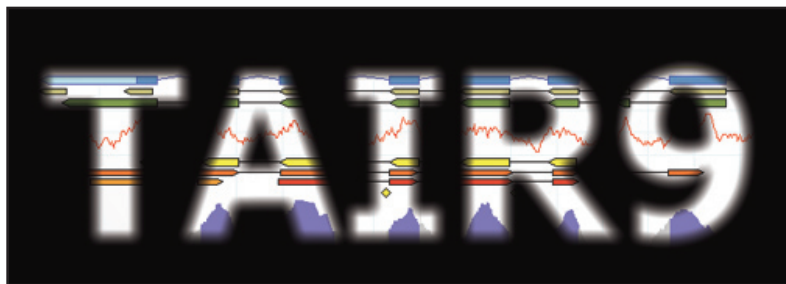


TAIR Newsletter Summer 2009

TAIR9 genome release

TAIR9 Statistics

The Arabidopsis Information Resource (TAIR) is pleased to announce the release of the latest version of the Arabidopsis genome annotation (TAIR9) at TAIR and NCBI on June 19, 2009. The latest release builds upon the gene structures of the previous TAIR8 release using new cDNA and EST data released in the intervening period as well as manual updates informed by cross species alignments, peptides and community input regarding missing and incorrectly annotated genes.



The TAIR9 release contains 27,379 protein coding genes, 4827 pseudogenes or transposable element genes and 1312 ncRNAs (33,518 genes in all, 39,640 gene models). A total of

282 new loci and 739 new gene models were added. Fourteen percent (4626) of Arabidopsis genes now have annotated splice variants. Updates were made to 1254 gene models of which 774 had CDS updates; a total of 1144 exons were modified and 1056 new exons incorporated. There were 13 gene splits and 46 gene merges.

Genome Sequence Update

In agreement with our reference genome policy (http://www.arabidopsis.org/doc/portals/genAnnotation/gene_structural_annotation/ref_genome_sequence/11413) corrections to the reference assembly were only made if supported by at least two independently derived sequence libraries from the Columbia ecotype. The following updates were made to the chromosome sequences for this release:

227 single nucleotide substitutions were made to the assembly sequence based on re-sequencing data provided by Richard Clark (Ossowski et al. 2008) and Joe Ecker.

341 indels were made to the assembly sequence based on re-sequencing data provided by Richard Clark and EST and cDNA sequences deposited in Genbank that supported the insertion/deletion.

We previously identified 14 regions in TAIR8 as either vector, *E.coli* or rice contamination, and substituted the existing sequence with the equivalent number of IUPAC ambiguity code 'N'. In TAIR9, we standardized (via deletion) these sequences to a set size of 100bp.

All five nuclear chromosomes were updated for TAIR9. Details of the golden path length of each chromosome can be found at http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/agicomplete.jsp Further details of these assembly changes can be found at ftp://ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/

Assembly updates and gap information can also be viewed in TAIR's GBrowse (see Assembly tracks section).

Gene confidence ranking

In order to give Arabidopsis researchers a better idea about the amount of evidence that supports their gene(s) of interest, we developed a ranking systems that attributes confidence scores to all exons and gene models in TAIR9 based on different types of experimental and computational evidence. The new confidence ranking will for example allow researchers to identify a set of gold standard confirmed structures or identify sets of exons where both donor and acceptor splice sites are supported. More details of the confidence ranks can be found at ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/Gene_Confidence.pdf

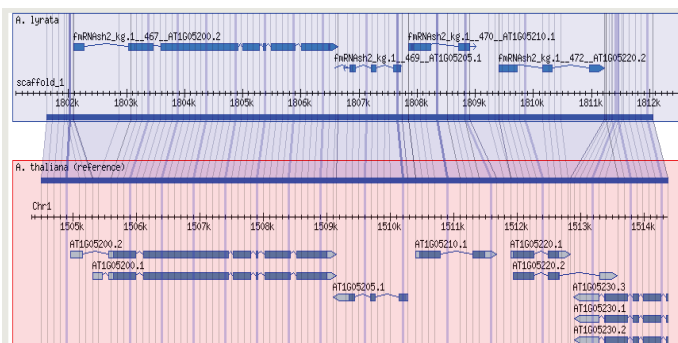
AraCyc 5.0 and Plantcyc 2.0 released

In March 2009, a new version of AraCyc was released, containing 20 new pathways curated from the literature, and 37 new pathways from computational prediction. In addition, two new compound sets were added from experimental results obtained through the Arabidopsis NSF2010 Metabolomics Project and the KNAp-SaCK Species-Metabolite Relationship Database. For more details on the AraCyc updates please visit <http://www.arabidopsis.org/biocyc/releasenotes.jsp>.

A new version of PlantCyc has also been released: Plantcyc 2.0 includes 83 new pathways. It now has orthologs from Arabidopsis, Medicago, tomato and rice. Please visit www.plantcyc.org to learn more about this database and to use it to investigate metabolism in Arabidopsis and many other plants.

Upcoming new viewers in TAIR

A **synteny viewer** (gbrowse_syn, a GMOD project) will be available at TAIR in July 2009. This tool will allow the user to compare syntenic regions between *A. thaliana* and *A. lyrata*. Additional plant genomes will be added in the future. We are also currently working on implementing **NBrowse**, a **protein interaction viewer** designed by Gunsalus et al, at TAIR. This viewer will display protein-protein interactions curated by TAIR, IntAct and BioGrid. Users can search for interactions found with their protein of interest, filter by type (experimental, predicted, interolog) or experiment (Y2H, Co-Immunoprecipitation,...) and can overlap their own interaction set with those displayed in NBrowse.



Meet TAIR curators at several plant biology conferences this summer

TAIR and PMN curators are taking to the road this summer to attend several meetings: We will attend the 20th International Conference on Arabidopsis Research (ICAR) in Scotland (June 30-July 4), the International Plant Biology Meeting in Hawaii (July 18-22), the 9th International Plant Molecular Biology Congress in St Louis (October 25-29), and the Plant Satellite Symposium of the SDB annual meeting in San Francisco (July 23). We will host a curation booth at ICAR and we encourage you to visit our booth and talk to the TAIR curators about any ideas, suggestions and questions you might have regarding TAIR. We will also present talks and workshops at all of the 4 above mentioned conferences. Details on the time and location of TAIR presentations at these conferences will be announced on the TAIR homepage prior to and during the time of each conference.

