739A: Zebrafish Genome Resources at NCBI

Nuala O'Leary; Terence Murphy ; Francoise Thibaud-Nissen; Valerie Schneider; Eukaryotic Genome Annotation Group; RefSeq Curation Group

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

U.S. National Library of Medicine
National Center for Biotechnology Information

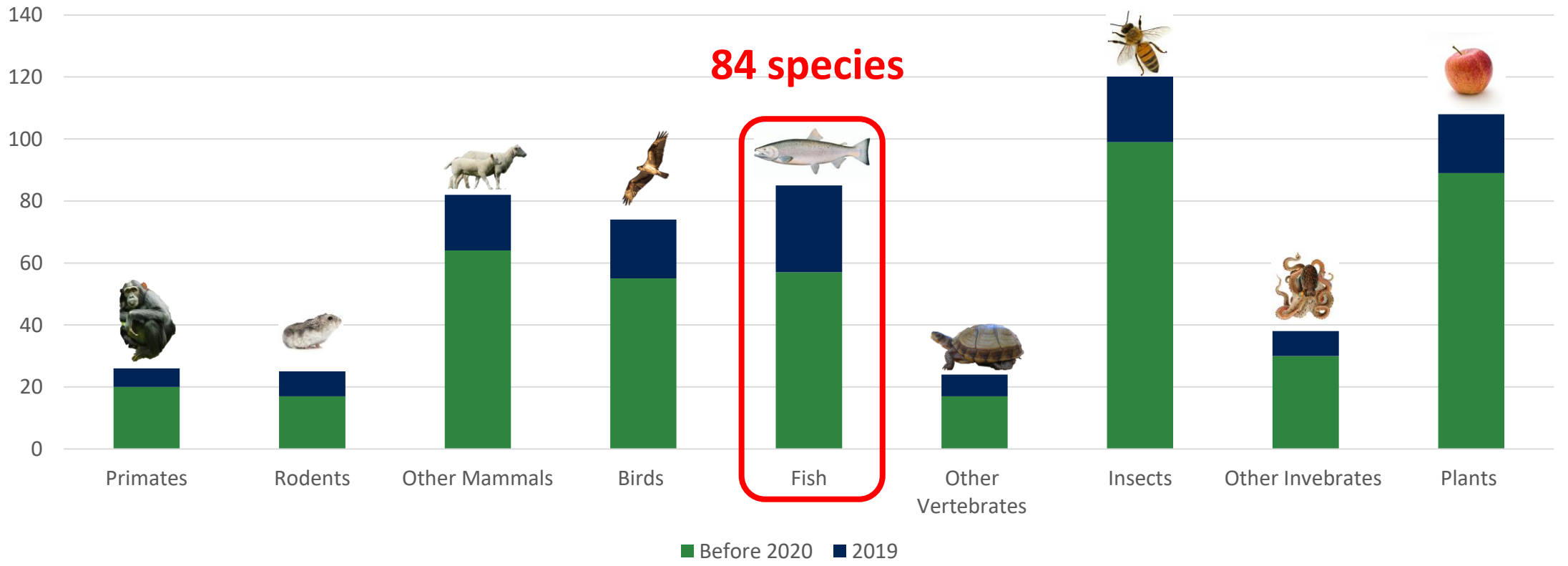
Abstract

Genomic research on model organisms such as zebrafish (Danio rerio) relies on the availability of high-quality annotated reference genomes to facilitate consistent reporting and mapping of genetic data. The Reference Sequence (RefSeq) project (https://www.ncbi.nlm.nih.gov/refseq/) at the NCBI provides a comprehensive annotation of the zebrafish Tuebingen strain reference genome assembly (GRCz11) maintained by the Genome Reference Consortium (GRC). The zebrafish RefSeq dataset is generated by a combination of computational analysis and manual curation that results in an annotation that focuses on representation of all full-length, non-redundant transcripts.

The primary sources of data used in this annotation pipeline include mRNAs, expressed sequence tags (ESTs), protein data, RNA-seq datasets, and protein homology. Zebrafish is one of a select group of vertebrates that are the major focus of RefSeq's manual curation efforts, which involves the in-depth review of sequence data to define new transcript variants, resolve sequence errors, and remove inaccurate information. We also collaborate with expert groups, including the Zebrafish Information Network (ZFIN) and UniProtKB, to provide appropriate annotation and nomenclature for both genes and proteins.

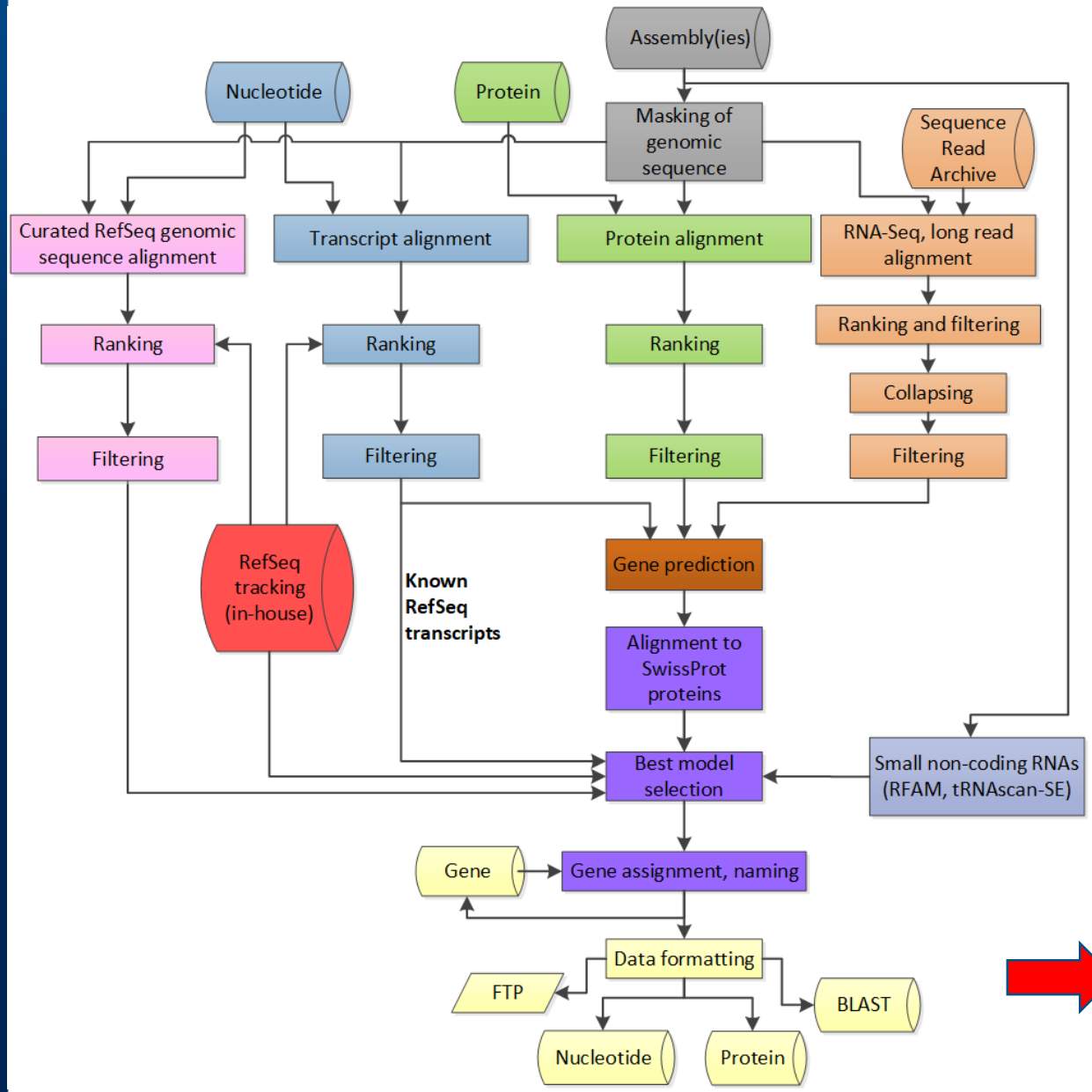
In addition to zebrafish, NCBI provides stable reference genome annotation for other fish species with high-quality genome assembly data submitted to NCBI's Assembly resource (https://www.ncbi.nlm.nih.gov/assembly/). To date, 80 fish species have RefSeq annotated genomes, providing a valuable resource for comparative genomic research. We have recently expanded our gene ortholog dataset for fish along with a new visualization tool for comparing homologous gene sets. In this poster presentation we will provide an overview of NCBI's zebrafish genome resources and highlight the utility of these resources to the zebrafish research community. We will also provide practical guidance on how to access RefSeq data and tools for analysis of individual genes as well as whole genome datasets.

Genomes annotated by NCBI



Eukaryotic Genome Annotation Pipeline

- Automated
- Evidence-based
- RNA-seq
- Long reads
- Proteins



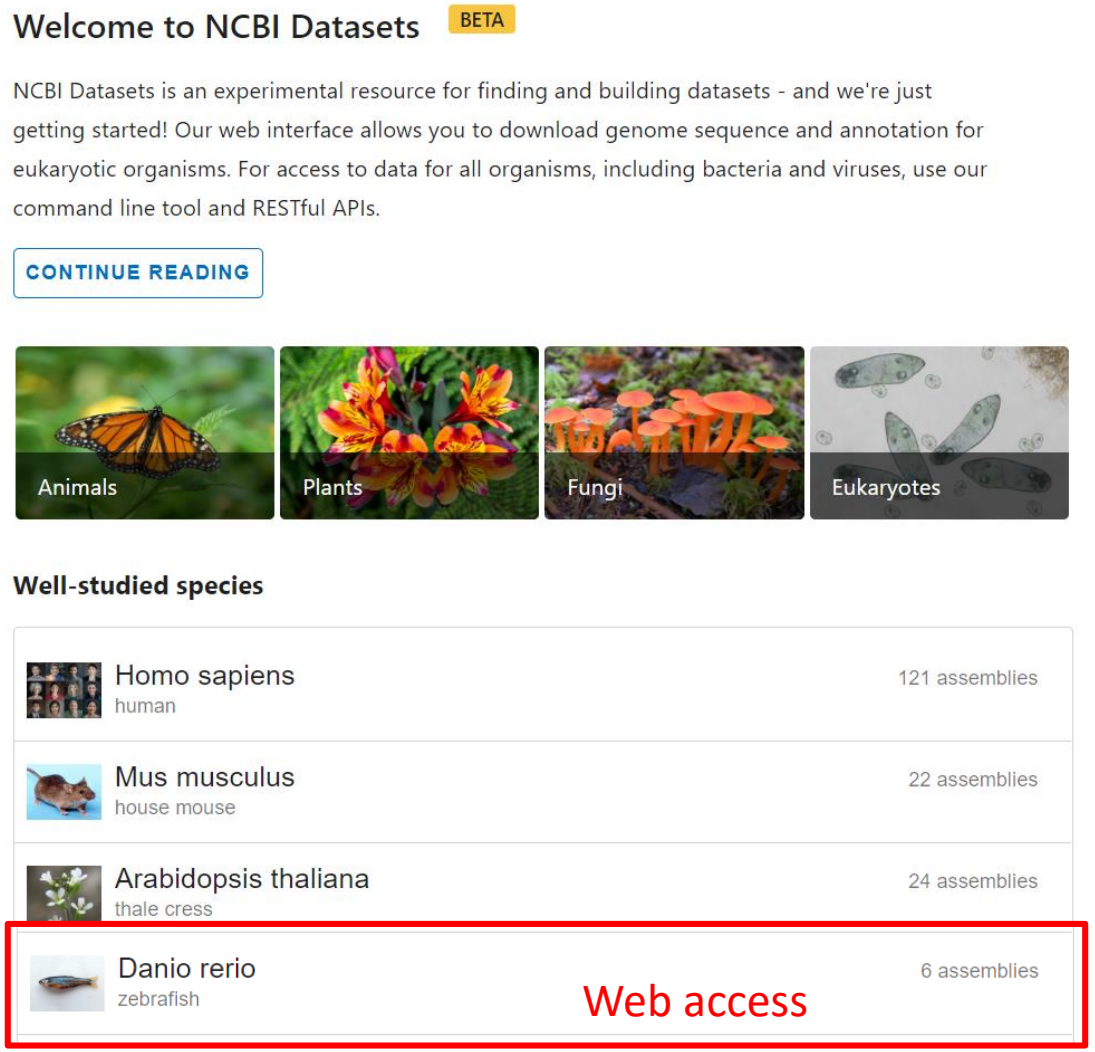
NEW

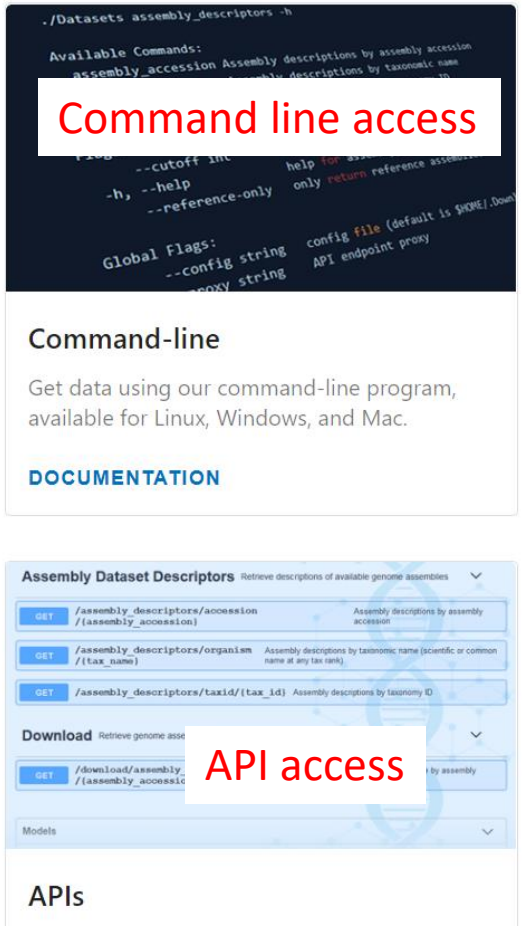
Access Genome sequence and annotation through **NCBI Datasets**

https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/

Introducing NCBI Datasets

A new way to access genome sequence and annotation at NCBI



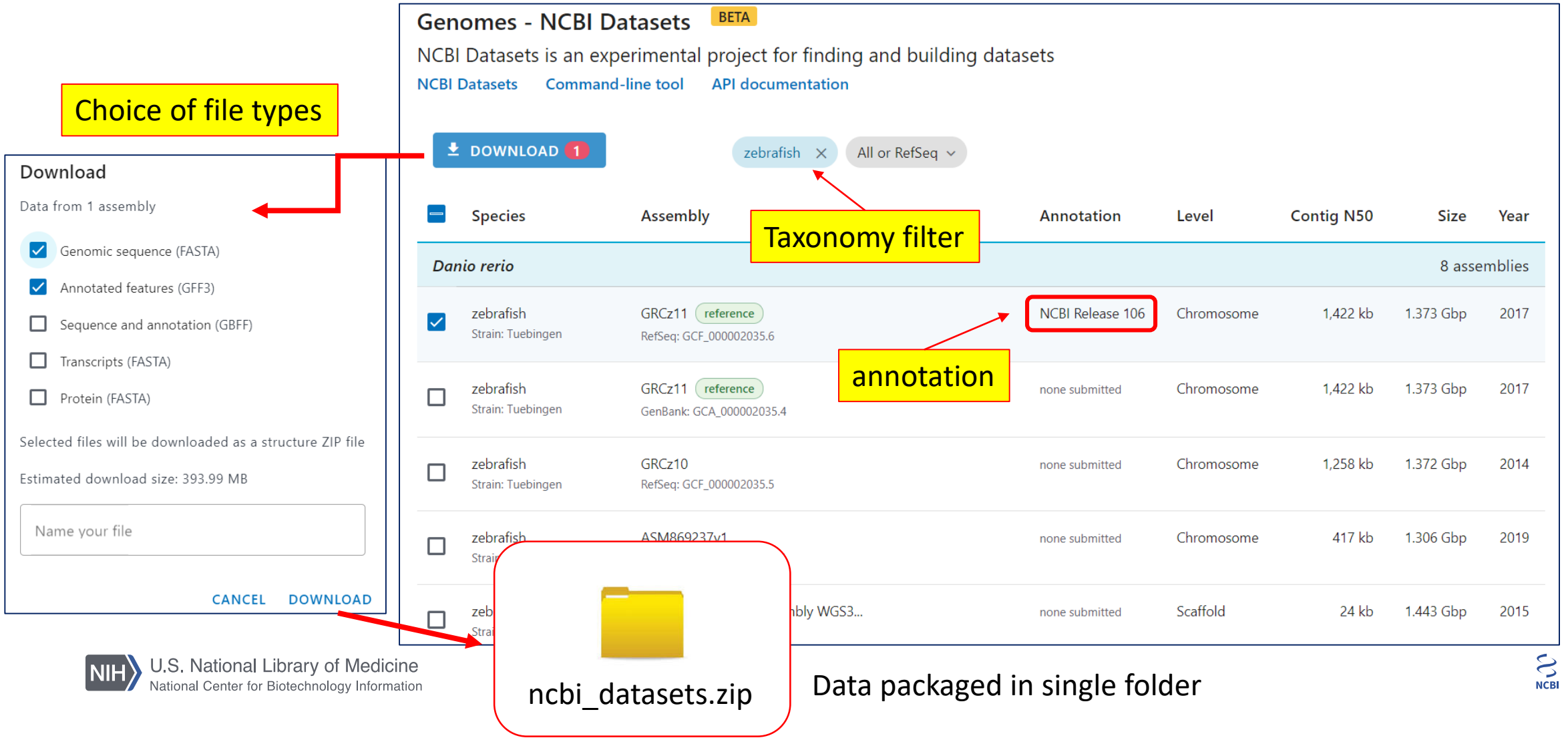


www.ncbi.nlm.nih.gov/datasets

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

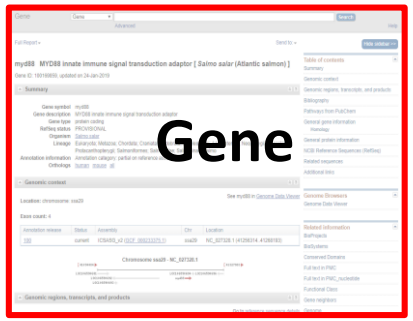
NCBI

Easy download of genome sequence and annotation



Data packaged in single folder

Finding genes



Search NCBI | zebrafish dmrt3a | Search

GENE

dmrt3a – doublesex and mab-3 related transcription factor 3a

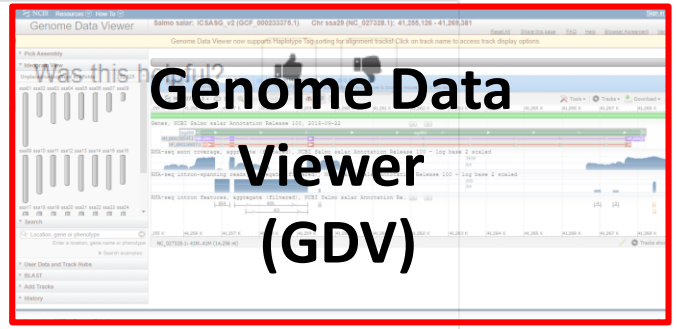
Danio rerio (zebrafish)

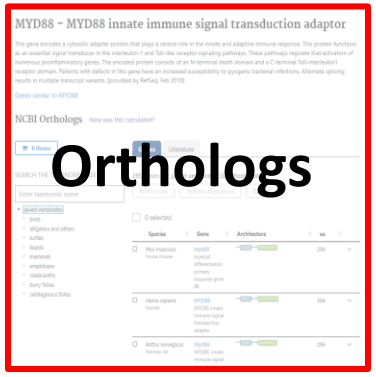
Also known as: dmrt3a1, zDmart3a, zgc:101766

GeneID: 450035

RefSeq transcripts (1) RefSeq proteins (1) PubMed (14)

Orthologs | Genome Browser | BLAST | Download

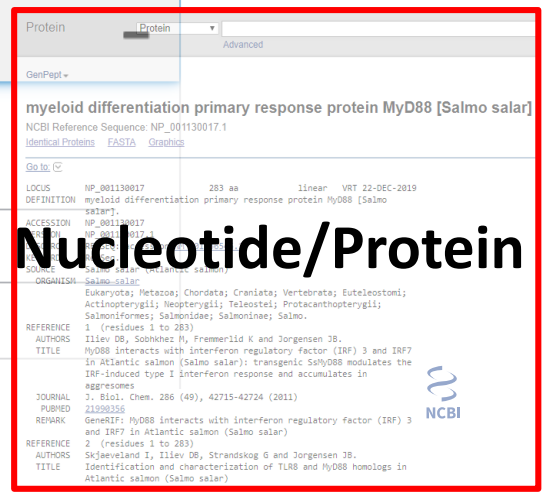




RefSeq Sequences

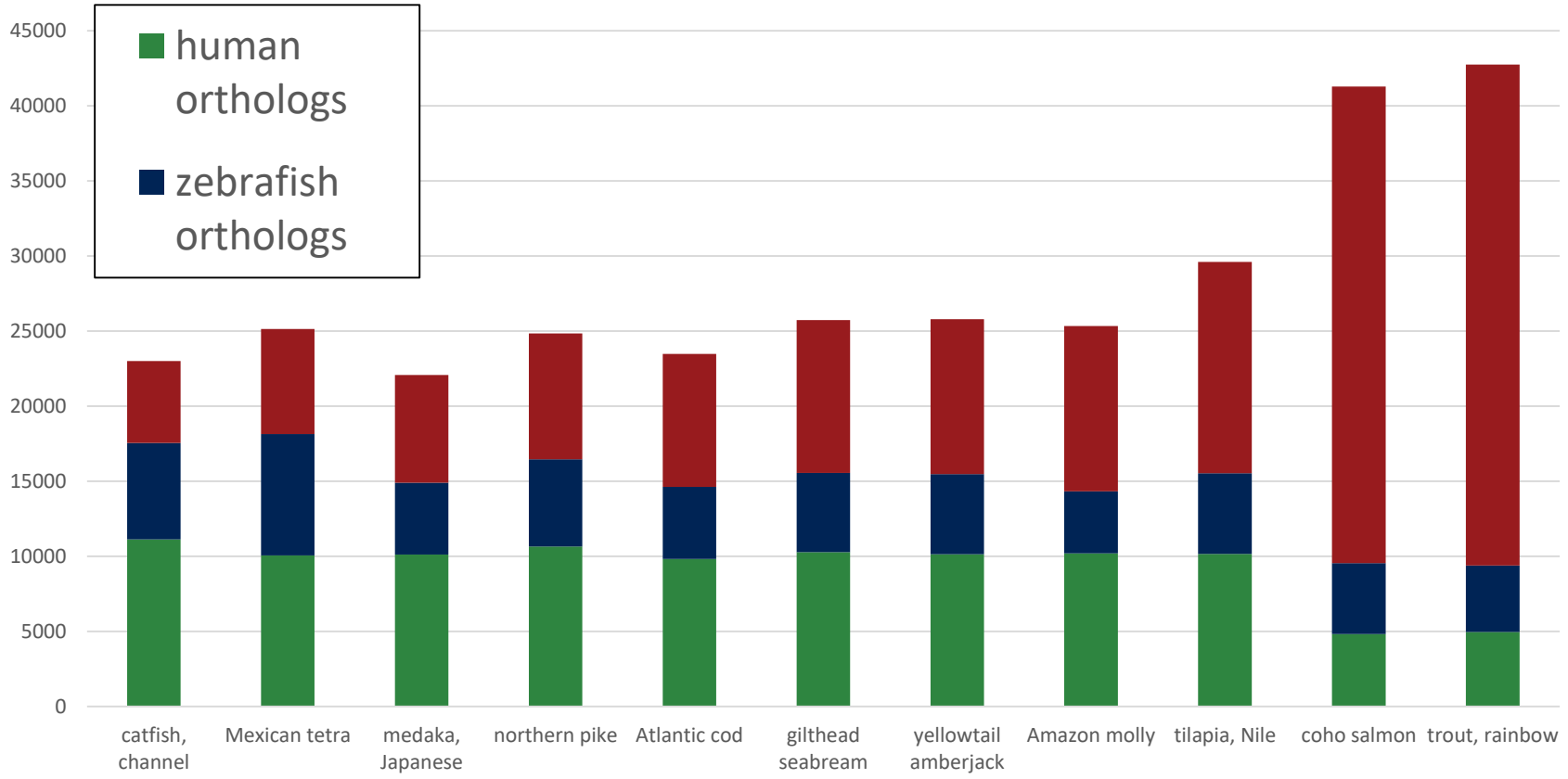
Showing 1 of 1 (by status, accession number)

Transcript	nt	Protein	aa	Isoform	Status
NM_001005779.2	1,920	NP_001005779.2	448		curated



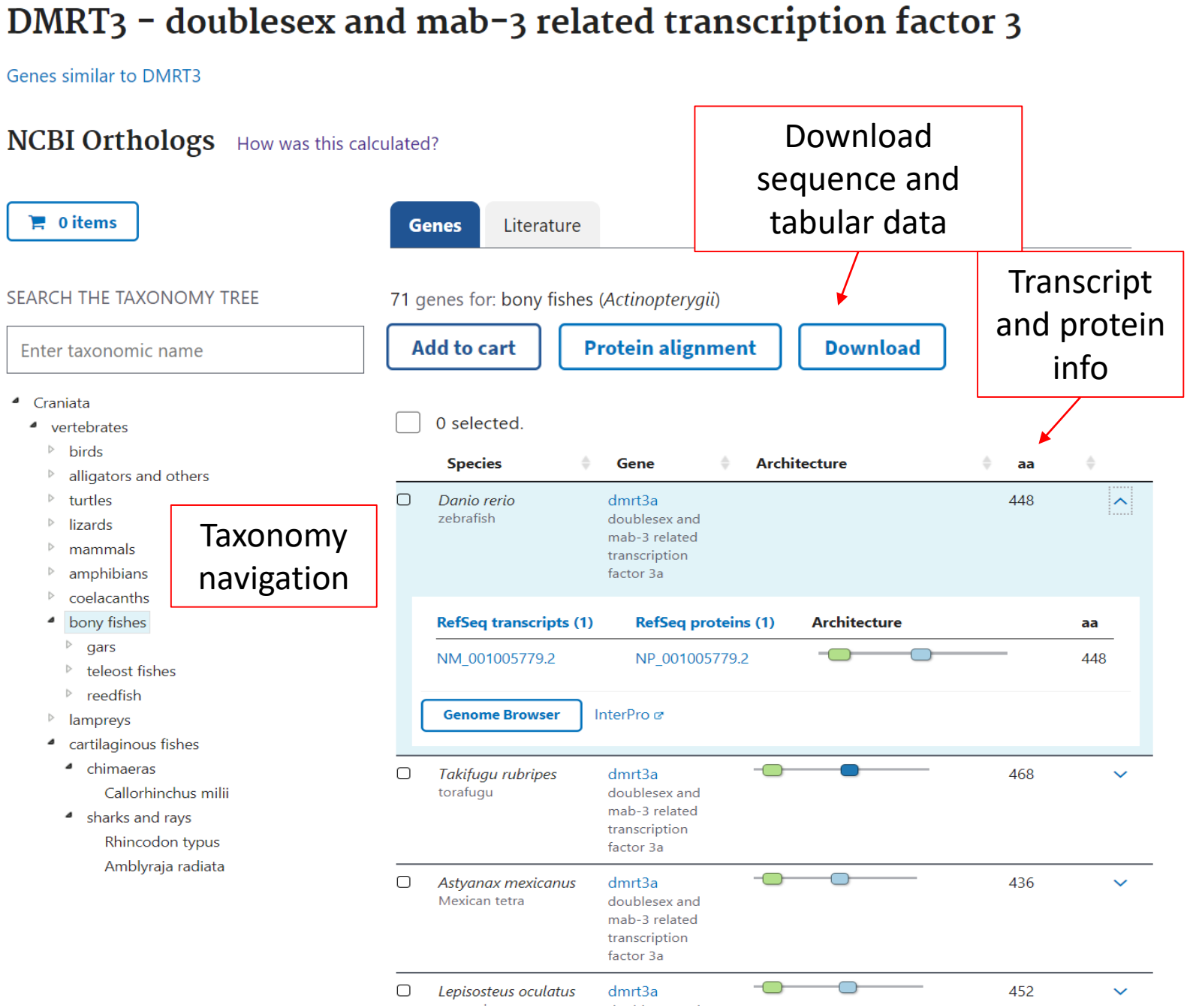
New! Fish ortholog sets

- 1:1 orthologs vs zebrafish
- Report human gene if identified
- Improve nomenclature through ortholog connections



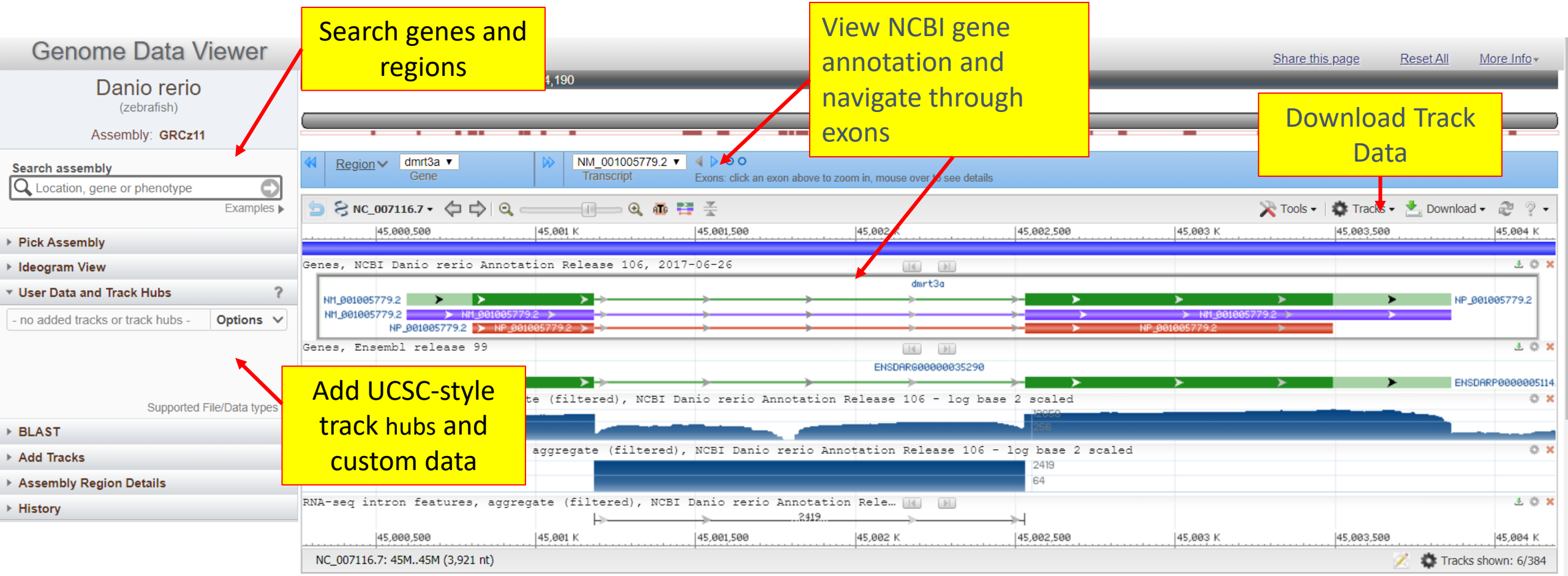
NCBI Orthologs

www.ncbi.nlm.nih.gov/gene/58524/ortholog/?scope=7898



U.S. National Library of Medicine
National Center for Biotechnology Information

Genome Data Viewer (GDV)



Acknowledgements

- This work was done as part of the Authors' official duties as NIH employees and is a Work of the United States Government. Therefore, copyright may not be established in the United States. 17 U.S.C. § 105

Send Feedback

https://www.ncbi.nlm.nih.gov/projects/RefSeq/update.cgi

Watch NCBI News for updates!

https://ncbiinsights.ncbi.nlm.nih.gov/

https://www.youtube.com/user/NCBINLM