



# Draft Genome Sequence and Annotation of the Lichen-Forming Fungus *Arthonia radiata*

Ellie E. Armstrong,<sup>a</sup> Stefan Prost,<sup>a,b</sup> Damien Ertz,<sup>c,d</sup> Martin Westberg,<sup>e</sup> Andreas Frisch,<sup>f</sup> Mika Bendiksby<sup>f</sup>

<sup>a</sup>Department of Biology, Stanford University, Stanford, California, USA

<sup>b</sup>Department of Integrative Biology, University of California, Berkeley, California, USA

<sup>c</sup>Department of Research, Botanic Garden Meise, Meise, Belgium

<sup>d</sup>Fédération Wallonie-Bruxelles, Direction Générale de l'Enseignement non Obligatoire et de la Recherche Scientifique, Brussels, Belgium

<sup>e</sup>Museum of Evolution, Uppsala University, Uppsala, Sweden

<sup>f</sup>NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

**ABSTRACT** We report here the draft *de novo* genome assembly, transcriptome assembly, and annotation of the lichen-forming fungus *Arthonia radiata* (Pers.) Ach., the type species for Arthoniomycetes, a class of lichen-forming, lichenicolous, and saprobic Ascomycota. The genome was assembled using overlapping paired-end and mate pair libraries and sequenced on an Illumina HiSeq 2500 instrument.

Here, we report the draft *de novo* genome assembly, transcriptome assembly, and annotation of *Arthonia radiata* (Pers.) Ach. (strain EZ20314). This lichen-forming fungus is the type species of *Arthonia* (Ach.) Ach., on which the Arthoniaceae Rchb. (Arthoniales, Arthoniomycetes), a large family of about 800 lichen-forming, lichenicolous, and saprobic Ascomycota, are based. *Arthonia radiata* is a common and polymorphic epiphyte on mainly smooth-barked deciduous trees throughout the Holarctic. It has also been reported in Africa and New Zealand (1). It has played a key role in recent efforts to develop a new classification of the Arthoniomycetes based on phylogenetic principles (2–4).

The genome was assembled using overlapping paired-end (PE) and mate pair (MP) libraries (with an average insert size of 5 to 8 kb) and sequenced with Illumina HiSeq 2500 v4 chemistry (2 × 125 bp). Assemblies were created using AllPaths-LG (5) and SPAdes (6). The best assembly was chosen based on assembly continuity and Benchmarking Universal Single-Copy Orthologs 2 (BUSCO2) (7) scores (using OrthoDB v9 data sets for fungi and ascomycetes, downloaded from <http://busco.ezlab.org>). AllPaths-LG provided the best assembly, with a contig  $N_{50}$  value of 1.2 Mb (46 contigs), a scaffold  $N_{50}$  value of 2.25 Mb (17 scaffolds), and a total sequence length of 33.5 Mb. Of all BUSCO2 genes, 99% for fungi (out of 290, none of which show duplication) and 94.8% for ascomycetes (out of 1,315, none of which show duplication) were present as complete genes in the assembly. This indicates the high quality of the presented genome.

Transcriptome assembly was performed using Trinity (8) and a combination of Hierarchical Indexing for Spliced Alignment of Transcripts v2 (HISAT2) (9) and StringTie (10). The best assembly was chosen based on assembly statistics and BUSCO2 scores. We further used different cleaning filters for read assembly with Trinity. First, we ran Trinity on the raw reads. Then, we filtered and removed only adapter sequences and low-quality bases. Last, we performed a full filtering for contaminants, adapter sequences, and low-quality bases. The combination of AdapterRemoval (11) and Trinity resulted in the best transcriptome assembly, with 27,220 contigs and an  $N_{50}$  value of 6.8 kb. BUSCO2 scores indicated the presence of 94% complete and 5.5% fragmented

Received 19 March 2018 Accepted 20 March 2018 Published 5 April 2018

**Citation** Armstrong EE, Prost S, Ertz D, Westberg M, Frisch A, Bendiksby M. 2018. Draft genome sequence and annotation of the lichen-forming fungus *Arthonia radiata*. Genome Announc 6:e00281-18. <https://doi.org/10.1128/genomeA.00281-18>.

**Copyright** © 2018 Armstrong et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mika Bendiksby, [mika.bendiksby@ntnu.no](mailto:mika.bendiksby@ntnu.no).

E.E.A. and S.P. contributed equally to this work.

transcripts (2 out of 290 are missing) for fungal BUSCO2 genes and 92.5% complete and 5.5% fragmented transcripts for ascomycete BUSCO2 genes (26 out of 1,315 are missing).

Next, we annotated repeats in the best genome assembly using a combination of RepeatMasker (homology-based annotation using the Repbase database) and Repeat-Modeler (*de novo* repeat annotation) (see <http://repeatmasker.org> for both programs). We found that 16.65% of the *A. radiata* genome constitutes repeat sequences, most of which were LTR elements (14.94% of the whole genome).

Gene annotation was performed using Maker3 (12), with simple repeats only soft masked in the repeat-masking step. We annotated 6,931 genes.

**Accession number(s).** This whole-genome shotgun project has been deposited at GenBank (assembly number GCA\_002989075), and all the data are available at NCBI (BioProject number PRJNA432823, BioSample number SAMN08462631) under the accession number [PSQN00000000](https://doi.org/10.1093/bioinformatics/btv351) (locus numbers PSQN01000001 to PSQN01000017).

## REFERENCES

- Coppins BJ, Aptroot A. 2009. *Arthonia* Ach. (1806), p 153–171. In Smith CW, Aptroot A, Coppins BJ, Fletcher A, Gilbert OL, James PW, Wolseley PA (eds), The lichens of Great Britain and Ireland. British Lichen Society, London, United Kingdom.
- Ertz D, Tehler A. 2011. The phylogeny of Arthoniales (Pezizomycotina) inferred from nuLSU and RPB2 sequences. *Fungal Divers* 49:47–71. <https://doi.org/10.1007/s13225-010-0080-y>.
- Ertz D, Tehler A, Irestedt M, Frisch A, Thor G, van den Boom P. 2015. A large-scale phylogenetic revision of Roccellaceae (Arthoniales) reveals eight new genera. *Fungal Divers* 70:31–53. <https://doi.org/10.1007/s13225-014-0286-5>.
- Frisch A, Thor G, Ertz D, Grube M. 2014. The Arthonialean challenge: restructuring Arthoniaceae. *Taxon* 63:727–744. <https://doi.org/10.12705/634.20>.
- Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494. <https://doi.org/10.1038/nprot.2013.084>.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357. <https://doi.org/10.1038/nmeth.3317>.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290. <https://doi.org/10.1038/nbt.3122>.
- Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 5:337. <https://doi.org/10.1186/1756-0500-5-337>.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.