A chromosome-level genome assembly of *Zasmidium syzygii* isolated from banana leaves

A.C. van Westerhoven[1,2*], R. Mehrabi[1,3*], R. Talebi[3], M.B.F Steentjes[1], B. Corcolon[4], Pablo A. Chong[5], G.H.J. Kema[1,#,*], M. F. Seidl[2,#,*]

[1]Laboratory of Phytopathology, Wageningen University & Research, the Netherlands

[2]Theoretical Biology & Bioinformatics, Department of Biology, Utrecht University, the Netherlands

[3]Keygene N.V, P.O. Box 216, 6700 AE, Wageningen, Netherlands

[4]Research, Information, Compliance Department, Tadeco Inc., Panabo, Philippines

[5]ESPOL Polythecnic University, Escuela Superior Politécnica del Litoral, Centro de Investigaciones Biotecnológicas del Ecuador, Laboratorio de Biología Molecular, Guayaquil, Ecuador

*Authors contributed equally

[#]Corresponding authors: Gert H.J. Kema and Michael F. Seidl

Keywords: *Zasmidium syzygii,* genome assembly, banana leaf disease, foliar blight, plant disease

Abstract

Accurate taxonomic classification of samples from infected host material is essential for disease diagnostics and genome analyses. Despite the importance, diagnosis of fungal pathogens causing banana leaf diseases remains challenging. Foliar diseases of bananas are mainly caused by three *Pseudocercospora* species, of which the most predominant causal agent is *P. fijiensis*. Here, we sequenced and assembled four fungal isolates obtained from necrotic banana leaves in Bohol (Philippines) and obtained a high-quality genome assembly for one of these isolates. The samples were initially identified as *P. fijiensis* using PCR diagnostics, however, the assembly size was consistently 30 Mb smaller than expected. Based on the ITS sequences, we identified the samples as *Zasmidium syzygii* (98.7% identity). The high-quality *Zasmidium syzygii* assembly is 42.5 Mb in size, comprising 16 contigs, of which 11 are complete. The genome contains 98.6% of the expected single-copy BUSCO genes and contains 14,789 genes and 10.3% repeats. The three short-read assemblies are less continuous but have similar genome sizes (40.4 – 42.4 Mb) and contain between 96.5% and 98.4% BUSCO genes. All four isolates have identical ITS sequences and are distinct from *Zasmidium* isolates that were previously sampled from banana leaves. We thus report the first continuous genome assembly of a member of the *Zasmidium* genus, forming an essential resource for further analysis to enhance our understanding of the diversity of pathogenic fungal isolates as well as fungal diversity.

Introduction

Genome sequencing is important for the diagnostics and monitoring of diseases and is an important step to understand the biology of pathogens and diseases. To enable sample identification and downstream genome analyses, accurate taxonomic classification and prevention of contamination in genome assemblies are essential (Francois et al., 2020; Lu & Salzberg, 2018; Rachtman et al., 2020). However, publicly available genome assemblies are occasionally reported to contain contaminants, which can lead to incorrect species classification (Cornet & Baurain, 2022; Kusch et al., 2023; Steinegger & Salzberg, 2020). Obtaining a clean genome assembly is especially challenging for pathogens that live in close association with their host; samples from these pathogens are often contaminated with host material or other organisms that proliferate in proximity to the host such as endophytic fungi (Kusch et al., 2023; Zaccaron & Stergiopoulos, 2021).

Banana is an important food crop providing food security in tropical and subtropical regions worldwide. Foliar blights are a major constraint to banana production and are mainly caused by a complex of three *Pseudocercospora* species with *P. fijiensis* as a major constituent that causes black leaf streak disease or black Sigatoka. Control of this disease is responsible for up to 25% of the total costs of banana production (Drenth & Kema, 2021). The two other species also cause foliar blights, but are currently less prevalent, *P. musae* causes yellow Sigatoka and *P. eumusae* causes eumusae leaf spot (Chang et al., 2016). However, besides these three *Pseudocercospora* species, other fungal species can appear in association with symptomatic banana foliage. For example, a recent study that analysed fungal isolates associated with banana foliar diseases revealed the presence of over 30 other fungal species, primarily belonging to the *Mycosphaerellacea* family (Crous et al., 2021).

Interestingly, before this study revealed the identity of these isolates, most of these fungal species were considered to belong to the *Pseudocercospora* genus, highlighting that accurate diagnosis of the causal agents of leaf symptoms remains challenging. Yet, accurate identification and classification of the obtained isolates is important, particularly for developing disease diagnostics and enhancing the effectiveness of disease management strategies (Kusch et al., 2023; Lu & Salzberg, 2018).

Here we sampled four fungal isolates from banana foliage showing blight symptoms in Bohol, Philippines, and identified the fungal isolates by disease symptoms, morphology, and diagnostic PCR assays (Arzanlou et al., 2008). Interestingly, initial observations classified the isolates as *P. fijiensis*, but further genome analyses revealed that these isolates rather represent *Zasmidium syzygii*. Until now, only three fragmented genome assemblies of *Zasmidium* species are publicly available (Haridas et al., 2020; Xu et al., 2017). Here, we assembled the first chromosome-level genome assembly of a representative of the *Zasmidium* genus. The availability of this genome assembly will help to improve molecular diagnostics for pathogenic *Pseudocercospora* spp. on banana foliage. Moreover, the genome adds to the diversity of available fungal genomes, providing a resource for future genomic studies.

88 **Methods & Materials**

89 *Fungal isolation and sequencing*

90 Banana leaves with symptoms of foliar disease were collected from a field in Bohol,
91 Philippines. Leaf samples were cut into 1 cm^2 segments and used to discharge
92 ascospores as described previously (Chong et al., 2019). Four single ascospore
93 isolates (P121, P122, P123, and P124) were collected and grown on potato dextrose
94 agar (PDA) plates supplemented with streptomycin (100 µg/ml) for three weeks at 25
95 °C. To obtain sufficient fungal biomass for DNA isolation, a piece of mycelium (2 cm^2)
96 was blended for 20 sec. at 6,000 rpm in an Ultra Turrax Tube Drive homogenizer (IKA,
97 Staufen, Germany) in 15 ml water using a sterile DT-20 tube (IKA, Staufen, Germany).
98 The fragmented mycelium was transferred into a flask containing 100 ml of PDB
99 amended with streptomycin (100 µg/ml) and kept at 25 °C on a rotary shaker with 150
100 rpm for about two weeks. The fungal mycelium was filtered through miracloth and
101 subsequently washed with sterile water. Fungal mycelium was freeze-dried overnight
102 and used for high-molecular-weight (HMW) DNA isolation based on the CTAB method
103 (Murray & Thompson, 1980). After adding isopropanol, HMW DNA was collected from
104 the extraction buffer using a sterile needle. AMPure XP purification kit (Becman Cluter
105 Life Sciences, USA) was used to clean up the DNA. DNA quality and quantity were
106 checked by gel electrophoresis, Nanodrop micro-volume spectrophotometers, and
107 Qubit Fluorometric Quantitation (Thermo Fisher Scientific, USA). The HMW DNA of
108 isolate P124 was sequenced using PromethION Oxford Nanopore (ONT) sequencing
109 technologies. Additionally, all isolates were sequenced using the Illumina HiSeq
110 platform, both sequencing platforms were located at Keygene B.V. (Wageningen, the
111 Netherlands).

112

113 *Fungal species diagnosis by PCR*

114 To identify the isolated fungal species, the DNA from each isolate was extracted and
115 subjected to *P. fijiensis*-specific diagnostic PCR according to a previously established
116 protocol (Arzanlou et al., 2008). The list of primers used in PCRs is shown in Table
117 S1. Actin (820 bp) was used as a positive control to ensure successful amplification
118 and to assess the quality of DNA.

119

120 *Pathogenicity assay*

121 The pathogenicity of isolate P124 was tested on two banana genotypes; Cavendish
122 cv. Grand Naine (AAA) and a diploid (AA) genotype 'Pisang Berlin'. The inoculum for
123 isolate P124 as well as for the *P. fijiensis* reference isolate P78 were prepared
124 similarly. A piece of 1 cm^2 mycelium from three-week-old colonies grown on PDA was
125 collected in an Eppendorf tube containing 3-4 metal beads (3 mm diameter) and was
126 blended for 20 seconds at 3,000 rpm in a Bead Beater Homogenizer. About one ml of
127 sterile water was added to each tube and fragmented mycelium was spread on PDA
128 plates amended with 100 µg/ml streptomycin. The plates were kept at 25°C for 3-4
129 weeks and then a piece of mycelium (10 cm^2) was blended for 40 sec. at 6,000 rpm in
130 an Ultra Turrax Tube Drive homogenizer (IKA, Staufen, Germany) in 15 ml of distilled
131 water using a sterile DT-20 tube (IKA, Staufen, Germany). The suspension was
132 passed through miracloth to remove non-fragmented mycelium. The collected mycelial
133 fragments were further diluted and adjusted to $5x10^5$ fragments ml^{-1} and
134 supplemented with 0.15% Tween 20. This suspension was used to inoculate two-
135 month-old banana plants on both sides of the leaves. Each treatment was repeated
136 tree times, as a control, water was used for mock inoculation. Inoculated plants were

137 kept for 48 h at 90% relative humidity at 25ºC in the dark in a growth cabinet and
138 subsequently for eight weeks in a greenhouse with >85% RH, and with a day length
139 of 12 hours.
140
141 De novo *genome assembly*
142 To construct a complete and continuous genome assembly, we assembled the ONT
143 long-read sequences of isolate P124. The raw Fast5 files were basecalled with Guppy
144 (v. 5.1.13) and filtered for a minimum quality score of 10. The filtered reads were
145 assembled *de novo* using Canu (v. 2.2, Koren et al., 2017) with the parameters '-
146 OvlMerThreshold=300' and '-corMaxEvidenceErate=0.15' to improve the assembly of
147 repetitive regions. The initial genome assembly was polished with Racon (v. 1.5.0
148 Vaser et al., 2017) using the raw ONT long reads. After polishing, missing telomeres
149 were reconstructed, when possible, using teloclip (v. 0.0.3,
150 https://github.com/Adamtaranto/teloclip) and a final polishing step was performed
151 based on shrot-read Illumina data using Pilon (v. 1.24, Walker et al., 2014). The
152 presence of telomeres (TTAGGG) and coverage of the chromosomes in the final
153 polished assembly were determined using tapestry (v. 1.0.0, Davey et al., 2020). To
154 estimate the completeness of the genome assembly, the presence of conserved
155 single-copy genes was determined using BUSCO with the Capnodiales database
156 (Simão et al., 2015). The assembly continuity and potential contaminations were
157 accessed using BlobTools (v. 4.1.5, Challis et al., 2020). To estimate the size of the
158 assembled genome, all Illumina sequences were analyzed using GenomeScope (v.
159 2.0, Vurture et al., 2017) with the 21-mer profile determined by JellyFish (v 2.3.0,
160 Marçais & Kingsford, 2011). Isolate P121 showed high k-mer coverage outliers,
161 therefore the maximal k-mer coverage was set to 350x, similar to the maximum k-mer

162 coverage observed in the other genome. Repeats were predicted using
163 RepeatModeler (v. 2.0.3, Flynn et al., 2020) and annotated using RepeatMasker (v.
164 4.1.2, Smit et al., 2015). Finally, protein-coding genes were annotated using
165 Funannotate (v. 1.8.9, Palmer & Stajich, 2019) with the BUSCO lineage
166 pezizomycotina utilizing predicted *Zasmidium cellare* proteins as evidence
167 (GCF_010093935.1, Haridas et al., 2020). Additionally, we assembled the Illumina
168 sequenced isolates P121-P123 with Spades (v. 3.13, Bankevich et al., 2012). To
169 assess the quality and completeness of these assemblies, we ran QUAST (v. 5.2,
170 Gurevich et al., 2013) and BUSCO (v. 5.3.2, Simão et al., 2015) with the Capnodiales
171 odb10 database.

172

173 *Identification of internal transcribed spaces sequence*

174 The internal transcribed spacer (ITS) sequences from the four assemblies (P121-
175 P124) were retrieved using ITSx (v. 1.1.3, Bengtsson-Palme et al., 2013) and blastn
176 was used to detect similar ITS sequences in the blast database (accessed 19th
177 November 2022). Additionally, a maximum-likelihood tree was constructed to compare
178 the four ITS sequences to 227 publicly available ITS sequences obtained from NCBI
179 (accessed 13th January 2023). Sequences were aligned using mafft (v.7. 453, Katoh
180 et al., 2002) and a maximum-likelihood phylogenetic tree was constructed using
181 RAxML, with 500 bootstrap replicates (-m GTRCAT -p 1234 -b 100 -N 500) (v. 8.2.12,
182 Stamatakis, 2014). The resulting phylogeny was visualized using iTOL (Letunic &
183 Bork, 2021).

184

185

186 *Genetic diversity*

187 To assess the genetic diversity among isolates, short-read sequencing data of P121-
188 P123 were aligned to the P124 long-read genome assembly using BWA-mem (v.
189 0.7.17, Li, 2013). GATK4 was used to call variants (Van der Auwera et al., 2013), and
190 these variants were subsequently filtered using GATK Variant Filtration based on the
191 GATK best practices (Van der Auwera et al., 2013). The filtering process involved
192 excluding variants from reads with low mapping quality, variants predominantly located
193 at the edge of reads, and variants exhibiting a bias towards reverse/forward strands.

Results and Discussion

Fungal isolates obtained from symptomatic banana foliage in Bohol

Foliar blights of banana is mainly caused by a complex of *Pseudocercospora* species (Chang et al., 2016; Drenth & Kema, 2021). However, additional fungal species predominantly belonging to the *Mycosphaerella* genus have also been associated with symptomatic banana leaves (Crous et al., 2021). To further analyze the fungal pathogens that cause disease on banana leaves, four samples were obtained from banana plants with necrotic lesions in Bohol, Philippines (Chong et al., 2021). For all four isolates (P121 – P124) we were able to amplify a PCR product using *P. fijiensis*-specific PCR primers (Arzanlou et al., 2008), indicating that the isolates can be identified as *P. fijiensis* (**Figure 1a**). To corroborate the identity of the isolates, we tested the pathogenicity of one representative (isolate P124) on banana cultivars Cavendish (cv. Grand Naine AAA) and Pisang Berlin (AA) and compared the pathogenicity to the reference *P. fijiensis* isolate P78 originating from Tanzania. No leave spot symptoms were observed on Cavendish upon inoculation with P124, in contrast to the necrotic lesions caused by P78 (**Figure 1b**). However, both isolates caused necrotic lesions on Pisang Berlin four weeks after inoculation (**Figure 1c**). The necrotic lesions caused by isolate P124 were less severe than the disease symptoms caused by *P. fijiensis* isolate P78. Although we observe a difference in pathogenicity between the two isolates differ in pathogenicity, both isolates but cause necrosis on banana foliage.

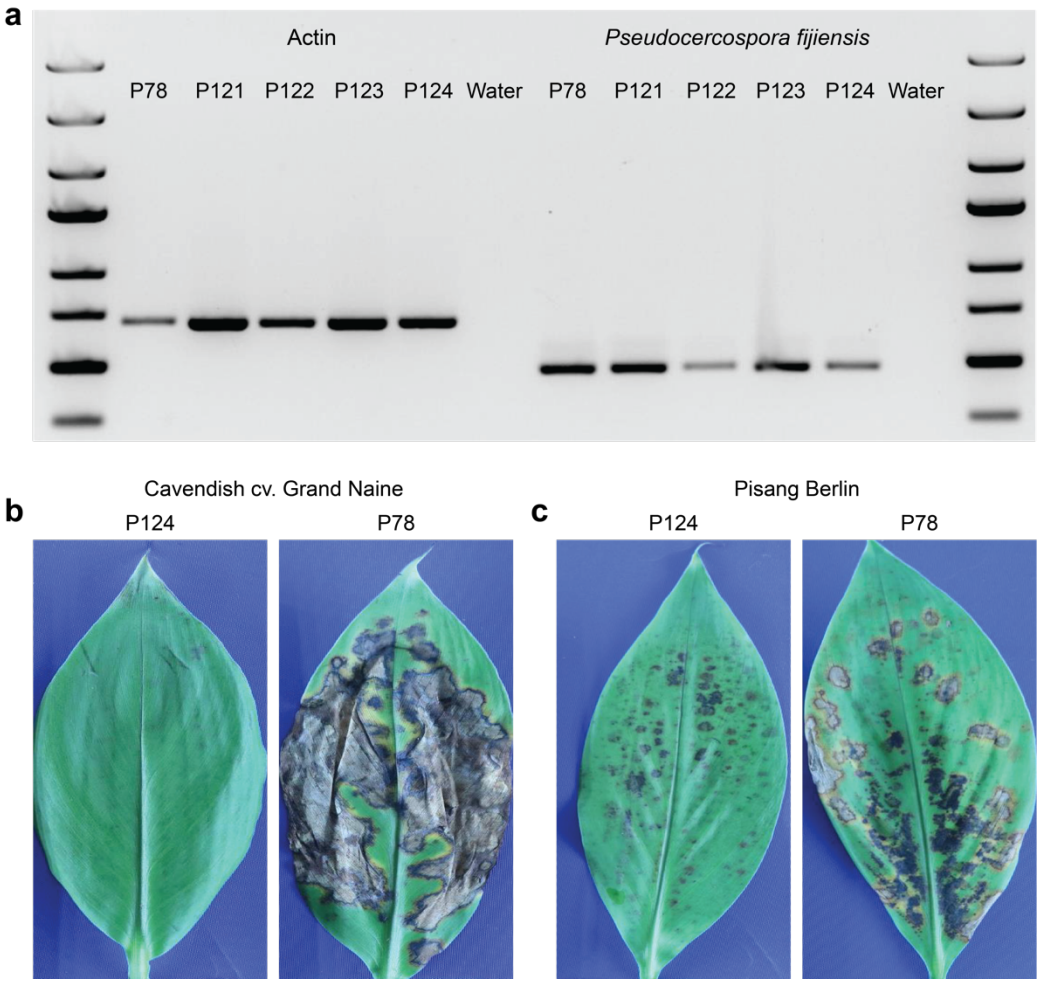


Figure 1. PCR diagnostics and pathogenicity assay of isolate P124 in comparison to the *Pseudocercospora fijiensis* reference strain P78 a) All isolates show amplification with actin primers (left panel). *Pseudocercospora*-specific primers amplified PCR products from *P. fijiensis* strain P78 (positive control) as well as from isolates P121, P122, P123, and P124 (right panel), suggesting that the isolates can be identified as *P. fijiensis*. b) Isolate P124 does not cause necrotic lesions on Cavendish banana, in contrast to necrotic symptoms caused by P78. c) Both P124 and P78 cause necrotic leaf symptoms on Pisang Berlin. Disease symptoms were scored eight weeks after inoculation.

Chromosome-level genome assembly of P124

To generate a high-quality genome assembly for the isolates from Bohol, we randomly selected strain P124 for sequencing using Oxford Nanopore Technology. This yielded 11.9 Gb of reads with an average size of ~9.1 kb and a read N50 of 12 kb, corresponding to a ~150x genome coverage based on the estimated genome size of 74 Mb (Arango et al., 2016). The *de novo* genome assembly resulted in 16 nuclear

232 contigs and one contig representing the mitochondrial genome. The assembly has an
233 N50 of 3.4 Mb and a total genome size of 42.5 Mb. Eleven of the 16 contigs had
234 telomeric sequences (TTAGGG) at both ends, and thus the assembly is highly
235 contiguous and mostly represents complete chromosomes (Figure 2). To assess
236 genome completeness, we queried the genome assembly for the presence of single-
237 copy BUSCO genes and identified 98.6% of the single-copy BUSCO genes that are
238 expected in the fungal order Capnodiales, indicating that the genome assembly of
239 P124 covers the conserved gene space (Figure 2a). We predicted a total of 14,789
240 protein-coding genes in the genome and identified that 10.3% of the genome consists
241 of repetitive elements (Figure 2b). *De novo* genome assembly of isolates P121 – P123,
242 sequenced with short-read sequencing technology only, resulted in genome
243 assemblies with similar sizes (40.4 – 42.5 Mb). Although the assemblies of P121 -
244 P123 are more fragmented compared to the assembly of the nanopore-sequenced
245 isolate P124, they approximately contain an equally high number of expected BUSCO
246 genes (Table 1).
247
248 **Table 1. The genome assembly statistics of four sequenced isolates obtained**
249 **from symptomatic banana foliage in Bohol, Philippines.**

	P124*	**P121****	**P122****	**P123****
Genome size (Mb) based on k-mer estimation (k=21)	40.6	39.8	39.9	41.9
Genome assembly size (Mb)	42.5	40.4	40.6	42.2
Number of contigs	16	1,225	1,563	796
Largest contig (Mb)	6.97	0.44	0.31	0.84
GC%	51.70	52.55	52.37	52.01

N50 (Mb)	3.40	0.98	0.60	1.9
N75 (Mb)	2.08	0.49	0.30	1.05
L50	2	134	201	67
BUSCO completeness (*n* = 3,578)	98.6%	97.8%	96.5%	98.4%
BUSCO fragmented genes	0.5%	0.9%	1.5%	0.5%
BUSCO missing genes	1.0%	1.3%	2.0%	1.1%

*P124 is assembled using long reads from Oxford Nanopore Technology (ONT). **P121-P123 are assembled using Illumina short reads.

The *P. fijiensis* reference genome is 73.6 Mb (Arango et al., 2016), and remarkably, the genome assemblies of the here sequenced fungal isolates are approximately 30 Mb smaller than expected. To verify the assembly size, we estimated genome sizes using k-mer profiles, which resulted in an estimated genome size ranging from 38.9 to 41.9 Mb, similar to the size of the *de novo* assembled genomes (Table 1). Furthermore, we mapped the sequencing reads of isolates P121 – P124 to the chromosome-level genome assembly of P124, mapping on average 95% of the reads with limited genetic diversity between isolates (on average 15 SNPs per kb; Figure 2b**).** An increased read mapping coverage is observed in most telomeric regions (Figure 2c), while the coverage drops in contig 1, contig 6, and contig 16, suggesting that these regions possibly contain assembly artefacts. Contig 16 is only 0.2 Mb in size and only contains a telomeric repeats at one end. The low coverage region together with read mappings that support a possible link to either contig 3 or 14, which also lack one of the telomeric repeats, suggests that this contig might be associated to one of these two contigs. Apart from these regions, mapping of the short reads of P124 to the P124 chromosome-level genome assembly revealed an overall constant

269 read mapping coverage over the chromosomes, indicating that the P124 genome
270 assembly does not contain extensively collapsed repetitive regions that could account
271 for the smaller genome size compared to *P. fijiensis*. This validates that, although the
272 genome assembly of strain P124 is smaller than expected, it is highly complete and
273 continuous, suggesting that the length variation is likely not due to assembly errors.
274 To capture possible contaminations in the genome assemblies, we queried all contigs
275 in the BLAST database to determine their identity. All contigs showed a high similarity
276 to members of the *Mycosphaerellacea* family; most contigs were similar to *Zasmidium*
277 species (13 contigs, 35 Mb) and only one contig (3.3 Mb) showed similarity to
278 *Pseudocercospora* species (**Figure S1**). Whole-genome alignments of our assembly
279 to the *P. fijiensis* reference genome assembly revealed that none of the contigs display
280 significant similarity (**Figure S2**). Therefore, we considered that the difference in
281 genome size between isolates P121 – P124 and the *P. fijiensis* reference isolate P78
282 is likely caused by the isolation of a different fungal species associated with necrotic
283 symptoms on banana foliage (Crous et al., 2021), most likely by a member of genus
284 *Zasmidium*.
285

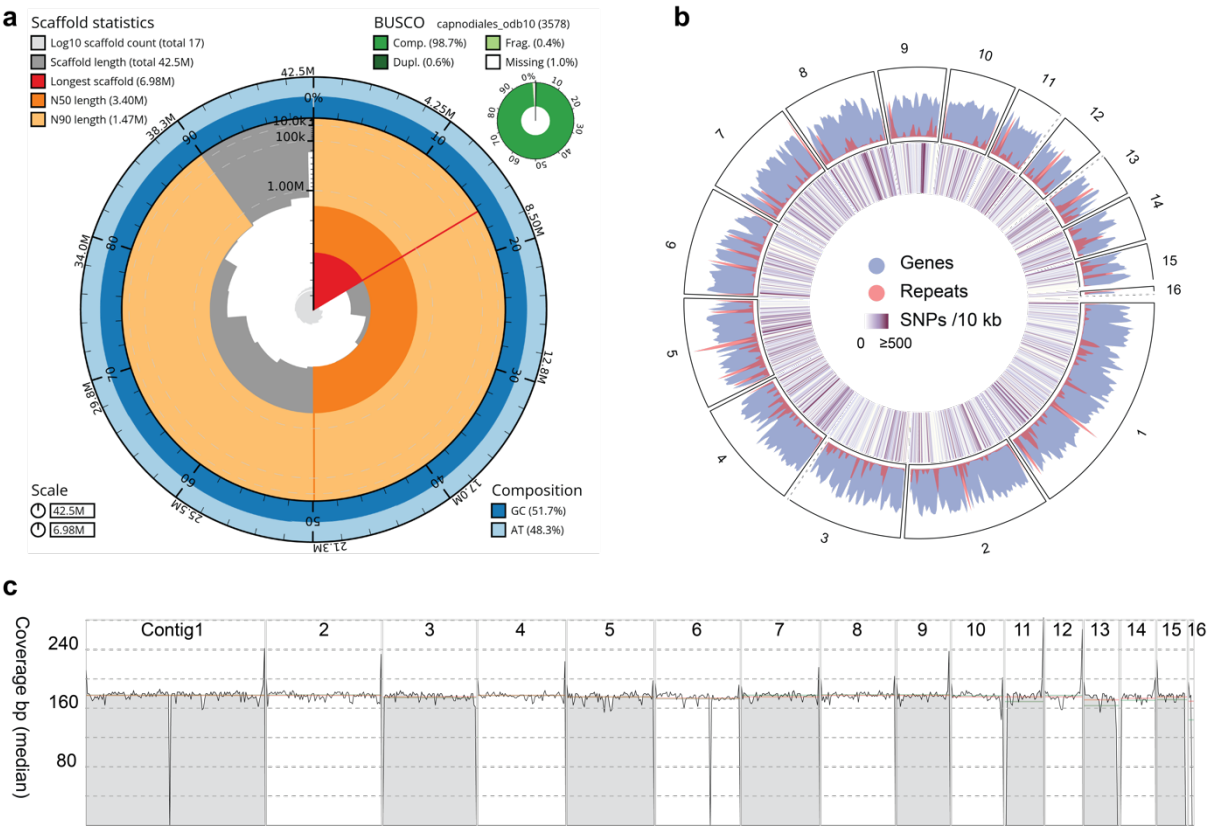


Figure 2. Chromosome-level genome assembly of *Zasmidium syzygii* isolate P124. a) Genome assembly statistics for the *de novo* assembly of *Z. syzygii* isolate P124 based on Oxford Nanopore Technology (ONT). The genome assembly has 17 contigs (16 nuclear contigs of which at least 11 are complete chromosomes, and the mitochondrial genome) with a total genome size of 42.5 Mb and contains 98.7% complete single-copy BUSCO genes. **b**) A circular representation of the contigs in P124. Dashed lines indicate missing telomeres for chromosome 3, 11, 13, 15, and 16. Genes (14,786) and repeats (10.3%) are distributed evenly over the chromosomes. Single nucleotide polymorphisms are found over all contigs, with two SNP dense regions on contig 8 and 9. **c**) Short-read coverage of P124 mapped to the assembled contigs of P124 shows minimal regions with exceptionally high or low coverage, suggesting that the assembly does not contain large repetitive regions that may have been collapsed during the assembly process. Mean coverage (green line) and Median coverage (red line) are indicated in the figure.

302 **The assembled genome sequence reveals that isolate P124 belongs to the**
303 **fungal genus *Zasmidium***
304 To determine the identity of isolate P124, we retrieved the ITS sequence from the
305 genome assembly and searched for related species in the non-redundant BLAST
306 database on NCBI using the P124 ITS sequence as a query. We obtained a highly
307 similar match (98.7% nucleotide identity) to *Zasmidium syzygii* (NR_111826.1),
308 supporting that the assembled isolate belongs to the *Zasmidium* genus. In line with
309 this finding, previous studies have found *Zasmidium* isolates on symptomatic banana
310 leaves (Arzanlou et al., 2008; Crous et al., 2021), and other *Zasmidium* species are
311 reported to cause leaf spot diseases on other plant species such as citrus (Aguilera-
312 Cogley et al., 2017; An et al., 2021; Han et al., 2015). The *Zasmidium* genome
313 assembly we report here is less fragmented and 4 Mb larger than the previous genome
314 assembly of a *Zasmidium* species; *Z. cellare*. The species, known as the 'wine cellar
315 fungus' because it thrives in walls and ceilings of wine cellars (Tribe et al., 2006), was
316 sequenced and its genome was assembled into >267 scaffolds with a genome size of
317 ~38 Mb (Haridas et al., 2020), illustrating that our genome offers a more continuous
318 genome representation of a member of the *Zasmidium* genus.
319 To validate the identity and to determine the diversity of the *Zasmidium* isolates,
320 we compared the ITS sequences of isolates P121-P124 with 43 other *Zasmidium* ITS
321 sequences from NCBI (20 December 2022) as well as ITS sequences of 120
322 *Mycosphaerella* strains obtained from banana leaves (Crous et al., 2021). A
323 maximum-likelihood phylogeny based on the aligned ITS sequences confirmed that
324 isolates P121 – P124 belong to the *Zasmidium* genus and shows that these four
325 isolates encode identical ITS sequences (Figure 3). Notably, the set of ITS sequences
326 also contains sequences from five other *Zasmidium* isolates that had been previously

327 sampled from banana leaves from different geographic locations (Martinique, Tonga,
328 and Gabon) (Crous et al., 2021). Interestingly, these do not cluster with isolates P121-
329 P124, suggesting that pathogenicity towards banana is a polyphyletic trait within the
330 *Zasmidium* genus (Figure 3). Although *Zasmidium* species have been linked to foliar
331 blights in various hosts (Aguilera-Cogley et al., 2017; Han et al., 2015; Osorio et al.,
332 2021), the pathogenicity and global spread of *Zasmidium* species has not been studied
333 in depth. Based on our data, we conclude that Z. *syzygii* occurs on banana leaves with
334 necrotic symptoms and can be the cause of mild necrotic lesions on the foliage.
335 However, the abundance and role of Z. *syzygii* as a banana pathogen remains
336 unknown, which requires further research to understand its prevalence, significance,
337 and potential impact on banana cultivation.

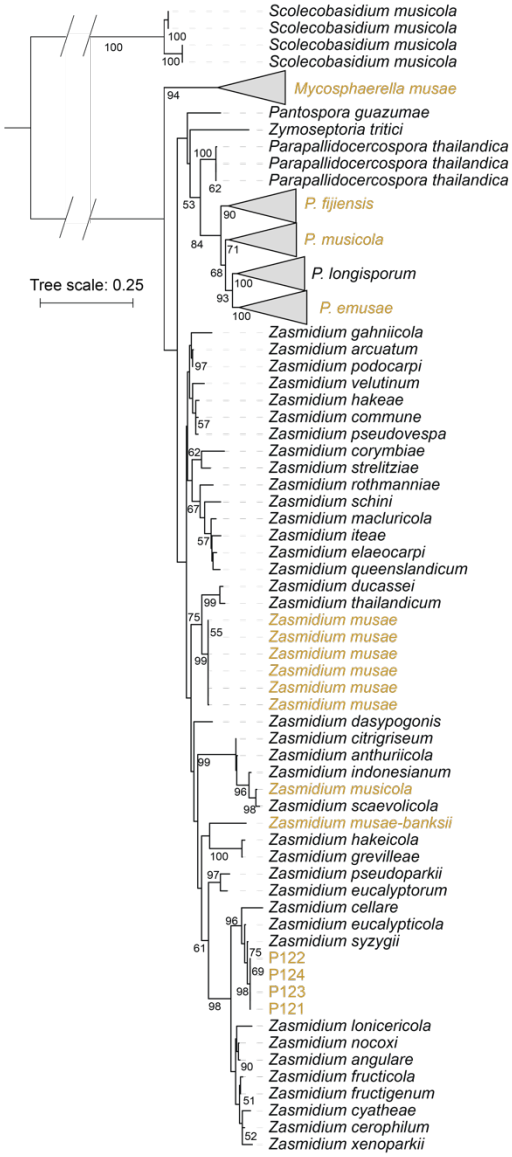


Figure 3. The four isolates sampled from Bohol (P121-P124) belong to the genus *Zasmidium*. A maximum-likelihood phylogenetic tree was constructed using the internally transcribed spacer (ITS) sequences of 43 *Zasmidium* species from NCBI and 120 *Mycosphaerellacea* isolates from banana leaves (Crous et al., 2021). The ITS sequence extracted from the chromosome-level genome assembly (P124) as well as from the other isolates (P121-123) associated with an isolate classified as *Z. syzgii* and are not related to *Zasmidium* strains previously isolated from banana leaves. *Zasmidium* isolates associated with infected banana leaves (yellow labels) are genetically diverse and are distributed across various branches of the phylogenetic tree.

Accurate diagnostics of pathogens is essential to detect the emergence and trace the dispersal of diseases, which is pivotal for effective disease management. However, our data reveal that *P. fijiensis* and *Z. syzygii* are indistinguishable with the current PCR diagnostic (Arzanlou et al., 2008). To compare the similarity of the PCR

354 primers between *P. fijiensis* and *Z. syzygii*, we *in silico* detected the amplicon of the
355 supposedly *Pseudocercospora*-specific primers in the *Zasmidium syzygii* P124
356 genome assembly and in the *P. fijiensis* reference genome assembly (Cirad86; Arango
357 et al., 2016). Both isolates possess the primer sequence used to distinguish *P. fijiensis*
358 (Table S1) from *P. musae* and *P. eumusae* and produce a similar sized amplicon of
359 480 bp in *Z. syzygii* P124 and 478 bp in *P. fijiensis* Cirad86, which explains the positive
360 result for *Z. syzygii* in our PCR assay (Figure 1). The amplicons share 88% sequence
361 identity and *Zasmidium* or *Pseudocercospora* isolates can therefore be distinguished
362 only upon amplicon sequencing. Thus, novel primer pairs need to be developed to
363 enable easy and accurate diagnosis of fungal species present in foliar blights of
364 banana.
365
366 **Conclusion**
367 Here, we report the first chromosome-scale genome assembly of a *Zasmidium*
368 species, this adds a high-quality genome sequence to the thus far limited genetic
369 resources available for this genus. Our data show that *Z. syzygii* occurs on banana
370 foliage in Bohol and can cause leaf necrosis, comparable to the foliar blight symptoms
371 observed for *P. fijiensis*. The availability of the genome assembly will facilitate further
372 research into the association of *Zasmidium, Pseudocercospora*, and possibly other
373 fungal species related to foliar blights of banana. Moreover, it will serve as a valuable
374 resource for developing novel molecular diagnostics, enabling the accurate
375 identification and characterization of these fungal species.
376

377 *Data Availability Statement*

378 All sequencing data are deposited in the NCBI database under accession number
379 PRJNA931964. Scripts used for the genome assembly and analysis are available on
380 github https://github.com/Anouk-vw/Zasmidium.

381 *Acknowledgments*

382 We thank the Research Support Group (RSG) at KeyGene B.V. (Wageningen, the
383 Netherlands) for performing the DNA isolation, library preparation and genome
384 sequencing. We thank Xiaoqian Shi-Kunne (Wageningen University & Research, the
385 Netherlands) for the data management and bioinformatics support.

386

387 *Funder Information*

388 ACW and GHJK were supported by the Bill and Melinda Gates Foundation, grant
389 number AG – 4425. Banana research at Wageningen University has been supported
390 by the Dutch Dioraphte Foundation, grant number 20 04 04 02.

391

392 *Conflict of Interest*

393 The authors report no conflict of interests.

394

References

Aguilera-Cogley, V. A., Berbegal, M., Català, S., Brentu, F. C., Armengol, J., & Vicent, A. (2017). Characterization of *Mycosphaerellaceae* species associated with citrus greasy spot in Panama and Spain. *PLOS ONE*, *12*(12), e0189585. https://doi.org/10.1371/journal.pone.0189585

An, Y.-Y., Zeng, X.-Y., Geng, K., Hyde, K. D., & Wang, Y. (2021). One new species and one new record of Zasmidium in China. *Biodiversity Data Journal*, *9*, e59001. https://doi.org/10.3897/BDJ.9.e59001

Arango, R. E. isaza, Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C. F., Jong, T. V. de, Vries, I. de, Dietrich, R., Farmer, A. D., Fereira, C. F., Garcia, S., Guzman, M., Hamelin, R. C., Lindquist, E. A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., … Kema, G. H. J. (2016). Combating a global threat to a clonal crop: Banana Black Sigatoka pathogen *Pseudocercospora fijiensis* (synonym *Mycosphaerella fijiensis*) genomes reveal clues for disease control. *PLOS Genetics*, *12*(8), e1005876. https://doi.org/10.1371/journal.pgen.1005876

Arzanlou, M., Groenewald, J. Z., Fullerton, R. A., Abeln, E. C. A., Carlier, J., Zapater, M.-F., Buddenhagen, I. W., Viljoen, A., & Crous, P. W. (2008). Multiple gene genealogies and phenotypic characters differentiate several novel species of *Mycosphaerella* and related anamorphs on banana. *Persoonia : Molecular Phylogeny and Evolution of Fungi*, *20*, 19–37. https://doi.org/10.3767/003158508X302212

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A.

(2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, *4*(10), 914–919. https://doi.org/10.1111/2041-210X.12073

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – interactive quality assessment of genome assemblies. *G3 Genes|Genomes|Genetics*, *10*(4), 1361–1374. https://doi.org/10.1534/g3.119.400908

Chang, T.-C., Salvucci, A., Crous, P. W., & Stergiopoulos, I. (2016). Comparative genomics of the Sigatoka disease complex on banana suggests a link between parallel evolutionary changes in *Pseudocercospora fijiensis* and *Pseudocercospora eumusae* and increased virulence on the banana host. *PLOS Genetics*, *12*(8), e1005904. https://doi.org/10.1371/journal.pgen.1005904

Chong, P., Essoh, J. N., Arango Isaza, R. E., Keizer, P., Stergiopoulos, I., Seidl, M. F., Guzman, M., Sandoval, J., Verweij, P. E., Scalliet, G., Sierotzski, H., de Lapeyre de Bellaire, L., Crous, P. W., Carlier, J., Cros, S., Meijer, H. J. G., Peralta, E. L., & Kema, G. H. J. (2021). A world-wide analysis of reduced

445 sensitivity to DMI fungicides in the banana pathogen *Pseudocercospora*
446 *fijiensis*. *Pest Management Science*, *77*(7), 3273–3288.
447 https://doi.org/10.1002/ps.6372
448 Chong, P., Vichou, A.-E., Schouten, H. J., Meijer, H. J. G., Arango Isaza, R. E., &
449 Kema, G. H. J. (2019). Pfcyp51 exclusively determines reduced sensitivity to
450 14α-demethylase inhibitor fungicides in the banana black Sigatoka pathogen
451 *Pseudocercospora fijiensis*. *PLOS ONE*, *14*(10), e0223858.
452 https://doi.org/10.1371/journal.pone.0223858
453 Cornet, L., & Baurain, D. (2022). Contamination detection in genomic data: More is
454 not enough. *Genome Biology*, *23*(1), 60. https://doi.org/10.1186/s13059-022-
455 02619-9
456 Crous, P. W., Carlier, J., Roussel, V., & Groenewald, J. Z. (2021).
457 Pseudocercospora and allied genera associated with leaf spots of banana
458 (Musa spp.). *Fungal Systematics and Evolution*, *7*, 1–19.
459 https://doi.org/10.3114/fuse.2021.07.01
460 Davey, J. W., Davis, S. J., Mottram, J. C., & Ashton, P. D. (2020). *Tapestry: Validate*
461 *and edit small eukaryotic genome assemblies with long reads* (p.
462 2020.04.24.059402). bioRxiv. https://doi.org/10.1101/2020.04.24.059402
463 Drenth, A., & Kema, G. (2021). The vulnerability of bananas to globally emerging
464 disease threats. *Phytopathology®*, *111*(12), 2146–2161.
465 https://doi.org/10.1094/phyto-07-20-0311-rvw
466 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit,
467 A. F. (2020). RepeatModeler2 for automated genomic discovery of
468 transposable element families. *Proceedings of the National Academy of*
469 *Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

470 Francois, C. M., Durand, F., Figuet, E., & Galtier, N. (2020). Prevalence and
471 implications of contamination in public genomic resources: A case study of 43
472 reference arthropod assemblies. *G3 Genes|Genomes|Genetics*, *10*(2), 721–
473 730. https://doi.org/10.1534/g3.119.400758
474 Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality
475 assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*,
476 *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086
477 Han, K. S., Park, M. J., Choi, I. Y., Park, J. H., & Shin, H. D. (2015). First report of
478 leaf spot caused by *zasmidium lythri* on *lythrum salicaria* in korea. *Plant*
479 *Disease*, *99*(5), 728–728. https://doi.org/10.1094/PDIS-10-14-1039-PDN
480 Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A.,
481 Andreopoulos, B., Baker, S. E., Barry, K., Bills, G., Bluhm, B. H., Cannon, C.,
482 Castanera, R., Culley, D. E., Daum, C., Ezra, D., González, J. B., Henrissat,
483 B., Kuo, A., … Grigoriev, I. V. (2020). 101 Dothideomycetes genomes: A test
484 case for predicting lifestyles and emergence of pathogens. *Studies in*
485 *Mycology*, *96*, 141–153. https://doi.org/10.1016/j.simyco.2020.01.003
486 Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). Mafft: A novel method for rapid
487 multiple sequence alignment based on Fast Fourier transform. *Nucleic Acids*
488 *Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436
489 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M.
490 (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer
491 weighting and repeat separation. *Genome Research*, *27*(5), 722–736.
492 https://doi.org/10.1101/gr.215087.116

493 Kusch, S., Vaghefi, N., & Kiss, L. (2023). The good, the bad, and the misleading:
494 How to improve the quality of ‘genome announcements’? *Molecular Plant-*
495 *Microbe Interactions®*. https://doi.org/10.1094/MPMI-01-23-0009-LE
496 Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for
497 phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1),
498 W293–W296. https://doi.org/10.1093/nar/gkab301
499 Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with*
500 *BWA-mem*. https://doi.org/10.6084/m9.figshare.963153.v1
501 Lu, J., & Salzberg, S. L. (2018). Removing contaminants from databases of draft
502 genomes. *PLOS Computational Biology*, *14*(6), e1006277.
503 https://doi.org/10.1371/journal.pcbi.1006277
504 Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
505 counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770.
506 https://doi.org/10.1093/bioinformatics/btr011
507 Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight
508 plant DNA. *Nucleic Acids Research*, *8*(19), 4321–4325.
509 https://doi.org/10.1093/nar/8.19.4321
510 Osorio, J. A., Linnakoski, R., Wingfield, M. J., de Beer, Z. W., & Roux, J. (2021). A
511 new species in the *Mycosphaerellaceae from Cecidomyiidae leaf galls* on
512 *Avicennia marina* in South Africa. *Antonie van Leeuwenhoek*, *114*(5), 515–
513 526. https://doi.org/10.1007/s10482-021-01537-3
514 Palmer, J., & Stajich, J. (2019). *nextgenusfs/funannotate: Funannotate v1.5.3*
515 [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2604804
516 Rachtman, E., Balaban, M., Bafna, V., & Mirarab, S. (2020). The impact of
517 contaminants on the accuracy of genome skimming and the effectiveness of

exclusion read filters. *Molecular Ecology Resources*, *20*(3), 649–661. https://doi.org/10.1111/1755-0998.13135

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0*. http://www.repeatmasker.org

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

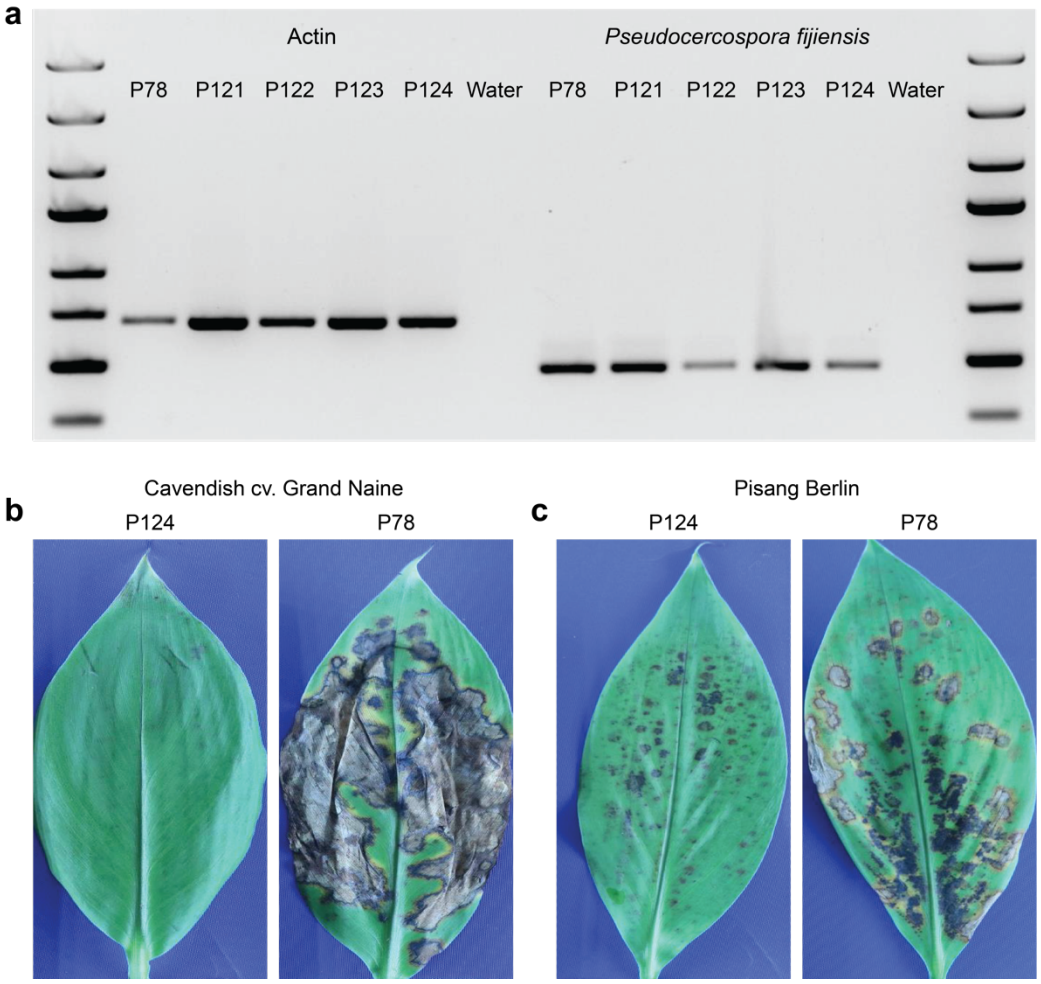
Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, *21*(1), 115. https://doi.org/10.1186/s13059-020-02023-1

Tribe, H. T., Thines, E., & Weber, R. W. S. (2006). Moulds that should be better known: The wine cellar mould, *Racodium cellare*. *Mycologist*, *20*(4), 171–175. https://doi.org/10.1016/j.mycol.2006.09.016

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

541 Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo
542 genome assembly from long uncorrected reads. *Genome Research*, *27*(5),
543 737–746. https://doi.org/10.1101/gr.214270.116
544 Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H.,
545 Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free
546 genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204.
547 https://doi.org/10.1093/bioinformatics/btx153
548 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo,
549 C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An
550 integrated tool for comprehensive microbial variant detection and genome
551 assembly improvement. *PLOS ONE*, *9*(11), e112963.
552 https://doi.org/10.1371/journal.pone.0112963
553 Xu, C., Zhang, R., Sun, G., & Gleason, M. L. (2017). Comparative genome analysis
554 reveals adaptation to the ectophytic lifestyle of sooty blotch and flyspeck
555 fungi. *Genome Biology and Evolution*, *9*(11), 3137–3151.
556 https://doi.org/10.1093/gbe/evx229
557 Zaccaron, A. Z., & Stergiopoulos, I. (2021). Characterization of the mitochondrial
558 genomes of three powdery mildew pathogens reveals remarkable variation in
559 size and nucleotide composition. *Microbial Genomics*, *7*(12), 000720.
560 https://doi.org/10.1099/mgen.0.000720
561
562

563 **Figures and tables**



564

565 **Figure 1. PCR diagnostics and pathogenicity assay of isolate P124 in**
566 **comparison to the *Pseudocercospora fijiensis* reference strain P78** a) All isolates
567 show amplification with actin primers (left panel). *Pseudocercospora*-specific primers
568 amplified PCR products from *P. fijiensis* strain P78 (positive control) as well as from
569 isolates P121, P122, P123, and P124 (right panel), suggesting that the isolates can
570 be identified as *P. fijiensis*. b) Isolate P124 does not cause necrotic lesions on
571 Cavendish banana, in contrast to necrotic symptoms caused by P78. c) Both P124
572 and P78 cause necrotic leaf symptoms on Pisang Berlin. Disease symptoms were
573 scored eight weeks after inoculation.
574

575 **Table 1. The genome assembly statistics of four sequenced isolates obtained**

576 **from symptomatic banana foliage in Bohol, Philippines.**

	P124*	**P121****	**P122****	**P123****
Genome size (Mb) based on k-mer estimation (k=21)	40.6	39.8	39.9	41.9
Genome assembly size (Mb)	42.5	40.4	40.6	42.2
Number of contigs	16	1,225	1,563	796
Largest contig (Mb)	6.97	0.44	0.31	0.84
GC%	51.70	52.55	52.37	52.01
N50 (Mb)	3.40	0.98	0.60	1.9
N75 (Mb)	2.08	0.49	0.30	1.05
L50	2	134	201	67
BUSCO completeness (*n* = 3,578)	98.6%	97.8%	96.5%	98.4%
BUSCO fragmented genes	0.5%	0.9%	1.5%	0.5%
BUSCO missing genes	1.0%	1.3%	2.0%	1.1%

577 *P124 is assembled using long reads from Oxford Nanopore Technology (ONT).

578 **P121-P123 are assembled using Illumina short reads.

579

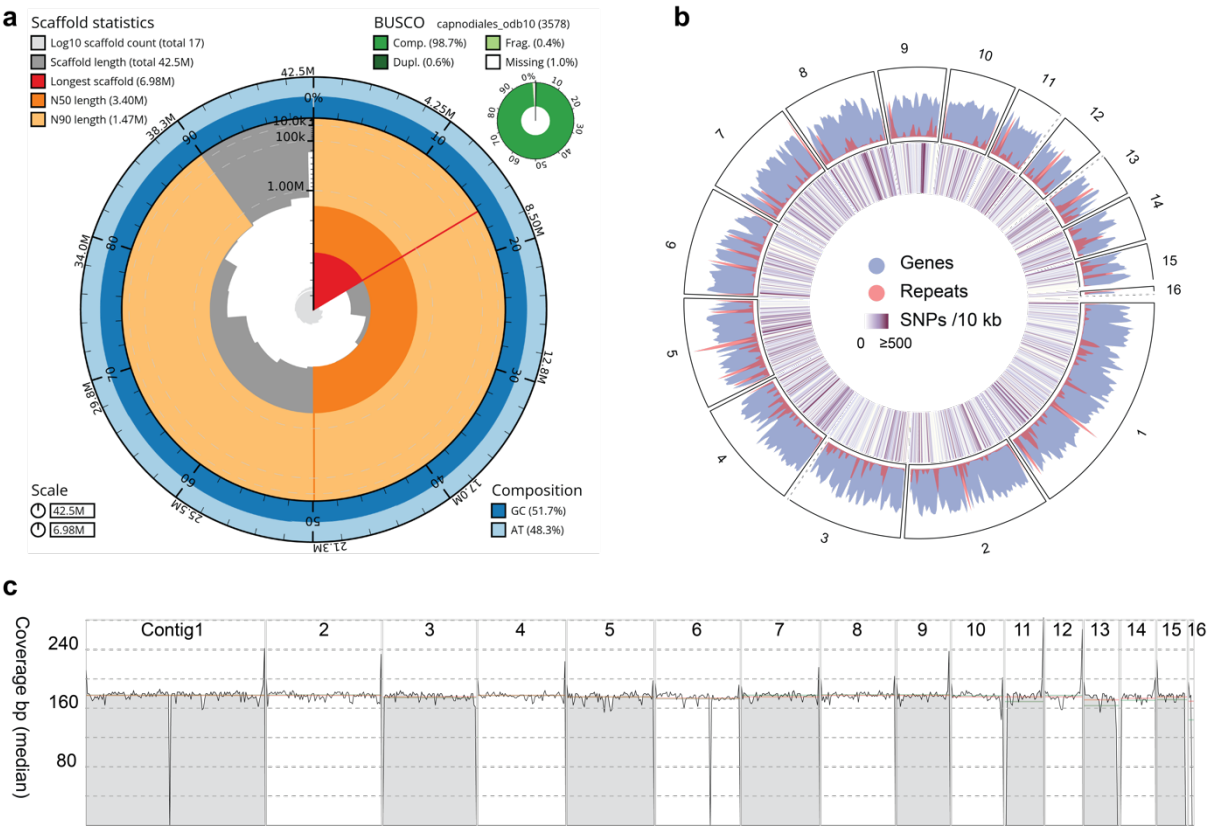


Figure 2. Chromosome-level genome assembly of *Zasmidium syzygii* isolate P124. a) Genome assembly statistics for the *de novo* assembly of *Z. syzygii* isolate P124 based on Oxford Nanopore Technology (ONT). The genome assembly has 17 contigs (16 nuclear contigs of which at least 11 are complete chromosomes, and the mitochondrial genome) with a total genome size of 42.5 Mb and contains 98.7% complete single-copy BUSCO genes. **b**) A circular representation of the contigs in P124. Dashed lines indicate missing telomeres for chromosome 3, 11, 13, 15, and 16. Genes (14,786) and repeats (10.3%) are distributed evenly over the chromosomes. Single nucleotide polymorphisms are found over all contigs, with two SNP dense regions on contig 8 and 9. **c**) Short-read coverage of P124 mapped to the assembled contigs of P124 shows minimal regions with exceptionally high or low coverage, suggesting that the assembly does not contain large repetitive regions that may have been collapsed during the assembly process. Mean coverage (green line) and Median coverage (red line) are indicated in the figure.

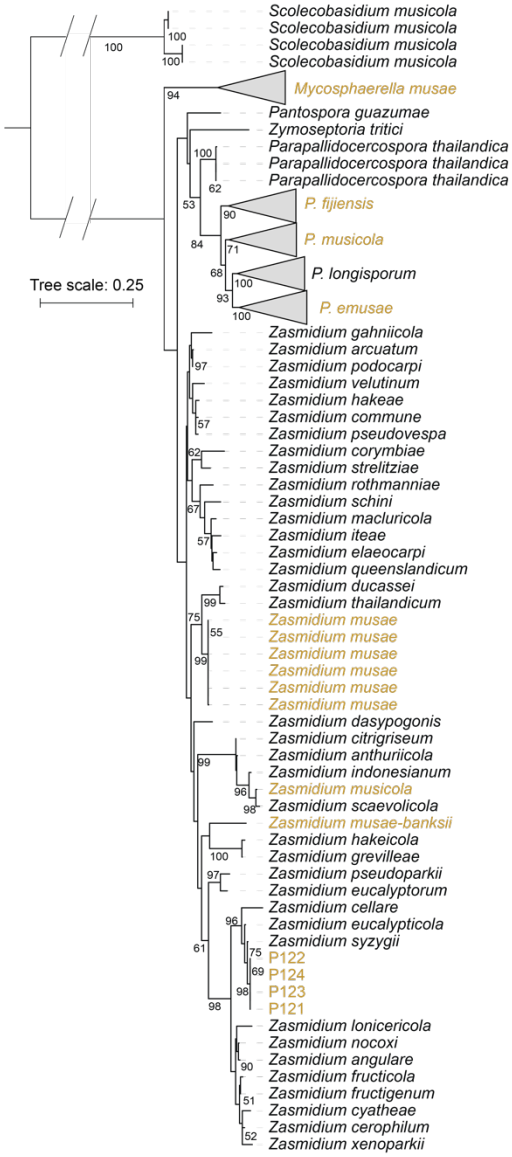


Figure 3. The four isolates sampled from Bohol (P121-P124) belong to the genus *Zasmidium.* Maximum-likelihood phylogenetic tree is constructed using the internally transcribed spacer (ITS) sequences of 43 *Zasmidium* species from NCBI and 120 *Mycosphaerella* isolates from banana leaves (Crous et al., 2021). The ITS sequence extracted from the chromosome-level genome assembly (P124) as well as from the other isolates (P121-123) associated with an isolate classified as *Z. syzgii* are not related to *Zasmidium* strains previously isolated from banana leaves. *Zasmidium* isolates associated with infected banana leaves (yellow labels) are genetically diverse and are distributed across various branches of the phylogenetic tree.