GABI – BEET: THE GERMAN SUGAR BEET GENOME INITIATIVE

GEORG KOCH

Strube-Dieckmann
A. Dieckmann-Heimburg, Kirchhorster Straße 16, 31688 Nienstädt, Germany

SUMMARY

GABI-BEET will provide new technologies for the molecular breeding of sugar beet. Marker technologies (RFLPs, RAPDs, SSRs, AFLPs) have already been employed successfully in practical beet breeding. Here, we aim to introduce new markers based on single nucleotide polymorphisms (SNPs) derived from both expressed genes and anonymous markers in combination with a high throughput system for marker analysis as a new tool for molecular breeding and research.

GABI-BEET is divided into four major projects: project A (sequencing of ESTs and bioinformatics) is the requirement for project B (high density mapping and allelic variation detected by SNP technologies). Project B will be achieved through the integration of conventional and functional markers. 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 (expressed sequence tags, ESTs) 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 polymorphisms (SNPs),

C) Comparison of the genome structure of wild and cultivated beet with respect to repetitive and mapped low-copy sequences,

D) Construction of a bacterial artificial chromosome (BAC) library from sugar beet genome.

The presentation will give an overview of the network of partners in the project, the goals and current status of the project.

ZUSAMMENFASSUNG

GABI-BEET wird neue Technologien für die molekulare Züchtung der Zuckerrübe zur Verfügung stellen. Verschiedene Marker-Technologien (RFLPs, RAPDs, SSRs, AFLPs) werden bereits erfolgreich in der praktischen Zuckerrübenzüchtung eingesetzt. Hier möchten wir neue Marker, basierend auf Einzelnukleotidpolymorphismen (SNPs) von exprimierten Genen und anonymen

Sequenzen, als neues Werkzeug der molekularen Züchtung und Forschung in Kombination mit einem Hochdurchsatz-System für die Markeranalyse einführen.

GABI-BEET ist unterteilt in vier große Projekte: Projekt A (Sequenzierung von ESTs und Bioinformatik) als Voraussetzung für Projekt B (hochdichte Kartierung und Detektion der allelischen Variation mittels SNP-Technologien) mit Integration konventioneller und funktioneller Marker. Dieser Ansatz wird vervollständigt durch eine vergleichende Kartierung und eine Genomanalyse mittels in situ Hybridisierung (Projekt C) und durch die Konstruktion einer repräsentativen Bibliothek mit großen Insertionen (Projekt D).

Das GABI-BEET Projekt ist fokussiert auf folgende Meilensteine:

A) Bereitstellung umfangreicher Sequenzinformationen von transkribierten Regionen (expressed sequence tags, ESTs) des Zuckerrübengenoms,

B) Erstellung einer integrierten Datenbank für molekulare und genetische Informationen und die Entwicklung von neuen Markersystemen basierend auf Einzelnukleotidpolymorphismen (SNPs),

C) Vergleich der Genomstruktur von Wildrüben und Zuckerrüben in Bezug auf repetitive und kartierte low-copy Sequenzen,

D) Aufbau einer Bibliothek aus artifiziellen Bakterienchromosomen (BACs) des Zuckerrübengenoms.

Die Präsentation wird einen Überblick vermitteln über das Netzwerk der Partner im Projekt, die Ziele und der gegenwärtige Status des Projektes.

ABRÉGÉ

Le programme GABI-BEET a pour but la production de nouveaux outils pour la sélection assistée par marqueurs moléculaires (SAM). Les RFLPs, RAPDs, SSRs et AFLPs ont déjà été utilisés avec succès en sélection de la betterave. L'objectif est d'introduire, pour la SAM et la recherche en général, un nouveau type de marqueurs : les SNPs dont le principe réside dans la détection de polymorphisme simple nucléotide, ceci à partir de gènes exprimés ou de marqueurs anonymes combinés avec un système complexe d'analyse de marqueur.

GABI-BEET est divisé en 4 projets principaux : le projet A (séquençage d'ESTs et bioinformatique) nécessaire au projet B (cartographie haute densité et variation allélique détectée par SNPs) ceci avec l'intégration de marqueurs conventionnels et fonctionnels. Cette approche est renforcée par la cartographie comparative et l'analyse des génomes par hybridation *in situ* (projet C) et la construction de banques représentatives à grands inserts (projet D).

Le programme GABI-BEET se déroule suivant les étapes suivantes :

A) Production en masse de séquences de régions transcrites (ESTs) du génome de la betterave,

B) Réalisation d'une banque de données intégrées contenant des informations moléculaires et génétiques et le développement nouveaux outils de marquage moléculaire (SNPs),

C) Comparaison de la structure des génomes de betteraves sauvages et cultivées en référence à l'utilisation de séquences répétées et de séquences faiblement répétées cartographiées,

D) Construction d'une banque BAC (Banque sur chromosome artificiel de bactérie) du génome de la betterave.

La présentation donnera une vue générale du réseau de partenaires participants aux projets, les objectifs et la situation actuelle des projets.

INTRODUCTION

GABI is an abbreviation for Genom Analysis in the Plant Biological System (www.gabi.de). This national genome programme collects and focuses all German plant genome projects funded by the BMBF, the German federal ministry of education and research. GABI phase I started end of 1998 and ends in 2003. Currently preparations for a second four year round of the project is underway. Joint projects with the French plant genome initiative, génoplante (www.genoplante.org) are already installed and will be extended in GABI II.

GABI I is focused to the model crops, *Arabidopsis thaliana* and barley, and a few other crops with model characters for specific traits and applications, e.g. rape seed, sugar beet, potato, wheat, rye, maize and poplar. Additionally, GABI I is funding resource centres to provide key technologies and materials and bioinformatics projects facilitating intelligent access to genomics data from within and from outside of the project. The funding of GABI I is approximately 50 Mio €. GABI II will continue with the basic ideas of GABI I but while applying more focus to bridging projects between the model plants and crops and placing more emphasis on applied research in crops with model traits of agronomic importance.

The funding structure of GABI touches two distinct areas of research: (I) Fundamental research is applied to research area I with up to 100 % federal funding; (II) Applied research is organised in research area II giving public institutes and life science companies more freedom to develop proprietary foreground rights related to the substantial contribution given by companies or institutes to the specific project with their background rights.

GABI-BEET is the only research area I project in GABI where companies are participating (www.gabi-beet.de). The partners in the project are given in Tab. 1.

Three further sugar beet projects has started so far in GABI research area II:

- GABI-BOLT: Cloning of the bolting gene B from sugar beet.
- SWEET-GABI: Crosslinking structural and functional genomics with regards to metabolism and development in sugar beet.
- BREATH-LESS GABI: Molecular physiology of storage organs.

All three projects are based in part on GABI-BEET by using tools and material developed in GABI-BEET, sharing material and genomic information and in some cases by overlap in stuff.

Tab. 1: Partner structure of GABI-BEET.

Institute of Crop Science and Plant Breeding, CAU Kiel	Prof. Dr. Christian Jung, coordinator
Subcontractors:	Dr. Lothar Frese
• Institute for Sugar Beet Research, IfZ, Göttingen	Prof. Dr. Bernward Märländer
• GABIinfo, GSF, Munich	Prof. Dr. Werner Mewes
MPI for Plant Breeding Research, Cologne	Dr. Katharina Schneider
• Subcontractor: ADIS	PD Dr. Bernd Weisshaar
Institute of Plant Breeding, MLU Halle	Prof. Dr. Eberhard Weber
TraitGenetics GmbH, Gatersleben	Dr. habil. Martin Ganal
KWS SAAT AG, Einbeck	Dr. Britta Schulz
Strube-Dieckmann, A. Dieckmann-Heimburg, Nienstädt	Dr. Georg Koch
Nordzucker AG, Braunschweig	Dr. Claudia Hemmerling

OBJECTIVES

GABI-BEET is a sugar beet project developing and providing major tools and access infrastructure to large scale genome analysis in sugar beet. Major objectives are

A) Supply of extensive sequence information of transcribed regions (*expressed sequence tags*, ESTs) 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 polymorphisms* (SNPs),

C) Comparison of the genome structure of wild and cultivated beet with respect to repetitive and mapped low-copy sequences,

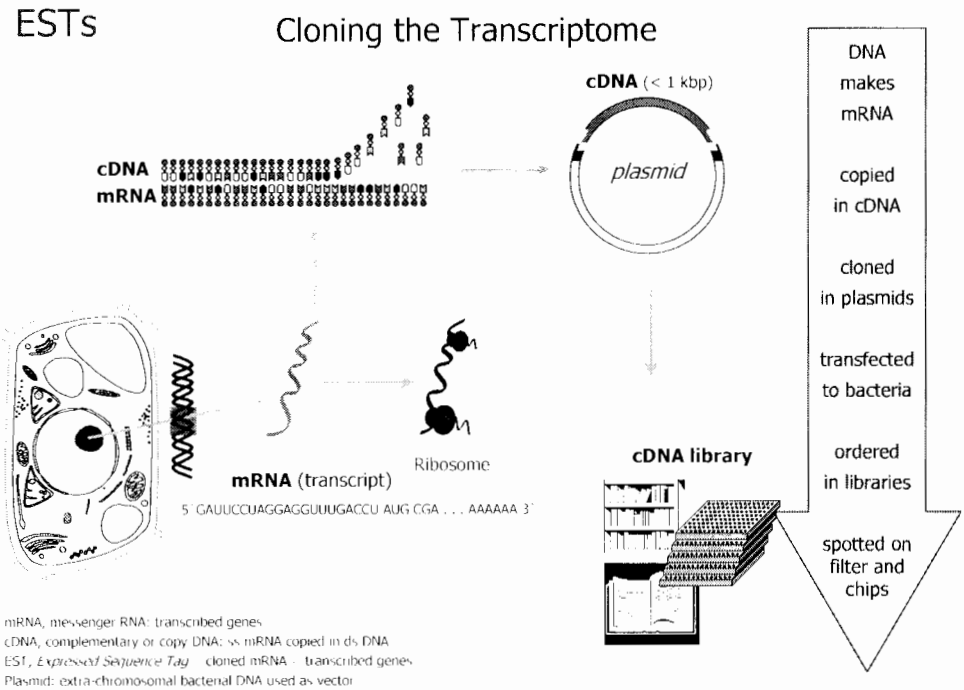
D) Construction of a *bacterial artificial chromosome* (BAC) library from sugar beet genome.

PROJECT A) UNRAVELLING THE TRANSCRIPTOME OF SUGAR BEET

The genome is build up by the linear sequence of the four nucleotides; A, T, C, G. The transcriptome describes the subset of sequence that is expressed via mRNAs, essentially the raw basic genes themselves. Putative gene fragments identified by sequencing such transcribed sequences are called ESTs, *expressed sequence tags*. ESTs are cloned in cDNA libraries (Fig. 1). The KWS project partner provided the consortium with 14,778 ESTs. Putatively unique sequences have been selected in the absence of sequence data from large collections of cDNAs by the oligonucleotide fingerprinting methodology (HERWIG *et al.* 2002). cDNA sequence clustering was achieved by correlating patterns of cross-hybridisation with a variety of oligonucleotide probes. The 14,778 ESTs for just over 10,000 distinct cDNAs were sequenced within GABI-BEET by ADIS

producing a unigene set of 9,730 sequences. This collection of oligonucleotide fingerprinted EST sequences is meanwhile publicly accessible in GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html).

Fig. 1: Cloning the transcriptome in a cDNA library.



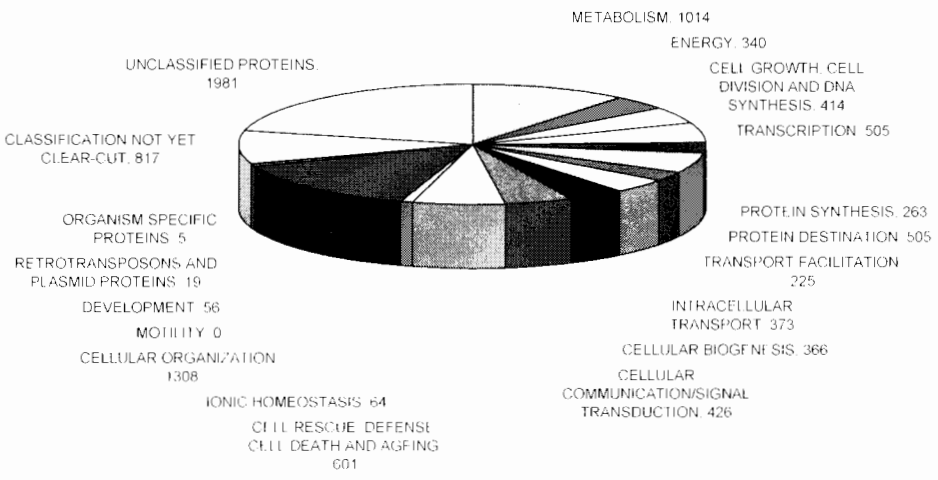
The GSF project partner was and remains one of the major contributors to annotation of the *Arabidopsis* genome. The functional catalogue (FunCat) developed by the group provides a resource that can be used to transfer sequence annotation between heterologous sequences. Based on the bioinformatic tools developed during the *Arabidopsis* genome project GSF was able to annotate putative functions to the ESTs of the unigene set and grouped them according to different metabolic pathways of the plant cell. Fig. 2 shows the distribution of the annotated functions of the unigene ESTs.

The functional annotation of the beet ESTs is available from the MIPS website (http://mips.gsf.de/proj/sputnik/).

PROJECT B) TOWARDS A FUNCTIONAL MAP OF THE SUGAR BEET GENOME

Molecular markers are the technological basis of molecular plant breeding. Molecular markers are landmarks on a genetic map of anonymous or preferable on genes of known function, e.g. candidate genes for agronomically important traits. Previously mapped RFLP markers serve as the framework on which the newly characterized functionally annotated ESTs can be mapped. According to the functional classification, each of three partners will map to the sugar beet

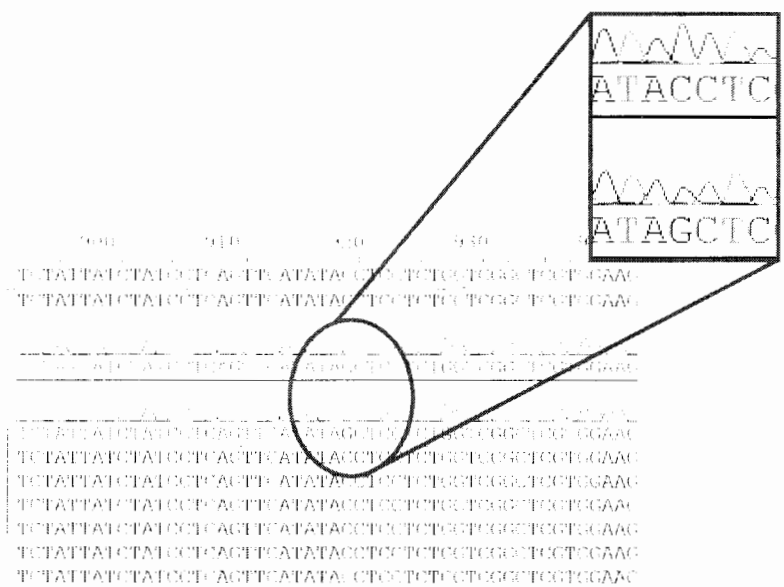
Fig. 2: Distribution of the annotated functions of the unigene EST set categorized according different functional compartments of the plant cell.



genetic genome an equal proportion of selected 1,000 ESTs. Partner CAU will map ESTs with a putative function in pathogen resistance or genes related to known resistance genes, so called *resistance gene analogues* or RGAs, TraitGenetics is focused towards stress-related sequences, and MPI selected candidate genes involved in carbohydrate metabolism and beet development.

Several approaches to map these ESTs onto the sugar beet map are possible. The most advanced and sophisticated technology uses SNPs, *single nucleotide polymorphisms*. Fig. 3 shows an example of such a SNP.

Fig. 3: Single nucleotide polymorphisms are the basis of SNP markers.

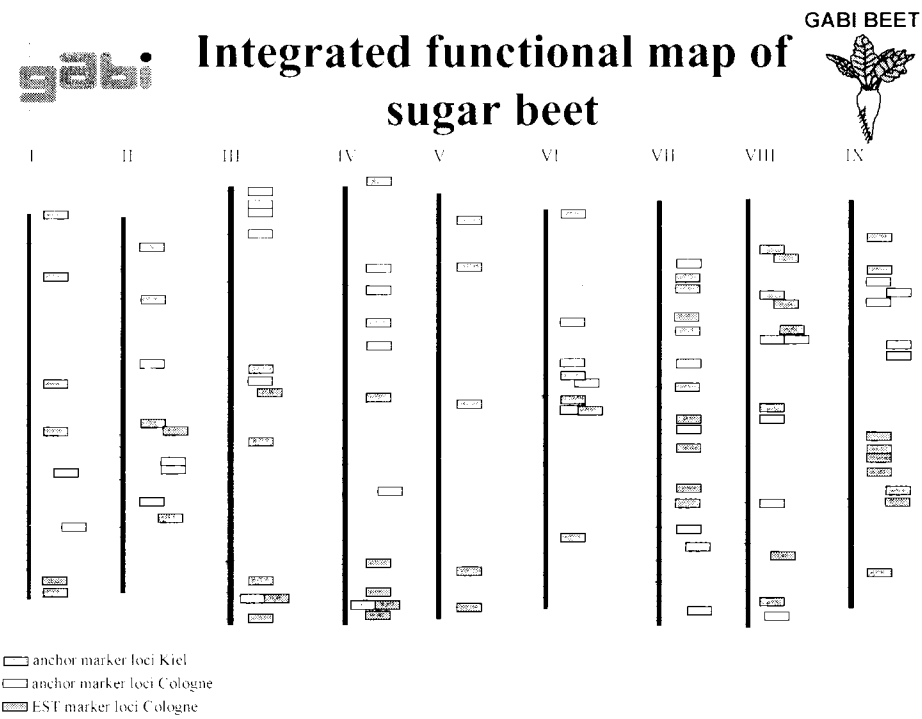


A SNP is a single base pair difference between two sequences derived of the same genetic locus, e.g. a gene, from two different plants. Fig. 3 shows part of a sequence from the same locus of 9 different plants differing by only one base in certain plants, a G/C SNP. GABI-BEET has adopted a panel of 16 different plants provided by the breeders that includes wild beets. These sugar beets largely represent the sugar beet gene pool and provide the opportunity to uncover many of the SNPs that exist in sugar beet. Sequencing of all these plants for each of the 1,000 targets, ESTs and some RFLP framework markers, is a laborious and time consuming task but will be finished by end of 2003. More than 400,000 bases have been sequenced and analysed already and more than 1,500 SNPs detected; approx. 1 SNP every 70 base pairs.

SNP mapping is another milestone hard to achieve. The specific advantages of using SNPs as the markers of the future are the potential of highly automated and high-throughput marker detection systems like MALDI-TOF, a chip-based method or the various very efficient primer-extension based scoring methods. Partner MPI will screen the available methods and make them available to the consortium. Model studies are already finished and large scale mapping can be started soon. However, only a small fraction of the detected SNPs are already mapped. The vast majority of the ESTs will be mapped as final task at the end of the project. The first 50 SNPs are mapped on the growing functional map of the sugar beet shown in Fig. 4.

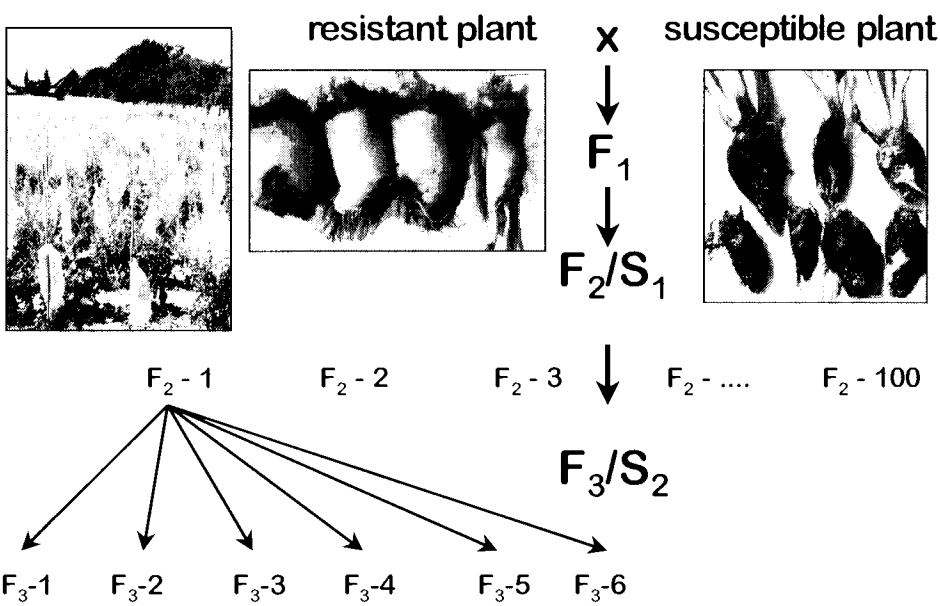
Progress reports were presented at Plant and Animal Genome Conferences 2003 and 2002 (www.pag-intl.org). A first publication out of the GABI-BEET consortium towards a functional map of the sugar beet genome was published by HUNGER *et al.* (2003).

Fig. 4: An integrated functional map of the sugar beet genome.



Touch stone of the SNP development and especially RGA mapping will be the mapping of the crown and root rot resistance genes. Disease resistance against the fungus *Rhizoctonia solani* is medium complex inherited. Several resistance genes may be involved but less than four major genes are expected (PANELLA & RUPPEL in SNEH *et al.*, 1996). Strube-Dieckmann produced several segregating plant populations for the specific needs of the GABI-BEET project (Fig. 5) and selected the most resistant one for the project (Fig. 6). Seed were handed over to IfZ that applied a newly developed, highly sophisticated bioassay to provide precise and reliable phenotypic scores for mapping the trait onto the functional map. Finally mapped RGAs can be proved as candidate genes for the mapped resistance QTL, *quantitative trait loci*. At present CAU is providing MLU with the corresponding marker information to fill up the existing map and to unravel *Rhizoctonia* resistance.

Fig. 5: Development of a mapping population segregating for Rhizoctonia *resistance at Strube-Dieckmann.*



PROJECT C) COMPARISON OF THE GENOME STRUCTURE OF WILD AND CULTIVATED BEET

The genome of *B. vulgaris* consists of approximately 60 % repeated DNA. The aim of this project performed by the CAU project partner is the identification, characterization and chromosomal localization of repetitive sequences. The experimental work was focussed on the following research activities:

LTR Retrotransposons. The sugar beet Ty1-*copia* element Tbv1 (5.2 kb) is flanked by LTRs of 584-614 bp. The variability of the LTR indicates a considerable divergence of Tbv1 within the *B. vulgaris* genome which is in line with the amplification and wide-spread distribution within *Beta* species. Fluorescent *in situ* hybridisation (FISH) revealed that Tbv1 is present on all *B. vulgaris* chromosomes and scattered throughout the genome (Fig. 7).

***Fig. 6**: Phenotypic characterization of three segregating F_2 populations produced for mapping* Rhizoctonia *resistance – selection of the best suited population for mapping in GABI-BEET at Strube-Dieckmann.*

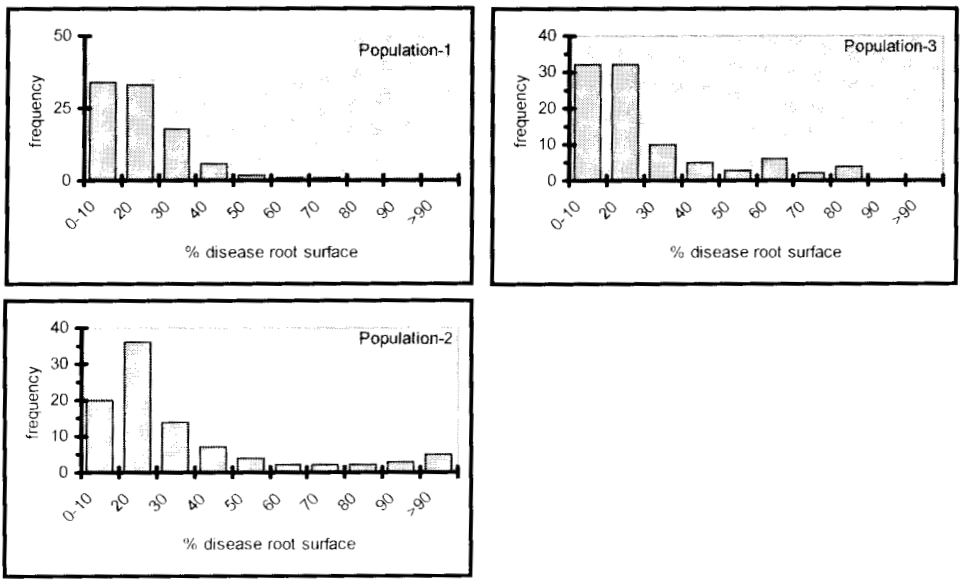
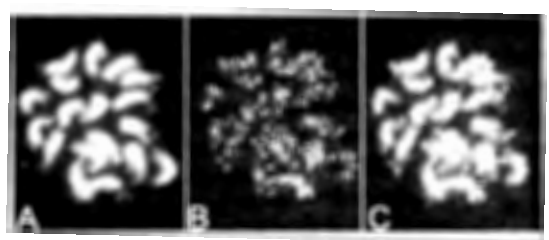


***Fig. 7**: Chromosomal localization of Tbv1.*

A) *Metaphase chromosomes of* B. vulgaris *(DAPI)*

B) *Yellow-green hybridization signals of Tbv1*

C) *Overlay*

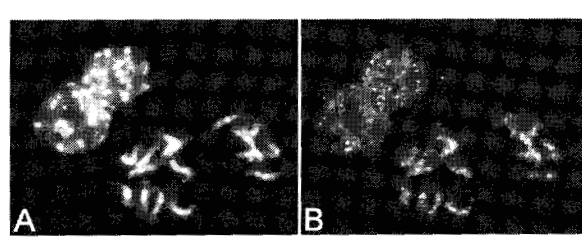


Class II Transposons. PCR assays were developed to detect *En/Spm*-like transposons in species of the genus *Beta*. Resulting PCR products have been cloned from *B. vulgaris*, *B. corolliflora*, *B. procumbens* and *B. nana*. Sequence analyses of 20 *En/Spm* fragments showed that these transposons are conserved (71.9 - 98.7 % identity at DNA level). Amplification of *En/Spm*-like transposons can also be demonstrated by FISH. *En/Spm*-like transposons are highly amplified and dispersed on *B. vulgaris* chromosomes (Fig. 8).

***Fig. 8:** Chromosomal distribution of En/Spm.*

A) *DAPI stained interphase and prometaphases of B. vulgaris*

B) *The same chromosomes after FISH with an* En/Spm *transposase fragment*



Physical mapping of BACs. Towards the integration of genetic and physical maps, the CAU project partner used their expertise in FISH to provide a technical platform for the physical mapping of candidate BAC clones, cf. Project D) above. In a pilot study in cooperation with GABI-BOLT CAU performed the physical mapping of three BACs around the bolting gene B by FISH (Fig. 9).

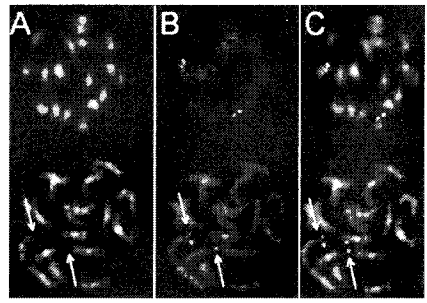


Fig. 9: BAC-FISH of a clone from a chromosomal region of the gene B

A) DAPI stained metaphase and prometaphase of B. vulgaris
B) Yellow-green signals of BAC (arrowed)
C) Overlay

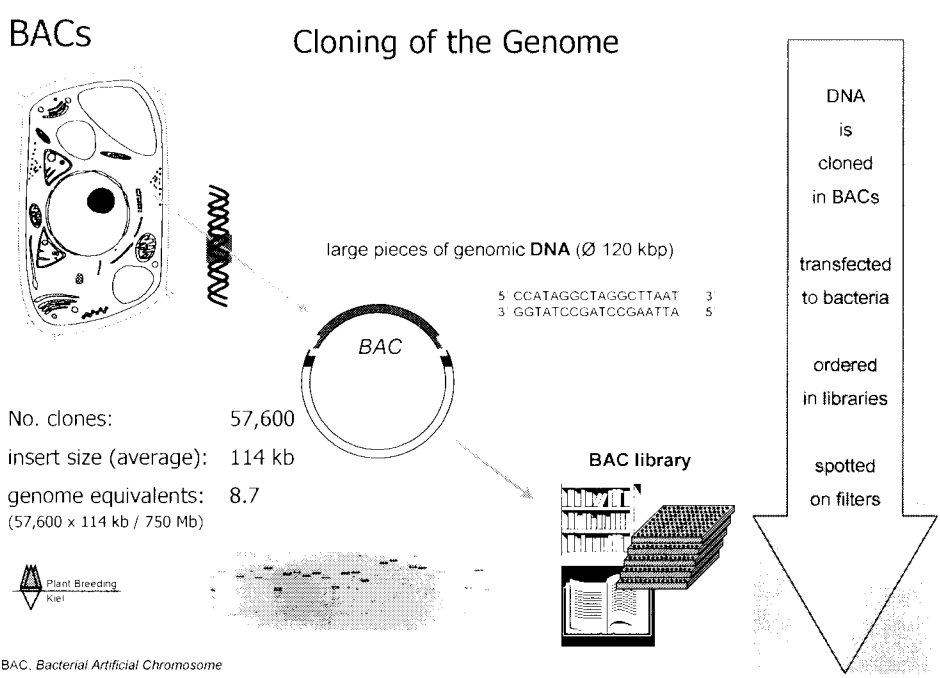
Fig. 7, 8 and 9 are courtesy of T. Schmidt; manuscript in preparation

Plant Breeding
Kiel

PROJECT D) CONSTRUCTION OF A BACTERIAL ARTIFICIAL CHROMOSOME (BAC) LIBRARY FROM SUGAR BEET GENOME

Bacterial artificial chromosomes, BACs, are molecular vectors to clone large DNA inserts. They are built from both plasmid and bacterial genetic elements. They are integrated and inherited in bacteria as wild type chromosomes. This

Fig. 10: Cloning the genome in BACs, bacterial artificial chromosomes.



BAC, *Bacterial Artificial Chromosome*

tool provides the means to clone complete plant genomes in manageable numbers of clones with realistic insert fragment sizes. Sugar beet DNA was cut into pieces using restriction enzymes and randomly cloned into such BACs. It is viable to store frozen BACs for the long-term in large arrays. In parallel BAC DNA was spotted on very dense ordered filter membrane grids for further hybridisation, characterization and identification (Fig. 10).

Partners CAU and TraitGenetics have put together their expertise and equipment to produce a genomic BAC library of 57,600 clones with an average insert size of 114 kbp (8,7 genome equivalents; HOHMANN *et al.* submitted). The BAC-library exists in multiple copies and is already in use by the GABI consortium (Fig. 11).

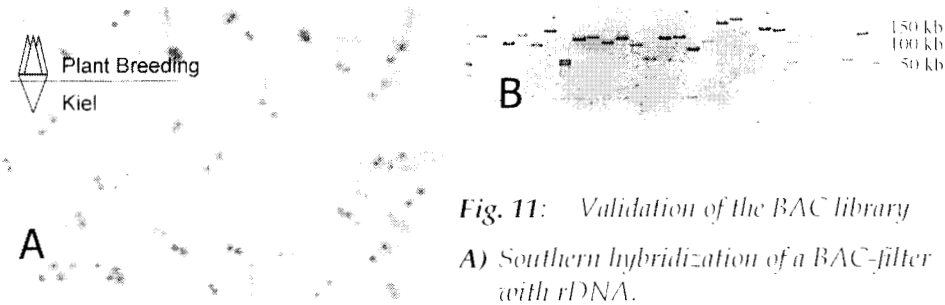


Fig. 11: Validation of the BAC library

A) Southern hybridization of a BAC-filter with rDNA.

B) Fragment size separation of random BAC inserts of genomic sugar beet DNA.

ACKNOWLEDGEMENTS

The GABI-BEET project is supported by German ministry BMBF (GABI grant 0312283) and the involved companies. GABI-BEET is a consortium driven project. All partners have contributed irregardless whether they are mentioned in this paper or not. I acknowledge to all the groups and group members their specific and unique contributions. Without their help this paper would not be made possible. Many thanks to Stephen Rudd who corrected the English and Bruno Desprez for the French translation.

REFERENCES

1. HERWIG, R. *et al.*: Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes. *The Plant Journal*, 32, 845-857, 2002.

2. HUNGER, S. *et al.*: Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris* L.). *Genome*, 46, 70-82, 2003.

3. PANELLA, L. and E. RUPPEL in SNEH *et al.* (eds.): *Rhizoctonia* species. Kluwer 1996.

4. HOHMANN, U. *et al.*: A bacterial artificial chromosome (BAC) library of sugar beet and a physical map comprising the bolting gene *B*. submitted.