

Module 1: Where do I start?

i - The Zebrafish Genome Project

Aims

- Introduce zebrafish genome project web pages
- Give examples of the services provided
- Show ways to navigate around these pages

Introduction

In spring 2001 the Wellcome Trust Sanger Institute started sequencing the genome of the zebrafish (*Danio rerio*). The strategy being used is clone mapping and sequencing from BAC and PAC libraries complemented with a whole genome shotgun (WGS) assembly. Some of these clones are selected for sequencing based on their location in the tiling path of the physical map. The released zebrafish assembly is based on the integration of the available finished clones with the WGS assembly contigs. The assembly is automatically annotated using the Ensembl pipeline and can be browsed on the Ensembl site. Assemblies are released once or twice a year depending on the available data. The current assembly is Zv6, which was released on March 31st, 2006. The assembly will eventually consist solely of finished clones, with no sequence from the WGS assembly.

Sequences from finished clones come through the sequencing pipeline on a daily basis. They currently cover around 80% (March 2006) of the estimated 1.6 Gb size of the zebrafish genome. In a collaboration with ZFIN, finished clones are manually annotated. The finished clones with manual annotation can be browsed in the Vega database. These data are updated regularly to reflect the changes in the physical map and to make public the annotation. In sections 2 and 3 the structure of the data in Ensembl and Vega is discussed in more detail.

The *Danio rerio* Sequencing Project Page

The main gateway to all the information regarding the zebrafish genome project is:

http://www.sanger.ac.uk/Projects/D_rerio

On the left-hand side of this page there is a quick-access toolbar with links to the services offered, and on the right-hand side there is a report with the jrecent news related to the project. The page is divided in five parts:

- FAQs and contact information
- clone mapping and sequencing
- assembly releases
- other services
- contacts and links

The screenshot shows the Sanger Institute website for the *Danio rerio* Sequencing Project. Several callout boxes highlight specific features:

- Quick-access toolbar**: Points to the top navigation bar containing links for Information, Projects, and Other Services.
- FAQs and contact**: Points to the 'Frequently asked questions' and 'Contact us' links in the left sidebar.
- news**: Points to the 'Vega release' news item on the right side of the page.
- clone mapping, sequencing and manual annotation**: Points to the 'Clone mapping and sequencing' section, which includes links to mapping and clone related pages.
- assembly releases**: Points to the 'Assembly releases' section, which includes links to WGS and assembly related pages.

The main content area displays the 'Current Sequencing Status' table:

Date	Unfinished	Finished	Total
09-May-2006	654,705,316	1,081,452,147	1,736,157,463

The table also lists 'Assembly releases' and 'Other services' such as online RepeatMasker, Genome Resources Workshop Dresden 2005, and Sanger part of Madison 2004 tutorial.

Contacting us

The email address for any enquiry regarding the project is:

zfish-help@sanger.ac.uk

There is also a link to FAQs page where a wide range of questions regarding the project are already answered.

Clone mapping and sequencing

This page lists all the relevant links to the zebrafish clones, from their mapping to the sequence. There are links to the Vega database, the FPC database and to a Blast server for searching all the available sequences from the project.

Assembly releases

This page has information about the current and previous assemblies with links to FTP sites from which the sequences can be downloaded. There is also a link to the trace repository. This is a database that features traces from several projects including all the zebrafish reads used in the whole genome shotgun assembly. These databases can be searched for alignments using SSAHA .

Other services

This section has links to an online RepeatMasker server and tutorials used in several courses and workshops . The repeat analysis is based on the same Repbase database used in the Vega/Ensembl analysis.

The output of this service returns the original sequence where repeats are masked by strings of Ns.

The screenshot shows the Zebrafish RepeatMasker Server web interface. A yellow box labeled "paste sequence or..." points to a large text area under the "SEQUENCE DATA" header. Another yellow box labeled "enter filename" points to a smaller text area below the "SEQUENCE DATA" header. The interface includes a sidebar with navigation links, a description of the service, and a "RESULTS" section at the bottom. A "Retrieve BLAST result" button is also visible.

paste sequence or...

enter filename

SEQUENCE DATA

OR upload a sequence file you wish to mask
OR enter a zebrafish clone accession you wish to mask
You may also specify a pipe ("|") separated list of accession's

RESULTS

Mask Reset

Reference for RepeatMasker: A.F.A. Smit & P. Green, unpublished data. RepeatMasker is written and supported by [Arian Smit](#). Queries regarding this RepeatMasker webserver should be sent to webmaster@sanger.ac.uk.

Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK Tel: +44 (0)1223 834244 webmaster@sanger.ac.uk
Registered charity number 210183 [Data Release Policy](#) [Conditions of Use](#) [Copyright](#)
Last Modified Tuesday, 19-Oct-2004 15:03:58 BST