



- A platform for genome annotation, analysis and browsing, developed jointly by the EBI and Wellcome Trust Sanger Institute
- Modules for handling variation data, comparative genomics, gene prediction etc.
- Multiple points of access to data: web-based application, Perl and REST-ful APIs, public mySQL databases, BioMart data mining tool, DAS client and server, FTP
- Upload your own data and compare to reference annotation
- Originally developed for vertebrate genomes, subsequently extended to non-vertebrate species

|--|

	Data
vertebrates (ensembl.org)	Genomic sequence Gene / transcript / protein models External references
bacteria.ensembl.org fungi.ensembl.org protists.ensembl.org	 Mapped cDNAs, proteins, microarray probes, BAC clones, repeats, markers etc. etc. Variation data: sequence variants structural variants Comparative data: gene trees, orthologues, paralogues pairwise whole genomic alignments, syntenic regions
CL 01.13 MGBC - Standards and Protocols PAG XXII EMBL-EBI	ЕМВІ-ЕВІ 🧼

Wheat data in Ensembl Plants (release 21)

- Version1.0 of the chromosome survey sequence assembly generated by the International Wheat Genome Sequencing Consortium.
- Version 2.0 of the IWGSC gene models called on the survey sequence
 - Version 2.1 due in next release (March 2014)
- Whole genome alignments against Brachypodium and rice
- Alignments to wheat Unigenes and RNA-seq data

19.02.2013 pkersey@ebi.ac.uk Plant and Animal Genomes 2014

EMBL-EBI

Chromosome survey sequence: plans for the next release (release 22, due March 2014)

- Whole genome alignments
 - Against barley, Triticum uratu and Aegilops taucshcii
 - Wheat AvB, AvD, BvD
- Alignment to wheat full length ESTs
- Additional RNA-seq alignments

Genome assemblies from Brenchley et al.

- The wheat genome assemblies generated by Brenchley et al. (PMID: 23192148) have also been aligned to the survey sequence, Brachypodium and barley
 - Homeolegous variants inferred between the 3 wheat genomes (from this work) are also displayed in the context of the gene models from Brachypodium and barley

Diploid progenitor genomes

- Aegilops tauschii and Triticum uratu are also included in Ensembl Plants
- These genomes have been aligned to rice, and to relevant RNA-seq reads.
- These genomes and other more distant relatives (e.g. barley, Brachypodium, and rice) are all included in gene-centric comparative analyses



Sequnence

- The chromosome survey sequence (and its annotations) are available to search via BLAST and other search alignment algorithms.
- EST-based search (with onward genomic mapping) also available

EnsemblPlants - Hatt Name Name Name Son S	entrana (1999 A Deconstructor		(c) - Sound Example Process			
Searchy (Advances)		Brand what percer	themseler rotations are been as			
ng Carbony's shall		Cataloghani Data from	The local share process in Execute Parch is the theorem and any ensemption for 200ant and any or Ohrom Development by the theorem Parch Reserved Reserved Description. The give match are provided to the beaution 1.6 and the theorem and the second			
ipider percense		Datas-tion	The star for electronegage of $\hat{\Sigma}(QG)$			
Andersteiningelie Underen Stationen	nan an		Find new much the assesses accounts and analyzes of based source provided by the part of them.			
			Encode Partie is developed in addition with other plant process and iterationation programs the LBX scale is the <u>loggPARM</u> consult The tracePLOP program is because by the <u>Encoders</u> Determined which in <u>SUPpresenterProgramme</u> , unler for low-scale area restored and 2008.			
Proceedings		barr 2010 2011 2011 2011 2011 2011 2011 20				
Log in to contention this law		EVENTS EVENTS				
2 percentes						
Solites a specific						
ns Sal Set at al Econoli Placts avecas		GRAMINE				
Nat's New in Nelsons 21			methanel by EMELER and is present by the Engineer automa system for the analysis and vaculation of process can a please (ICLNess)			
· Updated generative			Detter eta III (C'renner)			
· Updated gene models for Operation from storth-	Did you know?		Contraction of Contraction			
New Date	The Aspenditu Convolue Tast shows the conversion of a variety of amidiation					
 Core mainly for <u>200ard analogy</u> level which interveness service service services 	Some register for Pallinet analyzer from a short income a more taxanees. Surveys between official periods online of a					
 EAU (19) algoritants between tonal arteral and Daza aabaa, on well on Dechapadian deliathant 						
+ (Add-2003 alignments (where - Cardinan aviants (hering)) and the wild where progenity						
+ Addressi MM Depiters new aligned to breast effect as well as to its with progeniter						
· Montribulation chatters comparisons were also aligned to bread wheel as well as to be	and properties.					

IWGSC - Standards and Protocols PAG XXII

EMBL-EBI

	pkersey@ebi.ac.uk	Plant and Animal Genomes 2014

EMBL-EBI



EMB	L-I	EΒ		

New Search							
Search Ensembl Plants - New Search -> Gene (24)	Search resu	Its for 'RPOA'					
Ensembl Plants (24) Gene Tree	Showing 1-10 of	howing 1-10 of 12 Genes found in Ensembl Plants (filtered)					
Pan-taxonomic Compara	Filtered by specie	s: Triticum aestivum 🗙					
Configure this page	RPOA [Traos_5/	NS_6DFE37F47]					
🛃 Manage your data	Description	DNA-directed RNA polymerase subunit alpha [Source:Uniprot/SWISSPROT;Acc:P12073]					
💼 Export data	Gene ID	Traes_5AS_6DFE37F47					
	Species	Triboum aestivum					
Bookmark this page	Location	WGSC_CSS_5AS_scaff_1552572:38245-39772					
< Share this page	RPOA [Traes_38	3_1638411DC]					

9C_CSS_6AS_ece#_1652	572: 38,245-39,7	72				
al .						
				50	not CENTRAL Track	height 10160 DragSelect
			10.00			
			MARCHINE AND APPENDING			
		HERE CATE	Time 242, 144(0000		PETR - PETR -	100.011000
			(CMA)	1015101		
		- 430		-		
100	1000.00	3.9 6	2010 0216	8.810	20210	ene 400
Fuke only						
01.54540.001572.0046-0072	dere:					E-3E-8
	_	_		_	_	
			1374			Kagon
			1000 211 341 Juny 1010-	2.20	N.C.	3.41
100(4.00				_		
- marine					~~~~	
340						
	Files and Files and Files and Co. Sci Lett. 2010;2010;2010;2010;2010;2010;2010;2010					

26.02.2014	IWGSC - Standards and Protocols PAG XXII

EMBL-EBI

26.02.2014 IWGSC - Sandards and Protocols PAG XXII EMBL-EBI

Location: W050_000_545_acaR_101	0072-06.045-06.272 Game (1110	and the second second	XV.				
Scaffold IWGSC,	CSS_5AS_scatt_15525	72: 38,245-39,77	2				
0<2							
unica katokrivetec_eta							
Region in detail							
0<∋⊡≁					60	e HID het	height 1919 CraySoluti
	100	10.004	11.01.00	9.00 0.00 0.00 0	20.00 4	31.01.0	Nona Roma
naper - Confige Polater cooling Games -			en con	Tani, JAL Indiana y 170 York 200	1000 100 100 100 100 100 100 100 100 10	1000	Andread a Gana Kama Milita New Table 100 ANTEXTS
					C MILIN		Den IC 3.0 Loater Safer WIC (20.005.aut.10075
Her special polds.			1.47%		1000.0		Rate and
	10.0	10.00	100	D24 DE0	22.6		Eneral Revenue Analysis MPA
graduated in Gene Legend	Paker and the	9.84	186	D214 D.010	22.4		Pasterior Interum asstrum galast names annotation to 100
LOURIDE: MOR.CO.	4.508.003923046-0092 💽	Dens:		•			• · · · · · · · · · · · · · · · · · · ·
0 4 < 8 5							
	1.41	8.60		1010	8.01	0.47	Rept 1
Gentigen Dreitern, streiters Garra	Care -			Prost CHILING LINE, 1873			
100							
	242	2400	8.00	A.M.	8.00	241	242

1 5	26.02.2014	IWGSC - Standards and Protocols PAG XXII	EMBL-EBI

	Gene: RPOA Two AM									
ary states to (1)	Gene. HP ON Hall, SAS	Surger to								
ripl comparison fing evidence	Description	DNA-directed FNA-polymerane a								
	Location	Sorter WGSC CSS 549 acr	1 102272 36 245-08 772 mil	ine shand.						
andany Bruchare al references disc	Tranactipta	This game has 1 transcript (option	verlant) Hide transcript table							
alon.	Elevando columna		Pine .							
D' biological process (1)	Name Transcript ID			Internal I						
	APON DE TANK AND COFE	1547.2 1941 Yans SAL KO	1521fer.2 355 P	utein coding						
company prints alignments (mage)	Summary									
	oummary									
e Doo (mage) ana kwa Janij	Name	RPDA (IniProKill Dave Name)								
	Owner hape	Protein-anding								
ana painikas teo	Production Method	Triticum anativum-perses annotati	ni by MPG							
e Troe (mage)	Carto Region In	Detail for more tracks and nonligation of	ptions (e.g. zoenting)							
ana Two (ted) ana Two (alignment)	W									
cicip.el	0 (6 < 22									
i Variation					7.1(6-					
)es	No.	100	-	1.00	16.4	atten .	181
stural variation	Police sadeg (See	CENTRA ADV 1								
			m002 -	mmari-	P175-001-					
		PLD-311 - PD	Billion -	HEPB-DET-	PETS doi: -					
onal annotation	Contigen Promotion and the Continue			-				_		
079		-Tellipsont - College			CONTRACTOR IN					
ary e Natory								- Property		
ny nishoy Igure this page									_	
ary e Natory				_						
ory a history Igure this page oge your data	-	nrf 164 (96)	346	30	2.510	05	- 06			
ny nishoy Igure this page	- Contraction	194 De-	146	360	20.00	105	- 00			

1 6	26.02.2014	WGSC - Standards and Protocols PAG XXII	EMBL-EBI

Gene: RP				
	OA Trans. SAS. 60/00/787			
Description	Philcircuited (INR columns as a	burit sizha bara zeroz aberera r coro		
Location		1902172 04 Dat. 35 772 means shared		
Terecipie	Tvis pres has 1 transmipt (splice of	sectors) Hole homorist table		
Showhide	columns	Piler .		
Name	Transcript ID Langin (he) Protein ID			
PP04-002 3	TANK AND ROPERTURN THE TANK AND ROPE			
	the sea as an area in the sea of the			
GO: mole	roular function	and rearrang		
001110	cular function			
Acculated to	cular function			
Acculated to	cular function			
Acculated to	cular function			Char
Acceleration The following Electronic	cular function mean mean mean mean mean mean mean mea	in Sana Drivinge Beldenas	Anvestion Tource	020the 020the Terms Accessions
Annutational to The failure in	coular function mean forestry that g tense deaches the malecular, function of this only anteres	in Gene Drivinge	Assessment Source	ODER DOERS TARES
Acceleration The following Electronic	cular function mean mean mean mean mean mean mean mea	in Sana Drivinge Beldenas		008tin 008tin Terms Accessions 00282871 mitmake function

1 26.02.2014 WIGSC - Standards and Protecode PAG 303 EMBL-EBI

ene-based displays Summary	Gene: RPOA Trace EAS SOFE	17547			
Spice variants (1) Transcript comparison					
Supporting evidence	Description	DNA-directed RNA polymenase subunit al			
Sequence	Location	Scaffold M/OSC OSS 5AS scaff 15525	72-38.045-39.775	reverse shand.	
 Secondary Structure External references 	Transcripts	This gene has 1 transcript (splice variant)		able	
Pequiation					
Expression	Showhide columns		Filter		
Ontokozy	Neme 1 Transcript ID	Length (bp) : Protein ID	Length (se)	Dichor	
- GO: biological process (1)	RPOA 202 Trees 5A8 (OFE37F47)			Protein codino	
GO: molecular function (3) Plant Company	THE SHE CONSTRAINED	1015 505 507 527 47	200	1.1100 COLO \$	
- Genomic alignments (image)	GO: molecular function				
- Genomic alignments (text) Gene tree (mage)					
- Gone tree deats	Annotated territion Ancestry chart				
Gene tree (alignment)					
- Gene gainfoss tree - Orthologues (30)	Terms:				redecular function
Panalogues (44)	Annotated terms				
Pan-taxonomic Compana 					
- Gene Tree (mage)	Denario OO alim terma			catal	ytic activity binding
Gono Tree (alignment)				-	
- Orthologues Pherotype				transforase octivity	organic cyclic heterocyclic protein binding
Genetic Variation	Plant sim terms			CRAPTER OF COMPANY	compound binding compound binding
 Variation table 					
 Bructural variation Variation image 	Relations:			Itanifetase	
External data	t is_a + pat of			activity,	mucheic sold binding protein dimerization
Personal annotation	+ bec's			transferring	activity
ID History Gene history				containing groups	
Configure this page					
Configure Pes page				rac locitic/immediance	
Managa your data				activity	DNA binding
b Export data					
Bookmark this page				RNA polymetase activity	
Share this page				DNA directed DNA	
Downiced view as CBV				polymerape activity	

		36.249-28.772 Gene FEGA Transactor MPGA-202					
necript-based displays Summary Summary	Transcript: RPOA-20	22 Teen, AAA, KDPENTYCL					
Easter (7)	Executation	ONA directed FMA polymerane suburit sights back participations in a figure	0				
	Location	Souther INDEG. C55, 545, and 1952572, 38,245,39,751 www.wine	6				
Protein Idential Datasercas	Dene	This transmipt is a product of gave Taxes, MR. 824922742					
General Monthlers Oligo protes		This gave has 1 thereerist isplice variant) Index transmitt table					
GO Mological process (1) GO molecular function (3)	Showhide columns	The	3				
enetis Variation Variation table Variation image Passiellon properties	Name Temeripi I APGA-302 Tass SNI 6077		-				
Pspulation comparison Comparison image relate information	Protein summary						
Protein summery Domains & Instance (18)	Protein domains for Trees,	EAS_SOFE37F47.3.1					
	0						
Protein structure Hernal Sola Personal annetation I Makery	Topos Sel, AVYC ST. NLAP Met America Sources and Average	NATURA NATU			_	INTER-	_
	Part density PARCE-SPL density	REAL PROFESSION FROM FROM REPLACED TO THE ACTION OF THE AC				100,00	_
Protain Notrry				10		1000 0.000	
Pepale Hotory Configure this page	Garrel E. Annah	accepted become to	10				
Protein Noticy Configure this page Warnage your data	Distance	Re reside reside 11357 gind	¥				
Proside history Configure this page Wanage your data Export data	D took in:	* * U	<i>w</i>				

			5.7530			Forward al eard
Daal.Chr. 1	41.716Mb	41.717Mb	81.71 Mb	41.7199b	H.720Mb	41.721Mb
VSU gene.	LOC_Os01g71960 >					
ROSP gene	P0466H10.36 >					
Contige IRGSP gene			1.41700001.418000	0 -		_
VBU gene. D. sah T. aes ikisto net			<loc_0601g71990< td=""><td></td><td></td><td></td></loc_0601g71990<>			
Osat.Chr. 1	41.71 GMs	40.717365	41,71 8Mb 5,7510	41.719Mo	41.720Mb	41.721Mb
« 4 Q D 1 I Q	F #					
ave Scotloid IWGS.	3a	42	- 17916/ Sib		- 54	Powerd at text
Protein-cooking Gene	_SR_CADBR/SBB >	_				
	SB_DB#BFBHDC >					
Configs Prote in-cooking Gene			WOSC, CSS, SR, Kerky, S			
Scip-Tues Insta-cet			<t066_38_1e51< td=""><td>dices</td><td></td><td></td></t066_38_1e51<>	dices		
Des Scoffold WOB	34b	41	30 37910		ha	7
# 4 Q I Q + *		W				
kdis Chr. 2 Idio T.aes ikstoret	56.02.416	58.02.308	57510 58.0234b	Sectored	54 (2016)	Person stan
WPSUG gene.	BPAD to George >	_	_			
Contrigge			< 2.58000001.58100	000		
WPSUKI gene			-	BRADIDGeo730		_
3dis Chr. 2	58.02486	58.02.308	58.02345	56.02114	58.02018	

1 25.02.2014 INVGSC - Standards and Protocols PAG XXII EMBL-EBI

2 0	WGSC - Standards and Protocols PAG XXII	EMBL-EBI

Active tracks	Wheat SNPs and alignments	(m) + 0 + 0
Track order Beach reachs Begarree and assembly (JP4) Beach reachs Genes and tensorphis (JP4) December 2015 (JP4) Production tensorphis (JP4) Production tensorphis (JP4) Production tensorphis (JP4) Production tensorphis (JP4) Beach Academic (JP4) Minut Academic (JP4) Minut Academic (JP4) Minut Academic (JP4) Minut Academic (JP4) Minut Academic (JP4)	bateline and a final data and a difference of the second	* 6 * 4 * 4 * 6 * 4 * 6 * 6 * 6 * 6 * 6
Pagest regions (2014) Information and decorations (2010) Dealphy optimal Serve configuration as	Item Constantion Form For	÷ (
All your data	Please rule had be solited of extend tools is not the segmentality of the Diverted project. URL-based or DAS tools may other dow down our extend to branch properties OR may be unavailable as these are served and stand-from other servers abandwars on the Internet.	

nthan UsGen Iorne dia 878.20 100			1.00			1.22	1.00	Las	140	k.co	1.22	1.0	Lee	Last the second
100 March 100		Thur and a law					_	_	_	-				
lan as diser (919 130	there are the	Titue and a Class							_	_	_			
turosatur#03 130	No. of Lot	These service parts												
iner ne al cual de la	No. artes	The second per-			una la region I de la antechera		-							
101.04.05.00 ^{10.10}	No. artes	Terrancepe	reingener		your frame your	and the set	-	_	-		-			
10	-	The second stage					_	_	_					
120	there are the	Terre milita etal	an 20 has 1	and II day of	ngilikana ani 150			_	-					_
ur os deur PRO 1.0	they will be	Thus with a right		utt. young Re	-	(1)/2 Lat &	_		_					
eix ood og Dene					MACOUNT -					_				
Rat. Infrario Demone	L.	= .	1			hin	alad					-		
	and Barrette street		1.00	_	-00	1.00	1.40	Los	1.00	1.00	6.00	Lee	1.00	1.00
	petitin-coc				011C. CR8. ADL av									

2 26.02.2014 INVGSC - Standards and Protocols PAG XXII 2	EMBL-EBI
---	----------

Inter-homeologeous variants

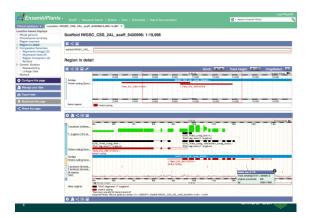
- Data from Brenchley et al. mapped onto chromosome survey sequence
- Data from chromosome survey sequence itself due in next release

		variants (mapped)	
A/D	102,593	101,570	99
В	22,959	29,638	99
Total	132,552	131,208	99

23 19.02.2013 pkersey@ebi.ac.uk Plant and Animal Genomes 2014

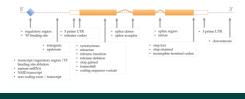
EMBL-EBI

EMBL-EBI 🌒



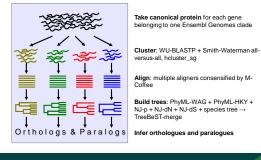
Variant Effect Predictor (VEP)

- · Predicts functional consequences of known and unknown variants
- · For substitutions, insertions, deletions and structural variants
- Web interface (for up to 750 variants), standalone Perl script, Perl API and REST API
- http://www.ensembl.org/info/docs/variation/vep/index.html



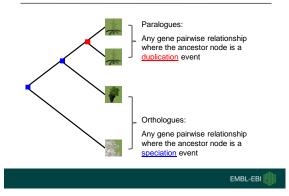
EMBL-EBI

Gene tree pipeline



EMBL-EBI

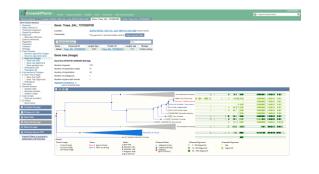
Orthologues and paralogues



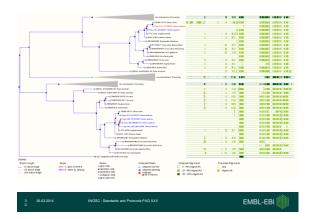
Orthology / paralogy types

- ortholog_one2one
- ortholog_one2many
- ortholog_many2many
- apparent_ortholog_one2one
- possible_ortholog (weakly supported duplication node)
- within_species_paralog
- other_paralog (too distant to be in the same tree)
- contiguous_gene_split (artefact)
- putative_gene_split (artefact)
- 94,280 wheat genes in inferred orthology relationships with 270216 genes from other creals

EMBL-EBI







Visualise your own data

Upload data

- Data saved by Ensembl
- 5 MB limit (and therefore not possible for large file formats)

Attach remote file

- URL-based (HTTP or FTP)
- No size limit
- Uploaded / attached tracks can be saved (to account)
- · Uploaded / attached tracks can be shared with other users
- · Only trivial security, do not use for sensitive data!
- <u>http://www.ensembl.org/info/website/upload/index.html</u>

EMBL-EBI

Possible formats

•	BAM	sequence alignments	(no upload)
---	-----	---------------------	-------------

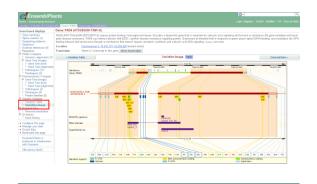
- BED genes / features
- BedGraph continuous-valued data
- BigBed genes / features (no upload)
- BigWig continuous-valued data (no upload)
- TrackHub collection of tracks
- Gbrowse genes / features
- GFF/GTF genes / features
 PSI sequence alignments
- PSL sequence alignments
 VCF variants (no upload)
- VCF variants (no upload) WIG continuous-valued data
- VIG continuous-valueu da

EMBL-EBI

EMBL-EBI

Next release

- · Improved representation of the polyploid genome
 - Preconfigured A v B v D compartive views
 - · Separation of 3 wheat genomes in the gene tree analysis
- Representation of inter-homelogue data as a full variation database
- Data structure can also accommodate inter-varietal SNPs, links to phenotypes, haplotype structure etc.



EMBL-EBI

Funding

- Ensembl Genomes Funded by
 - EMBL
 - EU (INFRAVEC, Microme, transPLANT, AllBio)
 - BBSRC (PhytoPath, wheat/barley/midge sequencing, UK-US collaboration, RNAcentral)
 - Wellcome Trust (PomBase)
 - NIH/NIAID (VectorBase)
 - NSF (Gramene collaboration)
 - Bill and Melinda Gates Foundation (wheat rust)

People

- James Allen, Irina Armean, Dan Bolser, Mikkel Christensen, Paul Davies, Lee Falin, Christoph Grabmueller, Kevin Howe, Malcolm Hinsley, Jay Humphrey, Arnaud Kerhornou, Julia Khobdova, Eugene Kulesha, Nick Langridge, Dan Lawson, Mark McDowall, Uma Maheswari, Gareth Maslen, Michael Nuhn, Chuang Kee Ong, Michael Paulini, Helder Pedro, Anton Petrov, Dan Staines, Mary Ann Tuli, Brandon Walts, Gary Williams
- The Ensembl teams @ EBI (Paul Flicek) and WTSI (Steve Searle)

	26.02.2014	pkersey	@ebi.ac

EMBL-EBI