




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

Anouk C. van Westerhoven,^{1,2,‡} Rahim Mehrabi,^{1,3,‡} Reza Talebi ,³ Maikel B.F. Steentjes ,¹ Benny Corcolon,⁴ Pablo A. Chong ,⁵ Gert H.J. Kema,^{1,*‡} Michael F. Seidl^{2,*‡}

¹Laboratory of Phytopathology, Wageningen University & Research, Wageningen 6700 AA, The Netherlands

²Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Utrecht 3584 CS, The Netherlands

³Keygene N.V., Wageningen 6700 AE, The Netherlands

⁴Research, Information, Compliance Department, Tadeco Inc., Panabo, Davao del Norte 8105, Philippines

⁵Escuela Superior Politécnica del Litoral, Centro de Investigaciones Biotecnológicas del Ecuador, Laboratorio de Biología Molecular, ESPOL Polytechnic University, Guayaquil 090112, Ecuador

*Corresponding author: Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands. Email: m.f.seidl@uu.nl;

*Corresponding author: Laboratory of Phytopathology, Wageningen University & Research, Droevendaalsesteeg 1, 6700 AA, Wageningen, The Netherlands.

Email: gert.kema@wur.nl

‡These authors contributed equally.

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 3 *Pseudocercospora* species, of which the most predominant causal agent is *Pseudocercospora 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 internal transcribed spacer (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 most likely complete chromosomes. The genome contains 98.6% of the expected single-copy BUSCO genes and contains 14,789 genes and 10.3% repeats. The 3 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 4 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.

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

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 (Lu and Salzberg 2018; Francois et al. 2020; Rachtman et al. 2020). However, publicly available genome assemblies are occasionally reported to contain contaminants, which can lead to incorrect species classification (Steinegger and Salzberg 2020; Cornet and Baurain 2022; Kusch et al. 2023). 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 (Zaccaron and Stergiopoulos 2021; Kusch et al. 2023).

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 3 *Pseudocercospora* species with *Pseudocercospora 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 and Kema 2021). The 2 other species also cause foliar blights but are currently less prevalent: *Pseudocercospora musae* causes yellow Sigatoka and *Pseudocercospora eumusae* causes eumusae leaf spot (Chang et al. 2016). However, besides these 3 *Pseudocercospora* species, other fungal species can appear in association with symptomatic banana foliage. For example, a recent study that analyzed fungal isolates associated with banana foliar diseases revealed the presence of over 30 other fungal species, primarily belonging to the *Mycosphaerellaceae* 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 are important, particularly for developing disease diagnostics and enhancing the

effectiveness of disease management strategies (Lu and Salzberg 2018; Kusch et al. 2023).

Here, we sampled 4 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 3 fragmented genome assemblies of *Zasmidium* species are publicly available (Xu et al. 2017; Haridas et al. 2020). 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.

Methods and materials

Fungal isolation and sequencing

Banana leaves with symptoms of foliar disease were collected from a field in Bohol, Philippines. Leaf samples were cut into 1-cm² segments and used to discharge ascospores as described previously (Chong et al. 2019). Four single ascospore isolates (P121, P122, P123, and P124) were collected and grown on potato dextrose agar (PDA) plates supplemented with streptomycin (100 µg/ml) for 3 weeks at 25°C. To obtain sufficient fungal biomass for DNA isolation, a piece of mycelium (2 cm²) was blended for 20 s at 6,000 rpm in an ULTRA-TURRAX Tube Drive Homogenizer (IKA, Staufen, Germany) in 15-ml water using a sterile DT-20 tube (IKA, Staufen, Germany). The fragmented mycelium was transferred into a flask containing 100 ml of PDB amended with streptomycin (100 µg/ml) and kept at 25°C on a rotary shaker with 150 rpm for about 2 weeks. The fungal mycelium was filtered through Miracloth and subsequently washed with sterile water. Fungal mycelium was freeze dried overnight and used for high-molecular-weight (HMW) DNA isolation based on the CTAB method (Murray and Thompson 1980). After adding isopropanol, HMW DNA was collected from the extraction buffer using a sterile needle. AMPure XP purification kit (Beckman Cluter Life Sciences, USA) was used to clean up the DNA. DNA quality and quantity were checked by gel electrophoresis, NanoDrop microvolume spectrophotometers, and Qubit Fluorometric Quantitation (Thermo Fisher Scientific, USA). The HMW DNA of isolate P124 was sequenced using PromethION Oxford Nanopore Technology (ONT) sequencing technologies. Additionally, all isolates were sequenced using the Illumina HiSeq platform; both sequencing platforms were located at Keygene B.V. (Wageningen, the Netherlands).

Fungal species diagnosis by PCR

To identify the isolated fungal species, the DNA from each isolate was extracted and subjected to *P. fijiensis*-specific diagnostic PCR according to a previously established protocol (Arzanlou et al. 2008). The list of primers used in PCRs is shown in Supplementary Table 1. Actin (820 bp) was used as a positive control to ensure successful amplification and to assess the quality of DNA.

Pathogenicity assay

The pathogenicity of isolate P124 was tested on 2 banana genotypes: Cavendish cv. Grand Naine (AAA) and a diploid (AA) genotype 'Pisang Berlin'. The inoculum for isolate P124 and inoculum for the *P. fijiensis* reference isolate P78 were prepared similarly. A

piece of 1-cm² mycelium from 3-week-old colonies grown on PDA was collected in an Eppendorf tube containing 3–4 metal beads (3 mm diameter) and was blended for 20 s at 3,000 rpm in a Bead Beater Homogenizer. About 1 ml of sterile water was added to each tube, and fragmented mycelium was spread on PDA plates amended with 100-µg/ml streptomycin. The plates were kept at 25°C for 3–4 weeks, and then a piece of mycelium (10 cm²) was blended for 40 s at 6,000 rpm in an ULTRA-TURRAX Tube Drive homogenizer (IKA, Staufen, Germany) in 15 ml of distilled water using a sterile DT-20 tube (IKA, Staufen, Germany). The suspension was passed through Miracloth to remove nonfragmented mycelium. The collected mycelial fragments were further diluted and adjusted to 5×10^5 fragments ml⁻¹ and supplemented with 0.15% Tween 20. This suspension was used to inoculate 2-month-old banana plants on both sides of the leaves. Each treatment was repeated 3 times; as a control, water was used for mock inoculation. Inoculated plants were kept for 48 h at 90% relative humidity at 25°C in the dark in a growth cabinet and subsequently for 8 weeks in a greenhouse with >85% RH and with a day length of 12 h.

De novo genome assembly

To construct a complete and continuous genome assembly, we assembled the ONT long-read sequences of isolate P124. The raw Fast5 files were basecalled with Guppy (v. 5.1.13) and filtered for a minimum quality score of 10. The filtered reads were assembled de novo using Canu (v. 2.2, Koren et al. 2017) with the parameters “-OvlMerThreshold = 300” and “-corMaxEvidenceErate = 0.15” to improve the assembly of repetitive regions. The initial genome assembly was polished with Racon (v. 1.5.0, Vaser et al. 2017) using the raw ONT long reads. After polishing, missing telomeres were reconstructed, when possible, using teloclip (v. 0.0.3, <https://github.com/Adamtaranto/teloclip>), and a final polishing step was performed based on short-read Illumina data using Pilon (v. 1.24, Walker et al. 2014). The presence of telomeres (TTAGGG) and coverage of the contigs in the final polished assembly were determined using tapestry (v. 1.0.0, Davey et al. 2020). To estimate the completeness of the genome assembly, the presence of conserved single-copy genes was determined using BUSCO with the Capnodiales database (Simão et al. 2015). The assembly continuity and potential contaminations were accessed using BlobTools (v. 4.1.5, Challis et al. 2020). To estimate the size of the assembled genome, all Illumina sequences were analyzed using GenomeScope (v. 2.0, Vurture et al. 2017) with the 21-mer profile determined by JellyFish (v 2.3.0, Marçais and Kingsford 2011). Isolate P121 showed high k-mer coverage outliers; therefore, the maximal k-mer coverage was set to 350x, similar to the maximum k-mer coverage observed in the other genome. Repeats were predicted using RepeatModeler (v. 2.0.3, Flynn et al. 2020) and annotated using RepeatMasker (v. 4.1.2, Smit et al. 2015). Finally, protein-coding genes were annotated using Funannotate (v. 1.8.9, Palmer and Stajich 2019) with the BUSCO lineage *Pezizomycotina* utilizing predicted *Zasmidium cellare* proteins as evidence (GCF_010093935.1, Haridas et al. 2020). Additionally, we assembled the Illumina sequenced isolates P121-P123 with Spades (v. 3.13, Bankevich et al. 2012). To assess the quality and completeness of these assemblies, we ran QUAST (v. 5.2, Gurevich et al. 2013) and BUSCO (v. 5.3.2, Simão et al. 2015) with the Capnodiales odb10 database.

Identification of internal transcribed spaces sequence

The internal transcribed spacer (ITS) sequences from the 4 assemblies (P121–P124) were retrieved using ITSx (v. 1.1.3,

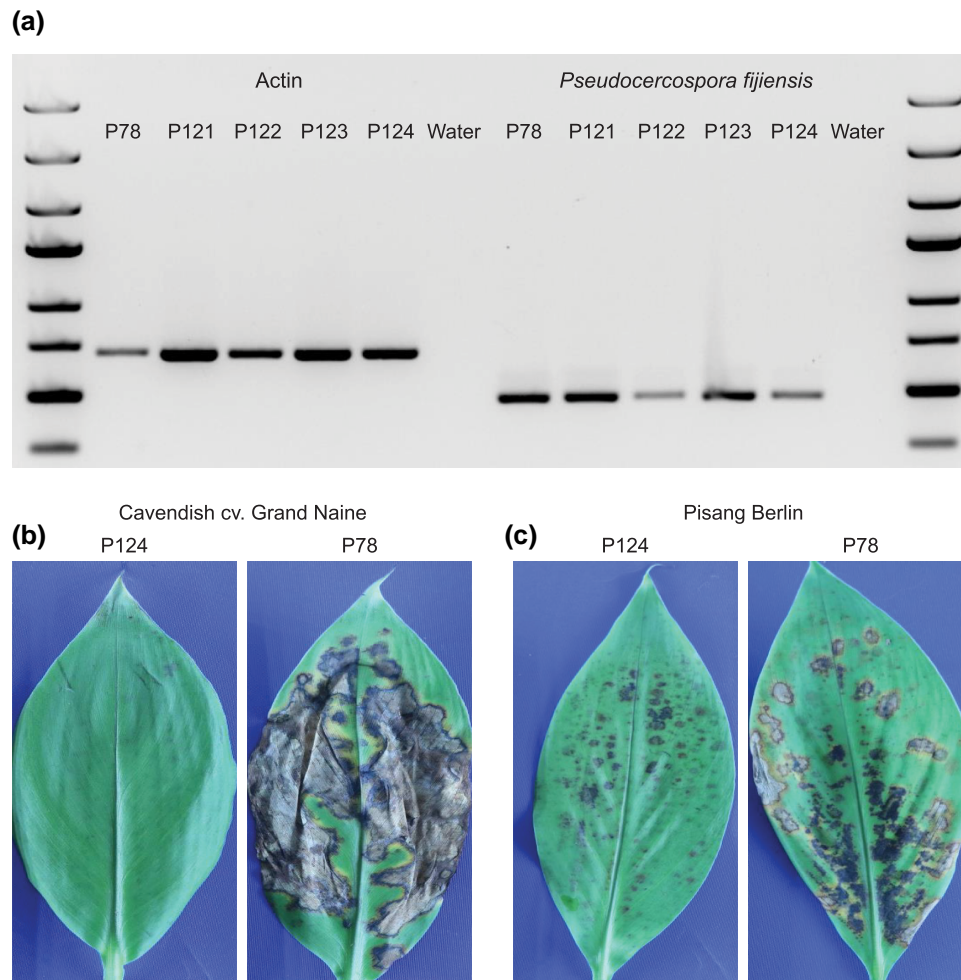


Fig. 1. PCR diagnostics and pathogenicity assay of isolate P124 in comparison to the *P. fijiensis* reference strain P78. a) All isolates show amplification with actin primers (left panel). *Pseudocercospora*-specific primers amplified PCR products from *P. fijiensis* strain P78 (positive control) as 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 8 weeks after inoculation.

Bengtsson-Palme *et al.* 2013), and BLASTn was used to detect similar ITS sequences in the BLAST database (accessed 2022 November 19). Additionally, a maximum-likelihood tree was constructed to compare the 4 ITS sequences to 227 publicly available ITS sequences obtained from NCBI (accessed 2023 January 13). Sequences were aligned using MAFFT (v.7. 453, Katoh *et al.* 2002), and a maximum-likelihood phylogenetic tree was constructed using RAxML, with 500 bootstrap replicates (-m GTRCAT -p 1234 -b 100 -N 500) (v. 8.2.12, Stamatakis 2014). The resulting phylogeny was visualized using iTOL (Letunic and Bork 2021).

Genetic diversity

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

Results and discussion

Fungal isolates obtained from symptomatic banana foliage in Bohol

Foliar blights of banana are mainly caused by a complex of *Pseudocercospora* species (Chang *et al.* 2016; Drenth and 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, 4 samples were obtained from banana plants with necrotic lesions in Bohol, Philippines (Chong *et al.* 2021). For all 4 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* (Fig. 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 leaf spot symptoms were observed on Cavendish upon inoculation with P124, in contrast to the necrotic lesions caused by P78 (Fig. 1b). However, both isolates caused necrotic lesions on Pisang Berlin 4 weeks after inoculation (Fig. 1c).

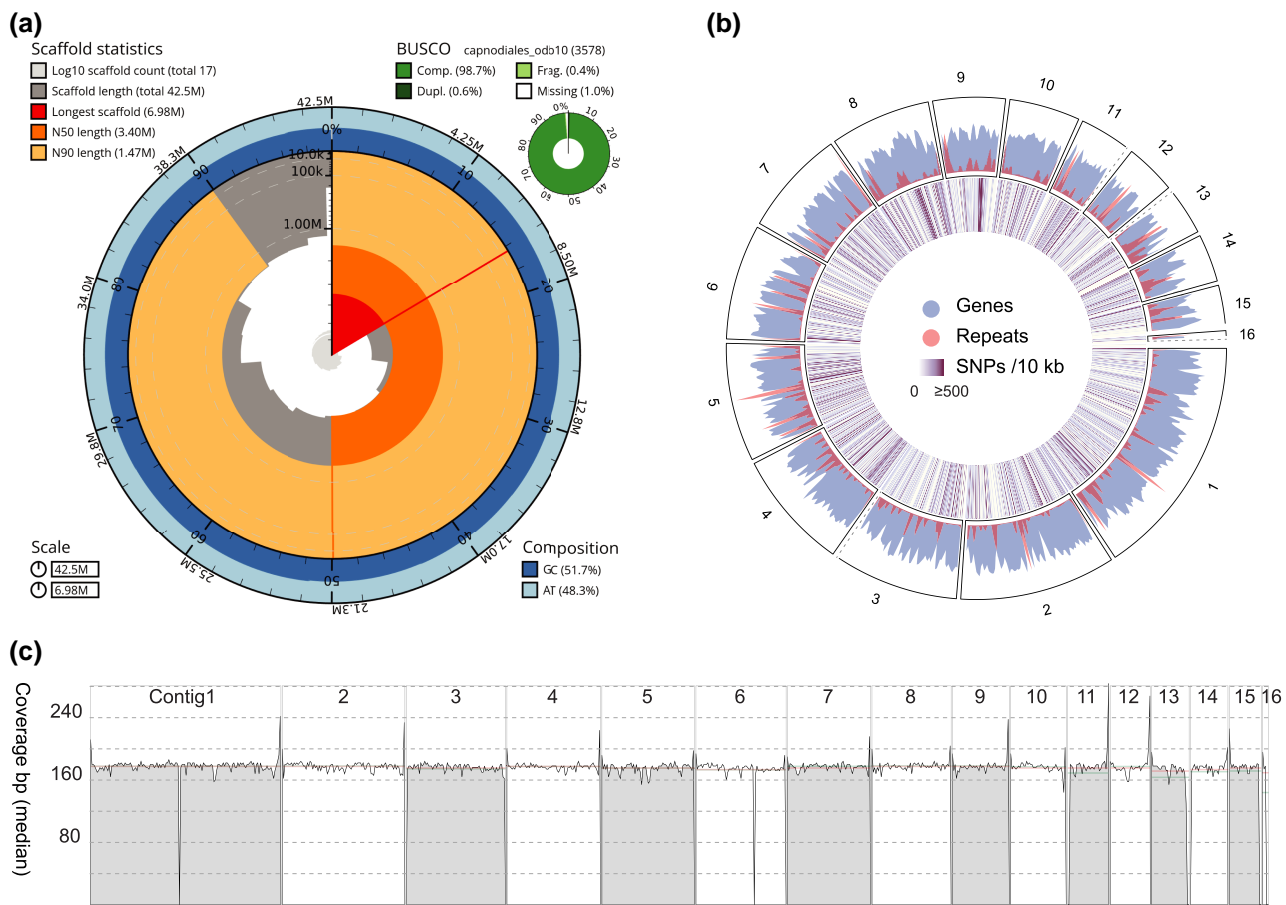


Fig. 2. Chromosome-level genome assembly of *Z. syzygii* isolate P124. a) Genome assembly statistics for the de novo assembly of *Z. syzygii* isolate P124 based on ONT. The genome assembly has 17 contigs (16 nuclear contigs of which at least 11 are complete chromosomes and the mitochondrial genome) with 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 contigs 3, 11, 13, 15, and 16. Genes (14,786) and repeats (10.3%) are distributed evenly over the contigs. SNPs are found overall contigs, with 2 SNP dense regions on contigs 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.

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 2 isolates differ in pathogenicity, both isolates cause necrosis on banana foliage.

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 ONT. 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 ~150× genome coverage based on the estimated genome size of 74 Mb (Arango et al. 2016). The de novo genome assembly resulted in 16 nuclear contigs and 1 contig representing the mitochondrial genome. The assembly has an N50 of 3.4 Mb and a total genome size of 42.5 Mb. Eleven of the 16 contigs had telomeric sequences (TTAGGG) at both ends, and thus, the assembly is highly contiguous and mostly represents complete chromosomes (Fig. 2). To assess genome completeness, we queried the genome assembly for the presence of single-copy BUSCO genes and identified 98.6% of the single-copy BUSCO genes that are expected in the fungal order Capnodiales, indicating that the genome assembly of P124 covers the conserved gene space (Fig. 2a). We predicted a total of 14,789 protein-coding genes in the genome and identified

that 10.3% of the genome consists of repetitive elements (Fig. 2b). De novo genome assembly of isolates P121–P123, sequenced with short-read sequencing technology only, resulted in genome assemblies with similar sizes (40.4–42.5 Mb). Although the assemblies of P121–P123 are more fragmented compared to the assembly of the nanopore-sequenced isolate P124, they approximately contain an equally high number of expected BUSCO genes (Table 1).

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 kilobase; Fig. 2b). An increased read mapping coverage is observed in most telomeric regions (Fig. 2c), while the coverage drops in contig 1, contig 6, and contig 16, suggesting that these regions possibly contain assembly artifacts. Contig 16 is only 0.2 Mb in size and only contains a telomeric repeat at one end. The low coverage region together with read mappings that support a

Table 1. The genome assembly statistics of four sequenced isolates obtained from symptomatic banana foliage in Bohol, Philippines.

	P124 ^a	P121 ^b	P122 ^b	P123 ^b
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%

^a P124 is assembled using long reads from ONT.

^b P121-P123 are assembled using Illumina short reads.

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 read mapping coverage over the contigs, indicating that the P124 genome assembly does not contain extensively collapsed repetitive regions that could account for the smaller genome size compared to *P. fijiensis*. This validates that, although the genome assembly of strain P124 is smaller than expected, it is highly complete and continuous, suggesting that the length variation is likely not due to assembly errors. To capture possible contaminations in the genome assemblies, we queried all contigs in the BLAST database to determine their identity. All contigs showed a high similarity to members of the *Mycosphaerellaceae* family; most contigs were similar to *Zasmidium* species (13 contigs, 35 Mb) and only 1 contig (3.3 Mb) showed similarity to *Pseudocercospora* species (Supplementary Fig. 1). Whole-genome alignments of our assembly to the *P. fijiensis* reference genome assembly revealed that none of the contigs display significant similarity (Supplementary Fig. 2). 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*.

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

To determine the identity of isolate P124, we retrieved the ITS sequence from the genome assembly and searched for related species in the nonredundant BLAST database on NCBI using the P124 ITS sequence as a query. We obtained a highly similar match (98.7% nucleotide identity) to *Z. syzygii* (NR_111826.1), supporting that the assembled isolate belongs to the *Zasmidium* genus. In line with this finding, previous studies have found *Zasmidium* isolates on symptomatic banana leaves (Arzanlou et al. 2008; Crous et al. 2021), and other *Zasmidium* species are reported to cause leaf spot diseases on other plant species such as citrus (Han et al. 2015; Aguilera-Cogley et al. 2017; An et al. 2021). The *Zasmidium* genome assembly we report here is less fragmented and 4 Mb larger than the previous genome assembly of a *Zasmidium* species, *Z. cellare*. The species, known as the “wine cellar fungus” because it thrives in walls and ceilings of wine cellars (Tribe et al. 2006), was

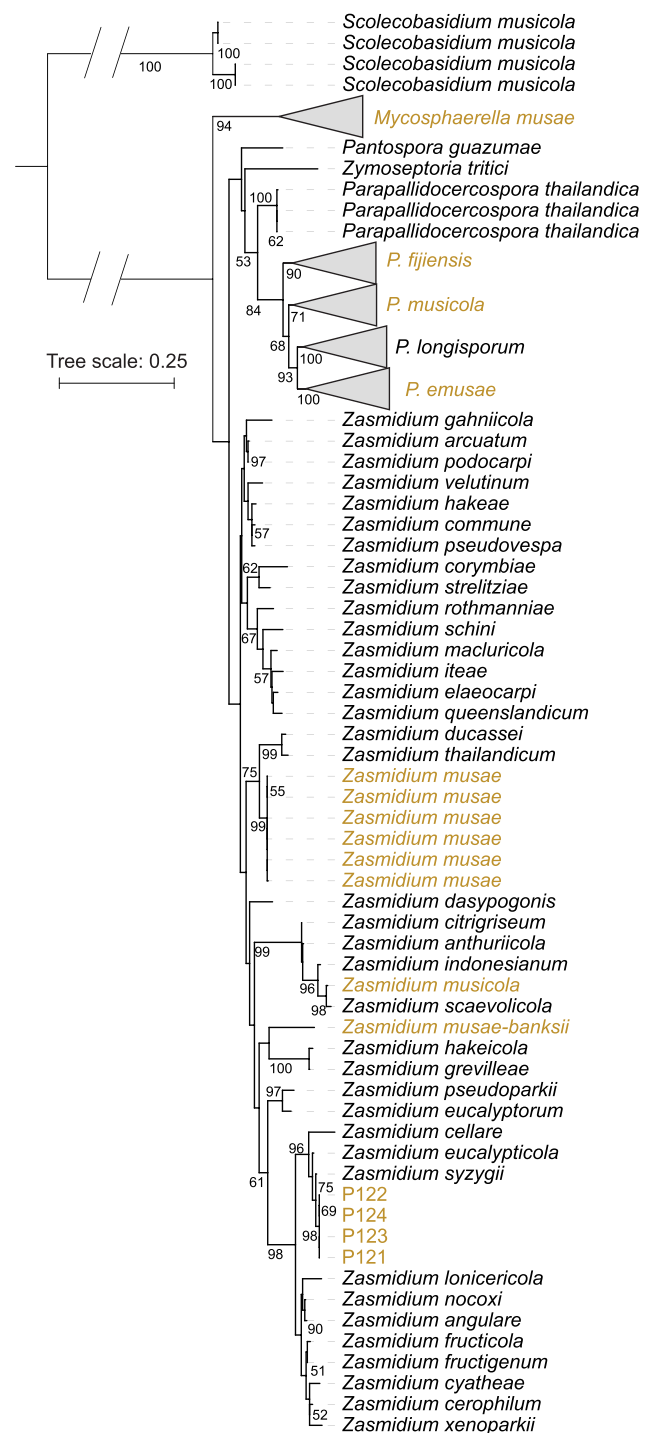


Fig. 3. The 4 isolates sampled from Bohol (P121–P124) belong to the genus *Zasmidium*. Maximum-likelihood phylogenetic tree is constructed using the ITS sequences of 43 *Zasmidium* species from NCBI and 120 *Mycosphaerella* isolates from banana leaves (Crous et al. 2021). The ITS sequence extracted from the chromosome-level genome assembly (P124) as well as from the other isolates (P121–123) associated with an isolate classified as *Z. syzygii* 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.

sequenced, and its genome was assembled into >267 scaffolds with a genome size of ~38 Mb (Haridas et al. 2020), illustrating that our genome offers a more continuous genome representation of a member of the *Zasmidium* genus.

To validate the identity and to determine the diversity of the *Zasmidium* isolates, we compared the ITS sequences of isolates P121–P124 with 43 other *Zasmidium* ITS sequences from NCBI (2022 December 20) as well as ITS sequences of 120 *Mycosphaerella* strains obtained from banana leaves (Crous et al. 2021). A maximum-likelihood phylogeny based on the aligned ITS sequences confirmed that isolates P121–P124 belong to the *Zasmidium* genus and shows that these 4 isolates encode identical ITS sequences (Fig. 3). Notably, the set of ITS sequences also contains sequences from 5 other *Zasmidium* isolates that had been previously sampled from banana leaves from different geographic locations (Martinique, Tonga, and Gabon) (Crous et al. 2021). Interestingly, these do not cluster with isolates P121–P124, suggesting that pathogenicity toward banana is a polyphyletic trait within the *Zasmidium* genus (Fig. 3). Although *Zasmidium* species have been linked to foliar blights in various hosts (Han et al. 2015; Aguilera-Cogley et al. 2017; Osorio et al. 2021), the pathogenicity and global spread of *Zasmidium* species has not been studied in depth. Based on our data, we conclude that *Z. syzygii* occurs on banana leaves with necrotic symptoms and can be the cause of mild necrotic lesions on the foliage. However, the abundance and role of *Z. syzygii* as a banana pathogen remains unknown, which requires further research to understand its prevalence, significance, and potential impact on banana cultivation.

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

Conclusion

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

Data availability

All sequencing data are deposited in the NCBI database under accession number PRJNA931964. Genome assembly, genome

annotation, and vcf files are available on figshare (<https://doi.org/10.25387/g3.24117504>). Scripts used for the genome assembly and analysis are available on github (<https://github.com/Anouk-vw/Zasmidium>).

Supplemental material available at G3 online.

Acknowledgments

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

Funding

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

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

- Aguilera-Cogley VA, Berbegal M, Català S, Brentu FC, Armengol J, Vicent A. 2017. Characterization of *Mycosphaerellaceae* species associated with citrus greasy spot in Panama and Spain. *PLoS One*. 12(12):e0189585. doi:10.1371/journal.pone.0189585.
- An Y-Y, Zeng X-Y, Geng K, Hyde KD, Wang Y, An Y-Y, Zeng X-Y, Geng K, Hyde K, Wang Y, et al. 2021. One new species and one new record of *Zasmidium* in China. *Biodiversity Data J*. 9:e59001. doi:10.3897/BDJ.9.e59001.
- Arango RE, Diaz-Trujillo C, Dhillon B, Aerts A, Carlier J, Crane CF, de Jong TV, de Vries I, Dietrich R, Farmer AD, et al. 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 Genet*. 12(8):e1005876. doi:10.1371/journal.pgen.1005876.
- Arzanlou M, Groenewald JZ, Fullerton RA, Abeln ECA, Carlier J, Zapater M-F, Buddenhagen IW, Viljoen A, Crous PW. 2008. Multiple gene genealogies and phenotypic characters differentiate several novel species of *Mycosphaerella* and related anamorphs on banana. *Persoonia*. 20(1):19–37. doi:10.3767/003158508X302212.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477. doi: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, et al. 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 Ecol Evol*. 4(10):914–919. doi:10.1111/2041-210X.12073.
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda)*. 10(4):1361–1374. doi:10.1534/g3.119.400908.
- Chang T-C, Salvucci A, Crous PW, 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 Genet.* 12(8):e1005904. doi:10.1371/journal.pgen.1005904.
- Chong P, Essoh JN, Arango Isaza RE, Keizer P, Stergiopoulos I, Seidl MF, Guzman M, Sandoval J, Verweij PE, Scalliet G, et al. 2021. A world-wide analysis of reduced sensitivity to DMI fungicides in the banana pathogen *Pseudocercospora fijiensis*. *Pest Manag Sci.* 77(7):3273–3288. doi:10.1002/ps.6372.
- Chong P, Vichou A-E, Schouten HJ, Meijer HJG, Arango Isaza RE, Kema GHJ. 2019. Pfcyp51 exclusively determines reduced sensitivity to 14 α -demethylase inhibitor fungicides in the banana black Sigatoka pathogen *Pseudocercospora fijiensis*. *PLoS One.* 14(10):e0223858. doi:10.1371/journal.pone.0223858.
- Cornet L, Baurain D. 2022. Contamination detection in genomic data: more is not enough. *Genome Biol.* 23(1):60. doi:10.1186/s13059-022-02619-9.
- Crous PW, Carlier J, Roussel V, Groenewald JZ. 2021. *Pseudocercospora* and allied genera associated with leaf spots of banana (*Musa* spp). *Fungal Syst Evol.* 7(1):1–19. doi:10.3114/fuse.2021.07.01.
- Davey JW, Davis SJ, Mottram JC, Ashton PD. 2020. Tapestry: Validate and edit small eukaryotic genome assemblies with long reads (p. 2020.04.24.059402). *bioRxiv.* <https://doi.org/10.1101/2020.04.24.059402>.
- Drenth A, Kema G. 2021. The vulnerability of bananas to globally emerging disease threats. *Phytopathology*. 111(12):2146–2161. doi:10.1094/phyto-07-20-0311-rvw.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Francois CM, Durand F, Figuet E, Galtier N. 2020. Prevalence and implications of contamination in public genomic resources: a case study of 43 reference arthropod assemblies. *G3 (Bethesda).* 10(2):721–730. doi:10.1534/g3.119.400758.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England).* 29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Han KS, Park MJ, Choi IY, Park JH, Shin HD. 2015. First report of leaf spot caused by *zasmidium lythri* on *lythrum salicaria* in Korea. *Plant Dis.* 99(5):728–728. doi:10.1094/PDIS-10-14-1039-PDN.
- Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, et al. 2020. 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. *Studies Mycol.* 96:141–153. doi:10.1016/j.simyco.2020.01.003.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066. doi:10.1093/nar/gkf436.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736. doi:10.1101/gr.215087.116.
- Kusch S, Vaghefi N, Kiss L. 2023. The good, the bad, and the misleading: how to improve the quality of ‘genome announcements’? *Mol Plant Microbe Interact.* 36:393–396. doi:10.1094/MPMI-01-23-0009-LE.
- Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49(W1):W293–W296. doi:10.1093/nar/gkab301.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-mem. <https://doi.org/10.6084/m9.figshare.963153.v1>.
- Lu J, Salzberg SL. 2018. Removing contaminants from databases of draft genomes. *PLoS Comput Biol.* 14(6):e1006277. doi:10.1371/journal.pcbi.1006277.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 27(6):764–770. doi:10.1093/bioinformatics/btr011.
- Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8(19):4321–4325. doi:10.1093/nar/8.19.4321.
- Osorio JA, Linnakoski R, Wingfield MJ, de Beer ZW, Roux J. 2021. A new species in the Mycosphaerellaceae from Cecidomyiidae leaf galls on *Avicennia marina* in South Africa. *Antonie van Leeuwenhoek.* 114(5):515–526. doi:10.1007/s10482-021-01537-3.
- Palmer J, Stajich J. 2019. nextgenusfs/funcannotate: Funcannotate v1.5.3 [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.2604804>.
- Rachtman E, Balaban M, Bafna V, Mirarab S. 2020. The impact of contaminants on the accuracy of genome skimming and the effectiveness of exclusion read filters. *Mol Ecol Resour.* 20(3):649–661. doi:10.1111/1755-0998.13135.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.
- Smit AFA, 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. doi:10.1093/bioinformatics/btu033.
- Steinberger M, Salzberg SL. 2020. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* 21(1):115. doi:10.1186/s13059-020-02023-1.
- Tribe HT, Thines E, Weber RWS. 2006. Moulds that should be better known: the wine cellar mould, *Racodium cellare*. *Mycologist.* 20(4):171–175. doi:10.1016/j.mycol.2006.09.016.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 43(1):11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5), 737–746. doi:10.1101/gr.214270.116.
- Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. Genomescope: fast reference-free genome profiling from short reads. *Bioinformatics.* 33(14):2202–2204. doi:10.1093/bioinformatics/btx153.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 9(11):e112963. doi:10.1371/journal.pone.0112963.
- Xu C, Zhang R, Sun G, Gleason ML. 2017. Comparative genome analysis reveals adaptation to the ectophytic lifestyle of sooty blotch and flyspeck fungi. *Genome Biol Evol.* 9(11):3137–3151. doi:10.1093/gbe/evx229.
- Zaccaron AZ, Stergiopoulos I. 2021. Characterization of the mitochondrial genomes of three powdery mildew pathogens reveals remarkable variation in size and nucleotide composition. *Microb Genom.* 7(12):000720. doi:10.1099/mgen.0.000720.