



e **ERG** **A**

BGC²₄

European Reference Genome Atlas

Biodiversity Genomics 2024
28 October 2024



Introduction

Navigating the European Reference Genome Atlas

Sampling for reference genomes across Europe with community help

The ERGA Genome Tracking Console

A Decentralised Method for Community-Reviewing Earth BioGenomes

ERGA Community Genomes

Q&A

ERGA Executive Board

Chiara Bortoluzzi

Astrid Böhne

Tyler Alioto

Diego de Panis

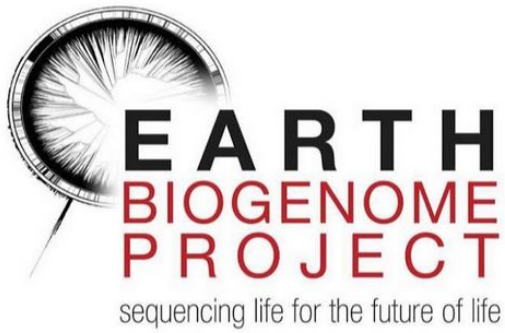
Robert Waterhouse



Welcome & Introduction

Rob Waterhouse





A GRAND CHALLENGE

The Earth BioGenome Project (EBP), a *moonshot* for biology, aims to sequence, catalog and characterize the genomes of all of Earth's eukaryotic biodiversity over a period of ten years.

A GRAND VISION

Create a new foundation for biology to drive solutions for preserving biodiversity and sustaining human societies.



Major projects:



Regional nodes:



ERGA Committees



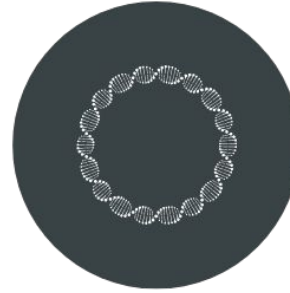
SSP

Sampling & Sample
Processing



SAC

Sequencing &
Assembly



Annot

Annotation
Committee



DAC

Data Analysis
Committee



ITIC

IT & Infrastructure
Committee



ELSI

Ethical, Legal and
Social Issues



Media

Communications and
public affairs



CS

Citizen Science
Committee



TKT

Training and
Knowledge Transfer



SJC

Social Justice
Committee

ERGA Council

EUROPEAN REFERENCE GENOME ATLAS



*as of January 2024



Navigating the European Reference Genome Atlas

Chiara Bortoluzzi





Navigating the European Reference Genome Atlas

Chiara Bortoluzzi



The 3-fold vision



An initiative that generates reference genomes and all needed resources



An umbrella network for all projects producing reference genomes from European biodiversity

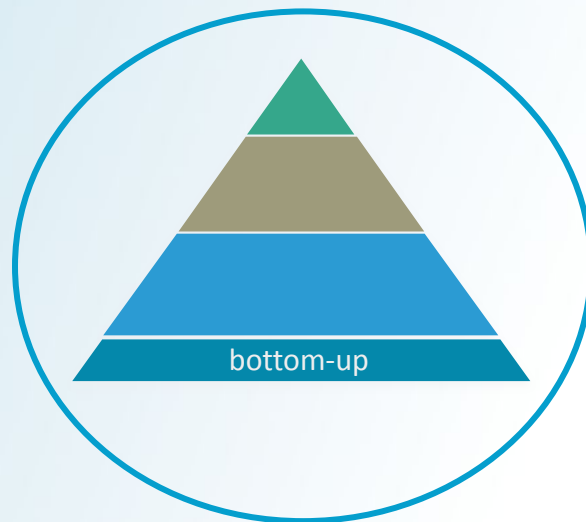


A knowledge hub for each step of reference genome generation and use

The European Reference Genome Atlas

A Pan-European BioGenome network promoting reference genome sequencing of Europe's biodiversity

- ❖ > 1000 Europe-based scientists of
- ❖ > 200 Research Institutions
- ❖ 38 countries



Increasing ERGA community participation

Which European species should be sequenced next?



Call for Species Proposals & Community sampling



CALL FOR EXPRESSION OF INTEREST FOR FINANCIAL SUPPORT

Developing biodiversity genomics applications for species in Europe to support the mission of Biodiversity Genomics Europe



CALL FOR EXPRESSION OF INTEREST FOR FINANCIAL SUPPORT

Sampling reference genomes of eukaryotic species from Biodiversity Hotspots in Europe to support the mission of BGE



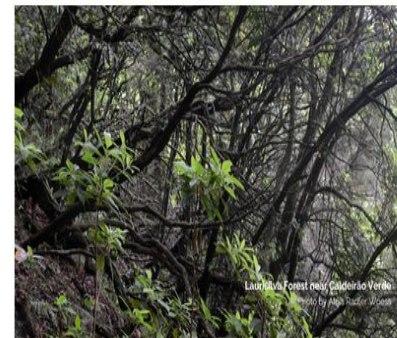
View to the Dolomites, Croatia



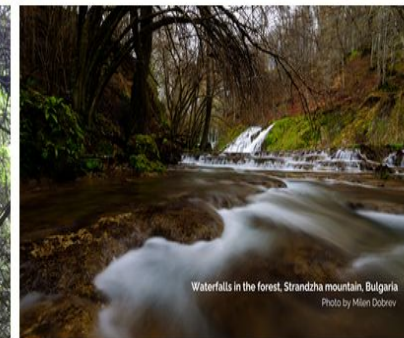
River Rioni, Georgia
Photo by K. Sathya



Cape Greco National Forest Park, Cyprus
Photo by Marina Mitsuola



Landscape Forest near Castell de Ferro
Photo by Ana Sofia Wilson



Waterfalls in the forest, Strandzha mountain, Bulgaria
Photo by Milan Dobrev



Tapolca, Transdanubia, Hungary
Photo by Anasztás Rószky

Biodiversity Genomics Europe Case Studies



Mapping the genomic basis of common thyme aromatic diversity and its adaptive significance for ecotype formation and climate change adaptation

Case Studies

Mapping the genomic basis of common thyme aromatic diversity and its adaptive significance for ecotype formation and climate change adaptation

October 21st, 2024 | [Case Studies](#)

[Read More >](#)



Annotating the Corema album Genome: A Step Toward Coastal Ecosystem Conservation

October 14th, 2024 | [Case Studies](#)

Case study Annotating the Corema album Genome: A Step Toward Coastal Ecosystem Conservation Researchers

[Read More >](#)



New reference genomes for European red algae

September 23rd, 2024 | [Case Studies](#)

Case study New reference genomes for European red algae
Establishing reference genomes for two red macroalgae (Porphyra dioica and Porphyra)



Enhancing the Long-term Viability and Reducing Extinction Risk of the Little Owl in Europe

September 16th, 2024 | [Case Studies](#)

Case study Studying Historical DNA to help protect the Little Owl Researchers from the Globe Institute at the

AND MANY MORE!

Read about all of them:
<https://biodiversitygenomics.eu/citizen-science/case-studies/>



Connecting past and future: range-wide genomic monitoring for Arnica

October 7th, 2024 | [Case Studies](#)

Case study Bringing Arnica conservation research to a new level A new reference genome for the popular medicinal plant Arnica



A High-Quality Reference Genome of one of the Rarest European Mammals, the Hungarian Birchmouse (Sicista trizona trizona, Sminthidae, Rodentia)

September 30th, 2024 | [Case Studies](#)

Case study A High-Quality Reference Genome of one of the

BGE - actions to enable Reference Genomes production

SOPs for specimen sampling,
metadata collection and shipment

Biobanking standards and guidelines

Computational workflows

Tutorials and training materials



1 Onboarding and Regulatory Framework

- a. Read and comply with the provided legal BGE-ERGA documentation
- b. Follow the legal and binding guidelines correspondent to your assigned Sequencing Facility and sign any required documents. If needed, contact the person in charge listed on the Table 1.
- c. Ensure that you possess all legally required permits. Those (if applicable) will be required in the following step.

2 Sample collection and storage

Before collecting the samples, ensure you are aware of all the required **Sample Manifest** fields to be filled in later (See Section 3)

- a. Collect the samples, following the requirements specified from your assigned Sequencing Facility (SOP, Table 1).
 - b. Separate a specimen Voucher, when possible (following taxon-specific standards) and take scaled pictures.
 - c. Dissect your specimens following the taxon/sequencing facility specific guidelines, and separate samples for:
 - Barcode your samples prior to sample shipping (1 tube only), following this SOP.
 - Biobank: Reserve one tissue tube to be stored at a Biobank Facility, following these requirements.
 - Sequencing: collect one lentil-sized piece of tissue per tube and place it in pre-chilled tubes. Collect 6-10 tubes*, if possible from different tissues.
 - d. Flash-freeze and store the tubes in:
 - dry ice, LN2 Dry Shippers (< -150°C) or at -80°C freezer
 - Wet ice and -20°C freezers are not suitable for the storage of FluidX tubes containing samples.
- Cell culture (optional):** If planning to send samples for Cell culture, please follow these instructions, and get in touch with ERGA-cells@upf.edu for guidance.



Biodiversity
Genomics
Europe



CONTRIBUTE TODAY: <https://knowledge.erga-biodiversity.eu/>



Is your favourite course or resource not listed here? If you find a problem please let us know. Any contribution to this collection is highly appreciated!

[CONTRIBUTE!](#)

Welcome to the ERGA Knowledge Hub - this is an open and global collection of educational resources in biodiversity genomics curated by the European Reference Genome Atlas Community. Submit and share resources to join our mission breaking down geographical barriers and facilitating knowledge transfer.

All resources 46 resources	Sampling and Sample Processing 2 resources	Genome Sequencing and Assembly 12 resources	Genome Annotation 14 resources	Data Analysis 14 resources	IT and Infrastructure 9 resources	Ethical, Legal, and Social Issues 7 resources	Citizen Science 2 resources	Training and Knowledge Transfer 1 resource
-------------------------------	---	--	-----------------------------------	-------------------------------	--------------------------------------	--	--------------------------------	---

Global search ("/" to focus)

45 Resources

PAGE 1 / 2

PER PAGE

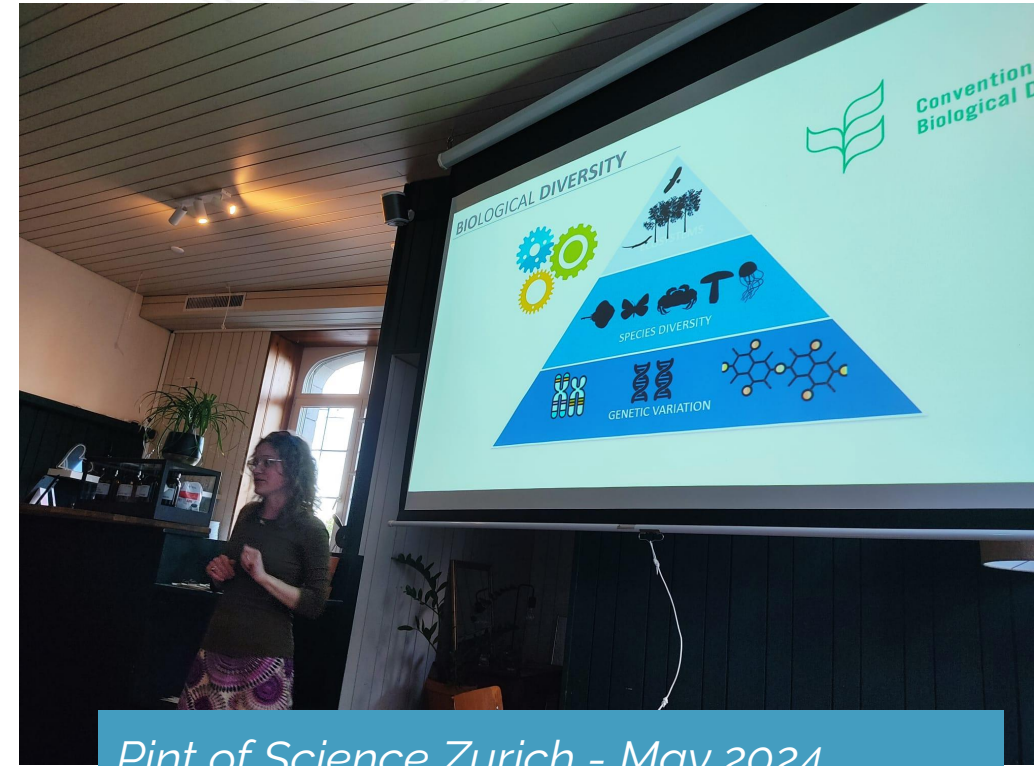
25

TYPE	RESOURCE	AUTHORS	TOPICS	LAST UPDATED	LICENSE
All	<input type="text" value="Search by name"/>	<input type="text" value="Search by author"/>		Max.	All

Genomics Next generation sequencing

TOPICS	
<input checked="" type="checkbox"/> IT & Infrastructure	3
<input type="checkbox"/> Unix/Linux	3
<input checked="" type="checkbox"/> Computational methods and pipelines	8
<input type="checkbox"/> Galaxy	5
<input type="checkbox"/> High performance computing	3
<input checked="" type="checkbox"/> Omics analysis	32
<input type="checkbox"/> Genomics	17
<input type="checkbox"/> RNA-seq	2
<input type="checkbox"/> Comparative genomics	3
<input type="checkbox"/> Genome assembly	8
<input type="checkbox"/> Population genetics	3
<input type="checkbox"/> Assembly Curation	2
<input type="checkbox"/> Next generation sequencing	8
<input type="checkbox"/> Variant analysis	5
<input type="checkbox"/> Genome annotation	13
<input type="checkbox"/> Long read sequencing	5
<input type="checkbox"/> Phylogenetics / Phylogenomics	3
<input type="checkbox"/> Assembly Evaluation	1
<input checked="" type="checkbox"/> Reproducibility and data management	9
<input type="checkbox"/> Data management	5
<input type="checkbox"/> FAIR data	2

ERGA Communications - Outreach & Engagement



Long Night of Science Berlin - July 2024

Pint of Science Zurich - May 2024



contact@erga-biodiversity.eu
www.erga-biodiversity.eu

@erga_biodiv

ERGA Communications - Community Engagement



SMBE Conference 2023 - Ferrara, Italy



XX International Botanical Congress - 2024

Parallel session organized by Katja Reichel



contact@erga-biodiversity.eu
www.erga-biodiversity.eu

@erga_biodiv

Follow us to stay updated!



www.erga-biodiversity.eu

contact@erga-biodiversity.eu



@erga_biodiv



@erga_biodiv@genomic.social



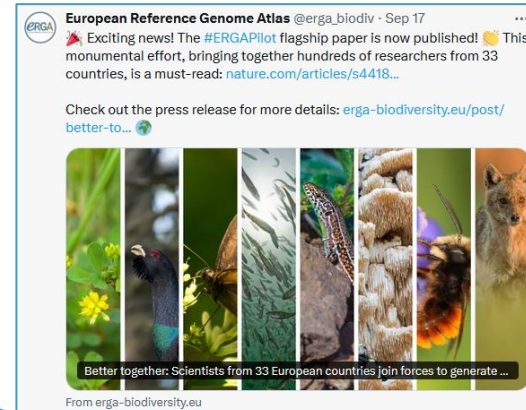
ERGA - European Reference Genome Atlas



ERGA-Consortium

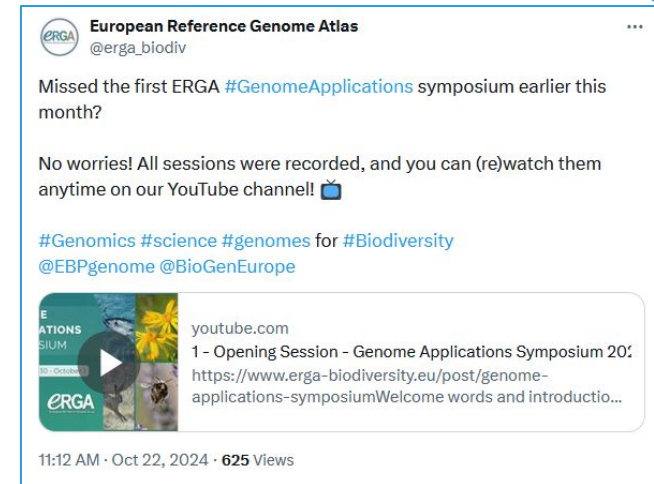
Our content

Reading Recommendations & Press Releases



We can also promote your ERGA-related publications and initiatives!

Promoting training opportunities, events & resources





Inclusion

Take actions that promote inclusion of researchers and citizens from European regions with low genomic resources and minorities of any kind.

ERGA is a fully open initiative and anyone is welcome to join!

- All nationalities
- All career stages
- All areas of expertise



[REGISTER NOW >](#)



contact@erga-biodiversity.eu

[@erga_biodiv](https://twitter.com/erga_biodiv)

www.erga-biodiversity.eu



Sampling for reference genomes across
Europe using community help

Astrid Böhne





ERGA

Sampling for reference genomes across
Europe using community help

Astrid Böhne





Biodiversity Genomics Europe

 Sampling Strategies

Community

2 ERGA Calls for European species nominations



Which European species should be sequenced next?



Call for Species Proposals & Community sampling



Hotspots



Species nominated from 3 European hotspot regions of the mediterranean basin





Bioblitz

Call for subcontracted* Bioblitz campaigns in European Hotspots



*ca. 30K EUR each



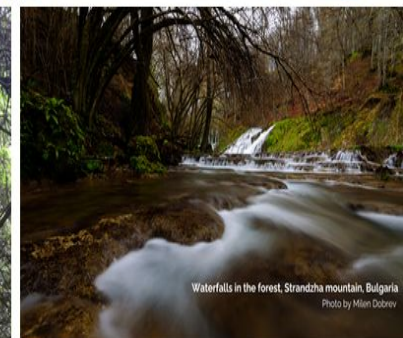
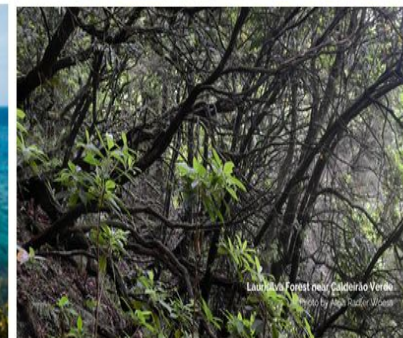
Bioblitz

Call for subcontracted* Bioblitz campaigns in European Hotspots



- Georgia
- Cyprus
- Croatia
- Portugal (Madeira)
- Hungary
- Bulgaria


*ca. 30K EUR each




 Sampling Strategies

Community

2 ERGA Calls for European species nominations




Which European species should be sequenced next?



Call for Species Proposals & Community sampling

Hotspots

Species nominated from 3 European Hotspot Regions of the mediterranean basin



Spain, Slovenia, Greece

Bioblitz

Call for subcontracted* Bioblitz campaigns in European Hotspots



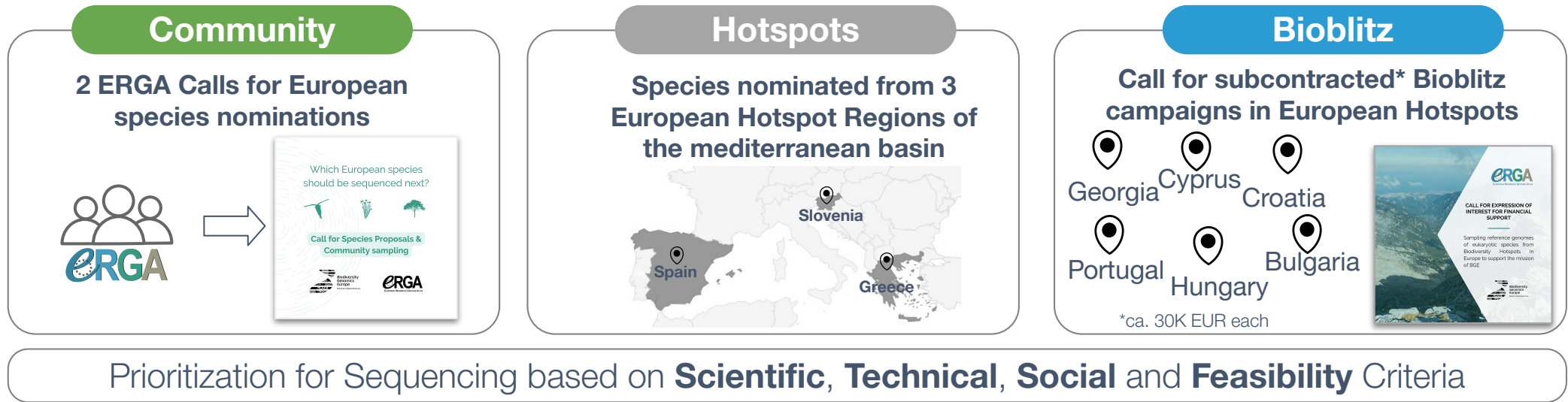
Georgia, Cyprus, Croatia, Portugal, Hungary, Bulgaria

*ca. 30K EUR each



 Sampling Strategies

 Species Selection




Sampling Strategies

Species Selection

Sample Collection

Community


2 ERGA Calls for European species nominations



Which European species should be sequenced next?
Call for Species Proposals & Community sampling

Hotspots

Species nominated from 3 European Hotspot Regions of the mediterranean basin



Spain, Slovenia, Greece

Bioblitz

Call for subcontracted* Bioblitz campaigns in European Hotspots

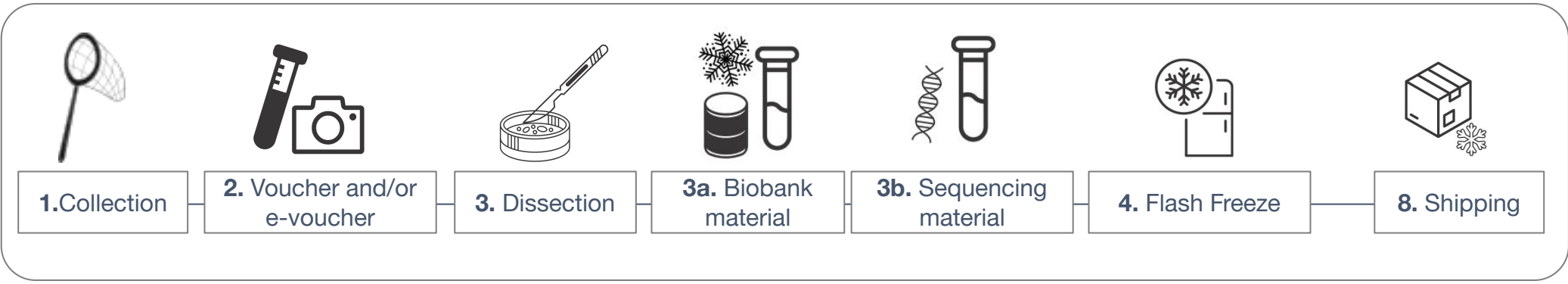


Georgia, Cyprus, Croatia, Portugal, Hungary, Bulgaria

*ca. 30K EUR each



Prioritization for Sequencing based on **Scientific, Technical, Social** and **Feasibility** Criteria

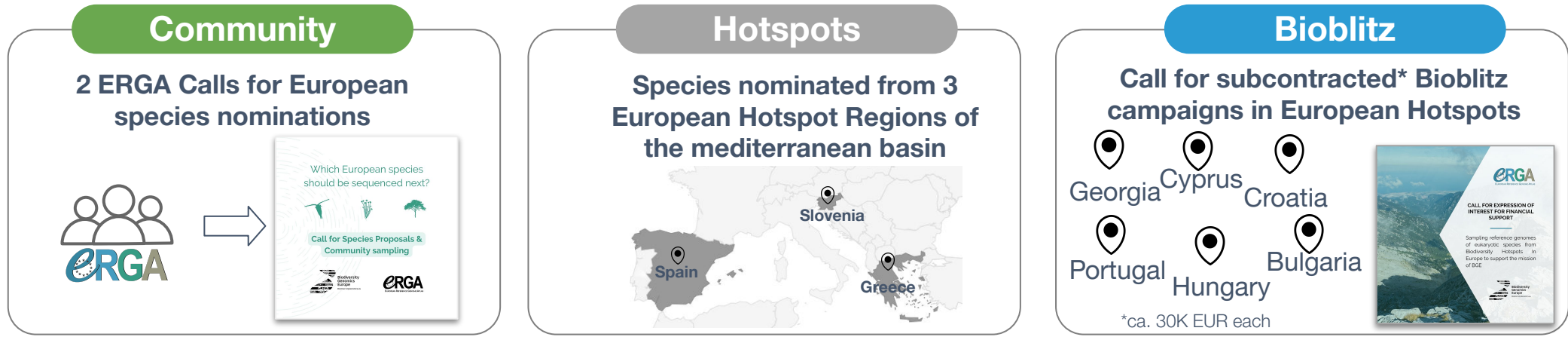


Sampling Strategies

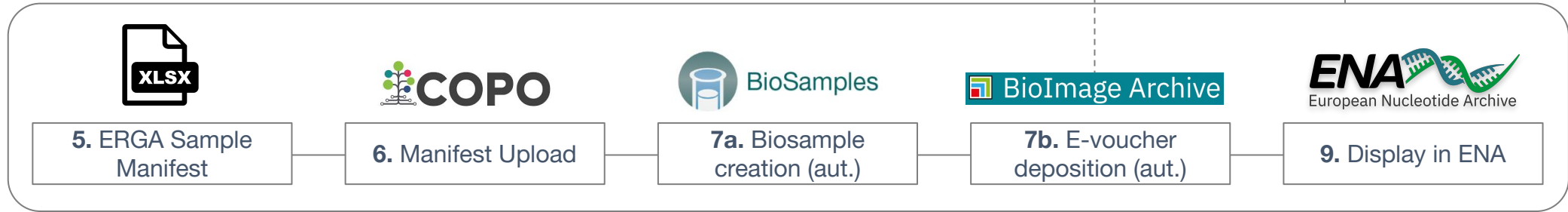
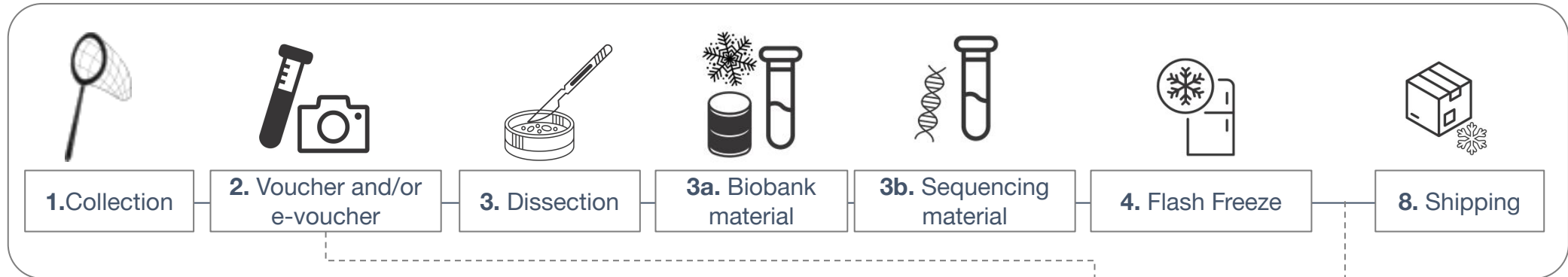
Species Selection

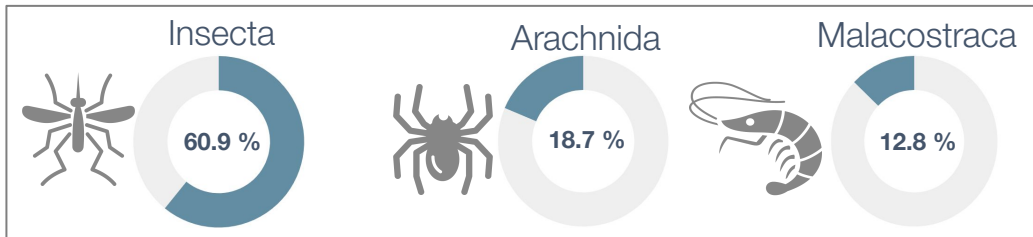
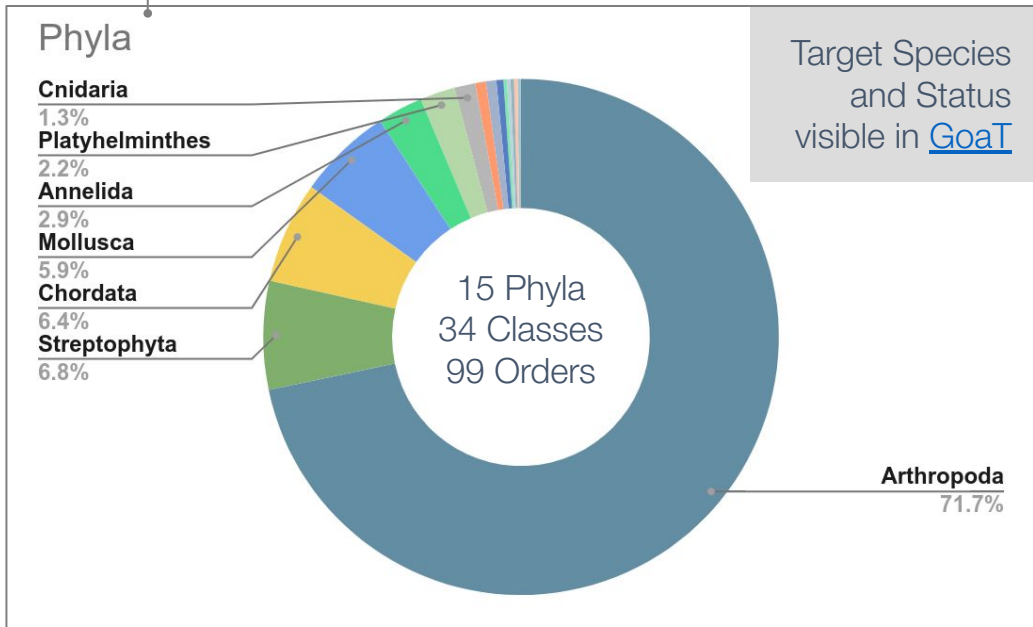
Sample Collection

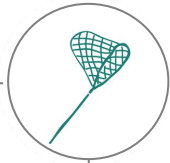
Metadata Records



Prioritization for Sequencing based on **Scientific, Technical, Social** and **Feasibility** Criteria







ca. **1111** collected species
Oversampling



22 countries sampled



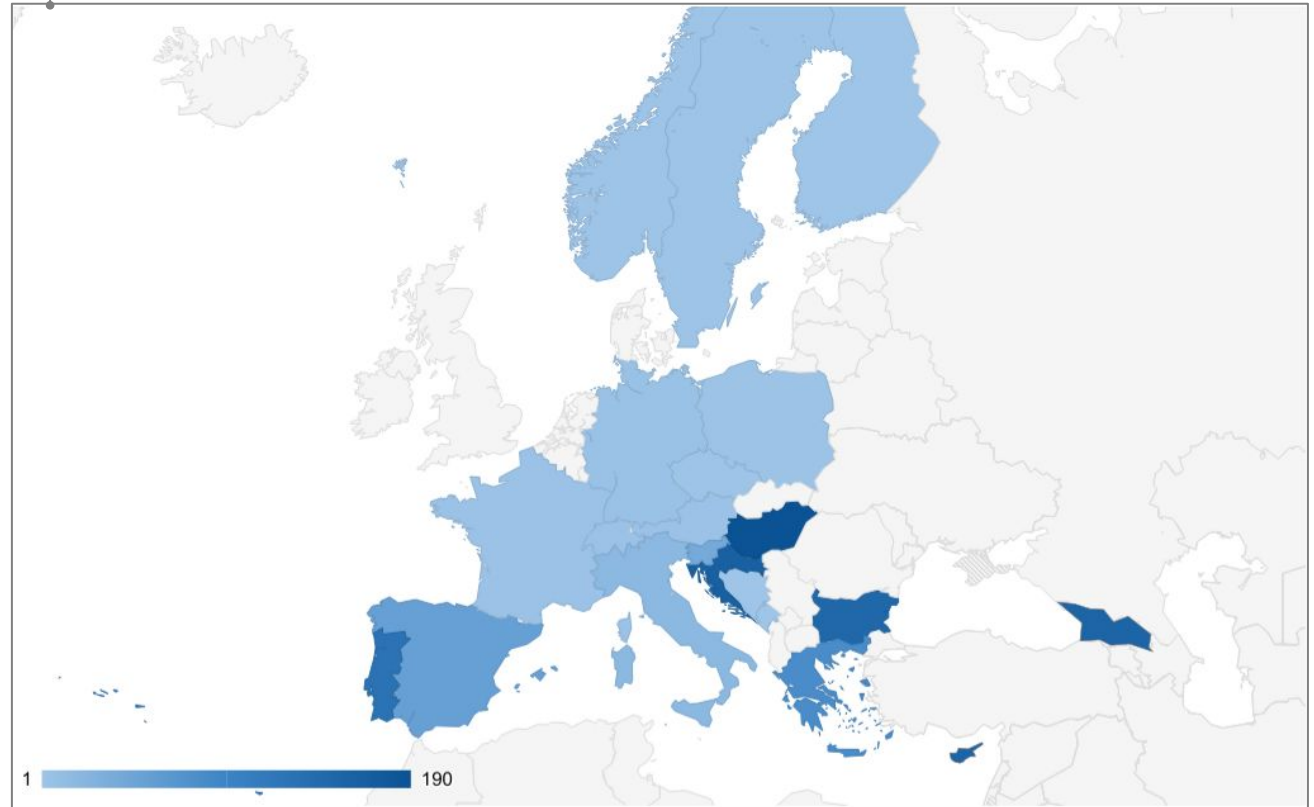
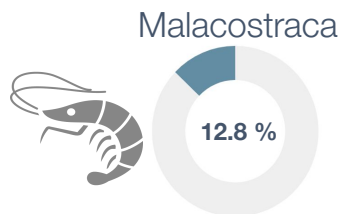
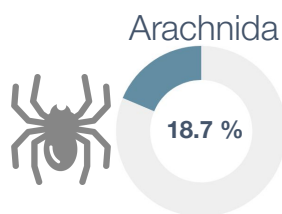
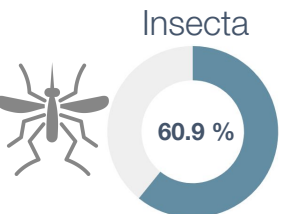
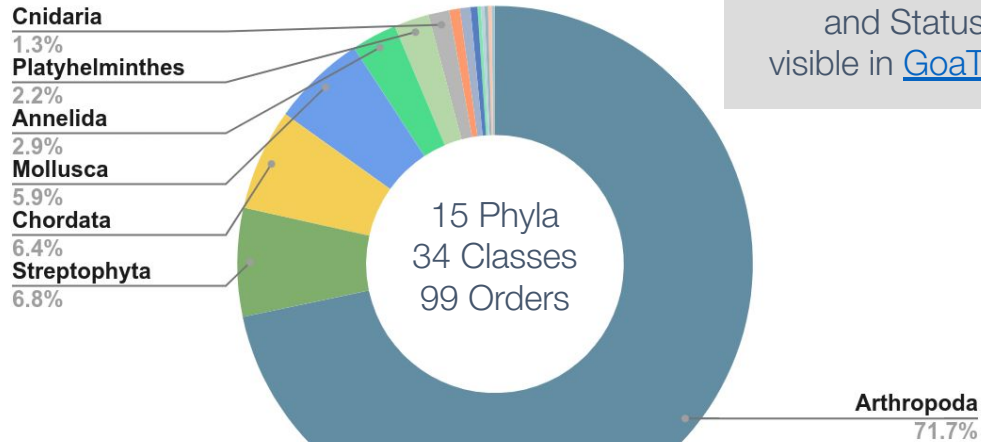
ca. **171** species shipped to Sequencing Facilities



Expected output

275 Gb of genome span produced

Phyla



- ➔ Community engagement presents challenges
- ➔ Strict deadline enforcement is essential
- ➔ Successfully engaged a wide network beyond BGE
- ➔ Coordinating sample shipments across Europe is complex
- ➔ Established clear workflows and SOPs
- ➔ Automation needed for ERGA Sample Manifest submissions

GENOME REPORT

ERGA-BGE genome of *Valencia hispanica* (Valenciennes, 1826): a critically endangered Iberian toothcarp.

Marc Ventura¹, Nati Franch², Rosa Fernández³, Javier Palma-Guerrero^{3,4}, Astrid Böhlme⁵, Rita Monteiro⁶, Laura Aguilera⁷, Marta Gut⁸, Tyler S. Alioto⁹, Francisco Cámara Ferreira², Fernando Cruz², Jéssica Gómez-Garrido¹, Leanne Haggerty², Fergal Martin², Thomas Brown^{10*}

¹ Centre for Advanced Studies of Blanes, Spanish National Research Council (CEAB-CSIC), Accés Cala Sant Francesc 14, 17200 Blanes, Catalonia, Spain
² Parc Natural del Delta de l'Ebre, Av. Catalunya, 44, 43580, Deltebre, Tarragona, Catalonia, Spain
³ Metazoa Phylogenomics Lab, Biodiversity Program, Institute of Evolutionary Biology (IBE, CSIC-UPF), Passeig marítim de la Barceloneta 37-49, 08003, Barcelona, Spain
⁴ Current address: FiBL, Research Institute of Organic Agriculture, Ackerstrasse 113, 5070 Frick, Switzerland
⁵ Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, 53119 Bonn, Germany
⁶ Centro Nacional de Análisis Genómico (CNAG), Baldiri Reixac, 4, 08028 Barcelona, Spain, Universitat de Barcelona (UB), Barcelona, Spain
⁷ European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
⁸ Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalle-Strasse 17, 10315 Berlin, Germany
⁹ Berlin Center for Genomics in Biodiversity Research (BCGenD), Konigin-Luise-Str 6-8, 14195 Berlin, Germany
¹⁰ To whom correspondence should be addressed: brown@izw-berlin.de

Abstract
 The reference genome of *Valencia hispanica*, a critically endangered actinopterygian species endemic to the Iberian Peninsula, is key to unravelling its genetic architecture and adaptation to freshwater ecosystems. This genomic resource will enable targeted conservation efforts and shed light on the species' essential role in ecological dynamics, including its contributions to algal biomass regulation and role in the aquatic food web while also highlighting the challenges it faces from habitat degradation and invasive species. Furthermore, it offers opportunities to gain valuable insights into the evolutionary paths within the Valenciidae family, significantly advancing our comprehension of genetic diversity and adaptability in aquatic ecosystems. The entirety of the genome sequence was assembled into 24 contiguous chromosomal pseudomolecules. This



Source: <https://paisajesvivos.org/ca/objetos/1105>



Image data available in the BioImage Archive (<http://www.ebi.ac.uk/bioimage-archive>) under accession number S-BIAD1012.





The ERGA Genome Tracking Console

Tyler Alioto

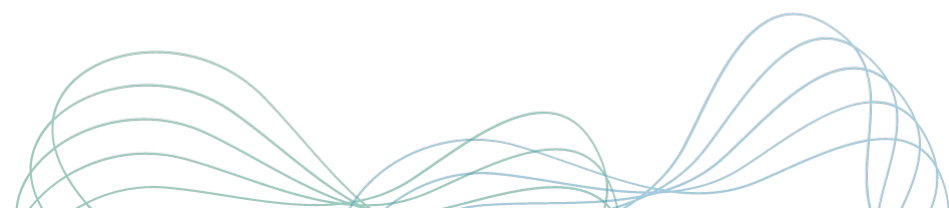
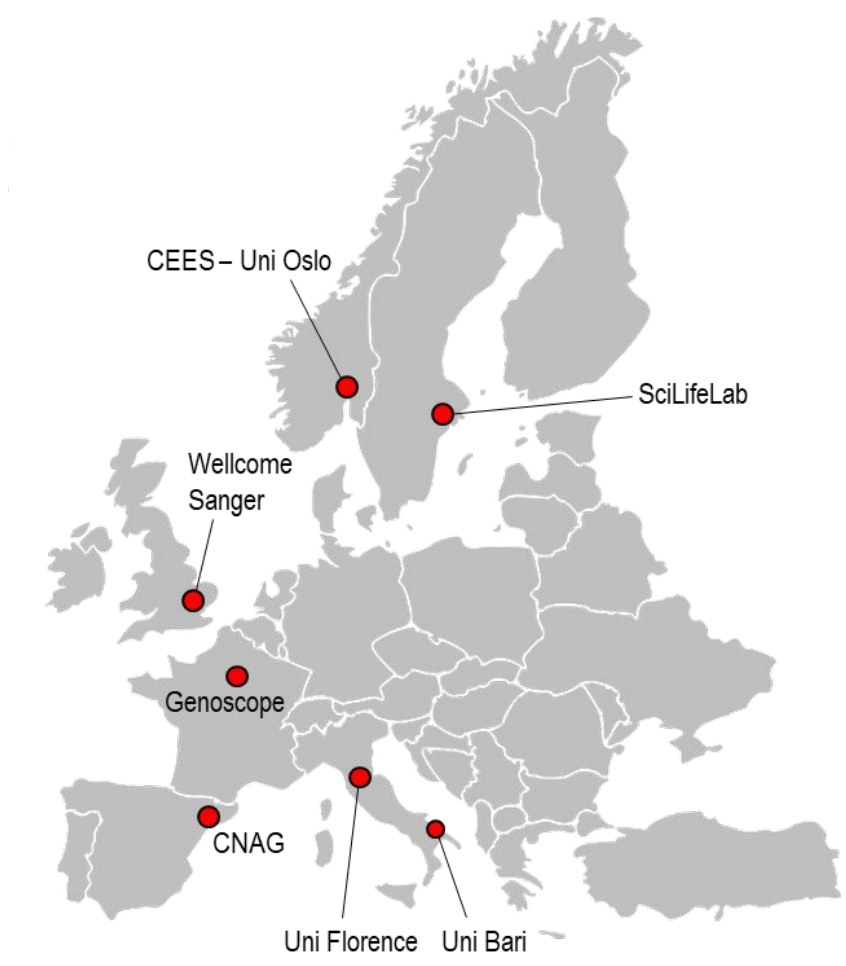
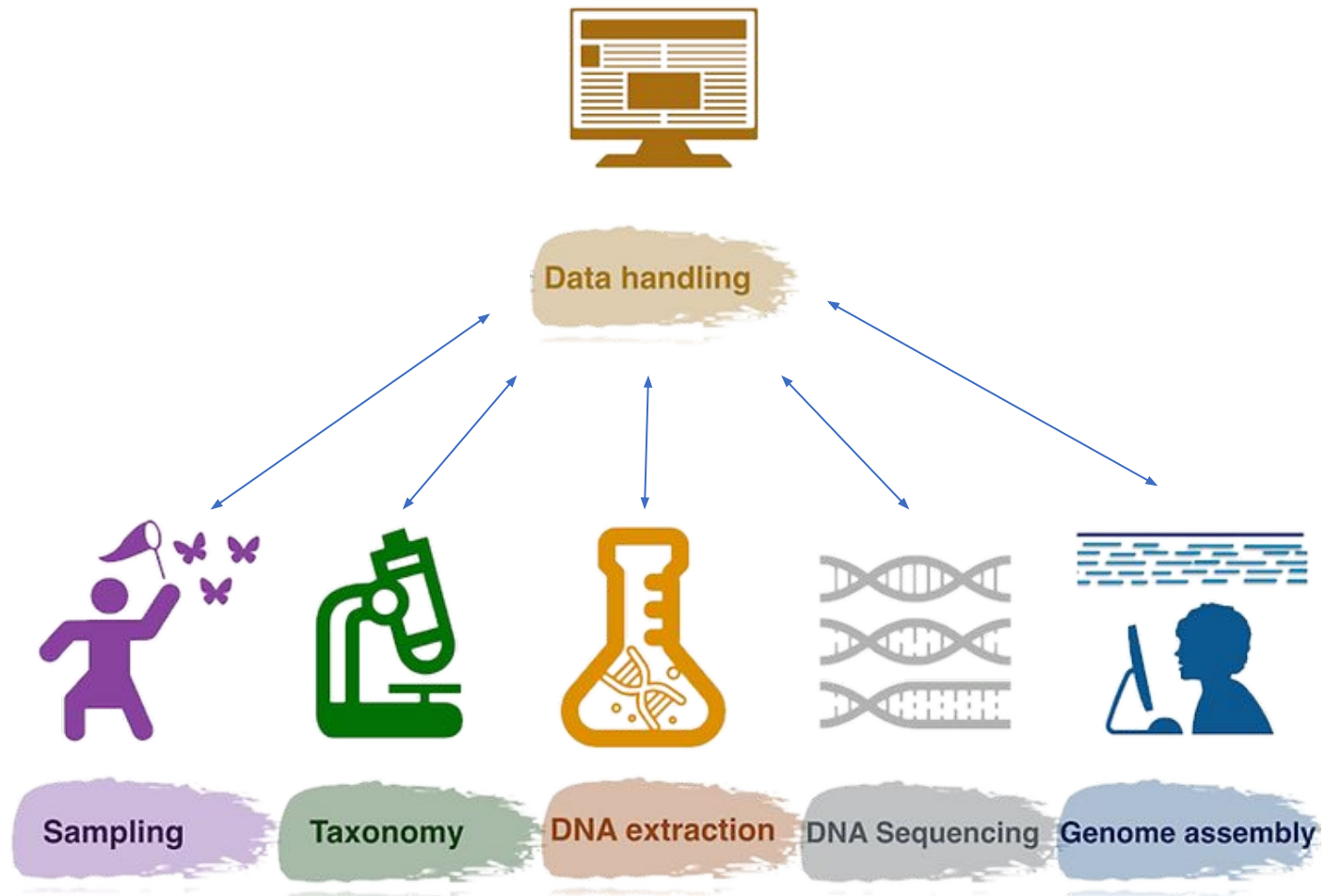




The ERGA Genome Tracking Console

Tyler Alioto

How do we manage and track the progress of reference genome generation?

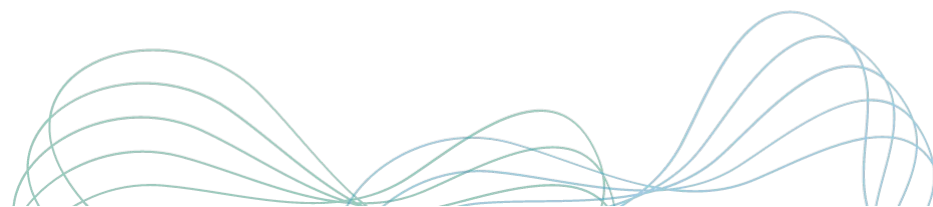


The background image is a blurred screenshot of a data table with the following columns: First Name, Last Name, Phone, Gender, DOB, Email, and Address. The rows contain various names and contact information, such as Helene Hutchison, Eric M.J., and Jane Smith.



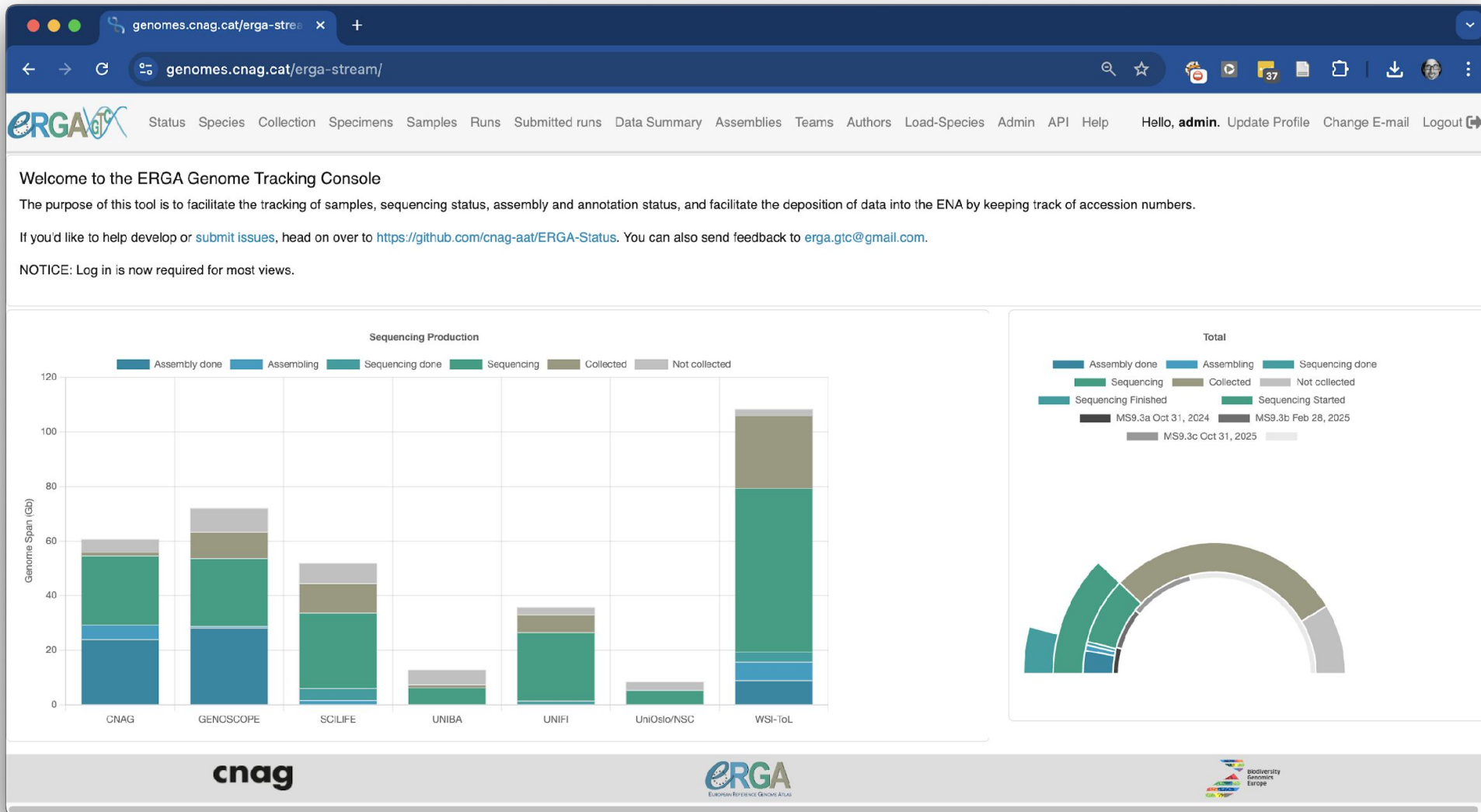
	A	B	C	D	E	F	G	H	I	J	K
1	TUBE_OR_WELL_ID	SPECIMEN_ID	PURPOSE_OF_SPECIMEN	SAMPLE_COORDINATOR	SAMPLE_COORDINATOR_AFFILIATION	SAMPLE_COORDINATOR_ORCID_ID	ORDER_OR_GROUP	FAMILY	GENUS	TAXON_ID	SCIENTIFIC_NAME
2	FF06763793	ERGA_FOS_4557_001	REFERENCE_GENOME	FELIX SHAW	EARLHAM_INSTITUTE	0000-0001-9649-5906	Capitellida	Arenicolidae	Arenicola	6344	Arenicola marina
3											

ERGA Manifest v2.5

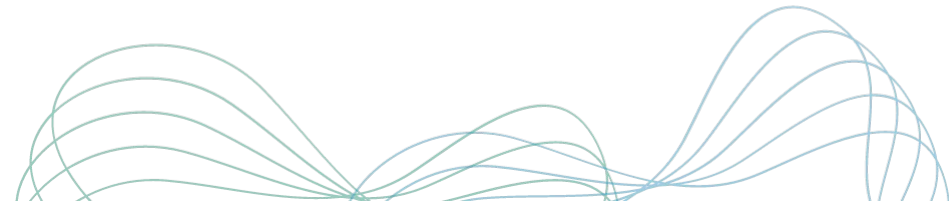
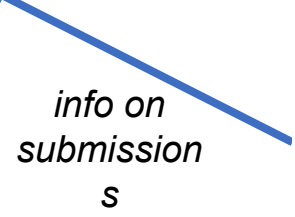
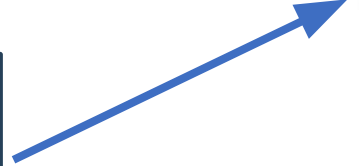
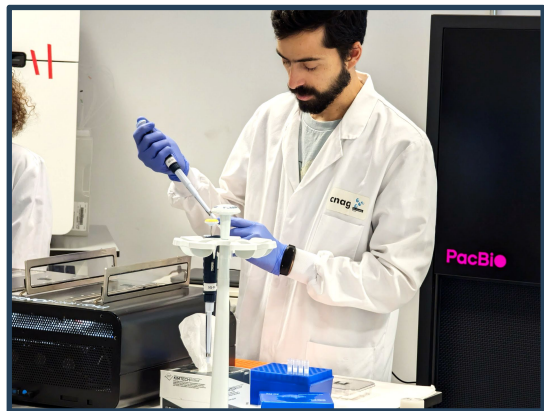
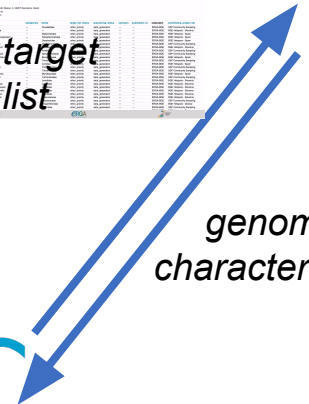
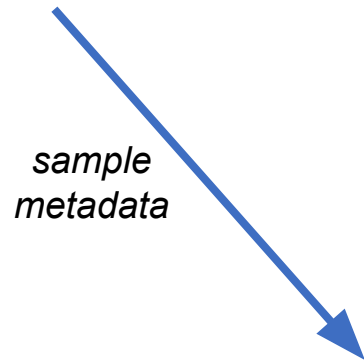


The screenshot displays a complex interface with multiple overlapping Excel spreadsheets. The primary spreadsheet, titled "BGE biodiversity hotspots species list phase 1A and 1B", contains a detailed list of species with the following columns: Species, Phylum/class/family, Kingdom, Source, NCBI taxon ID, Proposed sequencing centre, Sequencing status, Goat Stimated Genome Size, Phase, and Notes. The list includes species such as *Acomys cahirinus_minous*, *Apodemus agrarius*, *Astagobius angustatus*, and *Zelkova abelicea*.

Other visible spreadsheets include "ALLOCATIONS FIXED NOTE" (with a red header), "Bioblitz site" (listing Hungary), and "BGE-WP5_Species_list_GOAT". The interface also shows various Excel toolbars, search bars, and window management elements.

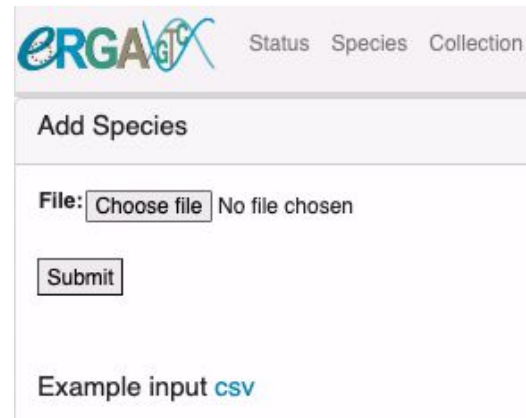
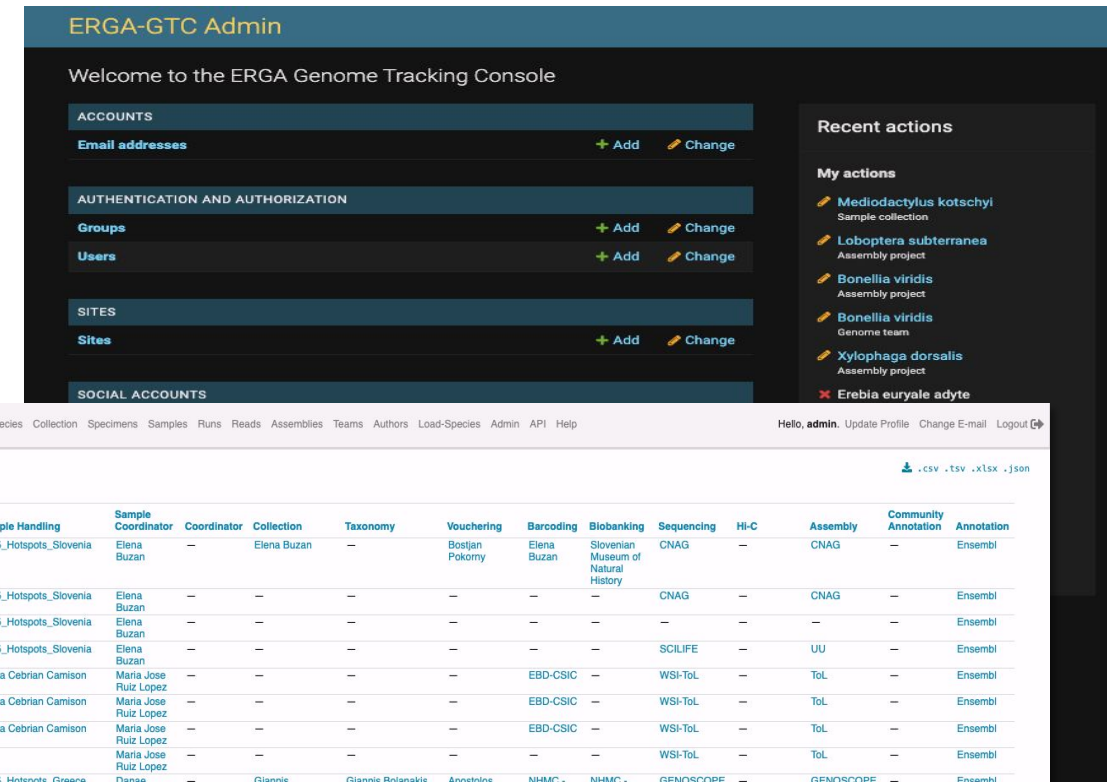


Programmatic interface - REST API
using Perl/Python for syncing with
different systems



Used by:

- **Sampling management**
 - Sample coordinators
 - Hotspot coordinators
 - Community sampling coordinators
 - Etc.
- **Genome team/project management**
- **Sequencing units**
 - SciLife/UU
 - UNIFI
 - WSI-ToL
 - GENOSCOPE
 - CNAG
 - UniOslo
- **Assembly teams**
- **Annotation teams**



The screenshot shows the main data table in the ERGA-GTC interface. The table has columns for Species, Sample Handling, Sample Coordinator, Coordinator, Collection, Taxonomy, Vouchering, Barcoding, Biobanking, Sequencing, Hi-C, Assembly, Community Annotation, and Annotation. It lists various species such as Rupicapra rupicapra, Apodemus agrarius, and Dendarus foraminosus, along with their respective handling and tracking details.

Species	Sample Handling	Sample Coordinator	Coordinator	Collection	Taxonomy	Vouchering	Barcoding	Biobanking	Sequencing	Hi-C	Assembly	Community Annotation	Annotation
Rupicapra rupicapra	WP5_Hotspots_Slovenia	Elena Buzan	-	Elena Buzan	-	Bosjan Pokorny	Elena Buzan	Slovenian Museum of Natural History	CNAG	-	CNAG	-	Ensembl
Apodemus agrarius	WP5_Hotspots_Slovenia	Elena Buzan	-	-	-	-	-	-	CNAG	-	CNAG	-	Ensembl
Umbra krameri	WP5_Hotspots_Slovenia	Elena Buzan	-	-	-	-	-	-	-	-	-	-	Ensembl
Astagobius angustatus	WP5_Hotspots_Slovenia	Elena Buzan	-	-	-	-	-	-	SCILIFE	-	UU	-	Ensembl
Culex laticinctus	Sonia Cebrían Camison	Maria Jose Ruiz Lopez	-	-	-	-	EBD-CSIC	-	WSI-ToL	-	ToL	-	Ensembl
Culex modestus	Sonia Cebrían Camison	Maria Jose Ruiz Lopez	-	-	-	-	EBD-CSIC	-	WSI-ToL	-	ToL	-	Ensembl
Culex perexiguus	Sonia Cebrían Camison	Maria Jose Ruiz Lopez	-	-	-	-	EBD-CSIC	-	WSI-ToL	-	ToL	-	Ensembl
Culex theileri	-	Maria Jose Ruiz Lopez	-	-	-	-	-	-	WSI-ToL	-	ToL	-	Ensembl
Dendarus foraminosus	WP5_Hotspots_Greece	Danae Karakasi	-	Giannis Bolanakis	Giannis Bolanakis	Apostolos Trichas	NHMC - Greece	NHMC - Greece	GENOSCOPE	-	GENOSCOPE	-	Ensembl
Dalognatha quadricollis	WP5_Hotspots_Greece	Danae Karakasi	-	Katerina Vardinogiannis	Katerina Vardinogiannis	Katerina Vardinogiannis	NHMC - Greece	NHMC - Greece	GENOSCOPE	-	GENOSCOPE	-	Ensembl
Scelopendra cretica	WP5_Hotspots_Greece	Danae Karakasi	-	Giannis Bolanakis	Giannis Bolanakis	Apostolos Trichas	NHMC - Greece	NHMC - Greece	GENOSCOPE	-	GENOSCOPE	-	Ensembl
Zeikova abelicea	WP5_Hotspots_Greece	Danae Karakasi	-	Eleftheria Antaloudaki	Eleftheria Antaloudaki	Eleftheria Antaloudaki	NHMC - Greece	NHMC - Greece	GENOSCOPE	-	GENOSCOPE	-	Ensembl
Acomys minous	WP5_Hotspots_Greece	Danae Karakasi	-	Manolis Papadimitrakis	Petros Lyberakis	Petros Lyberakis	NHMC - Greece	NHMC - Greece	GENOSCOPE	-	GENOSCOPE	-	Ensembl
Cladocora caespitosa	WP5_Hotspots_Slovenia	Elena Buzan	-	-	-	-	-	-	SCILIFE	-	UU	-	Ensembl
Pinna rudis	WP5_Hotspots_Spain_1	Javier Palma Guerrero	-	Diego Kersting (IATS-CSIC)	Diego Kersting (IATS-CSIC)	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	SCILIFE	-	UU	-	Ensembl
Siphonaria pectinata	WP5_Hotspots_Spain_1	Javier Palma Guerrero	-	Siphonaria pectinata	Fernando Fernández-Álvarez (ICM-CSIC)	Fernando Fernández-Álvarez	MNCN-CSIC	MNCN-CSIC	SCILIFE	-	UU	-	Ensembl
Tetrastemma malincoanthalum	WP5_Hotspots_Spain_1	Javier Palma	-	Tetrastemma malincoanthalum	Fernando Fernández-Álvarez	Fernando Fernández-Álvarez	MNCN-CSIC	MNCN-CSIC	SCILIFE	-	UU	-	Ensembl



ERGA-GTC Overview

genomes.cnag.cat/erga-stream/overview/

Status Species Collection Specimens Samples Runs Submitted runs Data Summary Assemblies Teams Authors Load-Species Admin API Help Hello, tyler@gmail. Update Profile Change E-mail Logout

.csv .tsv .xlsx .json

Scientific name	Toolid prefix	Phylum	Span	Task	Center	GoaT Status	COPO	ONT/HIFI WGS	Illumina WGS	Hi-C	RNA-Seq	Assembly	Anr
<i>Piacobdella costata</i>	wrPlaCost	Annelida	0.597	Community	WSI-ToL	insdc_open	Accepted	Submitted	Received	Submitted	Sequencing	Submitted	
<i>Marifugia cavatica</i>	wsMarCava	Annelida	0.734	Community	WSI-ToL	insdc_open	Accepted	Submitted	Done	Submitted	Submitted	Submitted	
<i>Chaetopelma lymberakisi</i>	qqChaLymb	Arthropoda	1.203	Greece	GENOSCOPE	insdc_open	Accepted	Submitted		Submitted	Prep	Submitted	
<i>Gluvia dorsalis</i>	qqGluDors	Arthropoda	1.100	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Loboptera subterranea</i>	lbLobSubt	Arthropoda	1.491	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Done	Submitted	
<i>Dendarus foraminosus</i>	icDenFora	Arthropoda	0.416	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Tarphius canariensis</i>	icTarCana	Arthropoda	0.450	Spain	WSI-ToL	insdc_open	Accepted	Sequencing	Done	Submitted	Submitted	Submitted	
<i>Culex laticinctus</i>	idCulLati	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	Submitted	Submitted	Submitted	
<i>Culex modestus</i>	idCulMode	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	Submitted	Submitted	Submitted	
<i>Culex perexiguus</i>	idCulPerx	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	Submitted	Submitted	Submitted	
<i>Erebia palarica</i>	ilErePala	Arthropoda	0.479	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Graellsia isabellae</i>	ilGraIsab	Arthropoda	0.650	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Valencia hispanica</i>	lValHis	Chordata	1.501	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Malpolor monspessulanus</i>	rMalMon	Chordata	1.477	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Done	Submitted	
<i>Rupicapra rupicapra</i>	mRupRup	Chordata	3.196	Slovenia	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Acomys minous</i>	mAcoMin	Chordata	2.582	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Sequencing	Submitted	Submitted	
<i>Apodemus agrarius</i>	mApoAgr	Chordata	3.042	Slovenia	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Diadema setosum</i>	eeDiaSeto	Echinodermata	1.174	Greece	GENOSCOPE	insdc_open	Accepted	Sequencing	Submitted	Sequencing	Submitted	Submitted	
<i>Arca noae</i>	xbArcNoae	Mollusca	1.623	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Sequencing	Submitted	Submitted	
<i>Pinctada radiata</i>	xbPinRadi	Mollusca	1.154	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Submitted	Sequencing	Submitted	
<i>Stigmatoteuthis arcturi</i>	xcStiArct	Mollusca	3.203	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Done	Submitted	
<i>Anisus vorticulus</i>	xgAniVori	Mollusca	1.311	Community	WSI-ToL	insdc_open	Accepted	Submitted	Submitted	Submitted	Submitted	Submitted	
<i>Phyllidia flava</i>	xgPhyFlav	Mollusca	1.956	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	Submitted	Done	Submitted	
<i>Deroceras lasithiense</i>	xgDerLasi	Mollusca	1.643	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Sequencing	Prep	Submitted	
<i>Albinaria teres</i>	xgAlbTers	Mollusca	2.846	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Sequencing	Issue	Submitted	
<i>Cymodocea nodosa</i>	laCymNodo	Streptophyta	0.421	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	Submitted	Issue	Submitted	

cnag ERGA University of Genoa, Europe



The screenshot displays the ERGA-GTC Overview interface. On the left, there is a sidebar with filters for Species, ToLID Prefix, Task, Center, GoaT Sequencing Status, Assembly Status, Annotation Status, Community Annotation Status, Kingdom, Phylum, Class, Order, Family, and Genus. The main area contains a table with the following columns: Scientific name, ToLID prefix, Phylum, Span, Task, Center, GoaT Status, COPO, ONT/HIFI WGS, Illumina WGS, and HI. The table lists various species and their associated genomic data and sequencing status.

Scientific name	ToLID prefix	Phylum	Span	Task	Center	GoaT Status	COPO	ONT/HIFI WGS	Illumina WGS	HI
Placobdella costata	wrPlaCost	Annelida	0.597	Community	WSI-ToL	insdc_open	Accepted	Submitted	Received	
Marifugia cavatica	wsMarCava	Annelida	0.734	Community	WSI-ToL	insdc_open	Accepted	Submitted	Done	
Chaetopelma lymberakisi	qqChaLymb	Arthropoda	1.203	Greece	GENOSCOPE	insdc_open	Accepted	Submitted		
Gluvia dorsalis	qqGluDors	Arthropoda	1.100	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	
Loboptera subterranea	ibLobSubt	Arthropoda	1.491	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	
Dendarius foraminosus	icDenFora	Arthropoda	0.416	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Tarphius canariensis	icTarCana	Arthropoda	0.450	Spain	WSI-ToL	insdc_open	Accepted	Sequencing	Done	
Culex laticinctus	icCulLati	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	
Culex modestus	icCulMode	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	
Culex perexiguus	icCulPerx	Arthropoda	0.763	Cases	WSI-ToL	insdc_open	Accepted	Submitted	Done	
Erebria palerica	ilErePala	Arthropoda	0.479	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	
Graellsia isabellae	ilGraIsab	Arthropoda	0.650	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	
Valencia hispanica	lValHis	Chordata	1.501	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	
Malpolon monspessulanus	rMalMon	Chordata	1.477	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	
Rupicapra rupicapra	mRupRup	Chordata	3.196	Slovenia	CNAG	insdc_open	Accepted	Submitted	Submitted	
Acomys minous	mAcoMin	Chordata	2.582	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Apodemus agrarius	mApoAgr	Chordata	3.042	Slovenia	CNAG	insdc_open	Accepted	Submitted	Submitted	
Diadema setosum	eeDiaSeto	Echinodermata	1.174	Greece	GENOSCOPE	insdc_open	Accepted	Sequencing	Submitted	
Arca noae	xbArcNoae	Mollusca	1.623	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Pincltada radiata	xbPinRadi	Mollusca	1.154	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Stigmatoteulhis arcturi	xcStiArct	Mollusca	3.203	Spain	CNAG	insdc_open	Accepted	Submitted	Submitted	
Anisus vorticulus	xgAniVori	Mollusca	1.311	Community	WSI-ToL	insdc_open	Accepted	Submitted	Submitted	
Phyllidia fleva	xgPhyFlav	Mollusca	1.956	Community	CNAG	insdc_open	Accepted	Submitted	Submitted	
Deroceeras lasithionense	xgDerLasi	Mollusca	1.643	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Albinaria teres	xgAlbTers	Mollusca	2.846	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	
Cymodocea nodosa	laCymNodo	Streptophyta	0.421	Greece	GENOSCOPE	insdc_open	Accepted	Submitted	Submitted	



The image shows a stack of browser windows displaying the ERGA-GTC interface. The windows are overlapping, showing different views of the system:

- ERGA-GTC Species:** A sidebar menu on the left lists various categories like Affiliations, Annotation team, Assemblies, etc. The main content area shows a list of species with columns for project name, subproject name, primary contact, date of update, and schema version.
- ERGA-GTC GoAT List:** A table showing subproject and species information. Columns include subproject (e.g., ERGA-BGE) and species (e.g., Troglodytes, Nelima, Tegenaria).
- ERGA-GTC Sample Collection:** A table showing specimen details. Columns include specimen id (e.g., ERGA_BJ_711X_0013) and species.
- ERGA-GTC GoAT Samples:** A detailed table of samples. The table has the following columns: CopolID, BioSample, SampleDerivedFrom, SampleSameAs, Purpose of specimen, GAL, Gal sample id, Collector sample id, Tube or well id, Corrected id, Date, and Status. The data shows two rows of samples, both with a status of 'accept'.

ERGA-GTC Genome Teams

genomes.cnag.cat/erga-stream/teams/?sort=-collection_team

ERGA GTC Status Species Collection Specimens Samples Runs Submitted runs Data Summary Assemblies Teams Authors Load-Species Admin API Help Hello, tyler-gmail. Update Profile Change E-mail Logout

.csv .tsv .xlsx .json

Species	Sample Handling	Sample Coordinator	Coordinator	Collection	Taxonomy	Vouchering	Barcoding	Biobanking	Sequencing	Hi-C	Assembly	Community Annotation	Annotation
Tetrastemma melanocephalum	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Tetrastemma melanocephalum	Fernando Fernández-Álvarez (ICM-CSIC)	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	SCILIFE	—	UU	—	Ensembl
Stigmatoteuthis arcturi	WP5_Hotspots_Spain_2	Nuria Escudero	—	Stigmatoteuthis arcturi	Fernando Fernández-Álvarez (ICM-CSIC)	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	CNAG	—	CNAG	—	Ensembl
Siphonaria pectinata	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Siphonaria pectinata	Fernando Fernández-Álvarez (ICM-CSIC)	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	SCILIFE	—	UU	—	Ensembl
Munidopsis polymorpha	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Rosa Fernandez	Rosa Fernandez	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	WSI-ToL	—	ToL	—	Ensembl
Conus guanche	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Rafael Zardoya	Rafael Zardoya	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	WSI-ToL	—	ToL	—	Ensembl
Gyrocaryum oppositifolium	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Pablo Vargas	Pablo Vargas	—	—	—	WSI-ToL	—	ToL	—	Ensembl
Obama nungara	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Obama nungara	Obama nungara	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	SCILIFE	—	UU	—	Ensembl
Graellsia isabellae	WP5_Community_Samples	Marta Vila	—	Neus Mari-Mena	Roger Vila	MNCN-CSIC	Daniel Garcia-Souto	MNCN-CSIC	CNAG	—	CNAG	—	Ensembl
Allolobophora mollerii	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Marta Novo	Alejandro Martínez Navarro	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	WSI-ToL	—	ToL	—	Ensembl
Valencia hispanica	WP5_Hotspots_Spain_1	Javier Palma Guerrero	—	Marc Ventura	Marc Ventura	MNCN-CSIC	MNCN-CSIC	MNCN-CSIC	CNAG	—	CNAG	—	Ensembl
Acomys minous	WP5_Hotspots_Greece	Danae Karakasi	—	Manolis Papadimitrakis	Petros Lymberakis	Petros Lymberakis	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Erebia palarica	WP5_Community_Samples	Marta Vila	—	Laura Torrado-Blanco	Nils Ryrholm	MNCN-CSIC	Marta Vila	MNCN-CSIC	CNAG	—	CNAG	—	Ensembl
Albinaria teres	WP5_Hotspots_Greece	Danae Karakasi	—	Katerina Vardinogiannis	Katerina Vardinogiannis	Katerina Vardinogiannis	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Dailognatha quadricollis	WP5_Hotspots_Greece	Danae Karakasi	—	Katerina Vardinogiannis	Katerina Vardinogiannis	Katerina Vardinogiannis	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Deroceras lasithionense	WP5_Hotspots_Greece	Danae Karakasi	—	Katerina Vardinogiannis	Katerina Vardinogiannis	Katerina Vardinogiannis	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Scolopendra cretica	WP5_Hotspots_Greece	Danae Karakasi	—	Giannis Bolanakis	Giannis Bolanakis	Apostolos Trichas	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Carabus banonii	WP5_Hotspots_Greece	Danae Karakasi	—	Giannis Bolanakis	Giannis Bolanakis	Apostolos Trichas	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Dendarus foraminosus	WP5_Hotspots_Greece	Danae Karakasi	—	Giannis Bolanakis	Giannis Bolanakis	Apostolos Trichas	NHMC - Greece	NHMC - Greece	GENOSCOPE	—	GENOSCOPE	—	Ensembl
Rupicapra rupicapra	WP5_Hotspots_Slovenia	Elena Buzan	—	Elena Buzan	—	Bostjan Pokorny	Elena Buzan	Slovenian Museum of Natural	CNAG	—	CNAG	—	Ensembl

cnag ERGA European Research Genomes Alliance

University Genomes Europe

The screenshot displays the ERGA-GTC web application interface across three overlapping browser windows. The top window, titled 'ERGA-GTC Assemblies', shows a table of genome assemblies with columns for project name, description, pipeline, type, span, contig/scaffold sizes, BUSCO scores, and reports. The middle window, 'ERGA-GTC Read Data', lists sequencing projects for various species. The bottom window, 'ERGA-GTC Runs', provides a detailed list of sequencing runs for each project.

Assembly project	Description	Assembly pipeline	Main options if not default	Type	Span (Gb)	Contig N50 (Mb)	Scaffold N50 (Mb)	Chr level	Pct. placed	BUSCO	BUSCO db	BUSCO version	QV	Report
Valencia hispanica	fValHis1.1 assembly for Valencia hispanica	CLAWS.2.1	—	Primary	1.292	38.297	56.938	✓	100.0	C:98.6% [S:97.8%,D:0.8%],F:0.4%,M:1.0%,n:3640	actinopterygii_odb10	5.4.0	50.60	https://github.com/ERGA-consortium/EARs/blob/main/Ass
Rupicapra rupicapra	mRupRup1.1 assembly for Rupicapra rupicapra	CLAWS.2.1	—	Primary	2.623	77.662	100.920	✓	—	C:97.7% [S:93.8%,D:4.1%],F:1.1%,M:1.2%,n:3354	vertebrata_odb10	5.4.0	55.40	https://github.com/ERGA-consortium/EARs/blob/main/Ass
Apodemus agrarius	mApoAgr2.1 assembly for Apodemus agrarius	CLAWS.2.1	—	Primary	2.614	35.189	119.125	✓	—	C:98.3% [S:96.4%,D:1.9%],F:0.5%,M:1.2%,n:3354	vertebrata_odb10	5.4.0	47.20	https://github.com/ERGA-consortium/EARs/blob/main/Ass
Erebia palarica	iErePala4.1 assembly for Erebia palarica	CLAWS.2.2	—	Primary	0.493	34.188	38.442	✓	—	C:97.9% [S:96.6%,D:1.3%],F:0.8%,M:1.3%,n:1013	arthropoda_odb10	5.4.0	44.68	https://github.com/ERGA-consortium/EARs/blob/main/Ass
Gluvia dorsalis	qqGluDors1.3 assembly for Gluvia dorsalis	CLAWS.2.2	—	Primary	0.787	37.604	198.510	✓	—	C:98.1% [S:97.3%,D:0.8%],F:1.4%,M:0.5%,n:1013	arthropoda_odb10	5.4.0	48.10	https://github.com/ERGA-consortium/EARs/blob/main/Ass
Culex laticinctus	idCulLati1.1 assembly for Culex laticinctus	—	—	Primary	0.834	0.434	291.895	✓	—	—	—	—	—	—
Culex modestus	idCulMode1.1 assembly for Culex modestus	—	—	Primary	0.698	0.313	246.157	✓	—	—	—	—	—	—
Culex perexiguus	idCulPerx1.1 assembly for Culex perexiguus	—	—	Primary	0.436	0.473	150.417	✓	—	—	—	—	—	—
Arca noae	xbArcNoae1 assembly for Arca noae	—	—	Primary	1.496	20.465	84.746	✓	—	—	—	—	—	—
Pinctada radiata	xbPinRad1 assembly for Pinctada radiata	—	—	Primary	0.931	8.065	63.837	✓	—	—	—	—	—	—
Dendarus foraminosus	idDenFora10 assembly for Dendarus foraminosus	—	—	Primary	0.594	24.404	51.873	✓	—	—	—	—	—	—
Anisus vorticulus	xgAniVori1.1 assembly for Anisus vorticulus	—	—	Primary	1.043	2.353	59.244	✓	—	—	—	—	—	—
Graellsia	iGralSab1.1	CLAWS.2.2	—	Primary	0.561	18.864	20.418	✓	—	98.9% Single	insecta_odb10	5.4.0	49.10	https://github.com/ERGA-consortium/EARs/blob/main/Ass

The screenshot shows a web browser window with the URL `genomes.cnag.cat/erga-stream/authors/?species=32&author=&role=`. The page header includes the ERGA logo and navigation links: Status, Species, Collection, Specimens, Samples, Runs, Submitted runs, Data Summary, Assemblies, Teams, Authors, Load-Species, Admin, API, Help. A user greeting says "Hello, tyler-gmail." with links for "Update Profile", "Change E-mail", and "Logout".

Below the header is a table with columns: Species, Role, Author, Affiliation, and ORCID. The table lists 17 authors for the species "Valencia hispanica".

Species	Role	Author	Affiliation	ORCID
Valencia hispanica	sequencing	Gut,Marta	Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain	0000-0002-4063-7159
Valencia hispanica	assembly	Alioto,Tyler S.	Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain	0000-0002-2960-5420
Valencia hispanica	assembly	Câmara Ferreira,Francisco	Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain	0000-0002-1971-5466
Valencia hispanica	assembly	Gómez-Garrido,Jêssica	Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain	0000-0001-6409-8009
Valencia hispanica	assembly	Cruz,Fernando	Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain	0000-0003-4098-8829
Valencia hispanica	sample handling	Monteiro,Rita	Leibniz Institute for the Analysis of Biodiversity Change - Zoological Research Museum Alexander Koenig	0000-0003-1374-4474
Valencia hispanica	sample handling	Boehne,Astrid	Leibniz Institute for the Analysis of Biodiversity Change - Zoological Research Museum Alexander Koenig	0000-0002-1284-3115
Valencia hispanica	sample handling	Palma Guerrero,Javier	Institut de Biologia Evolutiva (UPF-CSIC)	0000-0002-0867-7601
Valencia hispanica	sample handling	Fernández,Rosa	Institut de Biologia Evolutiva (UPF-CSIC)	0000-0002-4719-6640
Valencia hispanica	taxonomic identification	Ventura,Marc	CEAB-CSIC, Center for Advanced Studies of Blanes	0000-0003-1401-414X
Valencia hispanica	sample collection	Ventura,Marc	CEAB-CSIC, Center for Advanced Studies of Blanes	0000-0003-1401-414X
Valencia hispanica	biobanking	Conejero,María	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-1138-7162
Valencia hispanica	biobanking	Riesgo,Ana	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-7993-1523
Valencia hispanica	vouchering	Conejero,María	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-1138-7162
Valencia hispanica	vouchering	Riesgo,Ana	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-7993-1523
Valencia hispanica	barcoding	Conejero,María	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-1138-7162
Valencia hispanica	barcoding	Riesgo,Ana	Museo Nacional de Ciencias Naturales- CSIC, Madrid	0000-0002-7993-1523

Displaying 1 to 17 of 17.

The footer contains the logos for "cnag", "ERGA EUROPEAN REFERENCE GENOME ATLAS", and "Biodiversity Genomics Europe".





Production instance:

<https://genomes.cnag.cat/erga-stream>



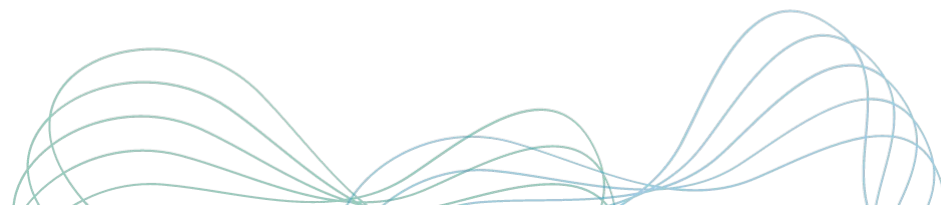
Development instance:

<https://genomes.cnag.cat/erga-stream-dev>



Code:

<https://github.com/cnag-aat/ERGA-Status>





A Decentralised Method for
Community-Reviewing
Earth BioGenomes

Diego de Panis

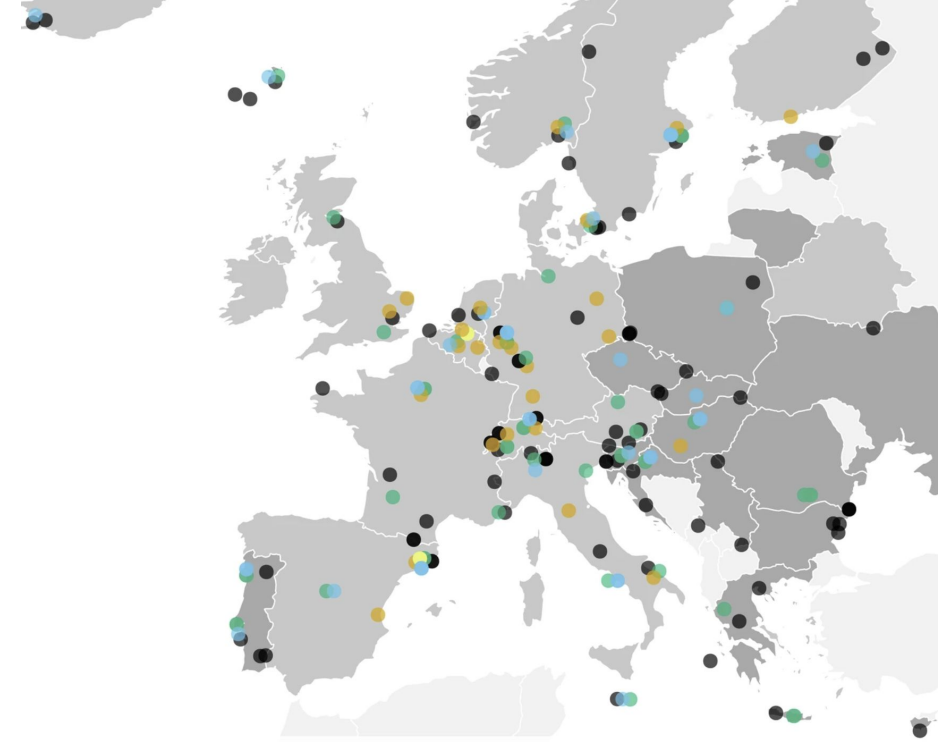




**A Decentralised Method for
Community-Reviewing Earth BioGenomes**

Diego De Panis

Genome assembly production takes place within a dynamic and varied scenario:



- Distributed across multiple sequencing centres, institutes, and labs
- Encompassing a diverse range of genera across the Tree of Life
- Using combinations of technologies and throughputs
- Applying unique selections of tools for data processing and evaluation

Drawing the same finish line for all:

What defines a high-quality genome assembly?

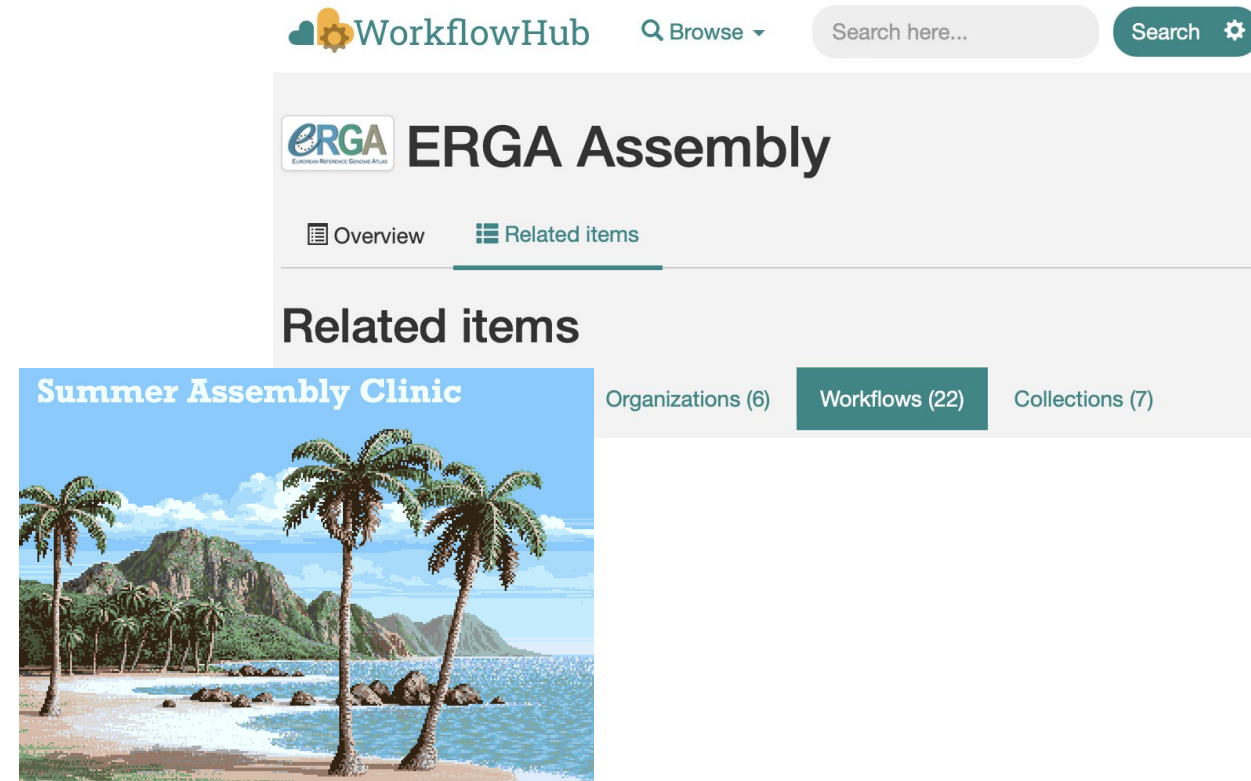
EBP Assembly Standard v5

sufficient DNA/tissue		limited DNA/tissue
chrom N50 \leq 1 Mpb	chrom N50 $>$ 1 Mpb	
C.C.Q40	6.C.Q40	5.C.Q40
<ul style="list-style-type: none"> < 1,000 Gaps/Gbp > 90% kmer completeness > 90% assigned to candidate chromosomes > 90% single copy BUSCO < 5% duplicated BUSCO 		

How to quickly and consistently assess the quality of a genome assembly?

The ERGA Sequencing and Assembly Community

- Facilitating collaborative spaces
- Promoting discussions
- Sharing knowledge
- Providing support



The screenshot shows the WorkflowHub interface for the ERGA Assembly community. At the top, there is a WorkflowHub logo, a 'Browse' dropdown, and a search bar. Below this is the ERGA logo and the title 'ERGA Assembly'. There are two tabs: 'Overview' and 'Related items', with 'Related items' being the active tab. Under the 'Related items' section, there are three filters: 'Organizations (6)', 'Workflows (22)', and 'Collections (7)'. The 'Workflows (22)' filter is highlighted. Below the filters is a large image titled 'Summer Assembly Clinic' showing a tropical beach scene with palm trees and a mountain in the background.

The community actively works on solutions to its own challenges

Which QC outputs are essential to evaluate the quality of a genome assembly?



Standardized PDF displaying key information to confirm that required quality standards are met

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	2053936
ToLID	xcStiArct1
Species	Stigmatoteuthis arcturi
Class	Cephalopoda
Order	Oegopsida

Genome Traits	Expected	Observed
Haploid size (bp)	3,669,337,499	3,249,387,216
Haploid Number	6 (source: ancestor)	46
Floidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7-Q43

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- Observed Haploid Number is different from Expected
- Kmer completeness value is less than 90 for collapsed

Curator notes

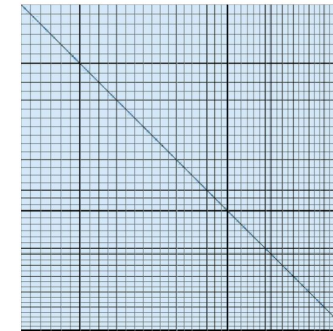
- Interventions/Gb: 9
- Contamination notes: "No clear contaminants were found. However, we observed that BlobTools produces best hits to species other than molluscs in a small percentage of the assembly. Upon closer scrutiny of some unlocs or unplaced scaffolds matching other taxa, we found that many of these sequences actually either do match mollusca sequence but with lower scores or match several different unrelated taxa. We suspect that the main issue is that the mollusca databases are quite incomplete, and most data represents a limited number of organisms that could be evolutionarily distant from this species"
- Other observations: "The curation process was complicated by the fairly poor quality of the Hi-C data we received, resulting in a sparser pretext map (approximately 3.5 million valid pairs, including cis >> 1kb and trans, per Gbase of sequence) compared to most projects with better Hi-C libraries. However, the pre-scaffolded NextDenovo assembly was already quite good, which helped us generate a draft assembly that meets most of the stringent EBP requirements. In this final version of the ERG report we addressed all of the reviewer concerns and produced a new pretext map and assembly based on that. The blobtools plot was not modified as that was deemed not to be necessary."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,249,385,616	3,249,387,216
GC %	34.92	34.92
Gaps/Gbp	234.81	237.28
Total gap bp	152,600	154,200
Scaffolds	503	497
Scaffold N50	73,833,701	74,550,748
Scaffold L50	19	19
Scaffold L90	42	42
Contigs	1,266	1,268
Contig N50	10,980,000	10,980,000
Contig L50	88	88
Contig L90	343	344
QV	43.3597	43.3596
Kmer compl.	79.2506	79.2506
BUSCO sing.	93.8%	93.8%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	1.9%	1.9%
BUSCO miss.	3.4%	3.4%

BUSCO 5.4.0 Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly



collapsed [\[LINK\]](#)

- . GoAT
- . NCBI
- . Genomescope
- . Curation notes
- . gfastats
- . Merqury
- . BUSCO
- . HiC contact maps
- . Kmer plots
- . Contamination plot
- . Data coverage
- . Tools/Pipelines used

Which QC outputs are essential to evaluate the quality of a genome assembly?



Standardized PDF displaying key information to confirm that required quality standards are met

K-mer spectra of curated assembly

Distribution of k-mer counts per copy numbers found in asm

Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening

collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	ONT	Illumina	OmniC
Coverage	56x	63x	60x

Assembly pipeline

```

- trim_galore
  | _ ver: 0.6.7
  | _ key param: --gzip
  | _ key param: =q 20
  | _ key param: --paired
  | _ key param: retain_unpaired
- filllong
  | _ ver: 0.2.1
  | _ key param: NA
- nextdenovo
  | _ ver: 2.5.0
  | _ key param: NA
- hypo
  | _ ver: 1.0.3
  | _ key param: NA
- purge_dups
  | _ ver: 1.2.6
  | _ key param: NA
- YaHS
  | _ ver: 1.2a
  | _ key param: NA
    
```

Curation pipeline

```

- PretextView
  | _ ver: 0.2.5
  | _ key param: NA
- GRIT_Rapid
  | _ ver: 2.0
  | _ key param: NA
    
```

Submitter: Francisco Camara
Affiliation: CNIG Barcelona
Date and time: 2024-10-16 10:43:32 CEST

- . GoAT
- . NCBI
- . Genomescope
- . Curation notes
- . gfastats
- . Merquy
- . BUSCO
- . HiC contact maps
- . Kmer plots
- . Contamination plot
- . Data coverage
- . Tools/Pipelines used



ERGA Assembly Reporting Tool (EAR) ☆ 🔗 ⌵ 📄 ▶ Run Tool

A tool to compile assembly reports and statistics from assembly pipeline
(Galaxy Version 24.10.15+galaxy0)

Tool Parameters

Sample Information ⌵

Input a ToLID * required

Enter the Species name * required

Enter the Species sex *

XX ⌵

Enter the Submitter name * required

Enter the Affiliation * required

Select a valid tag *

ERGA-BGE ⌵

Sequencing Data Information ⌵

Genome Profiling Information ⌵


Pre-Curation Assembly Information ⌵

Curated Assembly Data Information ⌵

Welcome to the EARs repo wiki!



github.com/ERGA-consortium/EARs/wiki

Create your  using the provided script or with the Galaxy app!



*Galaxy implementation by [Saim Momin](#)

An open, transparent, and traceable review system using EARs

mySpecies_EAR.pdf

- . Fork the EARs repo
- . Add the EAR.pdf to your fork
- . Create a Pull-Request

GitHub repo 

**Open & Interactive
Review takes place**

- . Fix issue
- . Share pretext savestate
- . Regenerate EAR.pdf

✗ Something is not right

✓ All good to upload genome to ENA

The participants of the EAR Review



the **Researcher**



the **Bot**



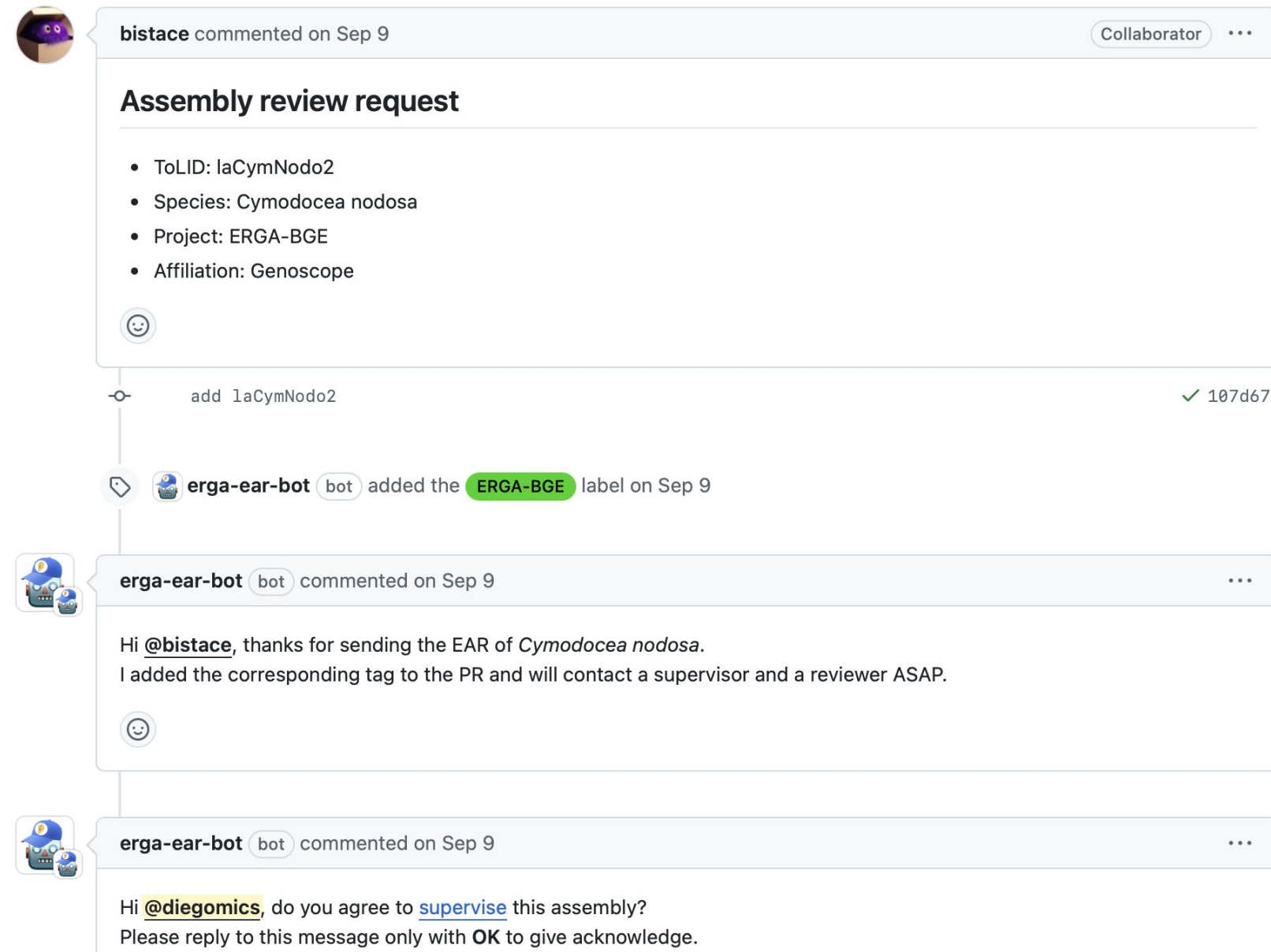
the **Supervisor**



the **Reviewer**

All persons from the community!

*bot implementation by [Arash Kadkhodaei](#)



bistace commented on Sep 9 Collaborator

Assembly review request

- ToLID: laCymNodo2
- Species: *Cymodocea nodosa*
- Project: ERGA-BGE
- Affiliation: Genoscope

add laCymNodo2 ✓ 107d677

erga-ear-bot bot added the **ERGA-BGE** label on Sep 9

erga-ear-bot bot commented on Sep 9

Hi **@bistace**, thanks for sending the EAR of *Cymodocea nodosa*. I added the corresponding tag to the PR and will contact a supervisor and a reviewer ASAP.

erga-ear-bot bot commented on Sep 9

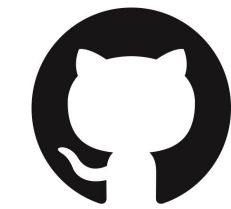
Hi **@diegomics**, do you agree to [supervise](#) this assembly? Please reply to this message only with **OK** to give acknowledge.



the **Researcher**



EAR.pdf



EARs repo



Open and traceable review conversations begin



erga-ear-bot bot commented on Sep 9



 EAR Reviewer Selection Process
 Date: 2024-09-09 05:55

All Eligible Candidates:

Github ID	Full Name	Institution	Total Reviews	Last Review	Active	Busy	Calling Score
SarahPelan	Sarah Pelan	Sanger	2	2024-08-20	Y	N	1001
joannacollins	Jo Collins	Sanger	2	2024-09-03	Y	N	1001
epaule	Michael Paulini	Sanger	2	2024-09-05	Y	N	1001
additive3	Jo Wood	Sanger	2	2024-06-15	Y	N	1000
tbrown91	Tom Brown	IZW	9	2024-07-05	Y	N	999
diegomics	Diego De Panis	IZW	9	2024-07-05	Y	N	997

Selected reviewer: Sarah Pelan (SarahPelan)
 The decision was based on:
 - different institution ('Sanger')
 - active ('Y')
 - not busy ('N')
 - oldest review and fewest reviews among the finalists (1051)

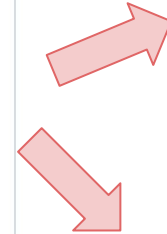


erga-ear-bot bot commented on Sep 9



Hi @SarahPelan, do you agree to review this assembly?
 Please reply to this message only with **Yes** or **No** by 16-Sep-2024 at 07:55 CET

- . From different institute
- . Active = Y
- . Busy = N
- . Highest Calling Score
- . NA = instant +50
- . Tie? Select NA or older review
- . Tie? Select less reviews
- . Tie? Random
- . initial = 1000
- . reviewing a genome = -1
- . rejecting review = 0
- . not answering on time = +1
- . genome from institute = +1 to all



	Github ID	Full Name	Institution	Supervisor	Total Reviews	Last Review	Active	Busy	Calling Score
1	gitcruz	Fernando Cruz	CNAG	N	5	2024-10-21	Y	Y	1005
2	jesgomez	Jessica Gomez Garrido	CNAG	N	2	2024-08-19	N	N	1008
3	talioto	Tyler Alioto	CNAG	N	2	2024-09-30	Y	Y	1006
4	bistace	Benjamin Istace	Genoscope	N	2	2024-10-11	Y	Y	1012
5	CaroB-M	Caroline Menguy	Genoscope	N	1	2024-06-14	Y	Y	1013
6	aurym	Jean-Marc Aury	Genoscope	N	1	2024-10-14	Y	N	1012
7	EmilieTeo	Emilie Teodori	Genoscope	N	0	NA	Y	Y	1013
8	andar27	Adama Ndar	Genoscope	N	1	2024-10-16	Y	N	1012
9	Idemirdj	Lola Demirdjian	Genoscope	N	1	2024-10-23	Y	N	1012
10	diegomics	Diego De Panis	IZW	Y	8	2024-10-22	Y	N	991
11	tbrown91	Tom Brown	IZW	Y	8	2024-07-05	Y	Y	994
12	DomAbsolon	Dom Absolon	Sanger	N	2	2024-09-23	Y	Y	1006
13	joannacollins	Jo Collins	Sanger	N	3	2024-10-23	Y	N	1005
14	additive3	Jo Wood	Sanger	N	3	2024-06-20	Y	Y	1005
15	epaule	Michael Paulini	Sanger	N	3	2024-10-16	Y	N	1005
16	SarahPelan	Sarah Pelan	Sanger	N	3	2024-10-04	Y	N	1005
17	tommthers	Tom Mathers	Sanger	N	3	2024-09-30	Y	N	1005
18	aersoares81	Andre Soares	SciLifeLab	N	0	NA	N	N	1000
19	gbdias	Guilherme Dias	SciLifeLab	N	0	NA	N	N	1000
20	mahesh-nanchal	Mahesh Binzer-Panchal	SciLifeLab	N	0	NA	N	N	1000
21									



erga-ear-bot bot commented on Sep 10



Thanks for agreeing!
I appointed you as the EAR reviewer.
I will keep your status as *Busy* until you finish this review.
Please check the [Wiki](#) if you need to refresh something. (and remember that you must download the EAR PDF to be able to click on the link to the contact map file!)
Contact the PR assignee for any issues.



SarahPelan commented on Sep 10

Collaborator ...

Hi [@bistace](#),
I have looked through laCymNodo2 and here's what I think:

SUPER_2
746-768 Kbp possible small retained haplotype
I think your SUPER_2_unloc_1 can join in the gap @ 18.3Mb in -ve orientation

SUPER_3
26.63 - end belongs on RH end of SUPER_5 in other orientation

SUPER_7
9.408-9.424Mb, possible small retained haplotype
Break in gap @ 19.97Mb and join this piece to LH end of SUPER_7

SUPER_10
4.55-4.59Mb approx, looks like retained haplotype


SUPER_12
10.891-10.907Mb, possible small retained haplotype

SUPER_13
5.35-5.37Mb, possible small retained haplotype





the **Reviewer**

The reviewer asks questions and points out things to improve

 **bistace** commented last month Collaborator Author ...

Hi, I reuploaded the map so the link should work correctly now!

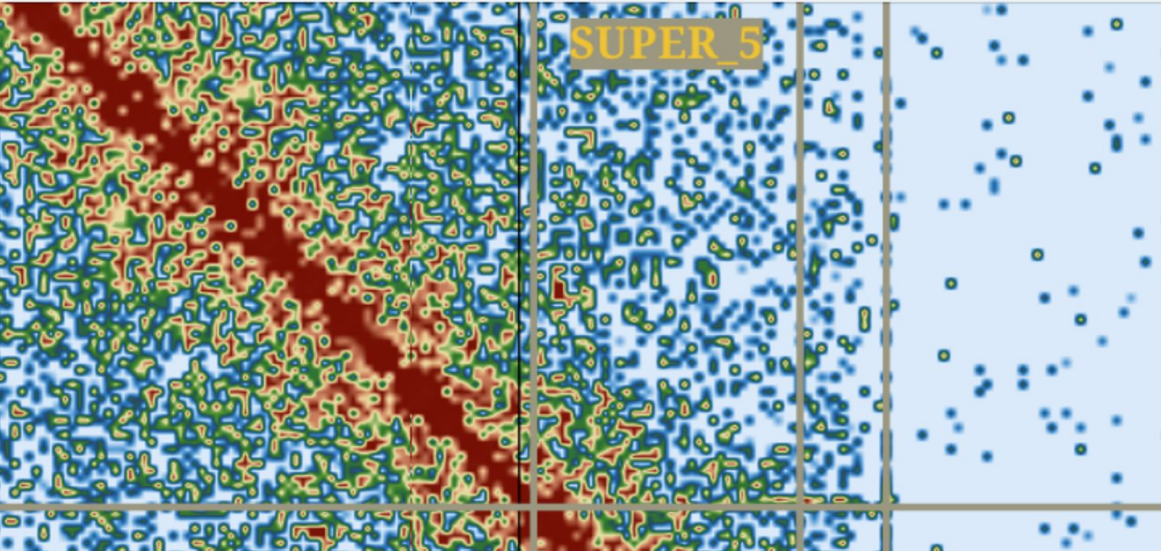


 **SarahPelan** commented last month Collaborator ...

Hi [@bistace](#)
I have rechecked, I think you missed off some unloc tags?
SUPER_2
27.4-end of chr, either needs to be broken and tagged as unloc, or joined in @ ~18.3Mb



SUPER_5
25.5-chr end, needs to be in other orientation. Then, the addition of the remaining small piece from the end of SUPER_3
looks good across the diagonal

laCymNodo2.pretext (2).map




The conversation space is flexible to provide links, attach images, bring other persons...


The process is enriching, equalises knowledge, and brings new perspectives to all participants

  **SarahPelan** approved these changes 3 weeks ago View reviewed changes

SarahPelan left a comment Collaborator ...


Hi **@bistace**
This now looks great, thanks for making all those changes!
Sarah




 **erga-ear-bot** bot commented 3 weeks ago ...


Thanks **@SarahPelan** for the review.
I will add a new reviewed species for you to the table when **@diegomics** approves and merges the PR ;)


Congrats on the assembly **@bistace**!
After merging, you can [upload the assembly to ENA](#).



 **diegomics** commented 3 weeks ago Member ...

Thanks **@bistace** for another nice Genoscope assembly! I want to take a moment to point out the extraordinary review by **@SarahPelan** 🙌



 **diegomics** merged commit **1218a96** into **ERGA-consortium:main** 3 weeks ago View details Revert
2 checks passed



the **Supervisor**

When the reviewer approves the report and everyone is satisfied, the supervisor merges the pull request to add the PDF to the stable repository of assembly reports



the **Bot**

After merging, points are counted, the conversation link and reviewer are recorded in a public table on the repo, and the newly approved assembly is announced



EAR bot APP 4:10 PM

New Assembly Finished! 🎉

Congratulations to Emilie Teodori and the Genoscope team for the high-quality assembly of *Macrothele cretica*

The assembly was reviewed by Jo Collins, and the process supervised by Tom Brown. The EAR can be found in the following link:

https://github.com/ERGA-consortium/EARs/blob/main/Assembly_Reports%2FMacrothele_cretica%2FqMacCret1%2FqMacCret1_EAR.pdf

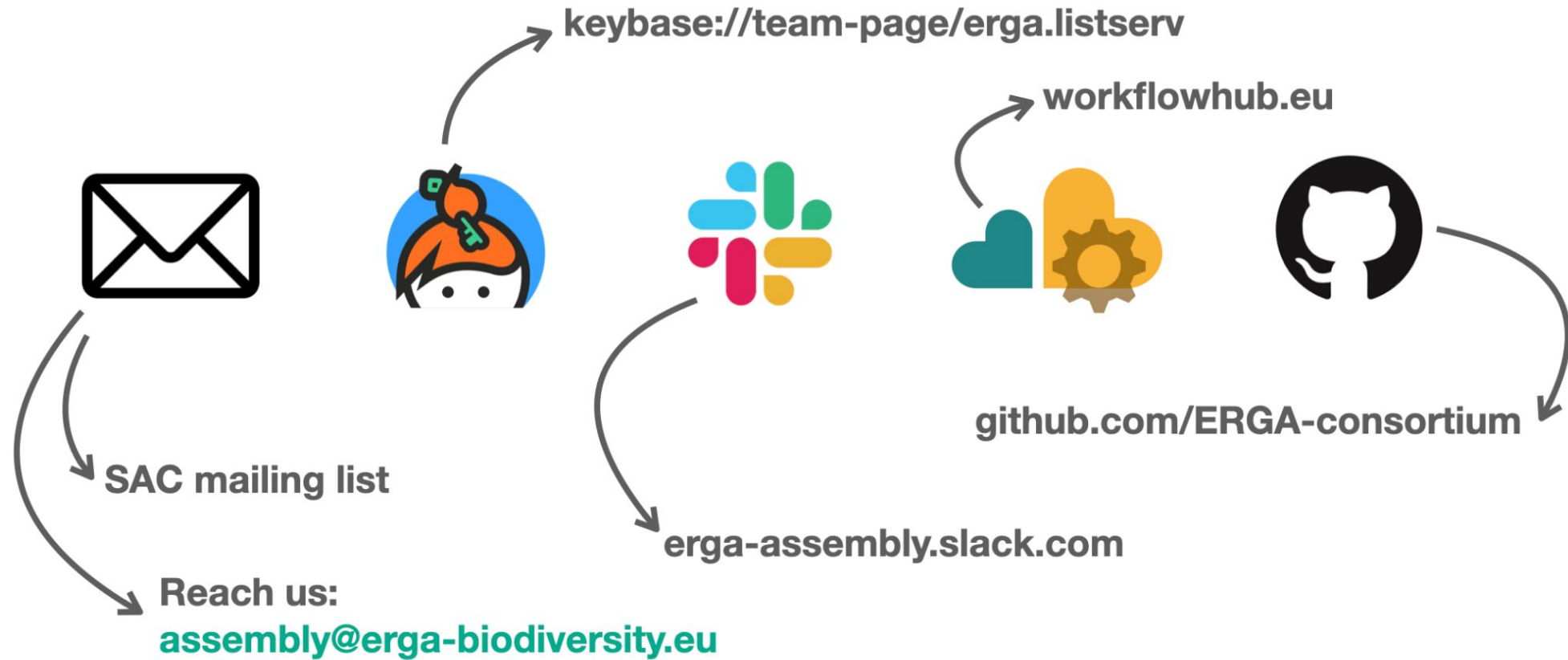


1	Reviewer Name	Institution	Reviewed Species Name	EAR reviewing conversation (PR) link
38	Dom Absolon	Sanger	Albinaria teres	https://github.com/ERGA-consortium/EARs/pull/51
39	Tom Mathers	Sanger	Deroceras lasithionense	https://github.com/ERGA-consortium/EARs/pull/50
40	Tyler Alioto	CNAG	Acomys minous	https://github.com/ERGA-consortium/EARs/pull/73
41	Fernando Cruz	CNAG	Diadema setosum	https://github.com/ERGA-consortium/EARs/pull/72
42	Sarah Pelan	Sanger	Cymodocea nodosa	https://github.com/ERGA-consortium/EARs/pull/66
43	Benjamin Istace	Genoscope	Placobdella costata	https://github.com/ERGA-consortium/EARs/pull/77
44	Jean-Marc Aury	Genoscope	Tarphius canariensis	https://github.com/ERGA-consortium/EARs/pull/74
45	Michael Paulini	Sanger	Loboptera subterranea	https://github.com/ERGA-consortium/EARs/pull/86
46	Adama Ndar	Genoscope	Stigmatoteuthis arcturi	https://github.com/ERGA-consortium/EARs/pull/82
47	Fernando Cruz	CNAG	Ebenus cretica	https://github.com/ERGA-consortium/EARs/pull/83
48	Fernando Cruz	CNAG	Chaetopelma lymerakisi	https://github.com/ERGA-consortium/EARs/pull/90
49	Diego De Panis	IZW	Marifugia cavatica	https://github.com/ERGA-consortium/EARs/pull/78
50	Lola Demirdjian	Genoscope	Cheirolophus tagananensis	https://github.com/ERGA-consortium/EARs/pull/103
51	Jo Collins	Sanger	Macrothele cretica	https://github.com/ERGA-consortium/EARs/pull/69

Bring your issues and ideas!



Meetings: 3rd Thursday of each month, 2 pm CET





ERGA Community Genomes

Robert Waterhouse

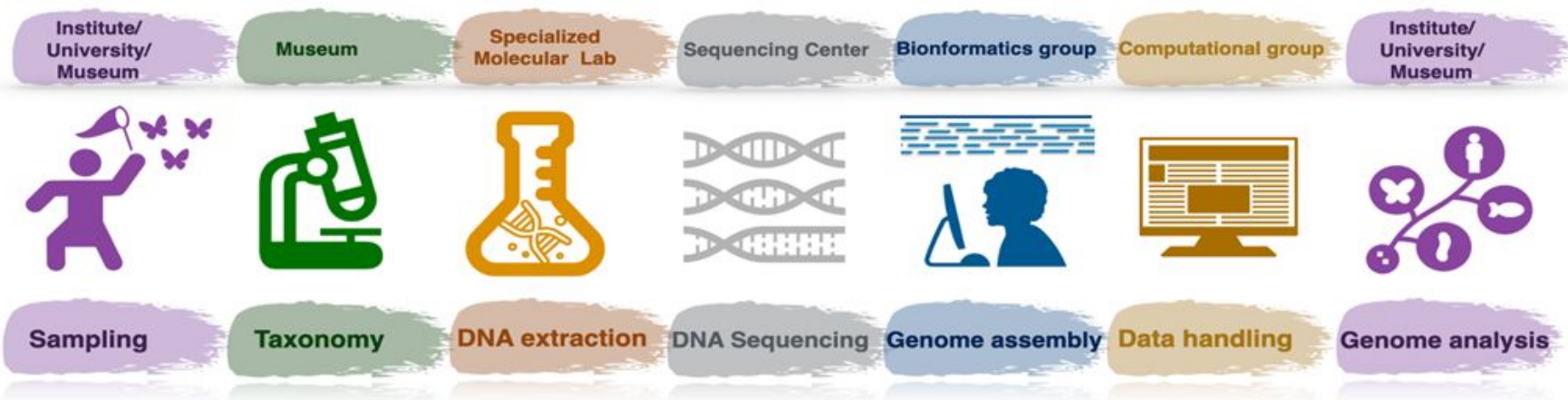
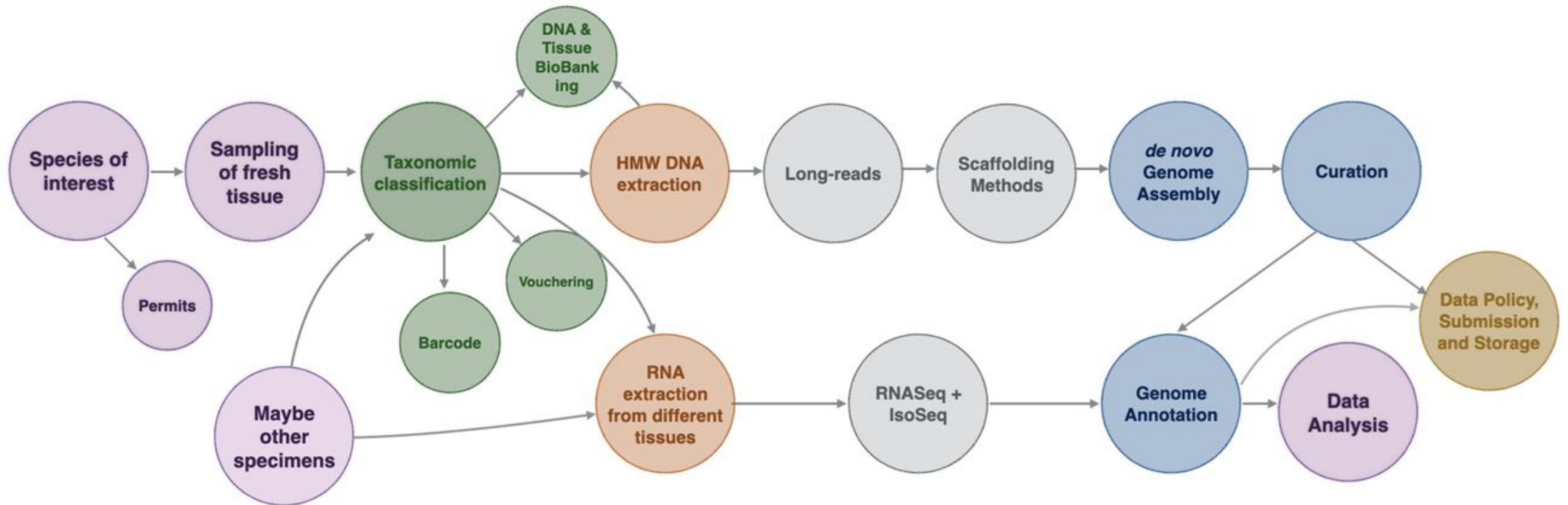




ERGA Community Genomes

Rob Waterhouse





Community Standards



SSP

Sampling & Sample Processing



SAC

Sequencing & Assembly



Annot

Annotation Committee



DAC

Data Analysis Committee



ITIC

IT & Infrastructure Committee

Metadata Collection

Practical Guidelines



ELSI

Ethical, Legal and Social Issues



Media

Communications and public affairs



CS

Citizen Science Committee



TKT

Training and Knowledge Transfer



SJC

Social Justice Committee

Support Systems

Analysis Workflows

Information Sharing

Standard Operating Procedures

Protocols & Best Practices



ABOUT ERGA

OUR COMMUNITY

RESOURCES

ERGA PROJECTS

NEWS & EVENTS

SUPPORT

JOIN & CONTACT



**Are you embarking on a reference genome project?
Do you want to learn about the steps required for success?**

Then join the growing family of ERGA Community Genomes!

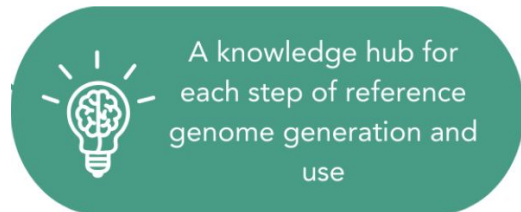
ERGA aims to coordinate the production of high-quality annotated genome assemblies that represent eukaryotic biodiversity in Europe. A key part of this is building capacity across European researchers and institutes by supporting the growing community of scientists in biodiversity genomics through the provision of guidelines, workflows, and best practices that explain and greatly facilitate the successful execution of the many steps required along the complex workflow for reference genome generation.

The guidelines below cover many of the main steps along the genome generation workflow, providing step-by-step advice and answers to frequently asked questions to help researchers navigate the complexities and find out where to turn for additional assistance:

Menu

- 1. Pre sampling
- 2. Sample Acquisition Strategy
- 3. DNA/RNA extraction
- 4. Libraries preparation
- 5. DNA sequencing data
- 6. RNA sequencing data
- 7. Assembly completed
- 8. Annotation completed
- 9. Downstream analysis

- Permits & ABS
- Permissions & MTAs
- Taxonomic Validation
- Specimen Vouchering
- Sample Biobanking
- Sample Shipping
- Metadata Manifest



The screenshot shows the ERGA portal interface. On the left, there are navigation panels for 'Data status', 'Symbionts', 'Projects', and 'Phylogeny/kingdom'. The 'Data status' panel shows: BioSamples - Submitted (544), Raw Data - Submitted (118), Assemblies - Submitted (46), and Annotation Complete - Done (11). The 'Projects' panel highlights 'ERGA' with 544 items. The 'Phylogeny/kingdom' panel shows: Metazoa (283) and Fungi (4).

The main 'Status Tracking' window is open, displaying a table with the following columns: Organism, Common Name, Metadata submitted to BioSamples, Raw data submitted to ENA, Assemblies submitted to ENA, and Annotation complete. The table lists 13 organisms with their respective submission statuses.

Organism	Common Name	Metadata submitted to BioSamples	Raw data submitted to ENA	Assemblies submitted to ENA	Annotation complete
Absolonia gigantea		Done	Waiting	Waiting	Waiting
Acanthodactylus beershebensis		Done	Waiting	Waiting	Waiting
Acanthodactylus schreiberi		Done	Waiting	Waiting	Waiting
Acipenser sturio		Done	Waiting	Waiting	Waiting
Acmaeoderella farinosa		Done	Waiting	Waiting	Waiting
Acmaeoderella pseudovirgulata		Done	Waiting	Waiting	Waiting
Acomys minous	Crete spiny mouse	Done	Done	Done	Waiting
Acrobeloides nanus		Done	Waiting	Waiting	Waiting
Aegopis spelaesus		Done	Waiting	Waiting	Waiting
Agallia sp. 1 PR-2024		Done	Waiting	Waiting	Waiting

At the bottom of the table, it shows 'Items per page: 15' and '1 - 15 of 544'.

Powered by EMBL-EBI <https://portal.erga-biodiversity.eu/home>

Logos for Biodiversity Genomics Europe, European Union, Swiss Confederation, and UK Research and Innovation are also present.

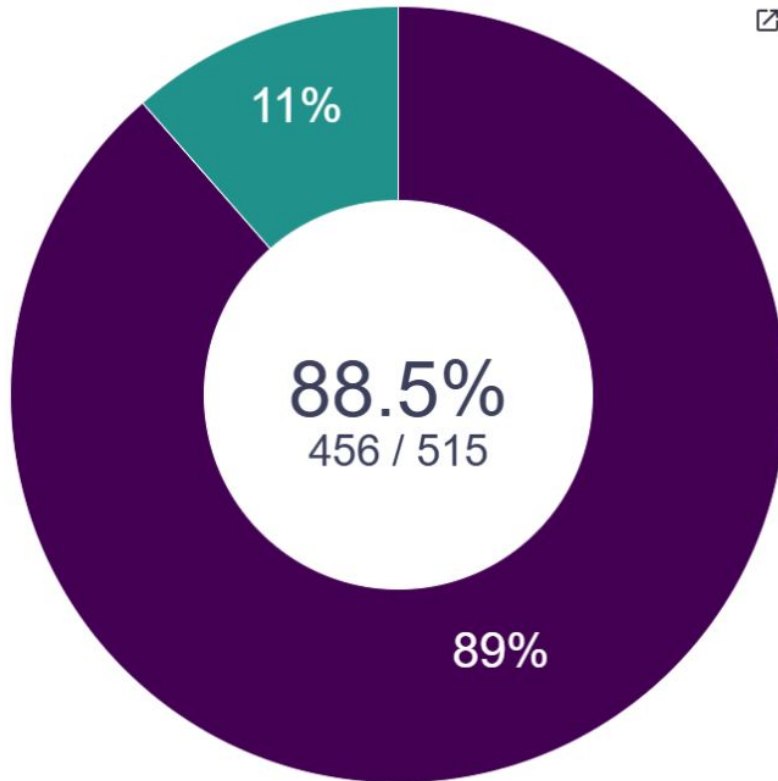


Component Projects

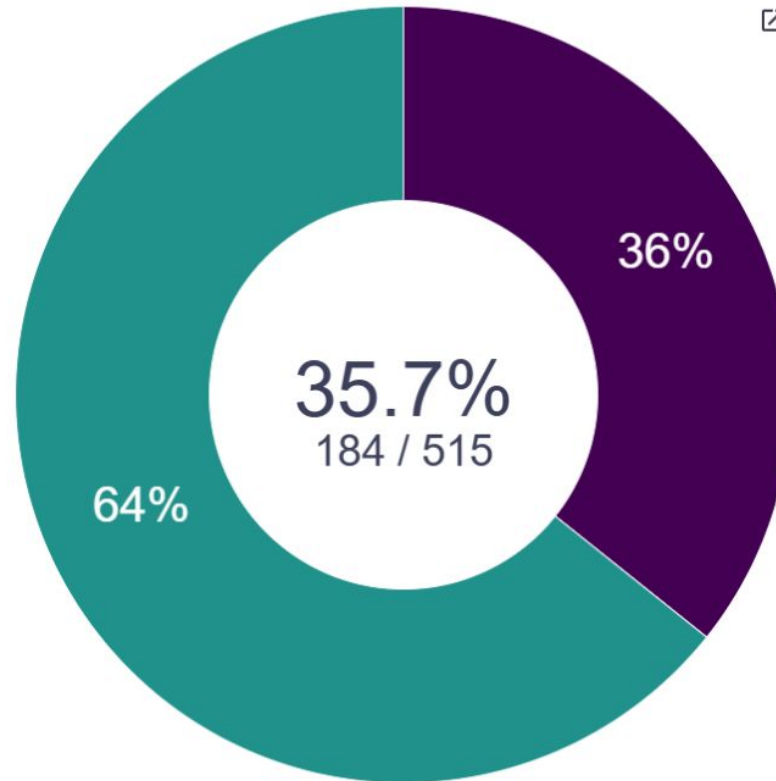
- 25 Genomes Project: Genome Data and Assemblies**
To celebrate the 25th Anniversary of the Wellcome Trust Sanger Institute, we sequenced and assembled... [See more](#)
- Project Psyche - Lepidopteran Genomes for Europe**
Lepidoptera are major pollinators, are important pests in agroecosystems, are key to the study of th... [See more](#)
- The Swiss node of the European Reference Genome Atlas (ERGA) initiative**
The European Reference Genome Atlas (ERGA) is a pan-European initiative to generate reference genome... [See more](#)
- European Reference Genome Atlas (ERGA): Biodiversity Genomics Europe (BGE) Project Genome Data and A...**
This pan-European effort, driven by the European Reference Genome Atlas (ERGA) consortium (<https://w...>) [See more](#)
- ATLSea : An Atlas of eukaryotic marine genomes**
Focused on the French Exclusive Economic Zone (EEZ), ATLSea will cover a wide range of marine speci... [See more](#)
- The European Reference Genome Atlas (ERGA) Community Genomes**
The European Reference Genome Atlas (ERGA) is a pan-European consortium with the goal of coordinat... [See more](#)
- The French node of the European Reference Genome Atlas (ERGA) initiative**
The European Reference Genome Atlas (ERGA) is a pan-European consortium with the goal of coordinat... [See more](#)
- Catalan Initiative for the Earth BioGenome Project (CBP)**
Repository for data generated from the Catalan Initiative for the Earth Biogenome Project (CBP), whi... [See more](#)
- ENDEMIXIT**
Genome assembly and population genomics of Italian endemics
- Darwin Tree of Life Project: Genome Data and Assemblies**
The Darwin Tree of Life Project (darwintreeoflife.org) generates high quality genome assemblies for ... [See more](#)

Coordination via ENA BioProjects under the ERGA Umbrella

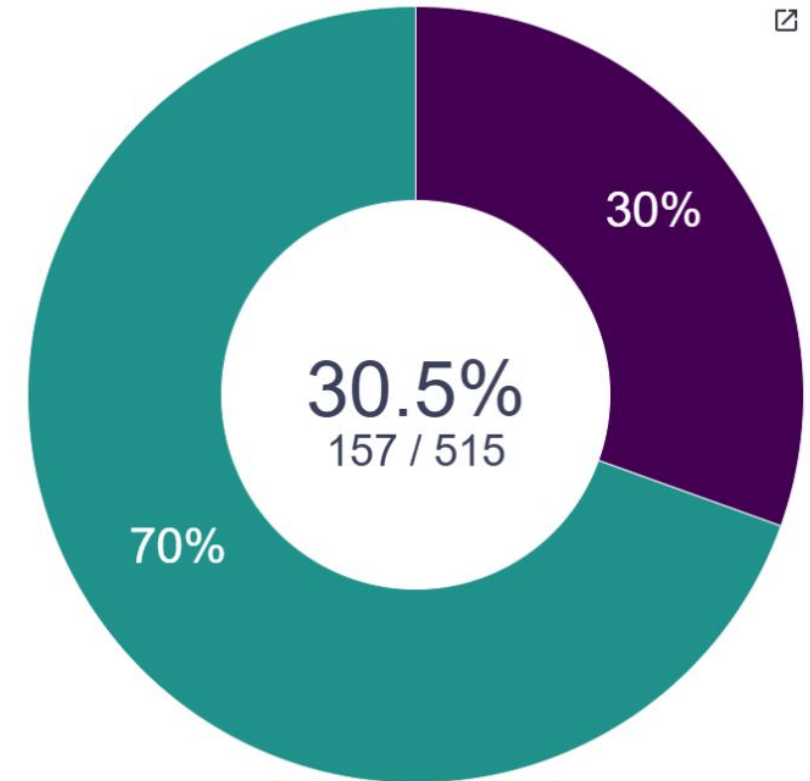
Sequencing Status of ERGA-BGE Species



Total of ERGA-BGE target species collected



Total of ERGA-BGE target species received by the assigned sequencing center



Total of ERGA-BGE target species with genome sequencing currently in progress



Coordination via ERGA at Genomes on a Tree (GoaT): e.g. ERGA-BGE

How to engage with Community Coordination Efforts ?

[1] Communicate which species you are working on || [2] Get your genome “ERGA Approved”

GoaT Spreadsheets

- Core Projects (e.g. Pilot, BGE, etc.)
- European Projects (ENDEMIXIT, CBP, etc.)
- Country GoaT Sheets - under development
- ERGA Community GoaT Sheet - under development



ERGA-CH - The Swiss node of the European Reference Genome Atlas (ERGA) initiative

GoaT Search Term: ERGA-CH

ERGA-CH is the Swiss node of the European Reference Genome Atlas initiative. ERGA-CH is coordinating national sampling, sequencing, and assembly efforts to contribute to reference genome generation as part of the ERGA community. For more information reach the contact info below.

Bioproject ID PRJEB49197

Main Contact: [switzerland\(at\)erga-biodiversity.eu](mailto:switzerland(at)erga-biodiversity.eu)

Metadata Validation (ERGA Manifest)

- SSP Committee - check that your metadata meet the required standards

Assembly Validation (EBP Standards)

- S&A Committee - check that your assembly meet required standards



Study Title:

The European Reference Genome Atlas (ERGA) Community Genomes

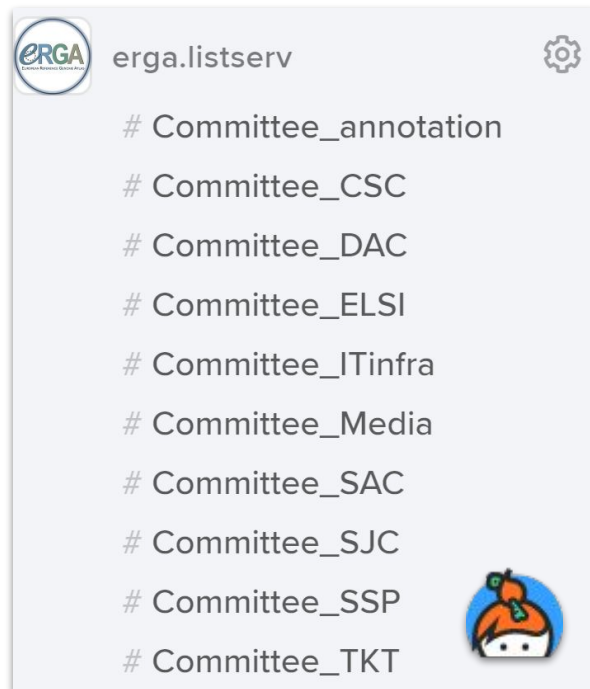
Center Name:

European Reference Genome Atlas

Got a question ? Are you stuck ? There's no need to struggle alone !

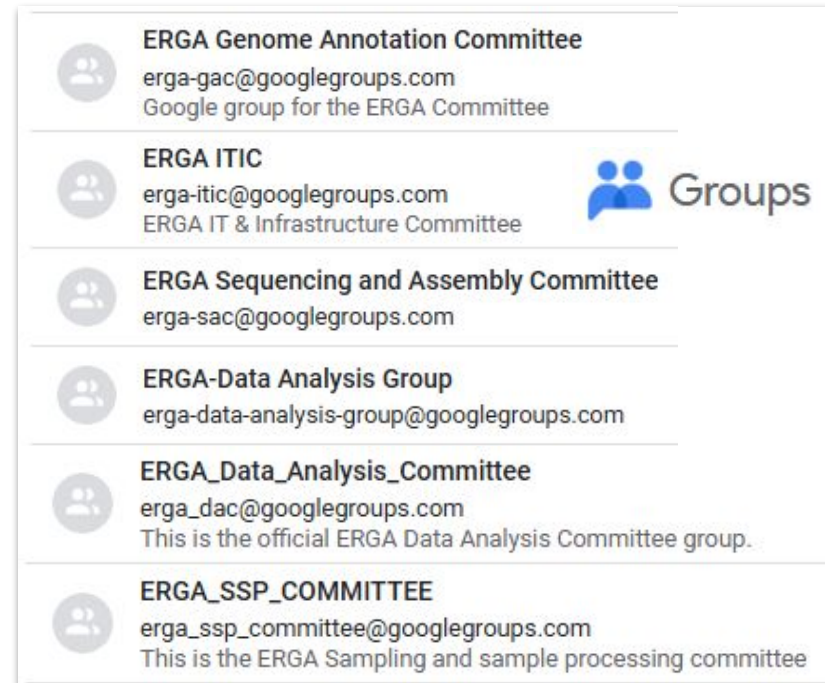
Keybase - ERGA's Community Chat Platform || ERGA Committee Group Mailing Lists

You don't have to be a regular Committee Participant to ask for help



erga.listserv

- # Committee_annotation
- # Committee_CSC
- # Committee_DAC
- # Committee_ELSI
- # Committee_ITinfra
- # Committee_Media
- # Committee_SAC
- # Committee_SJC
- # Committee_SSP
- # Committee_TKT



ERGA Genome Annotation Committee
erga-gac@googlegroups.com
Google group for the ERGA Committee

ERGA ITIC
erga-itic@googlegroups.com
ERGA IT & Infrastructure Committee

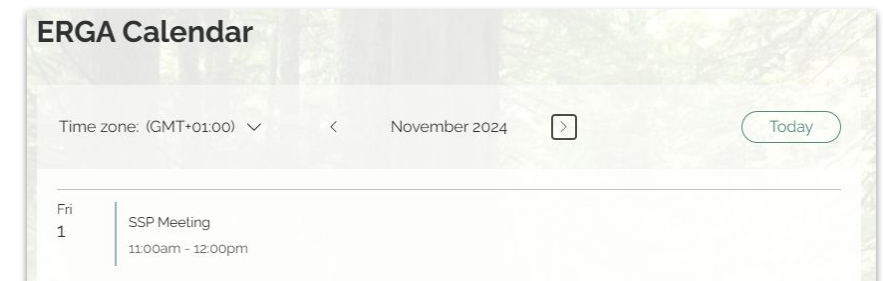
ERGA Sequencing and Assembly Committee
erga-sac@googlegroups.com

ERGA-Data Analysis Group
erga-data-analysis-group@googlegroups.com

ERGA_Data_Analysis_Committee
erga_dac@googlegroups.com
This is the official ERGA Data Analysis Committee group.

ERGA_SSP_COMMITTEE
erga_ssp_committee@googlegroups.com
This is the ERGA Sampling and sample processing committee

1. Post your question(s) on keybase in the most relevant channel
2. Email the Committee Steering Group with your question(s) [see website for emails]
3. Request a slot in a Committee meeting to discuss with experts



ERGA Calendar

Time zone: (GMT+01:00) < > November 2024 Today

Fri 1	SSP Meeting 11:00am - 12:00pm
-------	----------------------------------

Thank you!

