# Nanopore sequencing of the genome of an apple blossom isolate of *Metschnikowia pulcherrima*

**Yvonne Bösch, Jürg E. Frey, Florian Freimoser, Andreas Bühlmann**
Agroscope, CH-8820 Wädenswil; www.agroscope.ch

## Introduction

DNA Sequencing is of great importance in various applications of genetics and molecoular biology. Novel technologies such as Nanopore Sequencing allow sequencing of DNA fragments up to a size of 900kb. Here we aimed at testing Nanopore sequencing for sequencing of an apple blossom isolate of *Metschnikowia pulcherrima* with possible use in biocontrol of postharvest diseases.

### Principles of Nanopore Sequencing

Nanopores are nanoscale holes, formed by proteins sitting in a lipid bilayer. Nanopores are immerged in a buffer solution with dissolved ions. If voltage is applied across the bilayer, an electric current is created by ions passing through. Substances different from buffer solution passing through the pore produce a characteristic disruption in the ionic current. Tracing of this current disruption allows to define the characteristics of these molecules (Oxford Nanopore Technologies, 2016).

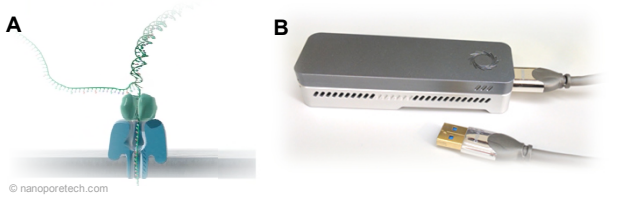


**Figure 1: A)** Illustration of a Nanopore placed in a biological memprane sequencing a DNA strand **B)** The Nanopore sequencing device MinION of Oxford Nanopore Technologies.

## Results

### Sequencing

Sequencing of *Metschnikowia pulcherrima* DNA using six flow cells (R9.4) yielded total sequence data of 1.3Gbp. The mean read length was 5800bp and the mean Phred score values were 13.6. While other users in the Nanopore community report read lengths of up to 900Mbp the longest mappable read generated within this study was 125kbp.

### *De novo* assembly and comparison to existing genomes

A *de novo* assembly approach in Canu v1.5 with only Nanopore data resulted in 13 contigs wit a N50 of 3.62Mbp. A comparison to an inhouse generated PacBio assembly resulted in a genome wide identity of 99.1%. Polishing of the initial assembly using 100x Illumina MiSeq data in Pilon resulted in an identity of 99.9%. In order to assess assembly completeness, the sequenced reads were mapped back at the *de novo* assembly. 98.2% of the Illumina reads mapped to the assembly while a *de novo* assembly of the unmapped reads using Velvet v1.2.10 did not yield contigs longer than 600bp suggesting that the present assembly is nearly complete. A comparison to existing sequencing projects on Metschnikowia shows a similar genome size and similar G+C contents, while the present assembly has the lowest contiguity (Table 1).

| Species | Contigs | Genome MB | G+C content % |
|---|---|---|---|
| *Metschnikowia pulcherrima* | 13 | 16.1 | 45.7 |
| *Metschnikowia australis* | 154 | 14.4 | 47.2 |
| *Metschnikowia bicuspidata* var. *bicuspidata* | 48 | 16.1 | 47.9 |
| *Metschnikowia frcuticola* | 93 | 26.1 | 45.9 |

**Table1:** Comparison of the present *M. pulcherrima* genome sequenced using Nanopore to existing *Metschnikowia* sequencing projects on NCBI.

A comparison of ribosomal rRNA ITS regions showed a close relationship of *M. fructicola* with *M. pulcherrima* (Fig 2A) whereas within species diversity showed the isolate sequenced within this study (red) being a representative isolate of the species *M. pulcherrima* (Fig 2B).

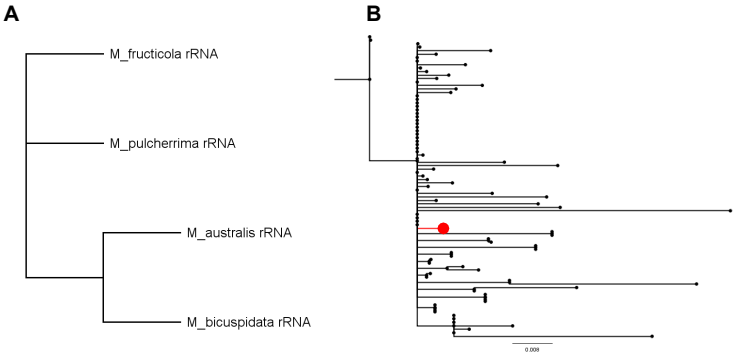


**Figure 2: A)** Maximum likelihood tree of the rRNA ITS regions of *Metschinkowia* species with sequenced genomes. **B)** Maximum likelihood tree of the rRNA ITS regions of all *Metschnikowia pulcherrima* isolates on NCBI.

## Conclusions

We found Nanopore sequencing being a promising sequencing technology bearing the potential to replace current standard sequencing techniques for various application. The advantages are:

- High contiguity assemblies
- *De novo* sequencing and assembly of yeast genomes to 99% accuracy
- Error correction with Illumina sequence data to 99.9% accuracy
- Future applications include direct RNA sequencing and direct sequencing of modified bases