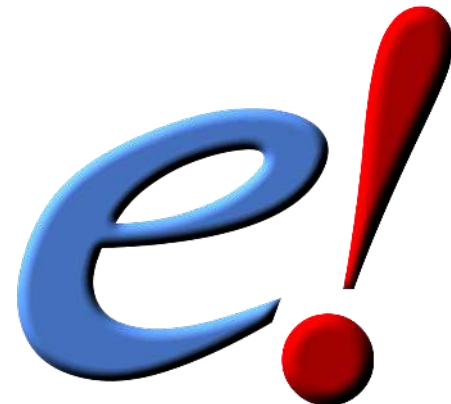


Ensembl Genome Browser

Comparative Genomics



Overview

- Rationale
- Comparative proteomics
 - Homologue predictions (Gene trees)
 - Protein family clustering
- Comparative genomics
 - Genome-wide DNA alignments
 - Syntenic block characterisation

Species Comparison

- From the user perspective
 - Where are orthologous/paralogous genes?
 - Is the gene order conserved?
 - Where are syntenic regions located?
 - How many genes are conserved?
 - What is missing in one species, but present in another?
Species or lineage-specific genes?
 - Where are potential regulatory regions?
- From the Ensembl perspective
 - Orthologous/paralogous gene links
 - Chromosome synteny links
 - Sequence comparison links
 - Protein family links

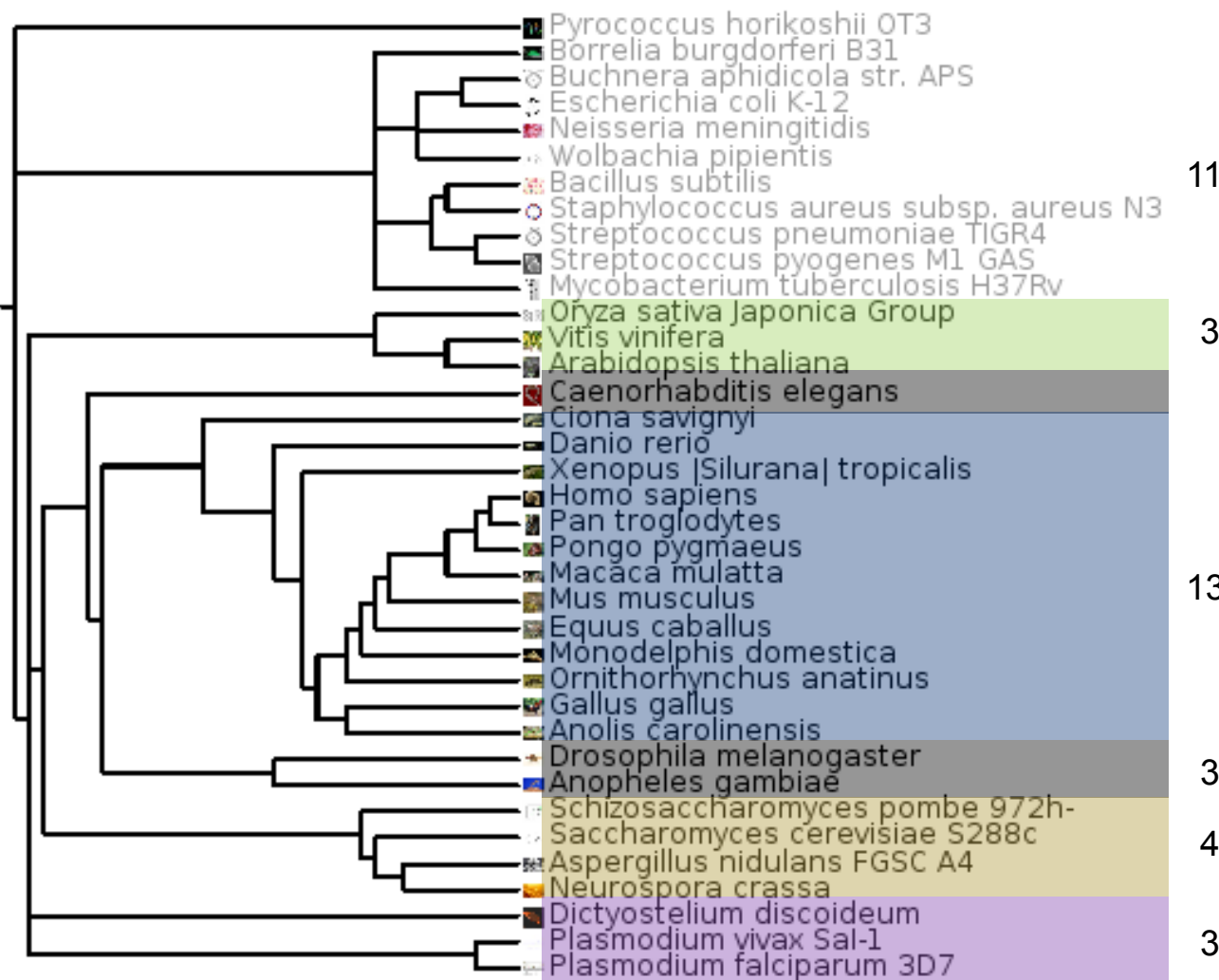
Species in Ensembl



Ensembl Species

 Alpaca <i>Vicugna pacos</i>	 Gorilla <i>Gorilla gorilla</i>	 Pika <i>Ochotona princeps</i>
 Anole Lizard <i>Anolis carolinensis</i>	 Guinea Pig <i>Cavia porcellus</i>	 Platypus <i>Ornithorhynchus anatinus</i>
 Armadillo <i>Dasypus novemcinctus</i>	 Hedgehog <i>Erinaceus europaeus</i>	 Rabbit <i>Oryctolagus cuniculus</i>
 Baboon (preview - assembly only) <i>Papio hamadryas</i>	 Horse <i>Equus caballus</i>	 Rat <i>Rattus norvegicus</i>
 Bushbaby <i>Otlemur garnettii</i>	 Human <i>Homo sapiens</i>	 Saccharomyces cerevisiae <i>Saccharomyces cerevisiae</i>
 Caenorhabditis elegans	 Hyrax <i>Procavia capensis</i>	 Sheep (preview - assembly only) <i>Ovis aries</i>
 Clona intestinalis	 Kangaroo rat <i>Dipodomys ordii</i>	 Shrew <i>Sorex araneus</i>
 Clona savignyi	 Lamprey (preview - assembly only) <i>Petromyzon marinus</i>	 Sloth <i>Choloepus hoffmanni</i>
 Cat <i>Felis catus</i>	 Lesser hedgehog tenrec <i>Echinops telfairi</i>	 Squirrel <i>Spermophilus tridecemlineatus</i>
 Chicken <i>Gallus gallus</i>	 Macaque <i>Macaca mulatta</i>	 Stickleback <i>Gasterosteus aculeatus</i>
 Chimpanzee <i>Pan troglodytes</i>	 Marmoset <i>Callithrix jacchus</i>	 Tarsier <i>Tarsius syrichta</i>
 Cow <i>Bos taurus</i>	 Medaka <i>Oryzias latipes</i>	 Tetraodon <i>Tetraodon nigroviridis</i>
 Dog <i>Canis familiaris</i>	 Megabat <i>Pteropus vampyrus</i>	 Tree Shrew <i>Tupaia belangeri</i>
 Dolphin <i>Tursiops truncatus</i>	 Microbat <i>Myotis lucifugus</i>	 Turkey (preview - assembly only) <i>Meleagris gallopavo</i>
 Duck (preview - assembly only) <i>Anas platyrhynchos</i>	 Mouse <i>Mus musculus</i>	 Wallaby <i>Macropus eugenii</i>
 Elephant <i>Loxodonta africana</i>	 Mouse Lemur <i>Microcebus murinus</i>	 Xenopus tropicalis
 Fruitfly <i>Drosophila melanogaster</i>	 Opossum <i>Monodelphis domestica</i>	 Zebra Finch <i>Taeniopygia guttata</i>
 Fugu <i>Takifugu rubripes</i>	 Orangutan <i>Pongo pygmaeus</i>	 Zebrafish <i>Danio rerio</i>
 Giant Panda <i>Ailuropoda melanoleuca</i>	 Pig <i>Sus scrofa</i>	

Ensembl Genomes Pan-taxonomic Comparison

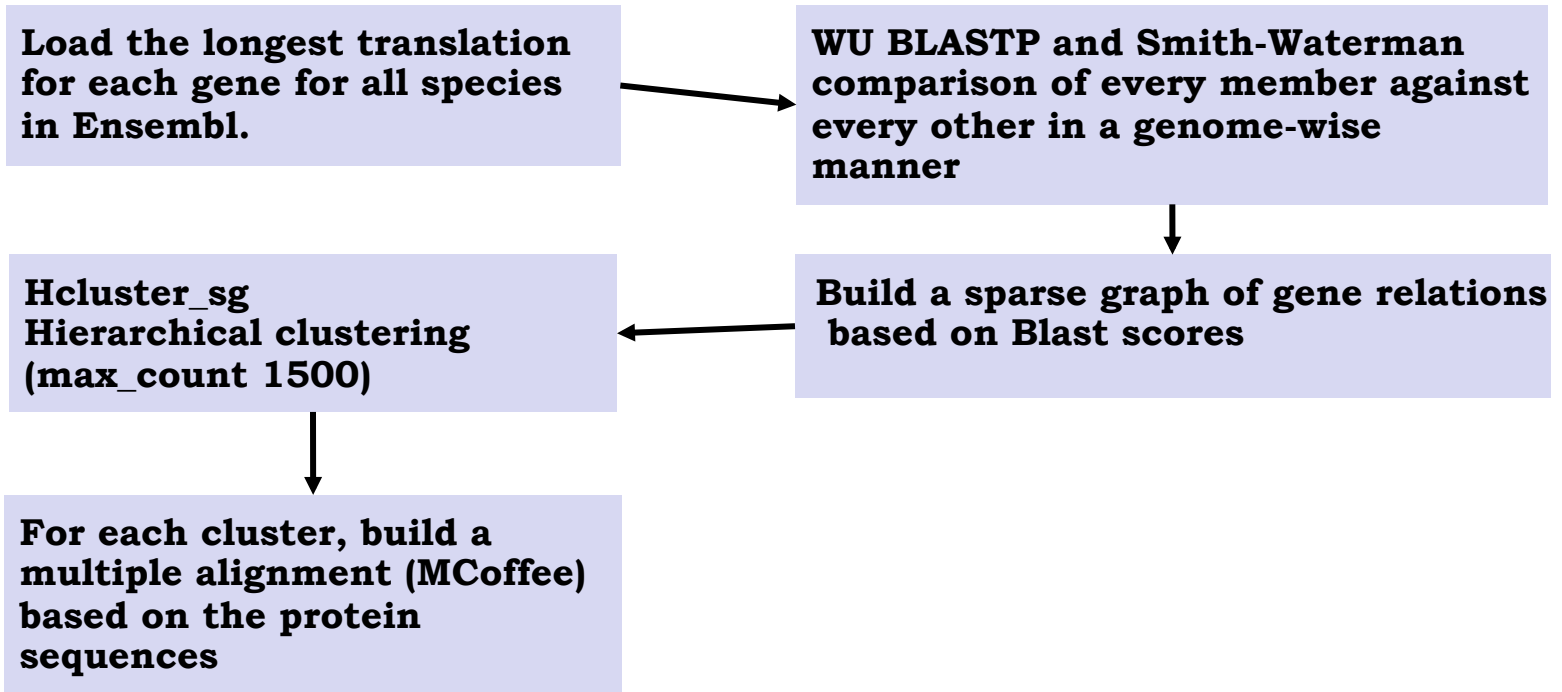


- Rationale
- Comparative proteomics
 - **Homologues predictions (Gene Trees)**
 - Protein family clustering
- Comparative genomics
 - Genome-wide DNA alignments
 - Syntenic block characterisation

Gene Trees

- Used for gene orthology and paralogy predictions
- Represent evolutionary relations of gene families, i.e. genes that diverged from a common ancestor
- Pipeline based on maximum likelihood phylogenetic gene trees (TreeBest)
- Trees reconciled with species trees to infer duplications and speciations
- For some dN/dS calculations

Gene Tree Prediction Pipeline

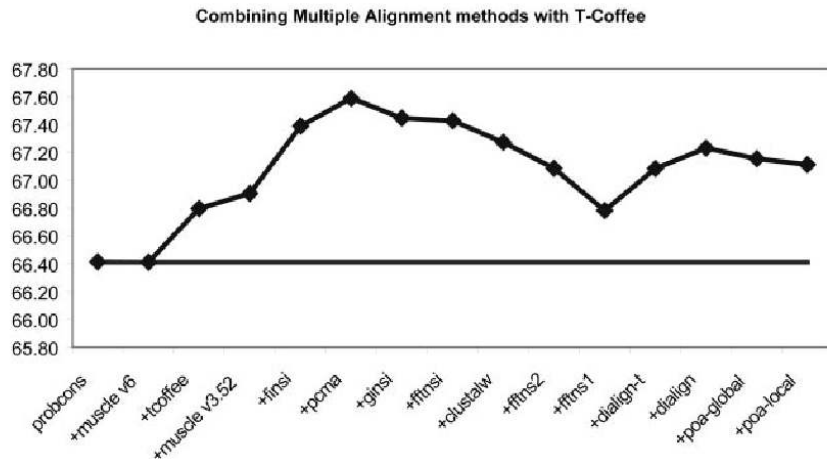


EnsemblCompara GeneTrees: Analysis of complete, duplication aware phylogenetic trees in vertebrates. Vilella AJ et al. Genome Res. 2008 Nov 24.

M-COFFEE

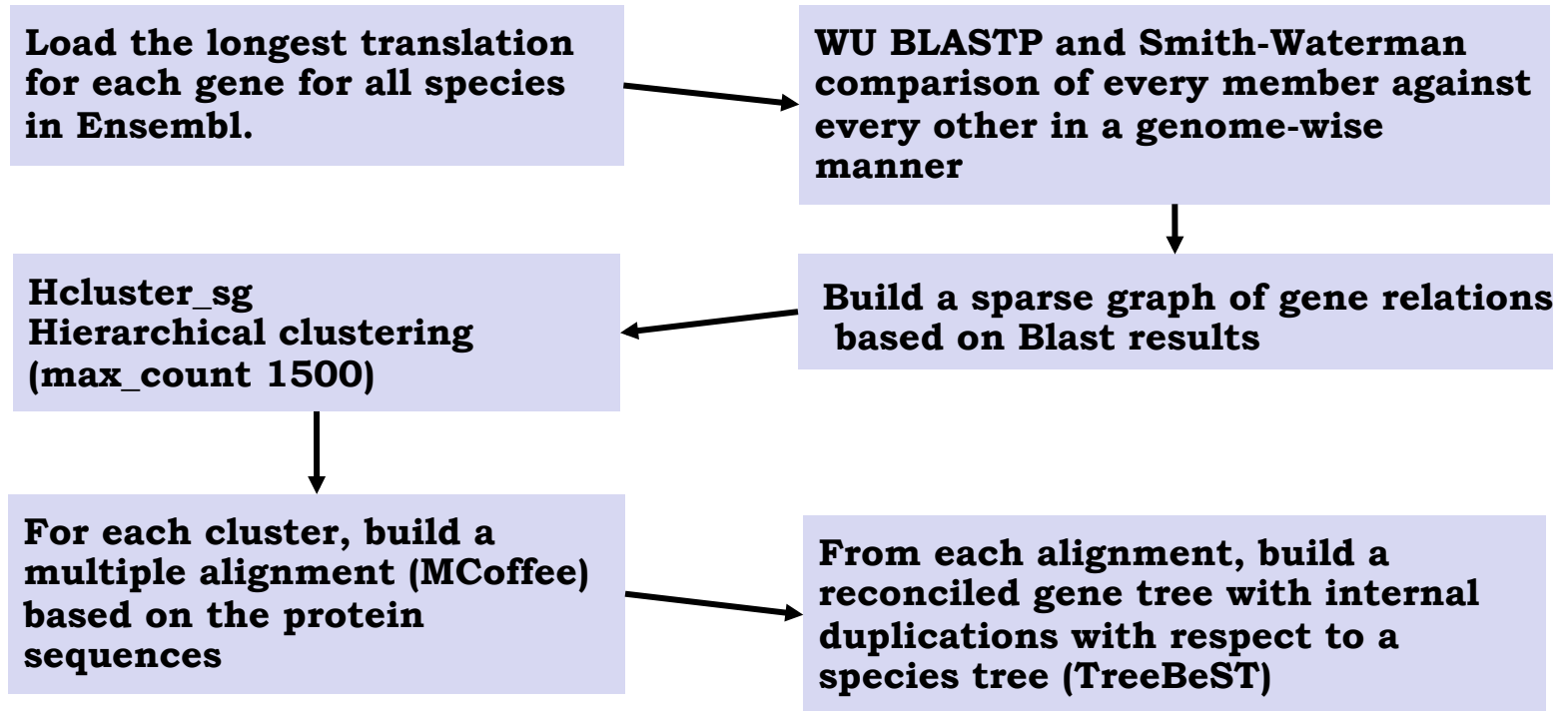
- Computes a **multiple sequence alignment (MSA)** and the associated **phylogenetic tree** by combining the output of several multiple sequence alignment packages
- T-coffee to produce final alignment

- By default, M-Coffee uses:
 - ClustalW2
 - Mafft
 - Dialign
 - POA
 - ProbCons
 - MUSCLE
 - T-Coffee
 - PCMA



Notredame et al. *T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment*. JMB, 302 (2000): 205-217

Gene Tree Prediction Pipeline

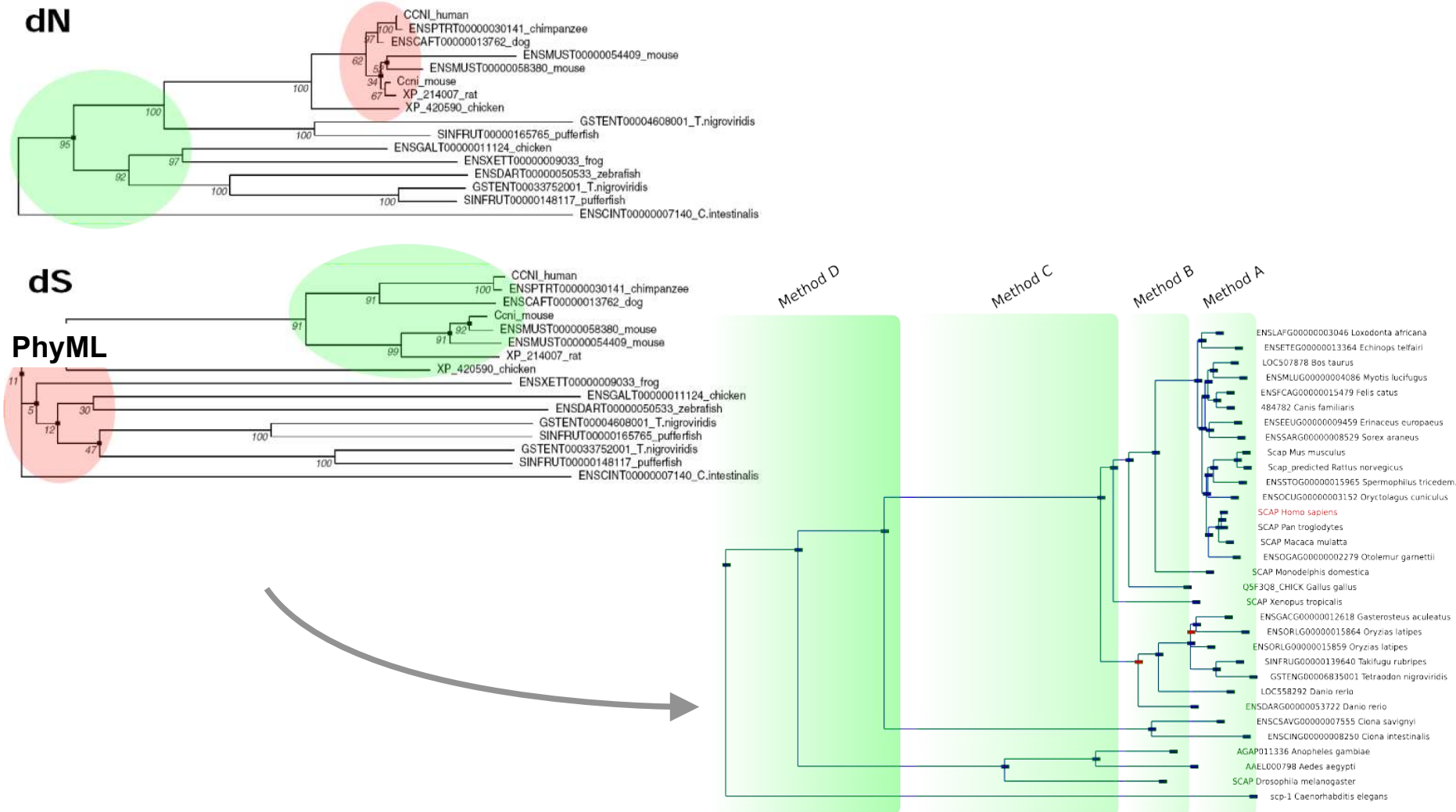


EnsemblCompara GeneTrees: Analysis of complete, duplication aware phylogenetic trees in vertebrates. Vilella AJ et al. Genome Res. 2008 Nov 24.

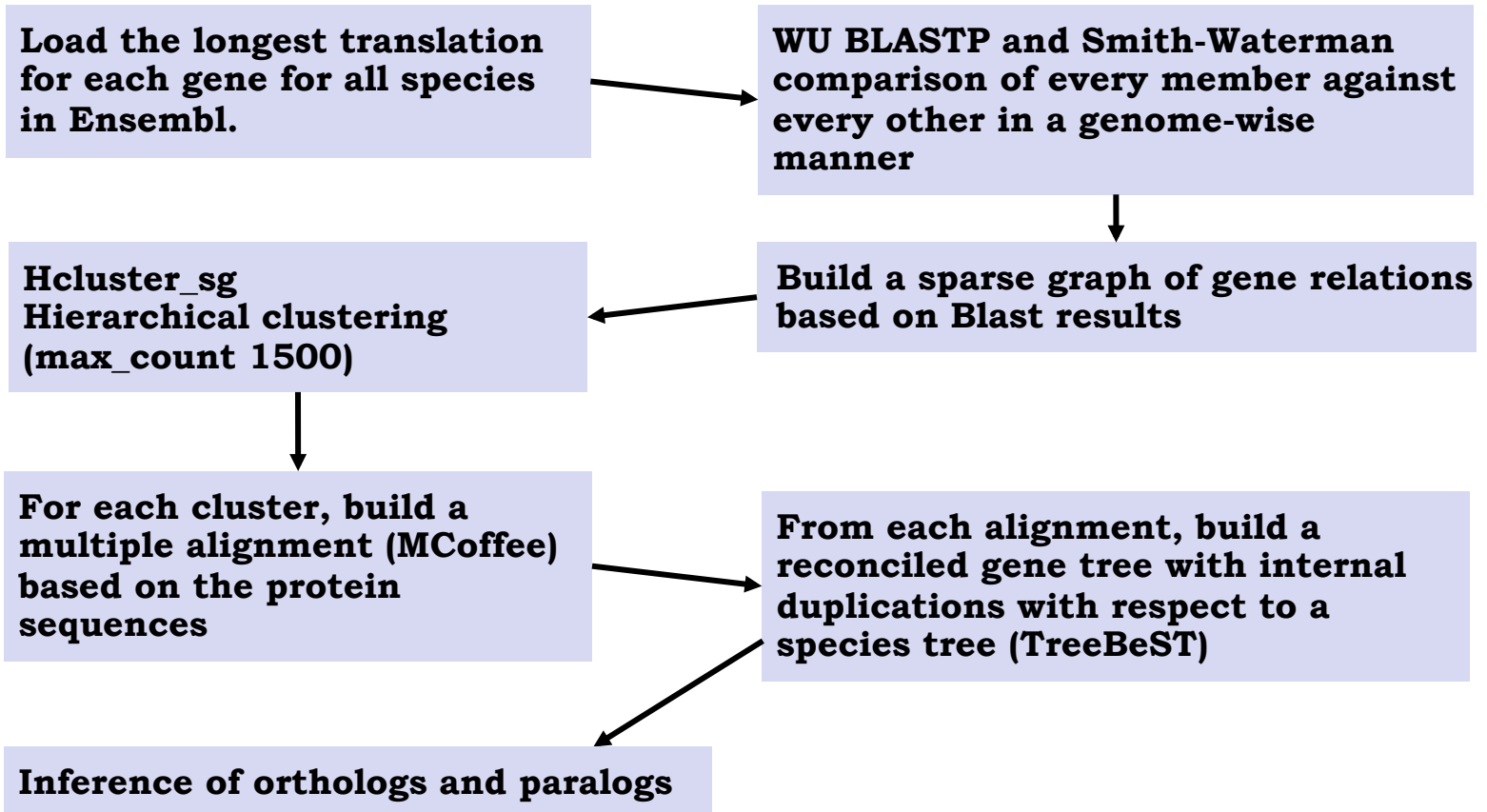
TreeBeST - treemerge

- ML-AA-WAG4 – WAG matrix amino acid model – Maximum Likelihood (PHYML)
- ML-NT-HKY85 – Hasegawa-Kishino-Yano nucleotide model – Maximum Likelihood (PHYML)
- NJ-NT-p-distance – any substitutions – neighbour-joining with bootstrap
- NJ-NT-dN – non-synonymous substitutions – neighbour-joining with bootstrap
- NJ-NT-dS – synonymous substitutions – neighbour-joining with bootstrap

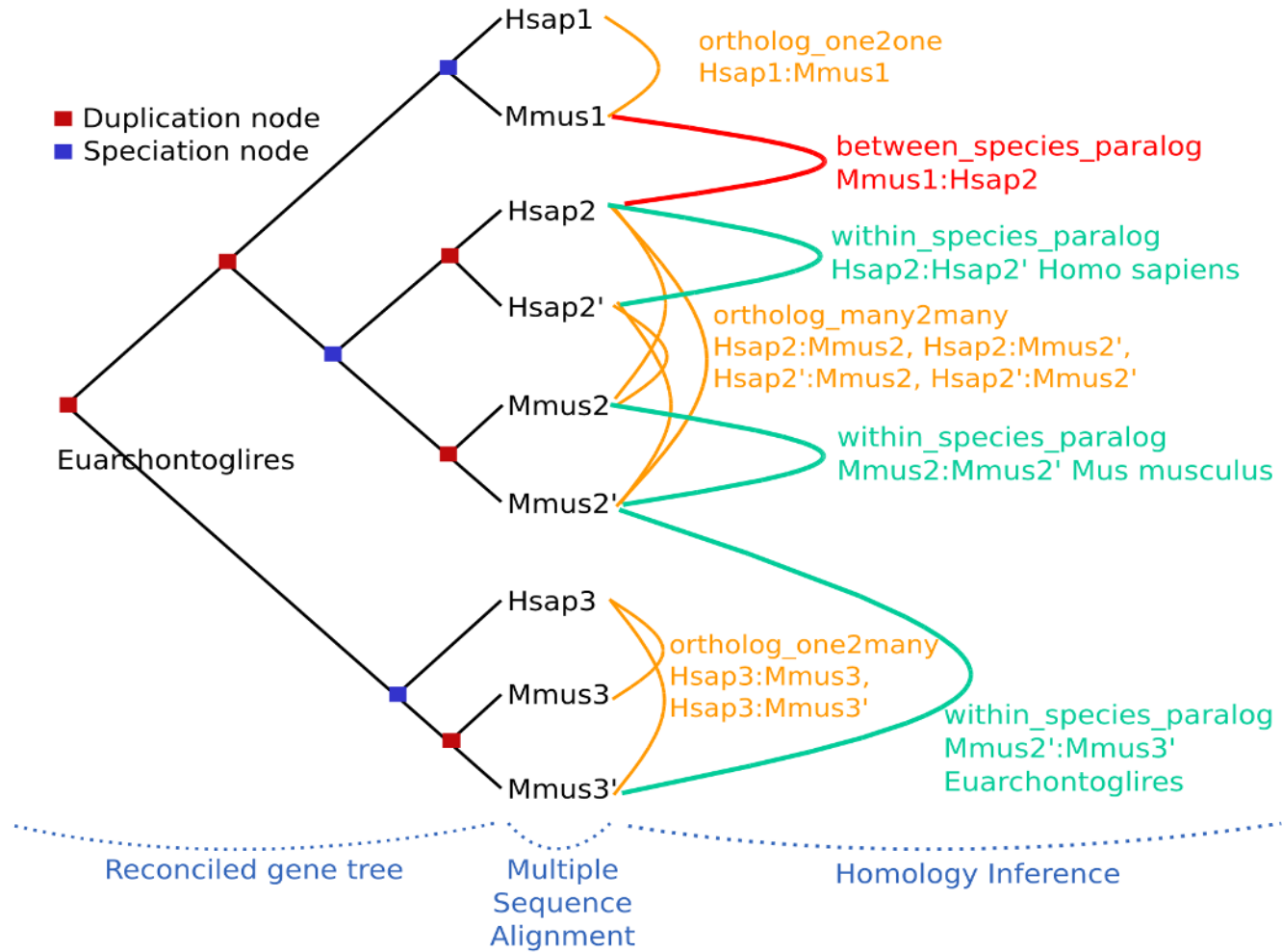
Each method performs best at a given level



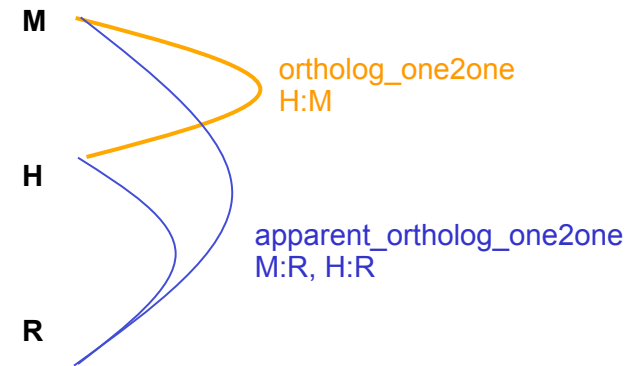
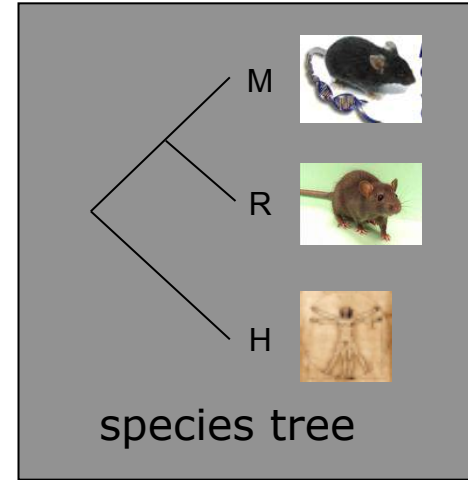
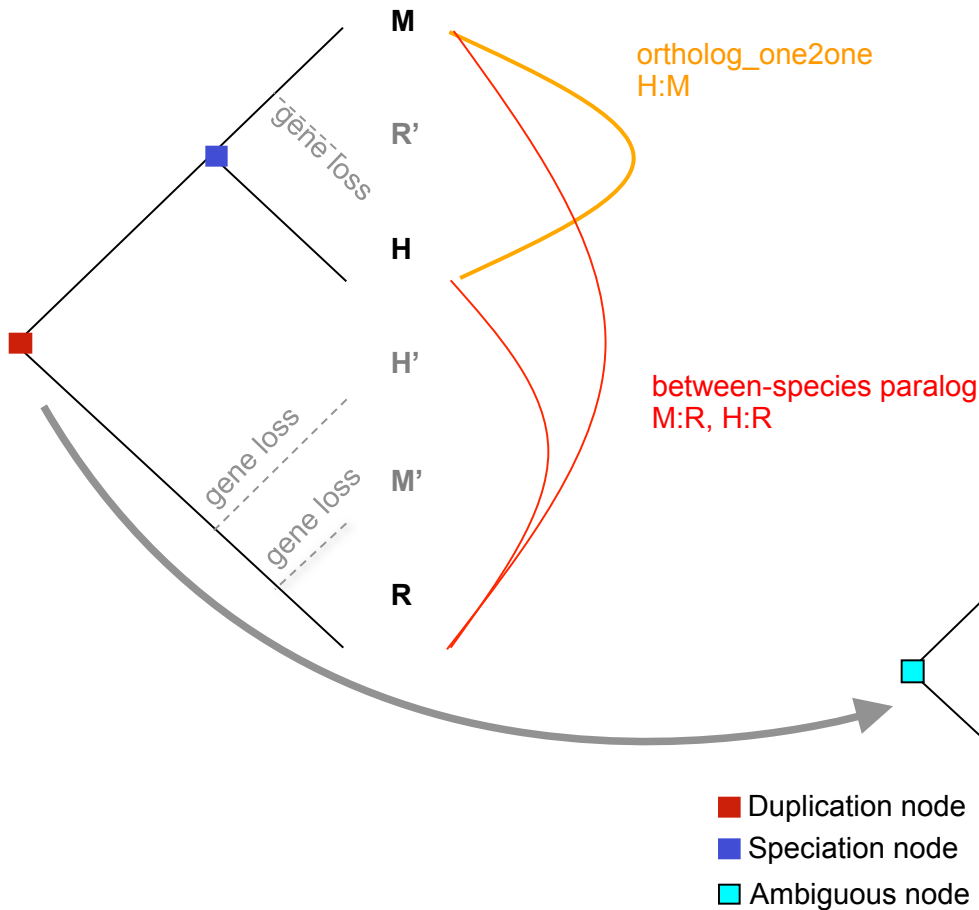
Gene Tree Prediction Pipeline



Homology inference



A special case of homology



Gene Tree Display

Ensembl
Home > Human
Location: 4:185,548,850-185,570,629 Gene: CASP3

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references (2)
- Regulation
- Comparative Genomics
 - Genomic alignments (3)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (71)
 - Paralogues (12)
- Protein families (1)
- 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: CASP3 (ENSG00000164305)

Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apoptain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17:Caspase-3 subunit p12] Source: UniProtKB/Swiss-Prot P42374

Location Chromosome 4: 185,548,850-185,570,629 reverse strand.

Transcripts There are 6 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
CASP3-001	ENST00000308394	ENSP00000211032	protein_coding
CASP3-002	ENST00000339365	ENSP00000377210	protein_coding
CASP3-003	ENST00000339368	ENSP00000377213	protein_coding
CASP3-004	ENST00000447121	ENSP00000407142	protein_coding
CASP3-201	ENST00000352650	ENSP00000296745	protein_coding
CASP3-202	ENST00000438467	ENSP00000380792	protein_coding

Genomic alignments Gene Tree (image) [help](#) Orthologues

Gene tree

Multiple Sequence Alignment Ideograms

View options:

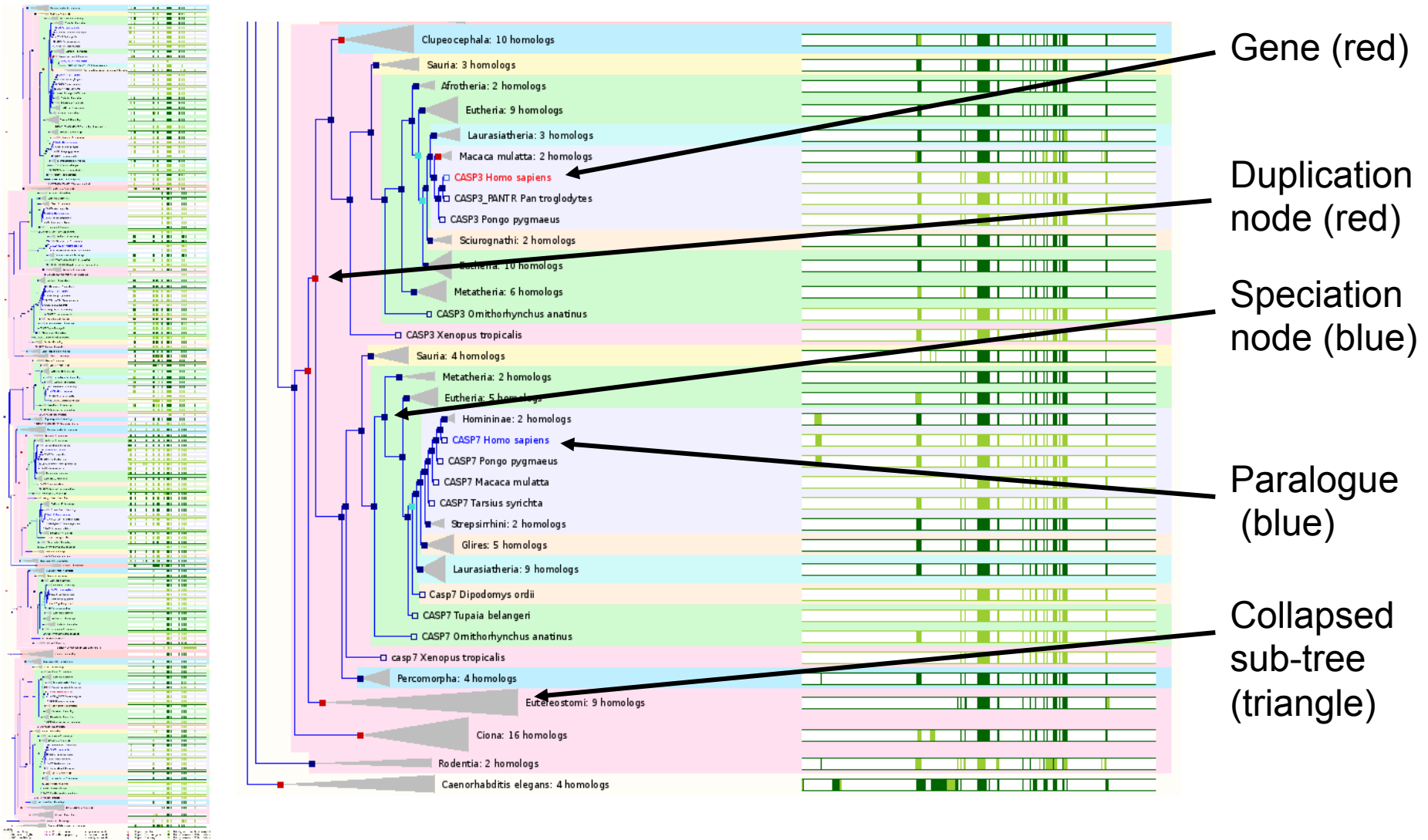
- View current gene only
- View paralogs of current gene
- View all duplication nodes
- View fully expanded tree

Use the 'configure page' link in the left panel to set the default. Further options are available from menus on individual tree nodes.

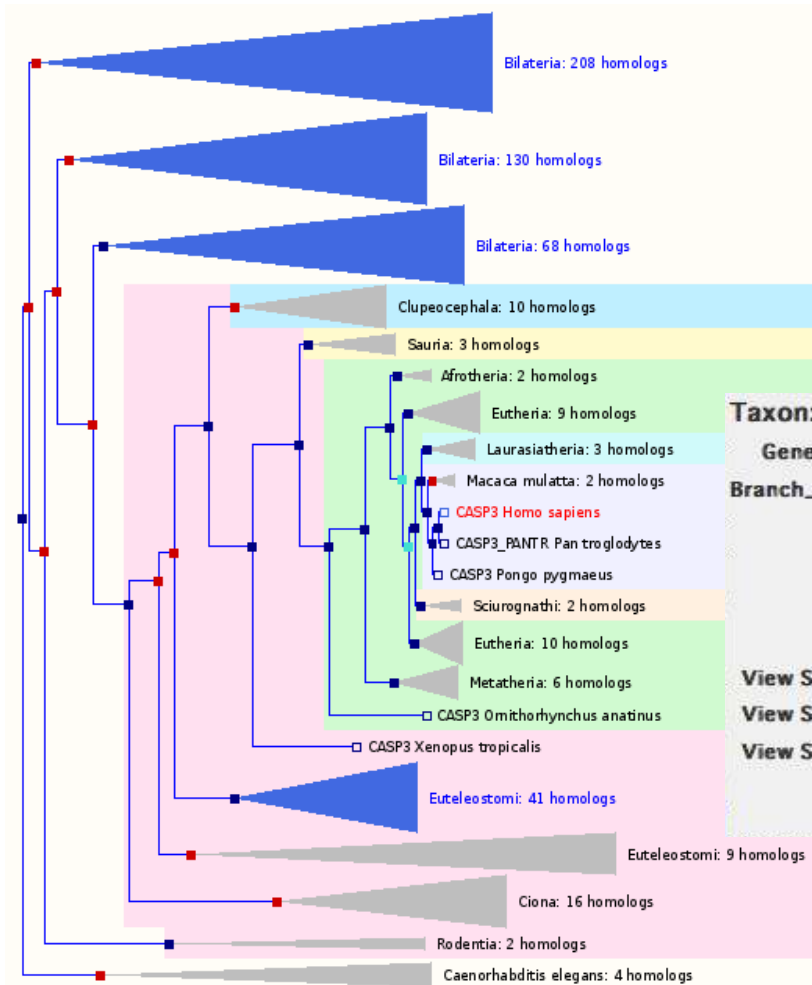
Ensembl release 55 - July 2009 © WTSI / EMBL
Permanent link - View in archive site

About Ensembl | Contact Us | Help

Gene Tree Annotation



Gene Tree Navigation



LEGEND

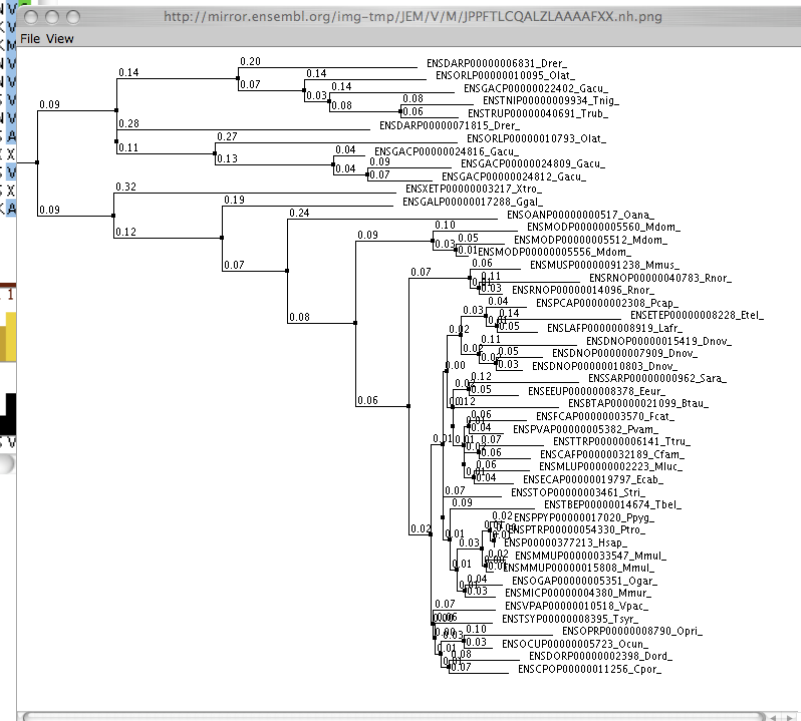
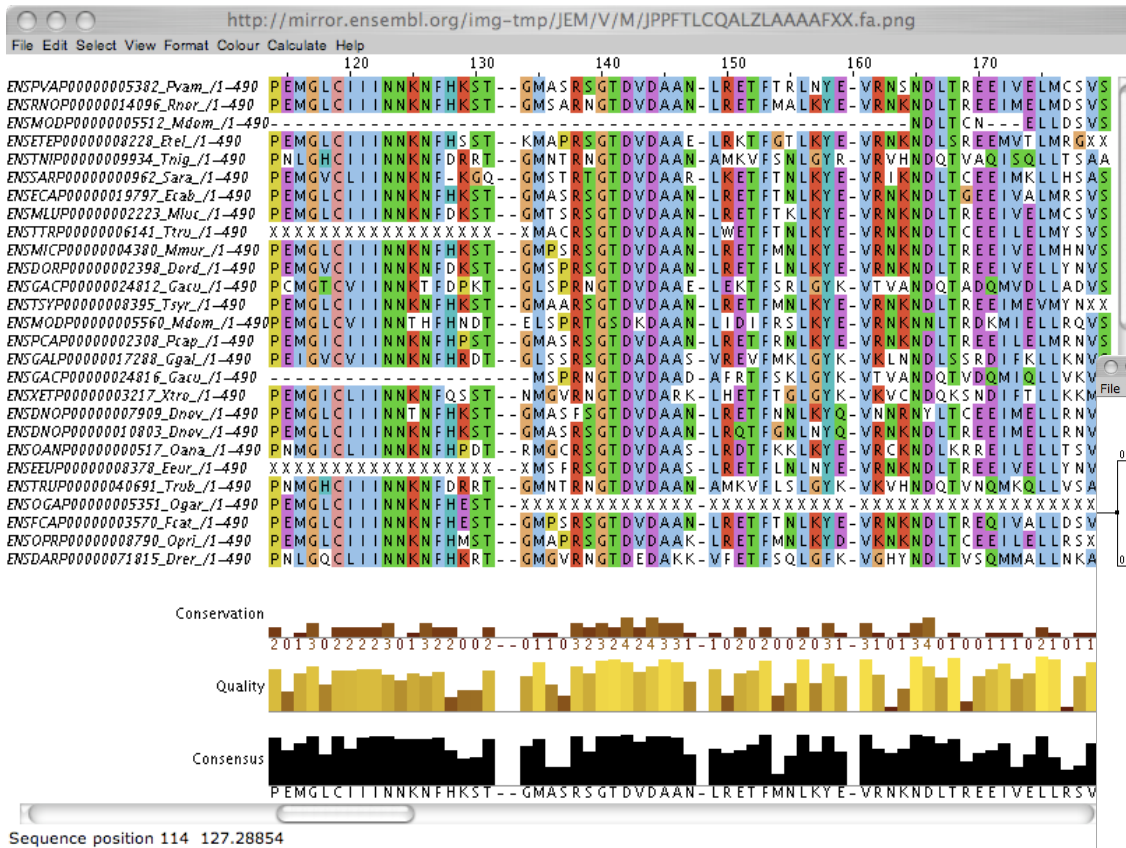
- x1 branch length
- x10 branch length
- x100 branch length
- Gene ID current gene
- Gene ID within-sp. paralog
- speciation node
- duplication node
- ambiguous node

Menu items to collapse or expand a sub-tree

Taxon: Eutheria (Placentals)	X
Gene_Count	15
Branch_Length	0.00444
Type	Speciation
Image	expand this sub-tree
Image	expand all sub-trees
Image	collapse other nodes
View Sub-tree	Alignment: FASTA
View Sub-tree	Tree: New Hampshire
View Sub-tree	[Requires Java]

Button to start the Jalview alignment editor

Jalview Alignment Editor



Orthologue Predictions

[Home](#) > [Human](#) [GRCh37]

Location: [4:185,548,850-185,570,629](#) | Gene: [CASP3](#)

Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
 - Genomic alignments (5)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (74)**
 - Paralogues (12)
 - Protein families (1)
- 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: CASP3 (ENSG00000164305)

Caspase-3 Precursor (CASP-3)([EC 3.4.22.56](#))(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot P42574](#)

Location [Chromosome 4: 185,548,850-185,570,629](#) reverse strand.

Transcripts There are 6 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
CASP3-001	ENST00000308394	ENSP00000311032	protein_coding
CASP3-002	ENST00000393585	ENSP00000377210	protein_coding
CASP3-003	ENST00000393588	ENSP00000377213	protein_coding
CASP3-004	ENST00000447121	ENSP00000407142	protein_coding
CASP3-201	ENST00000352650	ENSP00000296745	protein_coding
CASP3-202	ENST00000438467	ENSP00000390792	protein_coding

[« Gene Tree \(image\)](#)
[Orthologues help](#)
[Paralogues »](#)

The following gene(s) have been identified as putative orthologues:
(N.B. If you don't find a homologue here, it may be a "between-species paralogue". Please view the [gene tree info](#) to see more.)

Species	Type	dN/dS	Ensembl identifier	External ref.
Alpaca (<i>Vicugna pacos</i>)	1-to-1	na	ENSVPAG000000011298 Target %id: 86; Query %id: 61 [Multi-species view] [Align]	CASP3 Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot; acc: P42574]
Anole Lizard (<i>Anolis carolinensis</i>)	1-to-1	na	ENSACAG000000002733 Target %id: 69; Query %id: 64 [Multi-species view] [Align]	CASP3 Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot; acc: P42574]
Armadillo (<i>Dasypus novemcinctus</i>)	1-to-many	0.22486	ENSNOG000000013931 Target %id: 68; Query %id: 68 [Multi-species view] [Align]	Novel Ensembl prediction No description
	1-to-many	0.47405	ENSNOG000000019881 Target %id: 66; Query %id: 48 [Multi-species view] [Align]	Novel Ensembl prediction No description
Bushbaby (<i>Otolemur garnettii</i>)	1-to-1	na	ENSOGAG000000005985 Target %id: 78; Query %id: 78 [Multi-species view] [Align]	CASP3 Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot; acc: P42574]
Cat (<i>Felis catus</i>)	1-to-1	na	ENSFCAG000000003870 Target %id: 73; Query %id: 73 [Multi-species view] [Align]	CASP3 Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot; acc: P42574]
Chicken (<i>Gallus gallus</i>)	1-to-1	na	ENSGALG000000010638 Target %id: 63; Query %id: 64 [Multi-species view] [Align]	NP_990056.1 caspase 3 [Source:RefSeq peptide;Acc:NP_990056]
Chimpanzee (<i>Pan troglodytes</i>)	1-to-1	0.07500	ENSPTRG000000016640 Target %id: 99; Query %id: 99 [Multi-species view] [Align]	CASP3_PANTR Caspase-3 Precursor (CASP-3)(EC 3.4.22.56) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot; acc: Q5IS54]
Ciona savignyi	many-to-many na		ENSMSAVG000000001674 Target %id: 35; Query %id: 38 [Multi-species view] [Align]	Novel Ensembl prediction No description
	many-to-many na		ENSMSAVG000000001773 Target %id: 41; Query %id: 27 [Multi-species view] [Align]	Novel Ensembl prediction No description

Paralogue Predictions



Home > Human [GRCh37]
Location: 4:185,548,850-185,570,629 Gene: CASP3

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
 - Genomic alignments (5)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
- Orthologues (74)
- Paralogues (12)**
- Protein families (1)
- 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: CASP3 (ENSG00000164305)
Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot P42574](#)
Location [Chromosome 4: 185,548,850-185,570,629](#) reverse strand.
Transcripts There are 6 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
CASP3-001	ENST00000308394	ENSP00000311032	protein_coding
CASP3-002	ENST00000393585	ENSP00000377210	protein_coding
CASP3-003	ENST00000393588	ENSP00000377213	protein_coding
CASP3-004	ENST00000447121	ENSP00000407142	protein_coding
CASP3-201	ENST00000352650	ENSP00000296745	protein_coding
CASP3-202	ENST00000438467	ENSP00000390792	protein_coding

« Orthologues

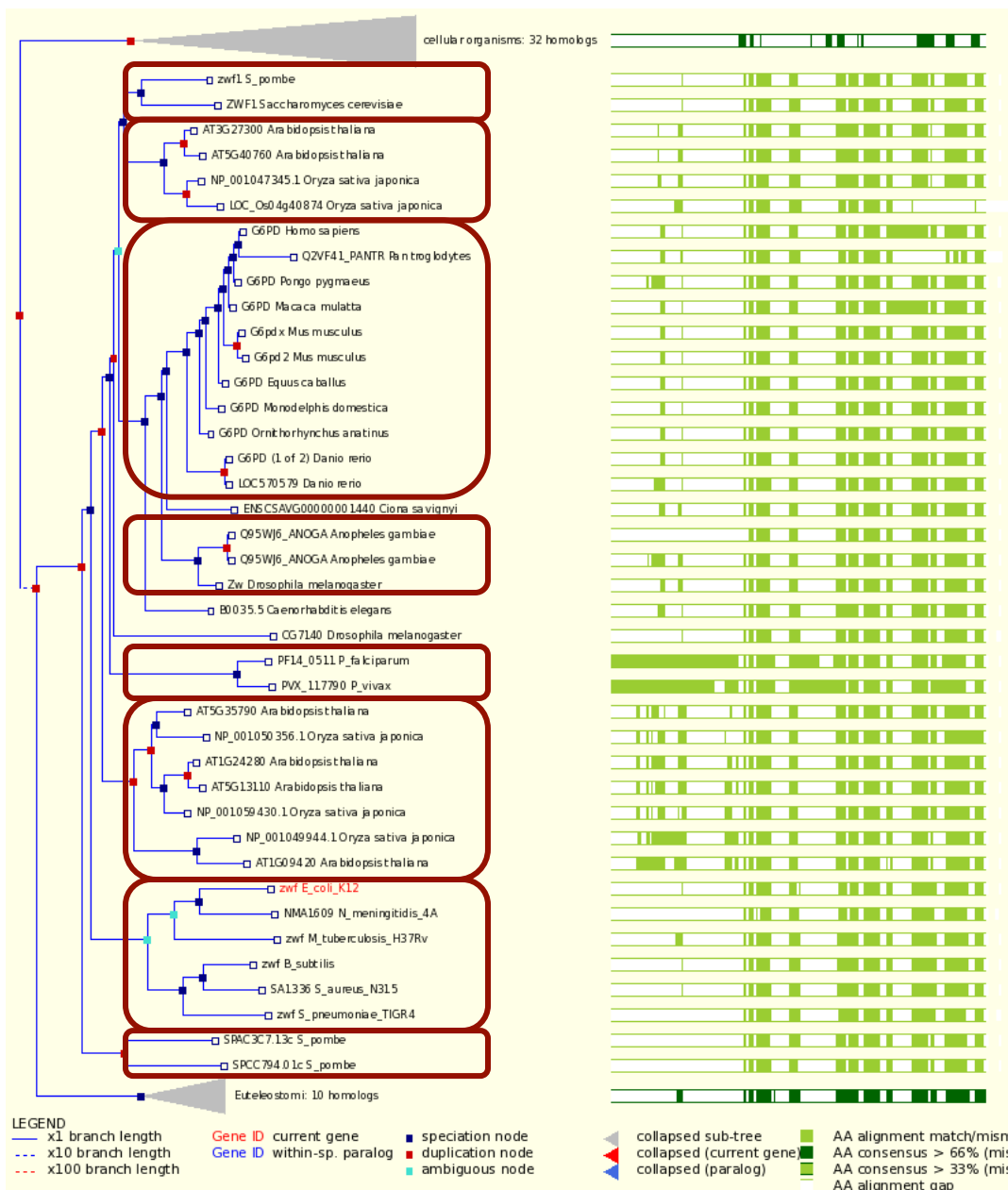
Paralogues [help](#)

Protein families »

The following gene(s) have been identified as putative paralogues (within species):

Taxonomy Level	Gene identifier
Euteleostomi	ENSG00000165806 (CASP7) [Multi-species comp.] [Align]
paralogue (within species)	Caspase-7 Precursor (CASP-7)(EC 3.4.22.60)(ICE-like apoptotic protease 3)(ICE-LAP3)(Apoptotic protease Mch-3)(CMH-1) [Contains Caspase-7 subunit p20;Caspase-7 subunit p11] [Source: UniProtKB/Swiss-Prot (P55210)] [Target %id: 42; Query %id: 51]
Chordata	ENSG00000138794 (CASP6) [Multi-species comp.] [Align]
paralogue (within species)	Caspase-6 Precursor (CASP-6)(EC 3.4.22.59)(Apoptotic protease Mch-2) [Contains Caspase-6 subunit p18;Caspase-6 subunit p11] [Source: UniProtKB/Swiss-Prot (P55212)] [Target %id: 35; Query %id: 37]
Bilateria	ENSG00000064012 (CASP8) [Multi-species comp.] [Align]
paralogue (within species)	Caspase-8 Precursor (CASP-8)(EC 3.4.22.61)(ICE-like apoptotic protease 5)(MORT1-associated CED-3 homolog)(MACH)(FADD-homologous ICE/CED-3-like protease)(FADD-like ICE)(FLICE)(Apoptotic cysteine protease)(Apoptotic protease Mch-5)(CAP4) [Contains Caspase-8 subunit p18;Caspase-8 subunit p10] [Source: UniProtKB/Swiss-Prot (Q14790)] [Target %id: 18; Query %id: 35]
Bilateria	ENSG00000003400 (CASP10) [Multi-species comp.] [Align]
paralogue (within species)	Caspase-10 Precursor (CASP-10)(EC 3.4.22.63)(ICE-like apoptotic protease 4)(Apoptotic protease Mch-4)(FAS-associated death domain protein interleukin-18-converting enzyme 2)(FLICE2) [Contains Caspase-10 subunit p23/17;Caspase-10 subunit p12] [Source: UniProtKB/Swiss-Prot (Q92851)] [Target %id: 16; Query %id: 30]
Undetermined	ENSG00000106144 (CASP2) [Multi-species comp.] [Align]
other paralogue (within species)	Caspase-2 Precursor (CASP-2)(EC 3.4.22.55)(ICH-1 protease)(Neural precursor cell expressed developmentally down-regulated protein 2)(NEDD-2) [Contains Caspase-2 subunit p18;Caspase-2 subunit p13;Caspase-2 subunit p12] [Source: UniProtKB/Swiss-Prot (P42575)] [Target %id: 15; Query %id: 25]
Undetermined	ENSG00000235505 (AP002004.1) [Multi-species comp.] [Align]
other paralogue (within species)	LOC643733 protein Fragment [Source: UniProtKB/TrEMBL (Q05BX3)] [Target %id: 17; Query %id: 5]
Undetermined	ENSG00000105141 (CASP14) [Multi-species comp.] [Align]
other paralogue (within species)	Caspase-14 Precursor (CASP-14)(EC 3.4.22.-) [Contains Caspase-14 subunit p19;Caspase-14 subunit p10] [Source: UniProtKB/Swiss-Prot (P31944)] [Target %id: 28; Query %id: 25]
Undetermined	ENSG00000228146 (AC108134.5) [Multi-species comp.] [Align]
other paralogue (within species)	No description [Target %id: 23; Query %id: 21]
Undetermined	ENSG00000137757 (CASP5) [Multi-species comp.] [Align]
other paralogue (within species)	Caspase-5 Precursor (CASP-5)(EC 3.4.22.58)(ICH-3 protease)(TY protease)(ICE(rel)-III) [Contains Caspase-5 subunit p20;Caspase-5 subunit p10] [Source: UniProtKB/Swiss-Prot (P51878)] [Target %id: 14; Query %id: 23]





Fungi

Plants

Vertebrates

Insects

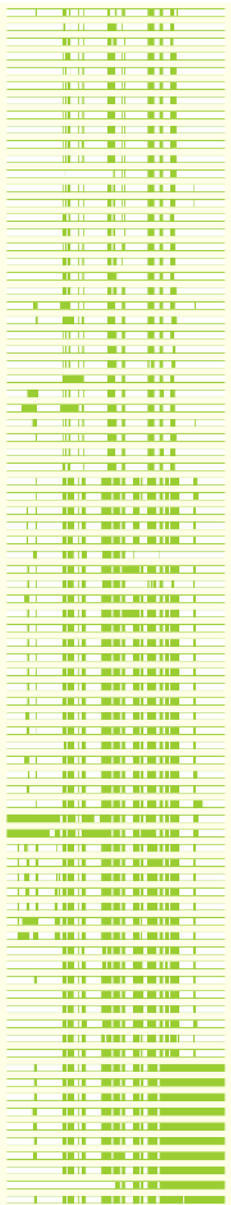
Protists

Plants

Bacteria

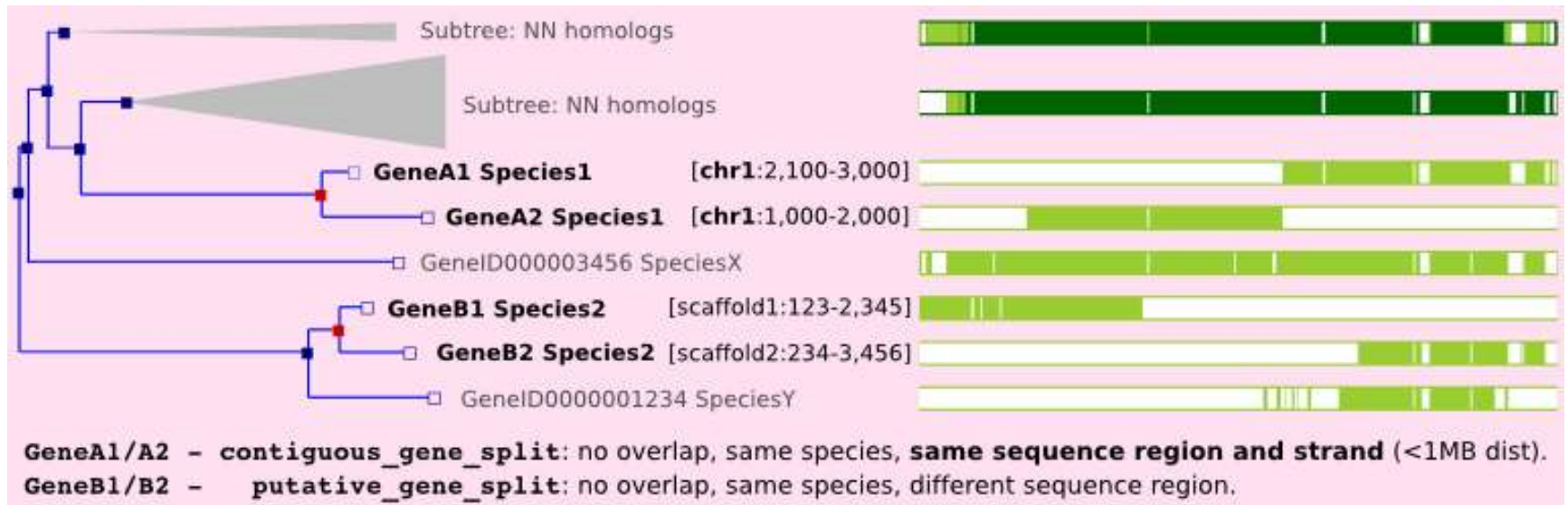
Fungi

Glucose-6-phosphate 1-dehydrogenase



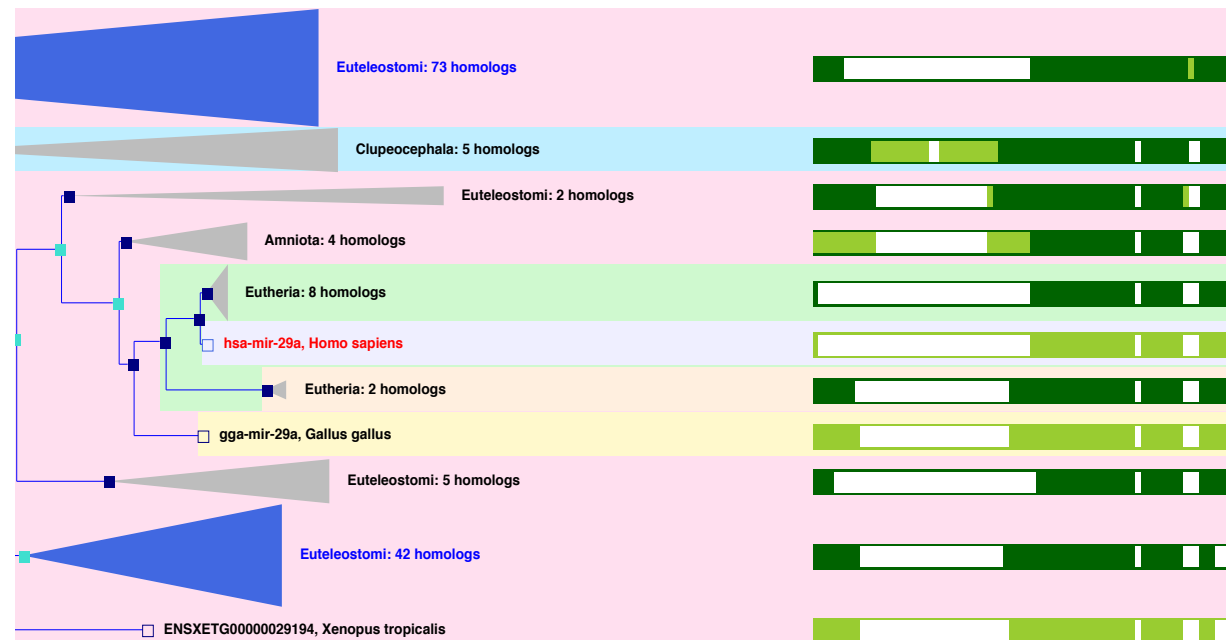
Paralogues: gene-split exception

- Split genes can happen because of problems in the assembly
- They will show in the same tree, but there won't be any overlap in the alignment
- We rename these cases as gene-split



Noncoding RNA trees

- Based on Rfam family IDs
- Alignment using both secondary structure information (Infernal) and genomic sequence





Homologues Exercise

1. Find the human MYL6 gene: go to its gene summary.
2. Find the MYL6 in the gene tree. Find a link to TreeFam. What is this resource useful for? Are there any other links to external comparative genomics resources?
3. How many paralogues does MYL6 have? Find them in the gene tree.
4. Which paralogue is closest to the human MYL6 gene? In what taxon is the common ancestor? Is the duplication score high or low?

- Rationale
- Comparative proteomics
 - Homologue predictions (Gene Trees)
 - **Protein family clustering**
- Comparative genomics
 - Genome-wide DNA alignments
 - Syntenic block characterisation

Protein Dataset

- About 1,500,000 proteins clustered:
 - All Ensembl protein predictions from all supported species
 - ~ 670,000 protein predictions
 - All metazoan (animal) proteins in UniProtKB:
 - ~ 80,000 UniProtKB/Swiss-Prot
 - ~ 830,000 UniProtKB/TrEMBL

Clustering Strategy

- BLASTP all versus all comparison
- Based on similarities Markov clustering (MCL)
- For each cluster:
 - Calculation of multiple sequence alignments (JalView)
 - Assignment of a consensus description (above 40%)
- Characterisation of protein domains
 - Internal QC data for checking consistency between orthologues, protein clusters and domain information
 - Cross-check data also available for our users

Ensembl Protein Families


[Home](#) > [Human](#) [GRCh37]

Location: [4:185,548,850-185,570,629](#) | **Gene: CASP3**

[Login / Register](#) | [BLAST/BLAT](#) | [BioMart](#) | [Docs & FAQs](#) | [Mirrors](#)

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
 - Genomic alignments (5)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (74)
 - Paralogues (12)
- Protein families (1)**
- 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: CASP3 (ENSG00000164305)
 Caspase-3 Precursor (CASP-3)([EC 3.4.22.56](#))(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot P42574](#)
Location [Chromosome 4: 185,548,850-185,570,629](#) reverse strand.
Transcripts There are 6 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
CASP3-001	ENST00000308394	ENSP00000311032	protein_coding
CASP3-002	ENST00000393585	ENSP00000377210	protein_coding
CASP3-003	ENST00000393588	ENSP00000377213	protein_coding
CASP3-004	ENST00000447121	ENSP00000407142	protein_coding
CASP3-201	ENST00000352650	ENSP00000296745	protein_coding
CASP3-202	ENST00000438467	ENSP00000390792	protein_coding

[« Paralogues](#)
[Protein families \[help\]\(#\)](#)
[Variation Table »](#)

Family ID	Consensus annotation	Other Human transcripts in this family	Multiple alignments
ENSM00500000269906 (2 genes) (all proteins in family)	CASPASE 3 PRECURSOR CASP 3 EC_3.4.22.56 [CONTAINS CASPASE 3 SUBUNIT P17; CASPASE 3 SUBUNIT P12]	• ENST00000352650 (CASP3-201) • ENST00000447121 (CASP3-004) • ENST00000438467 (CASP3-202) • ENST00000393588 (CASP3-003) • ENST00000393585 (CASP3-002) • ENST00000308394 (CASP3-001)	126 Ensembl members of this family JalView 215 members of this family JalView

 Ensembl release 56 - Sept 2009 © [WTSI](#) / [EBI](#)
[Permanent link](#) - [View in archive site](#)
[About Ensembl](#) | [Contact Us](#) | [Help](#)

Ensembl Protein Family-Genes

[Home](#) > [Human \(GRCh37\)](#)

Location: [4:185,548,850-185,570,629](#) **Gene: CASP3**

[Login / Register](#) | [BLAST/BLAT](#) | [BioMart](#) | [Docs & FAQs](#) | [Mirrors](#)

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
 - Genomic alignments (5)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (74)
 - Paralogues (12)
 - Protein families (1)
- 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: CASP3 (ENSG00000164305)

Caspase-3 Precursor (CASP-3)([EC 3.4.22.56](#))(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] [Source: UniProtKB/Swiss-Prot P42574](#)

Location [Chromosome 4: 185,548,850-185,570,629](#) reverse strand.

Transcripts There are 6 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
CASP3-001	ENST00000308394	ENSP00000311032	protein_coding
CASP3-002	ENST00000393585	ENSP00000377210	protein_coding
CASP3-003	ENST00000393588	ENSP00000377213	protein_coding
CASP3-004	ENST00000447121	ENSP00000407142	protein_coding
CASP3-201	ENST00000352650	ENSP00000296745	protein_coding
CASP3-202	ENST00000438467	ENSP00000390792	protein_coding

[Protein families](#)
HUMAN genes in this family [help](#)
[Variation Table](#)

Ensembl genes containing proteins in family ENSFM00500000269906

Gene ID and Location	Gene Name	Description(if known)
ENSG00000164305 Chromosome 4: 185,55m	CASP3	Caspase-3 Precursor (CASP-3)(EC 3.4.22.56)(Apopain)(Cysteine protease CPP32)(CPP-32)(Yama protein)(SREBP cleavage activity 1)(SCA-1) [Contains Caspase-3 subunit p17;Caspase-3 subunit p12] Source: UniProtKB/Swiss-Prot;Acc:P42574
ENSG00000165806 Chromosome 10: 115,44m	CASP7	Caspase-7 Precursor (CASP-7)(EC 3.4.22.60)(ICE-like apoptotic protease 3)(ICE-LAP3)(Apoptotic protease Mch-3)(CMH-1) [Contains Caspase-7 subunit p20;Caspase-7 subunit p11] Source: UniProtKB/Swiss-Prot;Acc:P55210

Ensembl release 56 - Sept 2009 © [WTSI](#) / [EBI](#)

[About Ensembl](#) | [Contact Us](#) | [Help](#)

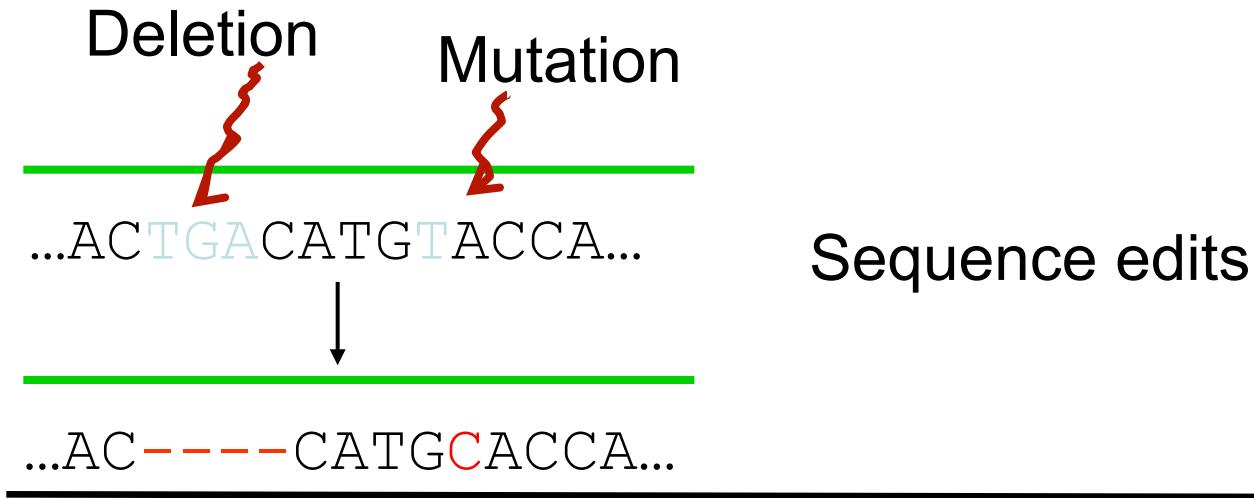
[Permanent link](#) - [View in archive site](#)

- Rationale
- Comparative proteomics
 - Orthologue predictions
 - Protein family clustering
- Comparative genomics
 - Genome-wide DNA alignments
 - Syntenic block characterisation

Whole Genome Alignments

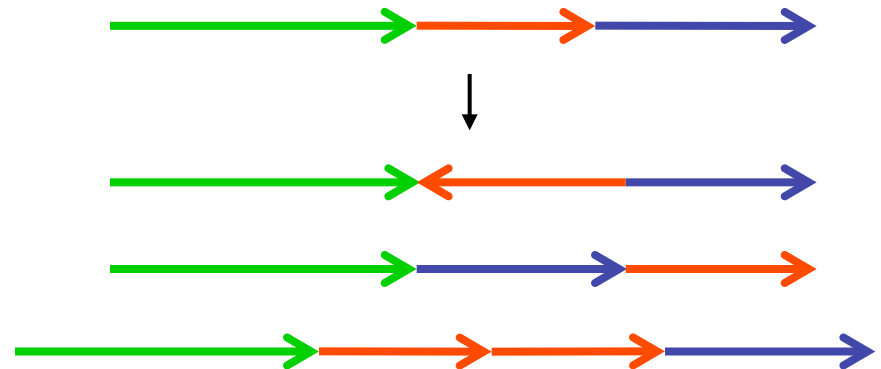
- Functional sequences evolve generally more slowly than non-functional sequences
- Therefore, sequences that remain conserved may exert a biological function
- Comparing DNA sequences from species at different evolutionary distances can help to identify biological function
 - Coding genes
 - Non-coding genes
 - Non-coding regulatory sequences

Evolution at the DNA level



Rearrangements

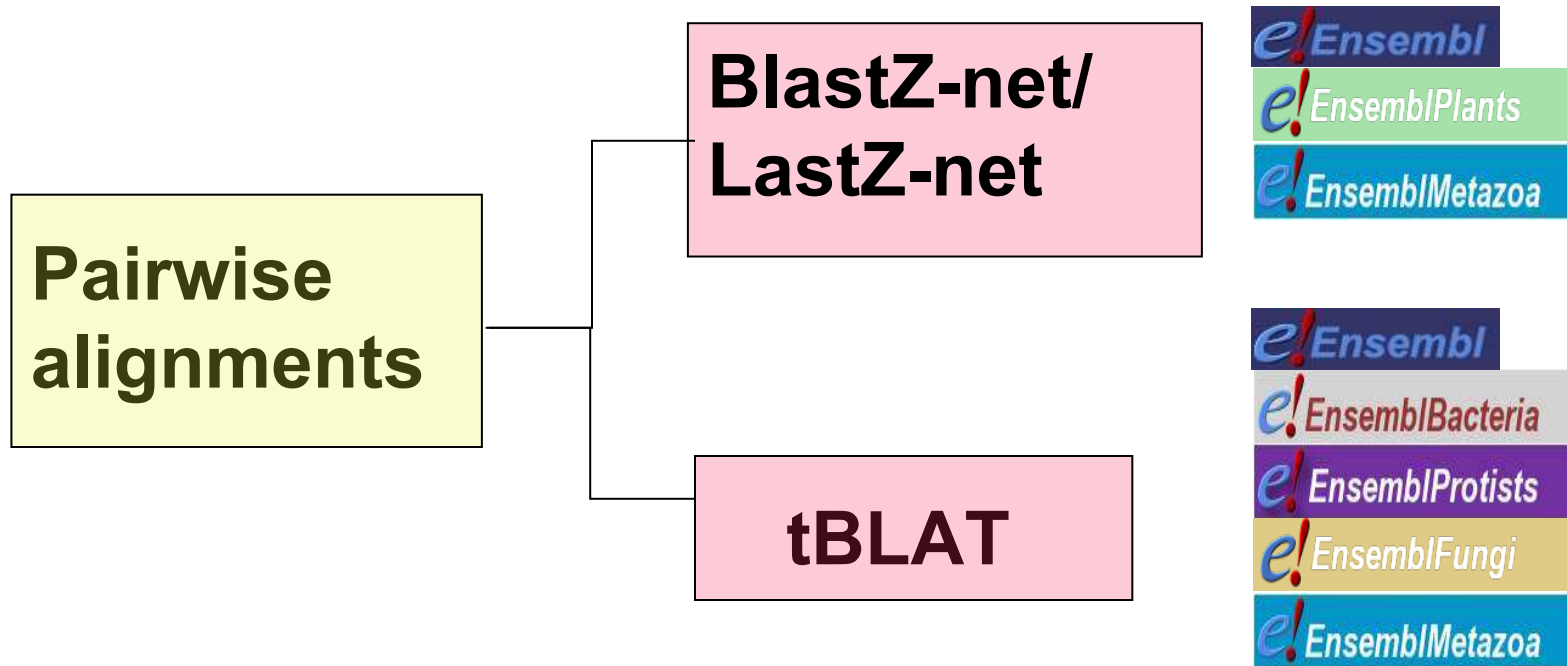
Inversion
Translocation
Duplication



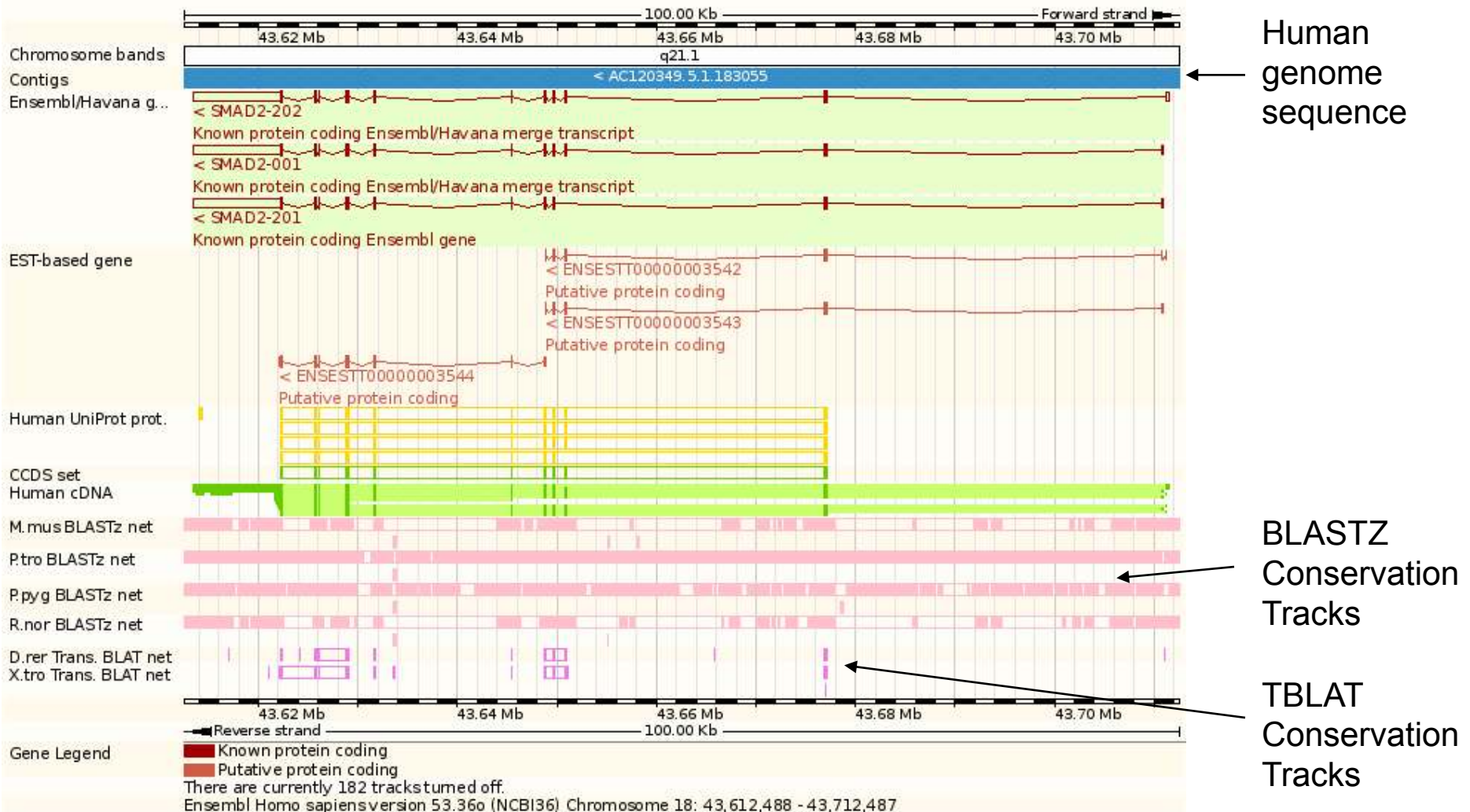
Pairwise Alignment Algorithm

- Should find all highly similar regions between two sequences
 - Orthologous, as well as paralogous
- Should cope with segments without similarity
- Should handle rearranged sequences correctly
- Need post-filtering to limit overlapping alignments

Pairwise whole genome alignments



Ensembl Location-View



Pairwise Alignment Export



Home > Human

Location: 18:43,631,947-43,632,461 Gene: SMAD2

Account | Logout | BLAST/BLAT | BioMart | Docs & FAQs

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Genomic alignments
 - Synteny (10)
- Genetic Variation
 - Resequencing (6)
 - Linkage Data
- Markers

- Configure this page
- Add custom data to page
- Export data
- Bookmark this page

Chromosome 18: 43,631,947-43,632,461



Export image

« Genomic alignments Synteny »

BLASTz net alignment between Homo sapiens chromosome 18 and Mus musculus chromosome 18

43631947	ATCTAATATTGACGGTTCCTTGGGTATCTTTCCAAAATACTCTATGAACACATAAGCAA	43632006
	A TA TG C TTCC G GTAT TCTAT AA ACAT AGC A	
76457901	GCTCACTACTGCCCTTTCCACGAGTAT-----TCTATAAAGTACATCAGCCA	76457842
43632007	ATACATTTTGTCTGTGTCCT-TTCTTTTCATATACAAGTGGTAACCGTTCCTTCATCTT	43632066
	ATA ATTT TGTT GTGC T CTT TCAT CAA GG A C T T TC T	
76457841	ATATATTCTGTTTAGTGTCATCCCTTCTCATGACCAACAGGAACACTTGAAGCTCAT	76457782
43632067	CTATAT----TATTCATCAGA-----TTGTTTCAATGTGTTTCCTTATTGTCTATC	43632126
	CT TAT TAT AT AGA TGT CA ATGT T TC TTG ATC	
76457781	CTGTATGAGGTATCTCATTAGACACCTTCTGTGCA-GATGTTTCTC-----TTGCTCATC	76457722
43632127	TACATATCAATTAGTCACATAACATCGCGATCCACAT-TGATGGATATTCAAGCTATTT	43632186
	TACATA C AT CA C A CC AC TG T G TA T A G ATT	
76457721	TACATACC-----TATGCCAACAGCAGCCTACTCATGGTCGCTAATAACGAGATTG	76457662
43632187	TCAATCTTTTGCTATAAGAACCAATGCTTTGAGATTGCAATAACTAATCTTGACCGTTTT	43632246
	T A CT TAAGAACCA A CT C TGA CG	
76457661	TTGA-----CTTTAAGAACCA-----AGCTGCGCTGATCGG---	76457602
43632247	GCACGCTTCCAATCACATCTGAAAGATACATTCTTCAAGAACTGCCAGGTCGAAAGGTT	43632306
	T CCA T TT T CAAG AA T CCAGTCA A GGT	
76457601	-----TGCCAGT-----TTTACAAGGAAGTACCAGGCTAGGGGTT	76457542
43632307	TGTGTATTGCAATTTTGATAGATTGGGTAATAATTATCATTATCAAGGTTGTACCACT	43632366
	GTG TT CA TTTGATAGAT GG AA TTATCAT CA AAGG TGTACC GT	
76457541	GGTGCCCTCCCAAGTTTGATAGATACTGGCTAAGTTATCATCCA-AAAGGCTGTACCGGT	76457482
43632367	TTACATTCCCACTCACACATAGTAAAG-----GGATGCTTGAGGTGTTATTGCTTCT	43632426
	ACATTCCCA T ACA CAT A G GGATG AGGT A TG	
76457481	ACACATTCCCAATTACACCATCAAGACGTTAACTTGGATGGCCAAGGTAGAACTGACCAC	76457422
43632427	AGATGAAAACATGGT---CACAAAGTACAGTTTAAATTTTGTAGTAGCTTACACATGT	43632483
	AGATG AAACATGGT CACAAG CAGTTT A T G T G TA A TGT	
76457421	AGATGCAACATGTTCTGCACAAG--CAGTTTCGAGCTGGGATGGTGGTATAGCTGT	76457365

Ensembl Location-Multi Species View

Staging
Home > Human (staging)
Locations: 7:116,992,721-117,384,165 [Genes: 138]
Login / Register | BLAST/BLAT | Docs & FAQs

Location-based displays:
 - Whole genome
 - Chromosome summary
 - Region overview
 - Region in detail
 - Comparative Genomics
 - Genomic sign size (kb)
 - Genomic sign size (Mb)
 - Multi-species comp.
 - Synteny (11)
 - Genes/Transcripts
 - Annotations (12)
 - Links/Refs
 - Maps
 - Other genomic browsers
 - NCBI
 - HGNC

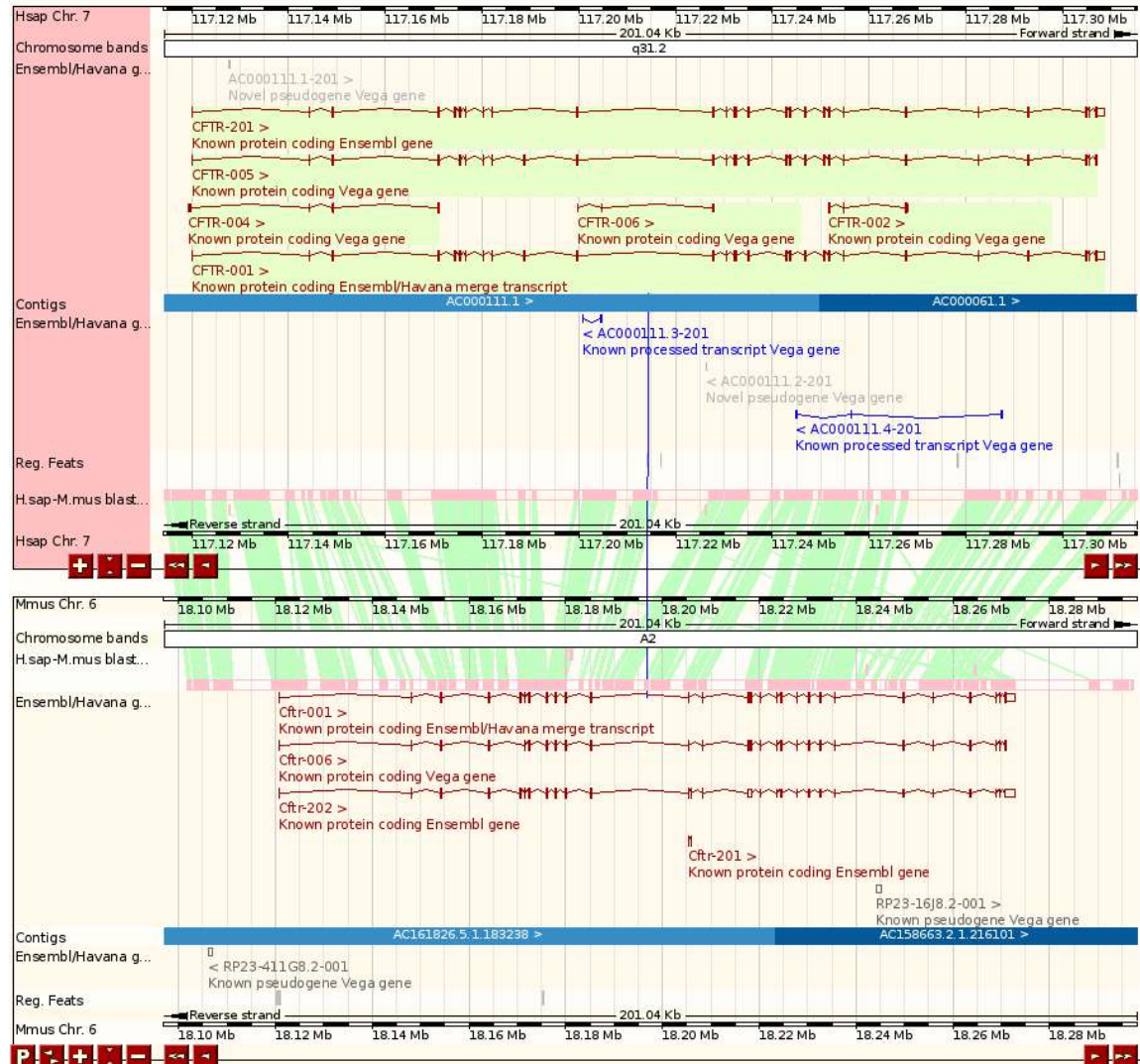
Chromosome 7: 116,992,721-117,384,165

Genomic alignments:
 - Add an alignment:
 - Remove an alignment:
 - Select a track:
 - Select a track to view:
 - Select a track to hide:
 - Select a track to zoom:
 - Select a track to zoom in:
 - Select a track to zoom out:
 - Select a track to zoom reset:
 - Select a track to zoom zoom in:
 - Select a track to zoom zoom out:
 - Select a track to zoom zoom reset:
 - Select a track to zoom zoom in:
 - Select a track to zoom zoom out:
 - Select a track to zoom zoom reset:

Locations: 7: 116,992,721-117,384,165 [Genes: 138]

Configure the display
 - Change the tracks you are displaying, use the "Configure this page" link on the left.
 - Request image

Ensembl release 88 - Top 2000 genes (7/14/14)
 Download Ensembl release 88



6-117,315,295 **Genes: CPTN** Login / Register | BLAST/BLAT | Docs & FAQs

Chromosome 7: 117,114,256-117,315,295

Genomic align slice [help](#) [Export image](#)

Region in detail [Genomic alignments](#)

Chromosome bands
Contigs
Ensembl-Havana g...

Genomic align slice [help](#) [Export image](#)

Alignment: Mouse (Mus musculus) - Ensembl [Go](#) [Go](#)

Location: 7 : 117114251 - 117131000 [Go](#)

Homo sapiens
Vega-Havana gene

Ensembl-Havana g...

Contigs
Ensembl-Havana g...

Vega-Havana gene

Gene legend

Mus musculus
Vega-Havana gene

Ensembl-Havana g...

Contigs
Vega-Havana gene

Gene legend

Ensembl release 50 - Sept 2009 (Genes: CPTN) [About Ensembl](#) | [Contact Us](#) | [Help](#)

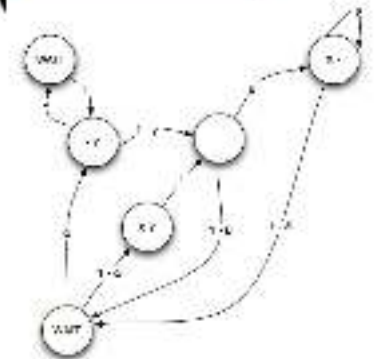
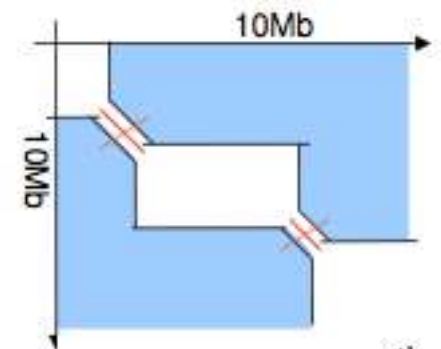
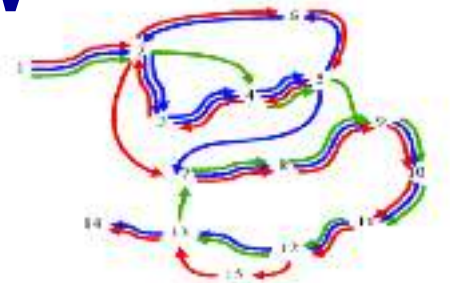


Multiple Alignment Sets

- Ensembl release 60:
 - 34 eutherian (placental) mammals (low coverage) E-P-O
 - Conservation scores
 - Constrained elements
 - 16 amniota vertebrates Pecan
 - Conservation scores
 - Constrained elements
 - 12 eutherian mammals E-P-O
 - 6 primates E-P-O
 - 5 teleost fish

EPO pipeline overview

- Enredo ¹
 - Defines blocks of colinear sequence
 - Supports segmental duplications
- Pecan ¹
 - Consistency based multiple aligner
 - Optimized to cope with long genomic sequences
- Ortheus ²
 - Ancestral sequences reconstructor (Tree Aligner)
 - Infers the history of insertion and deletions

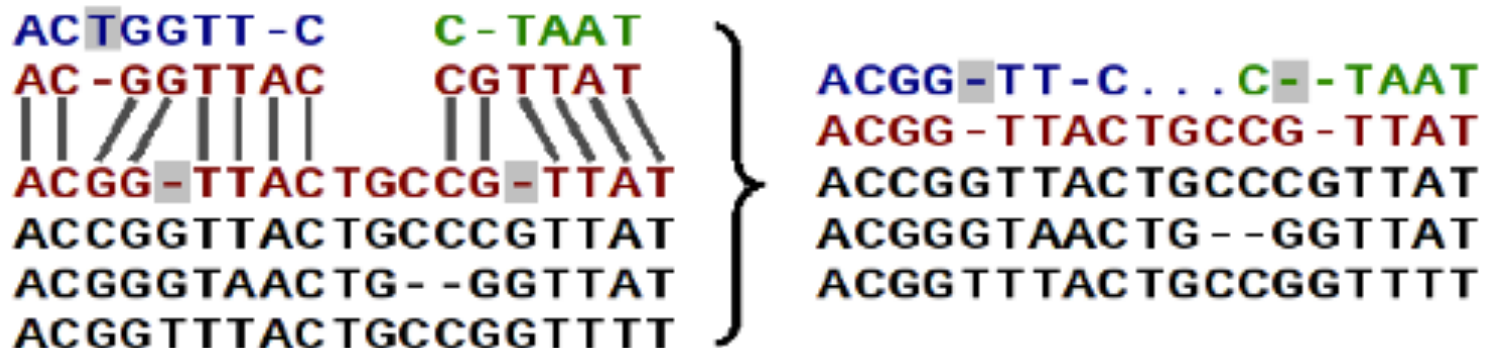


(1) Paten, Herrero et al. *Genome Res.* 2008. 18 (11):1814-28

(2) Paten et al. *Genome Res.* 2008. 18(11): 1829-43

Adding low-coverage (2x genomes)

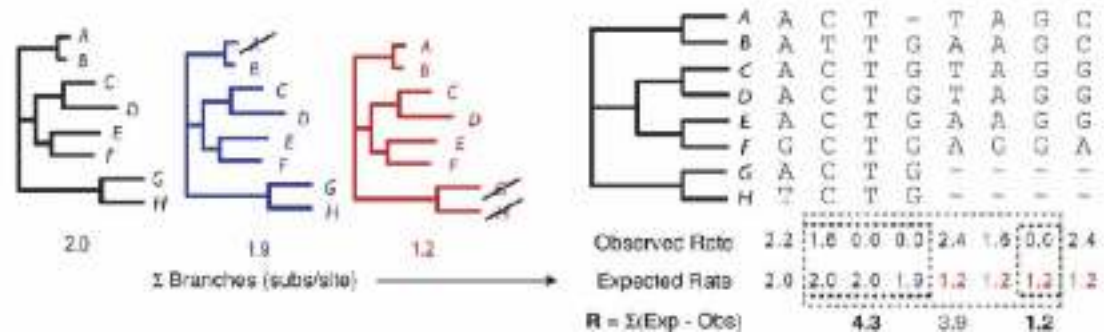
- Low coverage genomes cannot be fully assembled
- Resulting assembly is too scattered to be used with Enredo
- Run EPO on high-coverage genomes only
- Map 2x genomes using pairwise alignments



Genomic evolutionary Rate Profiling (Gerp ¹)

– constrained elements identification

- Stretches of the alignment with high conservation



- Conserved elements and coding exons
 - Majority of coding exons are associated with constrained elements
 - Only about a quarter of constrained elements are associated with coding exons
 - Other features (transcription start sites, regulatory features)

(1) Cooper et al. *Genome Research*, 2005

Ensembl Location-View

Constrained elements

Conservation Scores

Multiple Sequence Alignment



Ensembl Location-Genomic Alignments

e!Staging
Home > Human [GRCh37]

Login / Register | BLAST/BLAT | Docs & FAQs

Location: 7:117,174,227-117,175,648 Gene: CFTR

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Genomic align slice (49)
 - Genomic alignments**
 - Multi-species comp. (4)
 - Synteny (11)
- Genetic Variation
 - Resequencing (2)
 - Linkage Data
- Markers
- Other genome browsers
 - UCSC
 - NCBI

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

Chromosome 7: 117,174,227-117,175,648

chromosome 7

Export Image

« Genomic align slice Genomic alignments help Multi-species comparison »

Alignment: 4 catarrhini primates EPO Go>

Location: 7 : 117174227 - 117175648 Go>

THIS STYLE: Location of conserved regions (where >50% of bases in alignments match)
 THIS STYLE: Location of Ensembl exons
 THIS STYLE: Location of SNPs
 THIS STYLE: Location of inserts

Homo_sapiens > [chromosome:GRCh37:7:117174227:117175648:1](#)
 Pan_troglodytes > [chromosome:CHIMP2.1:7:117746362:117747781:1](#)
 Pongo_pygmaeus > [chromosome:PPYG2:7:114087241:114088661:1](#)
 Macaca_mulatta > [chromosome:MMUL_1:3:154864010:154865429:1](#)

Homo_sapiens AATAATGAATGCATAAATACTGAATT----AGTCATATTATAATTTACTTATAATATATTGTATTTTGTGTTGAAATTATCTAACT
 Pan_troglodytes AATAATGAGTGCATAAATACTGAATTACTCAGTCATATTATAATTTACTTATAATATATTGTATTTTGTGTTGAAATTATCTAACT
 Pongo_pygmaeus AATAATGAATGCATAAATACTGAATTACTCAGTCATATTATAATTTACTTATAATATATTGTATTTTGTGTTGAAATTATCTAACT
 Macaca_mulatta AATAATGAATGCATAAATACTGAATT----AGTCATATTATAATTTACTTATAATATATTGTATTTTGTGTTGAAATTATCTAACT

Homo_sapiens TTYCATTTTCTTTTAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAAATAAGTATTGGACAACCTGTTTGTCTCCTTTCCAACAACC
 Pan_troglodytes TTCCATTTTCTTTTAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAAATAAGTATTGGACAACCTGTTAGTCTCCTTTCCAACAACC
 Pongo_pygmaeus TCCCATTTTCTTTTAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAAATAAGTATTGGACAACCTGTTAGTCTCCTTTCCAACAACC
 Macaca_mulatta TCCCATTTTCTTTTAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAAATAAGTATTGGACAACCTGTTAGTCTCCTTTCCAACAACC

Homo_sapiens TGAACAAATTTGATGAAGTATGTACCTATTGATTTAATCTTTTAGGCACCTATTGTTATAAATTATACAACCTGGAAAGGCGGAGTTTTCCT
 Pan_troglodytes TGAACAAATTTGATGAAGTATGTACCTATTGATTTAATCTTTTAGGCACCTATTGTTATAAATTATACAACCTGGAAAGGCGGAGTTTTCCT
 Pongo_pygmaeus TGAACAAATTTGATGAAGTATGTACCTATTGATTTAATCTTTTAGGCACCTATTGTTATAAATTATACAACCTGGAAAGGCGGAGTTTTCCT
 Macaca_mulatta TGAACAAATTTGATGAAGTATGTACCTATTGATTTAATCTTTTAGGCACCTATTGTTATAAATTATACAACCTGGAAAGGCGGAGTTTTCCT

Homo_sapiens GGGTCAGATAATAGTAATTAGTGGTTAAGTCTTGCTCAGYCTAGCTTCCCTATTCTGGAACCTAAGAAAGGTCAATTGTATAGCAGAGC
 Pan_troglodytes GGGTCAGATAATAGTAATTAGTGGTTAAGTCTTGCTCAGCTCTAGCTTCCCTATTCTGGAACCTAAGAAAGGTCAATTGTATAGCAGAGC
 Pongo_pygmaeus GGGTCAGATAATAGTAATTAGTGGTTAAGTCTTGCTCAGCTCTAGCTTCCCTATTCTGGAACCTAAGAAAGGTCAATTGTATAGCAGAGC
 Macaca_mulatta GTATCAGAT-----AAGTGGTTAAGTCTTGCTCAGCTCTAGCTTCCCTATTCTGGAACCTGAGAAAGGTCAATTGTATAGCAGAAC



Alignments Exercise

Go to the Location tab, Region in Detail view for the human *BRCA2* gene. Customise the view to display pairwise alignments for chimpanzee, mouse, and the clawed frog.

1. What genomic region on chimpanzee is aligned? Is the alignment different for the clawed frog?
2. “View alignment” for one of the mouse exons.

Go to Location tab, Multi-species view for the human *APC* gene. Add chimpanzee and zoom on the first exon.

1. Are there any substitutions in the alignment?

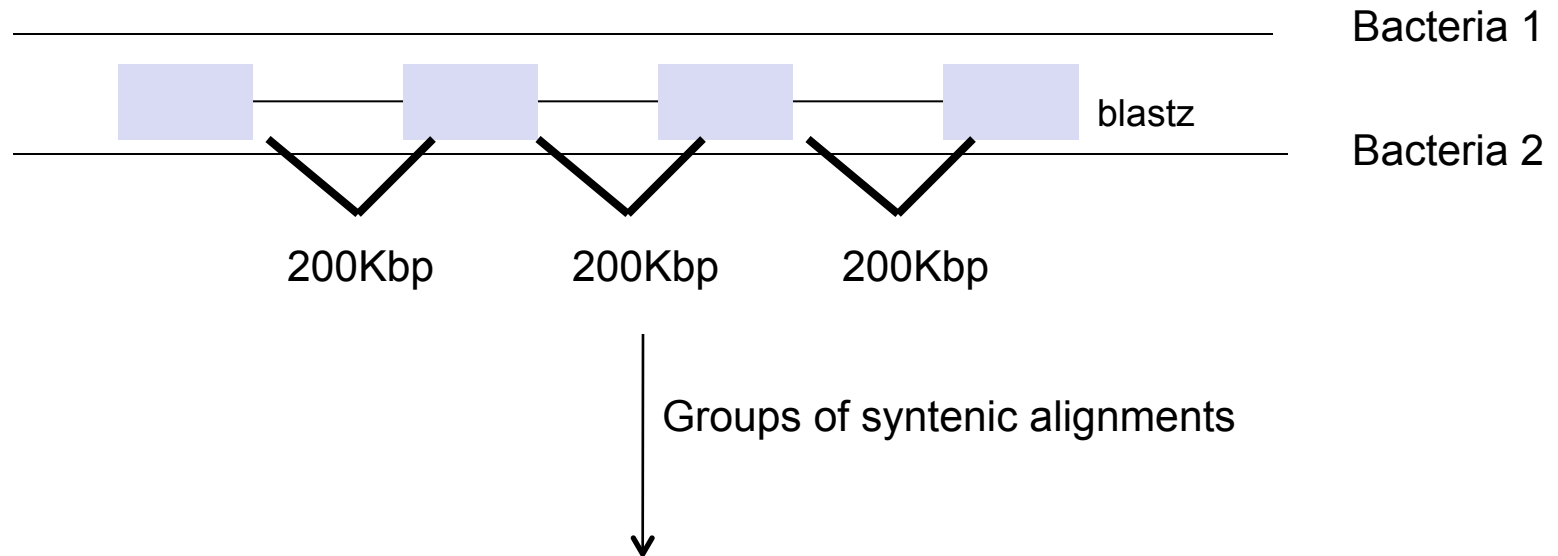
- Rationale
- Comparative proteomics
 - Orthologue predictions
 - Protein family clustering
- Comparative genomics
 - Genome-wide DNA alignments
 - Syntenic block characterisation

Syntenic Regions

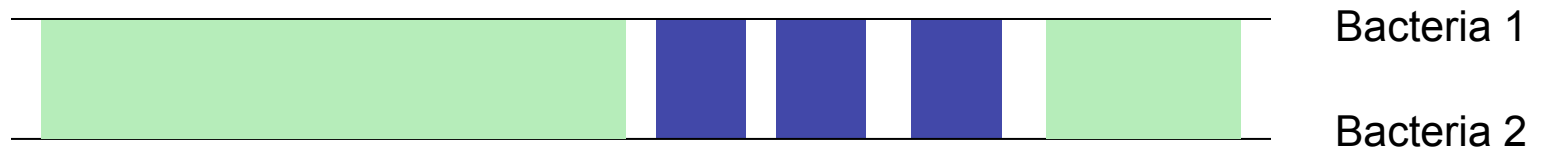
- Synteny
 - Regions of conserved sequence and gene order on chromosomes
 - result of descent from a common ancestor
- Genome alignments are refined into larger syntenic regions when the relative distance is $<3\text{Mbp}$ and order and orientation are consistent
- Based on BlasZnet/LastZ-net

Identifying syntenic regions

Step 1: Identification & grouping of syntenic alignments

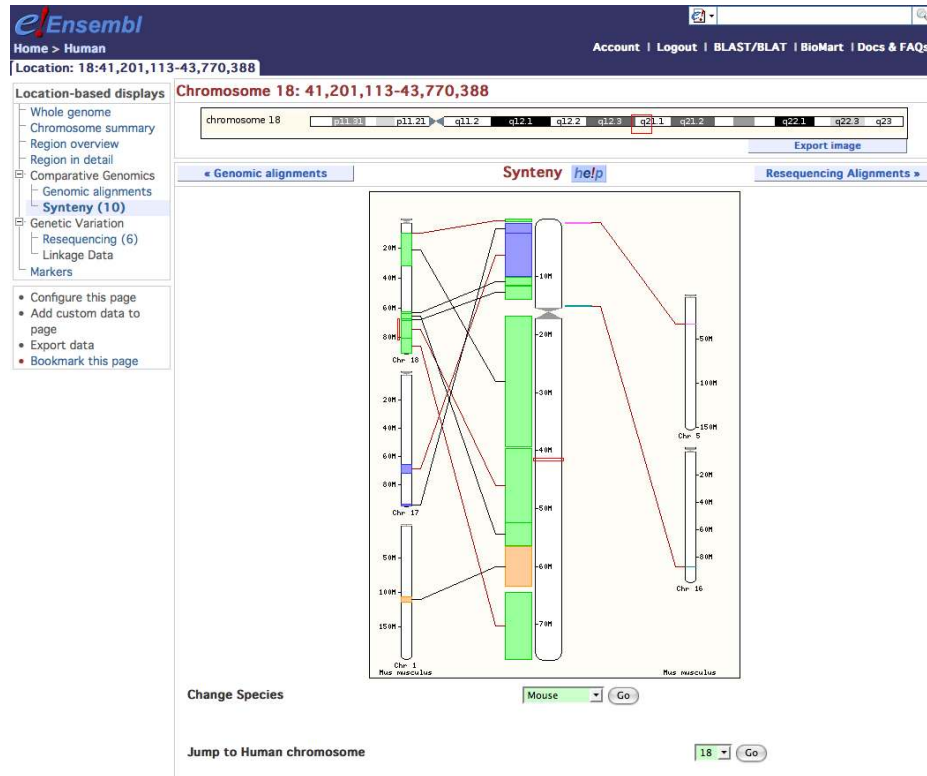


Step 2: Linking of groups to form syntenic blocks



- 3Mbp
- And no more than two non-syntenic groups between them

Ensembl Location-Synteny



« 15 upstream genes		Navigate homology	15 downstream genes »	
Homology Matches				
Homo sapiens Genes	Location		Mus musculus Homologues	Location
SLC14A2 (ENSG00000132874)	18:41448780-41517070	->	Slc14a2 (ENSMUSG00000024552)	18:78342889-78523153
SLC14A1 (ENSG00000141469)	18:41558147-41585046	->	Slc14a1 (ENSMUSG00000059336)	18:78296833-78320061
SIGLEC15 (ENSG00000197046)	18:41659543-41676519		No homologues	
KIAA1632 (ENSG00000152223)	18:41681573-41801303	->	S430411K18Rik (ENSMUSG00000039840)	18:78172736-78231750
PSTPIP2 (ENSG00000152229)	18:41817500-41906224	->	Pstpip2 (ENSMUSG00000025429)	18:78033042-78121618
ATP5A1 (ENSG00000152234)	18:41918108-41938197	->	Atp5a1 (ENSMUSG00000025428)	18:78012695-78021605
CCDC5 (ENSG00000152240)	18:41938323-41962296	->	Ccdc5 (ENSMUSG000000041840)	18:77996314-78006499
C18orf25 (ENSG00000152242)	18:42007986-42100952	->	8030462N17Rik (ENSMUSG000000047466)	18:77872020-77953222
RNF165 (ENSG00000141622)	18:42168185-42294781	->	Rnf165 (ENSMUSG00000025427)	18:77700452-77707018
LOXHD1 (ENSG00000167210)	18:42310933-42435624	->	Loxhd1 (ENSMUSG00000032818)	18:77546489-77649376
ST8SIA5 (ENSG00000101638)	18:42513079-42591037	->	St8sia5 (ENSMUSG000000025425)	18:77424592-77494057
PIAS2 (ENSG00000078043)	18:42646058-42751464	->	Pias2 (ENSMUSG00000025423)	18:77303947-77392539
KATNAL2 (ENSG00000167216)	18:42780785-42881656	->	Katnal2 (ENSMUSG00000025420)	18:77231939-77286035
TCEB3CL2 (ENSG00000221927)	18:42802656-42804296		No homologues	
TCEB3C (ENSG00000183791)	18:42808571-42810447		No homologues	
TCEB3B (ENSG00000206181)	18:42812941-42815990		No homologues	
HDHD2 (ENSG00000167220)	18:42887780-42930869	->	Hdhhd2 (ENSMUSG00000025421)	18:77168835-77210910
IER3IP1 (ENSG00000134049)	18:42935413-42956743		No homologues	
AC051635.7 (ENSG00000215474)	18:43011492-43029552		No homologues	
SMAD2 (ENSG00000175387)	18:43613464-43711510	->	Smad2 (ENSMUSG00000024563)	18:76401563-76464962

Q&A

EXERCISES