

Access to genes and genomes with Ensembl



Introduction and Worked Example

January 2007

CONTENTS

INTRODUCTION.....	2
WORKED EXAMPLE	7
BROWSING ENSEMBL	Error! Bookmark not defined.
Exercises	26
Answers.....	26
BIOMART	30
Exercises	30
Answers.....	31
COMPARATIVE GENOMICS.....	34
Exercises	34
Answers.....	37
EVALUATING GENES AND TRANSCRIPTS	38
Exercises	38
Answers.....	39
VARIATIONS.....	41
Exercises	41
Answers.....	41

Introduction

Ensembl is one of the world's primary resources for genomic research, a resource through which scientists can access the human genome as well as the genomes of other model organisms. Because of the complexity of the genome and the many different ways in which scientists want to use it, Ensembl has to provide many levels of access with a high degree of flexibility. Through the Ensembl website a wet-lab researcher with a simple web browser can for example perform BLAST searches against chromosomal DNA, download a genomic sequence or search for all members of a given protein family. But Ensembl is also an all-round software and database system that can be installed locally to serve the needs of a genomic centre or a bioinformatics division in a pharmaceutical company enabling complex data mining of the genome or large-scale sequence annotation.

The need for automatic annotation

Recent years have seen the release of huge amounts of sequence data from genome sequencing centres (figure 1). However, this raw sequence data is most valuable to the

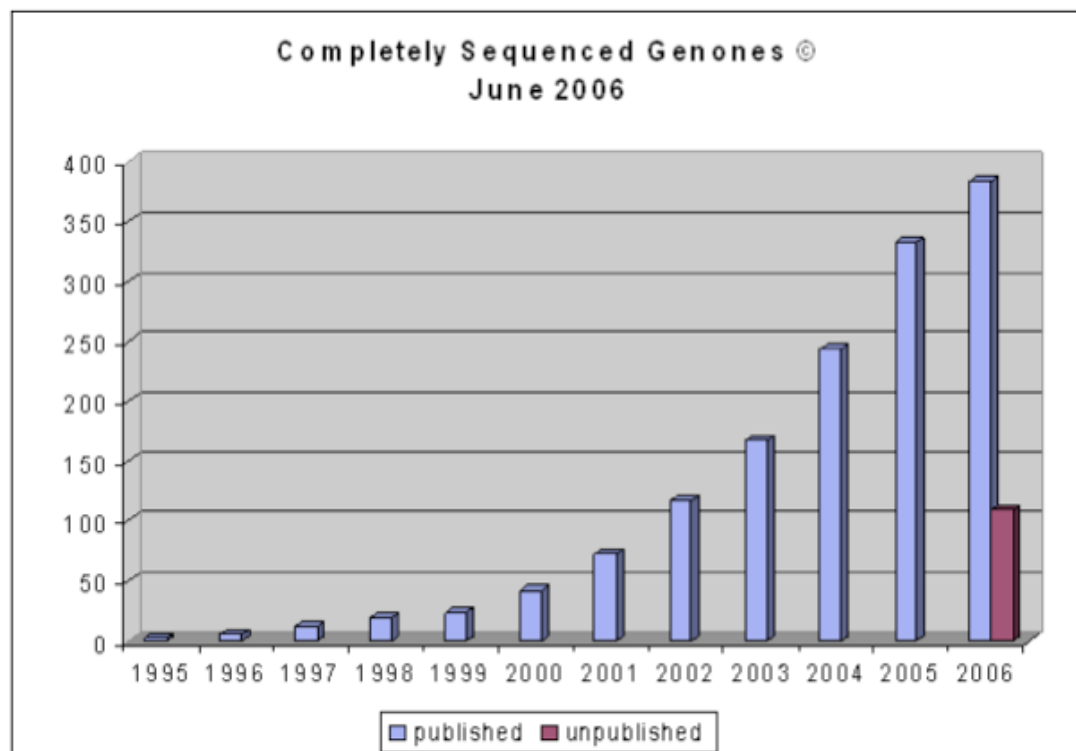


Figure 1. Completely sequenced genomes as of June 2006 (figure taken from <http://www.genomesonline.org>).

laboratory biologist when provided along with quality annotation of the genomic sequence. This information can be the starting point for planning experiments, interpreting Single Nucleotide Polymorphisms, inferring the function of gene products, predicting regulatory sites for gene expression and so on. The currently agreed 'gold standard' for the annotation of eukaryotic genomes is annotation made by a human being. This so-called "manual annotation" is based on information derived from sequence homology searches, the results of various *ab initio* gene prediction methods and literature searches. Annotation of large genomes (such as mouse and human) that meet this standard is slow and labour intensive, taking large teams of annotators years to complete. As a result, the annotation can almost never be entirely up-to-date and free of inconsistencies (as the annotation process usually begins before the sequencing process is complete). Hence, an automated annotation system is desirable since it is a relatively rapid process that allows frequent updates to accommodate new data. To meet this need, we produced the Ensembl annotation system by observing how annotators build gene structures and condensing this process into a set of rules.

The start of Ensembl

Ensembl's genesis was in response to the acceleration of the public effort to sequence the human genome in 1999. At that point it was clear that if annotation of the draft sequence was to be available in a timely fashion it would have to be automatically generated and that new software systems would be needed to handle genome data sets that were much larger, much more fragmented and much more rapidly changing than anything previous dealt with.

Ensembl was conceived in three parts: as a scalable way of storing and retrieving genomic data; as a web site for genome display; and as an automatic annotation method based around a set of heuristics. It was initially written for the draft human genome, which was sequenced clone-by-clone but has also been successfully used for whole genome shotgun assemblies. The storage and display parts of Ensembl are used for all the genomes currently present in Ensembl, while the automatic gene annotation has been run for most of the genomes with the exception of Takifugu, Tetraodon, Fruitfly, *C. elegans* and Yeast.

Over the past few years Ensembl has grown into a large scale enterprise, with substantial computing resources enabling it to process and provide live database access to currently more than 25 different genomes (figure 2) and a bimonthly update frequency to its website. It has a large community of users in both industry and academia, using it as a base for their individual organisation's experimental and computational genome based investigations, some of which maintain their own local installations.

Ensembl is a collaboration between the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute, both located on the Wellcome Trust Genome Campus in Hinxton, Cambridge, UK. Ensembl is funded

principally by the Wellcome Trust, with additional funding from the European Molecular Biology Laboratory (EMBL), the National Institutes of Health – National Institute of Allergy and Infectious Disease (NIH-NIAID) and the Biotechnology and Biological Sciences Research Council (BBSRC).

The Ensembl software and database system

As a software/database system Ensembl can be best described as a hybrid of a scripting programming language (Perl) and a relational database (MySQL, pronounced “My Ess Que Ell”).

Ensembl Perl software inherits from a tradition of biological object-design developed through BioPerl (<http://www.bioperl.org/>). This means that developers at Ensembl aimed at creating reusable pieces of software that would faithfully describe biological entities such as gene, transcript, protein, genomic clone or chromosome. Rules of usage and design of Ensembl and BioPerl objects can be best learned while using them, browsing their code and through a bit of trial-and-error. There is a comprehensive BioPerl tutorial available at the BioPerl website.

The Ensembl database is based on a relational database called MySQL. SQL in MySQL stands for ‘Structured Query Language’, a universal database programming language shared by many relational databases. Because MySQL is available free of charge for non-commercial developers, every academic centre can install its own local copy of MySQL (MySQL server) and download Ensembl data from the Ensembl ftp site. Simple queries of the database can be handled using the SQL language (see appendix), but for complex queries demanded by most biological analyses the Ensembl MySQL server is best accessed using Ensembl Perl objects.

The Ensembl annotation pipeline

The Ensembl analysis and annotation pipeline is based on a rule set of heuristics that a human annotator would use. All Ensembl gene predictions are based on experimental evidence, which is imported via manually curated UniProt/Swiss-Prot, partially manually curated NCBI RefSeq and automatically annotated UniProt/TrEMBL records. Untranslated regions (UTRs) are annotated to the extent supported by EMBL mRNA records. As there is no guarantee that UTR sequences in EMBL records are complete there is similarly no guarantee that the Ensembl genome analysis and annotation pipeline has enough biological evidence to predict complete UTR regions. For a limited number of species regulatory regions are annotated, but this annotation isn’t very extensive yet as the set of well-characterised promoters is still small and there is currently no algorithm yielding reliable results on a genomic scale.

The Ensembl website

Ensembl provides easy access to genomic information with a number of visualisation tools. The Ensembl website gives you for example the possibility to directly download data, whether it is a DNA sequence of a genomic contig

you are trying to identify novel genes in, or positions of SNPs in a gene you are working on. The key Ensembl web pages are called Views (e.g. GeneView, ContigView and SNPView), and will all be introduced appropriately later on. An updated version of the website is released bimonthly. Old versions are for at least two years accessible on the 'Archive!' website. Apart from that the 'Pre!' website provides displays of genomes that are still in the process of being annotated. There is also an ftp site to download large amounts of data from the Ensembl database, as well as the data-mining tool BioMart, that allows rapid retrieval of information from the databases. Finally, Ensembl BLAST offers the possibility to perform sequence searches against genomes and Ensembl gene and peptide sets.

Further reading

Hubbard, T.J.P. *et al.*

Ensembl 2007

Nucleic Acids Res. 2007 (*Database Issue*)

Birney, E. *et al.*

Ensembl 2006.

Nucleic Acids Res. 2006 Jan 34:D556-D561 (2006)

Hubbard, T. *et al.*

Ensembl 2005.

Nucleic Acids Res. 2005 33 D447-D453 (2005)

Birney, E. *et al.* *

An Overview of Ensembl.

Genome Research 14(5): 925-928 (2004)

Kasprzyk, A. *et al.*

EnsMart: a generic system for fast and flexible access to biological data.

Genome Research (2004) 14:1, 160-9.

Ashurst, J. L. *et al.*

The Vertebrate Genome Annotation (Vega) database.

Nucl. Acids Res. 33:D459-D465 (2005)

* This paper was part of the may 2004 issue of Genome Research which included an Ensembl special covering detailed aspects of the Ensembl web site, the underlying scalable database system for storing genome sequence and annotation information, as well as the automated genome analysis and annotation pipeline.

SPECIES		ASSEMBLY		GENEBUILD	
Mammals					
Human	<i>Homo sapiens</i>	NCBI 36	oct 2005	Ensembl	jul 2006
Chimpanzee	<i>Pan troglodytes</i>	PanTro 2.1	mar 2006	Ensembl	mar 2005
Rhesus macaque	<i>Macaca mulatta</i>	MMUL 1	feb 2006	Ensembl	aug 2006
Bushbaby*	<i>Otolemur garnettii</i>	BUSHBABY1			
Mouse	<i>Mus musculus</i>	NCBI m36	dec 2005	Ensembl	apr 2006
Rat	<i>Rattus norvegicus</i>	RGSC 3.4	dec 2004	Ensembl	feb 2006
Rabbit	<i>Oryctolagus cuniculus</i>	RABBIT	may 2005	Ensembl	aug 2006
Dog	<i>Canis familiaris</i>	CanFam 1.0	jul 2004	Ensembl	nov 2004
Cat*	<i>Felis catus</i>	CAT			
Cow	<i>Bos taurus</i>	Btau 2.0	mar 2005	Ensembl	dec 2005
Pig**	<i>Sus scrofa</i>				
Shrew*	<i>Sorex araneus</i>	sorAra1			
Hedgehog*	<i>Erinaceus europaeus</i>	eriEur1			
Microbat*	<i>Myotis lucifugus</i>	MICROBAT1			
Armadillo	<i>Dasypus novemcinctus</i>	ARMA	may 2005	Ensembl	aug 2006
Elephant	<i>Loxodonta africana</i>	BROAD E1	may 2005	Ensembl	aug 2006
Lesser hedgehog tenrec	<i>Echinops telfairi</i>	TENREC	may 2005	Ensembl	aug 2006
Opossum	<i>Monodelphis domestica</i>	MonDom 4.0	jan 2006	Ensembl	feb 2006
Platypus*	<i>Ornithorhynchus anatinus</i>	OANA 5			
Other species					
Chicken	<i>Gallus gallus</i>	WASHUC 1	mar 2004	Ensembl	dec 2005
<i>X. tropicalis</i>	<i>Xenopus tropicalis</i>	JGI 4.1	aug 2005	Ensembl	nov 2005
Zebrafish	<i>Danio rerio</i>	Zv 6	mar 2006	Ensembl	aug 2006
Fugu	<i>Takifugu rubripes</i>	FUGU 4.0	jun 2005	IMCB/JGI	may 2005
Tetraodon	<i>Tetraodon nigroviridis</i>	TETRAODON 7	apr 2003	Genoscope	sep 2004
Stickleback	<i>Gasterosteus aculeatus</i>	BROAD S1	feb 2006	Ensembl	aug 2006
Medaka	<i>Oryzias latipes</i>	HdrR 1	oct 2005	Ensembl	may 2006
<i>C. intestinalis</i>	<i>Ciona intestinalis</i>	JG 12	mar 2005	Ensembl	feb 2006
<i>C. savignyi</i>	<i>Ciona savignyi</i>	CSAV 2.0	oct 2005	Ensembl	apr 2006
Fruitfly	<i>Drosophila melanogaster</i>	BDGP 4	jul 2005	FlyBase	mar 2006
Anopheles	<i>Anopheles gambiae</i>	AgamP 3	feb 2006	VectorBase	oct 2005
Aedes	<i>Aedes aegypti</i>	AaegL 1	aug 2005	VectorBase	jun 2006
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>	WS 150	nov 2005	WormBase	nov 2005
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>	SGD 1	nov 2005	SGD	nov 2005

Figure 2 – Species in Ensembl, including name and date of their genome assembly and source and date of the genebuild. * = currently only available on the Pre! website, ** = only clone information available.

WORKED EXAMPLE – A walk through the main pages of the Ensembl browser, using the EPO (Erythropoietin precursor) gene as an example.

STEP 1:
Load Ensembl
www.ensembl.org

The screenshot shows the Ensembl genome browser homepage. Several green callout boxes highlight key features:

- Navigation column:** Points to the left sidebar containing links like 'Your Ensembl', 'Healthchecks', 'Help & Documentation', 'Select a species', and 'Ensembl Archive'.
- Search:** Points to the 'Search Ensembl' box at the top right, which includes a search input field and a 'Go' button.
- Help:** Points to the 'Help' link in the top right navigation bar.
- Help pages and Documentation:** Points to the 'Help & Documentation' section in the left sidebar.
- What's new:** Points to the 'Ensembl headlines: Release 42 (December 2006)' section in the main content area.

STEP 2: Click on "Homo sapiens"

The 'Popular genomes' section on the right lists various species, with 'Homo sapiens' highlighted. A yellow arrow points from the 'Homo sapiens' link to the 'STEP 2' callout box.

At the bottom of the page, there is a footer with copyright information: © 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.



STEP 3:
Type in 'EPO Gene'.
Click 'Go'.

Karyotype

e!Ensembl Human

Ensembl release 42

Search Ensembl *Homo sapiens*

Search: **Go**

e.g. chromosome X or 14:10000..200000 or BRCA2

Karyotype

Click on a chromosome for a closer view

Jump directly to sequence position

Chromosome: or region

From (bp):

To (bp): **Go**

Ensembl headlines: Release 42 (December 2006)

- New - User accounts** (all species)
- New species - Duck-billed Platypus** (*Ornithorhynchus anatinus*)
- New Dog assembly and genebuild** (*Canis familiaris*)
- New Chicken assembly and genebuild** (*Gallus gallus*)
- New Human Ensembl-Vega** (*Homo sapiens*)

[More news...](#)

Go to your account to customise this news panel

Statistics

Assembly:	NCBI 36, Oct 2005
Genebuild:	Ensembl, Aug 2006
Database version:	42.36d
Known genes:	21,774
Novel genes:	1,036
Pseudogenes:	1,069
RNA genes:	3,976
GENSCAN gene predictions:	69,195
Gene exons:	270,661
Gene transcripts:	44,676
Base Pairs¹:	3,253,037,807
Golden Path Length²:	3,093,120,360
Most common InterPro domains:	Top 40 Top 500

¹ Total number of base pairs = sum of lengths of DNA table
² Reference assembly (Golden path) length = sum of non-redundant top level seq regions

You are using the web team's integration server. [More...](#)

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

**Source and version of
assembly and genebuild**

Gene Model

STEP 5:
Click on 'Transcript Information'

Orthologues in other species

Matches in other databases

GO
(Gene Ontology)
terms

Your Ensembl

- Show account · Log out
- Save bookmark
- Save configuration as...

ENST00000252723

- Gene information
- Gene splice site image
- Gene regulation info.
- Genomic sequence
- Gene variation info.
- ID history
- Transcript information**
- Exon information
- Protein information
- Export transmembrane

Chromosome 7

- View of ChIP-seq tracks
- Graphical view of genomic tracks
- Graphical view of genomic tracks
- Export information for a specific region
- Export sequence data for a specific region
- Export Ensembl annotations for a specific region
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Healthchecks

- Health checks
- Old Health checks

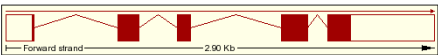
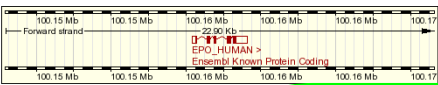
Ensembl Archive

- [View previous release of page in Archive!](#)
[Stable Archive! link for this page](#)



- 🔑 Logins
 - 💖 Bookmarks
 - 🔧 Settings
 - 👤 Groups
- User accounts**
New in Ensemble!

Ensembl Transcript Report

Transcript	EPO_HUMAN (UniProt/B/Swiss-Prot) To view all Exon links linked to the name click here This transcript is a member of the Human CCDS set CCDS85705
Ensembl Transcript ID	ENST00000252723
Transcript information	Exons: 5 Transcript length: 1,328 bps Translation length: 193 residues This transcript is a product of gene: ENSOG0000130427
Genomic Location	This transcript can be found on Chromosome 7 at location 100,156,359-100,159,257 . The start of this transcription is located in Contig AC009488.5.1.98876 .
Description	Erythropoietin precursor (Epoetin). Source: UniProtKB/SWISSPROT P01488
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise/Exonerate model from a database protein or a set of aligned cDNAs followed by an ORF prediction. GeneWise/Exonerate models are further combined with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al., Genome Res. 2004 14:942-50).
Similarity Matches	This Ensembl entry corresponds to the following database identifiers: CCDS: CCDS85705.1 BioMart: EPO_HUMAN [Target Nid: 100; Query Nid: 100] [align] NP_000790.2 [Target Nid: 100; Query Nid: 100] [align] NM_000799.2 [align] Q2M2L6_HUMAN [Target Nid: 100; Query Nid: 100] [align] 2056 A_14_P113914 [Target Nid: 3; Query Nid: 100] A_23_P145864 [Target Nid: 4; Query Nid: 100] A_23_P145869 [Target Nid: 4; Query Nid: 100] AC009488 [align] AF053256 [align] AF202306 [align] AF202307 [align] AF202308 [align] AF202309 [align] AF202310 [align] AF202311 [align] AF202312 [align] AF202313 [align] AF202314 [align] BC093628 [align] BC111937 [align] M11318 [align] S65458 [align] X02157 [align] X02158 [align] IPI: IPD0307226.3 [Target Nid: 100; Query Nid: 100] MIM gene: 133170 PDB: 1BUV 1CN4 1EEER Protein ID: AAAS2400.1 [align] AAC78791.1 [align] AAD13964.1 [align] AAF17572.1 [align] AAF23132.1 [align] AAF23133.1 [align] AAF23134.1 [align] AAH93628.1 [align] AAI1938.1 [align] AAPZ2357.1 [align] CAA26094.1 [align] CAAZ6095.1 [align] UniGene: Hs.2303 [Target Nid: 99; Query Nid: 98] Affymx Microarray FCG110: 207257_at Affymx Microarray HUG6110: 1023_at Affymx Microarray HUgeneFL: XD2158_rna1_at Affymx Microarray UT33: 207257_at 207257_s_at 207257_at 217254_s_at 217254_s_at Affymx Microarray U95: 1023_at GE Healthcare-Amershams Codelink WGA: OE79554 [Target Nid: 2; Query Nid: 100] Illumina Vt: OI_450358B-S [Target Nid: 3; Query Nid: 98]
GO	The following GO terms have been mapped to this entry via UniProt and/or RefSeq: GO:0001866 [From] [response to hypoxia]IEA GO:0005128 [enrhopoietin receptor binding]IEA GO:0005178 [hormone activity]IEA GO:0005573 [extracellular region]IEA GO:0005615 [extracellular space]TAS GO:0006950 [response to stress]TAS GO:0007165 [signal transduction]NAS GO:0007267 [cell-cell signaling]NR GO:0007275 [development]NR GO:0008015 [circulation]NAS GO:0030218 [from] [erythrocyte differentiation]IEA GO:0043249 [erythrocyte maturation]IEA
InterPro	IPR003013 Erythropoietin - View other genes with this domain IPR001323 Erythropoietin/thrombopoietin - View other genes with this domain
Protein Family	ENSP00000006225 : ERYTHROPOIETIN PRECURSOR This cluster contains 1 Ensembl gene member(s) in this species.
Transcript structure	
Transcript neighbourhood	
Transcript sequence	CCCCGGAGCGCAGCCGGGCCACCAGCGGCCTGTCTGTCCAGCACCGGCCCTTCCTGGACAG CGCCCTCTCTCTCCAGGCCCTTGATGGAGTTGGCTCTACAGCGCAACTCTCCGAGAATAGAG GCCCTCCGTGTGGTGGAGCGGCGCCCAGATTGCTCATGGAGCACCCGCGCAGCGGCGAG GATGGGGGTCACAAAATCTCTGTCTGGATGTGGCTCTCTGTGTCCCTGATGTGTGGTCCC TCTGGAGCTCTTAATCTGGAGCGCCACACAGAGCTCATCTGTACAGCAAGCTCTGGAG GAGTACCTCTCTGGAGCGCAGAGAGCGCCAGAAATATCACAGTAGGGCTGTCTGCAACTG CACTCTGAATAGAAATATACTGTCCAGACACAAGAAATTAATTCTATCTCTGGAAAAAG GATGGAGATCGGGACAGAGCTGTAAATCTGTGGCAGAGCTCTGGCTGTGTGTGGAGAGC TGTCTCGGGAGCGAGCTCTGTGTGAATCTCTTCCAGACGTGGAGAGCTCTGGCTCTGGAGCT GCATGTGATGAAGCTCATGTGGCTCTGGACAGAGCACTCTCTCTGGAGCTCTGGAGCG AGCCCGAAGGAGCACTATCCCTCCAGATGGGGCTCAAGTCTCAACTCCGAGAAC CACTGCTGACACTTCCGAGAACTCTCGAGGTCTACTCCAATTCTCTCCGGAGGAAGCT GAAGCTCTACAGAGGAGAGCTCTCGAGGACGGGAGCAATGACAGAGTGTGTCCACTG GGCATATCAACAGCTCTCTCAACAACTTGCTGTGGCAGCACTCTCCCGCGCACTCT GAACCTCTCGAGGGGCTCTCACTCAAGCGAGCTGTCCCATGGAACTCCAGTGGCCCA GCATATGACTCTCGAGGCGCAGAGAACTCTCAAGAGCTCAACTCTGAGTCTAAGATG TCACAGGGCGCAACTTAGGGGCCAGAGCAGAGCACTTAGAGAGCACTTAAATGACAG GGACAGAGCACTCTGGAGGAGCACTGTGACTCACTCGGCACTCGAAATATTATGCC AGACACACTTGGAGTGTATTAATCTTTTTCGACACTACATAGAGAGAGATGAC TGATATACTTAGGGAAGCTGTGAATCTCCAGGCTCACGGGCTGGGCACTCCCT GGTGGCAAGACCCCTTTGACACTGGAGTGGGGAACCTAGAGAGCAAGATGGGGGTG GGCTGTGGCTCTCAGGGGTCCAAATTTTTGTGTCTCAACTCACTGACAGAGACTG AAATCCAC

You are using the web team's integration server. [More →](#)
© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details

The diagram illustrates the analysis of a DNA sequence for various genetic variations. The sequence is shown in a multi-line format with different background colors for exons and introns. Annotations include:

- Exons - alternating text colour**: Exons are highlighted with alternating text colors (yellow and blue).
- Codons - alternating background colour**: Codons are highlighted with alternating background colors (yellow and blue).
- Synonymous SNP**: A single nucleotide change that does not alter the amino acid sequence.
- Non-synonymous SNP**: A single nucleotide change that alters the amino acid sequence.
- Other variation in coding sequence**: A variation that does not fit the standard codon reading frame.
- Translation**: The process of converting the DNA sequence into a protein sequence.
- Affected residue**: A specific amino acid in the protein sequence that is affected by a mutation. (Mouse over shows alternative codons)*
- Ambiguity code**: A code used to represent multiple possible nucleotides at a specific position.
- Other variation in UTR**: A variation in the untranslated region (UTR) of the DNA sequence. (Mouse over shows alleles)*
- UTR SNP**: A single nucleotide change in the UTR. (Mouse over shows alleles)*
- UTR (dark background)**: The untranslated region (UTR) of the DNA sequence, highlighted with a dark background.

Result of STEP 7:

e!Ensembl Human ExonView

Ensembl release 42 - Dec 2006 (ecs33307 - homo_sapiens_core_42_36d)

Search e!Human:

e.g. ENSE00001428812, ENSE00000837374

HOME - BLAST - BIOMART - SITMAP - HELP

Your Ensembl

- Show account - Log out
- Save bookmark
- Save configuration as...

ENSE00000252723

- Gene information
- Gene splice site image
- Gene regulation info
- Genomic tracks
- Gene variants
- ID history
- Transcript
- Exon info
- Protein info
- Export transcript

Chromosome 7
100,156,359 - 100,156,552

- View of Chromosome 7
- Graphical view
- Graphical overview
- Export information about region
- Export sequence as FASTA
- Export EMBL file
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Healthchecks

- Health checks
- Old Health checks

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Ensembl Exon Report

Transcript: **EPO_HUMAN** (UniProt/Swiss-Prot) To view all Ensembl genes linked to this transcript is a member of the Human CCDS set: [CCDS5705](#)

Ensembl Transcript ID: **ENSE00000252723**

Transcript information: Exons: 5 Transcript length: 1,328 bps Translation length: 193 residues
This transcript is a product of gene: [ENSG00000130427](#)
This transcript can be found on Chromosome 7 at location [100,156,359 - 100,156,552](#)
This transcript is located in [Contig AC009488.5:1,988,715](#)
Protein precursor (Epoetin). Source: UniProt/SWISSPROT P01988

Flanking sequence at either end of transcript: 50
Intron base pairs to show at splice sites: 25

Show full intronic sequence ☐
Show exons only ☐

Exon Information

No.	Exon / Intron	Chr	Strand	Start	End	Start Phase	End Phase	Length	Sequence
1	5' upstream sequence	7	1	100,156,359	100,156,552	-	-	-ccttgggcccacccgcccgcgtgcgtgcgtgcgcgcacccgcgtgtcct
2	Intron 1-2	7	1	100,156,553	100,157,116	-	1	-	gagccgggctggggcgtc.....ctctcagcctgggtatctgttctag
3	Intron 2-3	7	1	100,157,263	100,157,520	258	-	-	gtgagacccttcccccagcacatc.....actccccgggctgtgtgcatttcag
4	Intron 3-4	7	1	100,157,521	100,157,607	0	0	87	acgggctgtgtgaaacactgcagcttgaatgagaatatacactgtcccagacaccaaagt
5	Intron 4-5	7	1	100,158,537	100,159,257	0	-	721	gagtaggagcggacacttctgct.....cgacctcctgttttctccttggcag
6	3' downstream sequence	7	1	100,159,258	100,159,257	-	-	-	aatatgactcttggttttctgttttctgggaacctccaaatcccctggc.....

Supporting Evidence

The supporting evidence below consists of the sequence matches on which the exon predictions were based and are sorted by alignment score.

Score: 100 99 97 95 90 85 80 75 70 65 60 50 40 30 20 10 0 NO EVIDENCE

X02157.1
X02157.1 Human mRNA for fetal erythropoietin
P01988
P01988.1 EPO_HUMAN Erythropoietin precursor (Epoetin).

You are using the web team's integration server. [More](#)
© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

STEP 8:
Choose 'Flanking sequence at either end of transcript - 500', tick 'Show full intronic sequence' and click on [Go]

STEP 9:
Click on 'Graphical view'

Flank (green)

Intron (blue)

UTR (purple)

Coding sequence (black)

Supporting evidence

Result of STEP 8:

No. Exon / Intron	Chr	Strand	Start	End	Start Phase	End Phase	Length	Sequence
5' upstream sequence								
1	ENSE00001130431	7	1	100,156,359	100,156,552	-	1	194
Intron 1-2								
2	ENSE00001144077	7	1	100,157,117	100,157,262	1	0	146
Intron 2-3								
3	ENSE00001130423	7	1	100,158,223	100,158,402	0	0	180
Intron 3-4								
4	ENSE00001130416	7	1	100,158,403	100,158,536	0	0	134
5	ENSE00000894545	7	1	100,158,537	100,159,257	0	-	721
3' downstream sequence								

Flank
(green)

Intron
(blue)

Coding sequence
(black)

UTR
(purple)

Result of STEP 9:

The screenshot displays the Ensembl Human ContigView interface. At the top, it shows 'Ensembl release 42 - Dec 2006 (ecs3:3307 - homo_sapiens_core_42_36d)'. The main view is 'Chromosome 7' with a 'Chr. 7 band' showing a 1 Mb region. Below this, a 'Detailed view' of the EPO gene is shown, including 'EMBL mRNAs', 'Vega External gene', 'Ensembl trans.', 'DNA(contigs)', and 'Length'. The 'Gene models' section shows various transcripts and proteins. The 'Basepair view' is also visible at the bottom.

Chromosome

1 Mb region

STEP 12:
Go back one page in the browser to return to the EPO gene.
Select from the 'Features' drop-down menu 'SNPs', and 'Ensembl genes' and close the menu

STEP 11:
Click and drag the mouse to draw a box around another gene (Trip6). Zoom into the gene in Detailed View.

1 kb – 1 Mb region

Mapped proteins and cDNAs

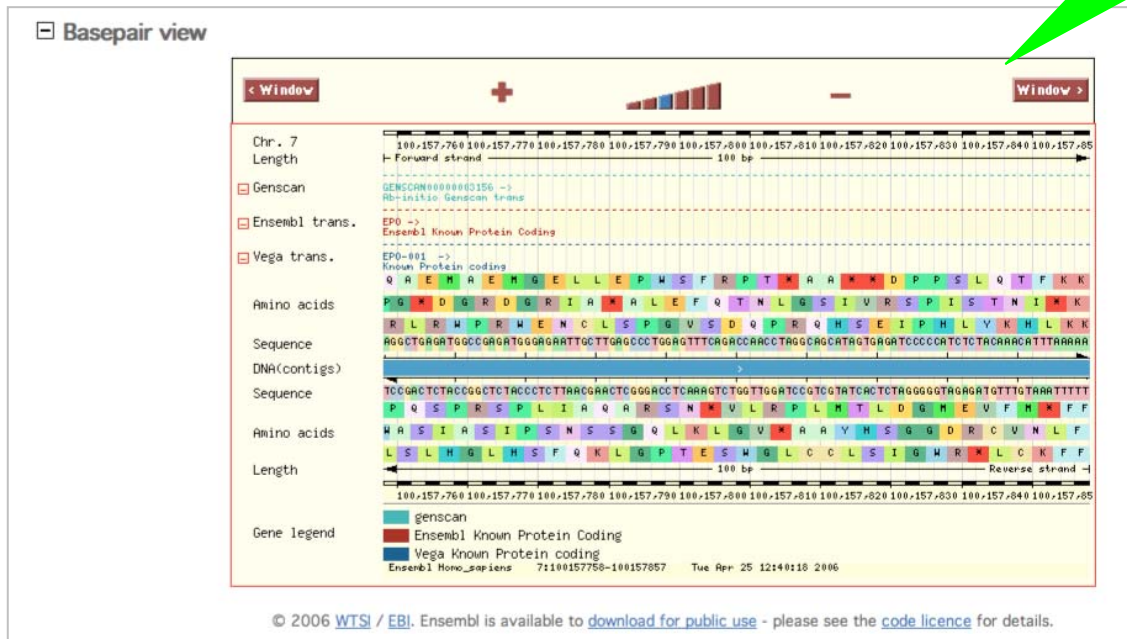
Gene models

Assembly

STEP 10:
Click on the '+' in front of 'Basepair view'

25 – 500 bp
region

Result of STEP 10:



Result of STEP 12:

Detailed view

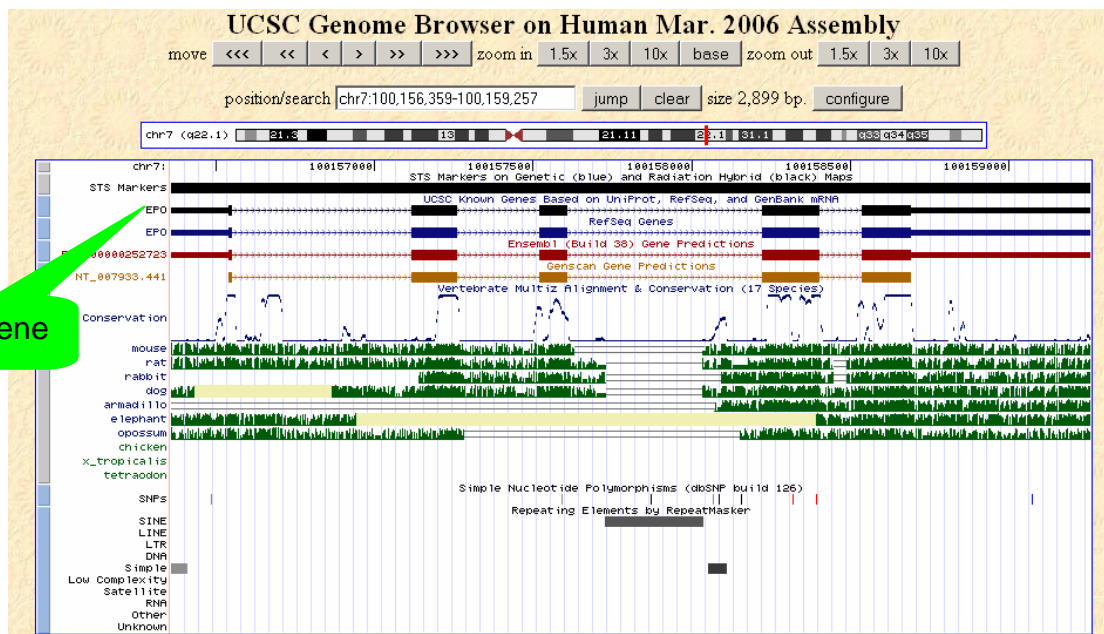


STEP 13:
Click on a SNP
(vertical line) and
subsequently on
'SNP properties'

[illegible]

STEP 14:
Go back to
ContigView with the
back button of the
internet browser.

STEP 15:
To see the same
chromosomal region
in the UCSC
genome browser,
click on 'Show in
UCSC browser' on
the left of the page.
A new window will
open.



STEP 16:
Once you see the EPO
gene and close this
window. (You can turn
on 'Ensembl genes' by
changing 'hide' to 'full')

Click on 'Graphical
Overview' on the left
hand of the ContigView
page to reach
CytoView.

e!Ensembl Human CytosView

Ensembl release

Your Ensembl

- Show account
- Save bookmarks
- Save configurations

Chromosome

99,657,808

100,657,808

101,657,808

102,657,808

103,657,808

104,657,808

105,657,808

106,657,808

107,657,808

108,657,808

109,657,808

110,657,808

111,657,808

112,657,808

113,657,808

114,657,808

115,657,808

116,657,808

117,657,808

118,657,808

119,657,808

120,657,808

121,657,808

122,657,808

123,657,808

124,657,808

125,657,808

126,657,808

127,657,808

128,657,808

129,657,808

130,657,808

131,657,808

132,657,808

133,657,808

134,657,808

135,657,808

136,657,808

137,657,808

138,657,808

139,657,808

140,657,808

141,657,808

142,657,808

143,657,808

144,657,808

145,657,808

146,657,808

147,657,808

148,657,808

149,657,808

150,657,808

151,657,808

152,657,808

153,657,808

154,657,808

155,657,808

156,657,808

157,657,808

158,657,808

159,657,808

160,657,808

161,657,808

162,657,808

163,657,808

164,657,808

165,657,808

166,657,808

167,657,808

168,657,808

169,657,808

170,657,808

171,657,808

172,657,808

173,657,808

174,657,808

175,657,808

176,657,808

177,657,808

178,657,808

179,657,808

180,657,808

181,657,808

182,657,808

183,657,808

184,657,808

185,657,808

186,657,808

187,657,808

188,657,808

189,657,808

190,657,808

191,657,808

192,657,808

193,657,808

194,657,808

195,657,808

196,657,808

197,657,808

198,657,808

199,657,808

200,657,808

201,657,808

202,657,808

203,657,808

204,657,808

205,657,808

206,657,808

207,657,808

208,657,808

209,657,808

210,657,808

211,657,808

212,657,808

213,657,808

214,657,808

215,657,808

216,657,808

217,657,808

218,657,808

219,657,808

220,657,808

221,657,808

222,657,808

223,657,808

224,657,808

225,657,808

226,657,808

227,657,808

228,657,808

229,657,808

230,657,808

231,657,808

232,657,808

233,657,808

234,657,808

235,657,808

236,657,808

237,657,808

238,657,808

239,657,808

240,657,808

241,657,808

242,657,808

243,657,808

244,657,808

245,657,808

246,657,808

247,657,808

248,657,808

249,657,808

250,657,808

STEP 18:
Click on 'View Syntenic regions ... with *Mus musculus*'

STEP 17:
Make sure '1Mb clones', '30k TPA clones', '32k clones' and 'Human tilepath clones' are selected under 'Decorations.' Zoom out 2 steps.

200 kb – 50 Mb region

BAC clones

Tiling path clones

Export cloneset information

Export data

Select Set of features to render

select

Output format

HTML

Select type to export


select

Export


Fields marked with * are required

You are using the web team's integration server. [More...](#)

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.



Human genes



Mouse homologues

e!Ensembl

Ensembl release 42

Your Ensembl

- Show account : Log out
- Save bookmark
- Save configuration as...

Chromosome 7


- View Chromosome 7
- View Chr 7 Synteny
- Map your data onto this chromosome

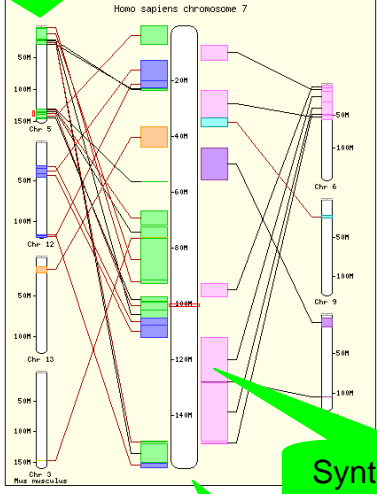
Healthchecks

- Health checks
- Old Health checks

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page





Mouse chromosomes

Human chromosome

Syntenic block

Homology Matches

Homo sapiens Genes	Mus musculus Homologues
EPO (0.10 Gb) [ContigView]	-> Epo (5: 137.71 Mb) [ContigView] [MultiContigView]
ZAN (0.10 Gb) [ContigView]	No homologues
EPHB4 (0.10 Gb) [ContigView]	-> Ephb2 (5: 137.58 Mb) [ContigView] [MultiContigView]
SLC12A9 (0.10 Gb) [ContigView]	-> Slc12a2 (18: 58.00 Mb) [ContigView] [MultiContigView]
	-> Slc12a9 (5: 137.54 Mb) [ContigView] [MultiContigView]
	-> Slc12a3 (8: 97.22 Mb) [ContigView] [MultiContigView]
	-> Slc12a4 (8: 108.83 Mb) [ContigView] [MultiContigView]
	-> Slc12a7 (13: 74.20 Mb) [ContigView] [MultiContigView]
	-> Slc12a8 (16: 33.44 Mb) [ContigView] [MultiContigView]
	-> Slc12a1 (1: 124.84 Mb) [ContigView] [MultiContigView]
	-> Slc12a5 (2: 164.66 Mb) [ContigView] [MultiContigView]
	-> Slc12a6 (2: 112.07 Mb) [ContigView] [MultiContigView]
	-> Zfx (6: 42.28 Mb) [ContigView] [MultiContigView]
	-> Wtp (7: 33.82 Mb) [ContigView] [MultiContigView]
	-> Trig6 (5: 137.54 Mb) [ContigView] [MultiContigView]
	-> Jub (14: 53.52 Mb) [ContigView] [MultiContigView]
	-> Lpp (16: 24.31 Mb) [ContigView] [MultiContigView]
	-> Lmd1 (9: 123.33 Mb) [ContigView] [MultiContigView]
	-> Fbln1 (4: 140.85 Mb) [ContigView] [MultiContigView]
ARS2_HUMAN (0.10 Gb) [ContigView]	-> Ars2 (5: 137.53 Mb) [ContigView] [MultiContigView]
NP_001015072.1 (0.10 Gb) [ContigView]	-> 2700038N03Rik (5: 137.52 Mb) [ContigView] [MultiContigView]
	-> 1810047C23Rik (8: 47.47 Mb) [ContigView] [MultiContigView]
ACHE (0.10 Gb) [ContigView]	-> Ache (5: 137.52 Mb) [ContigView] [MultiContigView]
	-> Bche (3: 73.72 Mb) [ContigView] [MultiContigView]
ENSG00000208819 (0.10 Gb) [ContigView]	No homologues
MUC3B (0.10 Gb) [ContigView]	No homologues
Q96MA9_HUMAN (0.10 Gb) [ContigView]	No homologues
MUC12 (0.10 Gb) [ContigView]	No homologues
ENSG00000205277 (0.10 Gb) [ContigView]	No homologues
Q9UKH1_HUMAN (0.10 Gb) [ContigView]	No homologues
MUC17 (0.10 Gb) [ContigView]	No homologues

Navigate Homology

[Upstream](#) (<0.10 Gb) [Downstream](#) (>0.10 Gb)

Change Chromosome

Chromosome 7

Fields marked with * are required

STEP 19:
Click on
[MultiContigView]

You are using the web team's integration server. [More](#) →

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

Your Ensembl

- Show account : Log out
- Save bookmark
- Save configuration as...

Chromosome 7
100,155,359 - 100,160,257

- View of Chromosome 7
- Graphical view of...
- Graphical overview
- View alignment with ...
- View alongside ...
- View Syntenic regions ...
- View region at UCSC
- View region in NCBI browser

Export data

- Export information about region
- Export sequence as FASTA
- Export EMBL file
- Get Gene info in region
- NP info in region
- NP info in region

STEP 20:
Click on 'Export
sequence as FASTA'

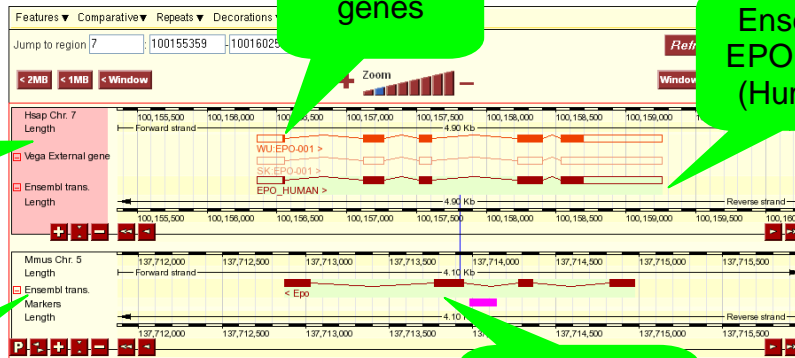
Top level



Navigational overview



Detailed View



Human

Vega EPO genes

Ensembl EPO gene (Human)

Mouse

Mouse EPO homologue



e!Ensembl Human ExportView

Search e!Human:

e.g. AL138722.15.1.44776, ENSG00000139618

Ensembl v100,153,359 - 100,160,257

STEP 21:
Click on
[Continue>>]

File output for FASTA format text file

Chromosome 7 100,155,359 - 100,160,257.

Continue >>

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

e!Ensembl Human ExportView

Search e!Human:

e.g. AL138722.15.1.44776, ENSG00000139618

HOME · BLAST · BIOMART · SITEMAP · HELP

STEP 22:
Select and copy a part of the sequence

STEP 23:
Click on 'BLAST'

Chromosome 7
100,155,359 - 100,160,257

View of Chromosome 7
Graphical View
Graphical overview
View alignment with ...
View alongside ...
View Syntenic regions ...
View region at UCSC
View region in NCBI
browser

Export data

- Export information about region
- Export sequence as FASTA
- Export EMBL file
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Healthchecks

- Health checks
- Old Health checks

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Saccharomyces cerevisiae
Yeast
Revised assembly

You are using the web team's integration server. [More ...](#)
© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

STEP 24:
Paste the copied sequence

STEP 25:
Select 'Homo_sapiens' and 'BLASTN' and click on [RUN>]

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

STEP 26:
Click on [Retrieve] to check for results

STEP 27:
Click on [VIEW]

Summary of BLAST search

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.



e!Ensembl Human BlastView

Search e!Human: Anything e.g. AL138722.15.1.44776, ENSG00000139618

Use Ensembl: Run a search, Data mining (BioMart), Upload your own data, Export data, Download data

Docs and downloads: Information, What's New, About Ensembl, Ensembl data, Software

Other links: Home, Sitemap, Vega, Pre Ensembl, View previous release of page in Archive!, Stable Archive! link for this page, Archive! sites, Trace server

Location of hits on the genome

alignments vs Homo_sapiens LATESTGP database
alignments of 107, sorted by Raw Score

Best hit

Alignment of hits to query sequence

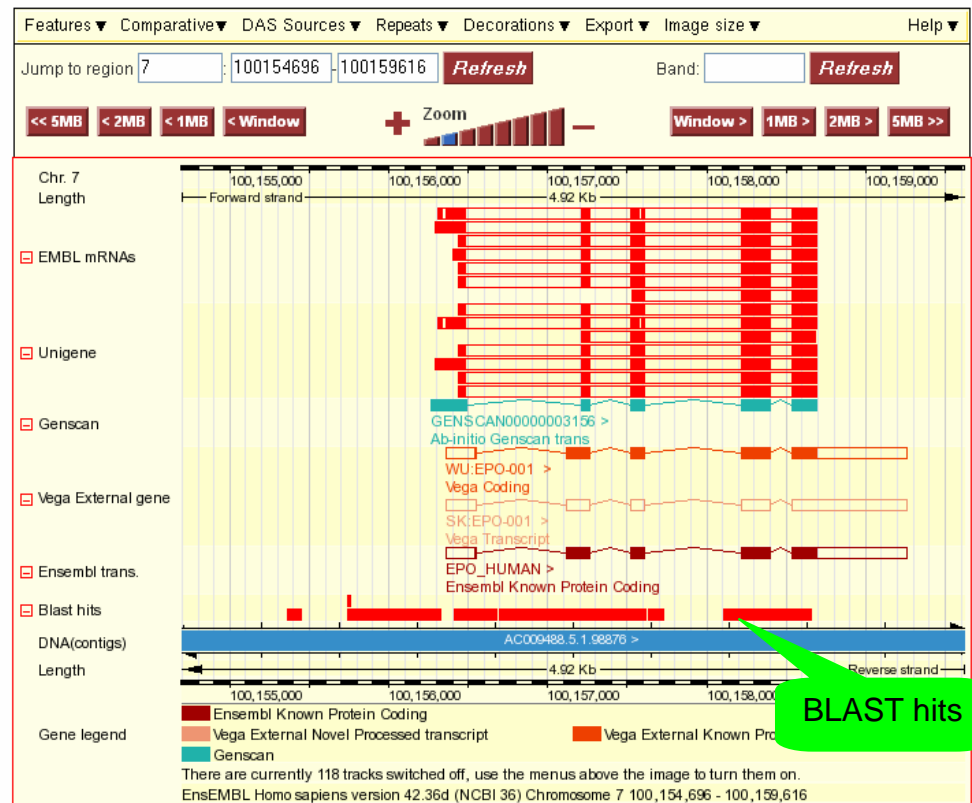
STEP 28:
Click on [C] in front of best hit

Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Query	Subject	Chromosome	Supercontig	Clone	Contig	Chromosome	Stats	Sort By
off Name Start	_off_ Name Start	_off_ Name Start	_off_ Name Start	_off_ Name Start	_off_ Name Start	_off_ Name Start	_off_ Name Start	>Chromosome <Score >Score
Links	Query	Start	End	Chromosome	Name	Start	End	Ori
[A] [S] [G] [C]	1538	2258	+	Chr-7	100156696	100157616	+	921 0. 100.00 921
[A] [S] [G] [C]	384	144	+	Chr-7	100155745	100156329	+	585 0. 100.00 585
[A] [S] [G] [C]	1051	1788	+	Chr-7	100156409	100156678	+	270 0. 100.00 270
[A] [S] [G] [C]	2753	2880	+	Chr-7	100158111	100158238	+	128 0. 100.00 128
[A] [S] [G] [C]	2274	2360	+	Chr-7	100157632	100157727	+	96 0. 100.00 96
[A] [S] [G] [C]	1	91	+	Chr-7	100155449	100155499	+	91 0. 100.00 91
[A] [S] [G] [C]	1571	1611	+	Chr-7	100157632	100157727	+	22 0.84 100.00 22
[A] [S] [G] [C]	2852	2912	+	Chr-7	100157632	100157727	+	22 0.84 93.33 30
[A] [S] [G] [C]	422	442	+	Chr-7	100157632	100157727	+	21 0.25 100.00 21
[A] [S] [G] [C]	2335	2352	+	Chr-7	100157632	100157727	+	21 0.88 100.00 21
[A] [S] [G] [C]	740	770	+	Chr-7	100157632	100157727	+	21 2.7 96.00 25
[A] [S] [G] [C]	1212	1212	+	Chr-7	100157632	100157727	+	21 3.2 96.00 25
[A] [S] [G] [C]	877	877	+	Chr-7	100157632	100157727	+	21 3.3 100.00 21
[A] [S] [G] [C]	807	807	+	Chr-7	100157632	100157727	+	21 3.3 96.00 25
[A] [S] [G] [C]	2814	2814	+	Chr-7	100157632	100157727	+	21 3.3 100.00 21
[A] [S] [G] [C]	1434	1434	+	Chr-7	100157632	100157727	+	21 3.3 96.00 25
[A] [S] [G] [C]	423	443	+	Chr-7	100157632	100157727	+	21 3.3 100.00 21
[A] [S] [G] [C]	1806	1826	+	Chr-3	47263809	47263829	+	21 3.3 100.00 21
[A] [S] [G] [C]	877	897	+	Chr-11	45147216	45147236	+	21 4.1 100.00 21
[A] [S] [G] [C]	899	918	+	Chr-4	10100112	10100131	+	20 0.074 100.00 20
[A] [S] [G] [C]	563	582	+	Chr-1	23045906	23045925	+	20 2.6 100.00 20
[A] [S] [G] [C]	875	897	+	Chr-17	2245254	2245277	+	20 2.8 95.83 24
[A] [S] [G] [C]	1072	1094	+	Chr-11	2404836	2404859	+	20 3.1 95.83 24
[A] [S] [G] [C]	2235	2254	+	Chr-2	55009141	55009160	+	20 3.2 100.00 20
[A] [S] [G] [C]	873	896	+	Chr-19	18119374	18119397	+	20 3.4 95.83 24
[A] [S] [G] [C]	809	828	+	Chr-12	21163734	21163753	+	20 4.0 100.00 20
[A] [S] [G] [C]	878	897	+	Chr-2	43261275	43261294	+	20 4.1 100.00 20
[A] [S] [G] [C]	661	679	+	Chr-1	18606381	18606399	+	19 0.19 100.00 19
[A] [S] [G] [C]	1768	1797	+	Chr-1	18656691	18656720	+	19 0.19 90.32 31
[A] [S] [G] [C]	1741	1770	+	Chr-10	45140401	45140430	+	19 1.2 90.32 31
[A] [S] [G] [C]	418	436	+	Chr-20	44598942	44598960	+	19 2.4 100.00 19
[A] [S] [G] [C]	2209	2227	+	Chr-14	46044909	46044927	+	19 3.2 100.00 19
[A] [S] [G] [C]	2804	2822	+	Chr-3	12983491	12983509	+	19 4.3 100.00 19
[A] [S] [G] [C]	881	899	+	Chr-14	104884074	104884092	+	19 5.3 100.00 19
[A] [S] [G] [C]	881	899	+	Chr-14	104884028	104884046	+	19 5.3 100.00 19
[A] [S] [G] [C]	881	899	+	Chr-14	104883982	104884000	+	19 5.3 100.00 19
[A] [S] [G] [C]	69	86	+	Chr-7	100155741	100155758	+	18 0. 100.00 18
[A] [S] [G] [C]	1905	1922	+	Chr-X	39633546	39633563	+	18 0.70 100.00 18
[A] [S] [G] [C]	1962	1979	+	Chr-X	70306710	70306727	+	18 3.9 100.00 18
[A] [S] [G] [C]	1700	1720	+	Chr-X	70247948	70247969	+	18 3.9 95.45 22
[A] [S] [G] [C]	2306	2323	+	Chr-1	150284632	150284649	+	18 6.0 100.00 18
[A] [S] [G] [C]	1801	1818	+	Chr-2	73022513	73022530	+	18 9.3 100.00 18
[A] [S] [G] [C]	74	90	+	Chr-2	29824827	29824843	+	17 0.25 100.00 17
[A] [S] [G] [C]	881	897	+	Chr-X	39611091	39611107	+	17 0.70 100.00 17
[A] [S] [G] [C]	881	897	+	Chr-1	20541390	20541406	+	17 1.1 100.00 17
[A] [S] [G] [C]	1798	1814	+	Chr-1	20511096	20511112	+	17 1.1 100.00 17
[A] [S] [G] [C]	1487	1503	+	Chr-10	43031647	43031663	+	17 2.2 100.00 17
[A] [S] [G] [C]	2100	2119	+	Chr-14	104633755	104633775	+	17 1.8 95.24 21
[A] [S] [G] [C]	1564	1580	+	Chr-9	139178040	139178056	+	17 2.2 100.00 17
[A] [S] [G] [C]	883	899	+	Chr-9	139163518	139163534	+	17 2.2 100.00 17
[A] [S] [G] [C]	2805	2825	+	Chr-14	46020025	46020044	+	17 3.2 95.24 21
[A] [S] [G] [C]	1577	1593	+	Chr-16	88233355	88233371	+	17 3.5 100.00 17
[A] [S] [G] [C]	1615	1629	+	Chr-19	47367644	47367660	+	17 5.4 100.00 17
[A] [S] [G] [C]	1805	1821	+	Chr-1	150273546	150273562	+	17 6.0 100.00 17
[A] [S] [G] [C]	737	756	+	Chr-6	34166916	34166936	+	17 7.0 95.74 21

Back in the contigview page...

Detailed view



END of the
Worked Example