

# Long read sequencing and assembly of wild diploid relatives and cultivars in support of banana breeding programs

Andrew Chen<sup>1, †</sup>, Xuefeng Zhou<sup>2, †</sup>, Keith Decker<sup>2</sup>, Ericka Havecker<sup>2</sup>, Zijin Du<sup>2</sup>, Allan F. Brown<sup>3</sup>, Elizabeth A. B. Aitken<sup>1</sup>, Shah Trushar<sup>4</sup>, Brigitte Uwimana<sup>5</sup>, Chelly Hresko<sup>2</sup>, Rony Swennen<sup>5, 6, \*</sup>, Rebecca F. Lowdon<sup>2, \*</sup>

<sup>1</sup> School of Agriculture and Food Sustainability, The University of Queensland, Brisbane, QLD. 4067, Australia. A.C. [a.chen2@uq.edu.au](mailto:a.chen2@uq.edu.au); E.A.B.A. [e.aitken@uq.edu.au](mailto:e.aitken@uq.edu.au)

<sup>2</sup> Bayer Crop Science, Chesterfield, MO, USA. X.Z. [xuefeng.zhou@bayer.com](mailto:xuefeng.zhou@bayer.com), K.D. [keith.decker@bayer.com](mailto:keith.decker@bayer.com); E.H. [ericka.havecker@bayer.com](mailto:ericka.havecker@bayer.com); Z.D. [zijin.du@bayer.com](mailto:zijin.du@bayer.com); C.H. [chelly.hresko@bayer.com](mailto:chelly.hresko@bayer.com); R.F.L. [rebecca.lowdon@bayer.com](mailto:rebecca.lowdon@bayer.com)

<sup>3</sup> International Institute of Tropical Agriculture, Arusha P.O. Box 447, Tanzania. [A.Brown@cgiar.org](mailto:A.Brown@cgiar.org)

<sup>4</sup> International Institute of Tropical Agriculture, Nairobi P.O. Box 30709-00100, Kenya. S.T. [tm.shah@cgiar.org](mailto:tm.shah@cgiar.org)

<sup>5</sup> International Institute of Tropical Agriculture, Kampala P.O. Box 7878, Uganda. B.U. [B.Uwimana@cgiar.org](mailto:B.Uwimana@cgiar.org); R.S. [r.swennen@cgiar.org](mailto:r.swennen@cgiar.org)

<sup>6</sup> Division of Crop Biotechnics, Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, 3001 Leuven, Belgium.

<sup>†</sup> These authors contributed equally to this work

<sup>\*</sup> Authors to whom correspondence should be addressed.

## Abstract

Banana is an important fruit and staple crop, which is vital for food and income security in developing countries. A genomics initiative was undertaken as part of the Global Alliance against TR4 (Tropical Race 4 of *Fusarium Wilt of Banana*) Alliance, with Bayer Crop Science, International Institute of Tropical Agriculture, and University of Queensland to collectively accelerate the breeding efforts in Eastern Africa. The goal of this project is to develop genomics resources to improve banana breeding, productivity, and quality traits amid a changing climate and shifting disease pressures for breeding. Here, we present the PacBio HiFi sequencing, assembly, and comparative analysis of seven genomes of wild and edible diploid banana accessions belonging to subspecies considered important progenitors to cultivated banana varieties and part of breeding programs. This banana genome resource will help power genome editing, a pangenome effort, and conventional breeding programs for germplasm improvement.

Key words: Banana breeding, PacBio HiFi, chromosome-scale genome assembly, banana, *Musa acuminata*, *Fusarium oxysporum* f. sp. *cubense*, Tropical race 4, Fusarium wilt of Banana.

## Introduction

Bananas (*Musa* spp.) are an important horticultural crop that is served as either staple food or fruit, providing sustenance in the form of essential nutrients to millions of people around the world. In 2021, 12 million hectares were under banana cultivation, producing 170 million

tonnes of bananas (FAO, 2023). Approximately 26.4 million tons of bananas, 15.5% of total production, are exported each year making it the most exported fruit in the world (FAO, 2023). The export market is dominated by Cavendish varieties. Its primary producers are based in Latin America and the Caribbean, which account for 80% of the export market estimated at 11 billion USD (FAO, 2023). 84.5% of bananas are consumed locally and are grown by small holder farmers (Solidaridad, 2023).

Cultivated bananas including plantains, highland cooking bananas, and Cavendish bananas are mostly derived from hybridizations between *Musa acuminata* (A genome,  $2n = 22$ ) and *Musa balbisiana* (B genome,  $2n = 22$ ) (Rouard et al., 2018). Increasingly, genome assemblies shed light on the genome structure and functional divergence of diploid and polyploid *Musa* spp. (D'Hont et al., 2012; Rouard et al., 2018; Wang et al., 2019; Martin et al., 2020b; Belser et al., 2021; Huang et al., 2023; Liu et al., 2023; Li et al., 2024). However, breeding programs require sequence knowledge of specific breeding accessions to facilitate marker-assisted breeding and genomic selection. The public availability of genome assemblies of parental accessions important to breeding programs are currently lacking.

Like other clonally propagated tropical crops, bananas are bred using a few elite cultivars and thus are subject to high disease and pest pressures in the field (Jones, 2018), with significant outbreaks of diseases attributed to bacterial (*Xanthomonas*) wilt (Tushemereirwe et al., 2004; Biruma et al., 2007; Reeder et al., 2007; Smith et al., 2008; Mbaka et al., 2009; Tripathi et al., 2009; Carter et al., 2010), fungal pathogens such as *Pseudocercospora fijiensis* (Churchill, 2011; Arango Isaza et al., 2016; Alakonya et al., 2018) and *Fusarium oxysporum* f. sp. *cubense* (Rutherford, 2001; Ploetz, 2015; Dita et al., 2018) as the respective causal agents for black sigatoka and Fusarium wilt, other crippling disease infections caused by the banana bunchy top virus (Dale, 1987; Su et al., 2003; Niyongere et al., 2012; Kumar et al., 2015), root-knot nematode (Speijer et al., 1999; Gaidashova et al., 2009; Onkendi et al., 2014), and banana weevil borer (Rukazambuga et al., 1998; Gold, 2000; Kiggundu et al., 2003; Ocan et al., 2008). In particular, Fusarium wilt of banana (FWB) is one of the most devastating diseases affecting banana plants. The Tropical race 4 (TR4) of FWB caused by the soil-borne fungus *Fusarium oxysporum* f. sp. *cubense* (*Foc*) has put major constraints on global banana productions (Ordonez et al., 2015; van Westerhoven et al., 2022). The monoculture practice, the use of contaminated materials for planting and the lack of viable long-term solutions thereof, all contribute to the exacerbation of the diseases that are especially taxing on small holder farmers.

Advancing banana breeding by exploiting the genetic diversity of diploid breeding parents is key to improving polyploid hybrid bananas. For example, differential susceptibility to Fusarium wilt of banana (*Foc* TR4) has been described for key banana germplasm (Kung'u and Jeffries, 2001; Zuo et al., 2018; Arinaitwe et al., 2019; Ndayihanzamaso et al., 2020). Genetic variation in fruit quality traits, such as parthenocarpy or nutrient content, can only be accessed through trait introgression from diploid breeding material (Sardos et al., 2016a). Commercially, bananas are sold as triploids, but breeding triploid material is a complicated process that requires crossing fertile diploids to generate tetraploids, crossed again to diploids to produce secondary triploids (Brown et al. 2017; Batte et al. 2019). In addition, many large structural rearrangements have been characterized in banana breeding germplasm, which pose complications for introgressing key traits through recombination (Marin, et al., 2017; Němečková et al., 2018; Martin, et al., 2020a; Martin, et al., 2020b; Šimoníková et al., 2020). Genomics-guided breeding methods can help address the above challenges by enabling variant discovery, genetic mapping, and genomic selection to fast-track important traits (Uwimana et

al., 2021; Chen et al., 2023a; Chen et al., 2023b; Bayo et al.; 2024). Accessing genomics-guided germplasm improvement requires investment in foundational genomics resources for banana.

To facilitate efforts to improve diploid banana breeding in quality and resistances against diseases and pests, an international team led by Bayer Crop Science, IITA, and the University of Queensland, through the TR4 Alliance, has made a genomics initiative to sequence the genomes of a collection of diploid *Musa* spp. and cultivars important for IITA's breeding program in Africa. These include 'Fu Des' (ITC0939) and 'Manameg Red' (ITC0966) from *M. acuminata* ssp. *banksii*; 'Pahang' (ITC0609), AA cv. 'Rose' (ITC0712), and 'Pisang Lilin' (ITC1121) from *M. acuminata* ssp. *malaccensis*; 'Maia Oa' (ITC0728) from *M. acuminata* ssp. *zebrina*; 'Pisang Tongat' (ITC0063) from *M. acuminata* ssp. *errans*. The selected germplasm span four key *M. acuminata* subspecies and are potential progenitors of the important banana cultivars 'Matooke', 'Mchare', Cavendish and other plantains (Němečková et al., 2018; Sardos et al., 2022; Li et al., 2024).

In this study, we used PacBio high-fidelity reads to generate haplotype resolved reference genomes for these phenotypically and taxonomically diverse *Musa* diploids. With these assemblies, we aim to increase the diversity and overall number of genome assemblies of key diploid banana parent lines. The genome data presented here complement the existing banana genome resources with complete genome sequences of agronomically-improved diploid hybrid parents. Given the undeterred spreading of the TR4 pandemic, threatening the major banana producing regions of the world (Martínez et al., 2023), this genomic resource will allow us to dissect TR4 resistance in divergent genotypes. This resource will enable fast tracking of resistance in conventional breeding through genomic prediction as well as gene editing of resistance traits as well as traits relating to fruit quality in breeding lines derived from these progenitors.

## Results

### *Sequencing and assembly of seven high-quality banana genomes*

High-molecular weight genomic DNA was extracted from cigar leaves of the seven banana cultivars in this study (**Table 1; Figure 1**) and sequenced on a PacBio Sequel IIe producing high fidelity (HiFi) reads (16.6-29.4 Gigabases (Gb) of sequence for each germplasm). The depth of coverage for the haplotype-resolved assemblies were in the range of 17 to 27x per assembly (**Table 2**). The 14 partially phased haplotype assemblies comprised of 247 to 1539 contigs, with the total assembly sizes ranging from 493 – 586 Megabases (Mb) (**Table 3**). The size of the DH-Pahang genome was estimated by flow cytometry at 523 Mb (D'Hont et al., 2012) and the latest long read assembly had a cumulative size of 485 Mb (Belser et al., 2021). Our haplotype phased 'Pahang' assemblies were comparably sized at 499 and 492 MB whereas the rest of the haplotype phased diploid assemblies had cumulative sizes between 506 Mb and 586 Mb, falling in the range of 94 to 112 % of the estimated genome size of 'DH-Pahang'. All haplotype phased assemblies were comprised of 600 contigs or less with the number of N90 contigs ranging from 47 to 157. Only the assemblies for 'Pisang Tongat' and 'Fu Des' accessions had more contigs and were less contiguous than the rest of the collection. Furthermore, both 'Pisang Tongat' and 'Fu Des', as well as cv. 'Rose' and 'Pisang Lilin' had high heterozygosity rates at 1.47%, 1.32%, 1.80% and 1.50%, respectively (**Table 4**).

The HiFi reads were assembled with 'hifiasm' (Cheng et al., 2021), and produced two partially haplotype-resolved assemblies for each germplasm. Hi-C paired-end reads will be used in the future to correct haplotype switch and produce fully phased genomes for these accessions.

Assessing telomeric repeats at the ends of the assembled chromosomes showed that most chromosomes have telomeres placed on both the 5' and 3' ends, supporting the completeness of the assemblies (**Table 5**).

### *K-mer analysis to estimate the genome size and heterozygosity rate*

K-mer analysis was used to estimate the genome size using long read datasets (Ranallo-Benavidez et al., 2020). 'Genomescope' ver2.0 was used to estimate heterozygosity while a custom analysis on k-mer frequency and counts was used to estimate genome size of each assembly (Methods).

The estimated genome size of each assembly was in the range of 523 to 628 Mb, which was larger than the actual sizes of these assemblies (except for AA cv. 'Rose'; **Table 4**). The relative heights of the peaks in a k-mer spectrum provided an estimate to the heterozygosity level of each accession (Ranallo-Benavidez et al., 2020) (Methods). The diploid assemblies had a heterozygosity range of 0.42-1.80%, which was in the range of expected heterozygosity for the sequenced material.

### *Gene model annotations*

Repeat-masking detected regions occupying 48.9% to 58.6% of the genome sequence that contain repetitive elements across all haplotype resolved assemblies. Out of these regions characterised, long terminal repeat elements occupied the majority at 39% to 51.3% of the genome sequence, whereas DNA elements and simple repeats respectively occupied 5.1% to

10.5% and 1% to 1.3% of the genome sequence. These results are comparable to the repetitive elements annotated totalling 52.6% of the genome in DH-Pahang (Belser et al., 2021).

'GeMoMa', a homology-based gene annotation program (Keilwagen et al., 2019), was used to annotate the haplotype resolved assemblies using gene models in five other *Musa* reference genomes. Firstly, reference genomes and annotations for *M. acuminata* Pahang v4, *M. beccarii* v1, *M. textilis* (abaca) v1, *M. schizocarpa* v1, *M. troglodytarum* v1, *M. balbisiana* DH-PKW v1, *M. itinerans* v1, *M. acuminata* spp. *banksii* v2, *M. acuminata* ssp. *burmannica* v1 and *M. acuminata* ssp. *zebrina* v1 were retrieved from Banana Genome Hub (BGH) (<https://banana-genome-hub.southgreen.fr/> accessed on 19 April 2024) and were subject to protein BUSCO analysis (Manni et al., 2021). Out of these references, five reference genomes showed a protein based BUSCO higher than 90%, including *M. acuminata* Pahang v4 (97.1%), *M. beccarii* v1 (94.8%), *M. textilis* (abaca) v1 (92.7%), *M. schizocarpa* v1 (91.6%), *M. troglodytarum* v1 (90%). These genomes were then used as reference species for 'GoMoMa' since its prediction quality depends on the quality of the target genome sequence (Keilwagen et al., 2019).

Transcript sets from these reference genomes were retrieved from BGH (<https://banana-genome-hub.southgreen.fr/> accessed on 19 April 2024) and were used to predict gene models for each of the 14 haplotype resolved assemblies using 'GeMoMa'. This resulted in the annotation of approximately 38,000 gene models in each new genome (**Table 6**). The number of gene models was higher than 36,979 genes annotated in DH-Pahang v4 (Belser et al., 2021). Despite not having RNA-seq reads from each accession to provide evidence in building transcript models, the BUSCO scores for the predicted amino acid sequences generated by 'GeMoMa' annotations showed 98.0-98.9% genome completeness for each of the 14 haplotype resolved genomes, in comparison to a 96.0% genome completeness assessed for the DH-Pahang v4 assembly.

Finally, we assessed the abundance of disease resistance genes particularly those encoding nucleotide-binding and leucine-rich repeat (NLR) proteins in our haplotype resolved assemblies. These immune receptors play important roles in plant surveillance of invading pathogens during infection and intracellularly triggers an immune response (Jones and Dangl, 2006). An NLR prediction software, namely the 'NLR-annotator' was used to annotate loci associated with protein motifs usually associated with NLRs (Steuernagel et al., 2020). We identified 105 to 165 loci containing NLR motifs in the haplotype resolved assemblies (**Table 6**). Variations in the number of NLR loci were evident in pairs of some of the haplotype resolved genomes, including 'Pisang Tongat', which had 112 NLR loci in one and 161 in the other haplotype resolved genome (**Table 6**).

### *Gene synteny analysis and detection of rearrangements*

Gene order and orthology amongst all haplotype resolved genomes were examined at the chromosome level using the software 'GENESPACE' (Lovell et al., 2022). The syntenic relationships were visualized based on the physical position of all genes across all haplotype phased genomes. Overall, conservation of gene content and order was evident amongst these assemblies at the chromosome level, and some local rearrangements and inversions were also detected (Figure 2).

There was no structural discordance indicative of incorrectly positioned contigs at the chromosome level. Further sequencing to detect long-range interactions (particularly Hi-C or

ultra long reads) will support scaffolding potentially address some of the finer discrepancies, particularly with respect to some of the well-known translocation events (Baurens et al., 2018; Dupouy et al., 2019; Martin et al., 2020a). A specific reciprocal translocation involving chromosome 1 and chromosome 4 was previously reported and characterised in 'Pisang Lilin' (Martin et al., 2017; Šimoníková et al., 2020) and is detected in the initial haplotype assemblies for 'Pisang Lilin' presented here.

## Discussion

In this study, we produced preliminary genome drafts of seven diploid (AA) banana accessions. Unlike the B genome progenitor *Musa balbisiana* which has its genome size estimated at 534-540 Mb using the 2C nuclear DNA content (Lysak et al., 1999), the AA genome appeared to be relatively larger compared to *M. balbisiana*, with wild relatives and cultivated diploids estimated at 591-615 Mb (Lysak et al., 1999), 580-640 Mb (Kamaté et al., 2001), and 523 Mb (D'Hont et al., 2012). 'Pahang' is the parent to 'DH-Pahang' or 'CIRAD930' and these have been shown to be genetically similar accessions in genotyping studies (Christelová et al., 2017; Ruas et al., 2017). Our haplotype resolved 'Pahang' assemblies are 492 and 499 Mb in size, and has a predicted size of 523 Mb, both of which are consistent with the characterisation of 'DH-Pahang' in previous studies. The other two *M. acuminata* ssp. *malaccensis* derived diploids, AA cv. 'Rose' and 'Pisang Lilin', had haplotype resolved assembly sizes of 544 and 566 Mb, and 506 and 540 Mb, respectively, which are larger than 'DH-Pahang' but smaller than the size of other AA diploids of *M. acuminata* determined by flow cytometry in previous studies (Lysak et al., 1999; Kamaté et al., 2001).

The two *M. acuminata* ssp. *banksii* derivatives, 'Fu Des' and 'Manameg Red' had haplotype resolved genome sizes of 544 and 586 Mb, and 523 and 538 Mb, respectively, which are considerably smaller than the estimated genome sizes of three distinct types of the wild species of *M. acuminata* ssp. *banksii* at 580-630 Mb (Kamaté et al., 2001). *M. acuminata* ssp. *zebrina* 'Maia Oa' produced 531 and 536 Mb haplotype resolved assemblies, which are shorter than the Illumina short read-based assembly of the same line at 623 Mb (Rouard et al., 2018) but comparable to an Oxford Nanopore Technology (ONT) long read assembly of *M. acuminata* ssp. *zebrina* v2.0, totalling 558 Mb in cumulative size (Li et al., 2024). Finally, *M. acuminata* ssp. *errans* derived 'Pisang Tongat' had haplotype resolved genome sizes of 512 and 554 Mb, which is smaller than the estimated size of *M. acuminata* ssp. *errans* at 606 Mb (Doležel et al., 1994). Additional sequencing data is expected to improve the phasing of these assemblies, which may alter the overall size of the haplotype resolved assemblies.

The cost of DNA sequencing has declined in the era of third generation long-read sequencing technologies, facilitating genome assemblies for large plant polyploid genomes (Sahu and Liu, 2023). However, technical hurdles still exist in producing fully phased plant genomes (Delorean et al., 2023). When an accession exhibits a high level of heterozygosity such as the ones presented in this study, the chance that the resulting assemblies contain chimeric regions of the individual's two haplotypes is high. These chimeric regions may provide misleading information in downstream analyses in gene discovery. Bionano optical maps, Chromatin conformation capture or Hi-C sequencing, ultra-long DNA sequencing reads, and genetic maps have been used as physical or genetic information to determine which assembly contigs or haplotigs group together to form a chromosome (Sedlazeck et al., 2018; Belser et al., 2021; Cochetel et al., 2023; Liu et al., 2023). Hi-C has been widely used to aid chromosome scale scaffolding in haplotype resolved plant genome assemblies (Šimková et al., 2024). In *Musa* spp., the latest genomes that have been fully haplotype-resolved have seen an increasing use

of Hi-C as the scaffolding tool to resolve haplotypes in diploid and triploid cultivars (Huang et al., 2023; Liu et al., 2023; Li et al., 2024). As evident in our assemblies, some of the translocation events will need to be distinguished from assembly errors by performing Hi-C or other long-range sequencing validation in our assemblies.

One of the haplotype phased assemblies in 'Pisang Lilin' and AA cv. 'Rose' appeared to carry reciprocal translocations involving chromosome 1 and 4. These translocations were previously reported in 'Pisang Lilin' (Martin et al., 2017; Šimoníková et al., 2020). The same translocations were also confirmed in 'Cavendish', 'Gros Michel' and 'Mchare' cultivars (Šimoníková et al., 2020). The precise location and the structure of the translocated fragments requires additional analysis to resolve in our assemblies.

In banana breeding, efforts have been made in the last decades to introgress traits relating to disease resistance in Africa (Lorenzen et al., 2010). This has been hindered by the long generation times of growth, fertility issues with the polyploids, and limited genetic diversity captured by the core *Musa* collection in breeding (Ortiz and Swennen, 2014). Often, banana breeding pipelines relied on diploids to increase genetic diversity of East African highland bananas through recurrent selection of important traits. Thus, improving the quality of diploids is an important activity for breeding. Here, we selected a set of diploids important in the IITA breeding program for sequencing. These diploids appear to show genetic diversity in whole genome diversity analyses (Sardos et al., 2016b; Christelová et al., 2017).

These assemblies when phased and structurally validated, will support banana breeding and trait discovery research programs. They will also serve as a key step in broadening genome sequence information for this important crop. Next steps for this project include generating more long-range interaction data to fully phase these assemblies and support the translocation events detected so far. Fully phased assemblies will also enable a closer examination of the TR4 resistance loci (Ahmad et al., 2020; Chen et al., 2023a; Chen et al., 2023b) and other key trait genes (Li et al., 2024).

## Materials and Methods

### *Source Material*

The seven *Musa* accessions sequenced in this study were as listed in Table 1. Lyophilized young leaves were shipped to Bayer Crop Science Precision Genomics Unit (St. Louis, MO, USA) for DNA extraction, library preparation and Pacific Biosciences HiFi sequencing.

### *Extraction, Library Preparation & Sequencing*

All lyophilized tissue was ground in liquid nitrogen to fine powder. To approximately 1 gram of leaf powder, 20 mL of pre-heated (65°C) modified Cetyltrimethylammonium bromide (CTAB) extraction buffer [2% CTAB (H-5882, Merck, Darmstadt, Germany); 100 mM Tris-HCl (pH8.0); 20 mM Ethylenediaminetetraacetic acid (EDTA, pH8.0); 1.4 M Sodium Chloride, freshly added 1% Polyvinylpyrrolidone and 2% (v/v) Tween-20] was added and vortexed gently until complete resuspension. The resuspended samples were incubated at 65°C for 45 minutes with gentle mixing every 10 minutes. After incubation, 20 µl of Proteinase K (P8107S, New England Biolabs, Ipswich, MA, USA) was added and incubated at 60°C for 15 minutes with gentle mixing every 5 minutes. DNA was purified using equal volumes of Phenol Chloroform Isoamyl (24:24:1) followed by Chloroform:Isoamyl alcohol extraction (24:1). DNA was purified by precipitation of aqueous phase by equal volume isopropanol and centrifugation.

Integrity and concentration of isolated genomic DNA was measured by Qubit (Thermo Fisher, Waltham, MA, USA), Nanodrop (Thermo Fisher, Waltham, MA, USA) and Femto Pulse system (Agilent, Santa Clara, CA, USA). 10 µg of resuspended DNA was subjected to a bead cleanup step (Genfind v3.0, Beckman Coulter, Brea, CA, USA) and 5 µg of DNA were used for HiFi library preparation. All HiFi libraries were prepared using Pacific Biosciences SMRTbell® Express Template Prep Kit 3.0 and sequenced on the Pacific Biosciences Sequel IIe system to generate Circular Consensus (CCS) reads.

### *K-mer analysis*

For each of the seven banana accessions, a k-mer analysis was performed on the HiFi reads with 'Jellyfish' v.2.2.4 (<https://github.com/gmarcais/Jellyfish>; (Marcais and Kingsford, 2011)) using a k-mer size of 55 and a histogram of the k-mer frequency was generated. The computation was run on an AWS EC2 r5.12xlarge node (48 vCPUs and 380Gb RAM). The output of jellyfish 'histo' program is a data set with counts of k-mers at given frequencies; this table was then used to estimate genome sizes for each accession. The genome size for a given assembly was estimated as the product of the total number of k-mers times the k-mer size (55 in this case), divided by the frequency mode from k-mer distribution. The same k-mer frequency table was used as the input for 'Genomescope' heterozygosity rate estimation.

### *Assembly with HiFi reads and pseudomolecule scaffolding*

All seven banana genomes were sequenced on a Pacific Biosciences Sequel IIe system. The sequencing adaptors and multiplexing barcodes were removed with PacBio SMRTLink version 11.1 after the demultiplexing was finished.



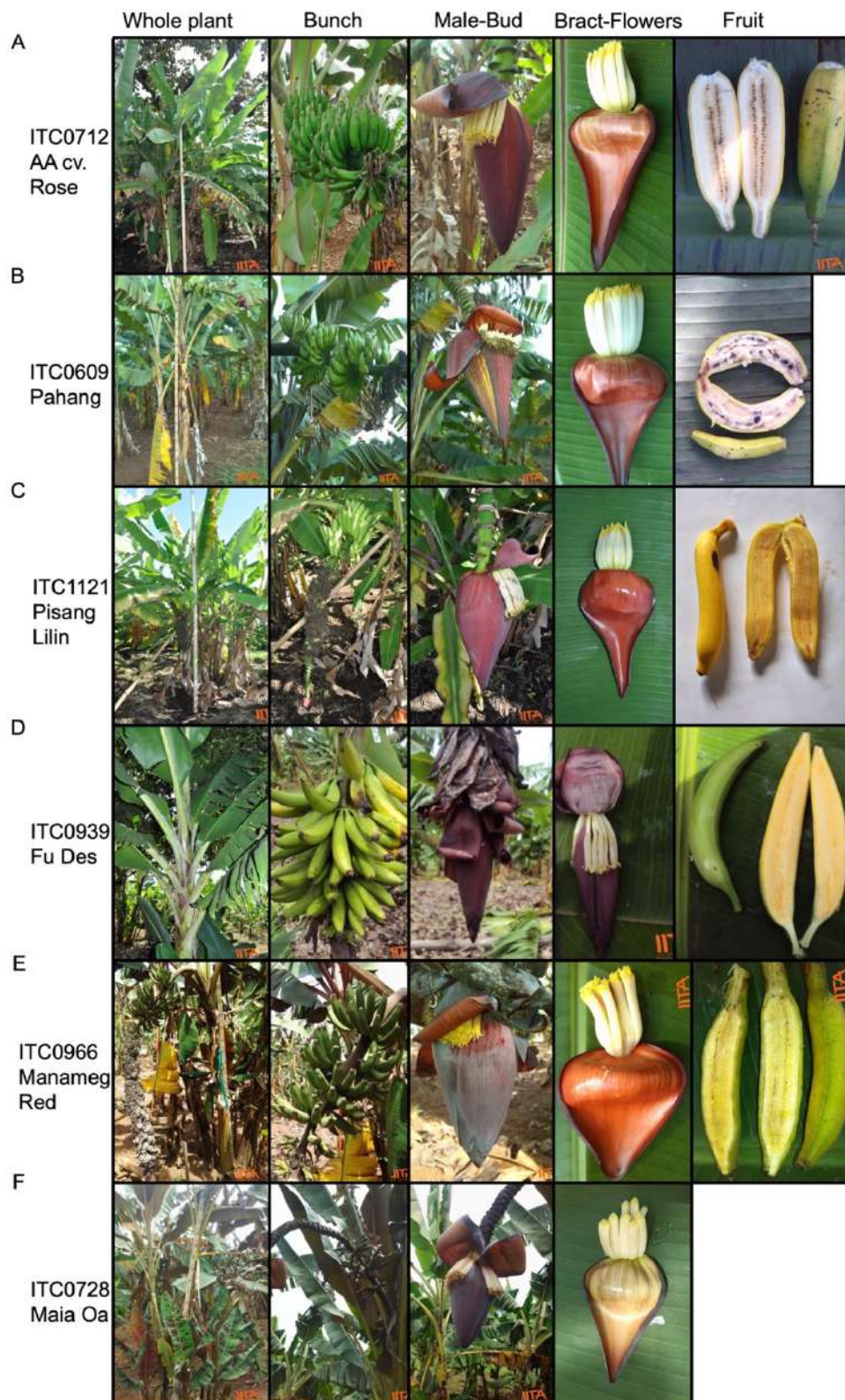
Each of these seven genomes was assembled using 'hifiasm' v. 0.18.2-r467 (Cheng et al., 2022). The assembly process of each genome was run on an AWS EC2 r5.12xlarge node (48 vCPUs and 380Gb RAM). The version 0.18.2 of 'hifiasm' produces two sets of partially phased haploid contigs by default. The heterozygosity rates estimated by Genomescope2.0 (Ranallo-Benavidez, 2020) were predicted to be lower than 0.5% in 'Pahang', 'Manameg Red' and 'Maia Oa', therefore the default parameter of 0.55 for setting -s for 'hifiasm' was used. The estimated heterozygosity rates of AA cv. 'Rose', 'Pisang Lilin', 'Fu Des', and 'Pisang Tongat' were all larger than 1.3%; therefore, the option -s for 'hifiasm' was set to 0.5, instead of 0.55, to avoid unbalanced sizes of partially phased haploid genomes.

To scaffold contigs into pseudomolecules, a reference-based scaffolding method, 'RagTag' (v2.1.0) (Alonge *et al.*, 2022), was adopted. The 'scaffold' tool was used to align the 'hifiasm' contigs with the DH-Pahang v4 reference (Belser *et al.*, 2021), and subsequently clustered contigs together into groups using the corresponding physical information from the 11 *Musa acuminata* chromosomes and mitochondrial genome of DH-Pahang v4. Then the contigs of each chromosome group were ordered based on their alignment positions and joined to form continuous pseudomolecules. Unscaffolded contigs for each haplotype assembly were concatenated onto chr10000001 with 100 "N" bases inserted between subsequent contigs.

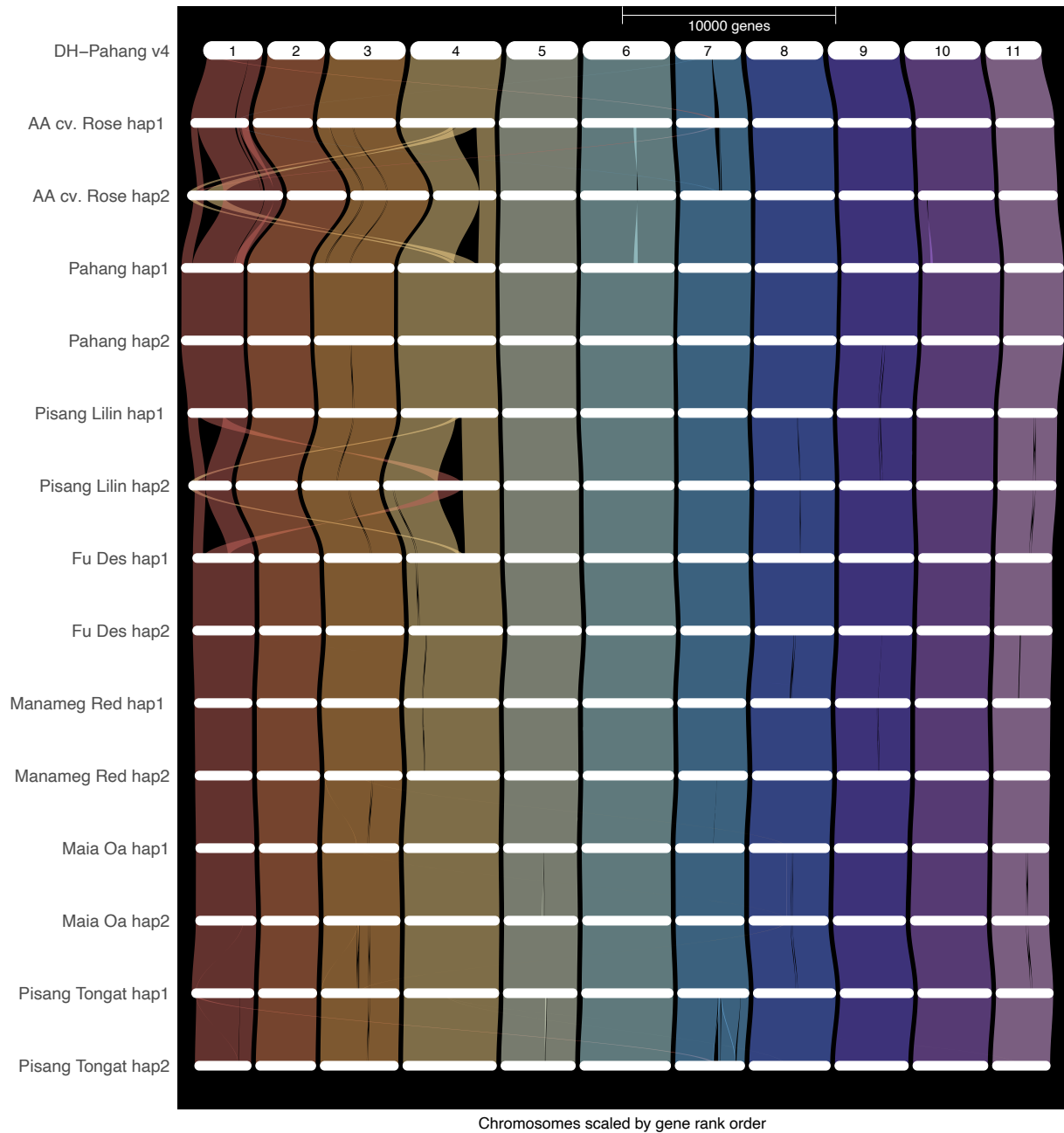
### ***Gene model annotation***

Non-redundant transposable element libraries were created for each genome using 'EDTA'(v2.0.0, --sensitive 0) (Ou et al., 2019). These libraries were then passed to 'RepeatMasker' (version open-4.0.9) (Smit et al., 2013) for repeat masking. 'GeMoMa' (v1.9) (Keilwagen et al., 2018) was used to map gene models from five public genome datasets: *Musa\_acuminata\_pahang\_v4*, *Musa\_beccarii\_1.0*, *Musa\_textilis\_abaca\_1.0*, *Musa\_schizocarpa\_1.0*, and *Musa\_troglodytarum* (GFF3 and fasta files were downloaded from the Banana Genome Hub (<https://banana-genome-hub.southgreen.fr/>, accessed in March 2024)). 'GeMoMa' annotation filter parameters were chosen to limit less conserved gene predictions (NaN(score) or score/aa $\geq$ 4). BUSCO scores were calculated using the BUSCO tool suite version 5.5.0 with the database 'liliopsida\_odb10' (creation date 2017-12-01) (Manni et al., 2021). The 'NLR-Annotator' (v2.1, default settings) was used to identify putative NLR loci (Stuernagel et al., 2020). 'GENESPACE' (v1.3.1, default settings) was used for gene synteny analysis (Lovell et al., 2022).

## Figures



**Figure 1.** Phenotypic characteristics of six accessions sequenced in this study (Pisang Tongat not shown). Images downloaded from <https://www.musabase.org> in February 2024.



**Figure 2.** GENESPACE plot of the 14 partially phased assemblies reported here and DH-Pahang v4.

## Tables

**Table 1.** The seven accessions sequenced in this study and their accession numbers, taxonomic classification, geographical location where these lines are likely originated from. "Edible" means the fruit is parthenocarpic; "wild" means the fruit only develop after pollination and will have seeds.

Accession number	Accession name	<i>Musa acuminata</i> subspecies	Geographical range <sup>1</sup>	Edibility
ITC0712	AA cv. Rose	<i>malaccensis</i>	Malaysia <sup>2</sup>	Edible
ITC0609	Pahang	<i>malaccensis</i>	Malaysia: Peninsular <sup>3</sup>	Wild
ITC1121	Pisang Lilin	<i>malaccensis</i>	Malaysia, Indonesia <sup>4</sup>	Edible
ITC0939	Fu Des	<i>banksii</i>	Papua New Guinea. Australia: Queensland. Samoa <sup>2</sup>	Edible
ITC0966	Manameg Red	<i>banksii</i>	Papua New Guinea. Australia: Queensland. Samoa <sup>5</sup>	Edible
ITC0728	Maia Oa	<i>zebrina</i>	Java/Indonesia <sup>6</sup> and Hawaii <sup>7</sup>	Wild
ITC0063	Pisang Tongat	<i>errans</i>	Philippines <sup>8</sup>	Some pulp

<sup>1</sup> Based on a compilation of published taxa (Sachter-Smith, 2023).

<sup>2</sup> A derivative of *Musa acuminata* ssp. *malaccensis*. AA cv. 'Rose' and its somaclone 'Gu Nin Chiao', also known as IDN110, are frequent genitors in world breeding programmes due to the resistance and fruit quality traits they carry (Bakry et al., 2021).

<sup>3</sup> Taxa data provided by Simmonds N.W., 1957. Unique traits include 'Bracts red, non-imbricate' (Sachter-Smith, 2023).

<sup>4</sup> A derivative of *Musa acuminata* ssp. *malaccensis*. 'Pisang Lilin' from the Pisang Lilin subgroup of diploids with AA genomes (Ploetz et al., 2007).

<sup>5</sup> Taxa data provided by Mueller F.J.H. von, 1864. Unique traits include 'Pendent bunch, bend in rachis, large brown blotches on pseudostem' (Sachter-Smith, 2023).

<sup>6</sup> Originated from Indonesia: Java (Nasution, 1991) with unique traits such as 'leaves with red-brown blotches, fruits not hairy' (Sachter-Smith, 2023).

<sup>7</sup> In Hawai'i, this type of seeded variety was introduced as a medicinal plant and plants with these phenotypic descriptive is called 'Maia Oa' (Ploetz et al., 2007).

<sup>8</sup> Taxa data provided by Valmayor R.V., 2001. A unique trait includes 'Bracts green, large bunches up to 26 hands' (Sachter-Smith, 2023).

**Table 2.** Statistics of the HiFi libraries.

Accession name	Total number of reads	Total bases	Average read length	N50 read length	Calculated depth of coverage of diploid genome <sup>1</sup>
AA cv. Rose	1,913,537	29,271,776,776	15,297	16,504	26
Pahang	1,060,907	16,580,962,367	15,629	16,661	17
Pisang Lilin	1,700,965	22,758,223,451	13,379	14,182	21
Fu Des	2,285,116	19,738,030,518	8,608	8,583	17
Manameg Red	1,448,673	17,945,069,754	12,387	12,804	17
Maia Oa	1,983,197	29,413,041,262	14,831	15,876	27
Pisang Tongat	2,508,435	25,685,496,835	10,239	10,230	24

<sup>1</sup> Total bases of reads divided by sum of total bases of both haplotype phased assemblies.

**Table 3.** Genome assembly quality statistics. The assembly statistics for each haplotype assembly are indicated as hap1 | hap2 in each cell.

Accession name	Number of contigs	Total bases	N50	Number of N50 contigs	N90	Number of N90 contigs
AA cv. Rose	430   254	566Mb   544Mb	10.46Mb   9.66Mb	17   20	2.41Mb   1.84Mb	57   64
Pahang	541   247	499Mb   493Mb	12.76Mb   11.66Mb	13   13	1.62Mb   2.05Mb	50   47
Pisang Lilin	625   311	540Mb   506Mb	7.36Mb   8.33Mb	24   19	1.44Mb   1.87Mb	91   70
Fu Des	1,539   1,020	586Mb   544Mb	1.60Mb   1.79Mb	91   71	0.28Mb   0.30Mb	411   357
Manameg Red	829   485	539Mb   523Mb	5.14Mb   3.74M	31   42	0.80Mb   0.82Mb	136   154
Maia Oa	603   401	537Mb   531Mb	5.01Mb   4.05Mb	33   36	0.69Mb   0.76Mb	154   157
Pisang Tongat	1,465   569	554Mb   512Mb	3.79Mb   4.08Mb	41   34	0.49Mb   0.77Mb	184   148

**Table 4.** The estimated and assembled genome sizes. Genome size estimates derived from k-mer analysis, as in Methods. The assembly statistics for each haplotype assembly are indicated as hap1 | hap2 in each cell where relevant. The heterozygosity rate was calculated using 'GenomeScope2.0' (Ranallo-Benavidez et al., 2020).

<b>Accession name</b>	<b>Estimated haploid genome (bp)</b>	<b>Assembly sizes (bp)</b>	<b>% difference to the estimated size</b>	<b>Heterozygosity (%)</b>
AA cv. Rose	540,120,967	566,453,526   544,298,584	4.88   0.77	1.80
Pahang	523,216,062	499,014,147   492,921,427	-4.63   -5.79	0.49
Pisang Lilin	596,438,694	540,429,291   506,266,108	-9.39   -15.12	1.50
Fu Des	594,313,610	586,153,264   544,383,587	-1.37   -8.40	1.32
Manameg Red	628,597,152	538,694,546   523,470,394	-14.30   -16.72	0.45
Maia Oa	574,587,557	536,922,034   531,126,841	-6.56   -7.56	0.42
Pisang Tongat	594,128,672	554,270,387   512,271,382	-6.71   -13.78	1.47

**Table 5.** Summary of telomeres detected in *de novo* partially phased genome assemblies for each cultivar. The telomeres detected for each haplotype assembly are indicated as hap1 | hap2 in each cell. In haplotype assemblies where both telomeres were not found, the result is indicated in bold text.

	AA cv. Rose	Pahang	Pisang Lilin	Fu Des	Manameg Red	Maia Oa	Pisang Tongat
<i>chr1</i>	<b>3'   3'</b>	<b>5'   5'</b>	both   <b>3'</b>	both   both	<b>5'</b>   both	both   both	both   <b>3'</b>
<i>chr2</i>	both   both	both   both	both   both	both   both	both   both	both   both	both   both
<i>chr3</i>	<b>5'   none</b>	both   both	both   both	both   both	<b>5'   5'</b>	<b>3'   none</b>	both   <b>5'</b>
<i>chr4</i>	both   both	both   both	<b>3'</b>   both	both   both	both   both	both   both	both   both
<i>chr5</i>	<b>5'</b>   both	both   <b>5'</b>	both   both	both   both	<b>3'   3'</b>	both   both	both   both
<i>chr6</i>	<b>3'   3'</b>	both   both	<b>5'</b>   both	both   <b>3'</b>	both   <b>3'</b>	<b>5'</b>   both	<b>3'   5'</b>
<i>chr7</i>	both   both	both   both	both   both	<b>5'</b>   both	both   both	both   both	both   both
<i>chr8</i>	both   both	both   both	both   <b>3'</b>	both   both	both   <b>none</b>	both   both	both   both
<i>chr9</i>	both   both	both   both	both   both	both   <b>5'</b>	both   both	<b>5'</b>   both	<b>3'   3'</b>
<i>chr10</i>	both   <b>5'</b>	<b>5'</b>   both	<b>5'</b>   both	both   <b>5'</b>	<b>5'   3'</b>	both   both	<b>5'</b>   both
<i>chr11</i>	both   both	both   both	<b>3'</b>   both	<b>3'   3'</b>	<b>5'   3'</b>	both   <b>none</b>	both   both

**Table 6.** Counts of gene model and transcript annotations and BUSCO Scores for each assembly obtained from 'GeMoMa' (see Methods).

Accession name	Haplotype	Annotated feature count (scaffolded contigs only)			BUSCO calculated based on protein sequence				
		gene	mRNA	putative NLRs	Complete	Single copy	Multi-copy	Fragmented	Missing
AA cv. Rose	hap1	38625	45180	142	98.1	78.8	19.3	0.7	1.2
AA cv. Rose	hap2	39046	45605	113	98.3	77.8	20.5	0.7	1.0
Pahang	hap1	39673	46302	144	98.8	80.7	18.1	0.6	0.6
Pahang	hap2	39683	46326	129	98.8	80.7	18.1	0.6	0.6
Pisang Lilin	hap1	38987	45610	136	98.7	79.7	19.0	0.7	0.6
Pisang Lilin	hap2	38884	45514	137	98.7	80.4	18.3	0.6	0.7
Fu Des	hap1	38307	44842	138	98.2	78.8	19.4	0.8	1.0
Fu Des	hap2	38292	44872	142	98.2	79.3	18.9	0.8	1.0
Manameg Red	hap1	38037	44625	114	98.5	79.8	18.7	0.8	0.7
Manameg Red	hap2	37929	44457	105	98.5	80.3	18.2	0.8	0.7
Maia Oa	hap1	37991	44518	162	98.5	80.3	18.2	0.9	0.6
Maia Oa	hap2	38018	44575	165	98.3	80.2	18.1	1.0	0.7
Pisang Tongat	hap1	38470	45077	112	98.6	80.3	18.3	0.7	0.7
Pisang Tongat	hap2	38104	44655	161	98.0	79.7	18.3	0.8	1.2

## Funding

Andrew Chen, Allan F. Brown, Elizabeth A. B. Aitken, Shah Trushar, Brigitte Uwimana, and Rony Swennen were supported by the Bill and Melinda Gates Foundation through its grant to the International Institute of Tropical Agriculture (IITA) under the project Accelerated Breeding of Better Bananas, grant number IITA 20600.15/0008-8—Phase II. Research funding support for Andrew Chen and Elizabeth Aitken was also provided by Hort Innovation Australia (BA 17006).



## References

- Ahmad, F., Martawi, N.M., Poerba, Y.S., De Jong, H., Schouten, H., and Kema, G.H.J. (2020). Genetic mapping of Fusarium wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. *Theor Appl Genet* 133, 3409-3418. doi:10.1007/s00122-020-03677-y
- Alakonya, A., Kimunye, J., Mahuku, G., Amah, D., Uwimana, B., Brown, A., and Swennen, R. (2018). Progress in understanding Pseudocercospora banana pathogens and the development of resistant *Musa* germplasm. *Plant Pathol* 67, 759-770.
- Arango Isaza, R.E., Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C.F., T, V.D.J., De Vries, I., Dietrich, R., Farmer, A.D., Fortes Fereira, C., Garcia, S., Guzman, M., Hamelin, R.C., Lindquist, E.A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., Reynolds, E., Scalliet, G., Souza, M., Jr., Stergiopoulos, I., Van Der Lee, T.A., De Wit, P.J., Zapater, M.F., Zwiers, L.H., Grigoriev, I.V., Goodwin, S.B., and Kema, G.H. (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, e1005876. doi:10.1371/journal.pgen.1005876
- Arinaitwe, I.K., Teo, C.H., Kayat, F., Tumuhimbise, R., Uwimana, B., Kubiriba, J., Swennen, R., Harikrishna, J.A., and Othman, R.Y. (2019). Evaluation of banana germplasm and genetic analysis of an F1 population for resistance to *Fusarium oxysporum* f. sp. *cubense* race 1. *Euphytica* 215, 1-11. doi:10.1007/s10681-019-2493-3
- Bakry F., Horry J.P., Jenny C. 2021. Making banana breeding more effective. In : Achieving sustainable cultivation of bananas. Volume 2 - Germplasm and genetic improvement. Kema Gert H.J. (ed.), Drenth André (ed.). Burleigh Dodds Series in Agricultural Science, 86. Burleigh Dodds Science Publishing, Cambridge. pp. 217-256. doi:10.19103/AS.2020.0070.12
- Batte, M., Swennen, R., Uwimana, B., Akech, V., Brown, A., Tumuhimbise, R., Hovmalm, H.P., Geleta, M., and Ortiz, R. (2019). Crossbreeding East African highland bananas: lessons learnt relevant to the botany of the crop after 21 years of genetic enhancement. *Frontiers in Plant Sci* 10, 421463. doi:10.3389/fpls.2019.00081
- Baurens, F.-C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., Ricci, S., Rouard, M., Habas, R., Lemainque, A., Yahiaoui, N., and D'hont, A. (2018). Recombination and large structural variations shape interspecific edible bananas genomes. *Mol Biol Evol* 36, 97-111. doi:10.1093/molbev/msy199
- Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., Lemainque, A., Wincker, P., D'hont, A., and Aury, J.-M. (2021). Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Commun Biol* 4, 1047. doi:10.1038/s42003-021-02559-3
- Biruma, M., Pillay, M., Tripathi, L., Blomme, G., Abele, S., Mwangi, M., Bandyopadhyay, R., Muchunguzi, P., Kassim, S., Nyine, M., Laban, T., and Eden-Green, S. (2007). Banana Xanthomonas wilt: A review of the disease, management strategies and future research directions. *Afr J Biotech* 6, 86.
- Brown, A. et al. (2017). Bananas and Plantains (*Musa* spp.). In: Genetic Improvement of Tropical Crops. Springer, Cham. doi: 10.1007/978-3-319-59819-2\_7
- Carter, B.A., Reeder, R., Mgenzi, S., Mwangi, M., Aritua, V., Smith, J., Beed, F.D., Kinyua, Z., Nakato, V., and Miller, S. (2010). Identification of *Xanthomonas vasicola* (formerly *X. campestris* pv. *musacearum*), causative organism of banana xanthomonas wilt, in Tanzania, Kenya and Burundi. In: *Plant Pathol* 59, 2, 403-403. Wiley. doi:10.1111/j.1365-3059.2009.02124.x

- Chen, A., Sun, J., Martin, G., Gray, L.-A., Hřibová, E., Christelová, P., Yahiaoui, N., Rounsley, S., Lyons, R., Batley, J. et al. (2023a). Identification of a major QTL-controlling resistance to the subtropical race 4 of *Fusarium oxysporum* f. sp. *cubense* in *Musa acuminata* ssp. *malaccensis*. *Pathogens* 12, 289. doi:10.3390/pathogens12020289
- Chen, A., Sun, J., Viljoen, A., Mostert, D., Xie, Y., Mangila, L., Bothma, S., Lyons, R., Hřibová, E., and Christelová, P. et al. (2023b). Genetic mapping, candidate gene identification and marker validation for host plant resistance to the race 4 of *Fusarium oxysporum* f. sp. *cubense* using *Musa acuminata* ssp. *malaccensis*. *Pathogens* 12, 820. doi:10.3390/pathogens12060820
- Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat methods* 18, 170-175. doi:10.1038/s41592-020-01056-5
- Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., Van Den Houwe, I., Sutanto, A., Kepler, A.K., Swennen, R., Roux, N., and Doležel, J. (2017). Molecular and cytological characterization of the global *Musa* germplasm collection provides insights into the treasure of banana diversity. *Biodivers Conserv* 26, 801-824. doi:10.1007/s10531-016-1273-9
- Churchill, A.C. (2011). *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Mol plant pathol* 12, 307-328. doi.org:10.1111/j.1364-3703.2010.00672.x
- Cochetel, N., Minio, A., Guarracino, A., Garcia, J.F., Figueroa-Balderas, R., Massonnet, M., Kasuga, T., Londo, J.P., Garrison, E., Gaut, B.S., and Cantu, D. (2023). A super-pangenome of the North American wild grape species. *Genome Biol* 24, 290. doi:10.1186/s13059-023-03133-2
- D'hont, A., Denoeud, F., Aury, J.-M., Baurens, F.-C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M. et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213-217. doi:10.1038/nature11241
- Dale, J.L. (1987). Banana bunchy top: an economically important tropical plant virus disease. *Adv Virus Res* 33, 301-325. doi:10.1016/S0065-3527(08)60321-8
- Delorean, E.E., Youngblood, R.C., Simpson, S.A., Schoonmaker, A.N., Scheffler, B.E., Rutter, W.B., and Hulse-Kemp, A.M. (2023). Representing true plant genomes: haplotype-resolved hybrid pepper genome with trio-binning. *Front Plant Sci* 14, 1184112. doi.org/10.3389/fpls.2023.1184112
- Dita, M., Barquero, M., Heck, D., Mizubuti, E.S., and Staver, C.P. (2018). Fusarium wilt of banana: current knowledge on epidemiology and research needs toward sustainable disease management. *Front Plant Sci* 9, 398832. doi:10.3389/fpls.2023.1184112
- Doležel, J., Doleželová, M., and Novák, F. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biol plant* 36, 351-357. doi:10.1007/BF02920930
- Dupouy, M., Baurens, F.-C., Derouault, P., Hervouet, C., Cardi, C., Cruaud, C., Istace, B., Labadie, K., Guiougou, C., and Toubi, L. (2019). Two large reciprocal translocations characterized in the disease resistance-rich *burmannica* genetic group of *Musa acuminata*. *Ann bot* 124, 319-329. doi:10.1093/aob/mcz078
- Fao (2023). <https://www.fao.org/faostat/>.
- Gaidashova, S.V., Van Asten, P., De Waele, D., and Delvaux, B. (2009). Relationship between soil properties, crop management, plant growth and vigour, nematode occurrence and root damage in East African Highland banana-cropping systems: a case study in Rwanda. *Nematology* 11, 883-894.

- García-Bastidas, F.A., Van Der Veen, A.J., Nakasato-Tagami, G., Meijer, H.J., Arango-Isaza, R.E., and Kema, G.H. (2019). An improved phenotyping protocol for Panama disease in banana. *Front Plant Sci* 10, 1006. doi:10.3389/fpls.2019.01006
- Gold, C.S. (2000). Biology and integrated pest management of banana weevil, *Cosmopolites sordidus* (Germar). *Advancing banana and plantain R&D in Asia and the Pacific* 10, 28-33.
- Huang, H.-R., Liu, X., Arshad, R., Wang, X., Li, W.-M., Zhou, Y., and Ge, X.-J. (2023). Telomere-to-telomere haplotype-resolved reference genome reveals subgenome divergence and disease resistance in triploid Cavendish banana. *Hortic Res* 10. doi:10.1093/hr/uhad153
- Jones, D.R. (2018). *Handbook of diseases of banana, abaca and enset*. CABI.
- Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. *Nature* 444, 323-329. doi:10.1038/nature05286
- Kamaté, K., Brown, S., Durand, P., Bureau, J.-M., Nay, D.D., and Trinh, T. (2001). Nuclear DNA content and base composition in 28 taxa of *Musa*. *Genome* 44, 622-627. doi:10.1139/g01-058
- Keilwagen, J., Hartung, F., Grau, J. (2019). GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. In: Kollmar, M. (eds) Gene Prediction. Methods in Molecular Biology, vol 1962. Humana, New York, NY. doi:10.1007/978-1-4939-9173-0\_9
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* 19, 189. doi:10.1186/s12859-018-2203-5
- Kiggundu, A., Gold, C.S., Labuschagne, M.T., Vuylsteke, D., and Louw, S. (2003). Levels of host plant resistance to banana weevil, *Cosmopolites sordidus* (Germar)(*Coleoptera: Curculionidae*), in African *Musa* germplasm. *Euphytica* 133, 267-277. doi:10.1023/A:1025725124849
- Kumar, P.L., Selvarajan, R., Iskra-Caruana, M.-L., Chabannes, M., and Hanna, R. (2015). Biology, etiology, and control of virus diseases of banana and plantain. *Adv virus res* 91, 229-269. doi:10.1016/bs.aivir.2014.10.006
- Kung'u, J.N., and Jeffries, P. (2001). Races and virulence of *Fusarium oxysporum* f. sp. *ubense* on local banana cultivars in Kenya. *Ann Appl Biol* 139, 343-349. doi:10.1111/j.1744-7348.2001.tb00148.x
- Li, X., Yu, S., Cheng, Z., Chang, X., Yun, Y., Jiang, M., Chen, X., Wen, X., Li, H., Zhu, W. et al. (2024). Origin and evolution of the triploid cultivated banana genome. *Nature Genet* 56, 136-142. doi:10.1038/s41588-023-01589-3
- Liu, X., Arshad, R., Wang, X., Li, W.-M., Zhou, Y., Ge, X.-J., and Huang, H.-R. (2023). The phased telomere-to-telomere reference genome of *Musa acuminata*, a main contributor to banana cultivars. *Sci Data* 10, 631. doi:10.1038/s41597-023-02546-9
- Lorenzen, J., Tenkouano, A., Bandyopadhyay, R., Vroh, B., Coyne, D. and Tripathi, L. (2010). Overview of banana and plantain (*Musa* ssp.) improvement in Africa: Past AND Future. *Acta Hort* 879, 595-603. doi:10.17660/ActaHortic.2010.879.66
- Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* 11, e78526. doi:10.7554/eLife.78526
- Lysak, M., Dolež Elová, M., Horry, J.-P., Swennen, R., and Dolež El, J. (1999). Flow cytometric analysis of nuclear DNA content in *Musa*. *Theore Appl Genet* 98, 1344-1350. doi:10.1007/s001220051201

- Manni, M., Berkeley, M.R., Seppy, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38, 4647-4654. doi:10.1093/molbev/msab199
- Martin, G., Baurens, F.-C., Hervouet, C., Salmon, F., Delos, J.-M., Labadie, K., Perdereau, A., Mournet, P., Blois, L., Dupouy, M., Carreel, F., Ricci, S., Lemainque, A., Yahiaoui, N., and D'hont, A. (2020a). Chromosome reciprocal translocations have accompanied subspecies evolution in bananas. *Plant J* 104, 1698-1711. doi:10.1111/tpj.15031
- Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., Fondi, E., Perrier, X., Glaszmann, J.-C., D'hont, A., and Yahiaoui, N. (2020b). Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *Plant J* 102, 1008-1025. doi:10.1111/tpj.14683
- Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., Derouault, P., Roques, D., Salmon, F., Rouard, M., Sardos, J., Labadie, K., Baurens, F.C., and D'hont, A. (2017). Evolution of the banana genome (*Musa acuminata*) is impacted by large chromosomal translocations. *Mol Biol Evol* 34, 2140-2152. doi:10.1093/molbev/msx164
- Martínez, G., Olivares, B.O., Rey, J.C., Rojas, J., Cardenas, J., Muentes, C., and Dawson, C. (2023). The advance of Fusarium wilt tropical race 4 in *Musaceae* of Latin America and the Caribbean: Current situation. *Pathogens* 12, 277. doi:10.3390/pathogens12020277
- Mbaka, J., Nakato, V., Auma, J., and Odero, B. (2009). Status of banana Xanthomonas wilt in western Kenya and factors enhancing its spread.
- Ndayihanzamaso, P., Mostert, D., Matthews, M.C., Mahuku, G., Jomanga, K., Mpanda, H.J., Mduma, H., Brown, A., Uwimana, B., Swennen, R. et al. (2020). Evaluation of Mchare and Matooke Bananas for Resistance to *Fusarium oxysporum* f. sp. *cubense* Race 1. *Plants* 9, 1082. doi:10.3390/plants9091082
- Němečková, A., Christelová, P., Čížková, J., Nyine, M., Van Den Houwe, I., Svačina, R., Uwimana, B., Swennen, R., Doležel, J., and Hříbová, E. (2018). Molecular and cytogenetic study of East African highland banana. *Front Plant Sci* 9, 1371. doi:10.3389/fpls.2018.01371
- Niyongere, C., Losenge, T., Ateka, E.M., Nkezabahizi, D., Blomme, G., and Lepoint, P. (2012). Occurrence and distribution of banana bunchy top disease in the Great Lakes region of Africa. *Tree and Forestry Science and Biotechnology* 6, 102-107.
- Ocan, D., Mukasa, H., Rubaihayo, P., Tinzaara, W., and Blomme, G. (2008). Effects of banana weevil damage on plant growth and yield of East African *Musa* genotypes. *J Appl Biosci* 9, 407-415.
- Onkendi, E.M., Kariuki, G.M., Marais, M., and Moleleki, L.N. (2014). The threat of root-knot nematodes (*Meloidogyne* spp.) in Africa: a review. *Plant pathol* 63, 727-737. doi:10.1111/ppa.12202
- Ordóñez, N., Seidl, M.F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B.P., Ploetz, R.C., and Kema, G.H. (2015). Worse comes to worst: bananas and Panama disease—when plant and pathogen clones meet. *PLoS pathog* 11, e1005197. doi:10.1371/journal.ppat.1005197
- Ortiz, R., and Swennen, R. (2014). From crossbreeding to biotechnology-facilitated improvement of banana and plantain. *Biotechnol Adv* 32, 158-169. doi:10.1016/j.biotechadv.2013.09.010
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., Jiang, N., Hirsch, C.N., and Hufford, M.B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* 20, 275. doi:10.1186/s13059-019-1905-y

- Ploetz, R.C. (2015). Management of Fusarium wilt of banana: A review with special reference to tropical race 4. *Crop Prot* 73, 7-15. doi:10.1016/j.cropro.2015.01.007
- Ploetz, R.C., Kepler, A.K., Daniells, J., and Nelson, S.C. (2007). Banana and plantain—an overview with emphasis on Pacific island cultivars. *Species profiles for Pacific Island agroforestry* 1, 21-32.
- Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Commun* 11, 1432. doi:10.1038/s41467-020-14998-3
- Reeder, R., Muhinyuza, J., Opolot, O., Aritua, V., Crozier, J., and Smith, J. (2007). Presence of banana bacterial wilt (*Xanthomonas campestris* pv. *musacearum*) in Rwanda. *Plant Pathol* 56, 1038-1038.
- Rouard, M., Droc, G., Martin, G., Sardos, J., Hueber, Y., Guignon, V., Cenci, A., Geigle, B., Hibbins, M.S., Yahiaoui, N. et al. (2018). Three New Genome Assemblies Support a Rapid Radiation in *Musa acuminata* (Wild Banana). *Genome Biol Evol* 10, 3129-3140. doi:10.1093/gbe/evy227
- Ruas, M., Guignon, V., Sempere, G., Sardos, J., Hueber, Y., Duvergey, H., Andrieu, A., Chase, R., Jenny, C., Hazekamp, T. et al. (2017). MGIS: managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. *Database (Oxford)* 2017. doi:10.1093/database/bax046
- Rukazambuga, N., Gold, C., and Gowen, S. (1998). Yield loss in East African highland banana (*Musa* spp., AAA-EA group) caused by the banana weevil, *Cosmopolites sordidus* Germar. *Crop Prot* 17, 581-589. doi:10.1016/S0261-2194(98)00056-8
- Rutherford, M. (1999). Fusarium wilt of banana in East Africa, In: Banana fusarium wilt management: towards sustainable cultivation. Proceedings of the International Workshop on the Banana Fusarium Wilt Disease, Genting Highlands Resort, Malaysia, 18-20 October 1999: International Network for the Improvement of Banana and Plantain), 86-94.
- Sachter-Smith, G. (2023). The Wild Bananas: a catalogue of wild *Musa* species and tribute to Markku Häkkinen. Bioversity International. [www.musalit.org/seeMore.php?id=21272](http://www.musalit.org/seeMore.php?id=21272) (accessed on 27 June 2024).
- Sahu, S.K., and Liu, H. (2023). Long-read sequencing (method of the year 2022): the way forward for plant omics research. *Mol Plant* 16, 791-793. doi:10.1016/j.molp.2023.04.007
- Sardos, J., Breton, C., Perrier, X., Van Den Houwe, I., Carpentier, S., Paofa, J., Rouard, M., and Roux, N. (2022). Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. *Front Plant Sci* 13, 969220. doi:10.3389/fpls.2022.969220
- Sardos J, Rouard M, Hueber Y, Cenci A, Hyma KE, van den Houwe I, Hribova E, Courtois B, Roux N. (2016a) A genome-wide association study on the seedless phenotype in banana (*Musa* spp.) reveals the potential of a selected panel to detect candidate genes in a vegetatively propagated crop. *PLoS One* 4, 11(5):e0154448. doi: 10.1371/journal.pone.0154448.
- Sardos, J., Perrier, X., Doležel, J., Hříbová, E., Christelová, P., Van Den Houwe, I., Kilian, A., and Roux, N. (2016b). DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (*Musa* spp.). *Ann Bot* 118, 1269-1278. doi:10.1093/aob/mcw170
- Sedlazeck, F.J., Lee, H., Darby, C.A., and Schatz, M.C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Revs Genet* 19, 329-346. doi:10.1038/s41576-018-0003-4
- Šimková, H., Câmara, A.S., and Mascher, M. (2024). Hi-C techniques: from genome assemblies to transcription regulation. *J Exp Bot*.

- Šimoníková, D., Němečková, A., Čížková, J., Brown, A., Swennen, R., Doležel, J., and Hřibová, E. (2020). Chromosome Painting in Cultivated Bananas and Their Wild Relatives (*Musa* spp.) Reveals Differences in Chromosome Structure. *Int J Mol Sci* 21, 7915. doi:10.3390/ijms21217915
- Smit, A.F.A., Hubley, R., and Green, P. (2013). *RepeatMasker Open-4.0*. Online. <http://www.repeatmasker.org> (Accessed on 27 June 2024).
- Smith, J., Jones, D., Karamura, E., Blomme, G., and Turyagyenda, F. (2008). An analysis of the risk from *Xanthomonas campestris* pv. *musacearum* to banana cultivation in Eastern, Central and Southern Africa. Montpellier, France. Bioversity International.
- Solidaridad (2023). Small Farmer Atlas: Evaluating the voice of small-scale farmers in high value markets. Online. <https://www.solidaridadnetwork.org/wp-content/uploads/2023/05/2023-Small-Farmer-Atlas-Solidaridad-5-3.pdf> (Accessed on 27 June 2024).
- Speijer, P., Boonen, E., Vuylsteke, D., Swennen, R.L., and De Waele, D. (1999). Nematode reproduction and damage to *Musa* sword suckers and sword sucker derived plants. *Nematropica*, 193-203.
- Steuernagel, B., Witek, K., Krattinger, S.G., Ramirez-Gonzalez, R.H., Schoonbeek, H.J., Yu, G., Baggs, E., Witek, A.I., Yadav, I., Krasileva, K.V., Jones, J.D.G., Uauy, C., Keller, B., Ridout, C.J., and Wulff, B.B.H. (2020). The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol* 183, 468-482. doi:10.1104/pp.19.01273
- Su, H.J., Tsao, L.Y., Wu, M.L., and Hung, T.H. (2003). Biological and molecular categorization of strains of Banana bunchy top virus. *J Phytopathol* 151, 290-296. doi:10.1046/j.1439-0434.2003.00721.x
- Tripathi, L., Mwangi, M., Abele, S., Aritua, V., Tushemereirwe, W.K., and Bandyopadhyay, R. (2009). *Xanthomonas* wilt: a threat to banana production in East and Central Africa. *Plant Dis* 93, 440-451. doi:10.1094/PDIS-93-5-0440
- Tushemereirwe, W., Kangire, A., Kubiriba, J., Nakyanzi, M., and Gold, C. (2004). Diseases threatening banana biodiversity in Uganda. *Afr Crop Sci J* 12, 19-26. doi:10.4314/acsj.v12i1.27658
- Uwimana, B., Mwanje, G., Batte, M., Akech, V., Shah, T., Vuylsteke, M., and Swennen, R. (2021). Continuous mapping identifies loci associated with weevil resistance [*Cosmopolites sordidus* (Germar)] in a triploid banana population. *Front Plant Sci* 12, 753241. doi:10.3389/fpls.2021.753241
- Van Westerhoven, A.C., Meijer, H.J.G., Seidl, M.F., and Kema, G.H.J. (2022). Uncontained spread of *Fusarium* wilt of banana threatens African food security. *PLOS Pathog* 18, e1010769. doi:10.1371/journal.ppat.1010769
- Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., Xu, C., Zhao, S., Fang, X., Jia, C., Wang, J. et al. (2019). *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nat Plants* 5, 810-821. doi:10.1038/s41477-019-0452-6
- Zuo, C., Deng, G., Li, B., Huo, H., Li, C., Hu, C., Kuang, R., Yang, Q., Dong, T., and Sheng, O. (2018). Germplasm screening of *Musa* spp. for resistance to *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (Foc TR4). *Eur J Plant Pathol* 151, 723-734. doi:10.1007/s10658-017-1406-3