transPLANT milestone report

MS13 (work package 5): DAS servers provided for sequence and annotation for 15 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 provides DAS through the Ensembl Plants interface: in the course of the first year of transPLANT, servers were provided for six additional genomes: the bread wheat D-genome progenitor *Aegilops tauschii,* barley (*Hordeum vulgare*), banana (*Musa acuminata*), barrel clover (*Medicago truncatula*), potato (*Solanum tuberosum*) and the bread wheat A-genome precursor *Triticum urartu*. These new servers take the total number of species for which DAS servers are available through Ensembl Plants to 25, of which 16 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).

Figure 1 Screenshot of DAS sources available in Ensembl Plants

Ensembl

DAS sources available

#	URI	Title	Species	Assembly	Maintainer	Taxon ID	Test range	Capabilities	Description
1	ENSEMBL PLANTS_1_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.reference	*Aegilops tauschii*	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 429891 top level entries.
2	ENSEMBL PLANTS_3_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.transcript	*Aegilops tauschii*	GCA_000347335.1	helpdesk@ensemblgenomes.org	37682		das1:stylesheet das1:features	Annotation source for Aegilops_tauschii transcript
3	ENSEMBL PLANTS_4_GCA_000347335.1	Aegilops_tauschii.GCA_000347335.1.translation	*Aegilops tauschii*	GCA_000347335.1	helpdesk@ensemblgenomes.org	37682		das1:stylesheet das1:features	Annotation source for Aegilops_tauschii translation
4	ENSEMBL PLANTS_1_v.1.0	Arabidopsis_lyrata.v.1.0.reference	*Arabidopsis lyrata*	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.
5	ENSEMBL PLANTS_3_v.1.0	Arabidopsis_lyrata.v.1.0.transcript	*Arabidopsis lyrata*	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheet das1:features	Annotation source for Arabidopsis_lyrata transcript
6	ENSEMBL PLANTS_4_v.1.0	Arabidopsis_lyrata.v.1.0.translation	*Arabidopsis lyrata*	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheet das1:features	Annotation source for Arabidopsis_lyrata translation
7	ENSEMBL PLANTS_7_v.1.0	Arabidopsis_lyrata.v.1.0.prediction_transcript	*Arabidopsis lyrata*	v.1.0	helpdesk@ensemblgenomes.org	81972		das1:stylesheet das1:features	Annotation source for Arabidopsis_lyrata prediction_transcript
8	ENSEMBL PLANTS_1_TAIR10	Arabidopsis_thaliana.TAIR10.reference	*Arabidopsis thaliana*	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.
9	ENSEMBL PLANTS_3_TAIR10	Arabidopsis_thaliana.TAIR10.transcript	*Arabidopsis thaliana*	TAIR10	helpdesk@ensemblgenomes.org	3702		das1:stylesheet das1:features	Annotation source for Arabidopsis_thaliana transcript
10	ENSEMBL PLANTS_4_TAIR10	Arabidopsis_thaliana.TAIR10.translation	*Arabidopsis thaliana*	TAIR10	helpdesk@ensemblgenomes.org	3702		das1:stylesheet das1:features	Annotation source for Arabidopsis_thaliana translation
11	ENSEMBL PLANTS_7_TAIR10	Arabidopsis_thaliana.TAIR10.prediction_transcript	*Arabidopsis thaliana*	TAIR10	helpdesk@ensemblgenomes.org	3702		das1:stylesheet das1:features	Annotation source for Arabidopsis_thaliana prediction_transcript
12	ENSEMBL PLANTS_1_v1.0	Brachypodium_distachyon.v1.0.reference	*Brachypodium distachyon*	v1.0	helpdesk@ensemblgenomes.org	15368		das1:entry_points das1:sequence das1:features	Brachypodium_distachyon Reference server based on v1.0 assembly. Contains 83 top level entries.
13	ENSEMBL PLANTS_3_v1.0	Brachypodium_distachyon.v1.0.transcript	*Brachypodium distachyon*	v1.0	helpdesk@ensemblgenomes.org	15368		das1:stylesheet das1:features	Annotation source for Brachypodium_distachyon transcript
14	ENSEMBL PLANTS_4_v1.0	Brachypodium_distachyon.v1.0.translation	*Brachypodium*	v1.0	helpdesk@ensemblgenomes.org	15368		das1:stylesheet	Annotation source for

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.