Published in final edited form as:
Mol Plant Microbe Interact. 2018 December ; 31(12): 1227–1229. doi:10.1094/MPMI-06-18-0163-A.

Draft Genome Resource for the Potato Powdery Scab Pathogen *Spongospora subterranea*

Stefan Ciaghi[1], **Sigrid Neuhauser**[1], and **Arne Schwelm**[2,†]

[1]University of Innsbruck, Institute of Microbiology, Technikerstraße 25, 6020 Innsbruck, Austria

[2]Swedish University of Agricultural Sciences, Department of Plant Biology, Uppsala BioCenter, Linnean Centre for Plant Biology, P.O. Box 7080, SE-75007 Uppsala, Sweden

Abstract

The Plasmodiophorida (Phytomyxea, Rhizaria) are a group of protists that infect plants. Of this group, *Spongospora subterranea* causes major problems for the potato industry by causing powdery scab and root galling of potatoes and as vector for the *Potato mop-top virus* (PMTV) (genus *Pomovirus*, family *Virgaviridae*). A single tuber isolate (SSUBK13) of this uncultivable protist was used to generate DNA for Illumina sequencing. The data were assembled to a draft genome of 28.08 Mb consisting of 2,340 contigs and an L50 of 280. A total of 10,778 genes were predicted and 93% of the BUSCO genes were detected. The presented genome assembly is only the second genome of a plasmodiophorid. The data will accelerate functional genomics to study poorly understood interaction of plasmodiophorids and their hosts.

Genome Announcement

Spongospora subterranea belongs to the Plasmodiophorida (Phytomyxea, Rhizaria) (Neuhauser et al. 2014; Sierra et al. 2016). This soilborne, obligate biotrophic protist is present in most potato-growing regions worldwide (Gau et al. 2013). It causes powdery scab of potato tubers, root galling in its host, reduces tuber yield, and transmits the *Potato mop-top virus* (PMTV) (genus *Pomovirus*, family *Virgaviridae*) (Balendres et al. 2016; Falloon et al. 2016). It can also diminish the growth of other solanaceous crops, such as tomatoes (*Solanum lycopersicum*) (Balendres et al. 2018). Because of their intracellular growth and obligate host–dependent life style, genomic data for plasmodiophorids are rare, although those pathogens cause substantial crop damage, e.g., in brassicas, sugar beet, or gramineous crops (Schwelm et al. 2018). So far, the only publicly available whole-genome data of plasmodiophorids are from the clubroot pathogen *Plasmodiophora brassicae* (Bi et al. 2016; Schwelm et al. 2015). Therefore, the here-presented *S. subterranea* data now enables comparative analyses between plasmodiophorid plant pathogens.

The genome data described here were generated from DNA of purified sporosori (characteristic aggregates of haploid resting spores) from potatoes originating from Lower Saxony, Germany (IB2016400). Sporosori from an individual potato tuber were collected to

†Corresponding author: Arne Schwelm; arne.schwelm@slu.se.

Europe PMC Funders Author Manuscripts

generate a single tuber isolate (SSUBK13). To reduce the amount of bacterial contamination, sporosori were immersed in lysis solution (GeneJET plasmid miniprep kit; Thermo-Scientific) for 5 min before washing twice by centrifugation in sterile water. Sporosori were purified by sequential filtering using CellTrics strainers (Sysmex). Material between 20 and 50 mm was collected and a subsample of the sporosori was visually checked for the presence of fungal and bacterial contamination prior DNA extraction. Sporosori were snap-frozen in liquid nitrogen and disrupted using a Fastprep 24 5G homogenizer (MP Biomedicals, LLC). For lysis, the material was incubated in cetyltrimethylammonium bromide buffer at 65°C for 10 to 12 h before iso-amylacohol and chloroform extraction and isopropanol precipitation of the DNA. A yellow-brown pigment coprecipitated with the DNA was separated from the DNA via gel electrophoresis. The genomic DNA of a size larger than 10 kb was purified from the gel using the ZYMO large-fragment DNA gel purification kit (Zymo Research Corporation). A strand-specific paired-end library (2× 150 bp) was constructed and was sequenced using HiSeq 2500 technique (Illumina) at the Vienna Biocenter Core Facilities sequencing platform. Reads were quality checked using FastQC and sequencing adapters, and low-quality reads (sliding window 5 bp; average quality score <20) were removed using Trimmomatic v0.36 (Bolger et al. 2014). Only surviving read pairs in which each read was longer than 70 bp were processed further. To filter out reads from contaminating organisms (e.g., bacteria) and the host, reads that matched to the potato genome (Potato Genome Sequencing Consortium et al. 2011) or the National Center for Biotechnology Information (NCBI) bacterial reference database using Bowtie2 v2.2.6 (Langmead and Salzberg 2012) were removed. The so-filtered reads were assembled using Velvet v1.2.10 (Zerbino and Birney 2008) and SPAdes v3.11.1 (Bankevich et al. 2012). Velvet assembled (kmer set to 31) contigs (≥1,000 nt, coverage ≥13) were used in the SPAdes assembly (kmer settings 23, 33, 41, 55, 63) as untrusted contigs. From this assembly only contigs longer than 500 bp were used and were manually sorted based on coverage, blastN hit, and length. Contigs with coverage above 30 were selected when their BlastN hit did not exclusively match bacterial genomes or the potato genome. All other contigs were analyzed by BlastX, using the NCBI nr database and *S. subterranea* protein models (Schwelm et al. 2015) and contigs with their best hits matching only to *P. brassicae* or *S. subterranea* protein models were also kept. Gapfiller v1.10 (Boetzer and Pirovano 2012) and SSPACE-standard v3.0 (Boetzer et al. 2011) were used to improve scaffolding.

The *S. subterranea* SSUBK13 genome draft presented here consists of 2,340 scaffolds with a total length of 28.08 Mb (229 N/100 kbp) and 1,150 scaffolds above 5 kb. The largest scaffold was 262,481 nt long (N50 value of 28,677; L50 value of 280). The GC content of 45.7% is lower than the 58% for *P. brassicae* (Schwelm et al. 2015) but comparable to the 46% GC content of the genome of the rhizarian *Bigellowiella natans* (Curtis et al. 2012). After repeat-masking (RepeatMasker program), a total of 10,778 genes were predicted using the BRAKER1 v2.0 pipeline (Hoff et al. 2016), a TopHat v2.1.0 (Kim et al. 2013)-created BAM-file, and available RNAseq reads (Schwelm et al. 2015; European Nucleotide Archive [ENA] BioProject PRJEB9159). The completeness of the genome assembly was estimated using BUSCO v2.0.1 (Simão et al. 2015), identifying 93% (88% complete, 5% partial) of the Eukaryota set. For *P. brassicae*, the only other sequenced Phytomyxea to date, 97% (94% complete, 3% partial) of BUSCO proteins are present in the isolate e3 (Schwelm et al.

2015). The presented genome draft contains the complete ribosomal DNA sequence (scaffold 678). The mitochondrial sequence was partially assembled from the unfiltered reads using MITObim (Hahn et al. 2013). Of the 10,780 predicted genes 1,991 could be categorized by KAAS (Moriya et al. 2007) to a KEGG category and 4,402 by eggNOG (Huerta-Cepas et al. 2016) to a COG (clusters of orthologous groups) category other than S, which is unknown function. SingalP v4.1 (Petersen et al. 2011) predicted 697 secreted proteins, of which 102 were classified as potential effectors by the fungal effector predictor EffectorP 2.0 (Sperschneider et al. 2018). Using dbCAN (Yin et al. 2012), 300 carbohydrate active enzymes (CAZymes) were predicted (e value < e^{-05}).

The whole-genome sequence assembly (accessions OUQQ01000001 to OUQQ01002340) and the complete data set used in this report, including the Illumina raw reads have been deposited at the European Nucleotide Archive (ENA) under BioProject PRJEB26377 (Table 1).

Acknowledgments

Library preparation and Illumina Sequencing was performed at the Vienna Biocenter Core Facilities Next Generation Sequencing Unit.

Funding

This work and Arne Schwelm were supported by FORMAS Swedish Research Council (grant 2015-1317). Stefan Ciaghi and Sigrid Neuhauser were funded by the Austrian Science Fund (grant Y0810-B16).

Literature Cited

Balendres MA, Tegg RS, Amponsah J, Wilson CR. Zoosporangial root infection of tomato by *Spongospora subterranea* in hydroponic and glasshouse culture results in diminished plant growth. J Phytopath. 2018; 166:412–419.

Balendres MA, Tegg RS, Wilson CR. Key events in pathogenesis of *Spongospora* diseases in potato: A review. Australas Plant Pathol. 2016; 45:229–240.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19:455–477. [PubMed: 22506599]

Bi K, He Z, Gao Z, Zhao Y, Fu Y, Cheng J, Xie J, Jiang D, Chen T. Integrated omics study of lipid droplets from *Plasmodiophora brassicae*. Sci Rep. 2016; 6

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27:578–579. [PubMed: 21149342]

Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012; 13:R56. [PubMed: 22731987]

Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–2120. [PubMed: 24695404]

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature. 2012; 492:59–65. [PubMed: 23201678]

Falloon RE, Merz U, Butler RC, Curtin D, Lister RA, Thomas SM. Root infection of potato by *Spongospora subterranea*: Knowledge review and evidence for decreased plant productivity. Plant Pathol. 2016; 65:422–434.

Gau RD, Merz U, Falloon RE, Brunner PC. Global genetics and invasion history of the potato powdery scab pathogen, *Spongospora subterranea* f.sp. *subterranea*. PLoS One. 2013; 8:e67944. [PubMed: 23840791]

Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. Nucleic Acids Res. 2013; 41:e129. [PubMed: 23661685]

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016; 32:767–769. [PubMed: 26559507]

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016; 44(D1):D286–D293. [PubMed: 26582926]

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007; 35:W182–W185. [PubMed: 17526522]

Neuhauser S, Kirchmair M, Bulman S, Bass D. Cross-kingdom host shifts of phytomyxid parasites. BMC Evol Biol. 2014; 14:33. [PubMed: 24559266]

Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8:785–786. [PubMed: 21959131]

Potato Genome Sequencing Consortium. Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011; 475:189–195. [PubMed: 21743474]

Schwelm A, Badstöber J, Bulman S, Desoignies N, Etemadi M, Falloon RE, Gachon CMM, Legreve A, Lukeš J, Merz U, Nenarokova A, et al. Not in your usual Top 10: Protists that infect plants and algae. Mol Plant Pathol. 2018; 19:1029–1044. [PubMed: 29024322]

Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja T, Bonilla-Rosso G, Karlsson M, Shevchenko A, Dhandapani V, Choi SR, Kim HG, et al. The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. Sci Rep. 2015; 5

Sierra R, Cañas-Duarte SJ, Burki F, Schwelm A, Fogelqvist J, Dixelius C, González-García LN, Gile GH, Slamovits CH, Klopp C, Restrepo S, et al. Evolutionary origins of rhizarian parasites. Mol Biol Evol. 2016; 33:980–983. [PubMed: 26681153]

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31:3210–3212. [PubMed: 26059717]

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol Plant Pathol. 2018; 19:2094–2110. [PubMed: 29569316]

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012; 40:W445–W451. [PubMed: 22645317]

Zerbino DR, Birney E. Velvet: Algorithms for *d. novo* short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. [PubMed: 18349386]

Author-Recommended Internet Resource

RepeatMasker program: http://www.repeatmasker.org

Europe PMC Funders Author Manuscripts

Table 1

Metrics of the *Spongospora subterranea* SSUBK13 genome draft

Feature	Numbers
Assembled genome size	28,084,886 nt
Number of scaffolds (>1 kb)	2,340 (2,255)
N50	28,677
L50	280
Largest scaffold	262,481 nt
Scaffolds >25 kb	339
BUSCOs	93% (271 complete, 15 partial)
Number of predicted genes	10,780
Number of predicted CAZymes	300
Number of predicted secreted proteins	697