DToL FAQs

Sampling

Q. I have collected a really special thing/I work on an amazing organism that hasn't been sequenced yet. How do I submit samples for genome sequencing through DToL?
  A. Contact DToL Project Management via contact@darwintreeoflife.org

Q. How do I find out what has been sampled already?
  A. Look at the [DToL Portal](#) which gives you full insight into everything in our pipeline.

Q: Do you sequence species that are not from the UK and Ireland?
  A. Our project focuses on the eukaryotes of Ireland and Britain, and we have limited resources to sequence additional species. If your species is found in continental Europe, the [European Reference Genome Atlas](#) project will be interested to hear from you. If your species is a vertebrate, the [Vertebrate Genomes Project](#) may be interested. Otherwise check the [Earth BioGenome Project affiliates page](#) and check if any of the projects may have an interest in your species.

Q. Do you have to kill the species you sequence?
  A. We follow a strict [ethical code of sampling](#). While it is true that for smaller animal species we do have to euthanize an individual to obtain enough sample for sequencing, for larger species we try to sample non-lethally wherever possible. For animals covered by Home Office legislation, we only sample in carefully controlled, explicitly permitted ways consistent with animal welfare regulations. For endangered and protected species we only collect where we have permission to do so, and in close consultation with wildlife managers.

Sequencing

Q. We intend to or already started sequencing species X, but now found out it is on the DToL list. What happens next?

  A. Just because it is on the DToL list does not mean a reference will appear soon. We suggest that you contact DToL Project Management via [contact@darwintreeoflife.org](#) to discuss the status and the best way forward. If you are on track to complete an assembly that will meet the EBP quality metrics and be openly released soon, and DToL has not yet collected data, then we may put our project on pause. If we have already collected data and are on track to release our assembly soon you may decide to put your project on pause (your choice). If it is not clear when DToL will be able to complete a genome sequence you may decide to continue sequencing

yourself in order to be confident of an assembly becoming available on a timeframe under your control.

Assembly

Q.

A.

Annotation

Q. Where can I get a gff file of the annotations for my favourite species?

A. [Darwin tree of life @ensembl](#) holds all annotations produced for the DToL project. This provides a convenient way of accessing all DToL Ensembl data. Alternatively you can also find all DToL species and their associated files on the [Ensembl Rapid Release site](#).

Q. I have unpublished transcriptomic or other data for a species you are sequencing. The data isn't public yet because I haven't published the study, how can I share our data with you to improve the genome annotation?

A. We can only work with publicly archived data that has been submitted to the ENA or SRA, and the data need to be publicly available ahead of the annotation process beginning. We do value extra annotation data, so if you plan to make the data public, please contact us on helpdesk@ensembl.org to make us aware of it.

Data access

Q. How far has my species of interest progressed and when will data be released?

A. You can check the status of the species in the DToL portal at https://portal.darwintreeoflife.org/tracking. This provides progress tracking from registering of the sample through to the annotation being completed. For a more detailed view of the data generated (in case there is any already) check ToLQC at https://tolqc.cog.sanger.ac.uk/ where you will also find links to any released raw data. If your species of interest has had an assembly built and it has been released to the public archives and possibly also already annotated and a genome note being published, the portal will link to the relevant sources. Whilst many species take about 6-9 months to go through the pipeline from extraction to release, we can't guarantee a certain time period as every species comes with its unique challenges.

Q. Can I have early access to intermediate assemblies of species X?

A. Whilst this creates quite an overhead, it is certainly possible. Please check the current status of your species of interest (see above) and if there is no assembly released yet, but having any assembly data is essential for your science, please

request an intermediate release via contact@darwintreeoflife.org. Be aware that it will be of inferior quality and likely to change substantially on release.

Q. We are working on a project with a certain species and need the data/assembly urgently, what can we do to speed it up?

A. Please contact contact@darwintreeoflife.org to describe your case. Our project management will aim to prioritise your species of interest, if that is possible and resources are available. Even if it is prioritised, we can not guarantee it will be completed rapidly – various technical issues can cause indefinite delays.

Q. Where can I 'browse' along a Darwin Tree of Life genome assembly?
A. Annotated DToL genomes can be browsed through the Ensembl Genome Browser, via our Rapid Release website. The easiest way to access the data is to use our Ensembl DToL Project Page. From here use the "View in browser column" to open your species of interest in the Ensembl Genome Browser.

Q. Where can I do a 'BLAST' search against a Darwin Tree of Life genome assembly?
A. You can BLAST against a genome of interest for any genome released in the Ensembl Genome Browser. On our Rapid Release website, you can access BLAST via the banner at the top of any page. This will allow you to select one or more species to include in the BLAST session. You may BLAST against either the genomes or the proteomes.

Data Issues

Q. I have been using DToL data and have identified some data or metadata issues. Who should I contact?

A. You can email dataupdates@darwintreeoflife.org for any issues regarding metadata or data.

Q. I have looked at the genome sequence (or the specimen photograph) and I disagree with your species identification. Who should I discuss this with?

A. You can email dataupdates@darwintreeoflife.org for any issues regarding species identification

DToL services

Q. What is ToLQC and how does it work?

A. ToLQC https://tolqc.cog.sanger.ac.uk/ provides an overview and QC analysis of all raw data produced for a certain species. You can use it to check on the sequencing progress of a species, have a look at the quality of produced data, download it (if it is already released to the public archives) and have a look at the QC of the possibly

already produced draft assemblies. There is a short video explaining how to use the resource [here](#).

Publications

Q. We would like to use DToL data in our publication. Is that ok and what do we need to cite?

A. Yes. All our assemblies and data are being generated for the public to make best use of them, no restrictions. We are releasing genome notes for all assembled species (https://wellcomeopenresearch.org/treeoflife or access via the portal at https://www.darwintreeoflife.org/genomes/genome-notes/), please cite these. If there is no genome note released yet, please cite the INSDC accession of the assembly used and acknowledge the Darwin Tree of Life Project. Where applicable please also cite the sample accessions of any project data used and the umbrella accession for the project (INSERT). This greatly assists us with understanding the impact of the data.

Collaborations

Q. We have generated data for species X and now found out DToL has, too. Can we work together to generate an assembly and analysis?

A. Contact DToL Project Management via contact@darwintreeoflife.org who will work out the best way forward with you.