

A Short-Read Genome Assembly Resource for *Leveillula taurica* Causing Powdery Mildew Disease of Sweet Pepper (*Capsicum annuum*)

Stefan Kusch,¹ Márk Z. Németh,² Niloofar Vaghefi,³ Heba M. M. Ibrahim,⁴ Ralph Panstruga,¹ and Levente Kiss^{2,3,†}

¹ Unit of Plant Molecular Cell Biology, Institute for Biology I, RWTH Aachen University, Aachen, Germany

² Plant Protection Institute, Centre for Agricultural Research, Budapest, Hungary

³ Centre for Crop Health, University of Southern Queensland, Toowoomba, Australia

⁴ Division of Plant Biotechnics, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Powdery mildew of sweet pepper (*Capsicum annuum*) is an economically important disease. It is caused by *Leveillula taurica*, an obligate biotrophic ascomycete with a partly endophytic mycelium and haustoria, i.e., feeding structures formed in the mesophyll cells of infected host plant tissues. The molecular basis of its pathogenesis is largely unknown because genomic resources only exist for epiphytically growing powdery mildew fungi with haustoria formed exclusively in epidermal cells of their plant hosts. Here, we present the first reference genome assembly for an isolate of *L. taurica* isolated from sweet pepper in Hungary. The short read-based assembly consists of 23,599 contigs with a total length of 187.2 Mbp; the scaffold N₅₀ is 13,899 kbp and N₉₀ is 3,522 kbp; and the average GC content is 39.2%. We detected at least 92,881 transposable elements covering 55.5 Mbp (30.4%). BRAKER predicted 19,751 protein-coding gene models in this assembly. Our reference genome assembly of *L. taurica* is the first resource to study the molecular pathogenesis and evolution of a powdery mildew fungus with a partly endophytic lifestyle.

Genome Announcement

Powdery mildew fungi (Erysiphales, Ascomycota) are obligate biotrophic phytopathogens infecting over 10,000 plant species including important crops. Most powdery mildew species have epiphytic mycelia with haustoria, i.e., feeding structures that are produced exclusively in the epidermal cells of their plant hosts. Members of the genera *Leveillula*, *Pleochaeta*, and *Phyllactinia*, represent a distinct phylogenetic lineage named the endophytic group (Takamatsu 2013). Representatives of this lineage have partly endophytic mycelia and their haustoria are exclusively found in plant mesophyll cells (Kunoh et al. 1979; Zheng et al. 2013). It is likely that the effector protein arsenal, the secretory pathways of these proteins, and other molecular features associated with the infection process are different in the members of the endophytic lineage compared with those known from epiphytically

Funding

This work was partly supported by a grant of the Australia-Germany Joint Research Co-Operation Scheme, and the German Academic Exchange Service (DAAD)-funded program PPP Australia 2019. It was further supported by German Research Foundation (Deutsche Forschungsgemeinschaft [DFG]) grant number 861/14-1 to R. Panstruga in the context of the DFG-funded priority program ‘Rapid evolutionary adaptation–potential and constraints’ SPP1819 and also by the University of Southern Queensland.

Keywords

Leveillula taurica, powdery mildew, sweet pepper

[†]Corresponding author: L. Kiss; levente.kiss@usq.edu.au

The sequencing datasets produced for this study, the *L. taurica* genome assembly, transcriptome, and genome sequencing data are deposited in the EMBL-EBI European Nucleotide Archive under the project reference PRJEB36538. The corresponding scripts used in this study are available on GitHub.

The author(s) declare no conflict of interest.

Accepted for publication 6 March 2020.

propagating powdery mildew fungi. To date, genomic resources have not been provided for any representative of the endophytic lineage.

Wherever grown, most sweet pepper (*Capsicum annuum*) cultivars suffer regularly from powdery mildew infection. If not controlled by fungicide applications, the disease may cause serious yield losses. The causal agent, *Leveillula taurica*, belongs to the endophytic lineage of the Erysiphales, and it is also known as a pathogen of other cultivated and wild plants across a number of plant families, including Solanaceae, Malvaceae, and Alliaceae (Braun and Cook 2012). Currently, it is considered that the binomial *L. taurica* refers to a species complex that includes several separate species with distinct and still unrevealed host ranges (de Souza and Café-Filho 2003; Khodaparast et al. 2001, 2007, 2012). Here, we present the *ab initio* annotated draft genome of a *L. taurica* isolate infecting *C. annuum*, obtained by short-read next-generation sequencing.

Powdery mildew-infected leaves of sweet pepper cv. Kard were collected in a commercial plastic tunnel house in Csongrád, Hungary, in June 2014. Conidia collected from a single colony were used to infect a 2-month-old, healthy potted sweet pepper plant of the same cultivar in an experimental glasshouse in Budapest, Hungary. Following the appearance of the first powdery mildew colonies on the infected plants, the isolate, designated LTCaps-HU, was identified as *L. taurica* based on morphology and ribosomal RNA internal transcribed spacer (ITS) sequence (GenBank accession number MT125856), which was determined as described by Kovács et al. (2011). The isolate was then maintained for 5 years on mature potted sweet pepper plants, cv. Totál, in the same glasshouse. Plants were replaced every 2 to 4 months with newly grown and freshly inoculated plants to ensure the continuous propagation of the isolate. A herbarium specimen was deposited at the Mycological Collection of the Hungarian Natural History Museum under accession number HNHN-MYC-006405 (111091BP). The epiphytic parts of the powdery mildew mycelium were removed from infected leaves collected in the glasshouse with cellulose acetate peelings as described by Frantzeskakis et al. (2019). Acetate peelings were collected in 50-ml plastic tubes placed on ice and were ground to fine powder with liquid nitrogen, using a mortar and pestle. To extract high-molecular weight DNA, the protocol described by Feehan et al. (2017) was followed, omitting the second DNA precipitation. The quality of the DNA was assessed by agarose gel electrophoresis (0.8% agarose, 60 V) and spectrophotometric measurement with a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, MA, U.S.A.). The quantity of DNA was determined using a Qubit2 fluorometer (Thermo Scientific) and the DNA Broad Range reagent (Invitrogen Thermo Scientific, Schwerte, Germany). Short-read DNA sequencing was conducted on a NovaSeq6000 machine by CeGaT (Tübingen, Germany).

Complementary whole-transcriptome shotgun sequencing (RNA-seq) was performed using isolate LTCaps-Aus, collected from commercially grown *C. annuum* cv. Plato in Stanthorpe, Queensland, Australia, in March 2019, and maintained on cv. Plato in the experimental glasshouse of the University of Southern Queensland in Toowoomba, Australia, as described above for LTCaps-HU. The ITS sequence of LTCaps-Aus (GenBank accession number MT125857) was identical to LTCaps-HU. A herbarium specimen for LTCaps-Aus was deposited at Queensland Plant Pathology Herbarium under accession number BRIP 70887. To perform RNA-seq, 1-cm² pieces were cut out of mature pepper cv. Plato leaves, were infected with LTCaps-Aus, were flash-frozen in liquid nitrogen, and were then ground with stainless steel beads (2.8 mm diameter; Sigma-Aldrich) in a FastPrep-24 (MP Biomedicals, Sydney, Australia) (6.5 m/s, 2 × 15 s). RNA was isolated using a RNeasy Mini kit (Qiagen, Melbourne, Australia) following the manufacturer instructions. Genomic DNA was removed via DNase treatment (Qiagen). Quality and concentrations of RNA were assessed via agarose gel electrophoresis and a Qubit 3.0 fluorometer using an RNA Broad Range kit. Short-read RNA sequencing was conducted by the Australian Genome Research Facility (Melbourne, Australia) on a NextSeq platform to acquire 150-bp paired-end reads. The reads were quality- and adapter-trimmed using Trimmomatic v0.36, applying the following settings: ILLUMINACLIP:TruSeq2-PE.fa:5:30:10 SLIDINGWINDOW:3:18 LEADING:6 TRAILING:6 MINLEN:90 (Bolger et al. 2014). The quality of the RNA-seq data was assessed via FastQC v0.11.5 (Babraham bioinformatics, Cambridge). The reads were mapped to the newly generated reference assembly of *L. taurica* using HISAT2 (Kim et al. 2015) with -max-intronlen 1000 -k 10; output files were parsed with samtools v1.7 (Li et al. 2009) and bedtools v2.25.0 (Quinlan and Hall 2010) to obtain BAM and coverage files.

Genome assembly was performed using SPAdes (Bankevich et al. 2012) and SOAPdenovo2 (Luo et al. 2012). Then, we used BlobTools (Laetsch and Blaxter 2017) with

Table 1. Genome assembly statistics for *Leveillula taurica* isolate LTCaps-HU, isolated from *Capsicum annuum* cv. Totál

Feature/method	SPAdes	SOAPdenovo2	High confidence ^a
			SOAPdenovo2
Genome [bp]	166,797,524	187,262,869	170,696,302
N count	0	2,326,588	2,009,231
Scaffold/contigs	19,422	23,599	18,294
Putative mitochondrial contigs	–	16	16
Mitochondrial genome [bp]	–	30,644	30,644
Largest contig [bp]	94,988	99,596	99,596
Average length [bp]	8,588.1	7,935.2	9,330.7
GC content	39.1%	39.2%	39.2%
N ₅₀	12,261	13,899	15,209
N ₉₀	3,899	3,522	4,454
BUSCO ^b	S: 85.0%	S: 87.9%	S: 86.0%
	D: 0.0%	D: 0.0%	D: 0.0%
	F: 1.3%	F: 1.4%	F: 1.4%
	M: 13.7%	M: 10.7%	M: 12.6%
Repeat elements	29.6%	30.4%	29.8%
Genes (BRAKER)	–	19,751	–

^a “High confidence” indicates contigs with BLASTN similarity to published powdery mildew genomes.

^b Using the database for ascomycete core genes (ascomycota_odb9). S = complete single-copy BUSCOs, D = duplicated, F = fragmented, and M = missing BUSCOs.

RNA-seq data mapped back to the draft assembly and the results from BLASTN against the National Center for Biotechnology Information GenBank Nucleotide (NT) database (updated 11/2019) to remove non-Leotiomyecete contaminant contigs; no-hit contigs with a GC content and mapping coverage in the range of the Leotiomyecete contigs were not removed. We assessed the completeness of the genome assembly with BUSCO v3 (Simão et al. 2015) against the ascomycete_odb9 database. RepeatMasker v4.0.7 (available online) with the repeat database version RepBase-20170127 was used to quantify the extent of repetitive sequences and transposable elements. We used BRAKER (Hoff et al. 2016) for unsupervised gene prediction, with Augustus v3.3.1 trained on the annotation of *Blumeria graminis* f. sp. *hordei* isolate DH14 (Frantzeskakis et al. 2018) and the RNA-seq data from *Capsicum annuum* infected with *L. taurica* as input.

The SPAdes assembly resulted in 798,194 contigs and a total length of 712,628,793 bp. Filtering contigs based on size (>2 kb), identity (retaining only Erysiphaceae and no-hit contigs) as well as GC content (0.35 to 0.55) and remapping coverage (>50× average contig coverage) removed presumed contaminant contigs, likely originating from the host plant (*C. annuum*) and microbial epiphytic communities. These microbes were sequenced together with the powdery mildew mycelium removed with the cellulose acetate peelings. The SPAdes *L. taurica* draft genome assembly contained 19,422 contigs totaling 166,797,524 bp (Table 1). The completeness of the draft genome was assessed using BUSCO (ascomycete_odb9 database, 1,315 core genes), which found 85% of ascomycete core genes to be completely and 1.3% to be partially present.

The SOAPdenovo2 assembly generated 374,920 contigs with a total length of 260,244,152 bp. We filtered the contigs similarly, using the same pipeline as described above, which yielded 187,262,869 bp (including 2,326,588 Ns) of genome sequence on 23,599 scaffolds or contigs. The predicted completeness of the genome was higher than in the SPAdes assembly (1,156 complete BUSCO core genes [87.9%] and 19 fragmented core genes [1.4%]) (Table 1). Due to the higher completeness and the somewhat improved N₅₀ statistics (13,899 versus 12,261), we selected the SOAPdenovo2-based assembly as the reference assembly for *L. taurica*. Then, we conducted genome size estimation, using jellyfish 30-mers (Marçais and Kingsford 2011), which resulted in a total estimated genome size of 192.7 Mbp, suggesting that our draft assembly recovered a high percentage of the *L. taurica* genome, albeit highly fragmented as a result of short-read sequencing. We performed a RepBase search with RepeatMasker v4.0.7, which masked 30.4% of the genome and identified 63,352 retroelements covering 47,367,299 bp, 14,994 DNA transposons comprising 3,632,182 bp, and 14,535 unclassified transposable elements totaling 4,514,616 bp of sequence. We downloaded a list of curated fungal mitochondrial genes (available online from the MitoFun database) and used BLASTN to find contigs that may be of mitochondrial origin. We identified 16 contigs (E value < 1E-5) that harbor at least one mitochondrial gene; these

contigs have a total length of 30,644 bp. Finally, we used BRAKER (Hoff et al. 2016) for *ab initio* gene prediction, using the RNA-seq sample recovered from the Australian *L. taurica* isolate LTCaps-Aus. BRAKER-Augustus predicted 19,751 gene models in the reference assembly of *L. taurica*.

Using BLASTN, we aligned the 23,599 scaffolds against all available powdery mildew genomes available to date (January 2020) to obtain additional evidence for their identity, and we found 18,294 to have reliable BLASTN hits (E value < 1E-25). These contigs contain 170,696,302 bp of sequence and comprise a slightly more contingent assembly (N_{50} = 15,209 bp); however, 1.9% of BUSCO ascomycete core genes were lost by this assembly (Table 1). Therefore, we did not apply further filtering to the *L. taurica* reference genome assembly.

To conclude, we generated the first draft genome assembly and *ab initio* gene models for *L. taurica*, a powdery mildew pathogen with partly endophytic mycelium and haustoria produced exclusively in the photosynthesizing mesophyll cells of its plant hosts. To date, all genomic resources available for powdery mildew fungi have been sourced from species with epiphytic mycelia and haustoria produced in the epidermal cells of host plant tissues (Barsoum et al. 2019; Bindschedler et al. 2016). The dataset produced here will be highly valuable for future studies toward understanding the differences between epiphytic and endophytic lifestyles of powdery mildew pathogens.

Acknowledgments

The authors are grateful to E. Tóth, I. Csorba, K. Hadzi, and L. Huth for excellent technical support, A. Pintye for her assistance with powdery mildew maintenance, and S. Tancred (Orchard Services Pty Ltd, Stanthorpe, Queensland, Australia) for his help with field collections of *L. taurica*.

Author-Recommended Internet Resources

GitHub, *Leveillula taurica* scripts: https://github.com/stefankusch/genome_Leveillula_2020
MitoFun database: <http://mitofun.biol.uoa.gr/index.html>

Literature Cited

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.
- Barsoum, M., Sabelleck, B., Spanu, P. D., and Panstruga, R. 2019. Rumble in the effector jungle: Candidate effector proteins in interactions of plants with powdery mildew and rust fungi. *Crit. Rev. Plant Sci.* 38:255-279.
- Bindschedler, L. V., Panstruga, R., and Spanu, P. D. 2016. Mildew-Omics: How global analyses aid the understanding of life and evolution of powdery mildews. *Front. Plant Sci.* 7:123.
- Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
- Braun, U., and Cook, R. T. A. 2012. Taxonomic Manual of the Erysiphales (Powdery Mildews). CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands.
- de Souza, V. L., and Café-Filho, A. C. 2003. Resistance to *Leveillula taurica* in the genus *Capsicum*. *Plant Pathol.* 52:613-619.
- Feehan, J. M., Scheibel, K. E., Bourras, S., Underwood, W., Keller, B., and Somerville, S. C. 2017. Purification of high molecular weight genomic DNA from powdery mildew for long-read sequencing. *J. Vis. Exp.* 121:e55463.
- Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and Panstruga, R. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* 19:381.
- Frantzeskakis, L., Németh, M. Z., Barsoum, M., Kusch, S., Kiss, L., Takamatsu, S., and Panstruga, R. 2019. The *Parauncinula polyspora* draft genome provides insights into patterns of gene erosion and genome expansion in powdery mildew fungi. *MBio* 10:e01692-19.
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. 2016. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767-769.
- Khodaparast, S. A., Niinomi, S., and Takamatsu, S. 2007. Molecular and morphological characterization of *Leveillula* (Ascomycota: Erysiphales) on monocotyledonous plants. *Mycol. Res.* 111:673-679.
- Khodaparast, S. A., Takamatsu, S., Harada, M., Abbasi, M., and Samadi, S. 2012. Additional rDNA ITS sequences and its phylogenetic consequences for the genus *Leveillula* with emphasis on conidium morphology. *Mycol. Prog.* 11:741-752.
- Khodaparast, S. A., Takamatsu, S., and Hedjaroude, G.-A. 2001. Phylogenetic structure of the genus *Leveillula* (Erysiphales: Erysiphaceae) inferred from the nucleotide sequences of the rDNA ITS region with special reference to the *L. taurica* species complex. *Mycol. Res.* 105:909-918.
- Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12:357-360.
- Kovács, G. M., Jankovics, T., and Kiss, L. 2011. Variation in the nrDNA ITS sequences of some powdery mildew species: Do routine molecular identification procedures hide valuable information? *Eur. J. Plant Pathol.* 131:135-141.
- Kunoh, H., Kohno, M., Tashiro, S., and Ishizaki, H. 1979. Studies of the powdery-mildew fungus, *Leveillula taurica*, on green pepper. II. Light and electron microscopic observation of the infection process. *Can. J. Bot.* 57:2501-2508.
- Laetsch, D. R., and Blaxter, M. L. 2017. BlobTools: Interrogation of genome assemblies. *F1000 Res.* 6:1287.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., and Wang, J. 2012. SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18.
- Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764-770.
- Quinlan, A. R., and Hall, I. M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Takamatsu, S. 2013. Molecular phylogeny reveals phenotypic evolution of powdery mildews (*Erysiphales*, Ascomycota). *J. Gen. Plant Pathol.* 79: 218-226.
- Zheng, Z., Nonomura, T., Bóka, K., Matsuda, Y., Visser, R. G. F., Toyoda, H., Kiss, L., and Bai, Y. 2013. Detection and quantification of *Leveillula taurica* growth in pepper leaves. *Phytopathology* 103:623-632.