The genome of *Fraxinus excelsior* (European Ash)

Elizabeth Sollars, Laura Kelly, Bernardo Clavijo, David Swarbreck, Jasmin Zohren, David Boshier, Jo Clark, Anika Joecker, Sarah Ayling, Mario Caccamo, Richard Buggs.

@LizzySollars
It’s a tough time to be an ash tree

**Ash Dieback** *(Hymenoscyphus fraxineus)*

---

**Emerald Ash Borer** *(Agrilus planipennis)*
Workflow

DNA
- Mitochondrial assembly
- Genome Assembly
  - Resequencing 37 European trees
  - Ab initio gene prediction
    - Gene Sharing Analysis
    - Functional annotation

RNA

Workflow

- Mitochondrial assembly
- DNA
- Genome Assembly
- Resequencing 37 European trees
- Ab initio gene prediction
- RNA
- Gene Sharing Analysis
- Functional annotation
We know...
23 chromosome pairs
880 Mbp genome
Selfed

We sequenced...
Wood tissue
200x coverage
Illumina HiSeq + 454
Pairs 200bp – 40 kbp

We assembled...
875 Mbp in 89,285 scaffolds
N50 of 99 kbp
19% in gaps
446 kbp mt. genome
33 complete, 20 incomplete mt. genes

We used...
gsAssembler for 454 reads
CLC Genomics Workbench
SSPACE scaffolder
SOAP’s GapCloser
35% of the genome is classed as repeats.

- **32%** LTR Gypsy
- **18%** LTR Copia
- **12%** LINE
- **2%** SINE
- **1%** Class II (DNA transposons)
- **0.01%** Class I: retrotransposons
- **2%** Unclassified
- **35%** Tandem repeats
Workflow

DNA
- Mitochondrial assembly
- Genome Assembly
  - Resequencing 37 European trees
  - Ab initio gene prediction
  - Gene Sharing Analysis
  - Functional annotation

RNA
- Ab initio gene prediction
Workflow

- DNA
  - Mitochondrial assembly
  - Genome Assembly
    - Resequencing 37 European trees
  - Ab initio gene prediction
    - Functional annotation
- RNA
  - Gene Sharing Analysis
Gene finding & Annotation

Repeat-masked genome + protein alignments

AUGUSTUS   AUGUSTUS with RNA-Seq     MAKER

Evidence Modeller

43,298 genes, 59,154 transcripts.

Annotated with BLAST and GO terms
Gene finding

Full-LengtherNEXT Statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sure Complete</td>
<td>60</td>
</tr>
<tr>
<td>Sure C-terminus</td>
<td>10</td>
</tr>
<tr>
<td>Sure N-terminus</td>
<td>5</td>
</tr>
<tr>
<td>Sure Internal</td>
<td>1</td>
</tr>
<tr>
<td>Sure Mis-assembled</td>
<td>0</td>
</tr>
<tr>
<td>Putative Coding</td>
<td>0</td>
</tr>
<tr>
<td>Putative Coding</td>
<td>0</td>
</tr>
<tr>
<td>Putative ncRNA</td>
<td>0</td>
</tr>
<tr>
<td>Putative Unknown</td>
<td>0</td>
</tr>
</tbody>
</table>
The open home of the British Ash Tree Genome project

- Genome Assemblies
- Transcriptome assemblies
- JBrowse
- BLAST tool
Workflow

DNA:
- Mitochondrial assembly
- Genome Assembly
  - Resequencing 37 European trees

RNA:
- Ab initio gene prediction
  - Gene Sharing Analysis
  - Functional annotation

European trees
Workflow

DNA
- Mitochondrial assembly
- Genome Assembly
- Resequencing 37 European trees
- Ab initio gene prediction
- Gene Sharing Analysis

RNA
- Ab initio gene prediction
- Functional annotation
Gene Sharing

Clustering of orthogroups using OrthoMCL

- FEXC: Fraxinus excelsior
- MGUT: Mimulus guttatus
- UGIB: Utricularia gibba
- SLYC: Solanum lycopersicum
- CROB: Coffea robusta

Total orthogroups: 18,870
Gene Sharing

Clustering of orthogroups using OrthoMCL

- FEXC: Fraxinus excelsior
- MGUT: Mimulus guttatus
- UGIB: Utricularia gibba
- SLYC: Solanum lycopersicum
- CROB: Coffea robusta

The largest ash-specific orthogroups are:
- Disease resistance proteins
- F-box proteins
- Expansin-like proteins

Total orthogroups: 18,870
Gene Sharing

Clustering of orthogroups using OrthoMCL:

- FEXC: *Fraxinus excelsior*
- CROB: *Coffea robusta*
- PTRI: *Populus trichocarpa*
- ATRI: *Amborella trichopoda*
- PITA: *Pinus taeda*

In 11-species comparison:

- 4,283 groups shared by all
- 7,562 shared by all angiosperms

A group of laccase proteins were expanded in the woody species: role in lignin degradation.
(ash, pine, & poplar)
Workflow

- DNA
  - Mitochondrial assembly
  - Genome Assembly
    - Resequencing 37 European trees
- RNA
  - Ab initio gene prediction
  - Functional annotation
    - Gene Sharing Analysis
Resequencing: Diversity Panel
Diversity Panel

In 355 Mb of the reference sequence:
- Reads from 37 trees, ~10x coverage each.
- Total: **28.7 million** polymorphic positions (~ 1 every 12 bp).
- Low Frequency Variant Caller in CLC Genomics Workbench

407,911 causing non-synonymous amino acid change
15,689 at splice sites
Diversity Panel

- SNP tree made using **SNPhylo**
- 18.7 million polymorphic positions
- Filtered set of **14,000**
Diversity Panel

- SNP tree made using **SNPhylo**
- 18.7 million polymorphic positions
- Filtered set of **14,000**
Diversity Panel
Future Work

• Gene family expansions
  – In ash alone
  – In woody species

• Diversity panel
  – Network analysis
  – SNP allele frequencies among individuals
  – SNP verification - KASPAR

• Optical mapping using BioNano
We are the borers… resistance is futile.

The Fraxinus genus

- **F. excelsior**
  - EAB resistance

- **F. chinensis**
  - EAB resistance

- **F. mandshurica**
  - EAB and dieback resistance

- **F. quadrangulata**
  - Medium EAB resistance

- **F. americana**
  - Dieback resistance

- **F. pennsylvanica**
  - Medium dieback resistance
Funders