In-house long read sequencing yields affordable superb fungal genome assemblies

Sørensen, Trine; Petersen, Celine; Westphal, Klaus Ringsborg; Sondergaard, Teis Esben; Sørensen, Jens Laurids; Nielsen, Kåre Lehmann

Publication date:
2020

In-house long read sequencing yields affordable superb fungal genome assemblies

Celine Petersen[1], Trine Sørensen[1], Klaus R. Westphal[1], Lavinia I. Fechete[1], Teis E. Søndergaard[1], Jens L. Sørensen[2], and Kåre L. Nielsen[1]

[1]Department of Chemistry and Bioscience, Fredrik Bajers Vej 7H, 9220 Aalborg, Aalborg University, Denmark
[2]Department of Chemistry and Bioscience, Niels-Bohrs Vej 8, 6700 Esbjerg, Aalborg University, Denmark

Introduction

The traditional way to discover novel secondary metabolites from fungi is to screen the compounds produced under permissive conditions. However, this is quite tedious and labor intensive since numerous conditions must be screened for each fungus and still no guarantee of achieving permissive conditions exist. An alternative approach is whole genome sequencing followed by genome mining based on predictions of conserved domains, which can be utilized to identify secondary metabolite gene clusters of interest – thus predicting the synthesis potential of filamentous fungi. Promising candidate gene clusters can then be overexpressed *in situ* or in a heterologous production host.

Aim

Establish an affordable, efficient, robust, and semi-automated sequencing pipeline for fungal genomes utilizing the MinION platform from Oxford Nanopore Technologies to generate long reads spanning entire gene clusters.

Methods

Fungal incubation

Penicillium spp. are grown on solid media for ~14 days. Then the fungi are transferred to liquid media for ~5 days.

DNA extraction

DNA is extracted using modified phenol-chloroform or Genomic DNA buffer set from QIAGEN and subsequently purified with QIAGEN Genomic-tip 20/G to ensure high-molecular-weight DNA.
The resulting DNA is evaluated using TapeStation, UV absorption (NanoDrop), and fluorescence spectroscopy (Qubit).

Third generation sequencing

Library preparations are performed using Ligation Sequencing Kit from Oxford Nanopore Technologies. If several fungi are multiplexed then Native Barcoding Expansion is used in addition. The libraries are sequenced on a MinION R9.4.1. or R10 flowcell.

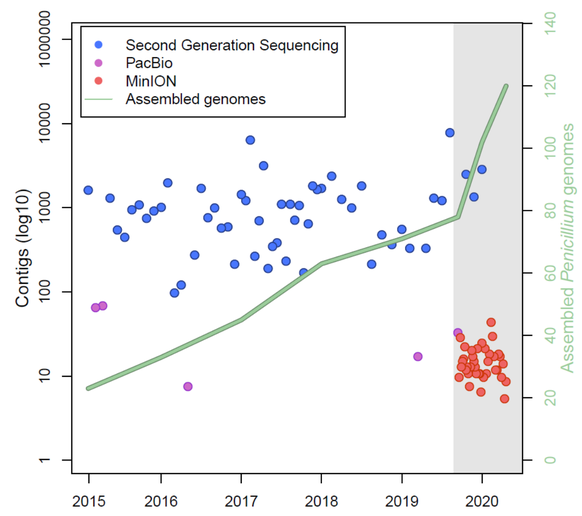
Data processing

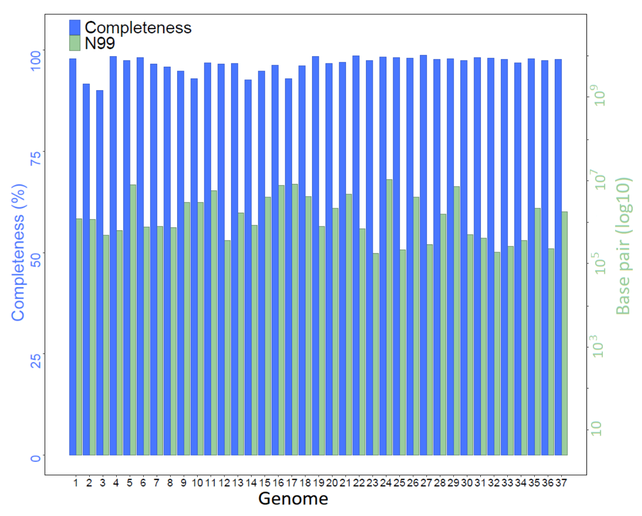
The raw reads are basecalled, trimmed, and assembled. Extensive polishing is performed to decrease error rate. The assemblies are subsequently evaluated based on continuity and completeness. AntiSMASH fungal version is used in initial genome mining.

Conclusion

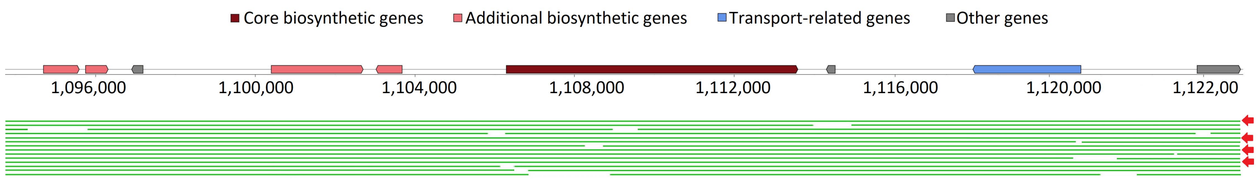
- The sequencing pipeline is straightforward and can be performed by non-specialists in their own laboratory.
- Assembled genomes are significantly less fragmented with comparable consensus sequence accuracy.
- All 37 genomes have a completeness above 90 %.
- Long single reads enables sequencing of entire gene clusters avoiding artefact chimeric gene clusters during bioinformatics analysis.

Results

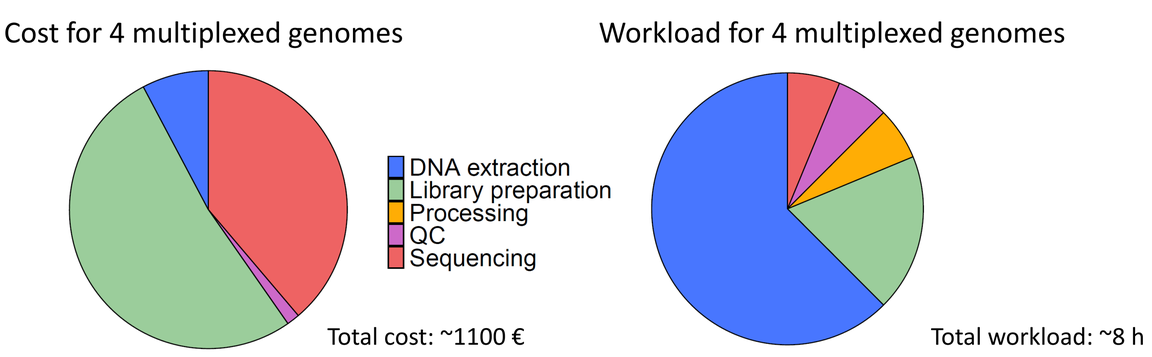




Continuity and completeness. With regards to continuity, the *Penicillium* spp. sequenced as part of this project at Aalborg University are superior to all *Penicillium* spp. published in NCBI from 2015 until 2020 with second generation sequencing. The completeness of the sequenced genomes is all above 90 % which are assessed from gene content in ascomycota. N99 is the minimum contig length needed to cover 99 % of the genome, and is in the range of 94,550 to 4,884,854.



Single reads span entire gene cluster. A major advantage of third generation sequencing is that single reads can cover entire gene clusters giving experimental validation of candidate gene clusters. This will exclude the possibility of artefact chimeric gene clusters, which is at higher risk in assemblies made by short reads. In the gene cluster (38,082 bp) above a section of the mapped reads are shown. In this section, four reads completely cover the gene cluster, indicated by the red arrows, besides the additional fragmented reads.



The cost and workload of the sequencing pipeline. MinION sequencing is affordable and efficient due to easy access and simplicity. The major costs are the flowcell and library preparation, but the start up price is minimal. In regards to workload, DNA extraction is the most time-consuming part, but necessary to ensure high-molecular-weight DNA with good quality.

cepe@bio.aau.dk
trso@bio.aau.dk

novo nordisk fonden

The data will be uploaded to GenBank following a period of 12 months.