Name of the fungal species

1. Abc
2. Def
3. Xyz

Sequencing data (Please write in short, the data that we have for assemblies)

Strategy for genome assembling (for each fungal genome)

Step 1:

Combine HiSeq & MiSeq reads from all lanes and both sonication

Step 2:

Generate several multiple assemblies using [assemblers🡪 genomic scaffolders 🡪 RNA scaffolders] and pipelines. Remove bacterial and eukaryotic contamination sequences.

Step 3:

Select best genome assembly using FGMP (<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2782-9>)

Step 4:

Perform genomic analysis:

* Estimate genome size (Check ‘Determination of the genome size’ in Results section of (Neu et al., 2017))
* Analyze repeats and TE

Step 5:

Annotate genome using FINDERv1

Functional annotation of the genes

Perform BUSCO analysis

Phylogenetic analysis of the genomes

Collect ideas from (Frantzeskakis et al., 2019)

List of pipeline, assemblers and scaffolders:

Assemblers

1. Masurca (<http://www.genome.umd.edu/masurca.html>)
2. ~~SWAP-Assembler2 (~~[~~https://ieeexplore.ieee.org/abstract/document/7573818~~](https://ieeexplore.ieee.org/abstract/document/7573818)~~)~~
3. SOAPdenovo
4. ALLPATHS-LG [Needs jumping reads]
5. ABySS 2.0 (<https://genome.cshlp.org/content/27/5/768>)
6. SPAdes
7. PASHA (Last Update on 2013) (<http://pasha.sourceforge.net/homepage.htm#latest>)
8. ~~ScalaDBG (~~[~~https://www-nature-com.proxy.lib.iastate.edu/articles/s41598-019-51284-9~~](https://www-nature-com.proxy.lib.iastate.edu/articles/s41598-019-51284-9)~~) [Lesser Priority – does not show comparison with other genomic assemblers]~~

Scaffolders

1. OPERA-LG (<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0951-y>)
2. InGAP (<https://academic.oup.com/nar/article/45/6/e43/2638393>)
3. iLSLS (<https://ieeexplore.ieee.org/abstract/document/8416733>)
4. AGOUTI (<https://academic.oup.com/gigascience/article/5/1/s13742-016-0136-3/2558793>) [Uses RNA-Seq reads]
5. Rascaf (<file:///Users/gsfuerst/Downloads/tpg-9-3-plantgenome2016.03.0027.pdf>) [Uses RNA-Seq reads]
6. P\_RNA\_scaffolder (<https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4567-3?optIn=false>)

Pipelines: (Includes assembly and scaffolding)

1. GATB-minia-pipeline (<https://github.com/GATB/gatb-minia-pipeline>)
2. HipMer (Contacted JGI to get software) (<https://people.eecs.berkeley.edu/~aydin/sc15_genome.pdf>)
3. Meraculous (<https://jgi.doe.gov/data-and-tools/meraculous/>)
4. IMAP (<http://github.com/jkimlab/IMAP>)

Points to discuss with MG: (24/03/2020)

* Conda installation – with masurca and abyss
* Masurca run over
* Soapdenovo run over
* Won’t do Allpaths-lg since they require jump reads

Date Sept 3rd, 2020

* Downloaded experimentally verified proteins from Uniprot (Type in database:(type:ensemblfungi) AND reviewed:yes in the search box) September 2020. Total of 28270 sequences were obtained.
* How to run BRAKER
* Create conda environment ‘conda create --name BRAKER2’
* conda activate BRAKER2
* conda install braker2 (This will take quite a long time so be patient. It will finish eventually!!)
* conda install -c bioconda genomethreader
* Check readme document about information about how to run BRAKER2

Reference:

Frantzeskakis, L., Németh, M. Z., Barsoum, M., Kusch, S., Kiss, L., Takamatsu, S., et al. (2019). The parauncinula polyspora draft genome provides insights into patterns of gene erosion and genome expansion in powdery mildew fungi. *MBio* 10. doi:10.1128/mBio.01692-19.

Neu, E., Featherston, J., Rees, J., and Debener, T. (2017). A draft genome sequence of the rose black spot fungus Diplocarpon rosae reveals a high degree of genome duplication. doi:10.1371/journal.pone.0185310.