

Module 2: Maps and Genome Sequence

(Sanger)

ii - The Vega Genome Browser

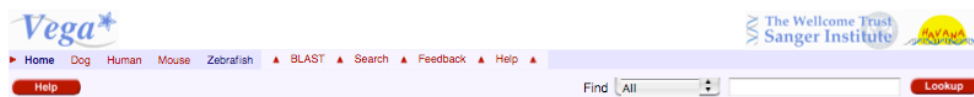
Aims

- Introduce the Vega genome browser
- Explain the source of the data in Vega
- Show the different Vega views stressing 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 currently covering approximately half of the genome. 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. This module gives a brief introduction to the Vega views emphasising the differences with Ensembl. Refer to module 2.i for more details.

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 it is blue.



Vega Genome Browser

Species	Last Updated
Human	12 May 2005
Mouse	14 Apr 2005
Zebrafish	15 Apr 2005
Dog	14 Feb 2005

What's new

Update to Vega Human
12th May 2005
This release of Vega contains manual annotation of five additional human chromosomes (1, 16, 18, 19 and Y) plus an extra Chr 6 MHC region haplotype (QBL). ... [more](#)

Update to Vega Zebrafish
25th Apr 2005
Updated annotation and assembly for Vega Zebrafish, plus code updates ... [more](#)

Feedback
We are keen to receive extra information on annotation, and to hear your comments, problems, and suggestions on Vega. Please send us [feedback](#).

**Zebrafish
Vega**

Follow the link to the zebrafish database.

Zebrafish Genome Annotation

This site allows browsing of zebrafish BAC and PAC clones that were sequenced within the zebrafish clone mapping and sequencing project. This is based on [tiling path information](#) from the [fpc database](#) (datafreeze 15.02.05).

The genome of the Tuebingen strain is currently displayed in chromosomes/linkages groups 1-25, plus one artificial chromosome, U, that contains all the clones with unknown chromosomal locations.

The AB 'chromosome' contains PAC clones from the AB strain, sorted out to avoid problems arising from high variations between the AB and the Tuebingen strain.

The manual annotation of the clones is ongoing - clones that haven't been annotated yet are displayed with all their features derived from automated computational analysis (repeat masking, ab initio gene predictions, Blast searches etc.) but are shaded in grey to avoid confusion with the annotated ones.

Acknowledgements

Zebrafish manual annotation is a collaboration between the Zebrafish sequence analysis group and the Havana Group at the Sanger Institute. Full acknowledgements are given [here](#).

For more details please see the Zebrafish Genome Project [homepage](#) and a Sanger Institute [press release](#). Data sets provided by external collaborators are acknowledged [here](#).

Please see the [conditions of use](#) for these data.

Zebrafish Sitemap

Click on the sitemap to explore Zebrafish Vega:

Select a Chromosomal Tile Path to Browse

Data Entry Points

Some example data points:

- ▶ Chromosome: Chromosome 18
- ▶ Gene: [hif1a](#)
- ▶ Transcript: [CH211-212G7.4-00](#)
- ▶ Exon: [CH211-212G7.4-00](#)
- ▶ Peptide: [CH211-212G7.4-00](#)
- ▶ Contig: [BX255914.3](#)

Data

Last update:	25-04-2005
Annotated loci:	3546
Annotated gene transcripts:	4457
Annotated gene exons:	29820
Clones:	4544
Annotated clones:	1108
Base Pairs:	675120918
Annotated Base Pairs:	143870195

'chromosomes' AB & U

There are two 'artificial' chromosomes 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 of the chromosomes do not represent an estimation of the size of the real chromosomes, but the amount of finished sequence. Chromosome 20 is top priority and that is reflected by the fact that it is the longest.

MapView and ContigView

Clicking on a chromosome links to the corresponding MapView page.

Unannotated regions

Chromosome 20	
Length:	44,535,767 bps
Annotated sequence length:	41,690,904 bps
Total number of clones:	348
Number of annotated clones:	332
Known genes:	120
Known genes (in progress):	2
Novel CDS:	699
Novel CDS (in progress):	6
Novel transcripts:	54
Total pseudogenes:	3
Processed pseudogenes:	1
Unprocessed pseudogenes:	2
Putative:	17

Definitions of indices
Shaded regions have not been manually annotated.

Change Chromosome

Chromosome:

Jump to Contigview

Click anywhere on the chromosome ideogram or one of the feature distribution plots to jump to a contig-level view of features at that point. Alternatively, you can jump to contigview between any two markers on this chromosome:

Between:

and:

Display contig-level view between any two features.

Map your data

Map your own data using KaryoView.

empowered Help Desk / Suggestions

The regions shaded in grey in the **MapView** pages indicate regions 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.

The **ContigView** is, like 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 16509110 to 16593534.

Vega Zebrafish ContigView
 The Wellcome Trust Sanger Institute

Home Dog Human Mouse **Zebrafish** BLAST Export Data Search Feedback Help

Help on ContigView Find All [e.g. AL590146.2, BX842684]

Chromosome 20

Chr. 20

Overview

DNA(contigs)
 Markers
 Zfish Genes
 Gene Legend

Detailed view

Jump to region: 20 bp 16509110 to 16593534

<< 2 Mb | 1 Mb | Window | Zoom | 1 Mb | 2 Mb >>

Features DAS Sources Repeats Decorations Export Jump to Image size Help

Length
 EMBL mRNAs
 Ensembl clones
 DNA(contigs)
 PolyA site
 PolyA signal
 Zfish trans.
 GenScan
 Fgenesh
 UniProt
 EMBL mRNAs
 Markers
 CpG islands
 Length
 Gene Legend

Basepair view

Length
 Ensembl clones
 Amino acids
 Sequence
 DNA(contigs)
 Sequence
 Amino acids
 Zfish trans.
 GenScan
 Fgenesh
 Restr. Enzymes
 Length
 Gene Legend

empowered Help Desk / Suggestions

Overview

Poly-A features

Manually annotated genes

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 DKEY-5P1.1-002.

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	OTTDARG0000005397
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 -czfish-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 1: DKEY-5P1.1-001 (OTTDART0000005844) [Transcript information] [Exon information & supporting evidence] [Protein information] 2: DKEY-5P1.1-002 (OTTDART0000005845) [Transcript information] [Exon information & supporting evidence] [Protein information]
Curated Transcripts	

Transcript/Translation Summary

DKEY-5P1.1-002	Stable ID: OTTDART0000005845 Versions: 1 Class: Coding Exons: 25 Transcript length: 5324 bp Translation length: 1216 residues [Transcript information] [Exon information & supporting evidence] [Protein information]
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/Serrate/lag-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] IPR006209 EGF-like domain - [View other Vega genes with this domain]
Transcript Structure	
Protein Features	
DKEY-5P1.1-001	Stable ID: OTTDART0000005844 Versions: 1 Class: Coding Exons: 26 Transcript length: 5438 bp Translation length: 1254 residues [Transcript information] [Exon information & supporting evidence] [Protein information]
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/Serrate/lag-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] IPR006209 EGF-like domain - [View other Vega genes with this domain]
Transcript Structure	
Protein Features	

e-powered Help Desk / Suggestions

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.

Transcript	DKEY-SP1.1-001 (Vega_transcript ID)
Vega Transcript ID	OTTDART0000005844
Version	1
Class	Coding [Definition]
Transcript Information	Exons: 25 Transcript length: 4150bp Translation length: 1254 residues This transcript is a product of gene: OTTDARG0000005397 [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] IPR001861 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/Serrate/lag-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] IPR006209 EGF-like domain - [View other Vega genes with this domain]
Export Data	Export transcript data in EMBL, GenBank or FASTA

Exon information link

Transcript cDNA Sequence
No markup [v] No numbers [v]

Transcript Structure
Graph showing exons and introns with a scale of 0 to 41.42 kb.

Transcript Neighbourhood
Genomic map showing the transcript location on chromosome 20 with a scale of 16.50 Mb to 16.55 Mb.

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 The Wellcome Trust Sanger Institute

Home Dog Human Mouse Zebrafish BLAST Export Data Search Feedback Help

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

Vega Exon Report

Transcript	DKEY-5P1.1-001 (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

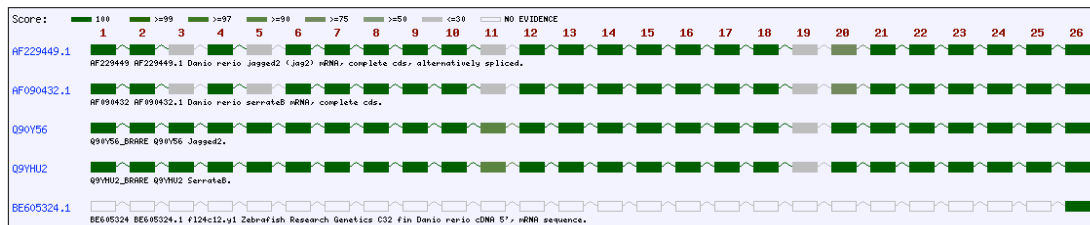
Exon Information

No.	Exon / Intron	Chr	Strand	Start	End	Length	Sequence
	5' upstream sequence					gagatctcctcaggcgtgtggtagccgtgaacatatacattgocgtgaa
1	OTTDARE00000033254	20	-1	16593319	16593534	216 bp	GTGATCAGACCGAGGAGAGATCAGCACAGACCATCACCGCGCAACACACCACGCTCGT GAATTTTGCATGTCAGGAACGGAGATCCTGTCGCGGTCATCGCCGCTTTCCATCTT CGCTTTAATACACATCAAAATCGCGGCATGTGGAATTGTATCAGGATTAGGAATGGCTC CCAATCGCCTGCCTGCTTTAACGATGTGACGAAC gtgagagctctgattgttattagtg.....ttctgtttttcaatgtatttttag
	Intron 1-2	20	-1	16592862	16593318	457 bp	
2	OTTDARE00000033260	20	-1	16592541	16592861	321 bp	GTGTCACGACTCCTCTGGCTATTTGAGCTGCAGCTGATTGCTGTAGAAAAATGTAACGGT GAGTTGTGGGACGGGAATGTTGGCAGCAGCGGGAACCTCAAGACCAGCGCTGCGTG CGGACGAGTCCGATACCTTTAAAGTGTCTGAAGGATACCACTGGAAGTCAAC ACCACTGGACAGTGCACCTTCGGCTCTGGATCTACCGACGTTCTGGTGGAAATATAAT TCTTTTAAGACCGCAAAAAACAGCCCAACAAACGAGCGAGCTGGGAAAGATCATCATC CCTTTTCACTTCGCCTGGCCG gtgagtcctcgtctctccatgggc.....gttattctctctctctctctctattag
	Intron 2-3	20	-1	16565147	16592540	2794 bp	
3	OTTDARE00000033271	20	-1	16565092	16565146	55 bp	CGATCTACACACTCATCTTGAAGCTGGGACTGGGATTAACCTCCACTCAGAACA

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 module 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. Open the ContigView page in Vega for clone AL772329. The current version of this clone in Vega is 14, but in Zv4 this clone appears with version 13. You can quickly check this information by turning on the track 'Ensembl clones'.