

Module 2 - Maps and Genome Sequence

Aims

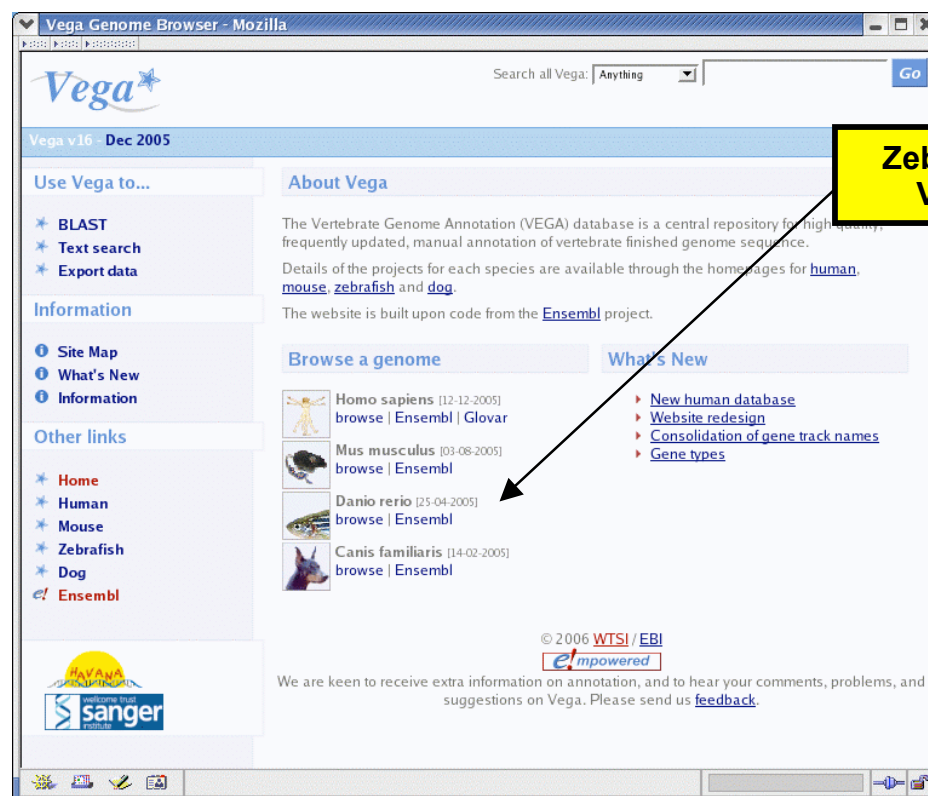
ii – The Vega Genome Browser

- Introduce the Vega genome browser
- Explain the source of the data in Vega
- Show the different Vega views stressing the differences to the Ensembl views

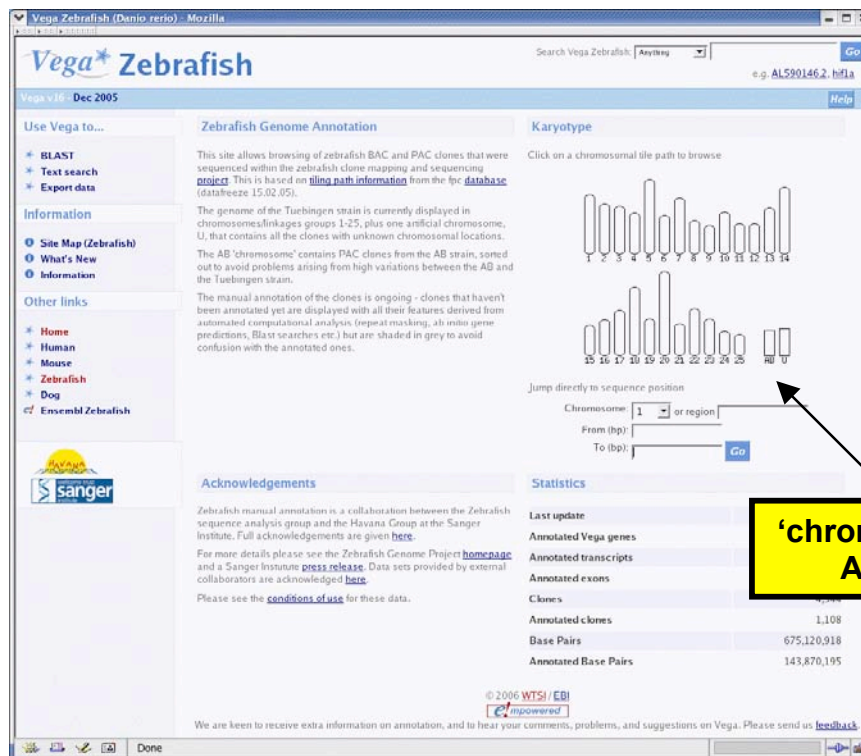
Introduction

The Vertebrate Genome Annotation (Vega) database is a central repository for high quality, frequently updated, manual annotation of vertebrate finished genome sequence. The *Danio rerio* Vega database contains all the finished clones. Unlike the *Danio rerio* Ensembl database, the Vega database only contains high-quality sequence with high-quality manual annotation. The annotation is undertaken in collaboration and synchronisation with the central zebrafish database ZFIN. The implementation of the Vega browser is based on the Ensembl code and so they share many features and functionality. This section gives a brief introduction to the Vega views emphasising the differences with Ensembl. Refer to section 2 for more details on the Ensembl views.

The main Vega page is <http://vega.sanger.ac.uk>. One obvious difference between Ensembl and Vega is the background colour. In Ensembl it is yellow whereas in Vega is blue.



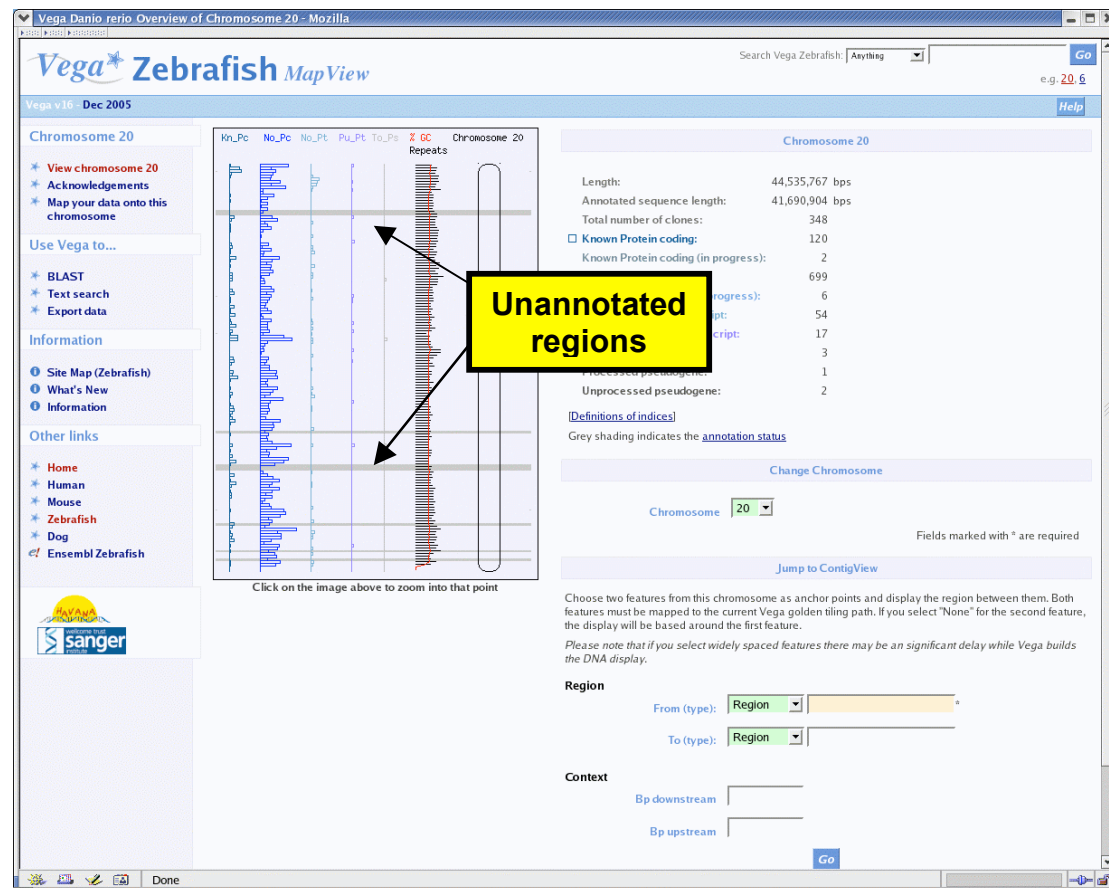
Follow the link to the zebrafish database.



This zebrafish page displays all the chromosomes plus two 'artificial' ones called AB and U. Chromosome AB groups all the sequenced clones for a PAC library made from the AB strain. Some of these clones have been manually annotated. Chromosome U contains finished clones which have not been placed in the physical map. The lengths do not represent an estimation of the size of the real chromosomes, but the amount of the current finished sequence. Chromosome 20 is top priority and that is reflected by the fact that it is the longest, ie the one with the most finished sequence.

MapView and ContigView

Clicking on a chromosome links to the corresponding MapView page.



The regions shaded in grey in the **MapView** pages indicate segments of the chromosome that have not been annotated yet. Check the difference between chromosomes 20 and 7 in terms of how much sequence has been annotated. As chromosome 20 is a priority one most of the sequence is not greyed, for chromosome 7 the situation is just the opposite.

The **ContigView** is, like in Ensembl, one of the main pages in Vega. The contents of the ContigView include some data specific to the manual annotation, for example there is a track for polyA signals. In order to facilitate the task of the annotators, the alignments of protein and ESTs is done more aggressively over finished clones than over the assembly. The most important track in the Vega ContigView page is the 'Zfish transcripts', the manually annotated transcripts. Observe that shaded regions do not contain this kind of transcript since they have not been annotated yet.

Jump to the ContigView for the region in chromosome 20 from 17925000 to 18135782.

The screenshot displays the Vega Zebrafish ContigView interface, which is a web-based tool for viewing genomic data. The interface includes a navigation bar with links to Home, Dog, Human, Mouse, Zebrafish, BLAST, Export Data, Search, Feedback, and Help. A search bar is located at the top right, with a 'Lookup' button and a search term 'e.g. AL590146.2, BX842684'. The main content area is divided into three panels:

- Overview:** This panel shows a genomic map of Chromosome 20. It includes a scale bar from 16.40 Mb to 16.70 Mb. Below the scale bar, there are tracks for DNA(contigs), Markers, Zfish Genes, and Gene legend. The Zfish Genes track shows several genes, including *jag2*.
- Poly-A features:** This panel provides a detailed view of the genomic region. It includes a scale bar from 16.51 Mb to 16.59 Mb. The left sidebar lists various features: Length, Ensembl clones, DNA(contigs), PolyA site, PolyA signal, Zfish trans., GenScan, Fgenesh, UniProt, EMBL nrNhs, Markers, CpG islands, and Length. The main content area shows a detailed view of the *jag2* gene, including its structure, sequence, and various annotations.
- Basepair view:** This panel shows the raw sequence data for the genomic region. It includes a scale bar from 16.551-200 to 16.551-260. The left sidebar lists various features: Length, Ensembl clones, Amino acids, Sequence, DNA(contigs), Amino acids, Zfish trans., GenScan, Fgenesh, Restr. Enzymes, and Length. The main content area shows the raw sequence data, including the DNA sequence, amino acid sequence, and various annotations.

Annotations in the Poly-A features panel include 'Manually annotated genes' and 'Poly-A features'. The Basepair view panel shows a detailed view of the *jag2* gene, including its structure, sequence, and various annotations.

This region contains a transcript labelled *jag2*. Look for this name in the 'Overview' panel.

GeneView, TransView, ExonView and ProteinView

Follow the link to the GeneView page from the *jag2* transcript *jag2-001*.

Vega Zebrafish GeneView The Wellcome Trust Sanger Institute

Home Dog Human Mouse Zebrafish [e.g. hif1a, OTTDARG0000004630]

Link to ZFIN

Type

Author

Transcript information links

Transcripts

Curated Locus Report

Curated Locus	jag2 (ZFIN ID) (to view all Vega genes linked to the name click here)
Locus ID	OTTDARG00000005397
Version	1
Date	Gene last modified on 15/06/2004 (Created on 15/06/2004)
Alternative Symbols	ZDB-GENE-011128-3
Type	Known [Definition]
Genomic Location	View gene in genomic location: 16509110 - 16593534 bp (16.5 Mb) on chromosome 20 This gene is located in sequence: chunk770
Description	jagged2
Author	This locus was annotated by zfish-zfish-help@sanger.ac.uk
Database Matches	ZFIN: jag2
Sequence Markup	View genomic sequence for this gene with exons highlighted
Export Data	Export gene data in EMBL, GenBank or FASTA
Curated Transcripts	<p>1: DKEY-5P1.1-001 (OTTDART0000005844) [Transcript information] [Exon information & supporting evidence] [Protein information]</p> <p>2: DKEY-5P1.1-002 (OTTDART0000005845) [Transcript information] [Exon information & supporting evidence] [Protein information]</p>

Transcript/Translation Summary

DKEY-5P1.1-002	<p>Stable ID: OTTDART0000005845 Version: 1 Class: Coding</p> <p>Exons: 25 Transcript length: 5324 bp Translation length: 1216 residues</p> <p>[Transcript information] [Exon information & supporting evidence] [Protein information]</p>
InterPro	<p>IPR001438 Type II EGF-like signature - [View other Vega genes with this domain]</p> <p>IPR001881 EGF-like calcium-binding - [View other Vega genes with this domain]</p> <p>IPR001687 ATP/GTP-binding site motif A (P-loop) - [View other Vega genes with this domain]</p> <p>IPR001774 Delta/Serrate/lag-2 (DSL) protein - [View other Vega genes with this domain]</p> <p>IPR007042 EGF-like domain, subtype 2 - [View other Vega genes with this domain]</p> <p>IPR001093 IMP dehydrogenase/GMP reductase - [View other Vega genes with this domain]</p> <p>IPR001052 Aspartic acid and asparagine hydroxylation site - [View other Vega genes with this domain]</p> <p>IPR006209 EGF-like domain - [View other Vega genes with this domain]</p>
Transcript Structure	
Protein Features	<p>Prints</p> <p>Profile</p> <p>Prosite</p> <p>Pfam</p> <p>Transmembrane</p> <p>Signal peptide</p> <p>Low complexity</p> <p>Peptide</p> <p>Scale (aa)</p>

DKEY-5P1.1-001	<p>Stable ID: OTTDART0000005844 Version: 1 Class: Coding</p> <p>Exons: 26 Transcript length: 5438 bp Translation length: 1264 residues</p> <p>[Transcript information] [Exon information & supporting evidence] [Protein information]</p>
InterPro	<p>IPR001438 Type II EGF-like signature - [View other Vega genes with this domain]</p> <p>IPR001881 EGF-like calcium-binding - [View other Vega genes with this domain]</p> <p>IPR001687 ATP/GTP-binding site motif A (P-loop) - [View other Vega genes with this domain]</p> <p>IPR001774 Delta/Serrate/lag-2 (DSL) protein - [View other Vega genes with this domain]</p> <p>IPR007042 EGF-like domain, subtype 2 - [View other Vega genes with this domain]</p> <p>IPR001093 IMP dehydrogenase/GMP reductase - [View other Vega genes with this domain]</p> <p>IPR001052 Aspartic acid and asparagine hydroxylation site - [View other Vega genes with this domain]</p> <p>IPR006209 EGF-like domain - [View other Vega genes with this domain]</p>
Transcript Structure	
Protein Features	<p>Prints</p> <p>Profile</p> <p>Prosite</p> <p>Pfam</p> <p>Transmembrane</p> <p>Signal peptide</p> <p>Low complexity</p> <p>Peptide</p> <p>Scale (aa)</p>

Every annotated gene in Vega has a ZFIN gene entry. Follow the link in the example by clicking on the name of the gene (jag2). Another special feature of the Vega GeneView page is the fields for the authors of the annotation, and

the type of gene. The gene type gives an indication of the confidence of the annotation based on the available evidence, for example:

- a gene has type **known** if it was listed by ZFIN at the moment of the annotation (eventually every annotated gene will have an entry in ZFIN), and
- a gene has type **novel CDS** if its product was similar to, but not identical to, a known protein from zebrafish or another organism.

Transcripts are also classified in several categories as well.

The gene *jag2* has been annotated with two transcripts. Follow the link labelled 'Transcript information' for the transcript OTTDART0000005844 to open the **TransView** page.

Vega Transcript Report


Transcript	DKEY-SP1.1-001 (Vega transcript ID)
Vega Transcript ID	OTTDART0000005844
Version	1
Class	Coding [Definition]
Transcript Information	Exons: 26 Transcript length: 5438bp Translation length: 1254 residues This transcript is a product of gene: OTTDART0000005837 [Exon information & supporting evidence] [Protein information]
Genomic Location	View transcript in genomic location: 16509110 - 16593534 bp (16.5 Mb) on chromosome 20 This transcript is located in sequence chunk770
Description	jagged2
Author	This locus was annotated by zfish-zfish-help@sanger.ac.uk
InterPro	IPR001438 Type II EGF-like signature - [View other Vega genes with this domain] IPR001881 EGF-like calcium-binding - [View other Vega genes with this domain] IPR001687 ATP/GTP-binding site motif A (P-loop) - [View other Vega genes with this domain] IPR001774 Delta/Serratelag-2 (DSL) protein - [View other Vega genes with this domain] IPR000742 EGF-like domain, subtype 2 - [View other Vega genes with this domain] IPR001093 IMP dehydrogenase/GMP reductase - [View other Vega genes with this domain] IPR000152 Aspartic acid and asparagine hydroxylation site - [View other Vega genes with this domain] IPR008209 EGF-like domain - [View other Vega genes with this domain]
Export Data	Export transcript data in EMBL, GenBank or FASTA

Transcript cDNA Sequence
[No markup] [No numbers]

Transcript Structure
[Diagram showing exons and introns]

Transcript Neighbourhood
[Diagram showing the transcript's position relative to other transcripts]

The **ExonView** page for this transcript gives more information about the sequence of the exons and introns and the supporting evidence used in the annotation. Follow the link labelled 'Exon information & supporting evidence' to open the ExonView page.

Vega Zebrafish *ExonView* 

Home Dog Human Mouse Zebrafish BLAST Export Data Search Feedback Help

Help on ExonView Find [e.g. CH211-212G7.4-001, OTTDART0000004952]

Vega Exon Report

Transcript	DKEY-SP1.1-001 (Vega_transcript ID)
Vega Transcript ID	OTTDART0000005844
Version	1
Class	Coding [Definition]
Transcript Information	This transcript is a product of Ensembl gene OTTDARG0000005397 [Transcript Information] [Supporting Evidence] [Peptide Information]
Genomic Location	View transcript in genomic location: 16509110 - 16593534 bp (16.5 Mb) on chromosome 20 This transcript is located in sequence: chunk770
Description	jagged2

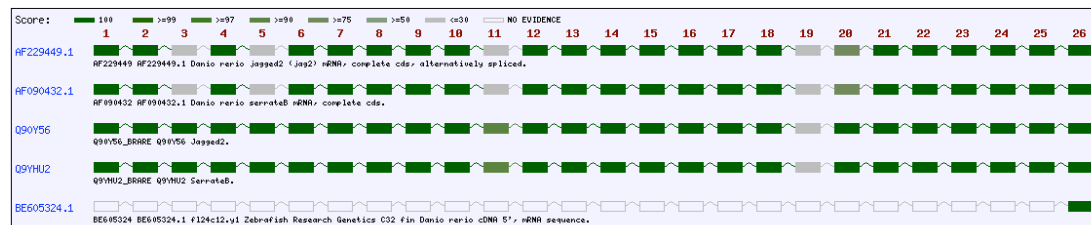
Exon Information

No.	Exon / Intron	Chr	Strand	Start	End	Length	Sequence
	5' upstream sequence					gagatctcctcaggcgtgtgtagccgtgaacatcatcattgccgtgaa
1	OTTDARE00000033254	20	-1	16593319	16593534	216 bp	GTGATCAGACCGAGGGAGAGATCAGCAGACACCATCACCGCGCAACACACCCACGCTCGT GAATTTTTCATGTCAGGAACGGAGATCCTGTCGCGGTCATCGGCCGTTTCCATCTT GCGTTTATACACATCAATCGCGGCATGTGGAAATGTATCAGGATTAGGAATGGCTC CCAATCGCGTGCCTGCTTTAAGCATGTGACGAAG gtgagagctctgattgttattagtg.....ttttgttttcaatgtatttttag
	Intron 1-2	20	-1	16592862	16593318	457 bp	
2	OTTDARE00000033260	20	-1	16592541	16592861	321 bp	GTGTCCAGTCTCTGGGTATTTGAGCTGCAGCTGATTGCTGTAGAAAATGTAAACGGT GAGTTGTGGGACGGGGAATGTTGCCAGCAGCAGCGGAACCTCAAGACACGCGCTGCGTG CGGACGAGTCCGATACCTACTTTAAAGTGTGCTGAAGGATACCACTCTGAAGTCACC ACCACTGGACAGTGCACCTTCGGCTCTGGATCTACGACGTTCTTGGTGGAAATATAAT TCTTTTAAGACCGCAAAAACAGCCCAAGCAAAACGAGCGACGTGGGAAAGATCATCATC CCTTTTCACCTCGCCTGGCCG gtgagtcctcgtctctccatgggc.....gttattctctctctctctctcttag
	Intron 2-3	20	-1	16565147	16592540	2794 bp	
3	OTTDARE00000033271	20	-1	16565092	16565146	55 bp	CGATCTACACACTCATCTTGAAGCTGGGACTGGGATACTCCACTCAGAACA

At the bottom of this page there is diagram showing the supporting evidence for this annotation.

Supporting evidence

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



In this example there is no evidence for exon 19, indicating that the annotator has 'built' this exon from other evidences such as splice sites, codon bias and ORFs. Compare this situation and what you see in the Ensembl predictions.

The link labelled 'Peptide Information' opens the **ProteinView** page. This data is generated automatically using the predicted translation in very much the same fashion as done for the Ensembl annotation.

Other views that we discussed in the module for the Ensembl browser are also present in Vega, for example, **ExportView** to download the data in files.

Exercises

1. Open the GeneView page for jag2 and visit the associated links. In particular open the ContigView page showing this gene.

2. Study the differences between the manual annotation for jag2 in Vega and the automatic annotation in Ensembl (see the Ensembl section for an example of how to open the GeneView page for jag2 in Ensembl).
3. One of the differences between the automatic prediction and the manually annotated jag2 is the number of exons and the UTR. Why do you think these data are different?
4. Customise the ContigView page to turn on the track for poly-A signals.
5. Another special track in Vega ContigView is 'Assembly tags'. This features information on special regions of the clones. These data are entered by the person in charge of finishing the sequence. A region that contains one of these tags is the finished clone 'AL928990'. Open a ContigView page for this clone in Vega and check the text for the assembly tag.
6. Many clones present in Vega are also placed in the assembly and therefore can be browsed in Ensembl. As Vega is updated more often there might be a difference in the versions. A sequenced clone may go through several updates from its first to its final submission, these are recorded via the version numbers. Check for clone present in the assembly and compare it to one in Vega.