GabiPD: The GABI Primary Database - a plant integrative ‘omics’ database

Diego Mauricio Riaño-Pachón[1,2], *Axel Nagel*[1], *Jost Neigenfind*[1], *Robert Wagner*[1], *Rico Basekow*[1], *Elke Weber*[1], *Bernd Mueller-Roeber*[2], *Svenja Diehl*[3], *Birgit Kersten*[1,2,§]

[1] GabiPD team, Bioinformatics group, Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.

[2] Department of Molecular Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, Haus 20, 14476 Potsdam-Golm, Germany.

[3] Former RZPD German Resource Center for Genome Research GmbH, Berlin, Germany.

[§] Corresponding author, kersten@mpimp-golm.mpg.de; Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

ABSTRACT

The GABI Primary Database, GabiPD (http://www.gabipd.org/), was established in the frame of the German initiative for Genome Analysis of the Plant Biological System (GABI). The goal of GabiPD is to collect, integrate, analyse and visualise primary information from GABI projects. GabiPD constitutes a repository and analysis platform for a wide array of heterogeneous data from high-throughput experiments in several plant species. Data from different ‘omics’ fronts are incorporated (i.e., genomics, transcriptomics, proteomics and metabolomics), originating from 14 different model or crop species. We have developed the concept of GreenCards for text based retrieval of all data types in GabiPD (e.g., clones, genes, mutant lines). All data types point to a central Gene GreenCard, where gene information is integrated from genome projects or NCBI UniGene sets. The centralised Gene GreenCard allows visualising ESTs aligned to annotated transcripts as well as displaying identified protein domains and gene structure. Moreover GabiPD makes available interactive genetic maps from potato and barley, and 2DE-gels from *Arabidopsis thaliana* and *Brassica napus.* Gene expression and metabolic profiling data can be visualised through MapManWeb. By the integration of complex data in a framework of existing knowledge, GabiPD provides new insights and allows for new interpretations of the data.

INTRODUCTION

Experimental studies in the post genomic era generate a very large amount of data from high-throughput experiments on biological systems. Current studies include, among others, expression and metabolite profiles, proteome and interaction data (e.g., DNA-protein and protein-protein interactions), collected at different space and time scales. This increasing flow of data requires computational systems that, beside managing efficiently the enormous quantity of data, are capable of integrating and displaying these disparate data collections in a meaningful and amicable way. We have developed the GABI Primary Database, GabiPD, in order to fulfil these requirements.

GabiPD is a web accessible database that was developed in the frame of the German initiative for Genome Analysis of the Plant Biological System (Genomanalyse im biologischen System Pflanze, GABI). GabiPD allows a seamless integration of varied ‘omics’ data types obtained from plant systems and follows the MIAME (1) and MIAMET (2) standards for storing gene expression and metabolic profiling data, respectively. Its flexible design allows for a high level of data integration, and eases cross-referencing the different GabiPD data types among each other (e.g., mapping information, sequences and SNP data, 2DE-gel images and protein information) and to public gene/protein-specific information which in turn provides the users a comprehensive overview of the available information for their particular gene or protein of interest. The integration with genome databases like TAIR (3) and general nucleotide databases like GenBank, as well as cross-links to secondary databases, such as ARAMEMNON (4), PlnTFDB (5), GABI-KAT (6), PhosPhAt (7), and ProMEX (8) further increase the usefulness of GabiPD.

METHODS AND CONTENTS

DESIGN AND IMPLEMENTATION

GabiPD’s web interface was developed using Perl and Java in combination with template processing to separate the visualisation from the application logic. Our applications are database driven, which means that the application interface logic is derived directly from the database structure (shown in blue in Figure 1). To achieve this, we deploy reverse engineering methods in combination with template processing to generate interfaces to programming languages like Perl or Java, thus supporting all the database-specific actions like ‘insert‘, ‘update’, ‘delete’ or ‘select’. These object-oriented interfaces are automatically generated from the database schema, supporting inheritance, automated key generation and advanced exception handling. The separation of database application interface and application logic (shown in yellow in Figure 1) facilitates fast adjustment to modifications of the data structure and diminishes the efforts on fixing existing application logic during larger database changes.

GabiPD CONTENT AND GENE-CENTRIC VIEWS

Currently GabiPD includes data originating from 14 different angiosperm species representing the most important lineages in the flowering plants (see Figure 2). *Arabidopsis thaliana* is the most widely represented model species, followed by the crop plants *Solanum tuberosum* (potato) and *Hordeum vulgare* (barley). In GabiPD, genomic, transcriptomic, proteomic and metabolomic data are integrated from those species. Genomic data comprise mapping information, sequences and SNP/InDel information. Transcriptomics is represented by a large number of ESTs and corresponding sequence trace files. ESTs are further analysed by BLAST and ORF analysis. For barley, in addition, EST clustering results and corresponding information on a new 27K unigene set are accessible and downloadable. As a type of proteomic data, annotated 2DE-gel images from *Arabidopsis thaliana* and *Brassica napus* are integrated. Moreover, transcript and metabolite profiling data are provided via MapManWeb.

Most entries of all GabiPD data types are pointing to the central Gene GreenCard and vice versa (see Figure 3 and next section). In the Gene GreenCard, gene information from genome annotation projects or NCBI UniGene sets is integrated and useful links to secondary databases are provided. Currently the genome annotation (TAIR version 7.0) for *A. thaliana* (3) is integrated. Annotations for other sequenced species will follow. In order to ease the transfer of knowledge from sequenced to non-sequenced species, i.e., crop plants, we have performed similarity-based mappings between closely related species, i.e., *Arabidopsis* and *Brassica* spp.

QUERYING THE DATABASE

We have developed the concept of **GreenCards** as a central entrance for text-based data queries and visualisation which grant public as well as credentials-based access to the integrated data in GabiPD. **GreenCards** enable users to comprehensively query GabiPD by genotype name, marker or gene name, keyword or GenBank sequence accession number. Searches can be restricted to selected species or data types, while wildcards can be used to broaden the scope of the query. The result of a **GreenCard** search is presented as a list of hits with links to complete descriptions, i.e., GreenCards. Figure 3 shows an example of this type of search, where the user had entered the gene name *'FLOWERING LOCUS T'* as a search term. This search retrieves, among others, an Arabidopsis **Gene GreenCard** (gene: AT1G65480.1) corresponding to the genome annotation project and **Plant GreenCards** representing mutant plant lines, e.g., plant 290E08, that has BLAST hits to the gene AT1G65480.1 and with seeds available from GABI-KAT (6). Moreover, several **Clone GreenCards** of cDNA clones (e.g., clone: MPMGp2011E01215) in which the keyword is found were retrieved by the search. A more strict relationship between the Gene GreenCard and the Plant and Clone GreenCards is established by similarity-based searches. The best BLAST hit of the sequence, e.g., relationship to a cDNA clone or a mutant plant line, appears in the section **Related with**, and the users can go from the clone or the plant line to the associated gene or vice versa.

Furthermore, the **Gene GreenCard**, which displays information from genome annotation projects and NCBI UniGene sets, has been extended to include links to secondary databases, such as ARAMEMNON (4), GABI-KAT (6), and ProMEX (8). Additionally, schematic representations of gene sequence features are provided to highlight protein domains identified using the latest PFAM library (9), exon-exon borders and untranslated regions (UTRs) identified by the genome annotation projects (Figure 3). These features are displayed onto a representation of the cDNA sequence.

Alternatively, users can enter their own amino acid or nucleotide sequence to identify by a BLAST search (10) similar sequences integrated in GabiPD.

In addition to the GreenCard and BLAST search functionality, users can browse and search the genetic maps and 2DE-gels stored in GabiPD, through specifically designed visualisation tools: (i) 2DEGelViewer by which 2DE-gel images can be viewed in an interactive way, which allows retrieving extra information on 2D-spots as identified by mass spectrometry (Figure 3); (ii) genetic mapping data can be visualised using YAMB (Yet Another Map Browser; Figure 4) with the possibility to view details on all mapped elements (11); (iii) MapManWeb (Figure 5), allows the visualisation and extraction of relevant information from transcript and metabolite profiling data and the graphical mapping of such data onto diagrams of metabolic pathways and other biological processes (12); and (iv) an extended version of JTrev (13) allows the display of sequence traces with integrated SNP information.

The GabiPD project page serves as an additional gateway to specific data by providing project specific views, such as BreedCAM or PoMaMo (11), where potato genomic data and Solanaceae function maps for pathogen resistance are accessible.

ADDITIONAL TOOLS AVAILABLE FROM GabiPD

In addition to the data and data visualisation available from our site, the newest versions of the following tools are made available for download:

MapMan - desktop version: a user-driven software tool that displays large datasets (e.g., gene expression data from Arabidopsis Affymetrix arrays) onto diagrams of biological processes, such as metabolic pathways (12).

SATlotyper: a software tool designed for inferring haplotypes and phased genotypes from unphased SNP data for polyploid and polyallelic heterozygous populations (14).

FUTURE DIRECTIONS

The presentation of a wide spectrum of different plant species in GabiPD paves the way for cross-species comparisons that are facilitated by the availability of BLAST hits between the GabiPD sequences and plant NCBI UniGene sets. To ease the transfer of knowledge from sequenced to non-sequenced plant species, the genome annotations of *Oryza sativa*, *Populus trichocarpa* and *Vitis vinifera* will be added and mapped to closely related species in the near future. By that useful information about orthologous genes will be included for cross-species studies. Moreover, we will extend our WebServices to provide programmatic access to multiple data types for all plant species in GabiPD.

FUNDING

This work was supported by the German Ministry for Education and Research (BMBF) [GABI I: 0312272, GABI II: 0313112 and GABI-FUTURE: 0315046], the former German Resource Center for Genome Research (RZPD) GmbH and the Max Planck Society.

ACKNOWLEDGEMENTS

The GabiPD team thanks the GABI and the WPG (Wirtschaftsverbund Pflanzengenomforschung GABI e.V.) community for providing data and supporting the continuous development of the database. Dr. Björn Usadel is acknowledged for helpful discussions and his support in the development of MapManWeb. We wish to thank Dr. Patrick Schweitzer, Dr. Lothar Altschmied, Dr. Uwe Scholz and Dr. Nils Stein for the collaboration in the generation of the new 27K barley unigene set, and Dr. Kathryn F. Beal for sharing the source code of the trace viewer JTrev. Özgür Demir and Sebastian Köhler are acknowledged for further developments of user interfaces and for data integration.

REFERENCES

1. Ball, C.A. and Brazma, A. (2006) MGED standards: work in progress. *Omics*, **10**, 138-144.
2. Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R. *et al.* (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol*, **22**, 1601-1606.

3. Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L. *et al.* (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res*, **36**, D1009-D1014.
4. Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flugge, U.I. and Kunze, R. (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol*, **131**, 16-26.
5. Riaño-Pachón, D.M., Ruzicic, S., Dreyer, I. and Mueller-Roeber, B. (2007) PlnTFDB: An integrative plant transcription factor database. *BMC Bioinformatics*, **8**, 42.
6. Li, Y., Rosso, M.G., Viehoever, P. and Weisshaar, B. (2007) GABI-Kat SimpleSearch: an *Arabidopsis thaliana* T-DNA mutant database with detailed information for confirmed insertions. *Nucleic Acids Res*, **35**, D874-878.
7. Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D. and Schulze, W.X. (2008) PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res*, **36**, D1015-1021.
8. Hummel, J., Niemann, M., Wienkoop, S., Schulze, W., Steinhauser, D., Selbig, J., Walther, D. and Weckwerth, W. (2007) ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics*, **8**, 216.
9. Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res*, **36**, D281-288.
10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.
11. Meyer, S., Nagel, A. and Gebhardt, C. (2005) PoMaMo-a comprehensive database for potato genome data. *Nucleic Acids Res*, **33**, D666-670.
12. Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Piques, M.C., Steinhauser, D. *et al.* (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol*, **138**, 1195-1204.
13. Bonfield, J.K., Beal, K.F., Betts, M.J. and Staden, R. (2002) Trev: a DNA trace editor and viewer. *Bioinformatics*, **18**, 194-195.
14. Neigenfind, J., Gyetvai, G., Basekow, R., Diehl, S., Achenbach, U., Gebhardt, C., Selbig, J. and Kersten, B. (2008) Haplotype inference from unphased SNP data in heterozygous polyploids based on SAT. *BMC Genomics*, **9**, 356.
15. Soltis, P.S. and Soltis, D.E. (2004) The origin and diversification of angiosperms. *Am J Bot*, **91**, 1614-1626.
16. Angiosperm Phylogeny Group. (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc*, **141**, 399-436.
17. Knapp, S. (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J Exp Bot*, **53**, 2001-2022.
18. Stein, N., Prasad, M., Scholz, U., Thiel, T., Zhang, H., Wolf, M., Kota, R., Varshney, R.K., Perovic, D., Grosse, I. *et al.* (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. *Theor Appl Genet*, **114**, 823-839.

FIGURES

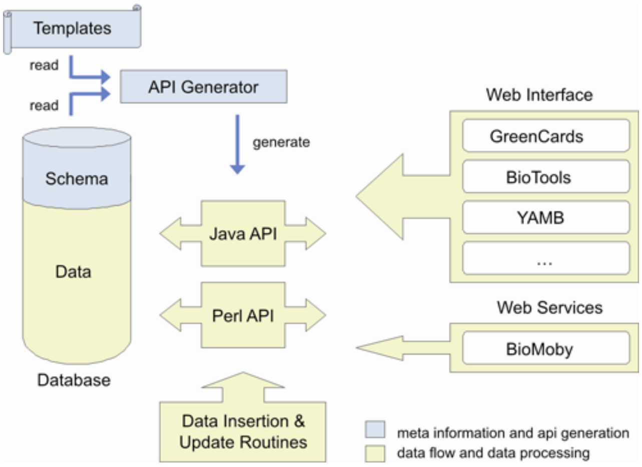


Figure 1. Schematic overview of the GabiPD application structure. The ´API generator´ translates the ´Templates´ using the database meta-information (shown in blue), generating the database application interface (Java and Perl API). The application logic (shown in yellow), i.e., Web Interface, WebServices and data manipulation routines, interacts with the ´Database´ through the database application interface.

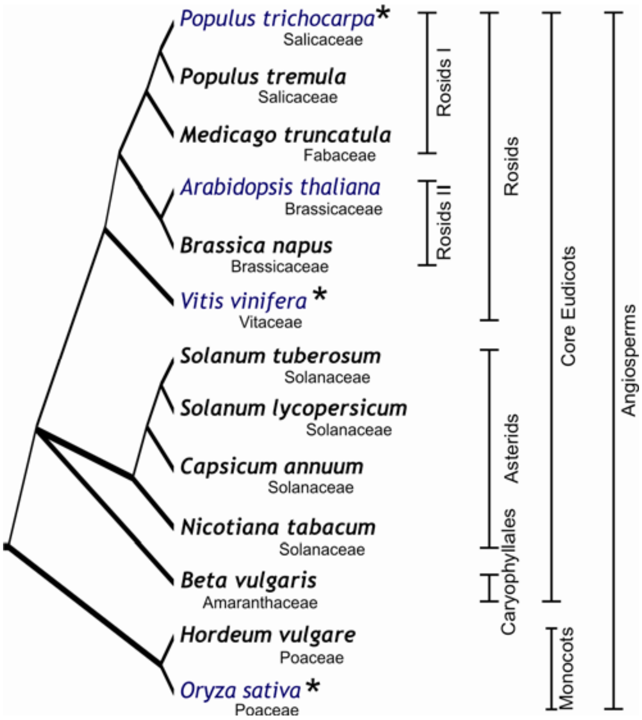
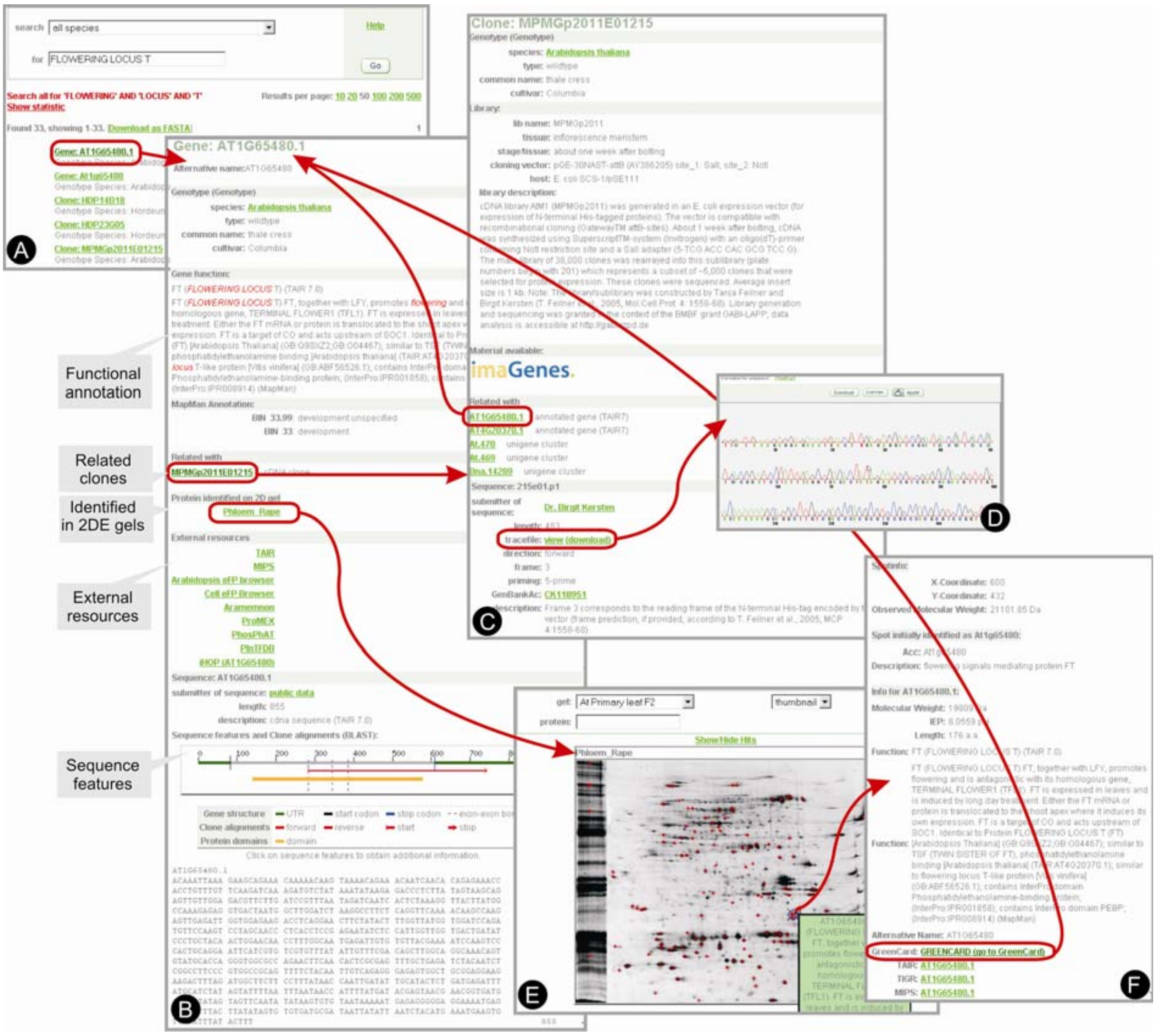


Figure 2. Phylogenetic tree depicting the evolutionary relationships among the species represented in GabiPD (15-17). Species for which whole-genome sequences and annotations are available are shown in blue. *Species that will soon be integrated in GabiPD. Species not shown: *Solanum bulbocastanum*, *S. demissum*, *S. phureja*, *S. spegazzinii*.



search: all species
for: FLOWERING LOCUS T
Search all for 'FLOWERING' AND 'LOCUS' AND 'T'
Show statistic
Results per page: 10 20 50 100 200 500
Found 33, showing 1-33. (Download as FASTA)

Gene: AT1G65480.1
Gene: At1g65480 — Genotype Species: Arabidopsis
Clone: HDP14G18 — Genotype Species: Hordeum
Clone: HDP23G05 — Genotype Species: Hordeum
Clone: MPMGp2011E01215 — Genotype Species: Arabidopsis

Gene: AT1G65480.1

Alternative name: AT1G65480

Genotype (Genotype)
species: Arabidopsis thaliana
type: wildtype
common name: thale cress
cultivar: Columbia

Gene function:
FT (FLOWERING LOCUS T) (TAIR 7.0)
FT (FLOWERING LOCUS T) FT, together with LFY, promotes flowering and … homologous gene, TERMINAL FLOWER1 (TFL1). FT is expressed in leaves … treatment. Either the FT mRNA or protein is translocated to the shoot apex … expression. FT is a target of CO and acts upstream of SOC1. Identical to Pr… (FT) [Arabidopsis Thaliana] (GB:Q9SXZ2;GB:O04467); similar to TSF (TWIN… phosphatidylethanolamine binding [Arabidopsis thaliana] (TAIR:AT4G20370… locus T-like protein [Vitis vinifera] (GB:ABF56526.1); contains InterPro doma… Phosphatidylethanolamine-binding protein; (InterPro:IPR001858); contains … (InterPro:IPR008914) (MapMan)

MapMan Annotation:
BIN 33.99: development unspecified
BIN 33: development

Related with
MPMGp2011E01215 cDNA clone

Protein identified on 2D gel
Phloem_Rape

External resources
TAIR
MIPS
Arabidopsis eFP browser
Cell eFP Browser
Aramemnon
ProMEX
PhosPhAT
PlnTFDB
iHOP (AT1G65480)

Sequence: AT1G65480.1
submitter of sequence: public data
length: 855
description: cdna sequence (TAIR 7.0)

Sequence features and Clone alignments (BLAST):
0 100 200 300 400 500 600 700

Gene structure: UTR, start codon, stop codon, exon-exon bo…
Clone alignments: forward, reverse, start, stop
Protein domains: domain
Click on sequence features to obtain additional information.

Clone: MPMGp2011E01215

Genotype (Genotype)
species: Arabidopsis thaliana
type: wildtype
common name: thale cress
cultivar: Columbia

Library:
lib name: MPMGp2011
tissue: inflorescence meristem
stage/tissue: about one week after bolting
cloning vector: pQE-30NAST-attB (AY386205) site_1: SalI, site_2: NotI
host: E. coli SCS-1/pSE111

library description:
cDNA library AM1 (MPMGp2011) was generated in an E. coli expression vector (for expression of N-terminal His-tagged proteins). The vector is compatible with recombinational cloning (GatewayTM attB-sites). About 1 week after bolting, cDNA was synthesized using SuperscriptTM-system (Invitrogen) with an oligo(dT)-primer containing NotI restriction site and a SalI adapter (5-TCG ACC CAC GCG TCC G). The main library of 38,000 clones was rearrayed into this sublibrary (plate numbers begin with 201) which represents a subset of ~5,000 clones that were selected for protein expression. These clones were sequenced. Average insert size is 1 kb. Note: The library/sublibrary was constructed by Tanja Feilner and Birgit Kersten (T. Feilner et al., 2005, Mol Cell Prot. 4: 1558-68). Library generation and sequencing was granted in the context of the BMBF grant GABI-LAPP; data analysis is accessible at http://gabi.rzpd.de

Material available:

Related with
AT1G65480.1 annotated gene (TAIR7)
AT4G20370.1 annotated gene (TAIR7)
At.470 unigene cluster
At.469 unigene cluster
Bna.14299 unigene cluster

Sequence: 215e01.p1
submitter of sequence: Dr. Birgit Kersten
length: 483
tracefile: view (download)
direction: forward
frame: 3
priming: 5-prime
GenBankAc: CK118951
description: Frame 3 corresponds to the reading frame of the N-terminal His-tag encoded by t… vector (frame prediction, if provided, according to T. Feilner et al., 2005, MCP 4:1558-68)

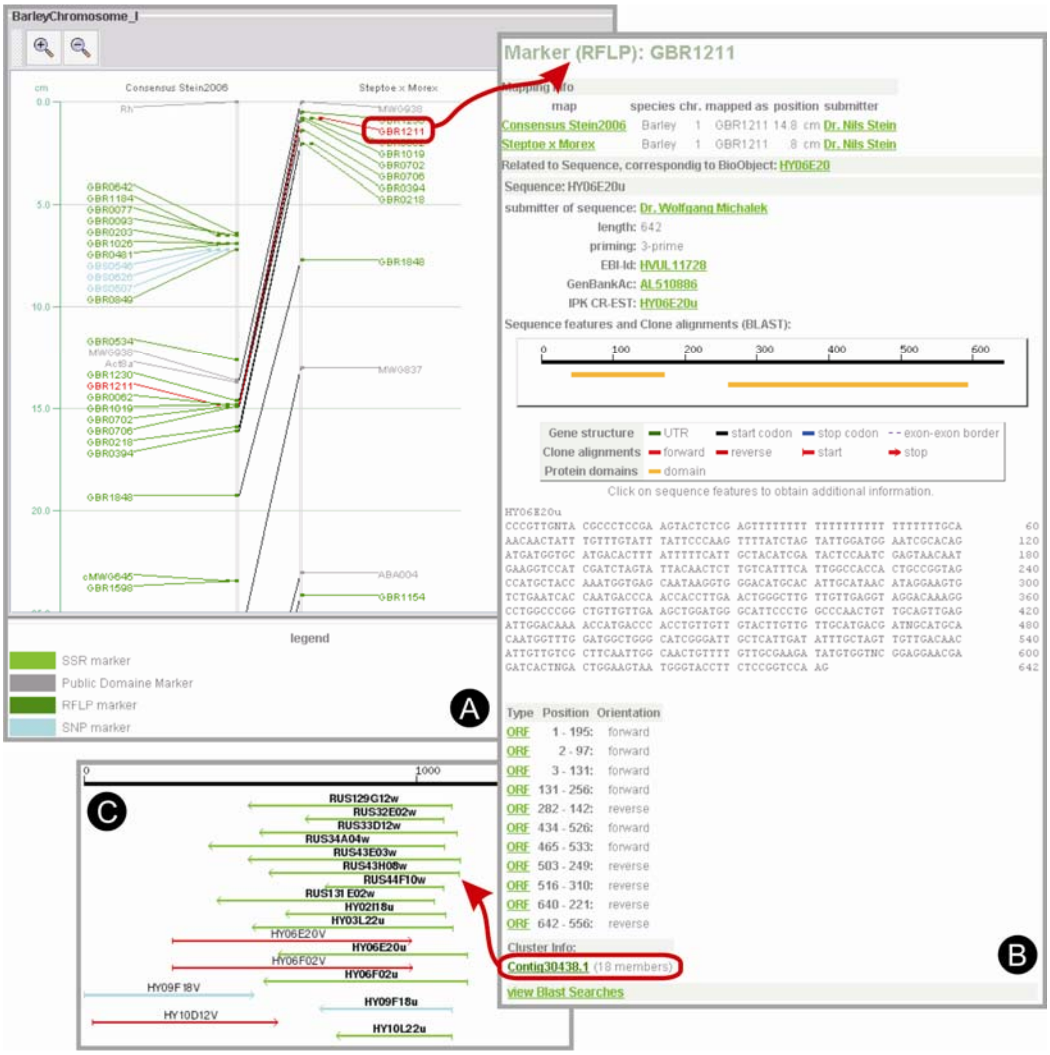
gel: At Primary leaf F2
thumbnail
protein:

Spotinfo:
X-Coordinate: 600
Y-Coordinate: 432
Observed Molecular Weight: 21101.85 Da

Spot initially identified as At1g65480:
Acc: At1g65480
Description: flowering signals mediating protein FT

Info for AT1G65480.1:
Molecular Weight: 19909 Da
IEP: 8.0559 pH
Length: 176 a.a.
Function: FT (FLOWERING LOCUS T) (TAIR 7.0)
Function: FT (FLOWERING LOCUS T) FT, together with LFY, promotes flowering and is antagonistic with its homologous gene, TERMINAL FLOWER1 (TFL1). FT is expressed in leaves and is induced by long day treatment. Either the FT mRNA or protein is translocated to the shoot apex where it induces its own expression. FT is a target of CO and acts upstream of SOC1. Identical to Protein FLOWERING LOCUS T (FT) [Arabidopsis Thaliana] (GB:Q9SXZ2;GB:O04467); similar to TSF (TWIN SISTER OF FT); phosphatidylethanolamine binding [Arabidopsis thaliana] (TAIR:AT4G20370.1); similar to flowering locus T-like protein [Vitis vinifera] (GB:ABF56526.1); contains InterPro domain Phosphatidylethanolamine-binding protein; (InterPro:IPR001858); contains InterPro domain PEBP; (InterPro:IPR008914) (MapMan)
Alternative Name: AT1G65480
GreenCard: GREENCARD (go to GreenCard)
TAIR: AT1G65480.1
TIGR: AT1G65480.1
MIPS: AT1G65480.1

Figure 3. Example of a keyword search using GreenCards. (A) The user had performed a search for the keywords 'FLOWERING LOCUS T', which retrieves links to the GreenCards of Genes (genome annotation projects), Clones (ESTs) and Plants (mutant plant lines). (B) Display of the Gene GreenCard, corresponding to the Arabidopsis annotated gene AT1G65480.1. Here the users find high confidence matching EST sequences displayed in the 'Related with' section. Sequence features and the sequences themselves are displayed as well. The selected gene has a matching EST (Clone: MPMGp2011E01215), this Clone GreenCard is shown in (C) and links back to the Gene GreenCard and to the original EST trace file, displayed by JTrev (D). A protein spot in rapeseed (*Brassica napus*) has been identified by 2-DE/MS as the protein encoded by the retrieved gene, and a link directs the user to the 2DE-gel with the identified protein spot highlighted with blue cross-hair (E). The spot identified links to a description of the protein (F) that provides links to the original Gene GreenCard.



Marker (RFLP): GBR1211

Mapping Info

map	species	chr.	mapped as	position	submitter
Consensus Stein2006	Barley	1	GBR1211	14.8 cm	Dr. Nils Stein
Steptoe x Morex	Barley	1	GBR1211	8 cm	Dr. Nils Stein

Related to Sequence, correspondig to BioObject: HY06E20

Sequence: HY06E20u

submitter of sequence: Dr. Wolfgang Michalek
length: 642
priming: 3-prime
EBI-Id: HVUL11728
GenBankAc: AL510886
IPK CR-EST: HY06E20u

Sequence features and Clone alignments (BLAST):

Gene structure	▬ UTR	▬ start codon	▬ stop codon	- - exon-exon border
Clone alignments	▬ forward	▬ reverse	▬ start	➡ stop
Protein domains	▬ domain			

Click on sequence features to obtain additional information.

```
HY06E20u
CCCGTTGNTA CGCCCTCCGA AGTACTCTCG AGTTTTTTTT TTTTTTTTTT TTTTTTTGCA    60
AACAACTATT TGTTTGTATT TATTCCCAAG TTTTATCTAG TATTGGATGG AATCGCACAG   120
ATGATGGTGC ATGACACTTT ATTTTTCATT GCTACATCGA TACTCCAATC GAGTAACAAT   180
GAAGGTCCAT CGATCTAGTA TTACAACTCT TGTCATTTCA TTGGCCACCA CTGCCGGTAG   240
CCATGCTACC AAATGGTGAG CAATAAGGTG GGACATGCAC ATTGCATAAC ATAGGAAGTG   300
TCTGAATCAC CAATGACCCA ACCACCTTGA ACTGGGCTTG TTGTTGAGGT AGGACAAAGG   360
CCTGGCCCGG CTGTTGTTGA AGCTGGATGG GCATTCCCTG GCCCAACTGT TGCAGTTGAG   420
ATTGGACAAA ACCATGACCC ACCTGTTGTT GTACTTGTTG TTGCATGACG ATNGCATGCA   480
CAATGGTTTG GATGGCTGGG CATCGGGATT GCTCATTGAT ATTTGCTAGT TGTTGACAAC   540
ATTGTTGTCG CTTCAATTGG CAACTGTTTT GTTGCGAAGA TATGTGGTNC GGAGGAACGA   600
GATCACTNGA CTGGAAGTAA TGGGTACCTT CTCCGGTCCA AG                      642
```

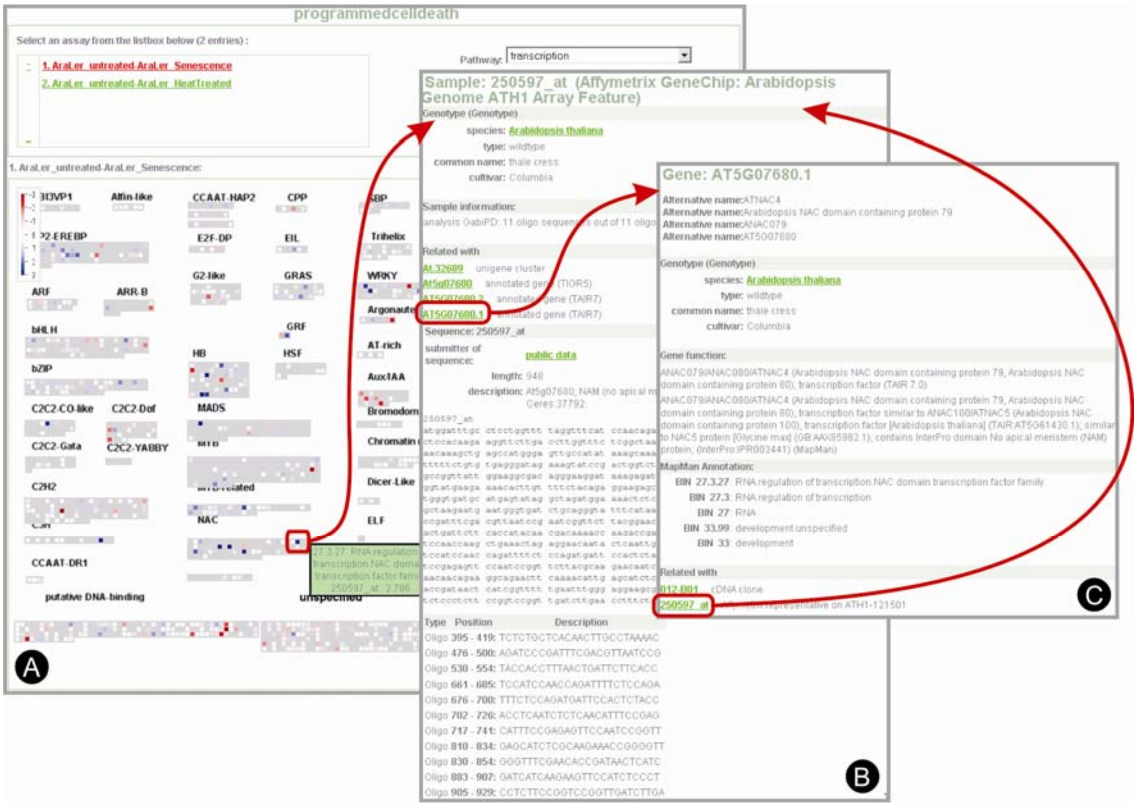
Type	Position	Orientation
ORF	1 - 195:	forward
ORF	2 - 97:	forward
ORF	3 - 131:	forward
ORF	131 - 256:	forward
ORF	282 - 142:	reverse
ORF	434 - 526:	forward
ORF	465 - 533:	forward
ORF	503 - 249:	reverse
ORF	516 - 310:	reverse
ORF	640 - 221:	reverse
ORF	642 - 556:	reverse

Cluster Info:

Contig30438.1 (18 members)

view Blast Searches

Figure 4. Visualisation of the genetic maps published by Stein et al. 2007 (18). (A) The region between 0cM and 25cM of barley chromosome I is shown; marker with single nucleotide polymorphisms (SNPs) are shown in light blue, marker with restriction fragment length polymorphisms (RFLPs) are shown in dark green. A selected marker is displayed in red, and links to the Marker GreenCard (B), which contains information on a related EST sequence therewith connecting genomic with transcriptomic information. With the EST description, cluster information is included (Contig30438.1) that links to the schematic representation of all ClusterContig members displayed onto the related consensus sequence (C). EST sequences that were selected from this ClusterContig as representatives for the new 27K barley unigene set are shown in turquoise.



programmedcelldeath

Select an assay from the listbox below (2 entries) :

1. AraLer_untreated-AraLer_Senescence
2. AraLer_untreated-AraLer_HeatTreated

Pathway: transcription

1. AraLer_untreated-AraLer_Senescence:

27.3.27: RNA.regulation transcription.NAC doma transcription factor fami 250597_at : 2.786

Sample: 250597_at (Affymetrix GeneChip: Arabidopsis Genome ATH1 Array Feature)

Genotype (Genotype)

species: Arabidopsis thaliana
type: wildtype
common name: thale cress
cultivar: Columbia

Sample information:
analysis GabiPD: 11 oligo sequences out of 11 oligo

Related with
At.32689 unigene cluster
At5g07680 annotated gene (TIGR5)
AT5G07680.2 annotated gene (TAIR7)
AT5G07680.1 annotated gene (TAIR7)

Sequence: 250597_at

submitter of sequence: public data
length: 948
description: At5g07680; NAM (no apical m Ceres:37792

```
250597_at
atggatttgc ctcctggttt taggtttcat ccaacaga
ctccacaaga aggttcttga ccttggtttc tcggctaa
aacaaagctg agccatggga gttgccatat aaagcaaa
tttttctgtg tgagggatag aaagtatccg actggtct
gccggttatt ggaaggcgac agggaaggat aaagagat
ggtatgaaga aaacacttgt tttctacaga ggaagagc
tgggtgatgc atgagtatag gctagatgga aaactctc
gctaagaatg aatgggtgat ctgcagggta tttcataa
ccgatttcga cgttaatccg aatcggttct tacggaac
actgattctt caccatacaa cgacaaaacc aagaccga
tccaaccaag ctgaaactag aggaacaata ctcaattg
tccatccaac cagattttct ccagatgatt ccactcta
tccgagagtt ccaatccggt tcttacgcaa gaacaatc
aacaacagaa ggcagaactt caaaacattg agcatctc
accgataact catcggtttt tgaatttggg aggaagcg
tctccctctt ccggtccggt tgatcttgaa cctttct
```

Type	Position	Description
Oligo	395 - 419:	TCTCTGCTCACAACTTGCCTAAAAC
Oligo	476 - 500:	AGATCCCGATTTCGACGTTAATCCG
Oligo	530 - 554:	TACCACCTTTAACTGATTCTTCACC
Oligo	661 - 685:	TCCATCCAACCAGATTTTCTCCAGA
Oligo	676 - 700:	TTTCTCCAGATGATTCCACTCTACC
Oligo	702 - 726:	ACCTCAATCTCTCAACATTTCCGAG
Oligo	717 - 741:	CATTTCCGAGAGTTCCAATCCGGTT
Oligo	810 - 834:	GAGCATCTCGCAAGAAACCGGGGTT
Oligo	830 - 854:	GGGTTTCGAACACCGATAACTCATC
Oligo	883 - 907:	GATCATCAAGAAGTTCCATCTCCCT
Oligo	905 - 929:	CCTCTTCCGGTCCGGTTGATCTTGA

Gene: AT5G07680.1

Alternative name:ATNAC4
Alternative name:Arabidopsis NAC domain containing protein 79
Alternative name:ANAC079
Alternative name:AT5G07680

Genotype (Genotype)

species: Arabidopsis thaliana
type: wildtype
common name: thale cress
cultivar: Columbia

Gene function:

ANAC079/ANAC080/ATNAC4 (Arabidopsis NAC domain containing protein 79, Arabidopsis NAC domain containing protein 80); transcription factor (TAIR 7.0)

ANAC079/ANAC080/ATNAC4 (Arabidopsis NAC domain containing protein 79, Arabidopsis NAC domain containing protein 80); transcription factor similar to ANAC100/ATNAC5 (Arabidopsis NAC domain containing protein 100), transcription factor [Arabidopsis thaliana] (TAIR:AT5G61430.1); similar to NAC5 protein [Glycine max] (GB:AAX85982.1); contains InterPro domain No apical meristem (NAM) protein; (InterPro:IPR003441) (MapMan)

MapMan Annotation:

BIN 27.3.27: RNA.regulation of transcription.NAC domain transcription factor family
BIN 27.3: RNA.regulation of transcription
BIN 27: RNA
BIN 33.99: development.unspecified
BIN 33: development

Related with
012-B01 cDNA clone
250597_at Affymetrix representative on ATH1-121501

Figure 5. Visualisation of expression profile data in MapManWeb. (A) The Affymetrix® NASC Array experiment on programmed cell death in Arabidopsis is displayed (NASCArrays reference number: 30). MapManWeb allows the visualisation of expressed genes in different biological processes; here only probesets (i.e., genes) involved in transcription regulation are shown. (B) Details for a strongly down-regulated probeset, with links to the related Gene GreenCard in *A. thaliana*. (C) The Gene GreenCard for the selected gene (ANAC79) links back to the probeset of the Affymetrix® ATH1 array.