



HAL
open science

BYTE-Sea: the digital infrastructure of ATLASea, the French marine genome sequencing programme

Karine Massau, Byte-Sea Consortium, Erwan Corre

► **To cite this version:**

Karine Massau, Byte-Sea Consortium, Erwan Corre. BYTE-Sea: the digital infrastructure of ATLASea, the French marine genome sequencing programme. JOBIM 2024 - Journées ouvertes en biologie, informatique et mathématiques, Jun 2024, Toulouse, France. pp.1-1, 2024. hal-04693601

HAL Id: hal-04693601

<https://hal.science/hal-04693601>

Submitted on 10 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

BYTE-Sea: the digital infrastructure of ATLASea, the French marine genome sequencing programme

Karine MASSAU¹, BYTE-Sea consortium² and Erwan CORRE^{1,3}

¹ ABiMS bioinformatic platform, FR2424, CNRS/Sorbonne Université, Station Biologique de Roscoff (SBR), 29680, Roscoff, France

² BYTE-Sea Consortium : https://gitlab.com/pepr-atlasea/byte-sea_consortium

³ IFB - Institut Français de Bioinformatique, CNRS UAR 3601, France



Priority research programs and equipment (PEPR)

Funded by the France 2030 framework program. Administered by ANR

Led by CNRS and CEA

Associated to the international ERGA and EBP projects

5 academic partners: Ifremer, MNHN, Aix-Marseille University, Paris Sciences & Lettres and Sorbonne University.

3 French infrastructures: France génomique, IFB, EMBRC-Fr

ATLASea Programme - an atlas of marine genomes

genome sequences of 4500 marine organisms along the French coast



Calls for expressions of interest (2026)

Study of metabolic pathways

Study of ecosystems subject to invasions

WHEEL-Sea (CNRS):
improve and support teaching and training in the field of biodiversity.

DIVE-Sea (MNHN):
sampling expeditions along the French mainland and overseas coasts and interactions with marine stations. Banking of samples.

SEQ-Sea (CEA):
preparation of high quality DNA, sequencing, assembly and annotation of genomes at Genoscope.

BYTE-Sea (IFB):
IT infrastructure for data storage, integration, display and sharing. Functional and expert annotation and comparative genomics

BYTE-Sea Work packages (WP)

WP1: Infrastructure for storage, computing and software

Partners: IFB, IFREMER

Infrastructures: BigOuest, IFB Core Cluster, DATARMOR

Storage: Data access and Data Ring

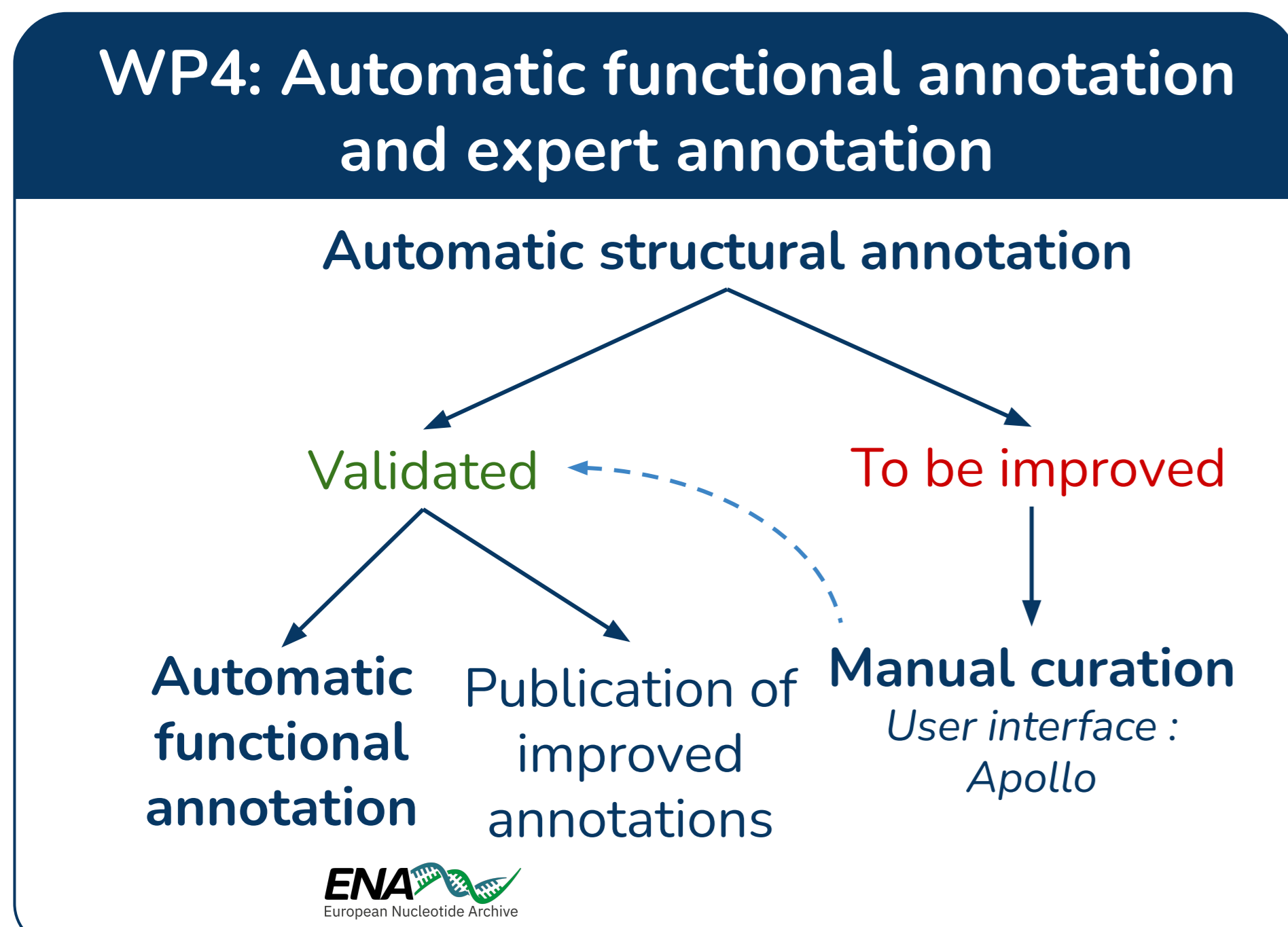
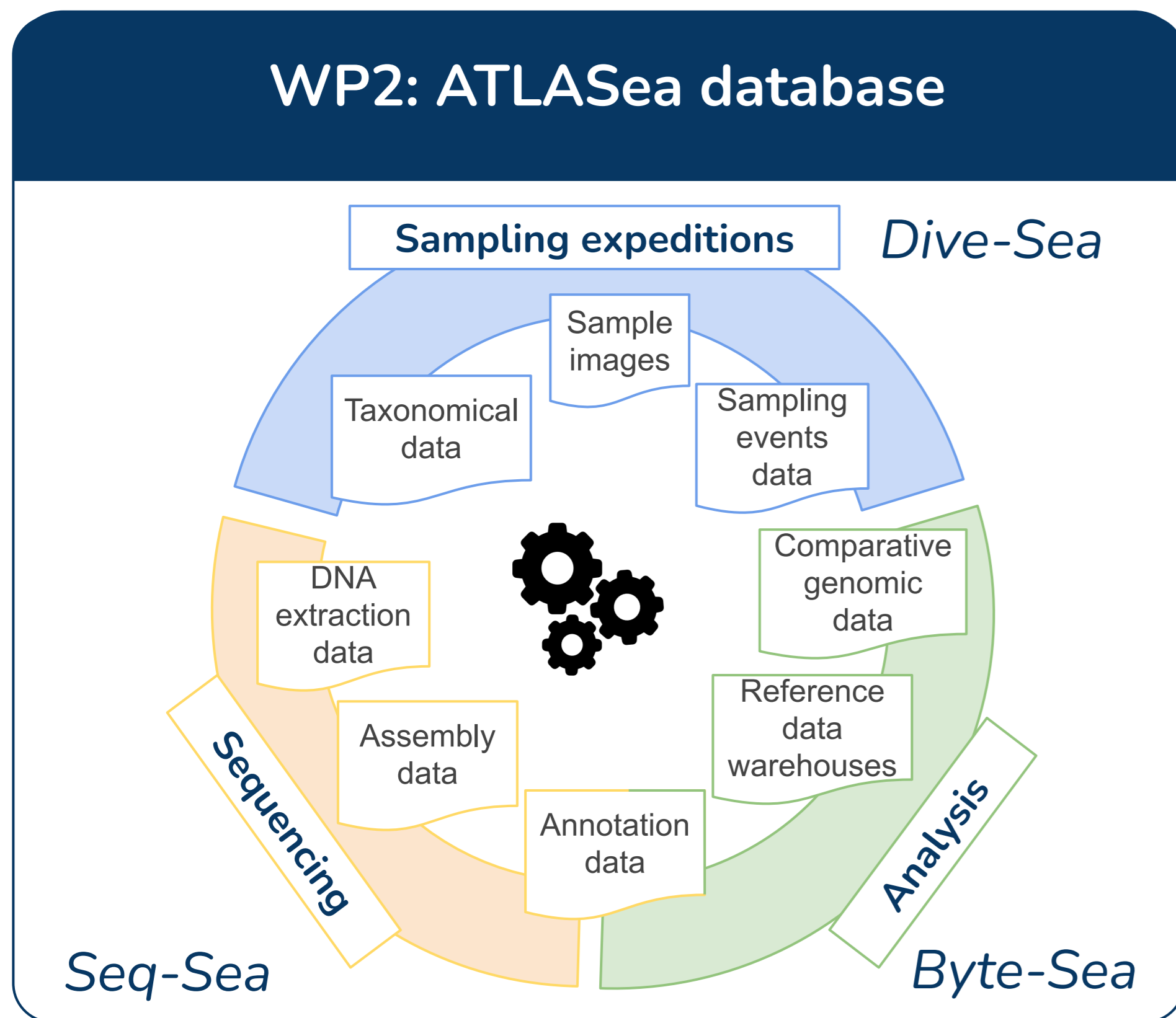
Computing: Functional annotation & comparative genomics

Software & services: GitLab repository: Tools and SOP

Database: Data integrity

Portal: Services: browsers, web site, API...

PEPR ATLASea
<https://gitlab.com/pepr-atlasea>



WP3: ATLASea portals for data and analysis tools

Tracking portal

Non-contractual image. Taken from the ERGA-Status tracking portal
<https://genomes.cnag.cat/erga-stream/overview/>

Data portal

<https://portal.atlasea.fr>

Genome portal

Biological data: *fasta, gff*

User input: *yaml*

ATLASea workflow: Deploy -> Lock data -> Derived data -> Validate & setup

GitLab CI/CD

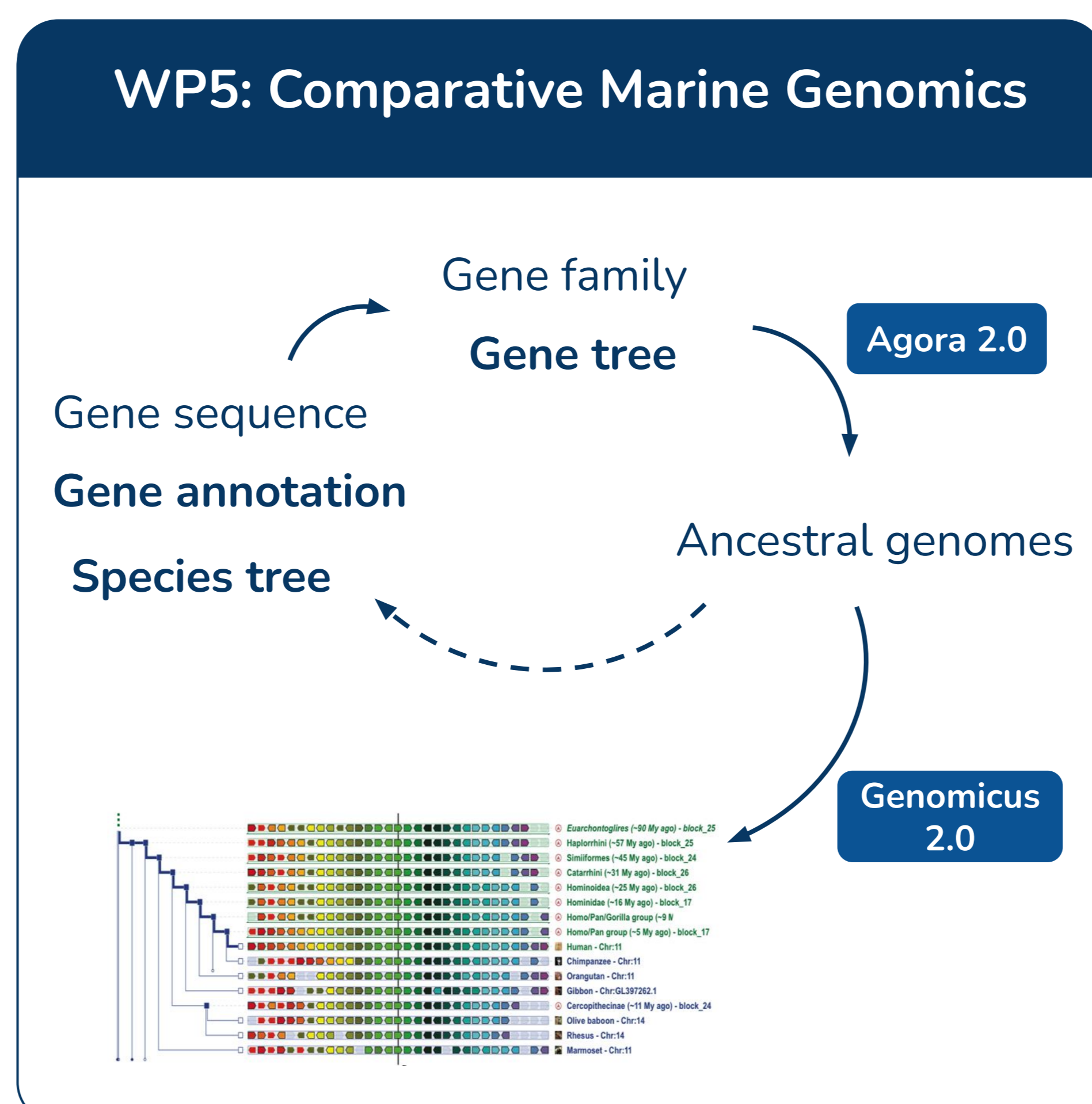
<https://gitlab.com/beaur1s/beauris>

Data access

ENEA European Nucleotide Archive

Galaxy

DATARMOR & IFB



WP6: Citizen Annotation

Manual curation of genome annotation

Citizens
High school / undergraduate students
Introduction to the world of genomics

Scientific output
Guided and gamified curation web UI
Release of improved gene annotation