

ERGA Community Genome Reports Guidelines

Prepared by Tom Brown with input from the ERGA Executive Board and Committees

This document lists the requirements for a Genome Report publication to be included within the European Reference Genome Atlas (ERGA) Community Genome Reports Collection, hosted at the Research Ideas and Outcomes (RIO) Journal. The Collection can be viewed at the following locations:

<https://doi.org/10.3897/rio.coll.280> https://riojournal.com/topical_collection/280/

Each submission to the Collection will be evaluated by the **Collection Editors**, to decide whether the Genome Report satisfies the criteria below. Note that to be included in the Collection at RIO, the Genome Report does not need to be published in the RIO Journal. If the article is published at another journal, a link to the article will be listed as part of the ERGA Community Genome Reports Collection and access to the manuscript will depend on the publication license of the publisher. If you wish instead to make use of the ARPHA writing tool to write your Genome Report manuscript and release your preprint or publish directly with RIO, you can use the [RIO submission portal](#) directly.

ERGA Community Genome Report Requirements

- ★ The Genome Report must be published with a platform that provides a permanent identifier, such as a Digital Object Identifier (DOI).
- ★ The Genome Report must detail the full methods pertaining to each step of the generation of the genome(s). E.g. sampling methods, required permits, biobanking and voucher information, protocols for extraction, quality checks, library preparation and sequencing of any genomic material, details of software, versions, and arguments used to produce and process sequencing data and generate the genome assembly and annotation in a manner that is fully reproducible.
- ★ Sample metadata accompanying the sequencing data must detail the collected sample, including sampling location, tissue or cell type and should ideally contain the mandatory fields outlined in the [ERGA Sample Manifest](#) and adhere to the [ENA Tree of Life Checklist](#).
- ★ The sample metadata, raw sequencing data, genome assembly or assemblies, and annotation(s) must be placed in a public repository with permanent accessions, such as the European Nucleotide Archive (ENA). These accessions must be listed in the Genome Report.
- ★ The species umbrella BioProject must be linked as a component project of the ERGA umbrella BioProject ([PRJEB43510](#)), e.g. via the ERGA Community Genomes BioProject ([PRJEB66264](#)), or projects (e.g. Pilot [PRJEB47820](#) or BGE [PRJEB61747](#)) or national nodes (e.g. ERGA-CH [PRJEB49197](#)). Please contact the ERGA IT & Infrastructure Committee ([ITIC](#)) for assistance.

- ★ While the specimen itself does not need to be sampled from Europe, the distribution of the sequenced species must have some overlap with Europe, for example falling within the [EEA biogeographical regions](#).
- ★ The Genome Report must describe the sequenced species in the introduction with some rationale (motivation) included for the generation of the genome.
- ★ The Genome Report must list statistics or display figures demonstrating the quality and quantity of the raw sequencing data, assembly, and annotation that can be independently evaluated by readers, particularly including the metrics listed in the Earth BioGenome Project (EBP) [Assembly standards documents](#). The genome itself should be of a quality considered “state-of-the-art” given the current technologies available for the sequenced sample.
- ★ The Genome Report must list the individuals involved in each step of the genome generation process and should ideally detail to which step(s) each individual contributed.

ERGA Community Genome Report Recommendations

- ★ The Genome Report publication should be fully open-access, however we acknowledge that due to multiple factors this is not always possible.
- ★ The genome [Assembly](#) and [Annotation](#) should meet the standards established by ERGA and the EBP, noting the exceptions provided for taxa or genomes for which the quality metrics are not yet realistic or attainable.
- ★ The Genome Report and any data, metadata, workflows, or protocols should be available, written, and detailed in an accessible and reusable manner in line with the [FAIR](#) principles. See documents such as [ERGA's Open Data Policy](#) for guidelines on how to ensure your project conforms with ERGA's guidelines.
- ★ The Genome Report should wherever relevant mention, e.g. in the acknowledgements section, any support received from ERGA, this could be in the form of technical support from the ERGA Committees or simply acknowledging the use of ERGA resources such as guidelines, protocols, workflows, etc. made available through the ERGA Community.

Note: This Collection at the RIO Journal is not for the publication of protocols, workflows, etc. instead we recommend using a permanent repository such as [protocols.io](#) or [WorkflowHub](#). In this way, these protocols or workflows can be versioned and cited in the Genome Reports.