

transPLANT milestone report

MS14 (work package 5): DAS servers provided for sequence and annotation for 20 reference genomes

The Distributed Annotation System (DAS; <http://www.biodas.org>) is a simple protocol for the publishing of reference entities and (positional or non-positional) annotation on those entities through the provision of a DAS server. The existence of DAS servers allows the development of DAS clients, integrative interfaces that selectively and dynamically integrate data from different sources. transPLANT partner EMBL-EBI has provided DAS through the Ensembl Plants interface: in the course of the third year of transPLANT, servers were provided for 14 additional genomes: *Amborella trichopoda* (a tropical shrub), *Brassica oleracea* (whose cultivars include cabbage, Brussels sprouts, broccoli, kohlrabi, kale and cauliflower), *Leersia perrieri* (a wild grass), *Ostreococcus lucimarinus* (a unicellular picoplankton), seven new rice genomes (*Oryza barthii*, *Oryza glumaepatula*, *Oryza meridionalis*, *Oryza nivara*, *Oryza punctata*, and *Oryza rufipogon*), *Prunus persica* (peach), *Theobroma cacao* (the cultivated chocolate tree), and *Triticum aestivum* (the bread wheat genome). These new servers take the total number of species for which DAS servers are available through Ensembl Plants to 39, of which 30 have been made available through transPLANT funding. We publish both the genomic sequence (allowing for its use as a reference by other annotation servers), and the annotation (genes, transcripts, translations) available within Ensembl Plants (as data on that reference) as DAS servers. Servers are also maintained for older versions of the genome, so that users can continue to visualise older annotation. Publishing annotation via DAS is based upon a common system for identification of reference sequence versions; allows for data sharing among consortium members and other sites and for visualisation in most commonly used genome browsers (for example, the Ensembl Genome Browser and Gbrowse are both DAS clients).

The availability of these DAS sources has been published at <http://plants.ensembl.org/das/sources> (see figure, below).

We will continue to publish additional DAS servers throughout the course of the project, while also developing new methods for users to integrate their data sets (which are increasingly based on new sequencing technologies, and consequently larger than traditional feature annotation) dynamically within Ensembl Plants and the other plant resources.

Figure 1 Screenshot of DAS sources currently available in Ensembl Plants

 DAS sources available Show downloads									
#	URI	Title	Species	Assembly	Maintainer	Taxon ID	Test range	Capabilities	Description
1	ENSEMBL_PLANTS_1_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.reference	<i>Aegilops tauschii</i>	GCA_000347335.1	helpdesk@ensemblgenomes.org	37682		das1:entry_points das1:sequence das1:features	Aegilops_tauschii Reference server based on GCA_000347335.1 assembly. Contains 429892 top level entries.
2	ENSEMBL_PLANTS_3_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.transcript	<i>Aegilops tauschii</i>	GCA_000347335.1	helpdesk@ensemblgenomes.org	37682		das1:stylesheets das1:features	Annotation source for Aegilops_tauschii transcript
3	ENSEMBL_PLANTS_4_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.translation	<i>Aegilops tauschii</i>	GCA_000347335.1	helpdesk@ensemblgenomes.org	37682		das1:stylesheets das1:features	Annotation source for Aegilops_tauschii translation
4	ENSEMBL_PLANTS_1_GCA_000471905.1	Amborella_trichopoda.GCA_000471905.1.reference	<i>Amborella trichopoda</i>	GCA_000471905.1	helpdesk@ensemblgenomes.org	13333		das1:entry_points das1:sequence das1:features	Amborella_trichopoda Reference server based on GCA_000471905.1 assembly. Contains 5745 top level entries.
5	ENSEMBL_PLANTS_3_GCA_000471905.1	Amborella_trichopoda.GCA_000471905.1.transcript	<i>Amborella trichopoda</i>	GCA_000471905.1	helpdesk@ensemblgenomes.org	13333		das1:stylesheets das1:features	Annotation source for Amborella_trichopoda transcript
6	ENSEMBL_PLANTS_4_GCA_000471905.1	Amborella_trichopoda.GCA_000471905.1.translation	<i>Amborella trichopoda</i>	GCA_000471905.1	helpdesk@ensemblgenomes.org	13333		das1:stylesheets das1:features	Annotation source for Amborella_trichopoda translation
7	ENSEMBL_PLANTS_1_v.1.0	Arabidopsis_lyrata.v.1.0.reference	<i>Arabidopsis lyrata</i>	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:entry_points das1:sequence das1:features	Arabidopsis_lyrata Reference server based on v.1.0 assembly. Contains 695 top level entries.
8	ENSEMBL_PLANTS_3_v.1.0	Arabidopsis_lyrata.v.1.0.transcript	<i>Arabidopsis lyrata</i>	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheets das1:features	Annotation source for Arabidopsis_lyrata transcript
9	ENSEMBL_PLANTS_4_v.1.0	Arabidopsis_lyrata.v.1.0.translation	<i>Arabidopsis lyrata</i>	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheets das1:features	Annotation source for Arabidopsis_lyrata translation
10	ENSEMBL_PLANTS_7_v.1.0	Arabidopsis_lyrata.v.1.0.prediction_transcript	<i>Arabidopsis lyrata</i>	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheets das1:features	Annotation source for Arabidopsis_lyrata prediction_transcript
11	ENSEMBL_PLANTS_1_TAIR10	Arabidopsis_thaliana.TAIR10.reference	<i>Arabidopsis thaliana</i>	TAIR10	helpdesk@ensemblgenomes.org	3702		das1:entry_points das1:sequence das1:features	Arabidopsis_thaliana Reference server based on TAIR10 assembly. Contains 7 top level entries.
12	ENSEMBL_PLANTS_3_TAIR10	Arabidopsis_thaliana.TAIR10.transcript	<i>Arabidopsis thaliana</i>	TAIR10	helpdesk@ensemblgenomes.org	3702		das1:stylesheets das1:features	Annotation source for Arabidopsis_thaliana