DATA NOTE

The genome sequence of the hazel leaf-roller, *Apoderus coryli* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Liam Crowley,1
University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

1Department of Zoology, University of Oxford, Oxford, UK

Abstract
We present a genome assembly from an individual male *Apoderus coryli* (the hazel leaf roller; Arthropoda; Insecta; Coleoptera; Attelabidae). The genome sequence is 428 megabases in span. The majority (98.90%) of the assembly is scaffolded into 20 chromosomal pseudomolecules, with the X and Y sex chromosomes assembled.

Keywords
Apoderus coryli, hazel leaf roller, genome sequence, chromosomal, Coleoptera

This article is included in the *Tree of Life* gateway.
Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Crowley L: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Crowley L et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Crowley L, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective et al. The genome sequence of the hazel leaf-roller, Apoderus coryli (Linnaeus, 1758) [version 1; peer review: awaiting peer review] Wellcome Open Research 2021, 6:315 https://doi.org/10.12688/wellcomeopenres.17380.1

First published: 18 Nov 2021, 6:315 https://doi.org/10.12688/wellcomeopenres.17380.1
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Attelabidae; Apoderinae; Apoderus; Apoderus coryli (Linnaeus, 1758) (NCBI:txid201766).

Background
The hazel leaf-roller, Apoderus coryli, is a distinctive leaf-rolling weevil which is widespread and locally common across Europe and western Asia. In the UK, it is common in deciduous woodland across England and Wales, with a few records from southern Scotland. The main host plant is common hazel, Corylus avellana, although it is also known to infrequently feed on alder, birch, hornbeam and beech. It may be considered a minor pest of hazelnut cultivation (Gantner, 2000). Adults are 6–8 mm with red elytra, pronotum and bases of the femora. The head, underside and remainder of the legs are black. It has a distinctive elongate head and ‘neck’. Adults are active from May, feeding on hazel leaves in which they leave characteristic round holes. Mating and oviposition occurs in early summer. Females lay a single bright orange egg around 1 mm in size on a leaf before cutting a line across the leaf with their mandibles and rolling it into a cylinder around the egg. The leaf rolling process may take 60–90 minutes and a single female may repeat this for up to 30 eggs across several days (Urban, 2014). Larvae develop and pupate within these leaf rolls. Second generation adults are active in late summer, the larvae of which overwinter within leaf rolls that drop to the ground and pupate in the spring. A previous study has shown that this species has 13 autosomal bivalents, that sex chromosomes exhibit a parachute association and that this species possesses a small amount of heterochromatin (Rozek et al., 2004).

Genome sequence report
The genome was sequenced from one male A. coryli (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.770, longitude -1.339). A total of 59-fold coverage in Pacific Biosciences single-molecule long reads and 93-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 57 missing/misjoins and removed 7 haplotypic duplications, reducing the assembly length by 0.41% and the scaffold number by 24.71%, and increasing the scaffold N50 by 24.75%.

The final assembly has a total length of 428 Mb in 128 sequence scaffolds with a scaffold N50 of 23.7 Mb (Table 1). The majority, 98.90%, of the assembly sequence was assigned to 20 chromosomal-level scaffolds, representing 18 autosomes (numbered by sequence length), and the X and Y sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.1.2 (Simão et al., 2015) completeness of 99.3% (single 97.7%, duplicated 1.6%) using the endopterygota_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods
Sample acquisition and DNA extraction
A single male A. coryli was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.770, longitude -1.339) by Liam Crowley, University of Oxford, using a pooter. The sample was snap-frozen on dry ice and stored using a CoolRack.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The icApoCory1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Tissue from the whole organism was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12-20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified
by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

**Sequencing**

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II and Illumina HiSeq X instruments. Hi-C data were generated from further whole organism tissue using the Arima v2 Hi-C kit and sequenced on a HiSeq X instrument.

**Genome assembly**

Assembly was carried out with Hifiasm ([Cheng et al., 2021](#)); haplotypic duplication was identified and removed with purge_dups ([Guan et al., 2020](#)) with the -e flag. One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with

---

**Table 1. Genome data for *Apoderus coryli*, icApoCory1.1.**

<table>
<thead>
<tr>
<th>Project accession data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly identifier</td>
</tr>
<tr>
<td>Species</td>
</tr>
<tr>
<td>Specimen</td>
</tr>
<tr>
<td>NCBI taxonomy ID</td>
</tr>
<tr>
<td>BioProject</td>
</tr>
<tr>
<td>BioSample ID</td>
</tr>
<tr>
<td>Isolate information</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Raw data accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific Biosciences SEQUEL II</td>
</tr>
<tr>
<td>10X Genomics Illumina</td>
</tr>
<tr>
<td>Hi-C Illumina</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genome assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly accession</td>
</tr>
<tr>
<td>Accession of alternate haplotype</td>
</tr>
<tr>
<td>Span (Mb)</td>
</tr>
<tr>
<td>Number of contigs</td>
</tr>
<tr>
<td>Contig N50 length (Mb)</td>
</tr>
<tr>
<td>Number of scaffolds</td>
</tr>
<tr>
<td>Scaffold N50 length (Mb)</td>
</tr>
<tr>
<td>Longest scaffold (Mb)</td>
</tr>
<tr>
<td>BUSCO* genome score</td>
</tr>
</tbody>
</table>

*BUSCO scores based on the endopterygota_odb10 BUSCO set using v5.1.2. C= complete [S= single copy, D= duplicated], F= fragmented, M= missing, n= number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/icApoCory1.1/dataset/CAJVRQ01/busc.
freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao et al., 2014) using SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016).
as described previously (Howe et al., 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021) and annotated with MitoFinder (Allio et al., 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Figure 3. Genome assembly of *Apoderus coryli*, icApoCory1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icApoCory1.1/dataset/CAJVRQ01/blob.
Figure 4. Genome assembly of *Apoderus coryli*, icApoCory1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icRhaFulv1.1/dataset/CAJPIC01.1/cumulative.

Ethics/compliance issues
The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this
Table 2. Chromosomal pseudomolecules in the genome assembly of *Rhagonycha fulva*, icRhaFulv1.1.

<table>
<thead>
<tr>
<th>INSDC accession</th>
<th>Chromosome</th>
<th>Size (Mb)</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OU452198.1</td>
<td>1</td>
<td>45.98</td>
<td>33.1</td>
</tr>
<tr>
<td>OU452199.1</td>
<td>2</td>
<td>45.50</td>
<td>32.8</td>
</tr>
<tr>
<td>OU452200.1</td>
<td>3</td>
<td>42.38</td>
<td>32.6</td>
</tr>
<tr>
<td>OU452201.1</td>
<td>4</td>
<td>30.93</td>
<td>32.1</td>
</tr>
<tr>
<td>OU452202.1</td>
<td>5</td>
<td>27.11</td>
<td>32.5</td>
</tr>
<tr>
<td>OU452203.1</td>
<td>6</td>
<td>23.69</td>
<td>32.7</td>
</tr>
<tr>
<td>OU452204.1</td>
<td>7</td>
<td>21.42</td>
<td>32.8</td>
</tr>
<tr>
<td>OU452205.1</td>
<td>8</td>
<td>20.46</td>
<td>32.7</td>
</tr>
<tr>
<td>OU452206.1</td>
<td>9</td>
<td>20.22</td>
<td>32.6</td>
</tr>
<tr>
<td>OU452207.1</td>
<td>10</td>
<td>18.89</td>
<td>32.9</td>
</tr>
<tr>
<td>OU452208.1</td>
<td>11</td>
<td>17.89</td>
<td>33.2</td>
</tr>
<tr>
<td>OU452209.1</td>
<td>12</td>
<td>17.13</td>
<td>33.6</td>
</tr>
<tr>
<td>OU452211.1</td>
<td>13</td>
<td>14.69</td>
<td>33.1</td>
</tr>
<tr>
<td>OU452212.1</td>
<td>14</td>
<td>13.58</td>
<td>33.5</td>
</tr>
<tr>
<td>OU452213.1</td>
<td>15</td>
<td>12.56</td>
<td>33.7</td>
</tr>
<tr>
<td>OU452214.1</td>
<td>16</td>
<td>11.87</td>
<td>33</td>
</tr>
<tr>
<td>OU452215.1</td>
<td>17</td>
<td>10.51</td>
<td>32.6</td>
</tr>
<tr>
<td>OU452216.1</td>
<td>18</td>
<td>10.30</td>
<td>32.9</td>
</tr>
<tr>
<td>OU452210.1</td>
<td>X</td>
<td>16.03</td>
<td>31.4</td>
</tr>
<tr>
<td>OU452217.1</td>
<td>Y</td>
<td>1.26</td>
<td>24.7</td>
</tr>
<tr>
<td>OU452218.1</td>
<td>MT</td>
<td>0.02</td>
<td>27.3</td>
</tr>
<tr>
<td>-</td>
<td>Unplaced</td>
<td>5.65</td>
<td>25.5</td>
</tr>
</tbody>
</table>

Figure 5. Genome assembly of *Apoderus coryli*, icApoCory1.1: Hi-C contact map. Hi-C contact map of the icApoCory1.1 assembly, visualised in HiGlass.
Table 3. Software tools used.

<table>
<thead>
<tr>
<th>Software tool</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hifiasm</td>
<td>0.12</td>
<td>Cheng et al., 2021</td>
</tr>
<tr>
<td>purge_dups</td>
<td>1.2.3</td>
<td>Guan et al., 2020</td>
</tr>
<tr>
<td>SALSA2</td>
<td>2.2</td>
<td>Ghurye et al., 2019</td>
</tr>
<tr>
<td>longranger align</td>
<td>2.2.2</td>
<td><a href="https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines">https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines</a></td>
</tr>
<tr>
<td>freebayes</td>
<td>1.3.1-17-gaa2ace8</td>
<td>Garrison &amp; Marth, 2012</td>
</tr>
<tr>
<td>MitoHiFi</td>
<td>2.1</td>
<td>Uliano-Silva et al., 2021</td>
</tr>
<tr>
<td>gEVAL</td>
<td>N/A</td>
<td>Chow et al., 2016</td>
</tr>
<tr>
<td>HiGlass</td>
<td>1.11.6</td>
<td>Kerpedjiev et al., 2018</td>
</tr>
<tr>
<td>PretextView</td>
<td>0.1.x</td>
<td><a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a></td>
</tr>
<tr>
<td>BlobToolKit</td>
<td>2.6.2</td>
<td>Challis et al., 2020</td>
</tr>
</tbody>
</table>

document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

The genome sequence is released openly for reuse. The A. coryli genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information
Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789929.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme collective are listed here: https://doi.org/10.5281/zenodo.5377053.


Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013542.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783559.

References
PubMed Abstract | Publisher Full Text | Free Full Text

PubMed Abstract | Publisher Full Text | Free Full Text

PubMed Abstract | Publisher Full Text | Free Full Text

PubMed Abstract | Publisher Full Text | Free Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

Publisher Full Text