

Consolidating, curating and using existing datasets for ash genomics

Project overview

Ash dieback is a highly destructive disease of ash trees, especially the European ash, one of the UK's most common woodland trees. Caused by an invasive fungus (*Hymenoscyphus fraxineus*), ash dieback is predicted to cost the UK £7.6 billion by 2030.

The aim of this project was to uncover the genetic basis of resistance to ash dieback, which can be used to inform potential ash breeding programmes and, ultimately, assist the long-term survival of native ash populations.

In this project, we (a) analysed genome sequencing data from a natural population of 580 ash trees affected by ash dieback in Surrey, (b) increased public accessibility to ash genomic data, (c) annotated a new, more complete, genome assembly for an eastern European ash tree, (d) re-analysed field trial data using the new genome assembly to identify possible genes associated with ash dieback resistance, and (e) investigated the contribution of structural variations in the ash genome to ash dieback resistance.

Key findings

(a) Analysis of a natural ash population:

- Analysis of genome sequencing data from a natural ash population in Surrey showed that natural selection appears to be increasing the population's resistance to ash dieback.
- The rate of natural selection may not be enough to rescue European ash populations from the disease, but adaptation could be accelerated by a breeding programme informed by genomic selection.
- Genomic Estimated Breeding Values, GEBV are a way of predicting a tree's offspring's resistance to ash dieback using genetic data. This is more accurate than predictions based on the parent tree's disease symptoms.
- A [paper](#) reporting this work has been published on BioRxiv and submitted to a journal.

(b) Increasing public accessibility to ash genomics data:

- Ash genome sequencing data previously hosted only on ashgenome.org have now been made available in an [ash community on Zenodo](#).
- The sequence data from the natural ash population were made available on the ENA short read archive under accession [PRJEB44697 \(ERP128769\)](#).

(c) Re-annotating ash reference genome sequences:

- A new, more complete, genome assembly for an eastern European ash tree, built by our Polish collaborators, has been annotated with descriptions of the predicted structures and functions of genes and elements within it.

(d) Re-analysis of previous ash dieback genome datasets using the new reference assemblies:

- Re-analysis of data from ash dieback field trials in the UK and Europe using the new genome assembly identified a larger, partially different, set of genes associated with ash dieback resistance compared with the previous study. This indicates that a more complete genome allows the identification of additional genes potentially involved with ash dieback resistance, and that the reference genome used for genomic studies can significantly affect which genes are identified as being associated with resistance.

(e) Analysis of the contribution of structural variants to ash dieback resistance:

- Preliminary analyses of structural variations in the ash genome (where DNA sequences are inverted, duplicated, deleted or moved elsewhere) suggest that structural differences are likely to be involved in resistance to ash dieback.

Ongoing and future work

Our results suggest that, to gain a fuller understanding of the genetic basis of ash dieback resistance, a pan-genome (a reference genome that incorporates information about the structural variation across a species) for ash is needed. This work is now being completed under the Centre for Forest Protection's 2022–2025 research plan, through a project titled, "[Transforming ash genomics: Creating a pangenome to understand ash dieback resistance](#)".