Complete Genome Sequence of *Spongospora subterranea* from North America

Augustina K. Arjarquah, Jatinder Singh, Kimberly Zitnick-Anderson, Binod Pandey, Ipsita Mallik, Upinder Gill, and Julie S. Pasche*
North Dakota State University, Department of Plant Pathology; Fargo, N, USA
Julie.Pasche@NDSU.edu

Introduction

Obtaining quality genomic DNA free from host and microbial contaminants of *Spongospora subterranea* f. sp. *subterranea* (Wallroth) Lagerheim is very challenging due to durability of the sporosori and the obligate pathogen life-style. A genome is available for this globally economically important pathogen, but little to no post-sequencing analyses were completed. Here, we present a complete genome with annotations by combining long- and short-read DNA sequences.

Methods

- Sporosori were collected from root galls on a single plant cultivar 'Russet Burbank' for DNA isolation followed by long-read sequencing using Oxford Nanopore Technology (ONT) (Figure 1).
- Illumina short-read sequencing was completed on DNA isolated from sporosori in pustules on tubers from a single commercial field in the same region.

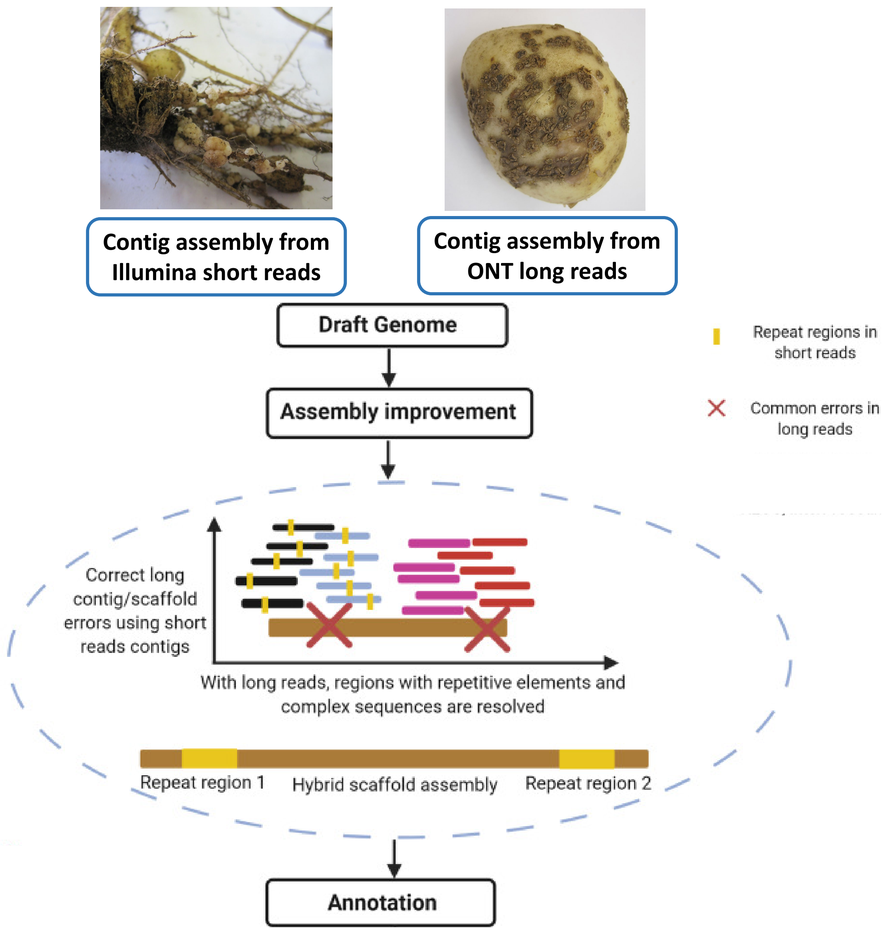

Figure 1. Schematic illustration of the *Spongospora subterranea Sss*MN22-1 whole genome assembly workflow. Adapted from Rodriguez-Anaya et al. 2021.

Results

Table 1. Genome assembly and annotation features of *Spongospora subterranea, Sss*MN22-1

Features	SssMN22-1
Sample	Root galls and tuber pustules
Location	Minnesota, USA
Sequencing method	ONT long-reads and Illumina 150 paired-end reads
Genome size (Mb)	31.5
Number of contigs	352
Number of scaffolds	346
Scaffold N_{50} (kb)	122
GC content (%)	45.7
Largest scaffold (bp)	2,088,320
Percent gaps (%)	0.001
Repeats and transposable elements (bp)	27.0%
BUSCO[a]	96.1%
Number of predicted genes	10,325
Proteins with KEGG[b] assignment	1981
Proteins with COG[c] assignment	2466
Proteins annotated with CAZymes[d]	135
Pfam[e]	2249
Number of secreted proteins predicted	700
Predicted fungal effectors from secreted proteins	321

[a]Benchmarking Universal Single-Copy Orthologs
[b]Kyoto Encyclopedia of Genes and Genomes
[c]Cluster of Orthologous Genes
[d]Carbohydrate Active Enzymes
[e]Protein Families

Table 2. Comparing genome assembly and annotation features of Plasmodiophorids

Feature	*S. subterranea*, *Sss*MN22-1 (Arjarquah et al. submitted)	*S. subterranea*, SSUBK13 (Ciaghi et al. 2018)	*P. brassicae* (Schwelm et al. 2015)
Sample	Root galls and tuber pustules	Resting spores	Single spore
Sequencing method	Nanopore-MinION and Illumina Miseq	Illumina Hiseq 2500	Roche 454 FLX and Illumina Hiseq 2100
Genome size (Mb)	31.51	28.08	25.5
Number of scaffolds	346	2,340	165
Scaffold N_{50} (kb)	122	28.6	473
GC content	45.70%	45.70%	58.50%
Percent gaps	0.00%	0.71%	1.43%
Repeats and TE[a]	26.97%	10.67%	5.45%
BUSCO[b]	96.10%	93.10%	97.00%
Predicted coding genes	10,325	10,778	9,730

[a]Transposable Element
[b]Benchmarking Universal Single-Copy Orthologs

Results

- The genome was predicted to have 10,325 protein-coding genes, including 321 potential fungal effectors, 700 signal peptide proteins, 135 carbohydrate active enzymes, 1981 KEGG assigned genes, and 2660 Clusters of Orthologous Genes (COGs).

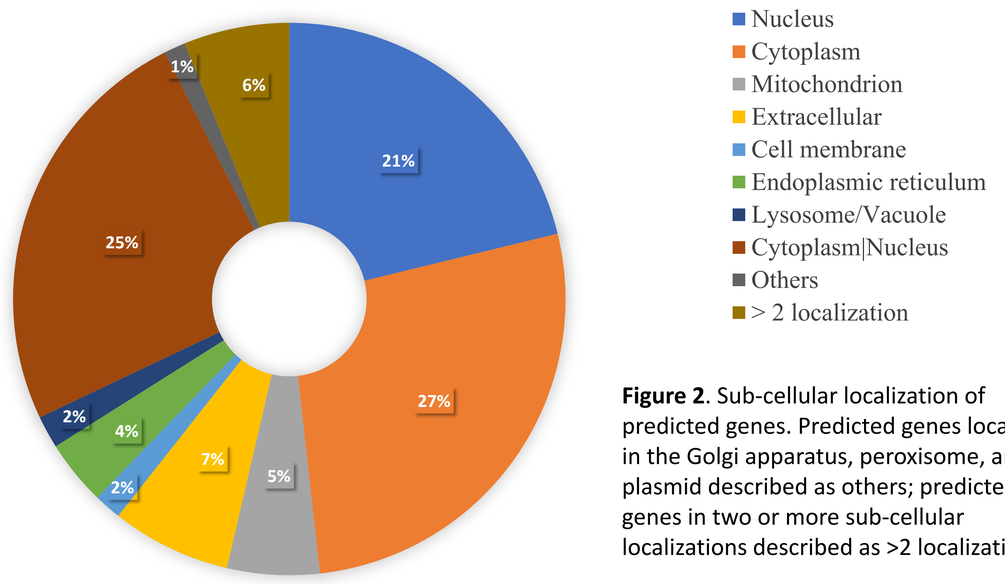

Figure 2. Sub-cellular localization of predicted genes. Predicted genes localized in the Golgi apparatus, peroxisome, and plasmid described as others; predicted genes in two or more sub-cellular localizations described as >2 localization.

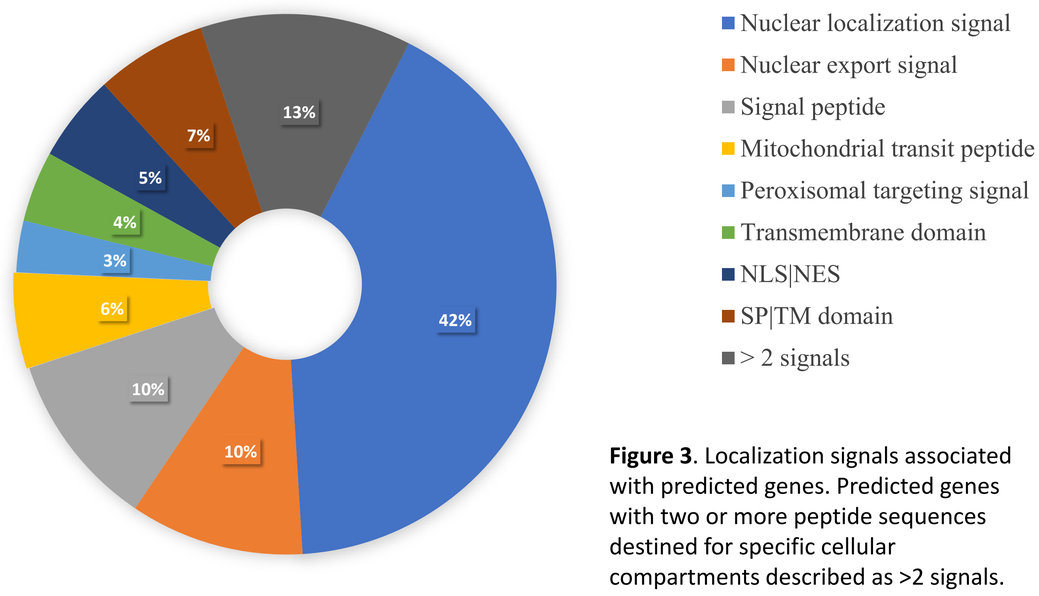

Figure 3. Localization signals associated with predicted genes. Predicted genes with two or more peptide sequences destined for specific cellular compartments described as >2 signals.

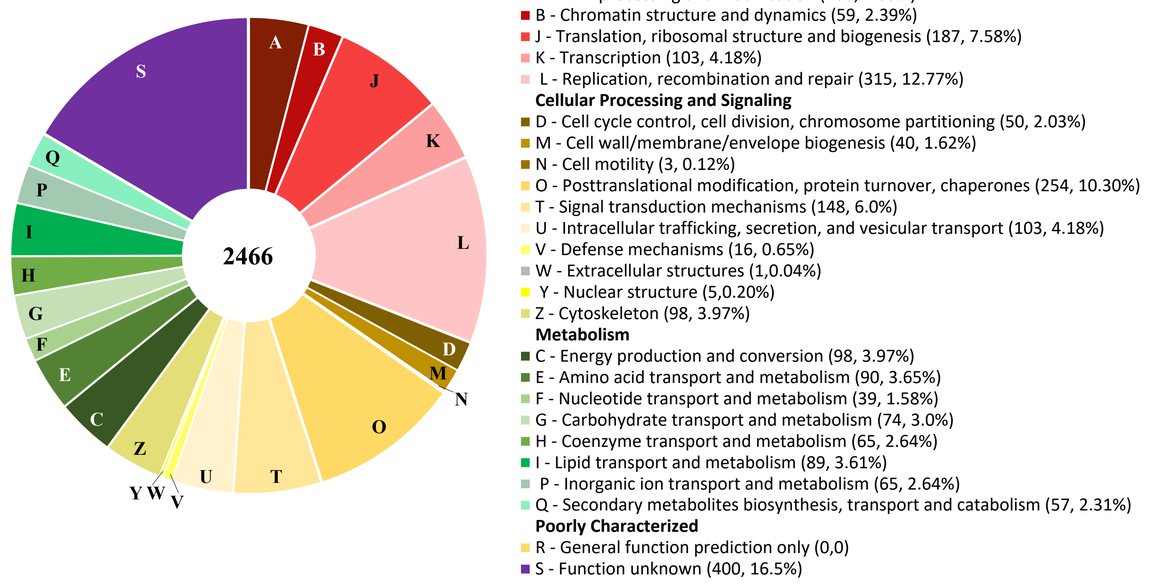

Figure 4. The phylogenetic classification of 2466 cluster of orthologous genes (COGs) from *Sss*MN22-1. Each functional category is denoted by a one-letter abbreviation.

Conclusions

The new draft reference genome resource described in this study:

- Provides a basis for increased understanding of international pathogen populations
- Enhances our understanding of the biology of *S. subterranea*
- Aids in the development of new strategies for managing powdery scab and PMTV of potato
- Help in identifying potential targets for disease control including the development of resistant crop varieties
- May increase knowledge of the evolution and spread of this economically significant plant pathogen
- Can reveal novel genes and pathways that could be exploited for biotechnology applications

Data Availability: Genome sequence data is available upon request

NDSU | NORTH DAKOTA STATE UNIVERSITY