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 that cause 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 chromosome-level genome assembly for one of these isolates. Using PCR diagnostics, the samples were initially identified as *P. fijiensis*, yet the assembly size was consistently 30 Mb smaller than expected. Based on the ITS sequences, we identified the samples as *Zasmidium syzygii*. The *Zasmidium syzygii* genome assembly is 42.5 Mb in size distributed over 16 chromosomes, eleven of which are complete. The genome contains 98.6% of the expected single-copy BUSCO genes and we predicted 14,789 protein-coding genes and 10.3% repeats. The three other genome assemblies based on short-read data 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 have been previously sampled from banana leaves. We thus report the first chromosome-level 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 forms an important step to help 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 other organisms; samples from pathogens are often contaminated with host material or other organisms that proliferate in or in proximity to host plants such as for example bacteria or endophytic fungi (Kusch et al., 2023; Zaccaron & Stergiopoulos, 2021).

Banana is an important food crop that provides 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*, the causal agent of black leaf streak disease or Black Sigatoka, being the dominant species. 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 Mycosphereallaceae family (Crous et al., 2021). Interestingly, before Crous and colleagues revealed the identity of these isolates, most of these fungi were considered to belong to the *Pseudocercospora* genus, highlighting that accurate diagnosis of the causal agents of leaf symptoms remains challenging. However, accurate identification and classification of the obtained isolates is important, especially when developing disease diagnostics or 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 belong to *Zasmidium syzygii*. Until now, only three fragmented genome assemblies of *Zasmidium* species are publicly available (Ametrano et al., 2022; 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 the community with a useful resource for future genomic studies.

91 **Methods & Materials**

92 *Fungal isolation and sequencing*

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

115

116 *Fungal species diagnosis by PCR*

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

122

123 *Pathogenicity assay*

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

140 kept for 48 h at 90% relative humidity at 25ºC in the dark in a growth cabinet and
141 subsequently for eight weeks in a greenhouse with >85% RH, and with a day length
142 of 12 hours.
143
144 De novo *genome assembly*
145 To construct a complete and continuous genome assembly, we assembled the ONT
146 long-read sequences of isolate P124. The raw Fast5 files were basecalled with
147 Guppy (v. 5.1.13) and filtered for a minimum quality score of 10. The filtered reads
148 were assembled *de novo* using Canu (v. 2.2, Koren et al., 2017) with the parameters
149 '-OvlMerThreshold=300' and '-corMaxEvidenceErate=0.15' to improve the assembly
150 of repetitive regions. The initial genome assembly was polished with Racon (v. 1.5.0)
151 using the raw ONT long reads. After polishing, missing telomeres were
152 reconstructed, when possible, using teloclip (v. 0.0.3,
153 https://github.com/Adamtaranto/teloclip) and a final polishing step was performed
154 based on shot-read Illumina data using Pilon (v. 1.24, Walker et al., 2014). The
155 presence of telomeres (TTAGGG) and coverage of the chromosomes in the final
156 polished assembly were determined using tapestry (v. 1.0.0, Davey et al., 2020). To
157 estimate the completeness of the genome assembly, the presence of conserved
158 single-copy genes was determined using BUSCO with the Capnodiales database
159 (Simão et al., 2015). The assembly continuity and potential contaminations were
160 accessed using BlobTools (v. 4.1.5, Challis et al., 2020). To estimate the size of the
161 assembled genome, all Illumina sequences were analyzed using GenomeScope (v.
162 2.0, Vurture et al., 2017) with the 21-mer profile determined by JellyFish (v 2.3.0,
163 Marçais & Kingsford, 2011). Isolate P121 showed high k-mer coverage outliers,
164 therefore the maximal k-mer coverage was set to 350x, similar to the maximum k-
165 mer coverage observed in the other genome. Repeats were predicted using

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

175

176 *Identification of internal transcribed spaces sequence*

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

186

187 *Genetic diversity*

188 To assess the genetic diversity among isolates, short-read sequencing data of P121-
189 P123 were aligned to the P124 long-read genome assembly using BWA-mem (v.

190 0.7.17, Li, 2013). GATK4 was used to call variants (Van der Auwera et al., 2013),
191 and these variants were subsequently filtered using GATK Variant Filtration based
192 on the GATK best practices (Van der Auwera et al., 2013). The filtering process
193 involved excluding variants from reads with low mapping quality, variants
194 predominantly located at the edge of reads, and variants exhibiting a bias towards
195 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 Mycosphaerellaceae family 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 (here called 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 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, however, less severe than the disease symptoms caused by *P. fijiensis* isolate P78. Even though we observe a difference in pathogenicity between the two isolates, both can cause necrosis on banana foliage.

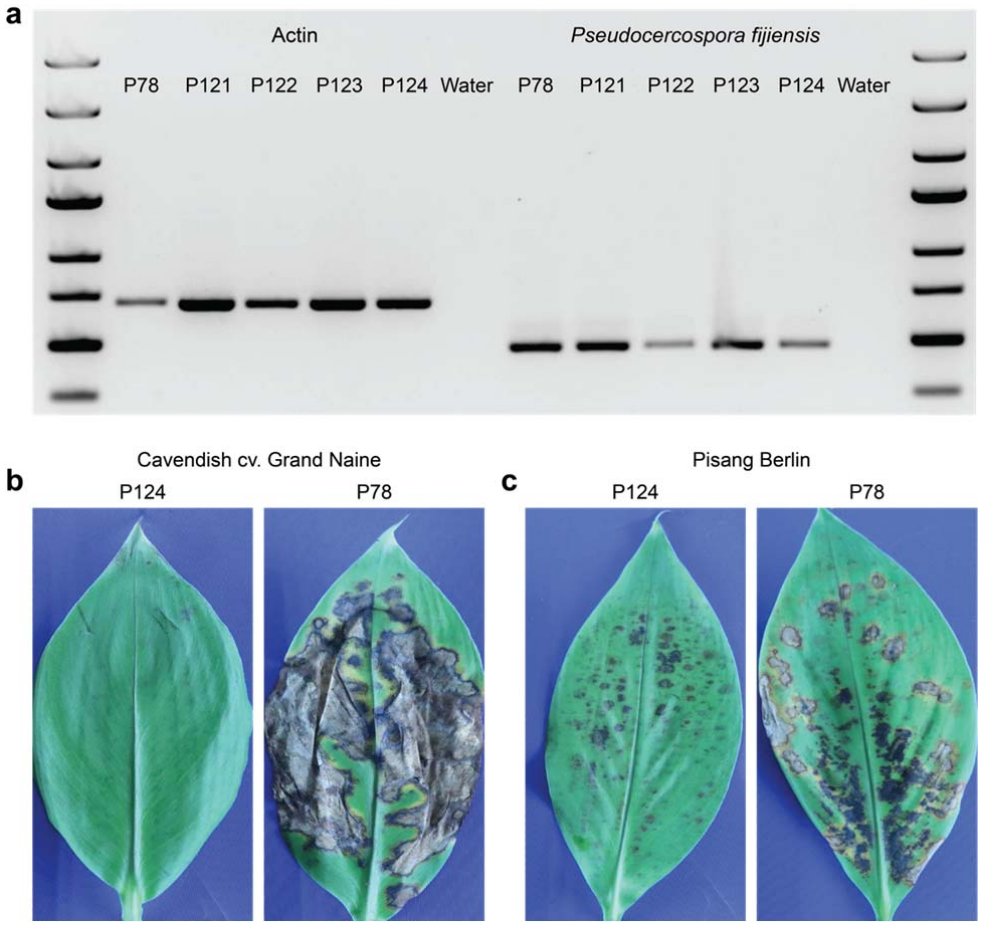


Figure 1. PCR diagnostics and pathogenicity assay of isolate P124 in comparison to the *Pseudocercospora fijiensis* reference strain P78 a) All isolates show amplification with actin primers (left panel). *Pseudocercospora*-specific primers amplified PCR products from *P. fijiensis* strain P78 (positive 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.

Complete genome assembly of P124

To generate a high-quality genome assembly for the isolates from Bohol, we sequenced strain P124 using Oxford Nanopore Technology, which yielded 11.9 Gb of reads with an average size of ~9.1 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 contigs with an N50 of 3.4 Mb and a total

234 genome size of 42.5 Mb. Eleven of the 16 contigs had telomeric sequences
235 (TTAGGG) at both ends, and thus the assembly is highly contiguous and mostly
236 represents complete chromosomes (**Figure 2**). To assess genome completeness,
237 we queried the genome assembly for the presence of single-copy BUSCO genes
238 and identified 98.6% of the single-copy BUSCO genes that are expected in the
239 fungal order Capnodiales, indicating that the genome assembly of P124 covers the
240 conserved gene space (**Figure 2a**). We predicted a total of 14,789 protein-coding
241 genes in the genome and identified that 10.3% of the genome consists of repetitive
242 elements (**Figure 2b**). *De novo* genome assembly of isolates P121 – P123,
243 sequenced with short-read sequencing technology only, resulted in genome
244 assemblies with similar sizes (40.4 – 42.5 Mb). Although the assemblies of P121 -
245 P123 are more fragmented compared to the assembly of the nanopore-sequenced
246 isolate P124, they approximately contain an equally high number of expected
247 BUSCO genes (Table 1).
248
249 **Table 1. The genome assembly statistics of four sequenced isolates obtained**
250 **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%

251 *P124 is assembled using long reads from Oxford Nanopore Technology (ONT).
252 **P121-P123 are assembled using Illumina short reads.
253
254 The *P. fijiensis* reference genome is 73.6 Mb (Arango et al., 2016), and remarkably,
255 the genome assemblies of the here sequenced fungal isolates are approximately 30
256 Mb smaller than expected. To verify the assembly size, we estimated genome sizes
257 using k-mer profiles, which resulted in an estimated genome size ranging from 38.9
258 to 41.9 Mb, similar to the size of the *de novo* assembled genomes (Table 1).
259 Furthermore, we mapped the sequencing reads of isolates P121 – P124 to the
260 chromosome-level genome assembly of P124, mapping on average 95% of the
261 reads with limited genetic diversity between isolates (on average 15 SNPs per kb;
262 **Figure 2b**). Moreover, mapping short reads of P124 to the P124 chromosome-level
263 genome assembly did not reveal any major coverage deviation (**Figure 2c**),
264 indicating that the P124 genome assembly does not contain extensively collapsed
265 repetitive regions that could account for the smaller genome size compared to *P.*
266 *fijiensis*. This validates that, although the genome assembly of strain P124 is smaller
267 than expected, it is highly complete and continuous, suggesting that the length
268 variation is likely not due to assembly errors. To capture possible contaminations in
269 the genome assemblies, we queried all contigs to the BLAST database to determine

their identity. All contigs showed high similarity to members of the Mycosphaerellaceae family; most contigs were similar to *Zasmidium* species (13 contigs, 35 Mb) and only one contig (3.3 Mb) showed similarity to *Pseudocercospora* (**Figure S1**). Whole-genome alignments of our assembly to the *P. fijiensis* reference genome assembly revealed that none of the contigs/chromosomes display significant similarity (**Figure S2**). Therefore, we considered that the difference in genome size between isolates P121 – P124 and the *P. fijiensis* reference isolate P78 is likely caused by the isolation of a different fungal species associated with necrotic symptoms on banana foliage (Crous et al., 2021), most likely by a member of genus *Zasmidium*.

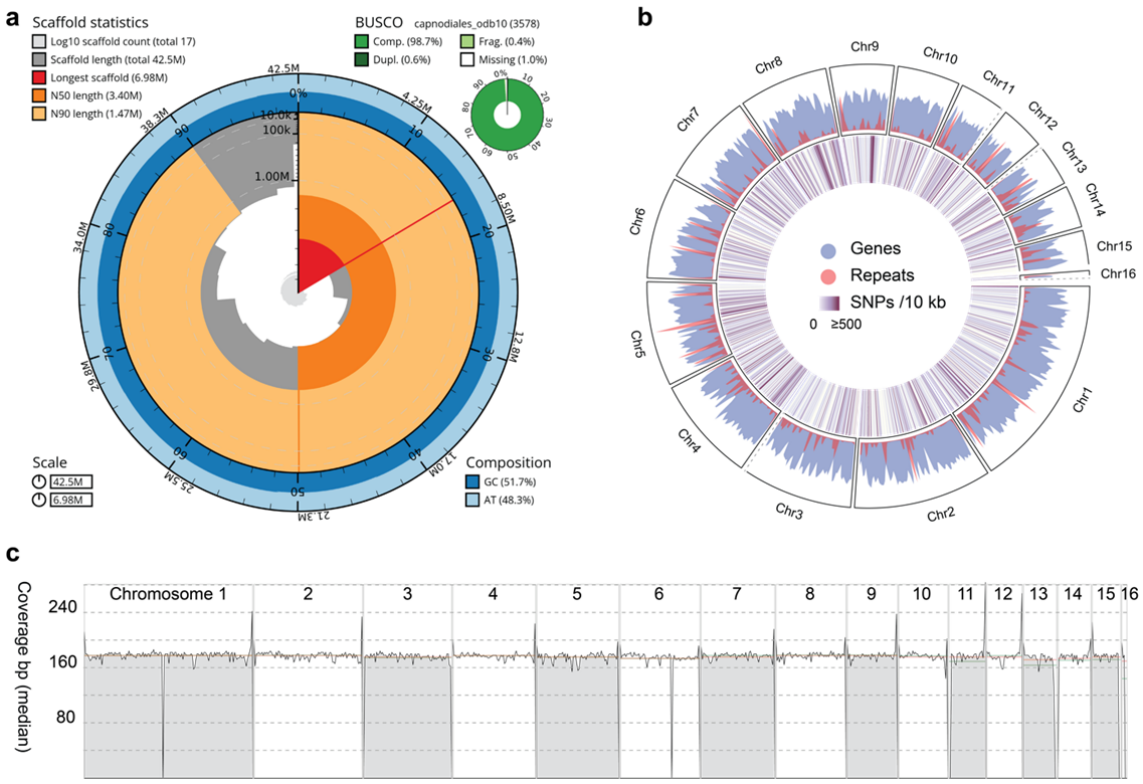


Figure 2. Chromosome-level genome assembly of *Zasmidium syzygii* isolate P124. a) Genome assembly statistics for the *de novo* assembly of *Z. syzygii* isolate P124 based on Oxford Nanopore Technology (ONT). The genome assembly has 17 contigs (16 nuclear contigs of which at least 11 are complete chromosomes, 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 chromosomes 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 chromosomes, with two SNP dense regions on chromosome 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.

297 **The assembled genome sequence reveals that isolate P124 belongs to the**
298 **fungal genus *Zasmidium***

299 To determine the identity of isolate P124, we retrieved the ITS sequence from the
300 genome assembly and searched for related sequences in the non-redundant BLAST
301 database on NCBI using the P124 ITS sequence as a query. We obtained a highly
302 similar match (98.7% nucleotide identity) to *Zasmidium syzygii* (NR_111826.1),
303 supporting our hypothesis that the assembled isolate belongs to the *Zasmidium*
304 genus. In line with this finding, previous studies have found *Zasmidium* isolates on
305 symptomatic banana leaves (Arzanlou et al., 2008; Crous et al., 2021), and other
306 *Zasmidium* species are reported to cause leaf spot diseases on other plant species
307 such as for example citrus (Aguilera-Cogley et al., 2017; An et al., 2021; Han et al.,
308 2015). The *Zasmidium* genome assembly we report here is less fragmented and 4
309 Mb larger than the previous genome assembly of a *Zasmidium* species, yet shares
310 extensive similarity to P124 (**Figure S2**). *Z. cellare* is known as the 'wine cellar
311 fungus' because it thrives in walls and ceilings of wine cellars (Tribe et al., 2006),
312 and was sequenced and its genome assembled into >267 scaffolds with a genome
313 size of ~38 Mb (Tribe et al., 2006), illustrating that our genome offers a more
314 continuous genome representation of a member of the *Zasmidium* genus.

315 To validate the identity and to determine the diversity of the *Zasmidium*
316 isolates, we compared the ITS sequences of isolate P121-P124 with 43 other
317 *Zasmidium* ITS sequences from NCBI (20 December 2022) as well as ITS
318 sequences of 120 Mycosphaerellaceae strains obtained from banana leaves (Crous
319 et al., 2021). A maximum-likelihood phylogeny based on the aligned ITS sequences
320 confirmed that isolates P121 – P124 belong to the *Zasmidium* genus and shows that
321 these four isolates encode identical ITS sequences (**Figure 3**). Notably, the set of

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

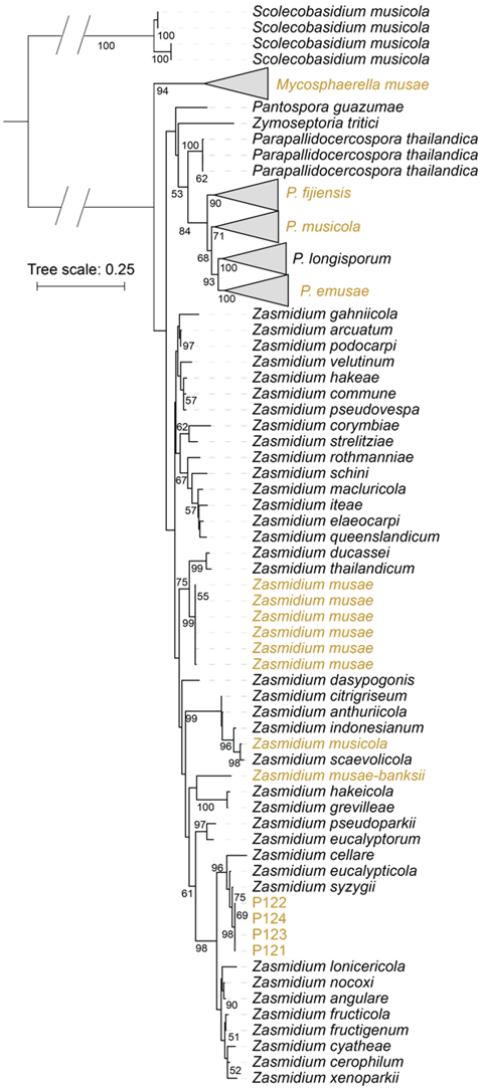


Figure 3. The four isolates sampled from Bohol (P121-P124) belong to the genus *Zasmidium.* A maximum-likelihood phylogenetic tree was constructed using the internally transcribed spacer (ITS) sequences of 43 *Zasmidium* species from NCBI and 120 Mycosphaerellaceae isolates from banana leaves (Crous 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. syzygii* 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

351 primers between *P. fijiensis* and *Z. syzygii*, we *in silico* detected the amplicon of the
352 supposedly *Pseudocercospora*-specific primers in the *Zasmidium syzygii* P124
353 genome assembly and in the *P. fijiensis* reference genome assembly (Cirad86;
354 Arango et al., 2016). Both isolates possess the primer sequence used to distinguish
355 *P. fijiensis* (**Figure S3**) from *P. musae* and *P. eumusae* and produce a similar sized
356 amplicon of 480 bp in *Z. syzygii* P124 and 478 bp in *P. fijiensis* Cirad86, which
357 explains the positive result for *Z. syzygii* in our PCR assay (**Figure 1**). The amplicons
358 share 88% sequence identity and *Zasmidium* or *Pseudocercospora* isolates can
359 therefore be distinguished only upon amplicon sequencing. Thus, novel primer pairs
360 need to be developed to enable easy and accurate diagnosis of fungal species
361 present in foliar blights of banana.
362

363 **Conclusion**

364 Here, we report the first chromosome-scale genome assembly of a *Zasmidium*
365 species, this adds a high-quality genome sequence to the thus far limited genetic
366 resources available for this genus. Our data show that *Z. syzygii* occurs on banana
367 foliage in Bohol and can cause leaf necrosis, comparable to the foliar blight
368 symptoms observed for *P. fijiensis*. The availability of the genome assembly will
369 facilitate further research into the association of *Zasmidium, Pseudocercospora*, and
370 possibly other fungal species related to foliar blights of banana. Moreover, it will
371 serve as a valuable resource for developing novel molecular diagnostics, enabling
372 the accurate identification and characterization of these fungal species.
373

374 *Data Availability Statement*

375 All sequencing data are deposited in the NCBI database under accession number
376 PRJNA931964.

377 *Acknowledgments*

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

382

383 *Funder Information*

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

387

388 *Conflict of Interest*

389 The authors report no conflict of interests.

390

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

Ametrano, C. G., Lumbsch, H. T., Di Stefano, I., Sangvichien, E., Muggia, L., & Grewe, F. (2022). Should we hail the Red King? Evolutionary consequences of a mutualistic lifestyle in genomes of lichenized ascomycetes. *Ecology and Evolution*, *12*(1), e8471. https://doi.org/10.1002/ece3.8471

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*

415 *Phylogeny and Evolution of Fungi*, *20*, 19–37.
416 https://doi.org/10.3767/003158508X302212
417 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S.,
418 Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V.,
419 Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A.
420 (2012). SPAdes: A new genome assembly algorithm and its applications to
421 single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
422 https://doi.org/10.1089/cmb.2012.0021
423 Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A.,
424 De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A.,
425 Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y.
426 J. K., Sanli, K., Eriksson, K. M., Vik, U., … Nilsson, R. H. (2013). Improved
427 software detection and extraction of ITS1 and ITS2 from ribosomal ITS
428 sequences of fungi and other eukaryotes for analysis of environmental
429 sequencing data. *Methods in Ecology and Evolution*, *4*(10), 914–919.
430 https://doi.org/10.1111/2041-210X.12073
431 Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit –
432 interactive quality assessment of genome assemblies. *G3*
433 *Genes*|*Genomes*|*Genetics*, *10*(4), 1361–1374.
434 https://doi.org/10.1534/g3.119.400908
435 Chang, T.-C., Salvucci, A., Crous, P. W., & Stergiopoulos, I. (2016). Comparative
436 genomics of the Sigatoka disease complex on banana suggests a link
437 between parallel evolutionary changes in *Pseudocercospora fijiensis* and
438 *Pseudocercospora eumusae* and increased virulence on the banana host.

439 *PLOS Genetics*, *12*(8), e1005904.
440 https://doi.org/10.1371/journal.pgen.1005904
441 Chong, P., Essoh, J. N., Arango Isaza, R. E., Keizer, P., Stergiopoulos, I., Seidl, M.
442 F., Guzman, M., Sandoval, J., Verweij, P. E., Scalliet, G., Sierotzski, H., de
443 Lapeyre de Bellaire, L., Crous, P. W., Carlier, J., Cros, S., Meijer, H. J. G.,
444 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

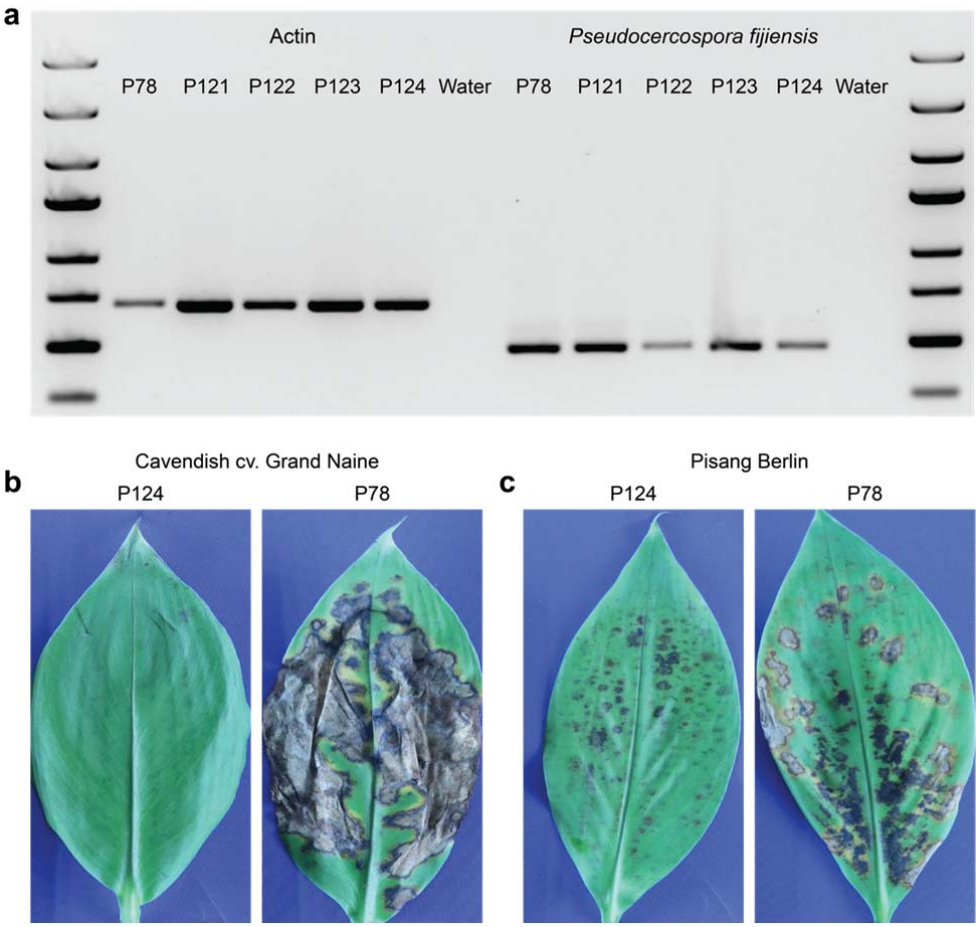
463 Drenth, A., & Kema, G. (2021). The vulnerability of bananas to globally emerging
464 disease threats. *Phytopathology®*, *111*(12), 2146–2161.
465 https://doi.org/10.1094/phyto-07-20-0311-rvw
466 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit,
467 A. F. (2020). 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 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M.
487 (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer

488 weighting and repeat separation. *Genome Research*, *27*(5), 722–736.
489 https://doi.org/10.1101/gr.215087.116
490 Kusch, S., Vaghefi, N., & Kiss, L. (2023). The good, the bad, and the misleading:
491 How to improve the quality of 'genome announcements'? *Molecular Plant-*
492 *Microbe Interactions®*. https://doi.org/10.1094/MPMI-01-23-0009-LE
493 Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for
494 phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1),
495 W293–W296. https://doi.org/10.1093/nar/gkab301
496 Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with*
497 *BWA-mem*. https://doi.org/10.6084/m9.figshare.963153.v1
498 Lu, J., & Salzberg, S. L. (2018). Removing contaminants from databases of draft
499 genomes. *PLOS Computational Biology*, *14*(6), e1006277.
500 https://doi.org/10.1371/journal.pcbi.1006277
501 Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
502 counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770.
503 https://doi.org/10.1093/bioinformatics/btr011
504 Osorio, J. A., Linnakoski, R., Wingfield, M. J., de Beer, Z. W., & Roux, J. (2021). A
505 new species in the Mycosphaerellaceae *from Cecidomyiidae leaf galls* on
506 *Avicennia marina* in South Africa. *Antonie van Leeuwenhoek*, *114*(5), 515–
507 526. https://doi.org/10.1007/s10482-021-01537-3
508 Palmer, J., & Stajich, J. (2019). *nextgenusfs/funannotate: Funannotate v1.5.3*
509 [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2604804
510 Rachtman, E., Balaban, M., Bafna, V., & Mirarab, S. (2020). The impact of
511 contaminants on the accuracy of genome skimming and the effectiveness of

512 exclusion read filters. *Molecular Ecology Resources*, *20*(3), 649–661.
513 https://doi.org/10.1111/1755-0998.13135
514 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M.
515 (2015). BUSCO: Assessing genome assembly and annotation completeness
516 with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.
517 https://doi.org/10.1093/bioinformatics/btv351
518 Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0*.
519 http://www.repeatmasker.org
520 Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: Large-scale
521 search identifies more than 2,000,000 contaminated entries in GenBank.
522 *Genome Biology*, *21*(1), 115. https://doi.org/10.1186/s13059-020-02023-1
523 Tribe, H. T., Thines, E., & Weber, R. W. S. (2006). Moulds that should be better
524 known: The wine cellar mould, *Racodium cellare*. *Mycologist*, *20*(4), 171–175.
525 https://doi.org/10.1016/j.mycol.2006.09.016
526 Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-
527 Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E.,
528 Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From
529 fastQ data to high-confidence variant calls: The genome analysis toolkit best
530 practices pipeline. *Current Protocols in Bioinformatics*, *43*(1), 11.10.1-
531 11.10.33. https://doi.org/10.1002/0471250953.bi1110s43
532 Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H.,
533 Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free
534 genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204.
535 https://doi.org/10.1093/bioinformatics/btx153

536 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo,
537 C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An
538 integrated tool for comprehensive microbial variant detection and genome
539 assembly improvement. *PLOS ONE*, *9*(11), e112963.
540 https://doi.org/10.1371/journal.pone.0112963
541 Xu, C., Zhang, R., Sun, G., & Gleason, M. L. (2017). Comparative genome analysis
542 reveals adaptation to the ectophytic lifestyle of sooty blotch and flyspeck
543 fungi. *Genome Biology and Evolution*, *9*(11), 3137–3151.
544 https://doi.org/10.1093/gbe/evx229
545
546

547 **Figures and tables**



548

549 **Figure 1. PCR diagnostics and pathogenicity assay of isolate P124 in**
550 **comparison to the *Pseudocercospora fijiensis* reference strain P78** a) All
551 isolates show amplification with actin primers (left panel). *Pseudocercospora*-specific
552 primers amplified PCR products from *P. fijiensis* strain P78 (positive control) as well
553 as from isolates P121, P122, P123, and P124 (right panel), suggesting that the
554 isolates can be identified as *P. fijiensis*. b) Isolate P124 does not cause necrotic
555 lesions on Cavendish banana, in contrast to necrotic symptoms caused by P78. c)
556 Both P124 and P78 cause necrotic leaf symptoms on Pisang Berlin. Disease
557 symptoms were scored eight weeks after inoculation.
558

559 **Table 1. The genome assembly statistics of four sequenced isolates obtained**
560 **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%

561 *P124 is assembled using long reads from Oxford Nanopore Technology (ONT).
562 **P121-P123 are assembled using Illumina short reads.
563

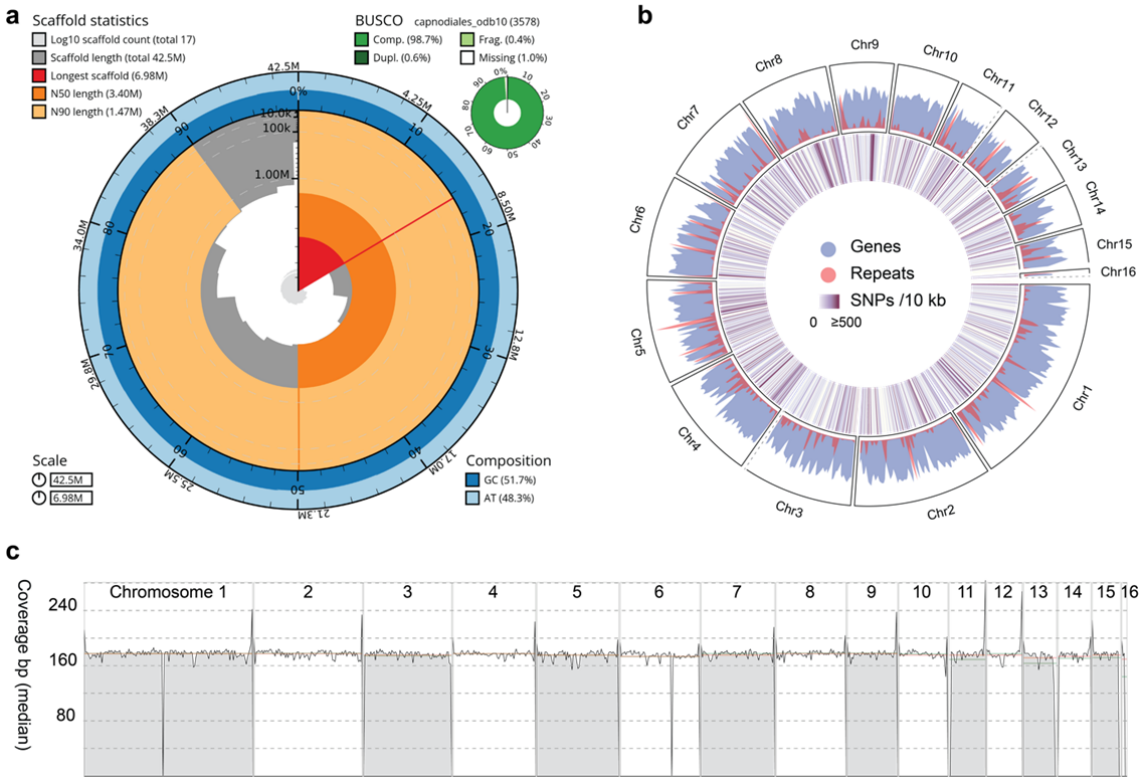


Figure 2. Chromosome-level genome assembly of *Zasmidium syzygii* isolate P124. a) Genome assembly statistics for the *de novo* assembly of *Z. syzygii* isolate P124 using Oxford Nanopore Technology (ONT). The genome assembly has 17 contigs (16 nuclear contigs of which at least 11 are complete chromosomes, 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 chromosomes 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 chromosomes, with two SNP dense regions on chromosome 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 repetitive regions that may have collapsed during the assembly process.

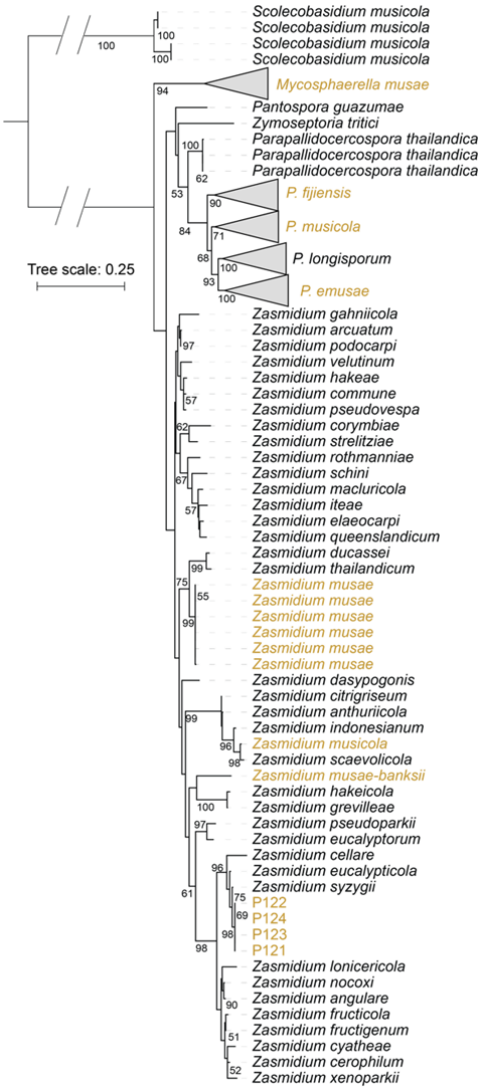
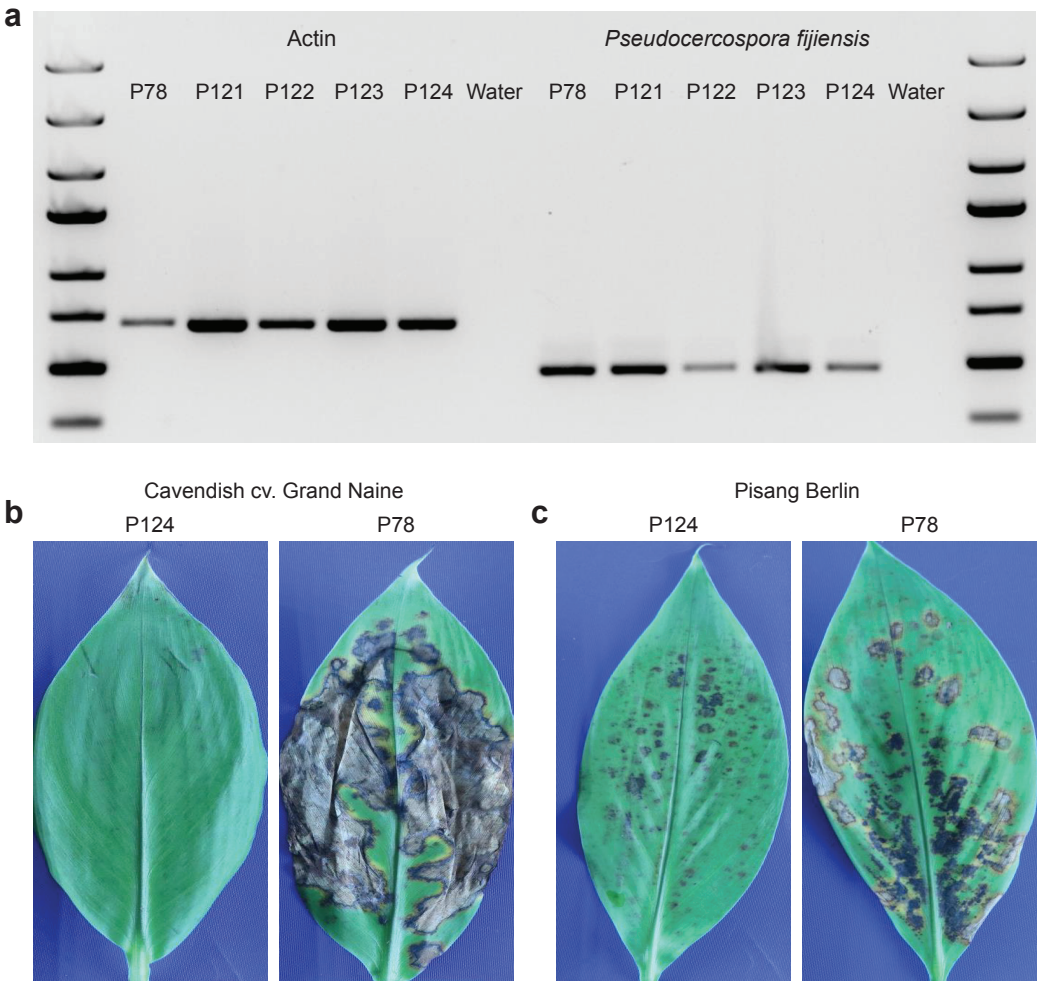
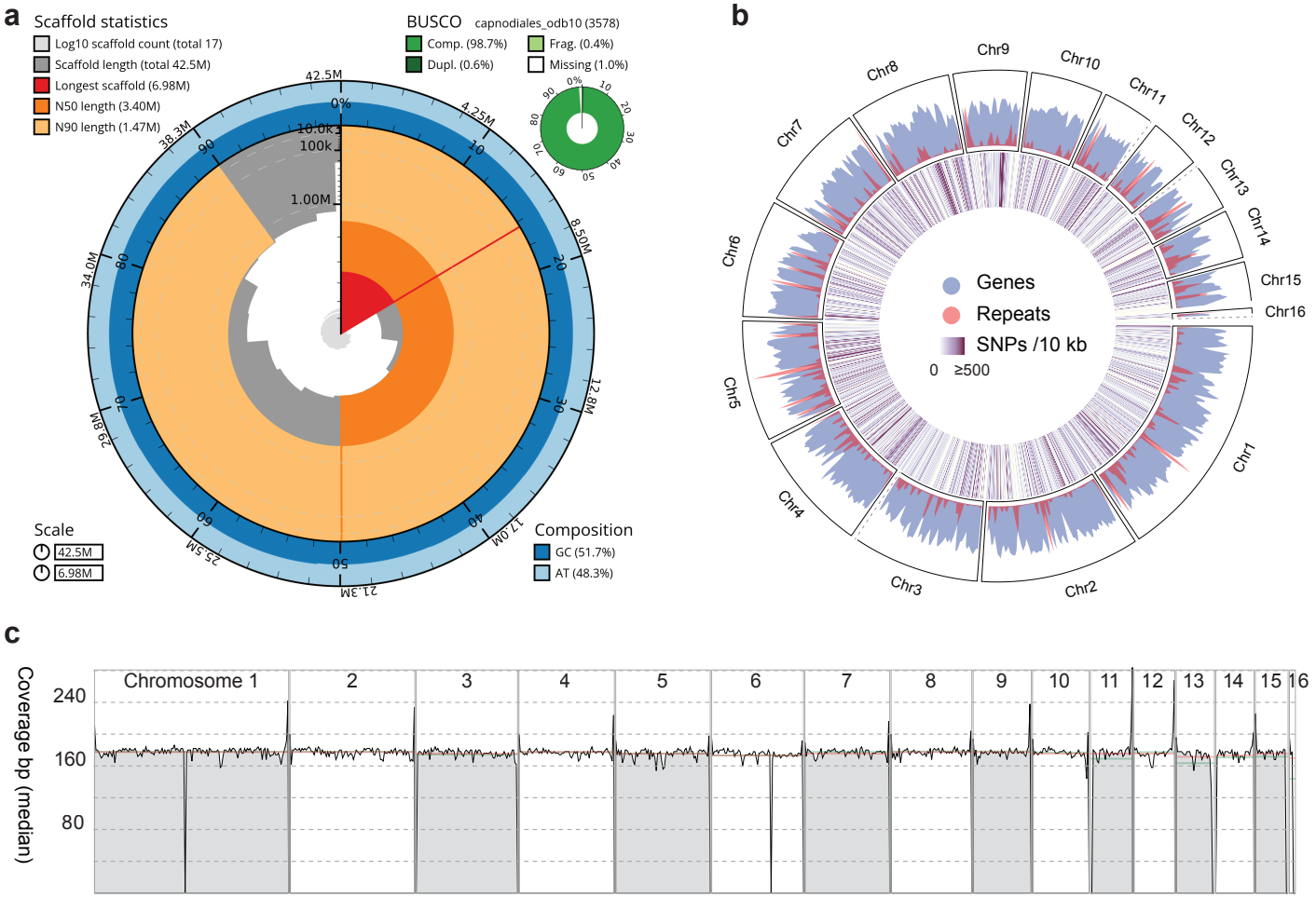
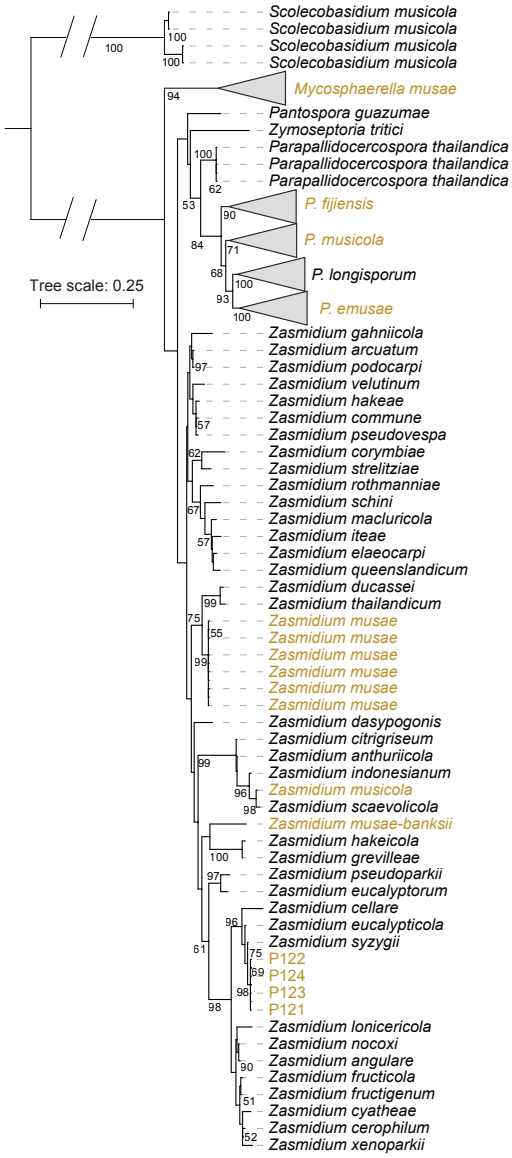


Figure 3. The four isolates sampled from Bohol (P121-P124) belong to the genus *Zasmidium.* Maximum-likelihood phylogenetic tree is constructed using the internally transcribed spacer (ITS) sequences of 43 *Zasmidium* species from NCBI and 120 Mycosphaerellaceae isolates from banana leaves (Crous et 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. syzygii* 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.


a
Actin
Pseudocercospora fijiensis
P78
P121
P122
P123
P124
Water
P78
P121
P122
P123
P124
Water
b
Cavendish cv. Grand Naine
P124
P78
c
Pisang Berlin
P124
P78


a
Scaffold statistics
Log10 scaffold count (total 17)
Scaffold length (total 42.5M)
Longest scaffold (6.98M)
N50 length (3.40M)
N90 length (1.47M)
BUSCO capnodiales_odb10 (3578)
Comp. (98.7%)
Frag. (0.4%)
Dupl. (0.6%)
Missing (1.0%)
Scale
42.5M
6.98M
Composition
GC (51.7%)
AT (48.3%)
b
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Genes
Repeats
SNPs /10 kb
0
≥500
c
Coverage bp (median)
240
160
80
Chromosome 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16


Scolecobasidium musicola
Scolecobasidium musicola
Scolecobasidium musicola
Scolecobasidium musicola
Mycosphaerella musae
Pantospora guazumae
Zymoseptoria tritici
Parapallidocercospora thailandica
Parapallidocercospora thailandica
Parapallidocercospora thailandica
P. fijiensis
P. musicola
P. longisporum
P. emusae
Tree scale: 0.25
Zasmidium gahniicola
Zasmidium arcuatum
Zasmidium podocarpi
Zasmidium velutinum
Zasmidium hakeae
Zasmidium commune
Zasmidium pseudovespa
Zasmidium corymbiae
Zasmidium strelitziae
Zasmidium rothmanniae
Zasmidium schini
Zasmidium macluricola
Zasmidium iteae
Zasmidium elaeocarpi
Zasmidium queenslandicum
Zasmidium ducassei
Zasmidium thailandicum
Zasmidium musae
Zasmidium musae
Zasmidium musae
Zasmidium musae
Zasmidium musae
Zasmidium musae
Zasmidium dasypogonis
Zasmidium citrigriseum
Zasmidium anthuriicola
Zasmidium indonesianum
Zasmidium musicola
Zasmidium scaevolicola
Zasmidium musae-banksii
Zasmidium hakeicola
Zasmidium grevilleae
Zasmidium pseudoparkii
Zasmidium eucalyptorum
Zasmidium cellare
Zasmidium eucalypticola
Zasmidium syzygii
P122
P124
P123
P121
Zasmidium lonicericola
Zasmidium nocoxi
Zasmidium angulare
Zasmidium fructicola
Zasmidium fructigenum
Zasmidium cyatheae
Zasmidium cerophilum
Zasmidium xenoparkii