peter Wehling
Federal Center for Breeding Research on Cultivated Plants
Institute of Agricultural Crops, Groß Lüsewitz

Introduction

Improving crops with enhanced disease resistance is a major objective in plant breeding. Two major classes of genes contribute to the plant disease resistance pathway: resistance (R) genes involved in the recognition of, as well as interaction with, elicitor molecules from pathogens, and genes involved in defense responses. In recent years R genes have been cloned and characterized from a variety of plants including both mono- and dicotyledonous species. This growing list of R genes has revealed that a majority belongs to the NBS/LRR class of genes which are characterized by a nucleotide-binding site (NBS) near the N-terminus and a leucine-rich repeat (LRR) region near the C-terminus.

The overall sequence homology among R genes of the NBS/LRR family is low, but the NBS contains several sequence motifs, P-loop, kinase2, kinase3 and GLPLAL, that are highly conserved among R genes even in distantly related plants (Leister *et al.* 1998). This conservation between different NBS/LRR resistance genes enables PCR-based strategies to isolate and clone R gene family members or analogs using degenerated or specific primers for the conserved regions.

Objectives

Identifying and mapping of resistance gene analogous (RGA) sequences in rye and other grasses could prove to be important in the identification of actual R genes as well as genome areas where R genes may reside. The identified RGAs will be a valuable resource as markers for molecular breeding programs as well as references for the identification of their orthologous genes in other cereal crops.

In an initial attempt we have analyzed populations segregating for leaf rust and powdery mildew in rye (*Secale cereale*), for leaf rust, powdery mildew and mild mosaic virus resistance in barley (*Hordeum vulgare*), and for crown rust resistance in ryegrass (*Lolium perenne*) using oligonucleotide primer pairs derived from different DNA sequence motifs of resistance- and defense-related genes. These primers identified 12 markers linked to resistance genes, e.g. effective towards BaMMV in barley (Fig 1), in the selected mapping populations. Most of the linked markers like those mapped in relation to crown-rust resistance of rye grass (Fig 2A) were based on NBS/LRR motifs. As a consequence, research was focussed on members of the NBS/LRR gene family in relation to defined leaf rust and powdery mildew resistance genes in rye.

Sequence tagged site (STS) primers were developed for 31 NBS/LRR genes from wheat, barley, rice and maize, which have previously been mapped on the short arm homeologous group 1 chromosomes in wheat and barley. Subgenomic fragments of the expected size could be amplified from the rye genome with 20 primer pairs. Within the selected mapping populations 50% of these STS revealed polymorphism. Their localization on the short arm of chromosome 1R and the origin of the sequence-specific primers as well as the observed fragment size indicate that orthologous RGA sites have been tagged in rye. As has been reported for other small grain cereals, our mapping data demonstrate a cluster of NBS/LRR genes located on rye chromosome 1RS (Fig 2B).

In a more general approach degenerated primers were used, which are designed from the conserved amino acids in the common structural NBS domains of cloned resistance genes. As in other grass species (Leister *et al.* 1998, Wang and Xiao 2002), these primers were successfully applied to amplify multiple resistance gene analogs in rye (Fig 1B). Actually, one RGA (IAC24) located in a gene-cluster on chromosome 1RS could be identified loosely linked to a mildew resistance gene locus (Fig 2B). Bulked segregant analysis of 12 populations is in progress and will demonstrate, if RGA-fragments can be identified closely linked to defined leaf rust and powdery mildew resistance genes in rye.

Conclusions

To date, the analysis of the polymorphic amplicons in the included mapping popu-lations yielded 34 markers, 5 cosegre-gating with resistance genes, 5 tightly linked markers (1.8-4.1 cM) and 14 moderately to loosely linked markers (6.7-27.5 cM). The remaining 10 poly-mor-phic markers mapped to the same chromosome as the respective resistance genes. These RGA-based sequence tagged site (STS) markers were developed for 2 leaf rust and 1 powdery mildew resistan-ces in rye, 1 powdery mildew resistance and 1 mild mosaic virus resistance (intro-gressed from *H. bulbosum*) in barley and 2 crown rust resistances in ryegrass. As a conclusion, information of resistance gene analogs represents a valuable resource for the develoopement of molecular tags for R genes in rye, barley, and ryegrass.

○ D. Leister, J. Kurth, D. A. Laurie *et al.* **Rapid reorganization of resistance gene homologues in cereal genomes.** Proc. Natl. Acad. Sci. (1998) 95, 370-375.
○ S. Wang and X. Xiao **Isolation and linkage mapping of disease-resistance-like sequences from various rice cultivars, containing different recognition specificities.** Plant Breed. (2002) 121, 203-209.

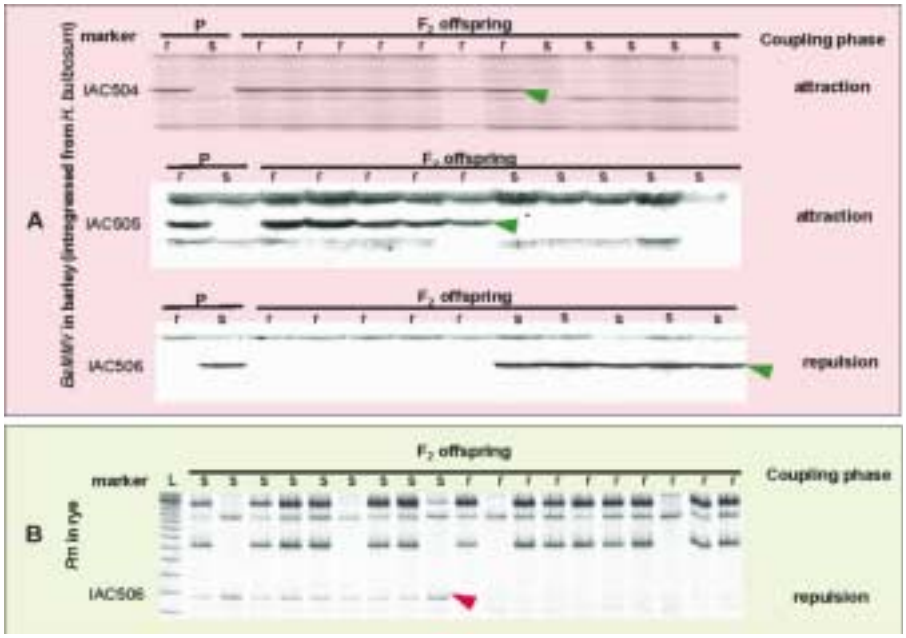


Fig 1: RGA-based polymorphism in barley and rye. *(A) STS markers linked to barley mild mosaic virus (BaMMV) resistance in barley (green arrows). (B) Polymorphic PCR amplification products of genomic DNA from rye using degenerated primers. Amplicons linked to powdery mildew (Pm) resistance are indicated by red arrows. s: susceptible and r: resistant individual plants; P: parents; L: 100-bp DNA size ladder.*

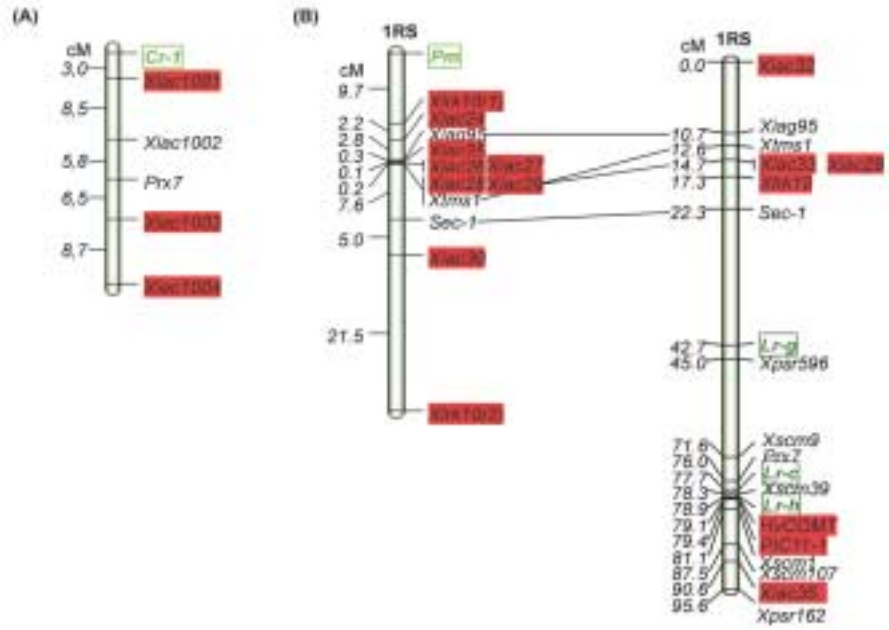


Fig 2: Mapping of RGA-based STS-markers relative to resistance genes. *(A) Crown-rust resistance gene Cr1 in rye grass (Lolium perenne). (B) Powdery mildew gene Pm and leaf-rust resistance genes Lr-c, Lr-g, and Lr-h (integrated linkage map) in rye (Secale cereale) on chromosome 1RS. Resistance genes are indicated by green and RGAs by red colour. The integrated linkage map was calculated from linkage data of three rye mapping populations. Common marker positions between the two maps in the distal part of 1RS are indicated by lines.*

Assessment of Agronomically Important Genome Regions from Exotic Germplasm in Hybrid Rye

GABI Rye

Subproject 1: 1 Bernd Hackauf, 2 Stephen Rudd, 1 Peter Wehling
1 BAZ, Institute of Agricultural Crops, Groß Lüsewitz
2 GSF Institute for Bioinformatics (MIPS), Neuherberg
Subproject 2a: Zoran Susic, Thomas Miedaner, Hartwig H. Geiger
University of Hohenheim, Stuttgart
Subprojects 2b, 2c: 3 Heinrich Wortmann, 4 Jörg Schondelmaier, 5 Peer Wilde, 5 Viktor Korzun
3 Hybro GmbH & CoKG, Bad Schönborn
4 Saaten-Union Biotech-Laboratory, Leopoldshöhe
5 Lochow-Petkus GmbH, Bergen

Introduction

Broadening the genetic base of elite breeding materials by introgressing exotic germplasm requires techniques which avoid reduction of productivity and which minimise interfering genetic interactions between recipient and donor. Both aims appear achievable by restricting the introgression to one or a few short donor chromosome (DC) segments. Molecular markers allow to considerably speed up the introgression process and to characterize the transferred segments.

Introgression libraries are considered a powerful and highly efficient tool to characterize and exploit genetic resources in rye (Miedaner *et al.* 2002). For developing introgression libraries microsatellites or simple sequence repeats (SSR) are particularly well suited because they are highly polymorphic, co-dominant and evenly distributed across the genome. EST-derived microsatellites combine the advantages of representing expressed genes and being solely based on PCR-methodology without further processing. Furthermore, the conserved nature of orthologous sequences between closely related species results in a higher level of trans-ferability of EST-derived SSRs.

Objectives

The objective of Subproject 1 was to develop and map at least 200 rye SSR markers to facilitate the establishment of rye introgression libraries in Subproject 2a and contribute to the characterization of agronomically superior introgression lines in Sub-projects 2b and 2c. The introgression libraries should be comprised of lines containing one to three DC segments each and, taken together, cover most of the donor genome. For comparison, phenotypically superior introgression lines should be identified in field trials and characterized by molecular markers a posteriori.

Development of SSR markers

In Subproject 1 more than 8000 rye ESTs were processed using a state-of-the-art EST clustering pipeline with novel algorithms for the characterisation of microsatellite-like features. The initial dataset of short, low-quality, single-pass sequences was refined down to a set of 5516 unique clusters, which could be assigned to genes with known or putative cellular function in the majority of cases. Based on the definition of an SSR with at least 12 bp, approximately 1500 perfect, imperfect, interrupted or compound di-, tri-, tetra- or pentanucleotide repeats could be identified within these EST clusters. Primers were derived from a subset of 276 microsatellite-bearing ESTs. PCR assays could be established for 176 SSRs. Together with the 157 recently described Secale cereale microsatellite (SCM) markers (Hackauf and Wehling 2002a) more than 330 SCM are available now, which allow for high-quality readings in the majority of cases and, thus, provide a valuable tool for genetic and applied purposes.

To support marker-assisted selection of DC segments in Sub-project 2a, the SCM were applied to the BC1 populations 9952, 9953 and 9954. Polymorphism could be observed for approximately 47% of the analyzed SCMs. A similarly high degree (52%) of polymorphic SCMs could be identified for the F2 mapping population IAC-1300. The selected mapping populations allow to map more than 200 SCMs till the end of this project. The established SCM markers appear to be quite evenly distributed throughout the rye genome, as can be deduced from initial mapping data (Hackauf and Wehling 2002b). Their integration into an AFLP framework map together with genomic rye SSR anchor markers (Saal and Wricke 1999; Korzun unpublished) allowed to address the anonymous linkage groups to the seven rye chromosomes and contributed to the development of introgression libraries in Subproject 2a (fig 1).

Establishment of introgression libraries by marker-aided backcrossing

In Subproject 2a, two rye introgression libraries were developed from F1 plants of crosses between an elite inbred line (L2053-N) and an exotic donor population (Altevogt 14160). The backcrossing procedure was initiated from three F1 plants. The first F1 plant gave rise to library F and the remainder two to a combined library G. AFLP and SSR markers were used to establish genetic maps in backcross generation 1 (BC1). The AFLP marker assays were conducted by Keygene N.V., Wageningen, The Netherlands. SSR data were assessed by project partners (BAZ, Lochow-Petkus GmbH, and SAATEN-UNION Resistance Laboratory GmbH). The SSR markers were generally very useful in bridging the gaps between AFLP marker positions and in extending the maps towards the distal regions of the chromosomes. The mapping resolutions were 3.48 cM in population F (196 markers, 683 cM total map length) and 2.74 cM in population G (250 markers on 685cM total map length).

The markers in the two BC1 maps were used to monitor DC segments and to select desired genotypes when advancing the material up to generation BC2S3. The number and chromosomal localisation of DC segments, as well as the percentage of recurrent parent genome served as selection criteria. The established introgression libraries comprise 80 BC2S3 near isogenic lines (40 per library, tab 1), jointly covering approximately 70% of the exotic donor genome (fig2).

Most of the introgression lines contain one to three homozygous DC

○ Hackauf B and Wehling P **Identification of microsatellite polymorphisms in an expressed portion of the rye genome.** Plant Breed (2002a) 121, 17-25. ○ Hackauf B and Wehling P **Development of microsatellite markers in rye: map construction.** Proc EUCARPIA Rye Meeting Radzików, Poland, (2002b), 333-340. ○ Miedaner T, Susic Z, and Geiger HH **Erweiterung der Diversität von selbstfertilen Roggenpopulationen durch Introgression genetischer Ressourcen.** Vortr. Pflanzenzüchtg. (2002) 54, 81-88. ○ Saal B and Wricke G **Development of simple sequence repeat markers in rye (*Secale cereale L.*).** Genome (1999) 42, 964-972. ○ M. Frisch, M. Bohn and A.E. Melchinger **PLABSIM: Software for simulation of marker-assisted backrossing.** Journal of Heredity (2000) 91, 86-87.

segments. Their length ranged from 1.5 to 71.3 cM averaging to 17.2 cM, while percentage of the pecurrent parent genome in established introgression lines reached 89-98% in library F and 91-99% in library G (tab 1).

Most of the introgression lines show overlap with others for one or more DC segments. This allows a more accurate estimation of DC effects and provides a certain risk assurance against disturbing effects of small DC segments that possibly remained undetected during the introgression procedure.

Development of introgression lines based on phenotypic selection

In Subprojects 2b and 2c two large BC2 populations were derived in greenhouse programs from crosses between the elite inbred line L2053-N as recurrent parent and self-incompatible population NEM1 (a Russian breeding strain) as donor (Subproject 2b) and between the elite line Lo57-N as recurrent parent and GLORIA (an old landvariety of middle Europe) as donor (Subproject 2c). In a field experiment at two sites each, superior BC2S1 lines were selected based on highly heritable and agronomically important traits. About 1000 BC2S1 lines per population were grown in micro-plots. The BC2S1 lines belonged to 50 BC2 families. Selection was practices within families only to preserve genetic variation. A remarkable range of variation between BC2 families was found for each trait (tab 2).

For most traits the best family clearly outperformed the recurrent parent. In the 2001/2002 season, BC2S3 progenies of selected BC2S1 lines were testcrossed with two cytoplasmic male-sterile (CMS) single crosses and multiplied in isolation cages. Sufficient seed was obtained for evaluating the combining ability and line performance in extensive field trials.

Theoretical work

Computer simulations (Plabsim, version 2, Frisch *et al.* 2000) were conducted to evaluate and optimize alternative strategies for developing an introgression library in rye (7 chromosome pairs, assumed genome size 665 cM). Six procedures differing in the number of backcross (BC) and selfing (S) generations were analysed. The BC3S1 procedure proved most recommendable. It allows to achieve close to 100% donor genome coverage with moderate population sizes in the individual generations and an acceptable total number of marker data points. Longer procedures were somewhat more cost efficient but too time demanding. The reverse was for shorter procedures. An optimal allocation of resources was achieved by starting with a small BC1 population and stepwise increasing the progeny size of the introgression lines in the succeeding generations.

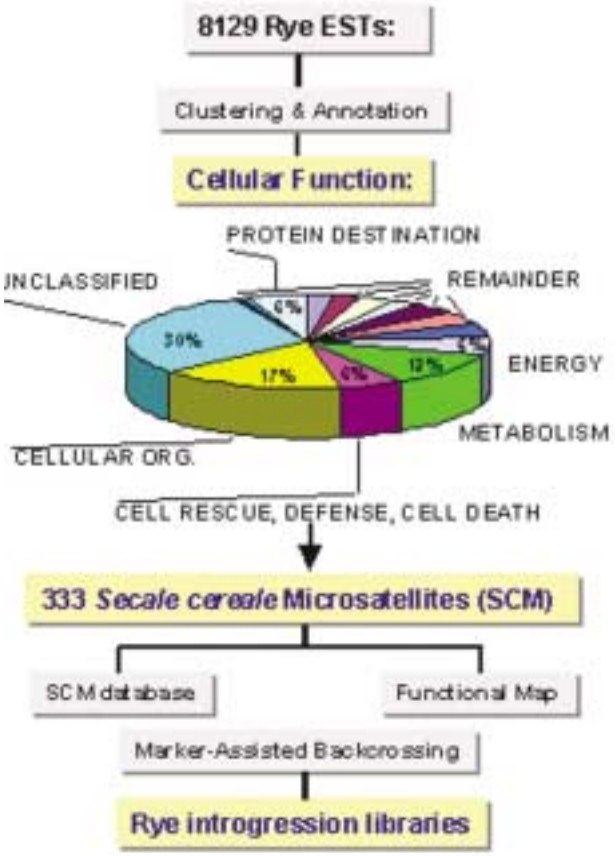


Fig 1: *Unlocking regions of the rye genome bearing agronomically important genes. Rye ESTs were exploited for the development of microsatellite markers, which contributed to establish the first introgression libraries in rye.*

Outlook

In the second phase (GABI 2) of the Subprojects 2a-c the introgression lines will be evaluated for combining ability and line performance in multi-environment field trials. The most promising introgression lines will be analysed by high-resolution mapping aiming at the isolation and functional characterization of economically important genes. Industrial partners (Hybro GmbH & CoKG, Lochow-Petkus GmbH) will directly use those lines in their hybrid rye breeding programs.

Conclusions

Marker-aided backcrossing resulted in the first two rye introgression libraries covering approximately 70% of the total donor genome (Subproject 2a). New SSR markers developed in Subproject 1 facilitated the introgression procedure in Subproject 2a and will contribute to the characterization of agronomically superior introgression lines developed in Sub-projects 2b and 2c. Results provide excellent opportunities to identify and characterize exotic DC segments and to proceed towards functional genomics in GABI 2.

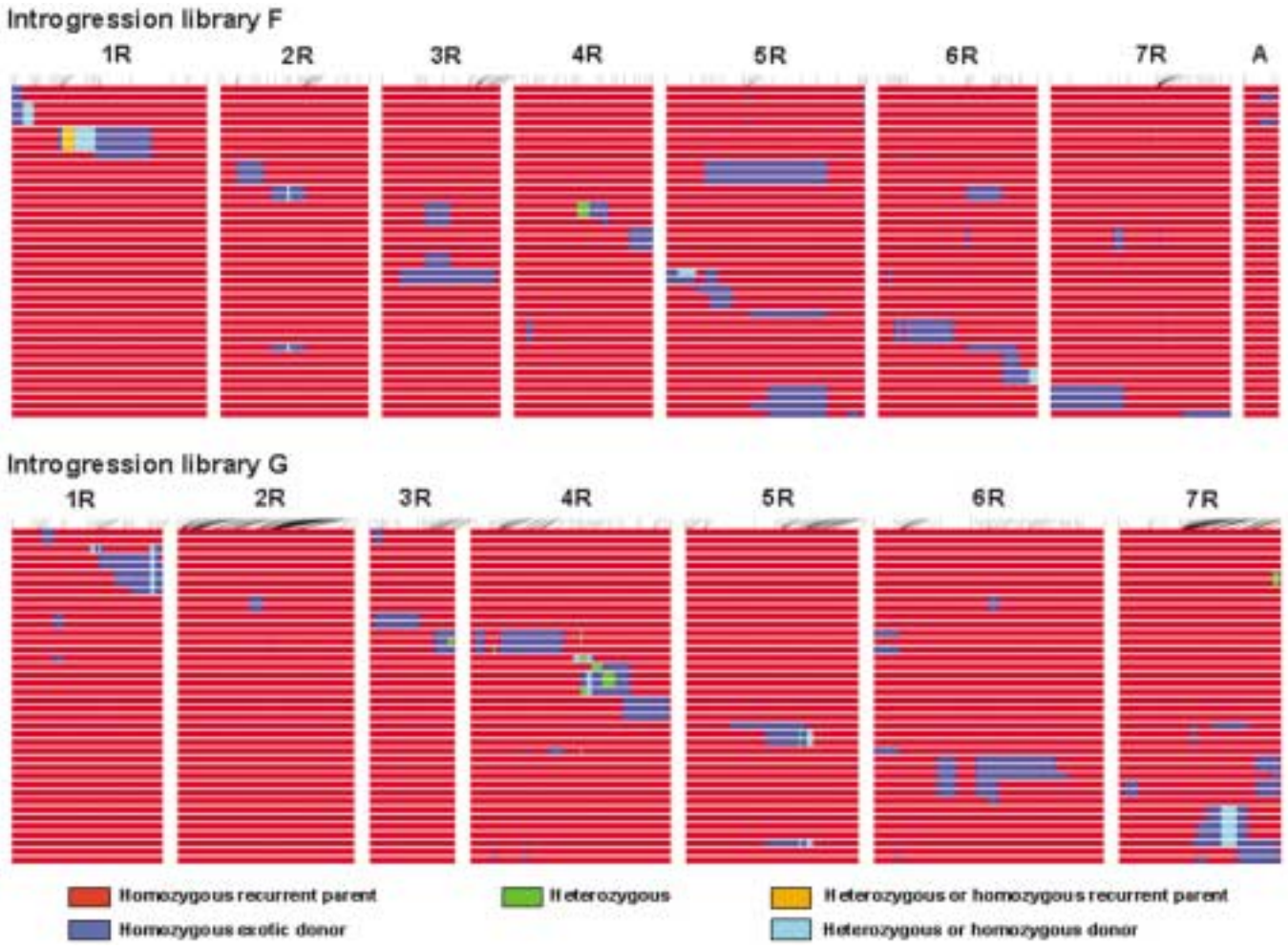


Fig 2: *Graphical genotypes of the two rye introgression libraries.*

Population	Generation	Number of plants		Number of primer combinations[§] used		Number of DC segments[§†]	% of the recurrent parent genome[†]
		Analysed	Selected	AFLP	SSR		
F	BC_1	68	9	13	-	6-9	76-85
	BC_2	154	19	5-8	-	1-4	82-95
	BC_2S_1	190	17	1-5	39	1-3	88-98
	BC_2S_2	256	40	1	20	1-3	89-98
G	BC_1	69	9	10	-	4-7	77-86
	BC_2	196	18	5-7	-	2-5	82-94
	BC_2S_1	133	22	2-5	36	1-3	88-98
	BC_2S_2	267	40	1-2	32	1-3	91-99

§ Per progeny.
† In the selected fraction

Tab 1: *General survey of the dimensioning and results of Subproject 2a.*

Trait	Recurrent parent	BC2S1-family	
		worst	best
Plant height [cm]	106	120	105
Heading date [1-9] §	3.9	3.4	4.6
Head blight [1-9] †	7.3	7.8	5.8
Falling number[sec]	175	156	229
Thousand-grain weight [g]	19	18	24

§ 1 = early, 9 = late.
† 1 = disease free, 9 = fully diseased

Tab 2: *Performance of the recurrent parent and the worst and best BC_2 family in Subproject 2b (micro-plots, 2 locations in 2001).*

AB-QTL analysis in wheat: Detection of favorable genes for quality and yield traits from wheat wild species and their introgression into the elite germplasm of cultivated wheat by means of DNA markers

Wheat
AB-QTL

4 Klaus Pillen, 1 Antje Kunert, 2 Anne Schubert, 3 Jens Léon,
1 Dept. of Crop Science and Plant Breeding, University of Bonn
Co-operating breeding companies
1 Deutsche Saatveredelung Lippstadt-Bremen GmbH, Lippstadt
2 Fr. Strube Saatzucht KG, Söllingen
3 Limagrain-Nickerson GmbH, Edemissen
4 Lochow-Petkus GmbH, Bergen
5 Saatzucht Josef Breun GdbR, Herzogenaurach
6 Saatzucht Schweiger OHG, Moosburg
7 W. von Borries Eckendorf GmbH & Co., Leopoldshöhe

Project aims

The project intends to use the AB-QTL (Advanced Backcross Quantitative Trait Locus) strategy developed by Tanksley & Nelson (1996) in order to improve our modern wheat (Triticum aestivum spp. aestivum) varieties. The particular aims of this AB-QTL project are: (1) The localization of QTLs in backcrosses between elite wheat varieties and exotic wheat accessions, (2) the detection of favorable, trait improving exotic alleles at these QTLs, (3) the production of near isogenic lines (NILs) containing favorable exotic QTL alleles and (4) the increase of the genetic diversity present in our modern wheat varieties through the introgression of favorable exotic QTL alleles.

Project description

After the successful application of the AB-QTL strategy in small grain species like rice and barley (Xiao *et al.* 1996, Pillen *et al.* 2003, Pillen *et al.* 2004 and Piepho & Pillen 2004), the present AB-QTL project uses hexaploid wheat as the recipient of exotic donor germplasm. For this, synthetic wheat accessions serve as the donor of wild species alleles for the wheat A, B and D genomes. The synthetic wheats have been generated by hybridizations of *T. turgidum ssp. dicoccoides* with *T. tauschii* (Lange & Jochemsen 1992). The AB-QTL study is conducted with four advanced backcross populations (Fig 1). Two spring wheat populations have been backcrossed and selfed until the BC2F4 generation was reached and two winter wheat populations were likewise treated up to the BC2F3 level. The wheat project is similar in regard to, both, the strategy and the goals of the barley project 0312278A which is also currently conducted in our laboratory.

The wheat project started in August 2002. It is scheduled for a period of four years and divided into two project phases (see Fig 2). During the first project phase (August 2002 through July 2003), the AB lines will be field propagated and, simultaneously, genotyped with SSR (simple sequence repeat) markers. During the second project phase (August 2003 through July 2006), the AB lines will be phenotyped and finally used for QTL detection. During the seasons 2004 and 2005, the AB lines will be field evaluated for agronomic and quality related traits as well as for pathogen resistances at eight locations in Germany. These locations include our own experimental station Dikopshof as well as the breeding stations of the seven co-operating breeding companies which are listed above. The evaluation of the AB lines cover the agronomic traits flowering time, plant height, lodging, maturation, spikes per square meter, grains per spike, thousand grain weight, total yield and harvest index. All agronomic traits will be measured under two levels of nitrogen (N) fertilization. The QTL analysis under the standard N-fertilization regime will assist in unraveling such exotic QTL alleles which can improve a trait under standard conditions of cultivation. Comparing the performances of the AB lines under the low N-fertilization level with the standard N-level will then be used in order to localize exotic QTL alleles which exhibit a high nitrogen efficiency. The latter QTLs can be used to breed low input wheat varieties which require less nitrogen than the standard wheat varieties. For quality analysis, the indirect parameters protein content, sedimentation and falling number will be measured at Strube, Söllingen. Direct baking quality parameters (milling and baking tests) will be assessed at the "Bundesanstalt für Getreideforschung", Detmold. Grain samples for measuring quality parameters will be taken from the standard N-fertilization plots at the Dikopshof.

In addition, the AB lines will be cultivated on a third plot without any fungizide treatment. Here, the possible field resistances of the AB lines against wheat diseases will be evaluated. The disease reactions investigated include powdery mildew, Septoria leaf spot, Drechslera leaf blotch, brown leaf rust, yellow stripe rust, Septoria head blotch and Fusarium head blight.

During the last six months of the second project phase, the phenotype and genotype data will be statistically analyzed, presumably resulting in the localization of significant QTLs for each parameter evaluated in the field. Based on successful executed AB-QTL projects in tomato, rice and barley it can be expected that favorable QTL alleles from the donor wild species will also be identified in our wheat study. At the end of the second project phase (July 2006), AB lines containing favorable wild species QTL alleles will be selected and made available to (i) co-operating wheat breeders as improved breeding material and to (ii) plant molecular biologists as a genetic resource tool for future studies on the molecular regulation of quantitative traits. After purification of the QTL alleles into QTL-NILs, these lines can be used as a starting point for succeeding studies of the differential expression of QTL-related genes which might ultimately lead to the map-based cloning of the underlying genes for quantitative traits.

The present state of the project

a) Propagation of the AB lines: In Summer 2002, approx. 400 BC2F4 lines of the spring wheat populations Triso*Synthetic84L and Devon*Synthetic84L as well as approx. 400 BC2F3 lines of the winter wheat populations Batis*Synthetic22L and Zentos*Synthetic86L were selected in Bonn. The four AB populations will be propagated in Bonn during the season 2002/03.

b) Polymorphism survey of the AB parents: A search for DNA polymorphism between the parents of the four crosses has already been conducted in co-operation with Dr. M. Röder, Gatersleben. In total, 195 SSRs from Röder *et al.* (1998) have been tested. Between 94 and 100 SSRs could be surveyed unambiguously. From this set, between 76 and 81 SSRs have been identified as polymorphic in a particular cross. The majority of polymorphisms (64) were simultaneously present in all four crosses.

○ K. Pillen, A. Zacharias and J. Léon **Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.).** Theor Appl Genet (2003) 107, 340-352. ○ W. Lange & G. Jochemsen **Use of the gene pools of Triticum turgidum ssp. dicoccoides and Aegilops squarrosa for the breeding of common wheat (*T. aestivum*), through chromosome-doubled hybrids I. Two strategies for the production of the amphiploids.** Euphytica (1992) 59, 197-212. ○ M. S. Röder *et al.* **A microsatellite map of wheat.** Genetics (1998) 149, 2007-2023. ○ S. D. Tanksley & J. C. Nelson **Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines.** Theor. Appl. Genet. (1996) 92, 191-203. ○ J. Xiao *et al.* **Genes from wild rice improve yield.** Nature (1996) 384, 223-224. ○ K. Pillen, A. Zacharias and J. Léon **Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare ssp. spontaneum*.** Theor Appl Genet (2004) 108, 1591-1601. ○ H.P. Piepho and K. Pillen **Mixed modelling for QTL x environment interaction analysis.** Euphytica (2004) 137, 147-153.

Future work

The polymorphic SSR markers plus additional SSR and SNP markers from other sources will be used to genotype the 800 AB lines. The genotype information and the phenotype information will finally be used in order to locate favorable exotic QTL alleles which improve quantitative agronomic traits.

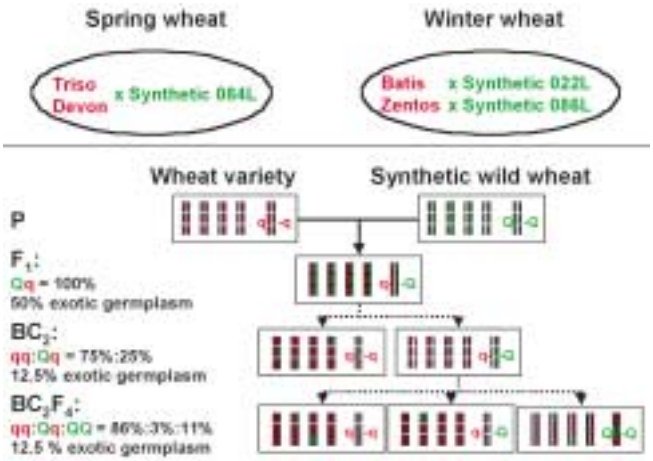


Fig 1: *Generation of four wheat AB populations from crossing spring wheat (Triso and Devon) and winter wheat (Batis and Zentos) elite varieties with synthetic exotic wheat accessions. The graphical genotypes to the right illustrate how the average exotic germplasm is reduced with each round of backcrossing and how the possible genotype classes at a marker locus or a QTL (QQ, Qq and qq) are distributed in each generation.*

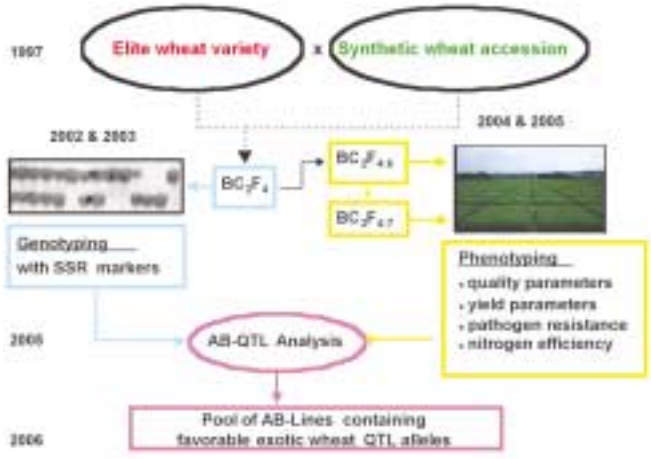


Fig 2: *The Strategy of the AB-QTL analysis in wheat. Synthetic wheat accessions will be used as donors in order to detect QTLs and to improve elite wheat varieties through the introgression of favorable, trait improving exotic QTL alleles.*

Maize Cool Overview

Maize is one of the most important crop plants in the world. Due to the fact that it has been domesticated in southwestern Mexico and because of its tropical origin the maize plant is relatively sensitive to chilling. Breeding maize for the cool and temperate regions of Europe one has until now relied on a chilling escape strategy. By selecting for photoperiod-insensitive early maturing varieties the sowing of maize could be delayed to late spring when chilling temperatures occur more rarely and the crop is harvested early in autumn before the onset of the cool weather.

Chilling in early autumn, i.e. in the grain-filling stage, leads to yield losses and decreases the energy content of silage maize. However, chilling temperatures are particularly deleterious in the initial growth stages of maize where they affect various developmental and physiological processes from germination to the early vegetative phase. Maize varieties with better chilling tolerance would not only improve yield due to a longer growing season but would also be highly desirable in aiming at a more sustainable agriculture. Sowing maize in the cool temperatures of early spring would result in a better ground cover and help to minimize erosion.

Although maize as a species is a relatively chilling-sensitive plant, its genome exhibits, nevertheless, quite a large degree of variation for chilling tolerance. For a better understanding of the genomics of chilling tolerance and in order to improve this trait, genetic variation and candidate gene approach are used to identify genomic regions conferring tolerance to chilling stress.

The first critical step in conducting a candidate gene approach is the choice of suitable genes that may play a relevant role in the process under investigation, i.e. chilling stress. As a matter of course up- or down-regulated genes identified experimentally by differential expression behavior under chilling stress are candidate genes par excellence. However, in addition, known candidate genes traced by extensive literature and database searches were incorporated into marker development.

The milestones concerning identification of candidate genes for chilling tolerance based on differential expression behavior were achieved.

Co-segregation between five candidate genes and QTL regions for fresh matter yield reveals the applicability and usefulness of the candidate gene approach as it was carried out. Such cases of co-segregation, especially with candidate genes derived from the gene expression profiling experiments, will serve as an important criterion when selecting chromosomal regions for further investigation.

Maize-Promoter Overview

The identification of tissue specifically or developmentally regulated promoter sequences is still of growing interest for the development of more efficient approaches in transgenic plant research. Therefore a promoter trapping system was developed that utilizes the maize specific transposable element Mutator (Mu) as a tool for transposition. Two regulatory genes of the anthocyanin biosynthesis pathway, R and C1, were used as reporter genes. After transformation and transposition of the cassette by the autonomous transposable element MuDR, a tagged promoter activity will induce the anthocyanin biosynthesis regulator gene(s) in the trapping construct. The specific promoter activity should be visible by color formation in the specific plant tissues that are responsible for promoter activity. The suitability of the chosen reporter genes, was successsfully demonstrated in stably transformed plants. The over-expression of each of the genes led to visible anthocyanin coloration in several different plant tissues. Plants obtained from crosses between C1 and R over-expressing lines showed strong anthocyanin coloration in most tissues. These experiments demonstrate that nearly all tissues of the maize plant are capable of color development after expression of both of the regulatory genes C1 and R. Therefore C1 and R fulfill the prerequisites to be suitable reporter genes in the system.

The cloning of a functional trapping construct will only be possible if the reporter gene can be expressed (although it is flanked by the two TIR of the transposable element) and if the reporter gene is not induced by a putative TIR specific promoter activity. To prove these conditions extensive transient and stable expression experiments with various TIR reporter gene constructs were carried out. It could be shown that the cloned TIR sequences possess no specific promoter activity. However, a reporter gene activity was reduced if a TIR sequence was cloned between the promoter and the reporter gene. To optimize the read through capability, a point mutation was introduced into the TIR. The mutation enhanced the reporter gene activity more than threefold. Trapping constructs containing one or both of the reporter genes were cloned, tested in transient experiments and introduced into suitable maize lines by biolistic transformation. The obtained plants were crossed with a line carrying the autonomous Mutator element MuDR. PCR-analysis of the progenies proved transposase mediated insertion events of the trapping construct in these plants. The current results suggest a high transposition frequency of the trapping construct and the convenient use of the constructs in the achieved promoter trapping system.

Genomics of Chilling Tolerance in Maize

Maize
GABI Cool

2 Thomas Presterl, 1 Karin Ernst, 2 Evelyn M. Möller, 3 Milena Ouzunova, 2 Hartwig H. Geiger, 1 Peter Westhoff
1 Heinrich-Heine University Düsseldorf, Department for Development and Molecular Biology, Düsseldorf
2 University of Hohenheim, Department of Plant Breeding, Seed Science and Population Genetics, Stuttgart
3 KWS SAAT AG, Einbeck

Scientific backgrounds and objectives

Although modern maize hybrids exhibit greater chilling tolerance than their tropical origins, the successful cultivation of maize in central Europe rests mostly on the high growth potential and on early maturity (3). Chilling temperatures are particularly deleterious in the initial growth stages where they affect various developmental and physiological processes from germination to the early vegetative phase (2). Maize varieties with better chilling tolerance would not only improve yield due to a longer growing season but would also be highly desirable in aiming at a more sustainable agriculture (3). For a better understanding and improvement of the genetic basis of chilling tolerance in maize, we firstly identified candidate genes for chilling tolerance by interrelating published information and by experiment due to their differential expression behavior under conditions of chilling stress. For these genes SNP-based markers suitable for linkage mapping were developed. Secondly, QTL for chilling tolerance were determined in a double haploid line mapping population derived from crosses between two contrasting elite inbred lines. Thirdly, the identified putative candidate genes were validated by co-localization with QTL (fig 1).

Identification of genomic regions involved in chilling tolerance

To generate a doubled haploid line mapping population, the chilling sensitive SL line and the chilling tolerant TH line were used as parents (Fig 2). From the F1 (SLxTH) a DH-line mapping population was produced by in vivo haploid induction. In total 720 DH-lines were increased and crossed to a flint tester line. The parents of the DH-line mapping population TH and SL were screened for polymorphisms using SSR markers evenly distributed over the genome. In total 146 SSR markers were mapped in the DH line population.

Phenotypic values of DH-lines per se and their testcrosses were assessed in field experiments across locations differing in their temperature conditions during the maize growing period. Plants were harvested at an early developmental stage (beginning of shoot elongation) in July to assess differences in adaptation to the chilly temperatures during May and June. Grain yield and yield related traits were determined to assess chilling tolerance in the late phase of plant development.

Chilly temperatures in the early maize growing periods of 2001 and 2002 led to a significant differentiation in early plant development of DH lines. The statistical analysis showed significant genotypic variation within the DH line population as well as between the corresponding testcrosses for fresh matter yield, grain yield and various plant traits. Heritabilities were medium to high for fresh matter and grain yield. Between 4 and 11 QTL with a LOD > 3 were detected for the different yield traits. The QTL with the largest LOD value was detected in the line per se experiments. This QTL explained 19% of the phenotypic variation in 2001 and 38% in 2002. The DH line per se trials of 2001 and 2002 showed four common QTL, while the testcross trials had three QTL in common. This indicates a good reproducibility of the results. In addition to the yield traits, QTL for several other chilling associated traits were detected. The comparison of the QTL positions of the various traits showed clusters in different chromosomal regions. Those QTL clusters will preferably be chosen for the second project phase.

The first set of candidate genes was mapped in the same DH line population. Co-segregation between five candidate genes and QTL regions for fresh matter yield was observed. All milestones for material development and QTL mapping were achieved according to the time schedule given in the original proposal.

Development of allele-specific SNP (single nucleotide polymorphism) markers for genetic mapping of candidate genes

The first critical step in conducting a candidate gene approach is the choice of suitable genes that may play a relevant role in the process under investigation, i.e. chilling stress. It is self-evident that up- or down-regulated genes identified experimentally by differential expression behavior under chilling stress are candidate genes par excellence. However, in addition, known candidate genes traced by extensive literature and database searches were incorporated into marker development (4). These genes are involved in:

- Antioxidative metabolism: superoxide dismutase, ascorbate peroxidase
- Lipid metabolism: desaturases
- Osmolyte synthesis: sugars, proline
- Stress proteins: dehydrins, LEA proteins, COR proteins, aquaporins
- Signal transduction: calcium-dependent protein kinases, mitogen-activated protein kinases
- Transcription factors: DREB1/2, CBF1

Depending on the type, number and distribution of SNPs encountered in the sequenced gene fragments IDP-based (Insertion-Deletion-Polymorphisms) or CAPS (cleaved amplified polymorphic sequence) markers were developed. A total of 33 SNP assays for candidate genes could be developed. So far five of the candidate genes chosen after surveying public information could be mapped in a QTL interval. This demonstrates that our approach to find candidate genes by literature surveys and data base researches have been successful. Such cases of co-segregation, especially with candidate genes derived from the gene expression profiling experiments, will serve as an important criterion when selecting chromosomal regions for project phase 2

Candidate genes for chilling tolerance: Identification by differential expression behaviour

To isolate cDNAs that were associated with chilling stress the early phase of the chilling response of the maize plant was chosen for experiments. At this stage the regulatory genes of the chilling response, like transcription factors and components of signal transduction chains (5), should be fully active but also the actual effector genes of the chilling response should be already switched on. On the other side, genes that may become activated due to secondary effects and therefore are not part of the primary chilling response should still be more or less silent. A high quality, non-amplified cDNA library

○ 1. L. Diatchenko *et al.* **Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries** Proc Natl Acad Sci USA (1996) 93, 6025-6030. ○ 2. J. A. Greaves **Improving suboptimal temperature tolerance in maize – The search for variation** J Exp Bot (1996) 47, 307-323. ○ 3. P. Stamp **Impact of chilling stress on maize – an overview.** In K. Dörffling, B. Brettschneider. (1994) ○ 4. M. F. Tomashow **Role of cold-responsive genes in plant freezing tolerance** Plant Physiol (1998) 118, 1-7. ○ 5. J. K. Zhu, **Cell signaling under salt, water and cold stresses** Curr Opin Plant Biol (2001) 4, 401-406.

was used as a central source for chilling-induced cDNAs. The library was prepared from the TH line and consisted of 40,000 individual clones that were arrayed in micro titer plates. Ten thousand of these clones were spotted onto nylon membranes to be used for colony hybridization with the subtracted cDNA probes (in collaboration with PD Dr. Bernd Weisshaar, MPIZ Köln). The suppression subtraction hybridization (SSH) method (1) was used to identify cDNAs whose corresponding mRNAs were induced or repressed by the chilling treatment. Since the SSH method generates fragmented cDNAs and on the other side full-size cDNAs were aimed for, the subtracted cDNAs were not cloned but used as hybridization probes for screening the arrayed cDNA library. Evaluation of the hybridization signals revealed that about 2% of the clones hybridized with the SSH probe that was enriched for chilling-induced genes, while 1% of the clones hybridized with the probe enriched for chilling-repressed genes. Taken together, 576 putative chill- induced cDNAs were isolated.

To confirm that the 576 putative candidates indeed represented chilling-induced genes a two-step verification procedure was pursued. Firstly, all isolated candidate sequences were analyzed by macro array hybridizations, and secondly verified candidates were re-examined by Northern hybridization. The macro array experiments confirmed that 130 cDNAs could be classified as chilling-regulated genes. All of these cDNAs were sequenced and subjected to standard bioinformatics analyses.

To verify the results of the macro array hybridization experiments Northern blot analyses were performed with 17 selected genes. All of these 17 tested genes could be confirmed as being cold induced. We conclude from these experiments that the chosen strategy which combines pre-selection of candidates for chilling-induced genes by suppression-subtraction hybridization and their verification by macro array hybridization was a powerful strategy for the isolation of chilling-induced genes.

To test whether the expression of the chilling-induced genes vary between chilling-sensitive and -tolerant maize lines both the sensitive SL and the tolerant TH line were grown side by side under the same growth conditions. Of the 96 genes assayed by macro array hybridization the expression of twelve of them was found to be different between the two lines. Taken together, these experiments support our working hypothesis that the chilling-sensitive and -tolerant maize differ in the expression levels of chilling stress-associated genes, and that suppression subtraction hybridization techniques are successful in unraveling these differences.

Summary

All milestones for material development and QTL mapping in the DH line mapping population were achieved according to the time schedule given in the original proposal. The DH line per se and testcross trials of 2001 and 2002 showed very good overlap of QTL positions confirming the good reproducibility and reliability of the QTL results in a large DH population. All together there exist good prerequisites to develop NILs with superior chilling tolerance QTL.

Overall, the chosen strategies and the methodical improvements done for the PCR with genomic maize DNA turned to be suited for development and mapping of allele specific SNP markers and 33 SNP assays for candidate genes could be developed. The milestones concerning identification of candidate genes for chilling tolerance based on differential expression behavior were also achieved according to the time schedule given in the original proposal. Subtractive cloning techniques combined with filter-based quantitative macro array hybridizations have proven to be highly efficient both in terms of sensitivity and cost-effectiveness to identify genes that are induced or repressed by chilling treatment of maize plants. We were also able to demonstrate that the chilling-tolerant (TH) and the chilling-sensitive (SL) maize lines differ in the expression of these genes under chilling stress. Co-segregation between five candidate genes and QTL regions for fresh matter yield reveals the applicability and usefulness of the candidate gene approach as it was carried out. Such cases of co-segregation, especially with candidate genes derived from the gene expression profiling experiments, will serve as an important criterion when selecting chromosomal regions for project phase 2.

Fig 1: *Candidate gene approach to identify genomic regions conferring tolerance to chilling stress*

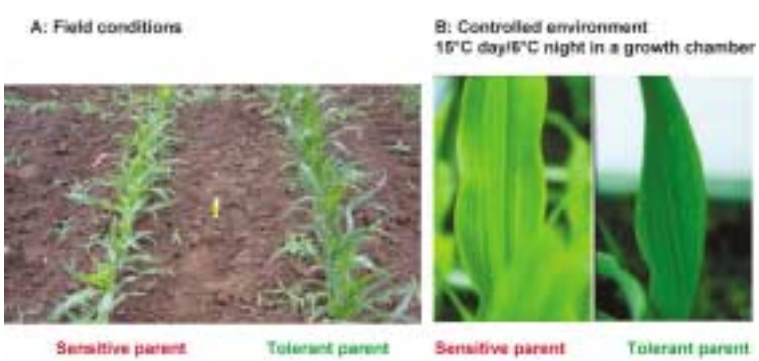


Fig 2: *Genetic variability for chilling tolerance between the parents of the mapping population (A) under field conditions and (B)under controlled environment.*

Isolation and analysis of monocotyledonous promoters from *Zea mays* L.

Maize
Promoter

1 René Lorbiecke, 1 Katja Müller, 1 Carina Paul, 1 Reinhold Brettschneider, 2 Günter Donn, 1 Udo Wienand
1 University of Hamburg, Section for Molecular Biology of Plants, Hamburg
2 Bayer Crop Science, Frankfurt/Main

Aim and scientific background

The tissue specific expression of genes is highly regulated and depends on their corresponding promoter sequences. The identification of such tissue and/or developmentally specific promoter sequences from crop plants is still a very important prerequisite for the establishment of transgenic plants. These promoters are necessary tools for the development of new lines, which can be used in new breeding strategies.

Maize is one of the most important crop plants worldwide and has been studied at the molecular level since decades. For the identification of tissue specifically expressed genes transposon-tagging strategies have been used frequently and successfully in maize. Transposons are families of short stretches of DNA that are present in nearly all organisms and that can relocate (transpose) within the genome. For transposition, "terminal inverted repeat" sequences (TIR, fig1) on both ends of the transposon have to be recognized by the enzyme transposase that afterwards mediates excision and new integration of the element. One member of a transposon family encodes the transposase. In maize, the Mutator (Mu) family is among the best-characterized transposable element families (Bennetzen 1996; Raizada *et al.* 2001).

The aim of the first GABI project phase (2000-03) was the establishment of a system that uses the maize specific transposable element Mutator as a convenient tool for the identification of tissue and/or developmentally specific promoter sequences in maize. The method is based on a promoter trapping system that uses a promoter-less reporter gene cloned between the Mu specific TIRs. Two regulatory genes of the anthocyanin biosynthesis pathway, R and C1 (Ludwig *et al.* 1989; Paz-Ares *et al.* 1986), are used as reporter genes. After transformation into the maize plant the transposition and/or duplication of the trapping construct will be mediated by the transposase activity of the endogenous autonomous Mutator element MuDR. After translocation of the trapping cassette behind a promoter by chance, the tagged promoter activity should be visible by the formation of red pigments. This is due to the activation of the anthocyanin pathway by the induction of the reporter genes C1 and/or R. Only such plant tissues that are responsible for this promoter activity would show color formation (fig1).

Results

The project was devided into two main subprojects: 1) An approach based on tissue culture and plant transformation to develop and establish the requirements for transformation and activity of the trapping constructs. 2) A molecular approach with focus on the development and testing of suitable trapping constructs.

In dependence on the genetic background of the maize lines, one or both of the regulatory genes C1 and R have to be active to induce color formation. Thus, a test requires the transformation of a maize line with the C1 gene and of another one with the R gene followed by subsequent cross of these two transgenic lines. In these experiments it could be proved that the reporter system is functional and that anthocyanin coloration can be induced in nearly all tissues of the maize plant if both reporter genes are expressed (Fig2a).

One other key condition for the success of the project was that the flanking TIRs should allow the expression of the reporter genes. This is particularly important, as one TIR will be positioned between a tagged promoter and the reporter gene sequence. For this purpose, the Mu flanking TIRs that are necessary for transposition, were cloned from two different Mu elements. The suitability of the TIR sequences was tested in extensive transient and stable expression studies with the commonly used reporter genes luciferase and beta-glucuronidase. It could be shown that, after a minor modification of the TIR sequence, this modified TIR is suitable for the use in a trapping construct.

Based on these results, two trapping cassettes that contain one or both reporter genes, respectively, were cloned (Fig2b). The capacity to activate the reporter genes inside the trapping construct was proved in assays that are based on the induction of color formation in immature maize embryos after transient biolistic transformation. It could be shown that the reporter genes are activated in both constructs when a promoter is present in front of the trapping constructs.

To establish maize lines that contain the trapping constructs, these were introduced by biolistic transformation into maize lines that are suitable for tissue culture regeneration.

Southern as well as PCR analyses verified the integration of the complete trapping constructs in seven independent lines. All plants were phenotypically normal, as was expected because trapping constructs without a promoter sequence were used. One transgenic plant containing the R-construct was reciprocally crossed with a wildtype line and the plants (F1) segregating for the transgene were tested by PCR and Southern analyses for transgene integration. It could be shown that in all cases the integration of the trapping cassette was stable and correlated with the herbicide resistance that was used as a selection marker for plant regeneration.

To induce transposition and/or duplication of the trapping construct, the presence of an active MuDR element is necessary. Because the MuDR element is not present in maize lines that can be used for plant transformation and regeneration, a MuDR active line was crossed with the F1 transgenic plants that contain the trapping cassette. The progenies were ana-

○ J. L. Bennetzen **The Mutator transposable element system of maize.** Curr. Top. Microbiol. Immunol. (1996) 204:195-229. ○ M. N. Raizada *et al.* **A maize MuDR transposon promoter shows limited autoregulation.** Mol. Genet. Genomics (2001) 265(1), 82-94. ○ S.R. Ludwig *et al.* **LC, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region.** Proc. Natl. Acad. Sci. (1989) 86, 7092-7096. ○ J. Paz-Ares *et al.* **Molecular cloning of the c locus of maize: a locus regulating the anthocyanin pathway.** EMBO Journal (1986), 5, 829-833.

lyzed for transgene integration as well as for MuDR activity by Southern analysis and PCR. To prove that the cloned trapping cassette can be transposed by the MuDR element, Southern analysis was carried out and a modified PCR-based AIMS method was developed. Using this method two independent new insertion events of the trapping construct could be proved in the one plant that has been analyzed so far. The data suggest that the trapping cassette behaves like a wildtype Mu transpososable element. Database comparison showed that one of the cloned integrations occured in a gene. This result points out that the trapping construct can integrate, as was expected, in transcribed regions of the genome.

In this project a functional trapping cassette based on the transposable element Mutator has been developed. The data suggest a high transposition frequency of the construct and thus a convenient use in the promoter trapping system. The regulators of anthocyanin biosynthesis, C1 and R are suitable reporter genes for this system.

According to the achieved aim of the project, all prerequisites for a successfull promoter trapping approach have been developed during the first period of GABI.

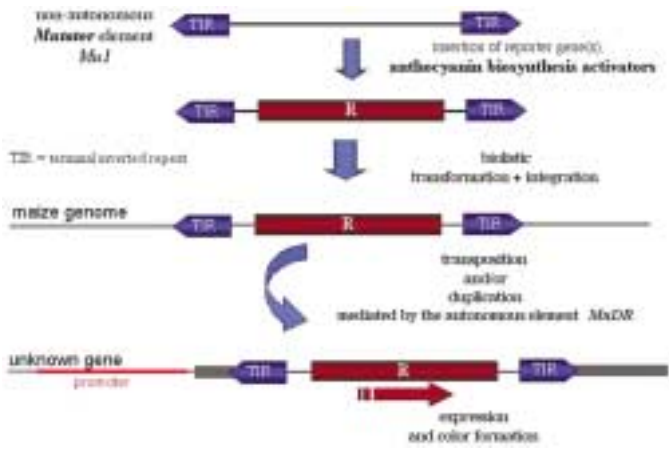


Fig 1: *Schematic drawing of a promoter trapping system in maize. The system is developed to identify tissue or developmentally specific regulated promoters. The trapping construct uses the transposable element Mutator and regulators of the anthocyanin biosynthesis as reporter genes. When a promoter is tagged due to relocation of the trapping cassette in the maize genome, the reporter gene wil be activated and color formation will be visible only in that tissues of the plant in which the promoter is active.*

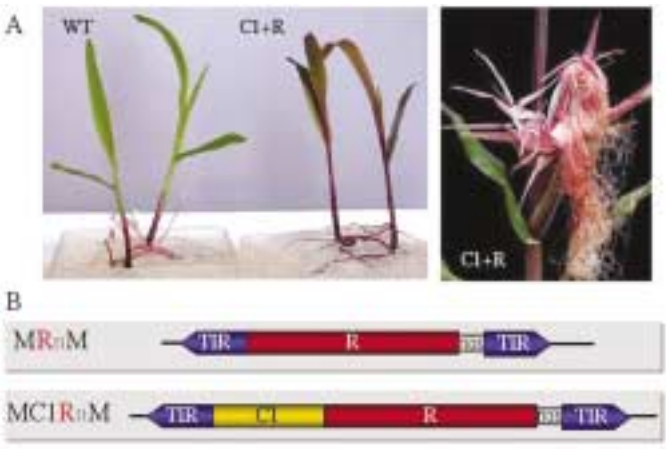


Fig 2: *a: Plants that express the anthocyanin regulatory genes C1 and R under the control of a constitutively active promoter (C1+R) show strong color formation in nearly all tissues if compared to wildtype plants (WT). b: Two constructs were cloned to be used as trapping cassettes in the system one construct with the regulatory gene R and one with the regulatory genes C1 and R, respectively. The transposition of the trapping cassette will be mediated by the terminal inverted repeats (TIR) derived from the Mutator transposon. nos= nopalin synthase terminator of transcription.*

Authors Index

A
Albrecht, Stefanie . . . 32
Altmann, Thomas . . . 62, 66, 70
Amende, Marion . . . 30

B
Bachmann, Konrad . . . 38
Badani, Ana Gloria . . .
Baetzel, Roland . . . 90
Becker, Heiko . . . 94
Berger†, Dieter . . . 66
Biemelt, Sophia . . . 150
Blattner, Frank . . . 38
Böhm, Jutta . . . 100
Borodina, Tatiana . . . 70
Brahm, Lutz . . . 100
Brettschneider, Reinhold . . . 166
Bringezu, Thomas . . . 154
Brumm-Scholz, Martina . . . 42
Bürkle, Lukas . . . 30
Busch, Heinrich . . . 96
Büttner, Georg . . . 106

C
Cahill, Dolores J. . . . 26
Choo, Tze Siang . . . 82
Clemens, Stephan . . . 86
Cossu, Roberto . . . 146

D
Dekker, Koen . . . 24
Desimone, Marcelo . . . 72
Deutsch, Frank . . . 124
Dewal, Gieta . . . 76
Donn, Günter . . . 166

E
Ebert, Berit . . . 84
Ecke, Wolfgang . . . 94
Eickhoff, Holger . . . 32
Ekanem, Joe-Gmem . . . 150
Ernst, Karin . . . 164

F
Feilner, Tanja . . . 26
Felk, Angelika . . . 150
Fensch, Rebekko . . . 150
Fisahn, Joachim . . . 84
Fischer, Dirk . . . 38
Fischer, Rita . . . 30
Fladung, Matthias . . . 124
Flügge, Ulf-Ingo . . . 72
Frank, Markus . . . 148
Freitag, Jens . . . 5, 12
Frese, Lothar . . . 106
Friedt, Wolfgang . . . 90
Frommer, Wolf B. . . . 72

G
Ganal, Martin . . . 110, 136, 146
Gebhardt, Christiane . . . 120
Geiger, Hartwig H. . . . 156, 164
Geiger, Michael . . . 92
Geistlinger, Jörg . . . 38
Giavalisco, Patrick . . . 28
Gobom, Johan . . . 28
Gonella, Giorgio . . . 72
Graner, Andreas . . . 34, 36, 134, 142, 146
Gubatz, Sabine . . . 140
Gustavsson, Niklas . . . 28

H
Hackauf, Bernd . . . 154, 156
Heim, Ute . . . 92
Henning, Simon . . . 150
Hensel, Götz . . . 44
Herbers, Karin . . . 92
Herz, Markus . . . 144
Heyl, Alexander . . . 30
Hofferbert, Hans-Reinhard . . . 120
Hohmann, Uwe . . . 106, 114
Horn, Renate . . . 96

J
Jansen, Carin . . . 150
Jung, Christian . . . 106, 114

K
Kehling, Carina . . . 110
Kehr, Julia . . . 84
Kersten, Birgit . . . 26
Kieselstein, Charlotte . . . 108
Klose, Joachim . . . 28
Koch, Georg . . . 112, 114
Kogel, Karl-Heinz . . . 150
Koprek, Thomas . . . 48
Korff Schmising, Maria von . . . 130
Korzun, Viktor . . . 156
Kota, Raja . . . 134
Kramer, Armin . . . 26
Krüger, Antje . . . 32
Kuhn, Eckehard . . . 30
Kulosa, Dagmar . . . 110
Kumar, Sandeep . . . 124
Kumlehn, Jochen . . . 44, 150
Kunert, Antje . . . 160
Kunze, Reinhard . . . 72

L
Lehrach, Hans . . . 26, 28, 30, 32
Léon, Jens . . . 130, 160
Li, Yong . . . 24
Lieckfeldt†, Elke . . . 84
Linz, Bodo . . . 154
Lorbiecke, René . . . 166
Lörz, Horst . . . 42
Lotz, Tina . . . 96
Lübeck, Jens . . . 120
Lüdecke, Christoph . . . 98
Lühs, Wilfried . . . 90

M

Malysheva, Ludmilla . . . 136
Märländer, Bernward . . . 106
Marschalek, Rubens . . . 94
Mayer, Klaus . . . 50
Mewes, Hans-Werner . . . 50, 106
Meyer, Rhonda . . . 62
Meyer, Svenja . . . 54
Miedaner, Thomas . . . 156
Mikolajewski, Sabine . . . 144
Mitchell-Olds, Thomas . . . 64, 70
Möhring, Silke . . . 108
Möller, Evelyn M. . . . 164
Mühlenbeck, Silke . . . 46
Müller, Katja . . . 166
Müller-Röber, Bernd . . . 74
Müssig, Carsten . . . 62

N

Nagel, Axel . . . 54
Neuhaus, Gisela . . . 132
Niemann, Maike . . . 106
Nietfeld, Wilfried . . . 32
Nordhoff, Eckhard . . . 28
Nürnberger, Thorsten . . . 148
Nyarsik, Lajos . . . 32

O

Obel, Nicolai . . . 82
Ordon, Frank . . . 132
Ouzunova, Milena . . . 164

P

Parker, Jane . . . 148
Paul, Carina . . . 166
Pauly, Markus . . . 82
Perovic, Dragan . . . 146
Pillen, Klaus . . . 130, 160
Polley, Andreas . . . 108
Poßling, Alexandra . . . 26
Potokina, Elena . . . 142
Prasad, Manoj . . . 34
Presterl, Thomas . . . 164

R

Radchuk, Volodymyr . . . 140
Reich, Susanne . . . 110
Reiss, Bernd . . . 24
Richter, Konstanze . . . 98
Röder, Marion . . . 136, 146
Rosleff Sörensen, Thomas . . . 68
Rosso, Mario . . . 24
Rudd, Stephen . . . 106, 156
von Rüden, Sibylle . . . 150

S

Saedler, Heinz . . . 24
Sagasser, Martin . . . 76
Schäfer, Wilhelm . . . 150
Schäfer-Pregl, Ralf . . . 107
Scheel, Dierk . . . 86, 148
Schmidt, Anja . . . 72
Schmidt, Jürgen . . . 86
Schmidt, Thomas . . . 106
Schneider, Anja . . . 72
Schneider, Katharina . . . 108, 116
Schneider, Uli . . . 30
Schondelmaier, Jörg . . . 132, 156
Schubert, Anne . . . 160
Schulz, Britta . . . 114, 116
Schulze-Kremer, Steffen . . . 54
Schulze-Lefert, Paul . . . 148
Schwacke, Rainer . . . 72
Schweizer, Günther . . . 144
Schweizer, Patrick . . . 148
Serazetdinova, Liliya . . . 42
Simon Rosin, Ulrike . . . 84
Snowdon, Rod . . . 96
Soldatov, Aleksey . . . 70
Sonnewald, Uwe . . . 150
Sreenivasulu, Nese . . . 140
Stein, Nils . . . 34, 36, 134, 142, 146
Steiner-Lange, Sabine . . . 24
Stelling, Dieter . . . 98
Stitt, Mark . . . 78
Stracke, Ralf . . . 68
Susic, Zoran . . . 156

T

Tacke, Eckart . . . 120
Thurau, Tim . . . 106
Törjek, Otto . . . 66

U

Usadel, Björn . . . 82

V

van der Graaff, Eric . . . 72
Vishnoi, Rajiv . . . 44

W

Wang, Huajun . . . 130
Wehling, Peter . . . 154, 155
Wehrmeyer, Silke . . . 26
Weisshaar, Bernd . . . 24, 68, 70, 76, 108
Weschke, Winfriede . . . 140
Wessjohann, Ludger . . . 86
Westhoff, Peter . . . 164
Weyen, Jens . . . 46, 134
Wienand, Udo . . . 166
Wilde, Peer . . . 156
Witt, Isabell . . . 74
Wobus, Ulrich . . . 140
Wortmann, Heinrich . . . 156
Wunder, Jörg . . . 108

Z

Zanor, Maria Ines . . . 74
Zhang, Hangning . . . 36
Zimmermann, Illona . . . 76
Zöller, Daniela . . . 84

T. Altmann and R.A. Narang
Phosphate acquisition heterosis in *Arabidopsis thaliana*: a morphological and physiological analysis.
Plant Soil (2001) 234, 91-97.

A.G Badani, R.J. Snowdon, R. Baetzel, F.D. Lipsa, B. Wittkop, R. Horn, A. De Haro, R. Font, W. Lühs and W. Friedt
A model for the inheritance of seed colour in oilseed rape (*Brassica napus* L.) based on analyses of segregation data, QTL and associated quality traits in two genetically distinct crosses.
Genetics (2005), submitted.

A.G. Badani, R. Snowdon, R. Baetzel, W. Lühs, R. Horn and W. Friedt
QTL mapping for yellow seed colour in oilseed rape (*Brassica napus*).
Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 85-87.

R. Baetzel
Stabilität und Genetik von wertbestimmenden Inhaltsstoffen gelbsamiger Rapslinien (*Brassica napus*) als Basis für die Züchtung von qualitativ hochwertigen Körnerölraps-Sorten (Dissertation Theses).
Agrarwissenschaften, Giessen, 2004.

R. Baetzel, W. Lühs, A.G. Badani and W. Friedt
Development of segregating populations in the breeding of yellow-seeded winter rapeseed (*Brassica napus* L.).
Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 238-242.

R. Baetzel, W. Friedt and W. Lühs
Seed colour assessment in rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy.
Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 243-246.

P.C. Bailey, C. Martin, G. Toledo-Ortiz, P.H. Quail, E. Huq, M.A. Heim, M. Jakoby, M. Werber and B. Weisshaar
Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana*.
Plant Cell (2003) 15, 2497-502.

D. Bellin, M. Werber, T. Theis, B. Schulz, B. Weisshaar and K. Schneider
EST sequencing, annotation and macroarray transcriptome analysis identify preferentially root-expressed genes in sugar beet.
Plant Biology (2002) 4, 700-710.

M. Bevan, K.F.X. Mayer, O. White, J. A. Eisen, D. Preuss, T. Bureau, S. L. Salzberg and H.-W. Mewes
Sequence and Analysis of the Arabidopsis genome.
Current Opinion in Plant Biology (2001) 4, 105-110.

E.D. Brenner, D.W. Stevenson, R.W. McCombie, M.S. Katari, S.A. Rudd, K.F.X Mayer, P.M. Palenchar, S.J. Runko, R.W. Twigg, G. Dai, R.A. Martienssen, P.N. Benfey and G.M. Coruzzi
Expressed sequence tag analysis in Cycas, the most primitive living seed plant.
Genome Biology (2003) 4, R78.

L. Bürkle, T. Feilner, W. Brenner, P. Giavalisco, H. Lehrach and B. Kersten
Genomic and proteomic approaches in plant research (book article).
In Pandalai S.G., ed., Recent Research Developments in Plant Molecular Biology (2003) 1, 239-263, Research Signpost, Kerala, India.

L. Bürkle, S. Meyer, H. Dortay, H. Lehrach and A. Heyl
In vitro recombination cloning of entire cDNA libraries in *Arabidopsis thaliana* and its application to the yeast two hybrid system.
Funct. Integr. Genomics (2005), submitted.

D.J. Cahill
Protein and antibody arrays and their medical applications.
Journal of Immunological Methods (2001) 250, 81-91.

E. Catoni, R. Schwab, M. Hilpert, M. Desimone, R. Schwacke, U.-I. Flügge, K. Schumacher and W.B. Frommer
A potential role of the mitochondrial succinate-fumarate translocator in lipid mobilization of seedlings and ethanolic fermentation in pollen of Arabidopsis.
FEBS Letters (2003) 534, 87-92.

T. Choo *et al.*
Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting.
Plant Physiology (2002) 130, 1754-1763.

T. Czechowski, R.P. Bari, M. Stitt, W.-R. Scheible and M.K. Udvardi
Real-Time RT-PCR profiling of 1500 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.
Plant J. (2004) 38, 366-379.

M. Desimone, E. Catoni, U. Ludewig, M. Hilpert, R. Kunzeb, U.-I. Flügge, W.B. Frommer and K. Schumacher
A novel superfamily of transporters for allantoin and other oxo-derivatives of N heterocyclic compounds in Arabidopsis.
Plant Cell (2002) 14 (4), 847-856.

V. Egelhofer *et al.*
Improvements in protein identification by MALDI-TOF-MS peptide mapping.
Analytical Chemistry (2000) 72, 2741-2750.

V. Egelhofer *et al.*
Protein identification by MALDI-TOF-MS peptide mapping: a new strategy.
Analytical Chemistry (2002) 74, 1760-71.

T. Feilner
Hochdurchsatz Generierung und Analyse von *Arabidopsis thaliana*-Proteinen (Dissertation, Theses).
Fachbereich Biologie, Chemie Pharmazie; Freie Universität Berlin; September 2004.

T. Feilner, C. Hultschig, J. Lee, S. Meyer, R. Immink, A. Koenig, A. Possling, A. Beveridge, D. Scheel, D.J. Cahill, H. Lehrach, J. Kreutzberger and B. Kersten
High-throughput identification of Arabidopsis MAP kinases substrates.
(2004) in preparation.

T. Feilner, J. Kreutzberger, B. Niemann, A. Kramer, A. Possling, H. Seitz and B. Kersten
Proteomic studies using microarrays (review).
Current Proteomics (2004) 1 (4), in press.

M. Fladung and S. Kumar
Gene stability in transgenic aspen-Populus – III. T-DNA repeats influence transgene expression differentially among different transgenic lines.
Plant Biology (2002) 51, 329-338.

M. Frisch, M. Bohn and A.E. Melchinger
PLABSIM: Software for simulation of marker-assisted backrossing.
Journal of Heredity (2000) 91, 86-87.

C. Gebhardt and J.P.T. Valkonen
Organization of genes controlling disease resistance in the potato genome.
2001 Phytopatholy (2001) 39: 79-102.

C. Gebhardt, A. Ballvora and A. Rickert
SaKa-Ragis Pflanzenzucht GbR, Windeby, BNA Zuchtgesellschaft mbH, Ebstorf; The German potato genome project GABI-CONQUEST: Results and Perspectives.
Vorträge zu Pflanzenzüchtung (2002) Supplement 1: 109.

C. Gebhardt
Rein in die Kartoffeln! Das GABI Projekt CONQUEST.
GenomXPress (2002) 2, 5-6.

P. Giavalisco *et al.*
Extraction of proteins from plant tissues for two-dimensional electrophoresis analysis.
Electrophoresis (2003) 24, 207-216.

P. Giavalisco, E. Nordhoff, T. Kreitler, K.-D. Kloeppel, H. Lehrach, J. Klose and J. Gobom
Proteome Analysis of *Arabidopsis thaliana* by 2-D Electrophoresis and Matrix Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry.
Proteomics (2005), in press.

Y. Gibon, H. Vigeolas, T. Tiessen, P. Geigenberger and M. Stitt
Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system.
Plant J. (2002) 30, 221-235.

Y. Gibon, O. Bläsing, N. Palacios-Rojas, D. Pankovic, J.H.M. Hendriks, J. Fisahn, M. Höhne, M. Günther and M. Stitt
Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars, and posttranslational activation of ADP glucose pyrophosphorylase in the following light period.
Plant J. (2004a) 39:847-862.

Y. Gibon, O. Bläsing, J. Hannemann, P. Carillo,
M. Höhne, J. Cross, J. Selbig and M. Stitt
A robot-based platform to measure multiple enzyme activities using a set of cycling assays: comparison of changes of enzyme activities and transcript levels in Arabidopsis during diurnal cycles and in prolonged darkness.
Plant Cell (2004b), in press.

J. Gobom *et al.*
Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics.
Analytical Chemistry (2001) 73, 434-438.

B. Hackauf and P. Wehling
Identification of microsatellite polymorphisms in an expressed portion of the rye genome.
Plant Breed (2002a) 121, 17-25.

L. Hartl, G. Schweizer, M. Herz and M. Baumer
Molekulargenetische Lokalisierung von QTL für die Malzqualität der Gerste.
Bericht über die Arbeitstagung 2000 der Vereinigung österreichischer Pflanzenzüchter, 21. bis 23. November 2000, (2000) 117-122.

B. Heidenreich, K.F.X. Mayer, H. Sandermann and D. Ernst
Mercury induced genes in *Arabidopsis thaliana*: identification of induced genes upon long-term mercuric-ion exposure.
Plant, Cell and Environment (2001) 24 (11), 1227-1234.

M. A. Heim, M. Jakoby, M. Werber, C. Martin, B. Weisshaar and P.C. Bailey
The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-wide Study of Protein Structure and Functional Diversity.
Molecular Biology and Evolution (2003) 20, 735-747.

J.H.M. Hendriks, A. Kolbe, Y. Gibon, M. Stitt and P. Geigenberger
ADP-glucose pyrophosphorylase is activated by post-translational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species.
Plant Physiol. (2003) 133, 838-849.

R. Herwig, B. Schulz, B. Weisshaar, S. Hennig, M. Steinfath, M. Drungowski, D. Stahl, W. Wruck, A. Menze, J. O'Brien, H. Lehrach and U. Radelof
Construction of a unigene cDNA clone set by oligonucleotide fingerprinting allows access to 25,000 potential sugarbeet genes.
Plant Journal 2002 32(5), 845-857.

M. Herz and S. Mikolajewski
Nucleic acid molecules involved in quality traits and methods for their isolation.
European patent application (2003) 03 010 518.13.

M. Herz, S. Mikolajewsi, G. Schweizer and M. Baumer
Identification of candidate cDNAs correlated to malting quality of barley by means of cDNA-AFLP analysis and differential genotype pooling.
Proceedings of the 29th EBC congress (2003), Dublin 2003

P. Hilson, I. Small and M. T. Kuiper
European consortia building integrated resources for Arabidopsis functional genomics.
Curr Opin Plant Biol (2003) 6, 426-429.

P. Hilson *et al.*
Versatile Gene-Specific Sequence Tags for Arabidopsis Functional Genomics: Transcript Profiling and Reverse Genetics Applications.
Genome Research (2004) 14, 2176–2189.

P. Hilson, J. Allemeersch, T. Altmann, S., Avon A. Aubourg, J. Beynon, R.P. Bhalerao, F. Bitton, M. Caboche, B. Cannoot, V. Chardakov, C. Cognet-Holliger, V. Colot, M. Crowe, C. Darimont, S. Durinck, H. Eickhoff, A.F. de Longevialle, E.E. Farmer, M. Grant, M.T. Kuiper, H. Lehrach, C. Leon, A. Leyva, J. Lundeberg, C. Lurin, Y. Moreau, W. Nietfeld, J. Paz-Ares, P. Reymond, P. Rouze, G.Sandberg, M.D. Segura, C. Serizet, A. Tabrett, L. Taconnat, V. Thareau, P. Van Hummelen, S. Vercruysse, M. Vuylsteke, M. Weingartner, P.J. Weisbeek, V. Wirta, F.R. Wittink, M. Zabeau and I. Small
Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications.
Genome Res. (2004) 14(10B), 2176-89.

U. Hohmann, G. Jacobs, A. Telgmann, D. Schmidt,
S. Alam, M. Ganal and C. Jung
A bacterial artificial chromosome (BAC) library of sugar beet and a physical map of the region encompassing the bolting gene B.
Molecular Genetics & Genomics (2003) 269(1), 126-136.

S. Hunger, G. di Gaspero, S. Möhring, D. Bellin, R. Schäfer-Pregl,
D. C. Borchardt, C.-E. Durek, B. Weisshaar, F. Salamini and K. Schneider
Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris* L.).
Genome (2003) 46(1), 70-82.

V. Hurry, R. Furbank and M. Stitt
The role of inorganic phosphate in the development of freezing tolerance and the acclimation of photosynthetic carbon metabolism to low growth temperature is revealed by studies of pho mutants of *Arabidopsis thaliana*.
Plant J. (2000) 24, 383-396.

W. M. Karlowski, H. Schoof, V. Janakiraman, V. Stümpflen and K.F.X. Mayer
MOsDB: an integrated information resource for rice genomics.
Nucleic Acids Research (2003) 31, 190-192.

B. Kersten, L. Bürkle, E.J. Kuhn, P. Giavalisco, Z. Konthur,
A. Lueking, G. Walter, H. Eickhoff and U. Schneider
Large-scale plant proteomics.
Plant Molecular Biology (2002) 48, 133-141.

B. Kersten, T. Feilner, A. Kramer, S. Wehrmeyer, A. Possling, I. Witt,
M.I. Zanor, R. Stracke, A. Lueking, J. Kreutzberger, H. Lehrach
and D.J. Cahill
Generation of Arabidopsis protein chips for antibody and serum screening.
Plant Molecular Biology (2003) 52, 999-1010.

B. Kersten, A. Possling, F. Blaesing, E. Mirgorodskaya,
J. Gobom and H. Seitz
Protein microarray technology and UV-crosslinking combined with mass spectrometry for the analysis of protein-DNA interactions.
Analytical Biochemistry (2004b) 331, 303-313.

B. Kersten, T. Feilner, P. Angenendt, P. Giavalisco, W. Brenner and L. Bürkle
Proteomic approaches in plant biology (review).
Current Proteomics (2004a) 1, 131-144.
K.-H. Kogel, J. Kumar, P. Schäfer, R. Hückelhoven,
G. Langen, H. Baltruschat, E. Stein and S. Nagarajan
Pathogen Profile: Bipolaris sorokiniana, a cereal pathogen of global concern. Cytological and molecular approaches towards better control.
Molecular Plant Pathology (2002) 3, 185-195.

J. Kopka, A.F. Fernie, W. Weckwerth, Y. Gibon and M. Stitt
Metabolite profiling in Plant Biology: Platforms and Destinations.
Genome Biology (2004) 5, 109.

M. von Korff, H. Wang, J. Léon and K. Pillen
Development of candidate introgression lines using an exotic barley accession (*H. vulgare ssp. spontaneum*) as donor.
Theor Appl Genet (2004) (DOI: 10.1007/s00122-004-1818-2).

M. von Korff, J. Plümpe, W. Michalek, J. Léon and K. Pillen
Insertion of 18 new SSR markers into the Oregon Wolfe Barley map.
Barley Genetics Newsletter (2004) (in press).

P. Kosarev, K.F.X. Mayer and C. S. Hardtke
Evaluation and classification of RING finger domains encoded by the Arabidopsis genome.
Genome Biology (2002) 3 (4), research 0016.1-0016.12.

R. Kota, M. Wolf, T. Thiel, W. Michalek and A. Graner
EST as a resource for the systematic development of novel markers in barley (*Hordeum vulgare* L.).
Hereditas (2001) 135 (2-3), 145-151.

R. Kota, M. Prasad, R.K. Varshney, M. Wolf, T. Thiel,
H. Zhang N. Stein and A. Graner
Barley ESTs: Resource for the development of a high density transcript map.
Australian Journal of Agricultural Research 2002

R. Kota, R. Kumar Varshney, T. Thiel, K. Dehmer,
K. Johannes and A. Graner
Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.).
Hereditas (2001a) 135, 145-151.

R. Kota, M. Wolf, W. Michalek and A. Graner
Application of DHPLC for mapping of single nucleotide polymorphisms (SNPs) in barley (*Hordeum vulgare* L.).
Genome (2001b) 44(4), 523-528.

R. Kota, S. Rudd, A. Facius, G. Kolesov, T. Thiel,
H. Zhang, N. Stein, K.F.X. Mayer and A. Graner
SniPping polymorphisms from the large barley EST collections.
Molecular and General Genomics (2003) 270, 24-33.

A. Kramer
Herstellung von Gerste-Protein-Chips mit Proteinen aus cDNA-Expressions-Bibliotheken und Testung ihrer Anwendbarkeit für Phosphorylierungsstudien in vitro (Diplomarbeit, Theses).
Fachbereich Biologie, Chemie Pharmazie; Freie Universität Berlin;
24.9.2003.

A. Kramer, T. Feilner, A. Possling, V. Radchuk,
W. Weschke, L. Bürkle and B. Kersten
Identification of barley CK2a targets by using protein microarray technology.
Phytochemistry (2004) 65, 1777-1784.

S. Kumar and M. Fladung
Transgene repeats in aspen: molecular characterisation suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome.
Molecular Genetics & Genomics (2000) 264, 20-28.

S. Kumar and M. Fladung
Gene stability in transgenic aspen (Populus). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen.
Planta (2001) 213, 731-740.

S. Kumar and M. Fladung
Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration.
Plant Journal (2002) 31, 543-551.

S. Kumar and M. Fladung
Somatic mobility of the maize element Ac and its usability for gene tagging in aspen.
Plant Molecular Biology (2003) 51(5), 643-650.

O. Lerouxel, T.S. Choo, M. Séveno, B. Usadel,
L. Faye, P. Lerouge and M. Pauly
Rapid strucutural phenotyping of plant cell wall mutants by enzymatic oligosaccharides fingerprinting.
Plant Physiology (2002) 130 (4), 1754-1763.

Y. Li, M.G. Rosso, N. Strizhov, P. Viehoever and B. Weisshaar
GABI-Kat SimpleSearch: a flanking sequence tags (FST) database for the identification of T-DNA insertion mutants in *Arabidopsis thaliana*.
Bioinformatics (2003) 19 (11), 1441-1442.

T. Lotz, R. Snowdon, R. Horn, G. Dewal, B. Weisshaar,
W. Friedt, H. Belcram, M. Caboche and B. Chalhoub
Molecular analysis of *Arabidopsis thaliana* transparent testa genes (tt) in *Brassica napus*.
Proc. 11th International Rapeseed Congress, July 2003, Copenhagen, Denmark, (2003) 1, 109-111.

U. Ludewig, N. von Wiren and W.B. Frommer
Uniport of NH4+ by the root hair plasma membrane ammonium transporter LeAMT1.
Journal of Biological Chemistry (2002) 277(16), 13548-13555.

K.F.X. Mayer *et al.*
Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*.
Genome Research (2001) 11, 1167-1174.

K.F.X. Mayer
Bioinformatische Analyse des Arabidopsis Genoms.
Vorträge für die Pflanzenzüchtung (2001) 52, 51-60.

K.F.X. Mayer, S. Rudd and H. Schoof
Arabidopsis and its use in cereal genomics.
In: Cereal Genomics (2004) Kluwer Academic Publishers, Dordrecht, Netherlands.

K.F.X. Mayer
Das Arabidopsis Genom, eine Blaupause für die Genome höherer Pflanzen.
GenomXPress (2001) 1/01, 3-5.

K.F.X. Mayer and H.-W. Mewes
How can we deliver the large plant genomes? Strategies and perspectives.
Current Opinion in Plant Biology (2002) 5, 173-177.

R.C. Meyer, O. Törjek, M. Becher and T. Altmann
Heterosis of biomass production in *Arabidopsis thaliana*; Establishment during early development.
Plant Physiology (2004) 134, 1813-1823.

S. Meyer, A. Nagel and C. Gebhardt
PoMaMo – a comprehensive database for potato genome data.
Nucleic Acids Research, Database issue 2005, in press.

W. Michalek *et al.*
EST analysis in barley defines a unigene set comprising 4,000 genes.
Theoretical Applied Gentics (2002) 104, 97-103.

Mikolajewski *et al.*
Untersuchung differentieller Genexpression im Verlauf der Vermälzung von Gerstenkörnern – ein Vergleich von Brau- und Futtergerstensorten mittels cDNA-AFLP-Technik.
Vortr. Pflanzenzüchtung (2002), 54, 405-408.

S. Möhring *et al.*
Multiplexed, linkage group-specific SNP marker sets for rapid genetic mapping and fingerprinting of sugar beet (*Beta vulgaris* L.).
Molecular Breeding (2004) 14:475-488.

B. Morgenstern, O. Rinner, S. Abdeddaim, D. Haase, K.F.X. Mayer, A. Dress and H.-W. Mewes
Exon Discovery by Genomic Sequence Alignment.
Bioinformatics (2002) 18(6):777-87.

R.A. Narang and T. Altmann
Phosphate accquisition heterosis in *Arabidopsis thaliana*: a morphological and physiological analysis.
Plant Soil (2001) 234, 91-97.

L. Narziß
Technologie der Malzbereitung.
Ferdinand Enke Verlag Stuttgart (1999).

E. Nordhoff, H. Eickhoff, M. Horn, T. Przewieslik, V. Egelhofer, P. Giavalisco, D. Theiss, H. Lehrach and J. Gobom
Large-Gel-2-D-MALDI-TOF-MS – An Analytical Challange for Studying Complex Protein Mixtures.
Electrophoresis (2001) 22, 2844-55.

K. Pillen, A. Zacharias and J. Léon
Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.).
Theor Appl Genet (2003) 107, 340-352.

K. Pillen, A. Zacharias and J. Léon
Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare ssp. spontaneum*.
Theor Appl Genet (2004) 108, 1591-1601.

H.P. Piepho and K. Pillen
Mixed modelling for QTL x environment interaction analysis.
Euphytica (2004) 137, 147–153.

E. Potokina *et al.*
Differential gene expression during seed germination in barley (*Hordeum vulgare* L.).
Funct. Integr. Genomics (2002) 2, 28-39.

E. Potokina, M. Caspers, M. Prasad, R. Kota, H. Zhang, N. Sreenivasulu, M. Wang and A. Graner
Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare* L.)
Molecular Breeding (2004) 14, 153-170.

L. Ramsay, M. Macaulay, S. Ivanissevich, K. MacLean, L. Cardle, J. Fuller, K.J. Edwards, S. Tuvesson, M. Morgante, A. Massari, E. Maestri, N. Marmiroli, T. Sjakste, M. Ganal, W. Powell and R. Waugh
A simple sequence repeat-based linkage map of barley.
Genetics (2002) 156, 1997-2005.

A.M. Rickert, A. Premstaller, C. Gebhardt and P.J. Oefner
Genotyping of SNPs in a polyploid genome by pyrosequencing.
Biotechniques (2002) 32(3), 592-593, 596-598, 600.

A.M. Rickert, J.H. Kim, S. Meyer, A. Nagel, A. Ballvora, P.J. Oefner and C. Gebhardt
First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome.
Plant Biotechnology Journal (2003) 1, 399-410.

A.M. Rickert
Entwicklung und Analyse von SNP-Markern (single nucleotide polymorphisms) in Pathogenresistenz-vermittelnden Regionen des Kartoffelgenoms (Dissertation, Theses).
Universität zu Köln, 2002.

E. von Roepenack-Lahaye, T. Degenkolb, M. Zerjeski, M. Franz, U. Roth, L. Wessjohann, J. Schmidt, D. Scheel and S. Clemens
Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry.
Plant Physiology (2004) 134, 548-559.

M. G. Rosso, Y. Li, N. Strizhov, B. Reiss, K. Dekker and B. Weisshaar
An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics.
Plant Molecular Biology (2003) 53, 247-259.

S. Rudd, H.-W. Mewes and K.F.X. Mayer
SPUTNIK: A database platform for comparative plant genomics.
Nucleic Acids Research (2003) 31, 128-132.

S. Rudd, M. Frisch, K. Grote, B.C. Meyers, K. Mayer and T. Werner
Genome-wide in silico Mapping of S/MARs in *Arabidopsis thaliana* suggests correlation of intragenic S/MARs with gene expression.
Plant Physiology (2004), 135, 715-722.

M. Sagasser *et al.*
***A. thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins.**
Genes & Development (2002) 16, 138-149.

M.A. Sattarzadeh
SNPs (Single nucleotide polymorphisms) and haplotypes linked to QTLs for resistance to Globodera pallida in tetraploid potato (Diplomarbeit, Theses).
Rheinische Friedrich-Wilhelms-Universität zu Bonn, 2003.

W.-R. Scheible, R. Morcuende, T. Czechowski, D. Osuna, C. Fritz, N. Palacios-Rojas, O. Thimm, M.K. Udvardi and M. Stitt
Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes and signalling infrastructure by nitrogen.
Plant Physiology (2004) 136, 2483-2499.

J. Schmidt, T. Rosleff-Sörensen, R. Stracke, O. Törjek, T. Altmann, T. Mitchell-Olds and B. Weishaar
Large scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*.
Genome Research (2003) 13(6), 1250-1257.

K. J. Schmid, S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar and T. Mitchell-Olds
A multilocus analysis of genetic variation in *Arabidopsis thaliana* reveals a genome-wide deviation from a neutral model of nucleotide polymorphism.
Genetics (2005), in press.

K.J. Schmid, T. Rosleff-Sörensen, R. Stracke, O. Törjek, T. Altmann, T. Mitchell-Olds and B. Weisshaar
Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*.
Genome Res. (2003) 13, 1250-1247.

K. Schneider, D. Bellin, M. Werber, T. Theis, B. Schulz and B. Weisshaar
EST sequencing, annotation and macroarray transcriptome analysis identify preferentially root-expressed genes in sugar beet.
Plant Biology (2002) 4(6):700-710.

H. Schoof, P. Zaccaria, H. Gundlach, K. Lemcke, S. Rudd, G. Kolesov, R. Arnold, H.W. Mewes and K.F.X. Mayer
MIPS *Arabidopsis thaliana* Database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome.
Nucleic Acids Research (2002) 30 (1), 91-93.

H. Schoof, W.M. Karlowski, V. Janakiraman, V. Stümpflen and K.F.X. Mayer
MOsDB: an integrated information resource for rice genomics.
Nucleic Acids Research (2003) 31, 190-192.

H. Schoof, R. Ernst, V. Nazarov, L. Pfeifer, H.-W. Mewes and K.F.X. Mayer
MIPS *Arabidopsis thaliana* Database (MAtDB): An integrated biological knowledge resource for plant genomics.
Nucleic Acids Research (2004), 32, Database issue D373-D376.

K. Schrick, D. Nguyen, W. Karlowski and K.F.X. Mayer
START Lipid/Sterol Binding Domains are amplified in plants and are predominantely associated with homeodomain transcription factors.
Genome Biology (2004) 5, R41.

R. Schwacke, A. Schneider, E. van der Graaff, K. Fischer,
E. Catoni, M. Desimone, W.B. Frommer, U.I. Flugge and Kunze
ARAMEMNON: A novel database for *Arabidopsis thaliana* integral membrane proteins.
Plant Physiology (2003) 131 (1), 16-26.
H. Seitz, M. Krause, C. Gotthold, U. Borgmeier,
B. Kersten, C. Luebbert and V. Heiser
Application of protein array technology to study protein-protein interactions and protein-DNA-interactions.
Molecular & Cellular Proteomics (2003) 2, No. 9, 939.

S.-H. Shiu, W. M. Karlowski, R. Pan, Y.-H. Tzeng, K.F.X. Mayer and W.-H. Li
Comparative analysis of the Receptor-Like Kinase Family in *Arabidopsis thaliana* and Rice.
Plant Cell (2004), 16, 1220-1234.

R.J. Snowdon and W. Friedt,
Molecular markers in Brassica oilseed breeding: current status and future possibilities.
Plant Breeding (2004) 123, 1-8.

N. Sreenivasulu *et al.*
Identification of genes specifically expressed in maternal and filial tissues of the barley caryopsis: a cDNA array analysis.
Molecular Genetics & Genomics (2002a) 266 (5), 758-767.

N. Sreenivasulu *et al.*
Mining functional information from cereal genomes – the utility of expressed sequence tags.
Current Science (2002) 83, 965-973.

N. Sreenivasulu *et al.*
Transcriptome changes in foxtail millet genotypes at high salinity: Identification and characterization of a PHGPX gene specifically upregulated by NaCl in a salt-tolerant line.
Journal of Plant Physiology (2004) 161, 467-477.

N. Sreenivasulu *et al.*
Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains.
Plant Journal (2004) 37,539-553.

M. Stitt and V. Hurry
A plant for all seasons: Arabidopsis as a model system to investigate the cold acclimation of photosynthetic carbon metabolism.
Curr. Opinion Plant Biol. (2002) 5, 199-206.

M. Stitt and A. R. Fernie
From Measurements of Metabolites to Metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions Current Opinion.
Current Opinion in Biotechnology (2003) 14(2), 136-144.

M. Stitt, C. Müller, P. Matt, Y. Gibon, P. Carillo,
R. Morcuende, W.-R. Scheible and A. Krapp
Steps towards an integrated view of nitrogen metabolism.
Journal of Experimental Botany (2002) 53(370), 959-970.

M. Stitt, V. Hurry and R. Furbank
The role of inorganic phosphate in the development of freezing tolerance and the acclimation of photosynthetic carbon metabolism to low growth temperature is revealed by studies of pho mutants of *Arabidopsis thaliana*.
Plant Journal (2000) 24, 383-396.

R. Stracke, M. Werber and B. Weisshaar
The R2R3-MYB gene family in *Arabidopsis thaliana*.
Plant Biology (2001) 4, 447-456.

N. Strizhov, Y. Li, M. G. Rosso, P. Viehoever, K. A. Dekker and B. Weisshaar.
High-throughput generation of sequence indexes from T-DNA mutagenized *Arabidopsis thaliana* lines.
BioTechniques (2003) 35, 1164-1168.

O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L.A. Müller, S.Y. Rhee and M. Stitt
MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.
The Plant Journal (2004) 37(6), 914-939.

T. Thiel, R. Kota, I. Grosse, N. Stein and A. Graner
Snp2caps: A snp and indel analysis tool for caps marker development.
Nucleic Acids Research (2004) 32.

O. Törjek, D. Berger, R.C. Meyer, C. Müsig, K. Schmid, T. Rosleff-Sörensen, B. Weisshaar, T. Mitchell-Olds and T. Altmann
Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis.
Plant J. (2003) 36, 122-140.

O. Törjek, D. Berger, R.C. Meyer, C. Müssig, K.J. Schmid, T. Rosleff-Sörensen, B. Weisshaar, T. Mitchell-Olds and T. Altmann
Establishement of a high-efficiency SNP-based framework marker set of Arabidopsis.
Plant J. (2003) 36, 122-140.

R.K. Varshney, T. Thiel, N. Stein, P. Langridge and A. Graner
In silico analysis on frequency and distribution of microsatellites in ests of some cereal species.
Cellular & Molecular Biology Letters (2002) 7, 537-546.

R.K. Varshney, R. Sigmund, A. Boerner, V. Korzun, N. Stein, M.E. Sorrells, P. Langridge and A. Graner
Interspecific transferability and comparative mapping of barley est-ssr markers in wheat, rye and rice.
Plant Science (2004), in press.

R. Varshney, H. Zhang, E. Potokina, N. Stein, P. Langridge and A. Graner
A simple hybridization-based strategy for the generation of non-redundant est collections – a case study in barley (*Hordeum vulgare* L.).
Plant Science (2004) 167, 629-634.

W. Weschke, R. Panitz, S. Gubatz, Q. Wang, R. Radchuk, H. Weber and U. Wobus
The role of invertases and hexose transporters in conrolling sugar ratios in maternal and filial tissues of barley caryopses during early development.
Plant Journal (2003) 33, 395-411.

T. Wicker, W. Zimmermann, D. Perovic, A.H. Paterson, M. Ganal, A. Graner and N. Stein
A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley hv-eif4e locus: Recombination, re-arrangements and repeats.
The Plant Journal (2004), in press
(doi: 10.1111/j. 1365-313x.2004.02285.x).

U. Wobus *et al.*
Molecular physiology and genomics of developing barley grains.
Recent Research Developments in Plant Molecular Biology (2004) 37/661 (2), 1-29.

H. Zhang, N. Sreenivasulu, W. Weschke, N. Stein, S. Rudd, V. Radchuk, E. Potokina, U. Scholz, P. Schweizer, U. Zierold, P. Langridge, R.K. Varshney, U. Wobus and A. Graner
Large-scale analysis of the barley transcriptome based on expressed sequence tags.
The Plant Journal (2004) 40, 276-290.

I.M. Zimmermann, M.A. Heim, B. Weisshaar and J.F. Uhrig
Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins.
Plant Journal (2004) 40, 22-34.

E. Zuther, K. Buchel, M. Hundertmark, M. Stitt, D.K. Hincha and A.G. Heyer
The role of raffinose in the cold acclimation of Arabidopsis.
FEBS Letters: (2004) in press.

European Union Chromosome 3 Arabidopsis Sequencing Consortium [M. Salanoubat, K. Lemcke,..., H.-W. Mewes and K.F.X. Mayer], The Institute for Genomic Research and Kazusa DNA Research Institute
Sequence and Analysis of chromosome 3 of the plant *Arabidopsis thaliana*.
Nature (2000) 408, 820-823.

The Kazusa DNA Research Institute, The Cold Spring Harbor and Washinghton University St Louis Sequencing Consortium & The European Union Arabidopsis Genome Sequencing Consortium [G. Murphy,....,K.F.X. Mayer,..., H.-W. Mewes and M. Bevan]
Sequence and Analysis of chromosome 5 of the plant *Arabidopsis thaliana*.
Nature (2000) 408, 823-826.

The Arabidopsis Genome Initiative
[Genome Sequencing Groups; Genome Analysis Group: K.F.X. Mayer, O. White *et al.*; Contributing Authors]
Sequence and analysis of the flowering plant *Arabidopsis thaliana*.
Nature (2000) 408, 796-815.

The Arabidopsis Genome Initiative
[Concept and Production: MIPS; technical Realisation: Biomax Informatics]
***Arabidopsis thaliana* Genome.**
CD, Nature (2000).

GABI Contacts

The index comprises contact details for each article, as a general rule those of the initial author.

Bachmann, Konrad
Institute of Plant Genetics and Crop Plant Research
Resourcecenter, Gatersleben
bachmann@ipk-gatersleben.de

Brahm, Lutz
Saatzucht Hadmersleben GmbH, Echzell
lutz.brahm@swseed.com

Bringezu, Thomas
Federal Centre for Breeding Research on Cultivated Plants
Institute of Agricultural Crops, Groß Lüsewitz
bafz-lk@bafz.de

Bürkle, Lukas
Max Planck Institute of Molecular Genetics
Dept. Lehrach, Berlin
buerkle@molgen.mpg.de

Clemens, Stephan
Leibniz Institute of Plant Biochemistry, Halle/Saale
sclemens@ipb-halle.de

Cossu, Roberto
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Dewal, Gieta
Max-Planck-Institute for Plant Breeding Research, Cologne
dewal@mpiz-koeln.mpg.de

Ecke, Wolfgang
Institute of Crop Science and Plant Breeding, Göttingen
wecke@gwdg.de

Fladung, Matthias
Institute for Forest Genetics and Forest Tree Breeding, Großhansdorf
mfladung@uni-hamburg.de

Gebhardt, Christiane
Max-Planck-Institute for Plant Breeding Research, Cologne
gebhardt@mpiz-koeln.mpg.de

Gobom, Johan
Max-Planck-Institute for Molecular Genetics, Berlin
gobom@molgen.mpg.de

Gubatz, Sabine
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Hackauf, Bernd
Federal Centre for Breeding Research on Cultivated Plants
Institute of Agricultural Crops, Groß Lüsewitz
b.hackauf@bafz.de

Heim, Ute
SunGene GmbH & Co KGaA, Gatersleben
ute.heim@sungene.de

Herz, Markus
Bavarian State Research Center for Agriculture, Freising
markus.herz@lfl.bayern.de

Jansen, Carin
Justus-Liebig-University of Giessen
Institute of Phytopathology and Applied Zoology, Giessen

Jung, Christian
Christian-Albrechts-University of Kiel
Dept. of Crop Science and Plant Breeding, Kiel
c.jung@plantbreeding.uni-kiel.de

Kersten, Birgit
Max-Delbrück-Centrum für Molekulare Medizin (MDC) Berlin-Buch
Dept. of Neuroproteomics, Berlin
b.kersten@mdc-berlin.de

Koch, Georg
Strube-Dieckmann, Nienstädt
g.koch@a-dieckmann.de

Koprek, Thomas
Max-Planck-Institute for Plant Breeding Research, Cologne
koprek@mpiz-koeln.mpg.de

Kota, Raja
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Kunze, Reinhard
Freie Universität Berlin, Dept. of Biology – Applied Genetics, Berlin
rkunze@zedat.fu-berlin.de

Kumlehn, Jochen
Institute of Plant Genetics and Crop Plant Research
Plant Reproductive Biology, Gatersleben
kumlehn@ipk-gatersleben.de

Lorbiecke, René
University of Hamburg, Section for Molecular Biology of Plants

Lüdecke, Christoph
Deutsche Saatveredlung Lippstadt-Bremen GmbH, Lippstadt

Malysheva, Ludmilla
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Mewes, Hans-Werner
National Research Center for Environment and Health
Institute for Bioinformatics, Neuherberg
mewes@gsf.de

Meyer, Rhonda
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
meyer@mpimp-golm.mpg.de

Meyer, Svenja
German Resource Center for Genome Research, Berlin
svenja@rzpd.de

Mitchell-Olds, Thomas
Max-Planck-Institute for Chemical Ecology, Jena
tmo@ice.mpg.de

Nietfeld, Wilfried
Max-Planck-Institute for Molecular Genetics, Berlin
nietfeld@molgen.mpg.de

Obel, Nicolai
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
obel@mpimp-golm.mpg.de

Ordon, Frank
Federal Centre for Breeding Research on Cultivated Plants, Aschersleben
f.ordon@bafz.de

Parker, Jane
Max-Planck-Institute for Plant Breeding Research, Cologne
parker@mpiz-koeln.mpg.de

Pauly, Markus
Max-Planck-Institute of Molecular Plant Physiology
pauly@mpimp-golm.mpg.de

Pillen, Klaus
University of Bonn, Dept. of Crop Science and Plant Breeding, Bonn
k.pillen@uni-bonn.de

Potokina, Elena
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Prasad, Manoj
Institute of Plant Genetics and Crop Plant Research, Gatersleben

Presterl, Thomas
KWS SAAT AG, Einbeck
T.Presterl@kws.de

Radchuk, Volodymyr
Institute of Plant Genetics and Crop Plant Research, Gatersleben
radchukv@ipk-gatersleben.de

Schneider, Katharina
National Research Center for Environment and Health
Institute of Radiation Protection, Neuherberg
Katharina.Schneider@gsf.de

Schmid, K.
Max-Planck-Institute for Chemical Ecology, Jena
schmid@ice.mpg.de

Schulz, Britta
KWS SAAT AG, Einbeck
b.schulz@kws.de

Simon Rosin, Ulrike
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
simon@mpimp-golm.mpg.de

Serazetdinova, Liliya
Sainsbury Laboratory, John Innes Centre, Norwich, UK
liliya.serazetdiova@sainsbury-laboratory.ac.uk

Schmidt, Dagmar
TraitGenetics GmbH, Gatersleben

Soldatov, Aleksey
Max-Planck-Institute for Molecular Genetics, Berlin
soldatov@molgen.mpg.de

Sreenivasulu, Nese
Institute of Plant Genetics and Crop Plant Research, Gatersleben
srinivas@ipk-gatersleben.de

Stitt, Mark
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
mstitt@mpimp-golm.mpg.de

Stracke, Ralf
Max-Planck-Institute for Plant Breeding Research, Cologne

Susic, Zoran
University of Hohenheim, Stuttgart

Törjek, Otto
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
Toerjek@mpimp-golm.mpg.de

Weisshaar, Bernd
Bielefeld University, Dept. of Biology, Bielefeld
bernd.weisshaar@uni-bielefeld.de

Weschke, Winfriede
Institute of Plant Genetics and Crop Plant Research, Gatersleben
weschke@ipk-gatersleben.de

Weyen, Jens
SAATEN-UNION Resistenzlabor GmbH, Leopoldshöhe
weyen@saaten-union-labor.de

Wobus, Ulrich
Institute of Plant Genetics and Crop Plant Research, Gatersleben
wobusu@ipk-gatersleben.de

Wortmann, Heinrich
HYBRO GbR, Bad Schönborn
hybro.badschoenborn@t-online.de

Witt, Isabell
Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm
Witt@mpimp-golm.mpg.de

Zhang, Hanging
Institute of Plant Genetics and Crop Plant Research, Gatersleben