**Supplementary text:** Programs and commands used for quality filtering, mapping, visualization, and SNP calling for genome sequencing data and assembly of long-read SMRT sequencing reads of the *Zymoseptoria ardabiliae* reference strain Za17.

**Quality filtering**

**Trimmomatic version 0.36** (Bolger *et al.*, 2014)

java -jar trimmomatic-0.36.jar PE -threads 8 -phred33 R1\_sample\_1.fastq.gz R2\_sample\_1.fastq.gz R1\_paired.fastq R1\_unpaired.fastq R2\_paired.fastq R2\_unpaired.fastq HEADCROP:10 CROP:149 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50

**Mapping**

**bowtie2 version 2.1.0** (Langmead & Salzberg, 2012)

bowtie2 -p8 -q -x reference -1 R1\_paired.fastq -2 R2\_paired.fastq -S sample\_1.sam

**Sorting and indexing**

**samtools version 1.3.1** (Li, 2011)

samtools view -Sb sample\_1.sam > sample\_1.bam

samtools sort sample\_1.bam > sample\_1\_sorted.bam

samtools index sample\_1\_sorted.bam

**Extraction of discordant reads mapping to different chromosomes**

samtools view -F 14 -h input\_sorted.bam | awk '($3!=$7 && $7!="=")' > output\_discordant.sam

**Visualization**

deeptools version 2.5.3 (Ramírez *et al.*, 2016)

data is normalized to 1x coverage  
**bamCoverage** --bam input\_sorted.bam -o output.bw --binSize 10 --normalizeTo1x 39686251

**SNP calling**

**samtools + bcftools version 1.3.1** (Li, 2011)

samtools mpileup -E -C50 -Q20 -q20 -uf reference.fasta sample\_1\_sorted.bam | bcftools call --ploidy 1 -vc -O u -o sample\_1.bcf

bcftools view sample\_1.bcf > sample\_1.vcf

bcftools filter -o sample\_1\_filtered.vcf -e'QUAL<20 | DP<10 | AF1<0.8' sample\_1.vcf

**Assembly of long-read SMRT sequencing *Zymoseptoria ardabiliae* reference strain Za17**

High molecular weight DNA of the *Z. ardabiliae* strain Za17 was isolated as described in (Allen *et al.*, 2006). Single molecule real time sequencing (SMRT) was performed on a PacBio RS II using 5 SMRT cells at the Functional Genomics Center Zürich, Switzerland.

The genome was assembled as previously described in (Plissonneau & Stürchler, 2016) using the assembler SMRTanalysis v 2.3.0 implemented in the Hierarchical Genome Assembly Process version 3 (HGAP3) (Chin *et al.*, 2013). Polishing was performed with Quiver that is part of the SMRTanalysis suite.

The following commands were used for the assembly:

> source Local\_SMRTanalysis/current/etc/setup.sh

> fofnTOSmrtpipeInput.py HGAP.input.fofn > HGAP.input.xml

> smrtpipe.py –D NPROC=2 –D MAX\_THREADS=2 –output=Result\_SMRT –params=HGAP.input.xml xml:HGAP.input.xml

|  |  |
| --- | --- |
| **Assembly statistics**  *Z. ardabiliae* (Za17) | |
|  |  |
| contigs (>= 0 bp) | 59 |
| contigs (>= 1000 bp) | 59 |
| total length (bp) | 38,963,581 |
| largest contig (bp) | 4,826,179 |
| GC (%) | 51.25 |
| N50 | 1,413,809 |
| N75 | 689,698 |
| L50 | 9 |
| L75 | 18 |

**References supplementary text**

**Allen GC, Flores-Vergara M a, Krasynanski S, Kumar S, Thompson WF**. **2006**. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature protocols* **1**: 2320–2325.

**Bolger AM, Lohse M, Usadel B**. **2014**. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

**Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, *et al.*** **2013**. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**: 563–569.

**Langmead B, Salzberg SL**. **2012**. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

**Li H**. **2011**. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

**Plissonneau C, Stürchler A**. **2016**. The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. *M Bio* **7**: e01231-16.

**Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T**. **2016**. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**: W160–W165.