

GIGA Members, Be Counted:

Register with GIGA-NCBI BioProject and utilize the GIGA-GoaT Resources.

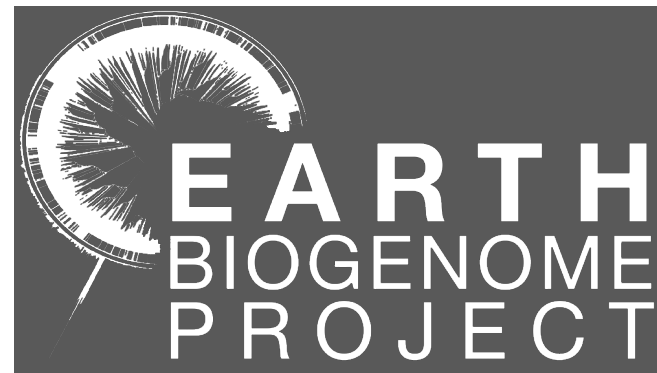
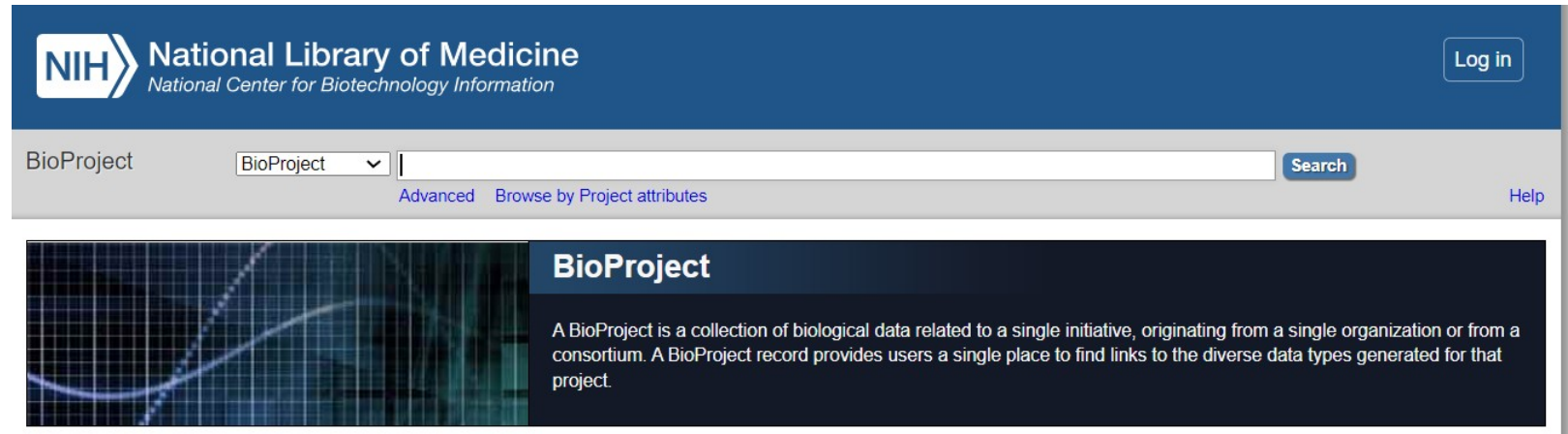


Jeffrey Robinson, UMBC



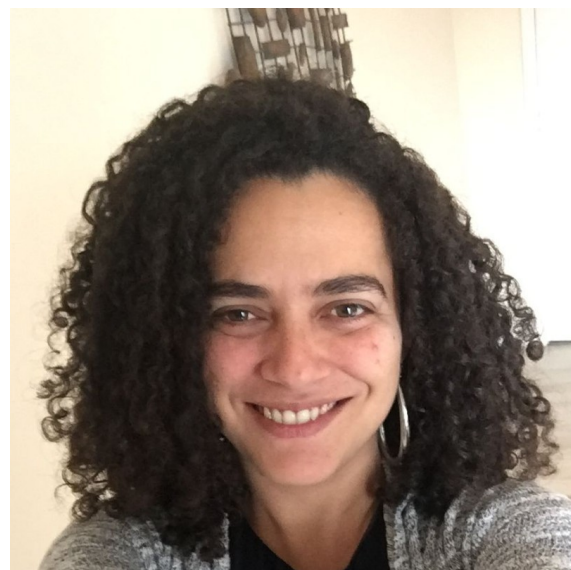
Cibele Sotero-Caio, Wellcome Sanger Institute

Fabrizio Ghiselli, University of Bologna



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Workshop Collaborators



Dr. Cibeles Sotero-Caio

Genomic Data Curator,
GoaT collaborator
Wellcome-Sanger Institute
United Kingdom

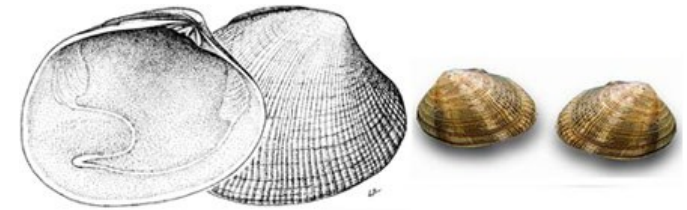


Dr. Fabrizio Ghiselli

Associate Professor
GIGA collaborator
Universita di Bologna
Italy



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Ruditapes philippinarum
(Adams & Reeve, 1850)
[Veneridae]

Querying genome metadata and sequencing projects using GoaT.

<https://www.youtube.com/watch?v=n-1chlrFi5fQ&t=6s>

Agenda Outline

I. OMICS Metadata Workflow Resources for GIGA members

- **GIGA NCBI BioProject**
- **Genomes on a Tree (GoaT) Resources**

II. Why GIGA members should link their BioProjects

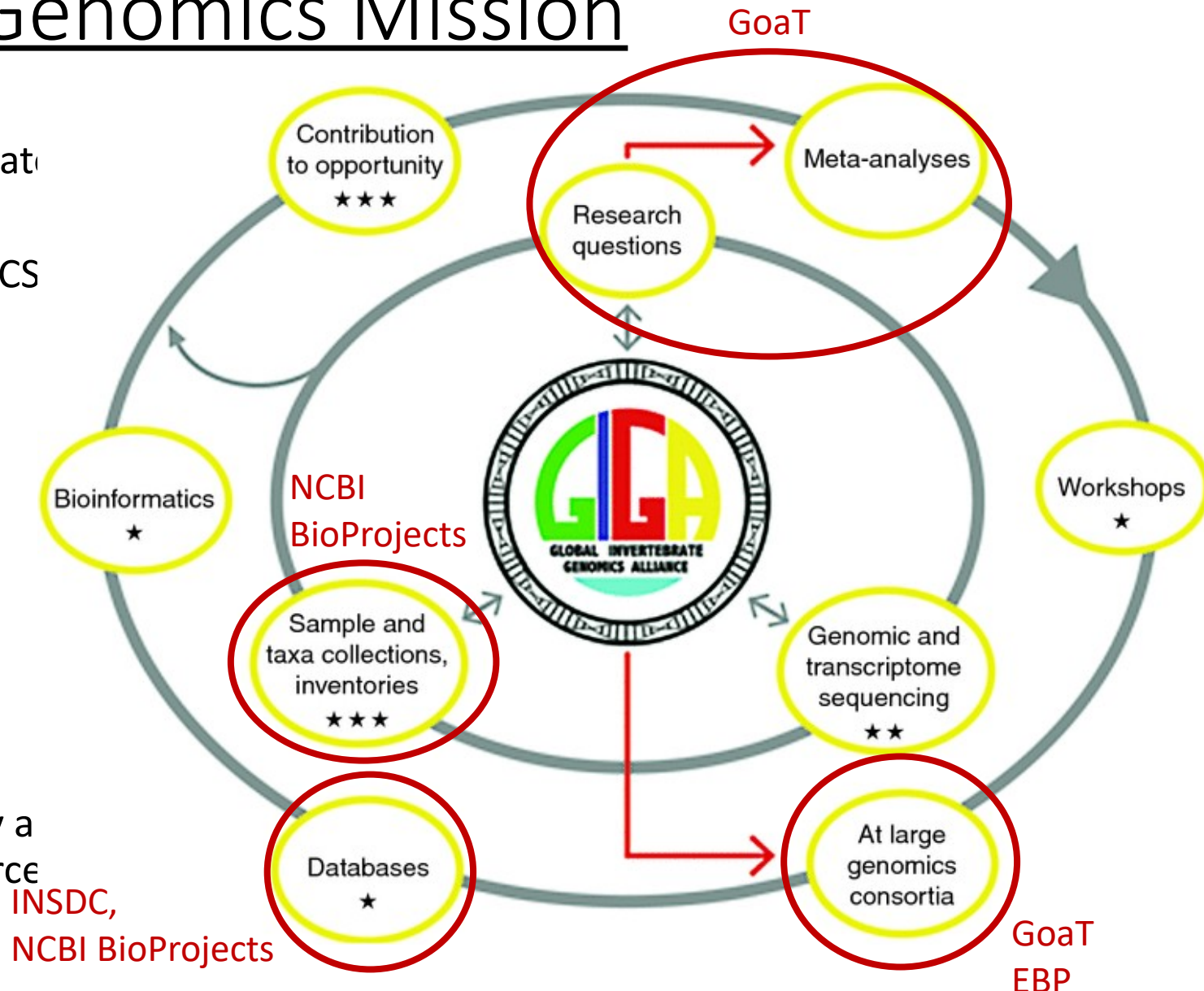
- **Contribute to the official GIGA content collection.**
- **Track progress of genome sequencing efforts for planning and collaborations using GoaT.**

III. GIGA-GoaT Use Case Example for GIGA Members

- **Fabrizio Ghiselli – *Ruditapes* genomics. Example of a GIGA BioProject linked to the GIGA Umbrella BioProject.**
- **Find species-associated OMICS projects at different taxonomic levels.**
- **Further Directions – Organize GIGA Specialty Groups based on research questions. Organize collaborative support for the GIGA-GoaT efforts.**

GIGA's Genomics Mission


- In Voolstra et al. 2017, GIGA organization was envisioned to function as a node for invertebrate OMICS research.
- The ability to perform the “Backend” of –OMICS research.
- Sample strategy and sequencing is just the beginning. What happens to BioSample and Sequence data after it has been collected.
- How does the data contribute to addressing research questions, and to framing Meta-analyses.
- Genomics requires a robust bioinformatics infrastructure.
- Databases and Bioinformatics necessarily play a major role. NCBI BioProjects and GoaT resource provide a system for this effort.



Advancing Genomics through the Global Invertebrate Genomics Alliance (GIGA). Voolstra CR et al. 2017. *Invertebr Syst.* 2017;31(1):1-7. doi:10.1071/is16059

GIGA's NCBI BioProject is an Umbrella for individual OMICS projects

- GIGA's NCBI BioProject page is a place where all GIGA contributors can link their BioProjects and data.
- Having a linked BioProject allows your OMICS data contribute to the goals of the larger genomics consortia, as well.
- Individual BioProjects can also be linked into the EBP umbrella.

**National Library of Medicine**
National Center for Biotechnology Information

BioProject

[Advanced](#) [Browse by Project attributes](#)

Display Settings: ▾ Send to: ▾

Aquatic non-vertebrate metazoa Accession: PRJNA649812 ID: 649812

The Global Invertebrate Genomics Alliance (GIGA) genomes and transcriptomes

GIGA is a collaborative network of diverse scientists to tackle major challenges (e.g., species selection, sample collection and storage, sequence assembly, annotation, analytical tools) associated with genome/transcriptome sequencing across a large taxonomic spectrum. It aims to promote standards that will facilitate comparative approaches to invertebrate genomics and collaborations across the international scientific community.

Accession	PRJNA649812
Type	Umbrella project
Publications	<ol style="list-style-type: none">1. Published online: "The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrate Genomes", <i>Journal of Heredity</i>, 2013;105(1):1-182. Published online: Voolstra C <i>et al.</i>, "Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA)", <i>Invertebrate Systematics</i>, 2017;31(1)
Submission	Registration date: 30-Jul-2020 Global Invertebrate Genomics Alliance - None
Related Resources	<ul style="list-style-type: none">• Earth Biogenome Project• Global Invertebrate Genomics Alliance
Relevance	Evolution

NAVIGATE UP

This project is a component of the Earth BioGenome Project (EBP)

NAVIGATE ACROSS

34 additional projects are components of the Earth BioGenome Project (EBP).


Related information

- [Assembly](#)
- [BioProject](#)
- [BioSample](#)
- [Data projects](#)
- [Full text in PMC](#)
- [PubMed](#)
- [SRA](#)
- [Umbrella projects](#)

Related Resources

- [Earth Biogenome Project](#)
- [Global Invertebrate Genomics Alliance](#)

Recent activity

-  [Aquatic non-vertebrate](#)

GIGA BioProject contents as of ~10/15/2023

- The GIGA BioProject links content from various NCBI, including Nucleotide (nucore), SRA, Protein, PubMed, PMC, BioSample, and Assembly databases.
- The GIGA BioProject contains both Genome and Gene Expression/Transcriptome datasets.
- Meta/Environmental genomics, and Epigenomics can also be accommodated.

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	94219
WGS master	6
TSA master	9
Genomic DNA	2
Transcript	88190
SRA Experiments	277
Protein Sequences	85123
PUBLICATIONS	
PubMed	11
PMC	9
OTHER DATASETS	
BioSample	213
Assembly	6

▼ SRA Data Details

Parameter	Value
Data volume, Gbases	2,454
Data volume, Tbytes	0.87

Aquatic non-vertebrate metazoa encompasses the following 11 sub-projects:

Project Type			Number of Projects
Genome sequencing			
<i>Highest level of assembly :</i>			
Chromosomes			1
Scaffolds			1
Contigs			4
SRA or Trace			1
Total			7
BioProject accession	Assembly level	Name	Title
PRJNA639707	SRA or Trace	A naturally heteroplasmic clam shows the effects of genetic bottleneck on paternal mtDNA	A naturally heteroplasmic clam shows the effects of genetic bottleneck on paternal mtDNA (University of Bologna)
PRJNA734913	Contigs	Amphibalanus amphitrite	Amphibalanus amphitrite: Genome Sequencing and assembly (Naval Research Laboratory)
PRJNA751628	Contigs	Amphibalanus amphitrite	Amphibalanus amphitrite RefSeq Genome sequencing and assembly (NCBI)
PRJNA498596	Contigs	Bugula neritina	Bugula neritina isolate:Kwan_BN1 Genome sequencing and assembly (Global Invertebrate Genomics...)
PRJNA983812	Chromosomes	Nematomorpha	Nematomorpha Genome sequencing and assembly (Harvard University)
PRJNA614970	Scaffolds	Pollicipes pollicipes	Pollicipes pollicipes isolate:AB1234 Genome sequencing and assembly (George Washington University)
PRJNA807867	Contigs	Ruditapes philippinarum	Ruditapes philippinarum isolate:M1 Genome sequencing and assembly (University of Bologna)

Focusing on Genomics *per se*, we see that there are relatively few linked BioProjects compared with the number of GIGA’s member labs.

For Gene Expression/Transcriptomics BioProjects, there are even fewer:

Transcriptome or Gene expression		4
BioProject accession	Name	Title
PRJNA663558	Cinachyrella	Effects of crude oil and dispersant exposure from the Deepwater Horizon oil spill on sponges from the Cinachyrella genus (Global Invertebrate Genomic...)
PRJNA672267	RNA-Seq of somatic tissues and gonads of Ruditapes philippinarum	RNA-Seq of somatic tissues and gonads of Ruditapes philippinarum (University of Bologna)
PRJNA170478	Ruditapes decussatus	Comparative transcriptomics in two bivalve species offers different perspectives on the evolution of sex-biased genes (University of Southern California)
PRJNA68513	Ruditapes philippinarum	De Novo assembly of the Manila clam Ruditapes philippinarum transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination (University of Southern California)

NCBI BioProjects Manual and Quickstart Guide.

(<https://www.ncbi.nlm.nih.gov/books/NBK54016/>)

From: NLM Support <nlm-support@nlm.nih.gov>
Date: Thursday, February 2, 2023 at 21:20
To: Fabrizio Ghiselli <fabrizio.ghiselli@unibo.it>
Subject: Re: case #CAS-1060807-Y1V8R0: Join a BioProject TRACKING:000300000010208
Dear Fabrizio Ghiselli,

Thank you for your email. Your BioProjects have been linked to the umbrella BioProject, as you requested. If you have other comments or questions, please reply to bioprojecthelp@ncbi.nlm.nih.gov.

Best regards,
Carol DeWeese Scott, Ph.D.
BioProject Curation Staff

* PLEASE DO NOT MODIFY THE SUBJECT LINE OF THIS EMAIL WHEN RESPONDING TO ENSURE CORRECT TRACKING *

Case Information:
Case #: CAS-1060807-Y1V8R0
Customer Name: Fabrizio Ghiselli
Customer Email: fabrizio.ghiselli@unibo.it
Case Created: 2/2/2023, 10:47:05 AM

Summary: Join a BioProject

Details: Dear Staff,

I would like to link some of the BioProjects I participated as a PI (listed below) to the BioProject PRJNA649812 (Aquatic non-vertebrate metazoa) which include genomes and transcriptomes of interest for the Global Invertebrate Genomics Alliance (GIGA), a non-profit community of scientists under the EBP umbrella. These BioProjects contain data that could be interesting for scientists working on aquatic invertebrates and I think linking them to the GIGA BioProject would help disseminating the data.
The BioProjects are: PRJNA170478, PRJNA68513, PRJNA639707, PRJNA672267, PRJNA807867.

Thank you very much.

Best Wishes,

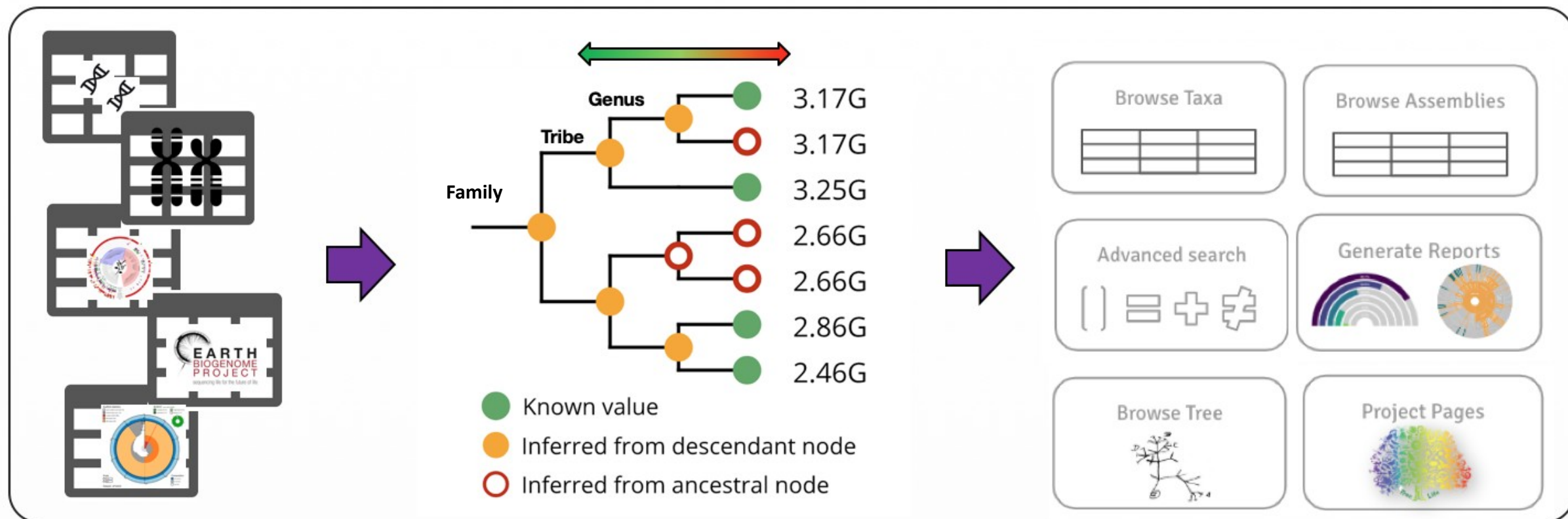
Fabrizio

Your NCBI BioProject can be linked to the GIGA Umbrella BioProject by sending a similar request to bioprojecthelp@ncbi.nlm.nih.gov . (The email would best come from the BioProject's listed PI.)



Genomes on a Tree - GoaT

GoaT is a powerful data aggregator created to provide a centralised source of genome metadata and help the overall coordination, definition, and tracking of sequencing projects.



<https://goat.genomehubs.org/>



Goatpeople +



Developer - Richard Challis

Data management - Sujai Kumar

Data curator - Cibeles Sotero-Caio

Goat-CLI developer - Max Brown

Conceptualization - Mark Blaxter

IT Infrastructure - Matthieu Muffato

IT Infrastructure - Guoying Qi



Contact:

goat@genomehubs.org



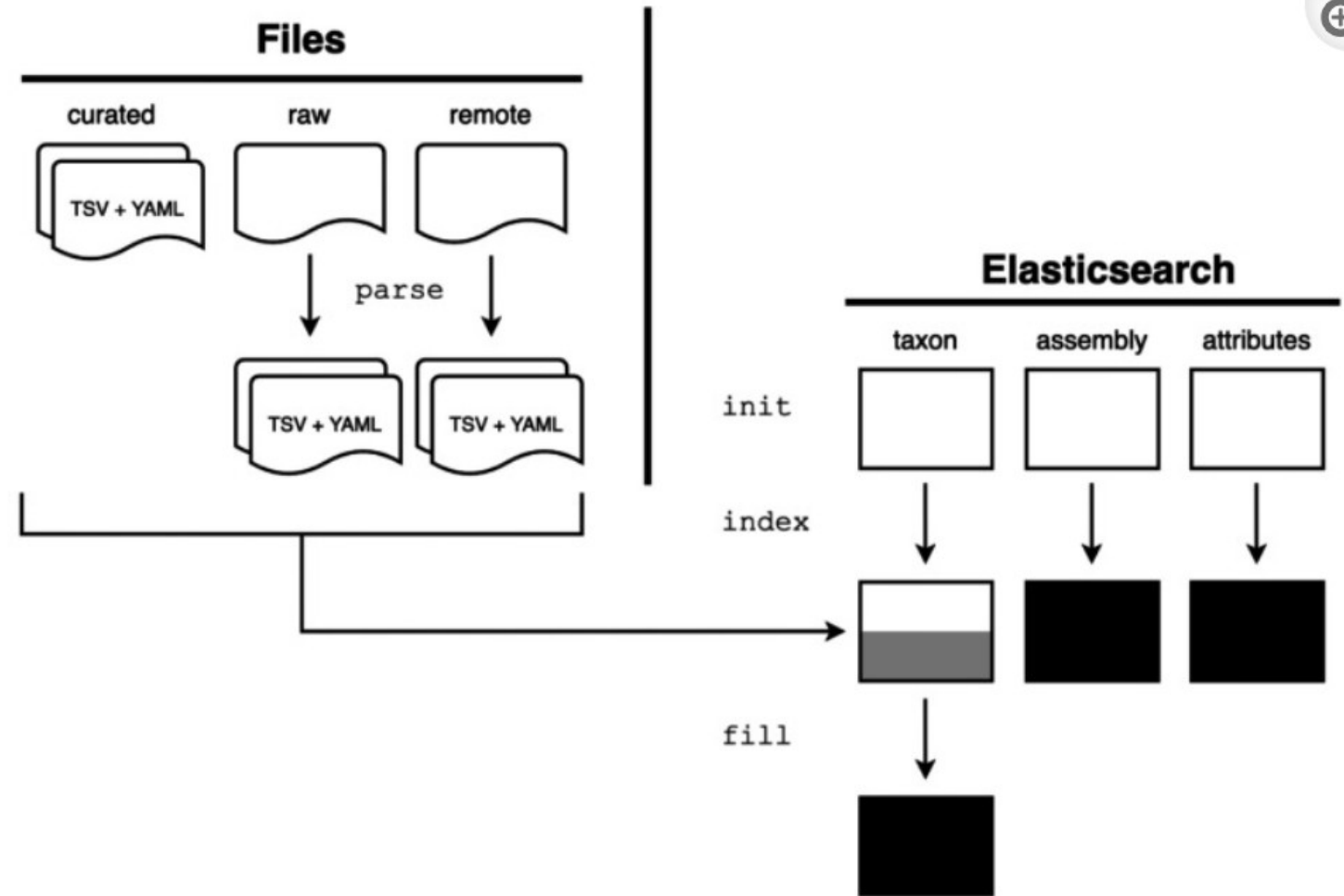
GenomeHubs 2.0 is the GoaT Backend

<https://www.sanger.ac.uk/tool/genomehubs/>

<https://github.com/genomehubs/genomehubs>

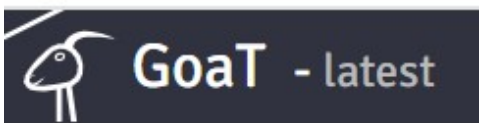


- GoaT is hosted by a backend (GenomeHubs.org) composed from a pipeline of Dockerized applications.
- GenomeHubs updates are described in the goat paper
- Data is imported as .tsv files from remote databases (INSDC, specialized databases and progress lists)
- Configuration files specify what to import from each .tsv for taxon IDs
- This backend framework is Open Source availability, so that cloned of the entire system can be built and customized.



For End-Users, GoaT Provides Subsetting, Filtering, and Dashboarding Functions for Genomics Meta-Data

Goat Homepage <https://goat.genomehubs.org/>



search about help api projects regions submissions sources

Genomes on a Tree (GoaT)

Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. Challis *et al.* 2023. Wellcome Open Res 2023, 8:24 [doi:10.12688/wellcomeopenres.18658.1](https://doi.org/10.12688/wellcomeopenres.18658.1)

GoaT has been built using [GenomeHubs](#) to help coordinate efforts across the [Earth Biogenome Project](#) (EBP) Network at all stages from planning through sequencing and assembly to publication. [read more...](#)

Search GoaT

The GoaT search window is powerful.

It allows subsetting, filtering, retrieval, and visualization of Genomics metadata from any taxon.

Q TAXON ▼

include descendants ☐ Off

include estimates ☒ On

empty columns ☐ Off

result columns

query builder

clear all

[show examples](#) or [browse tree](#)

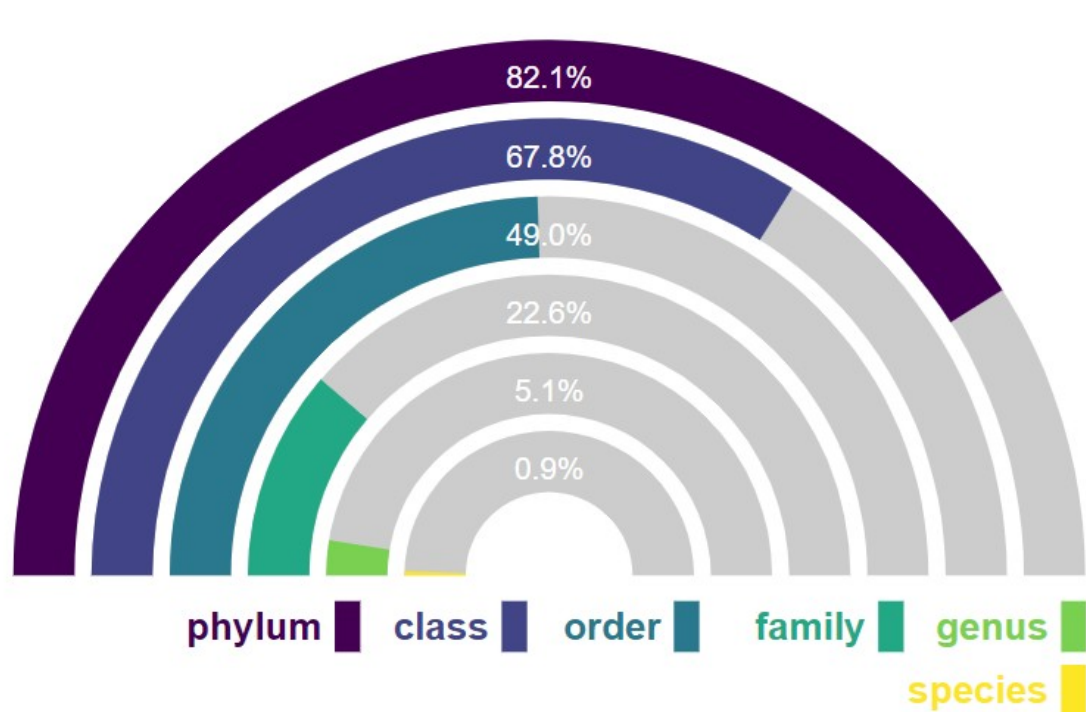
The GoaT main page shows a summary of all Eukaryotic Genome data (*per se*) retrieved from INSDC databases

<https://goat.genomehubs.org/>



Data summary

All search results in GoaT can be viewed in a table or as a set of interactive reports. The reports below present a live summary of the latest GoaT data. Click on a report to explore further.

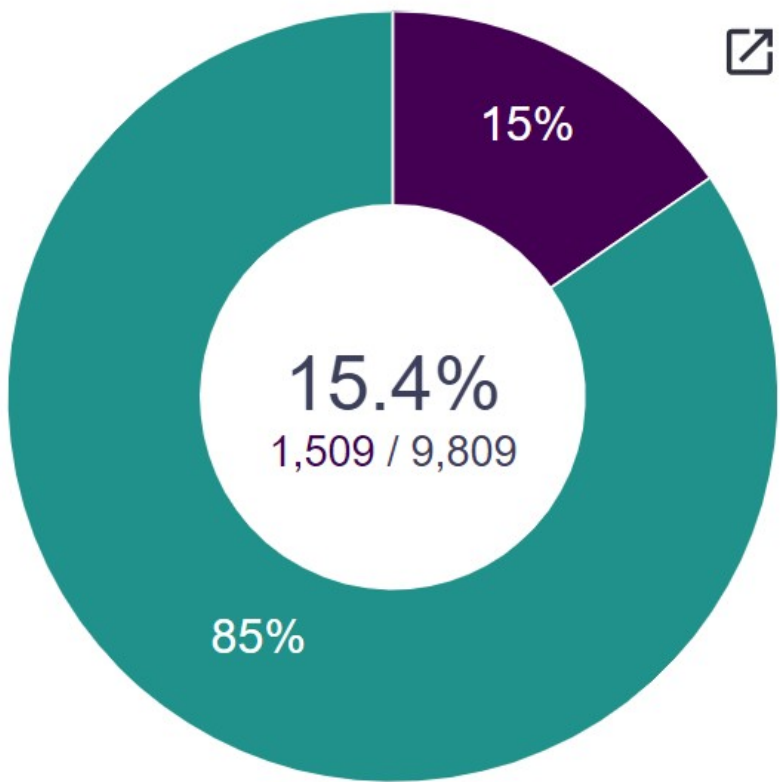


Taxa with assemblies out of all INSDC Eukaryotic taxa

1 superkingdoms
3 kingdoms
67 phyla
289 classes
1,424 orders
9,809 families
122,233 genera
1,622,497 species
33,625 subspecies

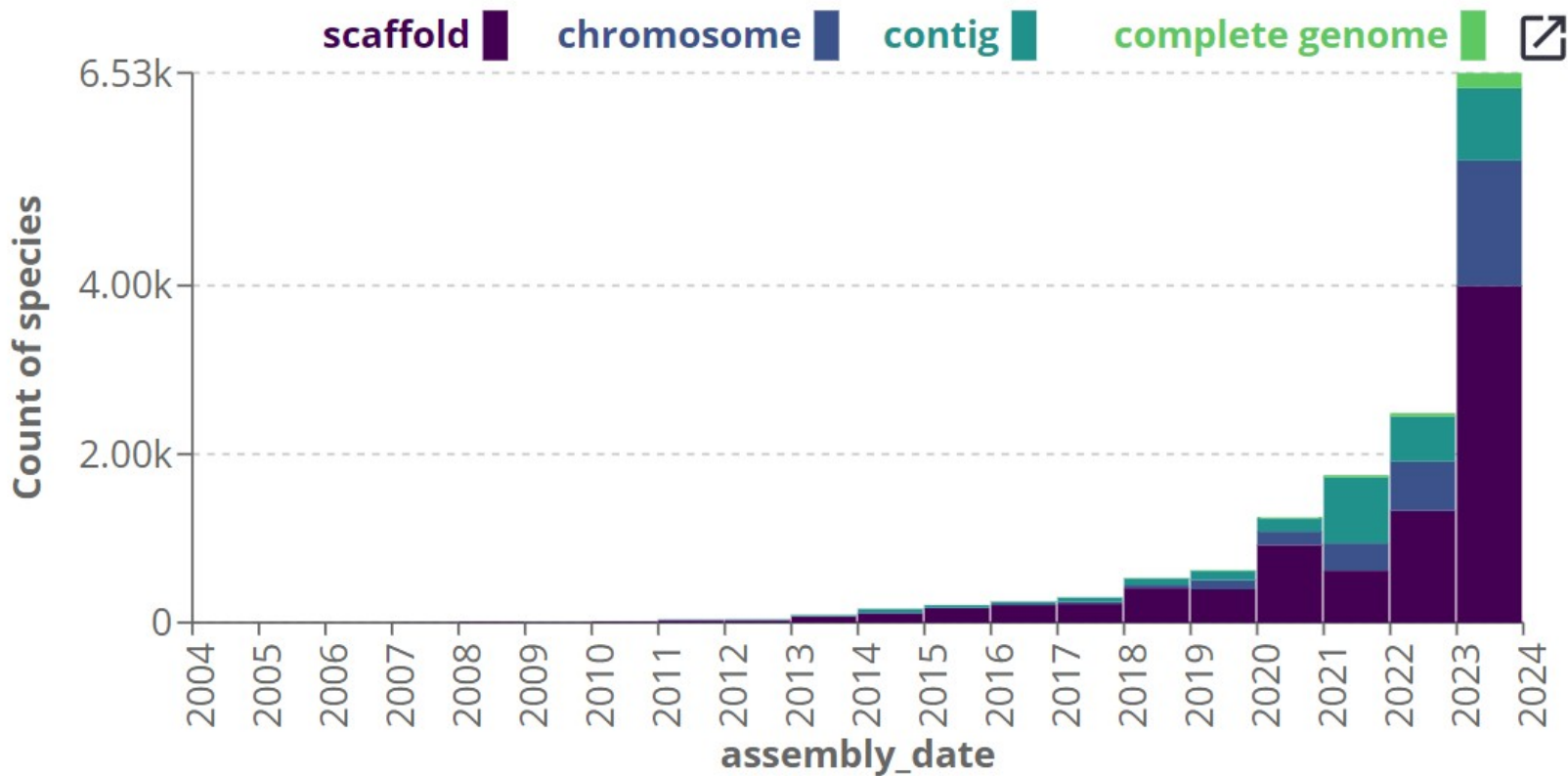
Counts of all taxa with information on GoaT

Genome Completeness Metrics include Chromosome Number 'not NULL' and 'Complete Genome' assembly level



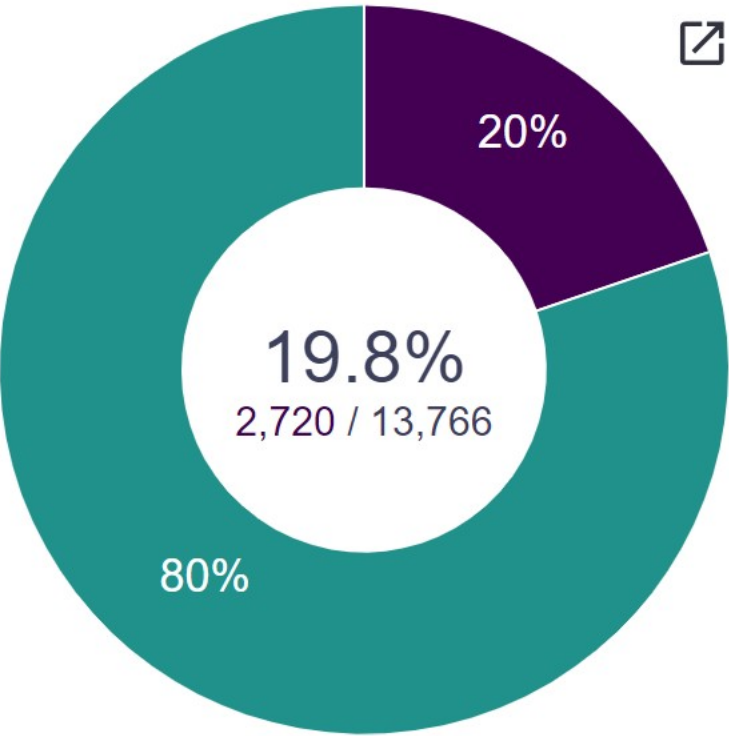
Total number of families with direct values of...

Total number of families with direct values of chromosome number

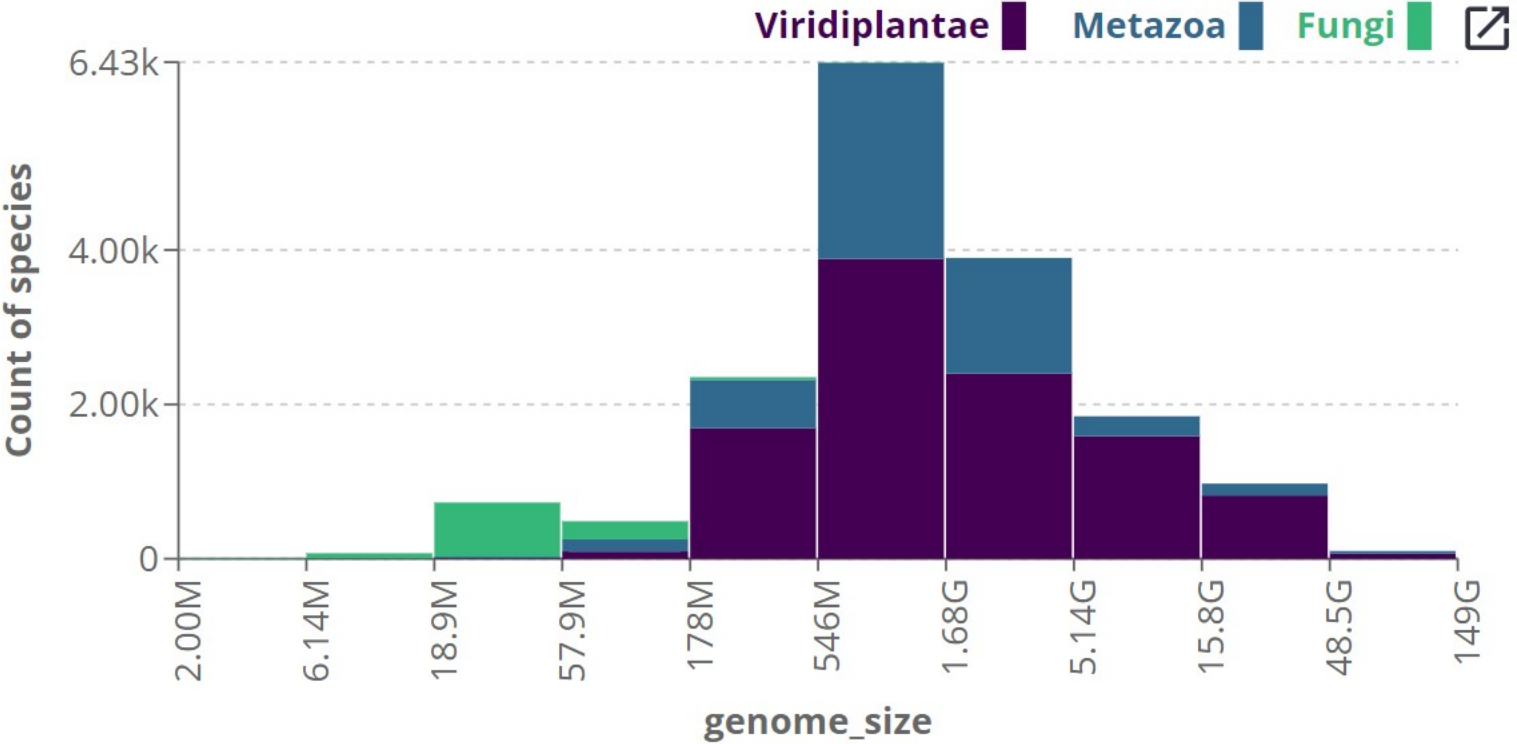


Progress of genome assemblies published on INSDC over time, by assembly level

Taxonomic Sampling and Comparative Genomics Metrics are also providing important information for Genomics Consortium project PI's and managers



Contribution of chromosome-level assemblies to the total...

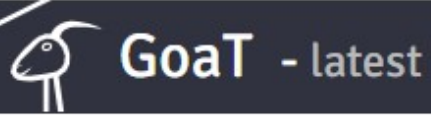


Frequency distribution of species genome sizes by kingdom

Contribution of chromosome-level assemblies to the total of species with available assemblies on INSDC

Genomics Consortium project PI's and managers highlighting progress in the [Earth BioGenome Project](#) and in individual affiliated projects such as the [Darwin Tree of](#)

Affiliated Genome Consortia have Individual Pages with Customized Target Taxa



search about help api projects regions submissions sources

<https://goat.genomehubs.org/projects>

Projects

GoaT hosts dedicated pages for the [Earth Biogenome Project](#) (EBP) and EBP-affiliated projects. Each project page contains information about the project, including its GoaT search term and BioProject ID together with reports of highlighting the planning, progress and completion of genome assemblies.

Earth Biogenome Project

- [EBP - Earth Biogenome Project \(bioproject PRJNA533106\)](#)

EBP Affiliated Project Networks

- [AFRICABP - African BioGenome Project \(AfricaBP, PRJNA811786\)](#)

Towards the goal of Genomics Meta-analysis and Database Resources, GIGA initialized a Genomes-on-a-Tree (GoaT) project page in 2022

GoaT is a resource developed by the Wellcome-Sanger Institute Tree of Life project.

<https://goat.genomehubs.org/>

<https://www.sanger.ac.uk/programme/tree-of-life/>

- **Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life.** Challis *et al.* 2023. Wellcome Open Res, 8:24 [doi:10.12688/wellcomeopenres.18658.1](https://doi.org/10.12688/wellcomeopenres.18658.1)

GoaT provides a resource for quick visualization of genome sequencing status and metadata within a taxonomic framework.

GIGA initialized a GoaT project page in 2022

<https://goat.genomehubs.org/projects/GIGA>

Type to search GoaT taxon index (e.g. Canidae)



include descendants



include estimates



result columns



query builder



clear all



Global Invertebrate Genomics Alliance (GIGA)

GoaT Search Term: GIGA

GIGA is building a network of scientists to tackle challenges associated with genome and transcriptome sequencing and promote standards that will facilitate comparative approaches to invertebrate genomics. For more information access GIGA's project page at <http://www.gigacos.org/>

Bioproject ID PRJNA649812

The Species Target List is the Basis for Customization of the GIGA-GoaT Project Page

The document is a relatively simple spreadsheet format

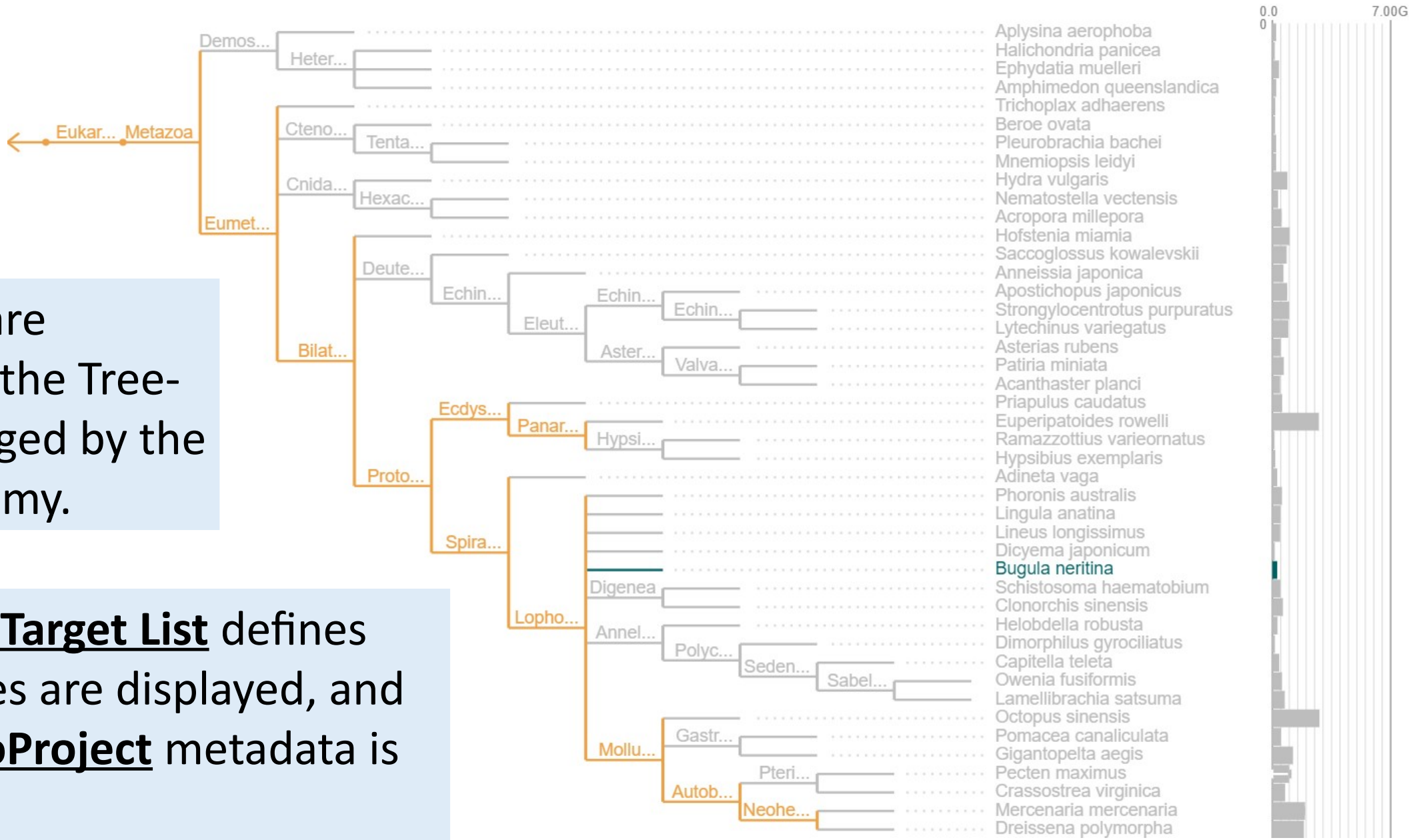
	A	B	C	D	E
1	# project_name	GIGA_phylums			
2	# subproject_name				
3	# primary_contact	Jeffrey Robinson			
4	# primary_contact_institution	UMBC			
5	# primary_contact_email	jrobin2@umbc.edu			
6	# date_of_update	2022-06-21			
7	# schema_version	ebp_species_goat_2.5			
8	ncbi_taxon_id	species	subspecies	phylum	class
9	6063	Halichondria panicea		Porifera	Demospongiae
10	289389	Aplysina aerophoba		Porifera	Demospongiae
11	6052	Ephydatia muelleri		Porifera	Demospongiae
12	400682	Amphimedon queenslandica		Porifera	Demospongiae

- The current species target list is limited to reference genomes from invertebrate taxa.
- These are placeholder taxa only, but the GIGA species target list is meant more to represent and track active sequencing efforts.

Target taxa are displayed in the Tree-of-Life arranged by the NCBI Taxonomy.

The **Species Target List** defines which species are displayed, and the **NBCI BioProject** metadata is overlaid.

<https://goat.genomehubs.org/projects/GIGA>



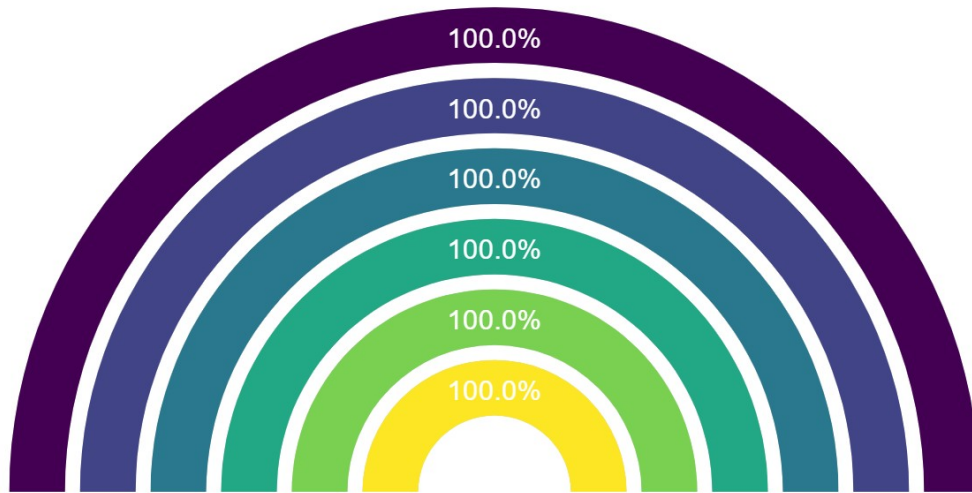
Here we see *Bugula neritina* highlighted because it is the only target-list species associated with the GIGA BioProject

GIGA Genomes Lists in the GoaT Dashboard

The GIGA-GoaT project page is initialized with a “Target List” of standard reference taxa from invertebrate Phyla, with an attempt to include the most common families. It appears as %100 completion rate because these genomes have already been sequenced”.

GIGA Project

<https://goat.genomehubs.org/projects/GIGA>

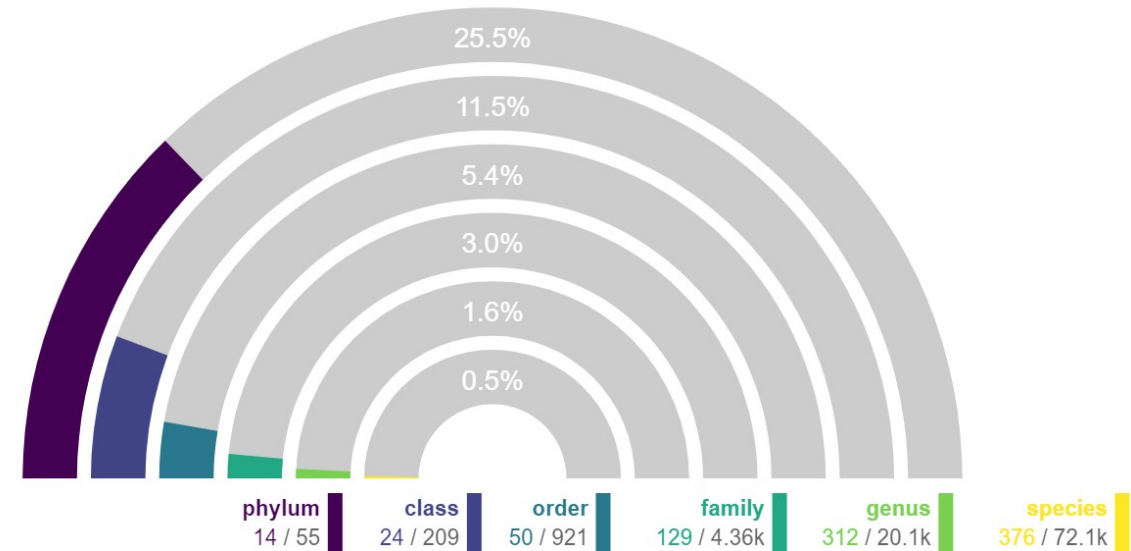


- An active target list would show as more similar to the Darwin Tree-of-Life project.
- Diversity of GIGA research, it may be suitable to organize “theme-centric” target lists.

1 kingdoms
19 phyla
23 classes
34 orders
44 families
44 genera
44 species

Darwin Tree-of-Life Project

<https://goat.genomehubs.org/projects/DTOL>



1 superkingdoms
3 kingdoms
55 phyla
209 classes
921 orders
4,362 families
20,109 genera
72,138 species

GoaT-GIGA Target List.

The Target List is a spreadsheet used to define target species for

	A	B	C	D	E
1	# project_name	GIGA_phylums			
2	# subproject_name				
3	# primary_contact	Jeffrey Robinson			
4	# primary_contact_institution	UMBC			
5	# primary_contact_email	jrobin2@umbc.edu			
6	# date_of_update	2022-06-21			
7	# schema_version	ebp_species_goat_2.5			
8	ncbi_taxon_id	species	subspecies	phylum	class
9	6063	Halichondria panicea		Porifera	Demospongiae
10	289389	Aplysina aerophoba		Porifera	Demospongiae
11	6052	Ephydatia muelleri		Porifera	Demospongiae
12	400682	Amphimedon queenslandica		Porifera	Demospongiae

F	G	H	I	J	K	L	M
family	target_list_status	sequencing_status	synonym	publication_id	Subproject_contributing_lab	Bioproject_umbrella	subproject_contact_person
Halichondriidae	family_representative						
Aplysinidae	family_representative						
Spongillidae	family_representative						
Niphatidae	family_representative	publication_available		10.1038/nature09201			
Edwardsiidae	family_representative	publication_available		10.1126/science.1139158			

UPDATED Target List

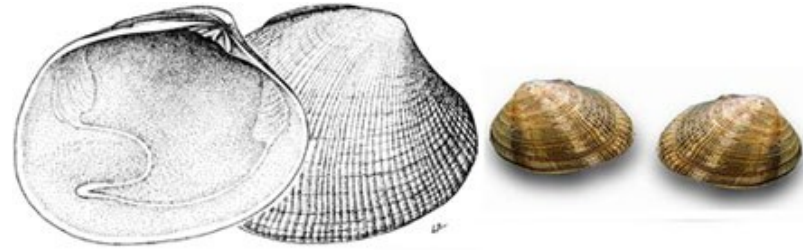
The great news is that Genome sequencing projects presented at GIGA V will now be included in the GIGA species target list!

Apologies there may have been errors or omissions. For example, I did not currently include taxa also in the Aquatic Symbiosis project or poster presenters.

Please contact me to update or correct the Target List details, or to update the subproject contact individual.

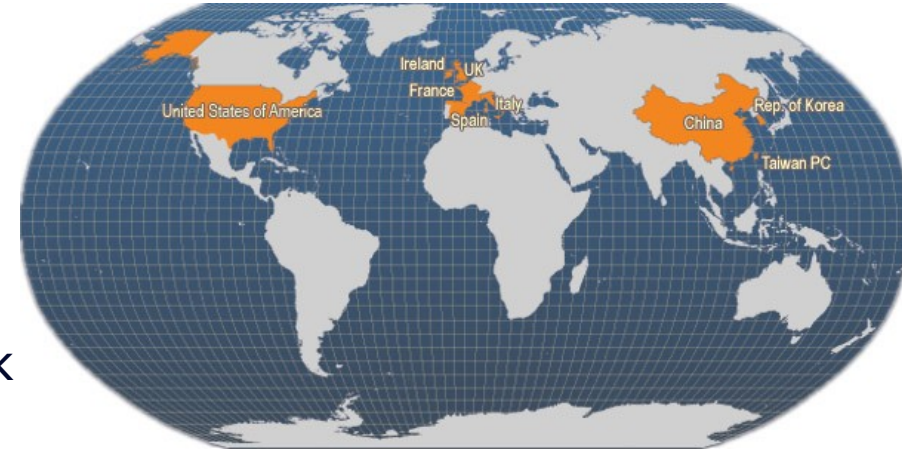
	A	B	C	D	E	F	
7	# schema version	ebp species goat 2.5					
8	ncbi_taxon_id	species	subspecies	phylum	class	family	subproject contact person
9	652874	Trapezia bidentata		Arthropoda	Malacostraca	Trapeziidae	Sarah Lemer
10	48498	Orbicella faveolata		Cnidaria	Anthozoa	Merulinidae	Natalia Andrade Rodriguez
11	39395	Cancer borealis		Arthropoda	Malacostraca	Cancridae	Dylan Comb
12	46700	Tubastraea coccinea		Cnidaria	Anthozoa	Trapeziidae	JoAo Gabriel Rodinho Nunes Ferrerira
13	351203	Scurria viridula		Mollusca	Gastropoda	Acmaeidae	Emily Giles/Pablo Saenz-Agudelo
14	351204	Scurria zebrina		Mollusca	Gastropoda	Acmaeidae	Emily Giles/Pablo Saenz-Agudelo
15	351200	Scurria scurra		Mollusca	Gastropoda	Acmaeidae	Emily Giles/Pablo Saenz-Agudelo
16	356393	Limnoperna fortunei		Mollusca	Bivalvia	Mytilidae	JoAo Gabriel Rodinho Nunes Ferrerira
17	928217	Madracis myriaster		Cnidaria	Anthozoa	Pocilloporidae	J.A. Baeza
18	3058964	Actuogordis ausrwaliensis		Nematomorpha	Gordioda	Gordiidae	Tauana Cuna
19	2528854	Nectonema munidae		Nematomorpha	Gordioda	Gordiidae	Tauana Cuna
20							

A Use-Case Example from a GIGA collaborator



Japanese carpet shell (AKA. small-neck Manila clam)

Ruditapes philippinarum
(Adams & Reeve, 1850) [Veneridae]



Dr. Fabrizio Ghiselli

Associate Professor
GIGA collaborator
Universita di Bologna
Italy

- Important aquaculture species in East Asia and Europe.
- Invasive in the USA
- Represented via multiple GIGA-linked BioProjects

Veneridae

Taxonomy ID: **6592** (for references in articles please use NCBI:txid6592)

current name

Veneridae

Genbank common name: **venus clams**
NCBI BLAST name: **bivalves**
Rank: **family**
Genetic code: [Translation table 1 \(Standard\)](#)
Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Lineage (full)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Protostomia](#); [Spiralia](#); [Lophotrochozoa](#); [Mollusca](#); [Bivalvia](#); [Autobranchia](#); [Heteroconchia](#); [Euheterodonta](#); [Imparidentia](#); [Neoheterodonte](#); [Venerida](#); [Veneroidea](#)

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	166,539	-
Protein	70,717	-
Structure	1 7	-
Genome	5	-
Popset	194	100
GEO Datasets	638	-
PubMed Central	8,292	231
Gene	58,984	-
SRA Experiments	1,949	-
Protein Clusters	12	-
Identical Protein Groups	56,712	-
BioProject	204	-
BioSample	2,054	-
Assembly	9	-
Taxonomy	367	1

In the NCBI Veneridae Taxonomy page, some important information includes:

1. Number of Accessioned Genomes (5)
2. Number of sub taxa (367, including Genus, species, and sub-species taxa)
3. NCBI Taxonomy ID (6592). **This is the only number we'll need to use GoaT.**

Back in the GoaT search window, we simply enter the taxID number, and GoaT generates a query string

We only need to paste the taxid number, and select the correct result, GoaT will generate a search string.

We want descendant taxa so we narrow the results by toggling the filters:

Type to search GoaT taxon index (e.g. Canidae)

tax_tree(6592[Veneridae])

include descendants
☒ On

include estimates
☐ Off

empty columns
☐ Off

result columns
⋮

query builder
⊕

clear all
↺

Change search index
TAXON
taxon
assembly
sample

13 hits

scientific_name	taxon_id	assembly_level ✓✓✓✓+	assembly_span ✓✓✓✓+	genome_size ✓✓✓✓+	chromosome_number ✓✓✓✓+	haploid_number
Leukoma staminea	462889	-	1.15G	1.64G	38	
Ruditapes decussatus	104385	-	1.27G	1.77G	38	
Ruditapes philippinarum	129788	Chromosome	1.27G	1.93G	38	
Tivela stultorum	345375	-	1.15G	939M	38	19 >
Saxidomus nuttalli	1129274	-	1.04G	1.08G	38	19 >
Saxidomus purpurata	311201	Scaffold	1.16G	1.08G	38	19 >
Saxidomus gigantea	410349	Scaffold	921M	1.08G	38	19 >

Our example starts with a taxon-centric search, but data can also be searched by Assembly ID and BioSample ID.

In the search settings, results columns can be selected and customized, including extensive assembly informations.

GoaT - latest

searchabouthelpapiprojectsregionssubmissionsources

☐ mitochondrion_assembly

☐ plastid_assembly

genome_size

genome_size

☐ regional_lists

☐ uk_legislation

☐ sample

☒ sequencing_status

sample_collected_by

number_acquired

sample_available

resampling_required

insdc_submitted

insdc_open

published

☒ assembly_level

☒ assembly_span

☐ biosample

☐ scaffold_n50

☐ contig_n50

☐ gene_count

☐ assembly_date

☐ ebp_metric_date

Saxidomus gigantea

410349

Scaffold

921M

1.08G

38

19

Cookie settings

Also, we can select the contributing laboratories.

atest search about help api projects regions submissions

☐ mitochondrion_assembly ☐ plastid_assembly

☐ genome_size ☐ regional_lists

☐ uk_legislation ☐ sample

☒ sequencing_status
sample_collected_by ☐ number_acquired ☐
sample_available ☐ resampling_required ☐
insdc_submitted ☐ insdc_open ☐ published ☐

☐ assembly
assembly_level ☐ assembly_span ☐



















































☐ karyotype
chromosome_number ☐ haploid_number ☐

Detailed information of responsible for species genome sequencing

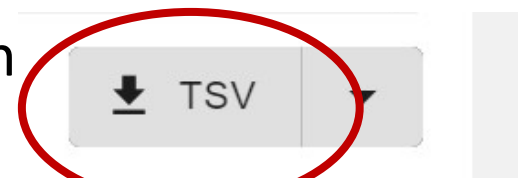
☐ contributing_project_lab
☐ other_priority
☐ family_representative
☐ long_list

Saxidomus gigantea 410349 Scaffold 921M 1.08G

- Run the search, and sort by assembly_level column, we'll see the most advanced genome sequencing projects from species in Family Veneridae:
- Check out the data on Genome Size, Chromosome Number.

scientific_name	taxon_id	common_name	assembly_level 	assembly_span	genome_size	chromosome_number	haploid_number	insdc_open	sequencing_status_giga	sequencing_status	sample_collected
<i>Ruditapes philippinarum</i>	129788	Japanese littleneck	 Chromosome	 1.27G	 1.93G	 38	 19	 GIGA	 insdc_open	 insdc_open	-
<i>Mercenaria mercenaria</i>	6596	northern quahog	 Chromosome	 1.86G	 1.96G	 38	 19	 GIGA	 published	 published	 GIGA
<i>Saxidomus purpurata</i>	311201	-	 Scaffold	 1.16G	 1.08G	 38	 19	-	-	 insdc_open	-
<i>Saxidomus gigantea</i>	410349	-	 Scaffold	 921M	 1.08G	 38	 19	-	-	 insdc_open	-
<i>Cyclina sinensis</i>	120566	-	 Scaffold	 903M	 1.79G	 38	 19	-	-	 insdc_open	-
<i>Leukoma staminea</i>	462889	Pacific littleneck	-	 1.15G	 1.64G	 38	 19	-	-	-	-
<i>Ruditapes decussatus</i>	104385	Venus clam	-	 1.27G	 1.77G	 38	 19	-	-	-	-
<i>Mysia undata</i>	1920014	-	-	 1.15G	 1.79G	 38	 19	-	-	 in_progress	 DTOL

Once we have selected the relevant columns, the results can be download in a spreadsheet format.



Once we have selected the relevant columns, the results can be download in a spreadsheet format for further analysis. Notice that some columns may not be populated.

AutoSaveOff

download

Search

Jeffrey RobinsonJR

Comments

Share

FileHomeInsertDrawPage LayoutFormulasDataReviewViewAutomateHelpAcrobat

Paste

Clipboard

Font

Alignment

Number

Styles

Cells

Analysis

Sensitivity

Add-ins

Adobe Acrobat

POSSIBLE DATA LOSS

Some features might be lost if you save this workbook in the text (.txt) format. To preserve these features, save it in an Excel file format.

Don't show again

Save As...

	A	B	C	D	E	F	G	H	I	
1	taxon_id	scientific_name	common_name	assembly_level	assembly_span	genome_size	contributing_project_lab	insdc_open	sequencing_status	sam
2	129788	Ruditapes philippinarum	["Japanese littleneck"]	Chromosome	1.27E+09	1.93E+09		GIGA	insdc_open	
3	6596	Mercenaria mercenaria	["northern quahog"]	Chromosome	1.86E+09	1.96E+09		GIGA	published	GIGA
4	311201	Saxidomus purpurata		Scaffold	1.16E+09	1.08E+09			insdc_open	
5	410349	Saxidomus gigantea		Scaffold	9.21E+08	1.08E+09			insdc_open	
6	120566	Cyclina sinensis		Scaffold	9.03E+08	1.79E+09			insdc_open	
7	462889	Leukoma staminea	["Pacific littleneck"]		1.15E+09	1.64E+09				
8	104385	Ruditapes decussatus	["Venus clam"]		1.27E+09	1.77E+09				
9	1920014	Mysia undata			1.15E+09	1.79E+09			in_progress	DTOI
10	380098	Clausinella fasciata			1.15E+09	1.79E+09			in_progress	DTOI
11	345375	Tivela stultorum	["Pismo clam"]		1.15E+09	9.39E+08			in_progress	CCGF
12	1129274	Saxidomus nuttalli	["California butterclam"]		1.04E+09	1.08E+09				
13	55715	Venus verrucosa			1.15E+09	1.79E+09			in_progress	ATLA
14	138984	Chamelea gallina			1.15E+09	1.79E+09			in_progress	DTOI
15	358437	Chamelea striatula			1.15E+09	1.79E+09			in_progress	DTOI
16	2602930	Protocallithaca adamsi	["Clam"]		1.15E+09	1.93E+09				
17	584776	Polititapes rhomboides			1.15E+09	1.79E+09			in_progress	DTOI
18	115816	Mercenaria campechiensis	["southern quahog"]		1.86E+09	2.25E+09				
19	380119	Timoclea ovata			1.15E+09	1.79E+09			in_progress	DTOI

download

AutoSave Off

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Acrobat

Clipboard Font Alignment Number Styles

Calibri 11

General

Conditional Formatting

Format as Table

Cell Styles

G9

	A	B	C	D	E	F
	taxon_id	assembly_level	common_name	sequencing_status	sample_collected	scientific_name
1	129788	Chromosome	["Japanese littleneck"]	insdc_open		Ruditapes philippinarum
2	6596	Chromosome	["northern quahog"]	published	GIGA	Mercenaria mercenaria
3	311201	Scaffold		insdc_open		Saxidomus purpurata
4	410349	Scaffold		insdc_open		Saxidomus gigantea
5	120566	Scaffold		insdc_open		Cyclina sinensis
6	462889		["Pacific littleneck"]			Leukoma staminea
7	104385		["Venus clam"]			Ruditapes decussatus
8	1920014			in_progress	DTOL	Mysia undata
9	380098			in_progress	DTOL	Clausinella fasciata
10	345375		["Pismo clam"]	in_progress	CCGP	Tivela stultorum
11	1129274		["California butterclam"]			Saxidomus nuttalli
12	55715			in_progress	ATLASEA	Venus verrucosa
13	138984			in_progress	DTOL	Chamelea gallina
14	358437			in_progress	DTOL	Chamelea striatula
15	2602930		["Clam"]			Protocallithaca adamsi
16	584776			in_progress	DTOL	Polititapes rhomboides
17	115816		["southern quahog"]			Mercenaria campechiensis
18	380119			in_progress	DTOL	Timoclea ovata
19	1028563		["Venus clam"]	in_progress	DTOL	Venerupis corrugata
20	368946		["Dosinia"]			Dosinia japonica
21	78549			sample_collected	ATLASEA	Callista chione
22						
23						
24						

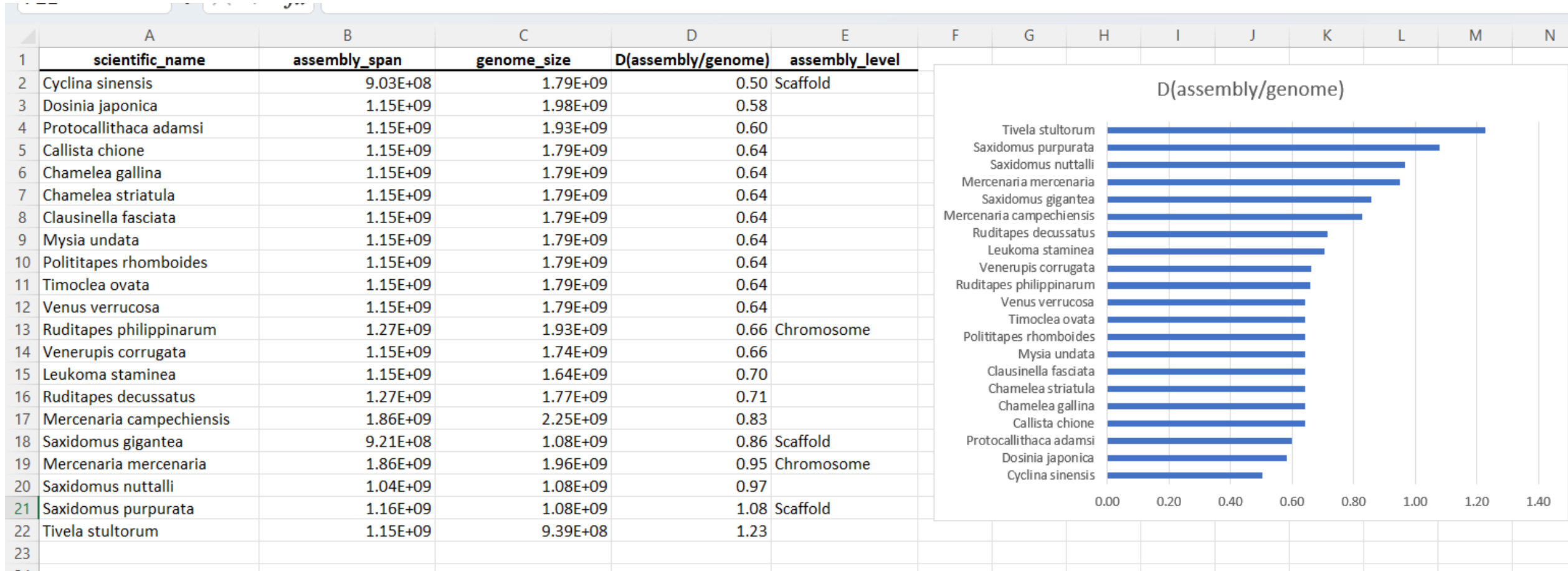
For the goal of identifying sequencing consortia for associated taxa, the **sample_collected** column shows the sample collecting consortia, if there was one.

Finally we'll analyze the assembly vs the known genome size

scientific_name	assembly_span	genome_size
Ruditapes philippinarum	1.27E+09	1.93E+09
Mercenaria mercenaria	1.86E+09	1.96E+09
Saxidomus purpurata	1.16E+09	1.08E+09
Saxidomus gigantea	9.21E+08	1.08E+09
Cyclina sinensis	9.03E+08	1.79E+09
Leukoma staminea	1.15E+09	1.64E+09
Ruditapes decussatus	1.27E+09	1.77E+09
Mysia undata	1.15E+09	1.79E+09
Clausinella fasciata	1.15E+09	1.79E+09
Tivela stultorum	1.15E+09	9.39E+08
Saxidomus nuttalli	1.04E+09	1.08E+09
Venus verrucosa	1.15E+09	1.79E+09
Chamelea gallina	1.15E+09	1.79E+09
Chamelea striatula	1.15E+09	1.79E+09
Protocallithaca adamsi	1.15E+09	1.93E+09
Polititapes rhomboides	1.15E+09	1.79E+09
Mercenaria campechiensis	1.86E+09	2.25E+09
Timoclea ovata	1.15E+09	1.79E+09
Venerupis corrugata	1.15E+09	1.74E+09
Dosinia japonica	1.15E+09	1.98E+09
Callista chione	1.15E+09	1.79E+09

- The difference between assembly_span vs genome_size can inform the completeness of sequencing strategy.

A quick Excel analysis shows the relative putative difference between assembly span and genome size.



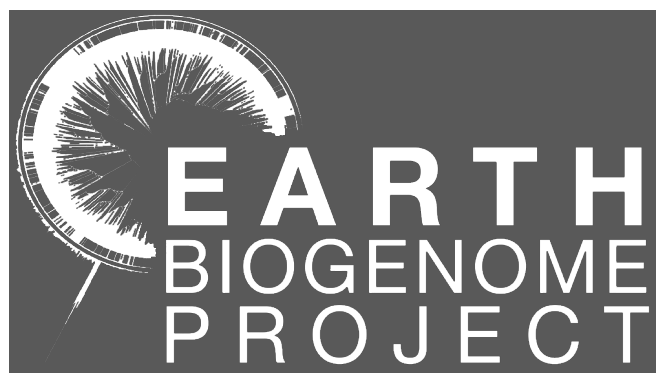
So, at least according to data found in NCBI, and processed through GoaT's infrastructure, it is

Tremendous Thanks!

- Dr. Joe Lopez
- Dr. Cibeles Sotero-Caio
- Dr. Fabrizio Ghiselli
- GIGA-V organizing committee
- Robinson Scientifics LLC



UMBC



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA