



1st transPLANT user training workshop

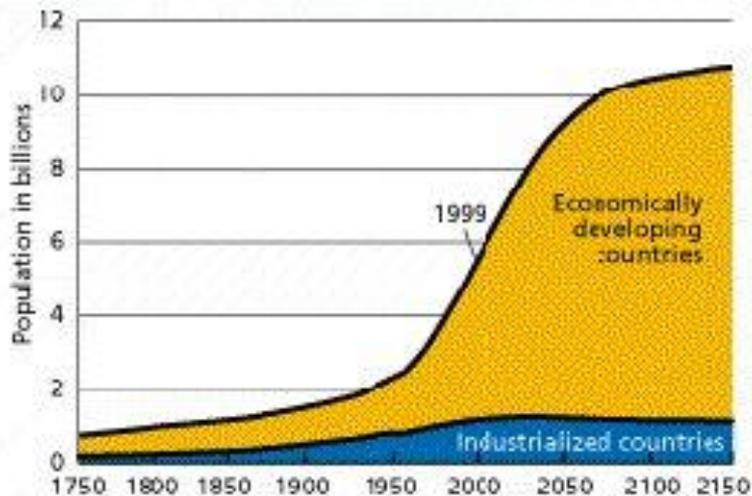
Versailles, 12th-13th November 2012



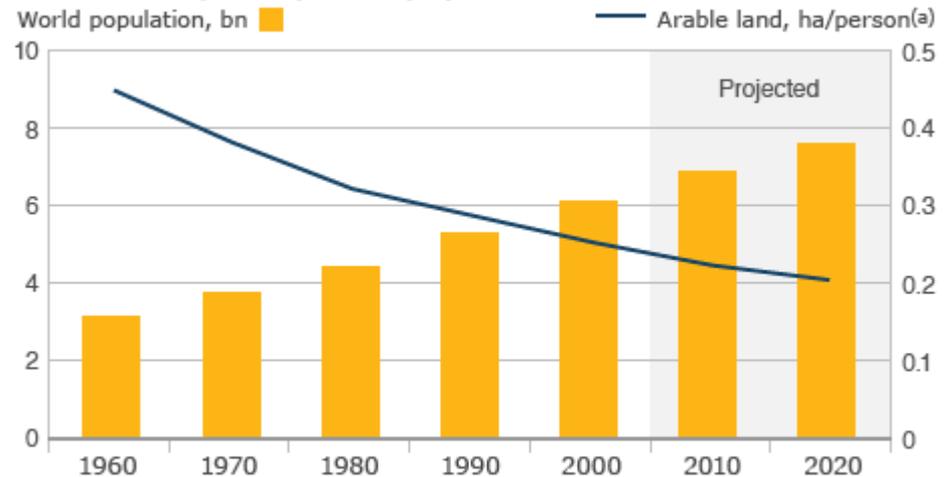
Paul Kersey, EMBL-EBI

More people, less land

World Population Growth, 1750-2150



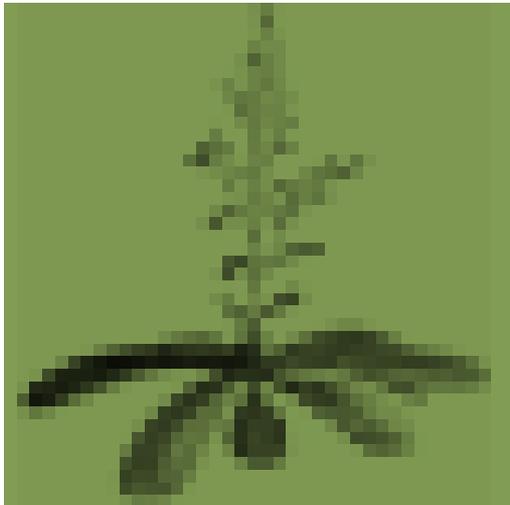
Arable land per capita vs population



(a) Hectares per person of arable land and land in permanent crops
Source: BHP Billiton investor presentation

Plant genome sequences, 2005

- *Arabidopsis thaliana* (2000)
- *Oryza sativa* (announced 2002, published 2005)



Why was plant genomics slow to take off?

- Smaller budget for plant research than for human health
 - Rising food and energy prices have since led to some re-adjustment of priorities
- Complex structure of genomes
 - Tetraploid: e.g. potato
 - Hexaploid: e.g. wheat
- Large size of genomes
 - Wheat genome (16 MB)
 - Loblolly pine genome (diploid) (40 MB)
 - Much of the size is accounted for by hard-to-assemble repeats

The Genomes are Coming!

- *Populus trichocarpa* (2006) 
- *Vitis vinifera* (2007) 
- *Zea mays* (2009) 
- *Brachypodium distachyon* (2010) 
- *Glycine max* (2010) 
- *Solanum tuberosum* L. (2011) 
- *Brassica rapa* (2012) 
- *Solanum lycopersicum* (2012) 
- Draft wheat, barley genome sequences in press
- 1001 (*Arabidopsis*) genome project (2009-2012)
- Massive sequencing and re-sequencing in crop plants to discover useful, breedable traits for new crop varieties

Automated phenotyping



www.transplantdb.eu

Barriers to unleashing the power of genomics

- Assembly and annotation
- Standards and databases
- Algorithms – smarter, more powerful (e.g. genome assembly complex trait analysis)
- Compute – capacity and interoperability
 - Within Europe and beyond
 - Towards “cloud” solutions

Some assemblies and their condition

- *Arabidopsis thaliana* (2000) 136 Mb, 7 chromosomes (140 contigs)
- *Oryza sativa* (2004) 487/488 Mb in 14 chromosomes (588 contigs)
- *Populus trichocarpa* (2006) 417 Mb in 2514 scaffolds (11048 contigs)
- *Vitis vinifera* (2007) 426/486 Mb in 19 scaffolds (14979 contigs)
- *Zea mays* (2009) 3.0/3.2 Gb in 12 chromosomes (19298 contigs)
- *Brachypodium distachyon* (2010) 2.71/2.72 Mb in 5 chromosomes (63872 contigs)

Some assemblies and their condition

- *Glycine max* (2010) 950/973 Mb in 11922 contigs 2.56..84 Gb in 10 chromosomes (15139 contigs)
- *Brassica rapa* (2012) 2.56/2.84 Gb in 10 chromosomes (50736 contigs)
- *Solanum lycopersicum* (2012) 781 Mb in 13 chromosomes (26917 contigs)

How well annotated are these genomes?

| Genome | Full length genes | Incomplete genes |
|--------------------------------|-------------------|------------------|
| <i>Arabidopsis thaliana</i> | 27399 | 17 |
| <i>Oryza sativa</i> | 57913 | 26 |
| <i>Populus trichocarpa</i> | 36483 | 4894 |
| <i>Vitis vinifera</i> | 29940 | 531 |
| <i>Zea mays</i> | 57449 | 5882 |
| <i>Brachypodium distachyon</i> | 26367 | 185 |
| <i>Glycine max</i> | 37653 | 8714 |
| <i>Brassica rapa</i> | 41018 | 0 |
| <i>Solanum lycopersicum</i> | 34363 | 312 |

Plant Ontology

Terms ▾

Jump To:

Details Visualization Notes (0) Term Mappings (11) Term Resources

- plant anatomical entity
 - plant anatomical space
 - plant structure
 - cardinal organ part
 - antheridium head
 - antiraphe
 - apical hook
 - archegonium head
 - archegonium neck
 - arilloid
 - auricle
 - awn
 - cambial zone
 - central zone of the leaf lamina
 - central zone of the petiole
 - connective
 - egg apparatus
 - elaiosome
 - fruit distal end

| | |
|-----------------|---|
| Preferred Name | seed coat epidermis |
| Synonyms | epidermis de la cubierta de la semilla (Spanish) 種子 種皮表皮 (Japanese) |
| Definitions | A portion of epidermis that is part of a seed coat. |
| ID | PO:0006048 |
| Full Id | http://purl.bioontology.org/ontology/PO/PO_0006048 |
| Exact Synonym | epidermis de la cubierta de la semilla (Spanish) 種子 種皮表皮 (Japanese) |
| Part Of | seed coat |
| Xref | OBO_SF_PO:3008016 |
| Xref Definition | POC:Ramona_Walls |

Schizosaccharomyces pombe

Location: I:534,120-537,869

Gene: pom1

Transcript: pom1

Gene-based displays

- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Ontology
 - Ontology graphs
 - phenotype (2)
 - cellular component (4)**
 - biological process (11)
 - molecular function (4)
 - protein modification (1)
 - Ontology table (22)
- Fungal Compara
 - Genomic alignments (image)
 - Genomic alignments (text)
 - Gene tree (image)
 - Gene tree (text)
 - Gene tree (alignment)
 - Orthologues (26)
 - Paralogues (20)
 - Protein families
- Pan-taxonomic Compara
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (63)
 - Paralogues (3)
 - List of species
 - Protein families (1)
- Phenotype
- Genetic Variation
 - Variation table
 - Structural variation
 - Variation image
- External data
 - Personal annotation
- ID History
 - Gene history

Gene: pom1 SPAC2F7.03c

Description

DYRK family protein kinase Pom1 [Source:PomBase Gene ID;Acc:SPAC2F7.03c]

Location

[Chromosome I: 534,120-537,869](#) reverse strand.

Transcripts

This gene has 1 transcript

| Name | Transcript ID | Length (bp) | Protein ID | Length (aa) | Biotype |
|------|-------------------------------|-------------|-----------------------------------|-------------|----------------|
| pom1 | SPAC2F7.03c.1 | 3750 | SPAC2F7.03c.1.pep | 1087 | Protein coding |

Ontology graph

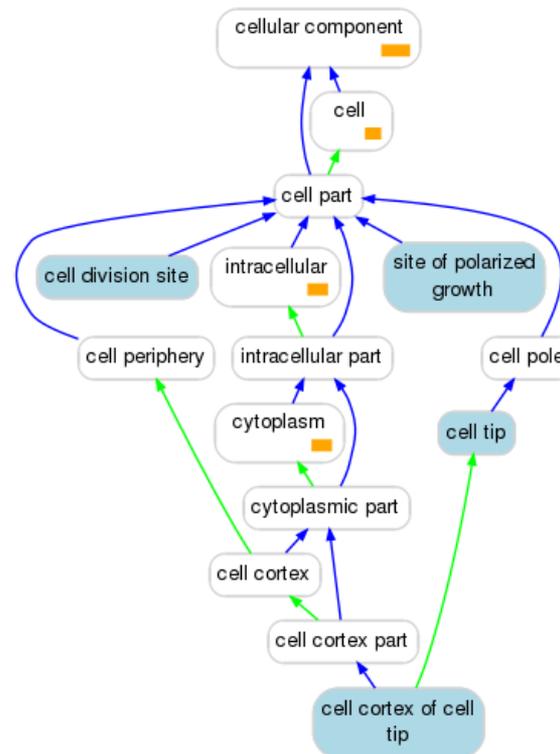
Terms:

Annotated terms

Generic GO slim terms

Relations:

- ↑ is_a
- ↑ part_of



- Configure this page
- Manage your data
- Export data
- Bookmark this page

How well annotated are these genomes?

| Genome | Number of genes | With biological process | With ≥ 1 process more specific than GO slim term (43 terms) |
|--------------------------------|-----------------|-------------------------|--|
| <i>Arabidopsis thaliana</i> | 27399 | 23490 | 12373 |
| <i>Oryza sativa</i> | 57939 | 22299 | 20467 |
| <i>Populus trichocarpa</i> | 41377 | 17494 | 15573 |
| <i>Vitis vinifera</i> | 30471 | 13937 | 12561 |
| <i>Zea mays</i> | 63331 | 19414 | 17073 |
| <i>Brachypodium distachyon</i> | 26552 | 12494 | 11094 |
| <i>Glycine max</i> | 46367 | 12073 | 11231 |
| <i>Brassica rapa</i> | 41018 | 23136 | 21187 |

How well annotated are these genomes?

| Genome | Number of genes | With at least 1 term annotated in all 3 GO ontologies (process, function, component) | With at least 1 sublim term annotated in all 3 ontologies |
|--------------------------------|-----------------|--|---|
| <i>Arabidopsis thaliana</i> | 27399 | 18350 | 4709 |
| <i>Oryza sativa</i> | 57939 | 10277 | 5923 |
| <i>Populus trichocarpa</i> | 41377 | 8952 | 4581 |
| <i>Vitis vinifera</i> | 30471 | 6442 | 3342 |
| <i>Zea mays</i> | 63331 | 9584 | 4967 |
| <i>Brachypodium distachyon</i> | 26552 | 6415 | 3441 |
| <i>Glycine max</i> | 46367 | 4653 | 2218 |
| <i>Brassica rapa</i> | 41018 | 17645 | 9619 |

www.transplantdb.eu



The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.



transPLANT

- A new 4 year EU FP7-funded project (DG INFOSOC) coordinated by EBI
- An I3 project: must have elements of coordination, service and RTD

What is transPLANT?

- We are proposing the establishment of new European consortium, transPLANT (**trans**-national infrastructure for **PLANT** genomic science), to design, implement, deploy and operate the software infrastructure critical to the future needs of plant scientists. transPLANT will exploit solutions already developed in adjacent domains (e.g. medical informatics), applying existing approaches to plants and extending them to deal with the specific challenges and opportunities unique to the domain of plant research; and thereby will maximise the potential of genomics to contribute to the second green revolution needed to keep the growing world population from hunger.

More specifically....

transPLANT will:

- Identify a common set of reference data to be shared between different researchers and service providers
- Construct of missing data archives
- Provide of tools to manipulate and mine plant genomic data
- Provide of an integrating point of interactive access to diverse data sets
- Provide of a compute environment for programatic access to plant genomic data
- Develop common standards for use within transPLANT and a wider community
- Train potential users
- Engage with other related communities to share experiences, tools and roadmaps

Coordination and Support Activities

- WP1 Management
- WP2 Interaction with relevant communities
- WP3 Standards development
- WP4 User training

Service Activities

- WP5 Services for computational access
- WP6 transPLANT portal: a single point of access to distributed data

RTD Activities

- WP7 A reference repository for genomes
- WP8 An architecture for plant genomic complexity
- WP9 An archive for plant variation data
- WP10 Linking genomes to phenotype
- WP11 Information retrieval systems
- WP12 Algorithm/tool evaluation, extension and development

From Genome to Variome

- A set of reference genomic data lies at the heart of the transPLANT (**WP7**)
- An archive for plant variation data (**WP9**) is the critical new component of the infrastructure
 - Archive will accession variant data, identify common variants between submissions, and project features forward between successive versions of an assembly
 - Based on submission of VCF files
 - **Looking for collaborators to help alpha-test the system**
- Representation (**WP8**), interpretation (**WP10**) and assembly (**WP12**) of variation data is a persistent theme throughout the RTD programme (and also a critical component of the standards work package, **WP3**)

A Virtual European Plant Database

- No single resource is adequately funded to run a “European Plant Database”
- Expertise in different crops, experimental approaches and analysis techniques are distributed throughout many countries
- Different interfaces serve different purposes

A Virtual European Plant Database

- This can be an advantage to users if:
 - Valuable data is persistently stored and remains accessible
 - Users can find the data they want
 - Users can combine data that resides in different places
 - Requires the use of common identifiers and descriptors for sequence, phenotype, cultivars, etc.
 - Users can compute against the data
 - Software and hardware components to this

Expertise of project partners

- The project partners bring together expertise in bioinformatic analysis and databases (**EBI, HGMU, TGAC**), agrobiolology (**INRA, IPG-PAS, DLO**), plant molecular biology (**GMI**), supercomputing (**BSC**), and also represent the commercial sector (**Biogemma, KeyGene**)

Triticeae in transPLANT

nature International weekly journal of science

Search Go [Advanced search](#)

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Research > Articles > Article

NATURE | ARTICLE OPEN

A physical, genetic and functional sequence assembly of the barley genome

The International Barley Genome Sequencing Consortium

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Nature (2012) | doi:10.1038/nature11543
Received 02 May 2012 | Accepted 30 August 2012 | Published online 17 October 2012

[PDF](#) [Citation](#) [Reprints](#) [Rights & permissions](#)

Abstract

[Abstract](#) • [Introduction](#) • [A sequence-enriched barley physical map](#) • [Repetitive nature of the barley genome](#) • [Transcribed portion of the barley genome](#) • [Regulation of gene expression](#) • [Natural diversity](#) • [Discussion](#) • [Methods](#) • [References](#) • [Acknowledgements](#) • [Author information](#) • [Supplementary information](#)

Barley (*Hordeum vulgare* L.) is among the world's earliest domesticated and most important crop plants. It is diploid with a large haploid genome of 5.1 gigabases (Gb). Here we present an integrated and ordered physical, genetic and functional sequence resource that describes the barley gene-space in a structured whole-genome context. We developed a physical map of 4.98 Gb, with more than 3.90 Gb anchored to a high-resolution genetic map. Projecting a deep whole-genome shotgun assembly, complementary DNA and deep RNA sequence data onto this framework supports 79,379 transcript clusters, including 26,159 'high-confidence' genes with homology support from other plant genomes. Abundant alternative splicing, premature termination codons and novel transcriptionally active regions suggest that post-transcriptional

[Journal home](#) | [Subscribe](#) | [Current issue](#) | [E-alert sign up](#) | [For authors](#) | [RSS feed](#)

98 days* **SCIENTIFIC REPORTS**

*For papers published in September 2012

Selected feature

The new map of science
In a special issue, *Nature* examines the changing global landscape of scientific research.

[See complete feature >](#)

Science jobs from naturejobs

Chief of the Division of Rheumatology at the David Geffen School of Medicine at UCLA
University of California, Los Angeles (UCLA)

Assistant Professor - Ecology and Evolution of Infectious Diseases
University of California, Berkeley

Senior Research Fellowship
Zoological Society of London

www.transplantub.eu

Triticeae in transPLANT

www.transplantdb.eu



The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.



Triticeae in transPLANT

- transPLANT partners working on important cereal crops including wheat, barely, rye, ryegrass (as well as the model system *Brachypodium*)
- The participating groups are mainly focused on bioinformatics, but are mostly embedded in institutions focused heavily on the genomics, biology and breeding of these species
- These institutions are at the heart of worldwide collaborative efforts to understand these species through genomics, and apply this understanding to the development of improved crops

Triticeae in transPLANT

International Wheat Genome Sequencing Consortium

News and Reports | Organization | Projects | Tools and Resources | General Documents

Bread wheat is grown on over 95% of the wheat growing area and its sequence holds the key to genetic improvements that will allow growers to meet the increasing demands for high quality food and feed produced in an environmentally sensitive, sustainable, and profitable manner. Further, because of its recent history, hexaploid wheat is a very good model to study **polyploidy**, a driving force for plant genome evolution.

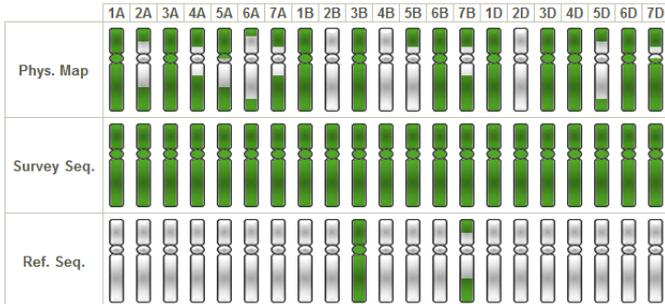
Hot topics

- BLAST now against the survey sequences of all individual chromosomes of bread wheat

The IWGSC was established by a group of plant scientists, breeders, and growers dedicated to sequencing the wheat genome to enhance our knowledge of the structure and function of the wheat genome. By gaining increased understanding of the biology of **agronomically important traits** and deploying state-of-the-art molecular tools, plant scientists and breeders will be able to accelerate wheat improvement to meet the challenges of the 21st century. The Consortium is committed to ensuring that the sequence of the wheat genome and the resulting DNA-based tools are available for all to use without restriction.

The IWGSC is governed by six co-chairs, a Coordinating Committee, and an executive director. The Coordinating Committee consists of representatives from laboratories involved in the development of resources and projects for sequencing and annotating the wheat genome, or from organizations providing direct funding for the consortium. **General membership** in the consortium is open to any individual, laboratory, or entity with an active interest in meeting the objectives of the IWGSC.

To achieve the vision of a sequenced wheat genome, the IWGSC establishes **strategic plans** with short- and mid-term goals, defines areas of **coordination**, facilitates and coordinates **research projects** and funding efforts at the national and international levels, develops and supports the design of research proposals, provides a framework for the establishment of common guidelines, protocols, and resources, and organizes scientific **meetings and workshops**.



News

- IWGSC Schedule scientific 2012 in Internati.
- IWGSC and a Bt Plant an Confere!
- G20 Agri the Inter Initiative (IRWI) T approve! Research Improve!



INTERNATIONAL
BARLEY
SEQUENCING
CONSORTIUM

[Home](#) | [Organization](#) | [Funded Projects](#) | [Sequence Resources](#) | [Barley Links](#) | [Documents](#) | [Contacts](#) | [News](#)

Mission Statement:

Barley (*Hordeum vulgare* L.) was one of the first domesticated cereal grains, originating in the Fertile Crescent over 10,000 years ago. Barley ranks fourth among the cereals in worldwide production and is widely cultivated in all temperate regions from the Arctic Circle to the tropics. In addition to its geographic adaptability, barley is particularly noted for its tolerance to cold, drought, alkali, and salinity.

The barley genome - with 5.3 billion letters of genetic code - is one of the largest in cereal crops and twice the size of the human genome. Barley is a true diploid, thus, it is a natural archetype for genetics and genomics for the Triticeae tribe, including polyploid wheat, and rye. Highly collaborative international efforts have produced a substantial body of genetic and genomic resources in the past several years.

The objective of the IBSC is to physically map and sequence the barley gene space, with the near-term need being the identification of the remainder of ~50,000 genes, including the 5' and 3' regulatory regions, and the longer-term goal an ordered physical map linked to the genetic map to accelerate crop improvement.

Latest News:

- October 17, 2012: Barley genome assembly published in Nature [\[http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11543.html\]](http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11543.html)

- IBSC News Release: [A new future for an old crop: Barley enters the genomics age](#)

IBSC Business Meeting at Plant Animal Genome XXI: Crescent



Triticeae in transPLANT

- Tremendous progress has been made, but assemblies are still fragmented
 - Disconnected contigs anchored on genetic and physical maps
- Gene models may be incomplete
- What can we do with this data, now?

[transPLANT user training workshop Monday, November 12, 2012 to Tuesday, November 13, 2012. INRA URGI campus, Versailles, France](#)

Resources

Find descriptions of the various plant-based databases, services, and software maintained by the transplant partners.



Many plant genomes are large and have complex evolutionary histories, making their analysis theoretically challenging and highly demanding of computational resources. Issues include genome size, polyploidy, and the quantity, diversity and dispersed nature of data in need of integration.

transPLANT is a consortium of 11 European partners gathered to address these challenges and to develop a trans-national infrastructure for plant genomic science. Bringing together groups with strengths in data analysis, plant science, computer science, and from the academic and commercial sectors, transPLANT will develop integrated standards and services and undertake new research and development needed to capitalise on the sequencing revolution, across the spectrum of agricultural and model plant species.

transPLANT is committed to establishing the broadest international collaborations for data and standards. Explore the project's aims in more detail on this website, or contact us on transplant_help@ebi.ac.uk.

Meetings and Events

[transPLANT user training workshop Monday, November 12, 2012 to Tuesday, November 13, 2012. INRA URGI campus, Versailles, France](#)

News

→ Article [Report on the Crop Plant Trait Ontology Workshop](#) added Wednesday, October 3, 2012 - 15:55

- Plant breeders, biologists, and bioinformaticians from ten countries, seven US states, and two plant agribusinesses gathered in Corvallis, Oregon, September 13th-15th, 2012 for a Crop Plant Trait Ontology Workshop.

→ Article [Ensembl Plants, Release 15](#) added Thursday, September 6, 2012 - 17:40

- We have added a set of homoeologous SNPs between wheat A, B and D genomes using wheat contigs aligned to *Brachypodium distachyon* as a reference framework, a structural variation dataset for *Sorghum bicolor* has been imported from dGVA, and a variety of small improvements to our assembly, annotation and variation datasets have been incorporated. See the [Ensembl Plants homepage](#) for details.

→ Publication [IDPredictor: predict database links in biomedical database](#) Mehlhorn H, added Thursday, July 5, 2012 - 15:14

Homepage

Static project summary, links to more info., dynamic listings: project related news / relevant upcoming meetings and workshops. RSS feeds for dynamic content (in and out).

Two workshops are planned for interacting with the target community. In these workshops, we will present our results, gather requirements, and explore avenues by which external stakeholders may take advantage of the transPLANT project. Round table discussions will be scheduled to promote exchange of perspectives and generate mutual feedback on the material presented.

Upcoming meetings

| Title | Description | Location | Start date | Type |
|-----------------------------------|--|--------------------------------------|---|----------|
| transPLANT user training workshop | <p>The workshop will focus on triticeae (wheat and barley mainly) data resources and tools, and include tutorials on data access and use from INRA, EBI-EMBL, IPK and MIPS.</p> <p>This workshop is targeted at (experimental) biologists, breeders, bio-informaticians and all other data users as no prior (informatics) knowledge or skills are required.</p> <p>For more information, a preliminary agenda and registration details please review the attached invitation.</p> | INRA URGI campus, Versailles, France | Monday, November 12, 2012 to Tuesday, November 13, 2012 | Workshop |

Previous meetings

| Title | Description | Location | Start date | Type |
|---|---|--|---|----------|
| Crop Plant Trait Ontology Workshop, Oregon State University | <p>The goals of the workshop are to develop ontology terms and definitions for the Plant Ontology (http://www.plantontology.org/) and the Reference Plant Trait Ontology (under development) and applications that utilize those resources.</p> <p>The ultimate goal is to establish a semantic framework for meaningful cross-species queries across gene expression and phenotype data sets from plant genomics and genetics experiments.</p> | Oregon State University | Thursday, September 13, 2012 to Saturday, September 15, 2012 | Workshop |
| Training workshop in plant pathogenic genomics | A hands-on training workshop in plant pathogenic genomics, focused on the new resource PhytoPath and including coverage of Ensembl Fungi, Ensembl Protists and PHI-base will be held at the EBI in September, aimed at PhD students and post-doctoral researchers who are working on fungal and oomycete-induced disease in plants. | EMBL-EBI, Hinxton, Cambridge, CB10 1SD, UK | Wednesday, September 19, 2012 to Thursday, September 20, 2012 | Workshop |
| | | | Thursday, | |

Meetings

Details of all meetings, past and future. Past events include associated resources, such as presentations and reports.

| | |
|---|--|
|  <p>The European Molecular Biology Laboratory Germany</p> <p>http://www.embl.org/</p> <p>An academic research institute located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK), part of the European Molecular</p> <p>Read more (and see members list)</p> |  <p>Helmholtz Zentrum Muenchen - Deutsches Forschungszentrum für Gesundheit und Umwelt Germany</p> <p>http://www.helmholtz-muenchen.de/</p> <p>A research group among the most reputed in plant genome oriented bioinformatics and has been highly influential in shaping the field by i</p> <p>Read more (and see members list)</p> |
|  <p>Gregor-Mendel-Institut für Molekulare Pflanzenbiologie Austria</p> <p>http://www.gmi.oeaw.ac.at/</p> <p>Established in 2000 by the Austrian Academy of Sciences to promote research excellence in the field of plant molecular biology, is locate</p> <p>Read more (and see members list)</p> |  <p>Leibniz Institut für Pflanzengenetik und Kulturpflanzenforschung Germany</p> <p>http://www.ipk-gatersleben.de</p> <p>A large and internationally renowned plant research institute in Gatersleben, Germany.</p> <p>Read more (and see members list)</p> |
|  <p>Institut National de la Recherche Agronomique France</p> <p>http://www.international.inra.fr/</p> <p>The first institute for agronomical research in Europe and the second</p> |  <p>Instytut Genetyki Roślin Polskiej Akademii Nauk Poland</p> <p>http://www.igr.poznan.pl/</p> <p>A centre of zoobiology and molecular genetics (100 scientific and technical</p> |

Partners

Complete table of project partners, including logos, short descriptions, and links to homepages. Partner pages include listings of project members and partner resources.

Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species.

Kersey PJ, Staines DM, Lawson D, Kulesha E, Derwent P, Humphrey JC, Hughes DS, Keenan S, Kerhornou A, Koscielny G, Langridge N, McDowall MD, Megy K, Maheswari U, Nuhn M, Paulini M, Pedro H, Toneva I, Wilson D, Yates A, Birney E

Nucleic Acids Res. 2012;40:D91-97

PMID:22067447

DOI:10.1093/nar/gkr895

An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M

Nature Genetics 2012;44:825-830

PMID:22067447

DOI:10.1038/ng.2314

IDPredictor: predict database links in biomedical database

Mehlhorn H, Lange M, Scholz U, Schreiber F

Journal of Integrative Bioinformatics 2012;9:190

PMID:22736059

DOI:10.2390/biecoll-jib-2012-190

Publications (and reports)

Listings of the outputs of the project, including publications and reports.



permalink

Brief description: BMRP (Bayesian Markov Random Fields for protein function prediction; Kourmpetis et al 2011, <http://www.ncbi.nlm.nih.gov/pubmed/21098674>) is a method for protein

ices, protein-contains predicted uncharacterized



The Crop EST Database (CR-EST) is a public available online resource providing access to sequence, classification, clustering, and annotation data of crop EST projects at the IPK.

Ensembl Plants
<http://plants.ensembl.org>



The Ensembl Genomes project produces genome browsers for important species from across the taxonomic range, using the Ensembl software system. Five sites are now available: Ensembl Bacteria, Ensembl Fungi, Ensembl Metazoa, Ensembl Plants, and Ensembl Protists.

Ensembl Plants includes reference genome assemblies and gene builds for 15 plant species, including functional, comparative, and variation data for important crop species.



rs for important Ensembl Fungi, s.

nd gene builds for variation data for



iptors list, originating from

almost 57 countries, are represented by advanced and primitive cultivars, breeding material, landraces, weedy and wild material.

Genebank Information and Order

European Barley Database (EBDB)
<http://pgrc-35.ipk-gatersleben.de/ebdb/>



The EBDB contains about 156,000 records, among them 1126 accessions of the International Barley Core Collection (BCC).

Genetic and Genomic Information System



<http://urgi.versailles.inra.fr/gnpis/>

Garlic ar <http://www.gatersleben>

General ir bulb struc passport : sequence

GnpIS is a powerful multispecies centralized database information system dedicated to plant, trees and their bioagressors (fungi). It is composed of a set of relational databases, each optimized according to one data domain and connected together to be search as a whole or by theme. Its originality is to bridge genetic and genomic data, allowing researchers and breeders to cross genetic information (i.e Genetic maps, QTL, markers, SNPs, Germplasms, Genotypes) with genomic data (i.e. physical maps, genome

Partner resources

Listing of all relevant resources maintained by the partners broken into three sections: Databases, Services and Software. Table provides descriptions and links.



Resource Instances

Species Type of data

Keywords

| Organization | Resource | Instance | URL | Source URL | Description | Species | Common name | Type of data | Tools | Keywords | Version | Release |
|---|----------------|----------------------|-----|------------|---------------------------------------|----------------------|------------------------|--------------|---------|----------------------------------|--|--|
| JGI | Phytozome | Apple | | | | Malus domestica | Apple | genomic | GBrowse | comparative genomics of plants | GDR prediction v1.0 on Malus x domestica assembly v1.0 | Phytozome version 8.0 - Jan 2012 |
| The European Molecular Biology Laboratory | Ensembl Plants | Arabidopsis lyrata | | source | Arabidopsis lyrata Ensembl Instance | Arabidopsis lyrata | Lyre-leaved rock cress | genomic | | Ensembl | Araly1 | Ensembl Plants release 13 - March 2012 |
| Gramene | Gramene | Arabidopsis lyrata | | source | Arabidopsis lyrata Ensembl Instance | Arabidopsis lyrata | Lyre-leaved rock cress | genomic | | Ensembl | Araly1 | Gramene release 34 - Jan 2012 |
| VIB Ghent | PLAZA | Arabidopsis lyrata | | source | | Arabidopsis lyrata | Lyre-leaved rock cress | genomic | | Green plant comparative genomics | JGI 1.0 | PLAZA 2.5 |
| The European Molecular Biology Laboratory | Ensembl Plants | Arabidopsis thaliana | | source | Arabidopsis thaliana Ensembl Instance | Arabidopsis thaliana | Thale Cress | genomic | | Ensembl | TAIR 10 | Ensembl Plants release 13 - March 2012 |
| Gramene | Gramene | Arabidopsis thaliana | | source | Arabidopsis thaliana Ensembl Instance | Arabidopsis thaliana | Thale Cress | genomic | | Ensembl | TAIR 10 | Gramene release 34 |

Resource registry

Categorised registry of community-wide resources of interest for plant bioinformatics. List can be browsed or filtered on specific facets of the data.



Integrated search

- Provide a single search for multiple resources
- Results from all data types are integrated into a simple table
- Links provided to view each result in the source database

transPLANT search

Enter terms

« first < previous ... 9 10 11 12 13 14 15 16 17 ... next > last »

| Provider | Type | ID | Species | Description |
|----------|-------------------------|------------------------------|-----------------|--|
| PlantsDB | transcript | Os04g0658300 | Oryza sativa | Similar to Ribulose biphosphate carboxyla (RA). - seq_id "AK067399";locus_ "0.999";coverage "0.999";Description "Simil chloroplast precursor (RuBisCO activase) (f "GO:0005524 : ATP binding";InterPro_id "IF INSD Accession #" |
| PlantsDB | transcript | Os09g0411650 | Oryza sativa | putative protein - seq_id "J065163B06";locu "NONE";coverage "NONE";Description "Sin subunit N- methyltransferase, chloroplast p lysine N-methyltransferase) (RuBisCO metl "Q43088";category "II";GO_id "No significar "J065163B06, 5/3'-EST" |
| CR-EST | expressed sequence tags | H105M01r | Hordeum vulgare | gi 18700270 gb AAL77745.1 At1g73110/F3 [Arabidopsis; gi 19387266 gb AAL87177.1 gi 50929327 ref XP_474191.1 OSJNBa0011 gi 5903101 gb AAD55659.1 Unknown prote activase 2 [Mesembryanthemum crystallinu |



- A modular suite of software for genome analysis and visualisation developed jointly by the Wellcome Trust Sanger Institute and the European Bioinformatics Institute
- Now used for genomes from across the taxonomic space (including plant hosts)
- Offers a standard set of interfaces to a wide range of genome-scale data, including:
 - Web-based GUI
 - Public MySQL server
 - Perl API

vertebrates

bacteria

fungi

protists

plants

metazoa

www.transplantdb.eu



The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.



Ensembl and agriculture

- Scope of Ensembl Plants includes important crop and model species
- Ensembl Metazoa, Protists, Fungi and Bacteria potentially include important agricultural pathogens, pests, pollinators, symbiotes
- Important farm animals (cow, sheep, pig, chicken etc.) among the key species in vertebrate Ensembl

- Location-based displays**
- Whole genome
 - Chromosome summary
 - Region overview
 - Region in detail**
 - Comparative Genomics
 - Genomic alignments (7)
 - Synteny (0)
 - Genetic Variation
 - Resequencing (0)
 - Linkage Data
 - Markers
 - Other genome browsers
- Configure this page
 Manage your data
 Export data
 Bookmark this page
- Ensembl Plants is produced in collaboration with Gramene
- DB built by NASC

Chromosome 3: 19,431,371-19,434,403

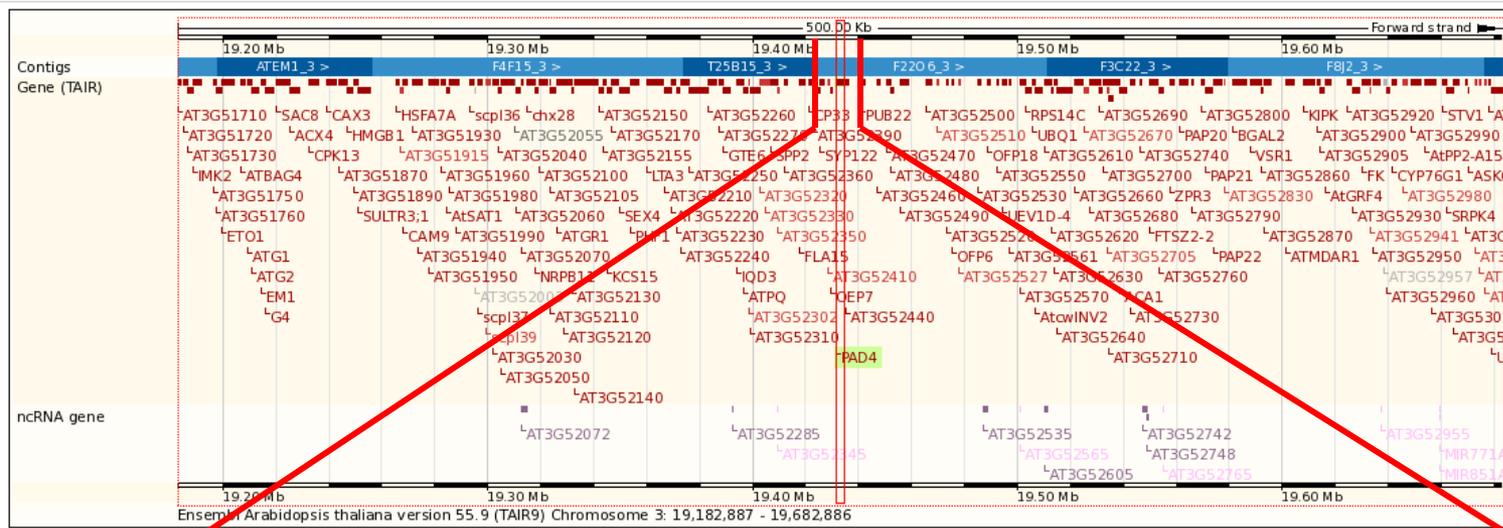
chromosome 3

[Export image](#)

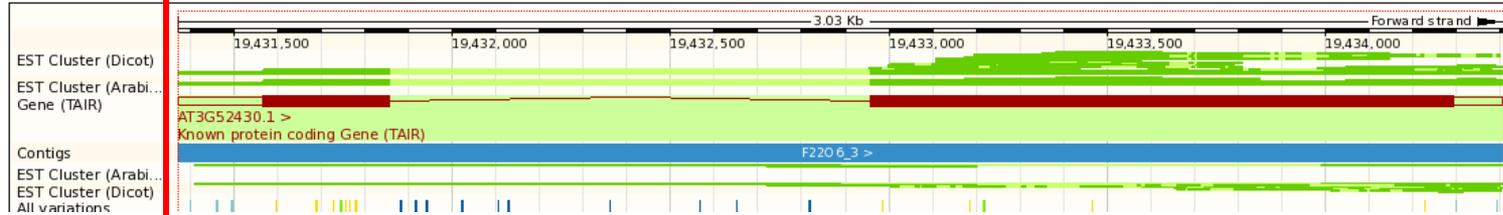
« Region overview

Region in detail [help](#)

[Genomic alignments »](#)



Location: 3 19431371 - 19434403 [Go>](#)



Gene-based displays

- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Plants Compara
 - Genomic alignments (7)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (12)
 - Paralogues (5)
- Pan-taxonomic Compara
 - Gene Tree (image)**
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (6)
 - Paralogues (5)
 - Protein families (0)
- Genetic Variation
 - Variation Table
 - Variation Image
- External Data
 - Personal annotation
- ID History
 - Gene history

- Configure this page
- Manage your data
- Export data
- Bookmark this page

Ensembl Plants is produced in collaboration with Gramene

DB built by NASC

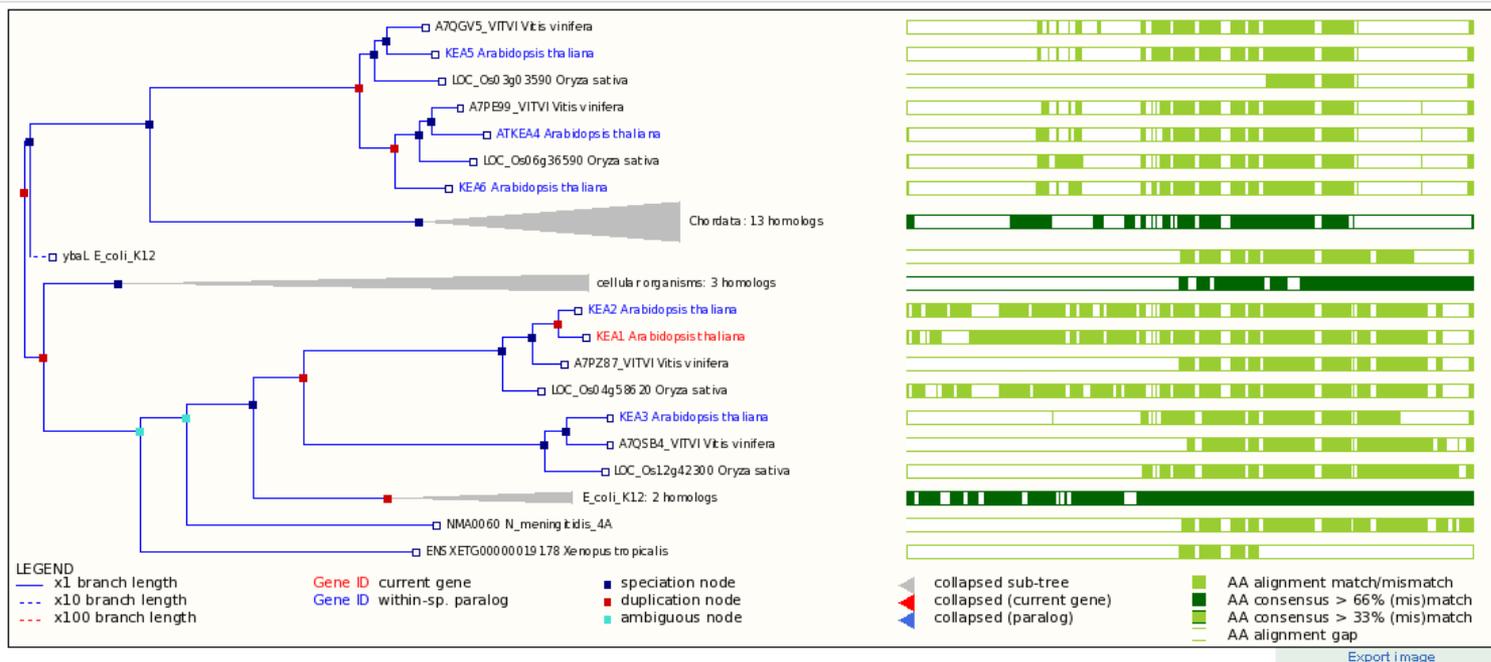
Gene: KEA1 (AT1G01790-TAIR-G)

KEA1 (K EFFLUX ANTIPORTER 1); potassium ion transmembrane transporter/ potassium:hydrogen antiporter; K efflux antiporter KEA1 source: TAIR KEA1

Location [Chromosome 1: 284,781-291,094](#) forward strand.

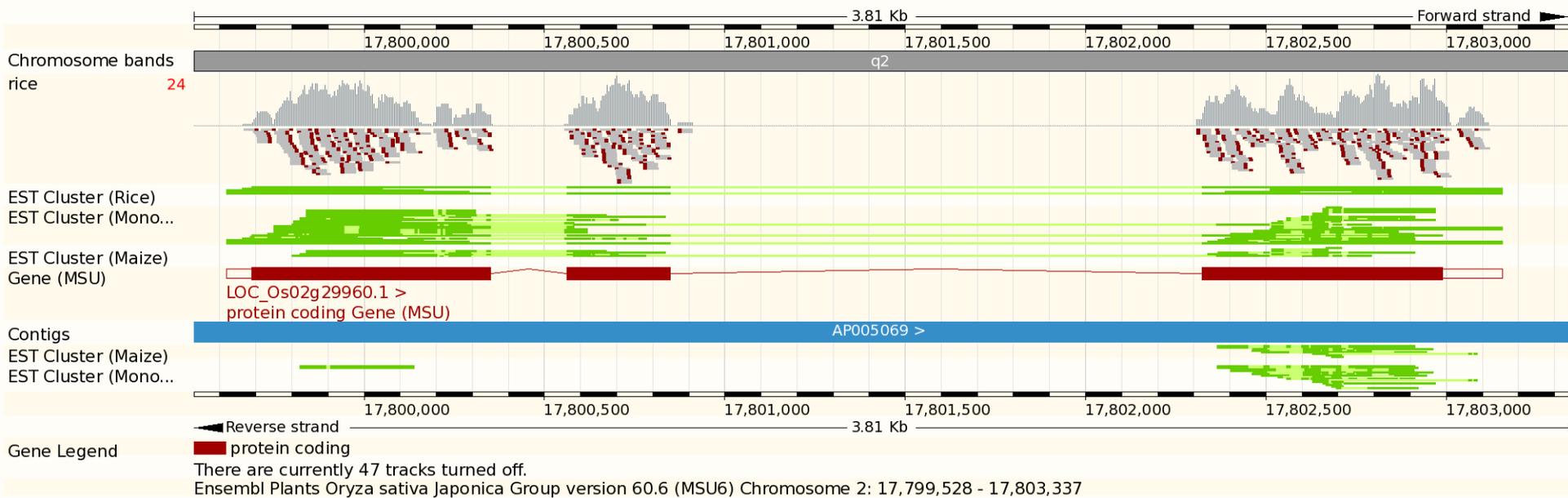
Transcripts There is 1 transcript in this gene: [show transcripts](#)

« Paralogues Gene Tree (image) Orthologues »



View options:

- [View current gene only](#)
- [View paralogs of current gene](#)
- [View all duplication nodes](#)



Gene-based displays

- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references (4)
- Regulation
- Plants Compara
 - Genomic alignments (7)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (11)
 - Paralogues (3)
- Pan-taxonomic Compara
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (2)
 - Paralogues (3)
 - Protein families (0)
- Genetic Variation
 - Variation Table
 - Variation Image**
 - External Data
 - Personal annotation
 - ID History
 - Gene history
- Configure this page
- Manage your data
- Export data
- Bookmark this page

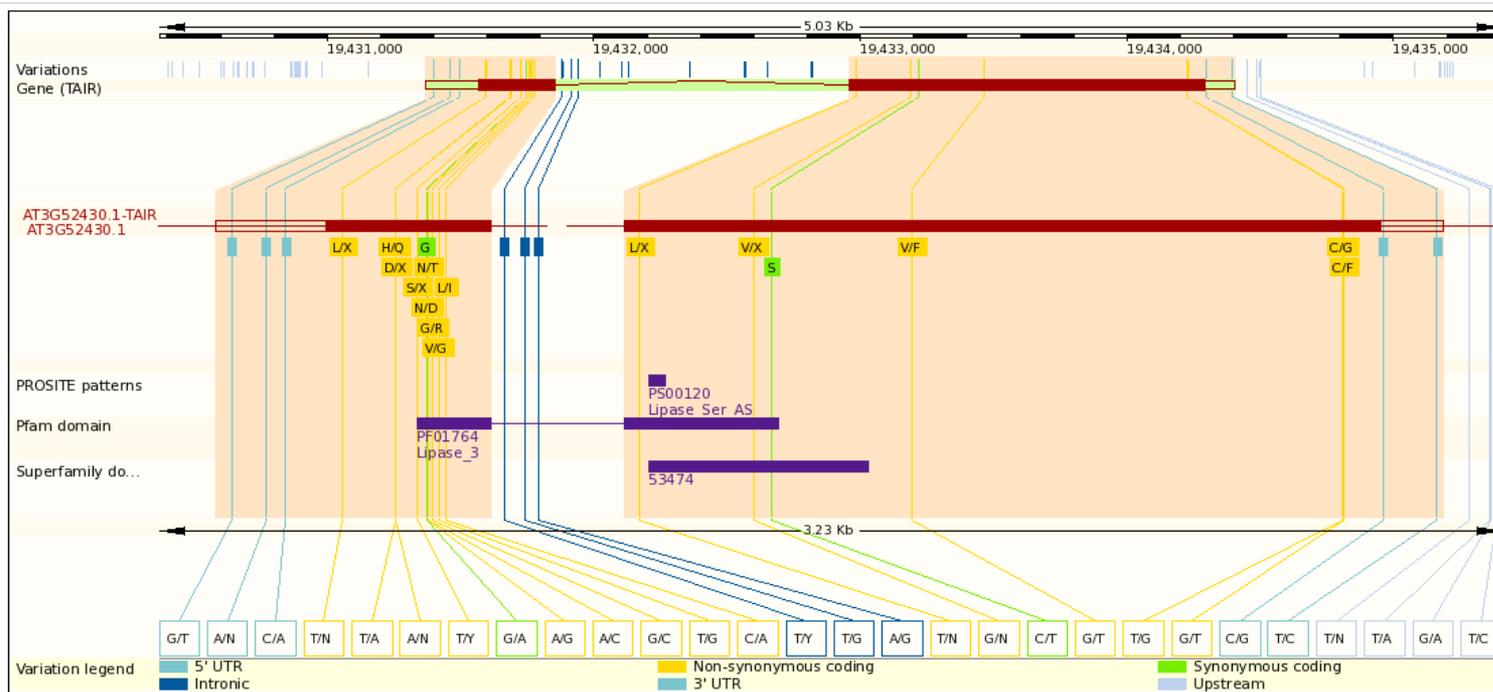
Gene: PAD4 (AT3G52430-TAIR-G)

PAD4 (PHYTOALEXIN DEFICIENT 4); lipase/ protein binding / triacylglycerol lipase; Encodes a lipase-like gene that is important for salicylic acid signaling and function in resistance (R) gene-mediated and basal plant disease resistance. PAD4 can interact directly with EDS1, another disease resistance signaling protein. Expressed at elevated level in response to green peach aphid (GPA) feeding, and modulates the GPA feeding-induced leaf senescence through a mechanism that doesn't require camalexin synthesis and salicylic acid (SA) signaling. source: TAIR PAD4

Location Chromosome 3: 19,431,371-19,434,403 forward strand.

Transcripts There is 1 transcript in this gene: [show transcripts](#)

« Variation Table Variation Image help External Data »



Acknowledgements

- Dan Bolser, Arnaud Kerhornou, Uma Maheswari, ChuangKee Ong, Dan Staines
 - Jay Humphrey, Eugene Kulesha, Nick Langridge
- transPLANT is funded by the European Commission within its 7th framework programme under the thematic area “infrastructures” Contract Number 283496
- United Kingdom Biosciences and Biotechnology Research Council funding for work on wheat, barley and UK-US collaboration

