

Darwin Tree of Life Open Data Release

Version 1.04

Data Release Policy: Darwin Tree of Life project

The Darwin Tree of Life project (DToL) is a partnership between The Wellcome Sanger Institute, The Natural History Museum, London (NHM), the Royal Botanic Gardens at Kew (RBGKew), the Royal Botanic Gardens Edinburgh (RBGE), the Marine Biological Association, the Earlham Institute, the European Bioinformatics Institute (EBI), the University of Oxford, the University of Edinburgh and the University of Cambridge, assisted by many collaborators.

DToL intends that all the sequence data generated by the project will be openly available for reuse, and is therefore depositing raw and assembled data in the public sequence databases (European Nucleotide Archive [ENA], and from there the other International Nucleotide Sequence Database Collaboration [INSDC] nodes - GenBank and the DNA Data Bank of Japan). Many physical specimens will also be deposited in the relevant collections (NHM, RBGKew, RBGE and others).

DToL data are released freely for reuse for any purpose upon deposition in ENA, and the DToL partners encourage such community reuse. Our intention is to rapidly publish all submitted assemblies as Wellcome Open Research notes, which can be cited (see, for example, Boyes *et al.* (2024) The genome sequence of the Figure of Eighty moth *Tethea ocellaris* Linnaeus, 1767 <https://wellcomeopenresearch.org/articles/9-348>), and we expect that users of the data will give appropriate acknowledgement and citation.

DToL may also make available for download intermediate data and assemblies *via* the project website at <https://www.darwintreeoflife.org>. These data and assemblies are provided as is as a service to the community, and we make no assurances as to their completeness or quality. Please note that intermediate assemblies will be improved before final submission to ENA and we cannot guarantee persistence or availability of intermediate files in the long term. We strongly recommend that published analyses are based on data and assemblies submitted to ENA/INSDC. The genome sequences submitted to ENA by DToL will be presented through the EBI Ensembl database, and the annotations presented through Ensembl should be regarded as the official DToL versions.

DToL collectively, and individual partners, will be publishing integrated analyses across the data we produce and we encourage researchers planning large scale analyses to contact us (contact@darwintreeoflife.org) so that we can collaborate effectively to mutual advantage.

About this document

Author: Mark Blaxter

Date: 08 06 2024

Version: 1.04,

Purpose: To deliver a data release statement for both pre-submission and post-submission DToL genome data that is clear and precise.

Status:

- 0.9 Released to DToL Leadership Team for comment and editing. 20 02 2020
- 1.0 Taken to DToL Steering Group. 26 02 2020
- 1.02 Circulated to DToL and ToL leadership 28 02 2020
- 1.03 Comments received to 20 03 2020 incorporated

- 1.04 Revised to clarify how intermediate data can be requested and update cited Genome Note 08 06 2024