



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

## Genome sequencing data for wild and cultivated bananas, plantains and abacá



Christine Sambles<sup>a</sup>, Lakshmipriya Venkatesan<sup>a</sup>,  
Olanrewaju M. Shittu<sup>a</sup>, James Harrison<sup>a</sup>, Karen Moore<sup>a</sup>,  
Leena Tripathi<sup>b</sup>, Murray Grant<sup>c</sup>, Rachel Warmington<sup>d</sup>,  
David J. Studholme<sup>a,\*</sup>

<sup>a</sup> Biosciences, University of Exeter, Exeter EX4 4QD, United Kingdom

<sup>b</sup> International Institute of Tropical Agriculture, P.O. Box 30709, Nairobi, Kenya

<sup>c</sup> Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

<sup>d</sup> Eden Project, Bodelva, Cornwall PL24 2SG, UK

## ARTICLE INFO

## Article history:

Received 11 August 2020

Revised 15 September 2020

Accepted 17 September 2020

Available online 23 September 2020

## Keywords:

Musa

Genome

Plant

Diversity

Sequence analysis

DNA

## ABSTRACT

We performed shotgun genome sequencing on a total of 19 different *Musa* genotypes including representatives of wild banana species *Musa acuminata* and *M. balbisiana*, allopolyploid bananas and plantains, Fe'i banana, pink banana (also known as hairy banana) and abacá (also known as hemp banana). We aligned sequence reads against a previously sequenced reference genome and assessed ploidy and, in the case of allopolyploids, the contributions of the A and B genomes; this provides important quality-assurance data about the taxonomic identities of the sequenced plant material. These data will be useful for phylogenetics, crop improvement, studies of the complex story of intergenomic recombination in AAB and ABB allotriploid bananas and plantains and can be integrated into resources such as the Banana Genome Hub.

© 2020 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

\* Corresponding author.

E-mail address: [d.j.studholme@exeter.ac.uk](mailto:d.j.studholme@exeter.ac.uk) (D.J. Studholme).

Social media: (C. Sambles), (O.M. Shittu), (L. Tripathi), (R. Warmington), (D.J. Studholme)

## Specifications Table

Subject	Biology
Specific subject area	Genomics of crop plants
Type of data	Deoxyribonucleic acid (DNA) sequence
How data were acquired	Shotgun genomic DNA sequencing was performed using Illumina HiSeq 2500, Illumina NovaSeq and BGISEQ-500 platforms
Data format	Raw sequencing reads
Parameters for data collection	DNA was extracted from leaf material
Description of data collection	Shotgun genomic DNA sequencing was performed using Illumina HiSeq 2500, Illumina NovaSeq and BGISEQ-500 platforms
Data source location	Institution: University of Exeter City: Exeter Country: United Kingdom Latitude and longitude (and GPS coordinates) for collected samples/data: Plant samples were collected from the Eden Project at 50.3601° N, 4.7447° W (50.357165238 -4.740163)
Data accessibility	Repository name: NCBI BioProject Data identification numbers: PRJNA540118, PRJNA413600 Direct URLs to data: <a href="https://www.ncbi.nlm.nih.gov/bioproject/540118">https://www.ncbi.nlm.nih.gov/bioproject/540118</a> <a href="https://www.ncbi.nlm.nih.gov/bioproject/413600">https://www.ncbi.nlm.nih.gov/bioproject/413600</a>

## Value of the Data

- This genomic resequencing data will inform studies of *Musa* evolution, biodiversity, speciation and allopolyploidy.
- Genome-wide sequence data are presented for abacá (*Musa textilis*), the Fe'i banana (*M. troglodytarum*) and the pink banana (*M. velutina*) as well as edible and wild bananas and plantains belonging to the species *M. acuminata* and *M. balbisiana* and their interspecific hybrids.
- This is a useful resource for breeders, researchers as well as science communicators engaging with the general public about the germplasm collection at the Eden Project.
- The data can be mined for polymorphisms with value as markers for breeding strategies.
- These data can be integrated into banana genomics resources such as the Banana Genome Hub [1].
- Since some samples were sequenced using more than one method, the data can be used to compare performances of alternative sequencing platforms [2].

## 1. Data Description

Genomic shotgun sequencing data was generated using BGISEQ-500 (Table 1), Illumina HiSeq 2500 using libraries of two different sizes (Tables 2 and 3) and Illumina NovaSeq 6000 (Table 4). This generated a total of 505.69 GB and 120.95 GB raw read data for the Eden Project and IITA accessions respectively. Raw data is available at NCBI's Sequence Read Archive [3] via BioProjects PRJNA540118 and PRJNA413600.

An important quality control step is to check whether the sequence data are consistent with the botanical identifications of the source material. Therefore, we assessed observed against expected levels of ploidy. For allopolyploids purported to originate from interspecific hybrids between *Musa acuminata* and *Musa balbisiana*, we assessed the relative contributions of these respective "A" and "B" genomes compared against the expected characteristics of each sample as described under Experimental Design, Materials, and Methods. The resulting quality-control metrics are summarised in Table 5 and in Fig. 1. Accessions 2012-1152 (SAMN11522021), 1999-2846 (SAMN11522023) and 2011-0950 (SAMN11522017) were expected to be allopolyploids containing contributions from both the A and B genomes but sequence data appeared to be exclusively

**Table 1**

Genomic sequencing data generated using BGISEQ-500 (2 × 150 bp reads, 300-bp insert size).

BioSample	SRA accession	Eden project identifier	Received as	Depth of coverage
SAMN11522014	SRR8989628, SRR9734077	2012-1161	<i>Musa acuminata</i> 'Green-Red'	59 ×
SAMN11522015	SRR8989629	2012-1156	<i>Musa acuminata</i> 'Paka'	28 ×
SAMN11522016	SRR8989630, SRR9734074	2012-1173	<i>Musa acuminata</i> subsp. <i>zebrina</i>	54 ×
SAMN11522017	SRR8989631, SRR9734078	2011-0950	<i>Musa acuminata</i> × <i>balbisiana</i> 'Congo 2' (plantain subgroup)	59 ×
SAMN11522018	SRR8989632	2012-1154	<i>Musa acuminata</i> subsp. <i>malaccensis</i>	28 ×
SAMN11522019	SRR8989633, SRR9734079, SRR9850640	2001-1027	<i>Musa balbisiana</i>	52 ×
SAMN11522020	SRR8989634, SRR9734076, SRR9850639	2012-1164	<i>Musa acuminata</i> 'Calypso'	54 ×
SAMN11522021	SRR8989635	2012-1152	<i>Musa acuminata</i> × <i>balbisiana</i> 'Safet Velchi' (Ney Poovan subgroup)	30 ×
SAMN11522022	SRR8989636	2011-0952	<i>Musa acuminata</i> × <i>balbisiana</i> "One Hand Planty"	28 ×
SAMN11522023	SRR8989637	1999-2846	<i>Musa</i> × <i>paradisica</i> <sup>a</sup>	31 ×
SAMN11522024	SRR8989638	1998-2307	<i>Musa acuminata</i> 'Pisang Mas' (Sucrier subgroup)	32 ×
SAMN11522025	SRR8989639, SRR9850642	1999-0524	<i>Musa textilis</i>	
SAMN11522026	SRR8989640, SRR9734080, SRR9850641	1999-0158	<i>Musa troglodytarum</i> 'Wain' (F'ei group)	36 ×
SAMN11522027	SRR8989641, SRR9734075	2012-1166	<i>Musa velutina</i>	47 ×

<sup>a</sup> Accession 1999-2846 was received as *Musa* × *paradisica* but genome sequence data suggest that it is *Musa acuminata*.

**Table 2**

Genomic sequencing data generated using Illumina HiSeq (2 × 150 bp reads, 800-bp insert size).

BioSample	SRA accession	Eden project identifier	Received as	Depth of coverage
SAMN11522025	SRR9696635	1999-0524	<i>Musa textilis</i>	23 ×
SAMN11522021	SRR9696636	2012-1152	<i>Musa acuminata</i> × <i>balbisiana</i> 'Safet Velchi' (Ney Poovan subgroup)	36 ×

**Table 3**

Genomic sequencing data generated using Illumina HiSeq (2 × 125 bp reads, 300-bp insert).

BioSample	SRA accession	Received as	Depth of coverage
SAMN07758499	SRR6147591	<i>Musa acuminata</i> × <i>balbisiana</i> 'Sukali Ndiizi' (AAB group)	53 ×
SAMN07758501	SRR6147590	<i>Musa acuminata</i> × <i>balbisiana</i> 'Gonja Manjaya' (AAB group)	18 ×
SAMN07758502	SRR6147593	<i>Musa acuminata</i> 'Cavendish' (AAA group)	23 ×
SAMN07758503	SRR6147592	<i>Musa balbisiana</i>	24 ×
SAMN07758500	SRR6147589	<i>Musa acuminata</i> × <i>balbisiana</i> 'Pisang Awak' (ABB group)	28 ×

**Table 4**

Genomic sequencing data generated using Illumina NovaSeq 6000 (2 × 150 bp reads, 300-bp insert size).

BioSample	SRA accession	Eden project identifier	Received as	Depth of coverage
SAMN11522021	SRR9015638	2012-1152	<i>Musa acuminata</i> × <i>balbisiana</i> 'Safet Velchi' (Ney Poovan subgroup)	30 ×
SAMN11522022	SRR9015637	2011-0952	<i>Musa acuminata</i> × <i>balbisiana</i> 'One Hand Planty'	28 ×

**Table 5**Ploidy prediction and estimated composition of 16 accessions of *Musa* spp.<sup>a</sup>

BioSample	Name	Expected ploidy	Observed ploidy according to nQuire (if different to expected)	Expected composition	SNP data consistent with expected composition?
SAMN11522018	<i>Musa acuminata</i> subsp. <i>malaccensis</i>	2		AA	Yes
SAMN11522015	<i>Musa acuminata</i> 'Paka'	2		AA	Yes
SAMN11522014	<i>Musa acuminata</i> 'Green-Red'	3		AAA	Yes
SAMN11522016	<i>Musa acuminata</i> subsp. <i>zebrina</i>	2	4	AA	Yes
SAMN07758502	<i>Musa acuminata</i> 'Cavendish'	3		AAA	Yes
SAMN11522020	<i>Musa acuminata</i> 'Calypso'	4		AAAA	Yes
SAMN11522021	<i>Musa acuminata</i> × <i>balbisiana</i> 'Safet Velchi' (Ney Poovan subgroup)	2	3	AB	No: appears to be exclusively A
SAMN07758499	<i>Musa acuminata</i> × <i>balbisiana</i> 'Sukali Ndiizi'	3		AAB	Yes
SAMN07758501	<i>Musa acuminata</i> × <i>balbisiana</i> 'Gonja Manjaya'	3		AAB	Yes
SAMN11522022	<i>Musa acuminata</i> × <i>balbisiana</i> 'One Hand Planty'	3		AAB	Yes
SAMN07758500	<i>Musa acuminata</i> × <i>balbisiana</i> 'Pisang Awak'	3	4	ABB	Yes
SAMN11522019	<i>Musa balbisiana</i>	2	4	BB	Yes
SAMN07758503	<i>Musa balbisiana</i>	2	4	BB	Yes
SAMN11522024	<i>Musa acuminata</i> 'Pisang Mas' (Sucrier subgroup)	2		AA	Yes
SAMN11522017	<i>Musa acuminata</i> × <i>balbisiana</i> 'Congo 2' (plaintain subgroup)	3		AAB	No: appears to be exclusively A
SAMN11522023	<i>Musa</i> × <i>paradisica</i>	2	3	AAB or ABB	No: appears to be exclusively A

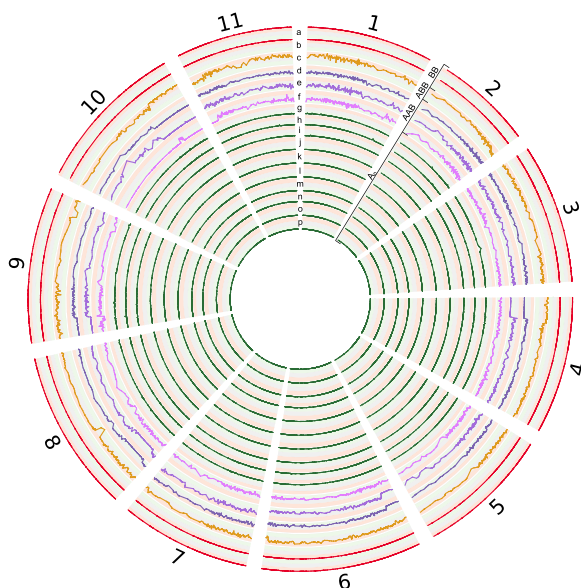
<sup>a</sup> Ploidy analysis was only performed on *M. acuminata*, *M. balbisiana* accessions and their hybrids. Consequently, *Musa textilis* (SAMN11522025), *Musa troglodytarum* 'Wain' (Fei group) (SAMN11522026) and *Musa velutina* (SAMN11522027) were excluded.

from the A genome, suggesting that these three plants had been mis-identified. Further, there were discrepancies between the expected ploidy levels versus the empirically inferred levels in several accessions.

## 2. Experimental Design, Materials and Methods

Fresh leaf material was obtained from five accessions from the IITA (International Institute of Tropical Agriculture) [4] accessions and 14 from the Eden Project. DNA was extracted from fresh leaf material and sequenced using a combination of Illumina HiSeq 2500, Illumina NovaSeq 6000 and BGISEQ-500 platforms. This yielded at least 20 × coverage of each genome and was sufficient for calling single-nucleotide polymorphisms, detecting presence/absence polymorphisms and cataloguing patterns of heterozygosity.

From the 14 plant accessions from the Eden Project, cigar leaves were cut from the plant and lyophilised in a freeze dryer before sending to BGI Tech Solutions (Hong Kong) Co., Limited, where DNA extraction and sequencing was performed.



**Fig. 1.** Circos representation of informative SNP variants identified in the 11 chromosomes of *M. acuminata*. The lines represent the LOESS smoothed percentage of B allele of 16 sequenced *Musa* accessions (*M. acuminata*, *M. balbisiana* and their hybrids). *Musa* accessions with the highest percentage of A genome are at the centre graduating to those with the highest percentage of B genome on the outside, according to the 1542 identified SNPs. Background colours represent percentage of B allele: green (0–33%), grey (33–66%) and red (66–100%). Tracks from outer (B allele dominant) to inner (A allele dominant) are: a. *M. balbisiana* (SAMN11522019), b. *M. balbisiana* (SAMN07758503), c. *M. acuminata* × *balbisiana* ‘Pisang Awak’ (SAMN07758500), d. *M. acuminata* × *balbisiana* ‘One Hand Planty’ (SAMN11522022), e. *M. acuminata* × *balbisiana* ‘Gonja Manjaya’ – AAB group (SAMN07758501), f. *M. acuminata* × *balbisiana* ‘Sukali Ndiizi’ (SAMN07758499), g. *Musa* × *paradisica* (SAMN11522023), h. *M. acuminata* × *balbisiana* ‘Safet Velchi’ – Ney Poovan subgroup (SAMN11522021), i. *M. acuminata* ‘Calypso’ (SAMN11522020), j. *M. acuminata* × *balbisiana* ‘Congo 2’ – plantain subgroup (SAMN11522017), k. *M. acuminata* ‘Pisang Mas’ – Sucrier subgroup (SAMN11522024), l. *M. acuminata* subsp. *malaccensis* (SAMN11522018), m. *M. acuminata* ‘Paka’ (SAMN11522015), n. *M. acuminata* ‘Green-Red’ (SAMN11522014), o. *M. acuminata* subsp. *zebrina* (SAMN11522016), p. *M. acuminata* ‘Cavendish’ – AAA group (SAMN07758502).

\*  $A_n$  describes A genome autopolyploidy i.e. AA or AAA or AAAA.

For the five accessions from the IITA (International Institute of Tropical Agriculture), genomic DNA was isolated using a modified CTAB (hexadecyltrimethylammonium bromide) extraction method [5]. The University of Exeter’s Sequencing Service prepared Illumina sequencing libraries after fragmenting 500 ng of DNA to an average size of 500 bp, using the NEXTflex 8-barcode Rapid DNaseq kit sequencing (Perkin Elmer) with adapters containing indexes and 5–8 cycles polymerase chain reaction (PCR) [6]. Library quality was determined using D1000 screen-tapes (Agilent) and libraries were either sequenced individually or combined in equimolar pools. Sequencing was performed on a single lane of a high-output v4 flow-cell on the Illumina HiSeq 2500 at the University of Exeter, yielding pairs of 125-bp reads.

This yielded at least  $20 \times$  coverage of each genome, sufficient for calling single-nucleotide polymorphisms, detecting presence/absence polymorphisms and cataloguing patterns of heterozygosity. Reads were also generated with longer inserts using the Illumina HiSeq ( $2 \times 150$  bp reads, 800-bp insert size) for two of the samples, which potentially aids resolution of sequence repeats if data are used in *de novo* assembly of genomes.

The quality of the sequencing reads was evaluated using FASTQC [7]. Before further analyses, reads were trimmed and adapters removed using TrimGalore [8] with command-line options “-q 30 --paired”. Trimmed and filtered reads were aligned to the *M. acuminata* genome [9] using BWA [10] to generate binary alignment map (BAM) files [11].

As a prerequisite for plotting the relative contributions of the A and B genomes in allopolyploids, we first identified a set of informative SNPs that distinguish A (*M. acuminata*) from B (*M. balbisiana*) as follows utilising SAMtools' *mpileup* function, BCFtools [11,12] and custom scripts available at <https://github.com/davidjstudholme/SNPsFromPileups>. First, the relevant BAM alignment files were converted into uncompressed VCF format using SAMtools v1.6 (*mpileup* function), selecting for variant sites only (*-v*) using the alternative model for multiallelic and rare-variant calling (*-m*). Potential SNPs were filtered using the filter function of BCFtools (v1.6), excluding potential SNPs that were within 100 base pairs of an indel (*--SnpGap 100*) and had a quality score of less than 35 (*QUAL>=35*) with a depth of 5 or more reads (*MIN(DP)>=5*). The minimum number of reads supporting an indel was set to two (*MIN(IDV)>=2*). Variants that were flagged as indels were excluded (*INDEL=0*). The resulting filtered VCF files contained the positions of candidate SNPs that distinguished the B genome [13] versus the A reference genome [14]. At each of these informative SNPs, we quantified the relative abundance of the A- and B- alleles, only considering sites where the depth was between 10 and 50. When plotting, the resulting percentage of the B allele was smoothed in R using the LOESS package [15]. The percentages of the B alleles at each SNP were visualised using Circos [16] (Fig. 1).

We used nQuire [17] to estimate ploidy from the BAM files (of genomic reads aligned against the *M. acuminata* reference genome). After de-noising to remove noise from mis-mapping due to highly repetitive regions, we assessed ploidy level using the *lrdmodel* command of nQuire to produce delta log-likelihoods of diploidy, triploidy or tetraploidy. The lowest delta log-likelihood was taken to indicate the most likely ploidy level (Table 5). To infer ploidy levels, we used nQuire [17] to predict ploidy using BAM alignment files generated with BWA. The ploidy model yielding lowest value of  $\Delta\log L$  was chosen as the inferred ploidy. The command lines used were as follows:

```
nQuire create -b example.bam -o example
for i in *.bin; do echo $i; nQuire denoise $i -o $i\_denoised; done
for i in *_denoised.bin; do echo $i; nQuire lrdmodel -t 8 $i; done
```

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

## Acknowledgements

David Studholme and Christine Sambles were supported by "MUSA: Microbial Uptakes for Sustainable management of major bananA pests and diseases" (Grant ID 727624, European Union (Horizon 2020)). DNA sequencing costs were supported by a grant from the Gatsby Charitable Foundation entitled "Banana Genetic Resources at Eden project" (GAT3587). We are grateful to Peggy Dousseaud for assistance with lyophilising leaf material and to Hetty Ninnis for expert assistance in collecting plant material at the Eden Project. This project made use of the University of Exeter's high-performance computing facility, Isca. This project utilised DNA sequencing equipment (Illumina HiSeq) funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1).

## References

- [1] G. Droc, D. Larivière, V. Guignon, N. Yahiaoui, D. This, O. Garsmeur, A. Dereeper, C. Hamelin, X. Argout, J.-F. Dufayard, J. Lengelle, F.-C. Baurens, A. Cenci, B. Pitollat, A. D'Hont, M. Ruiz, M. Rouard, S. Bocs, The banana genome hub, Database 2013 (2013) bat035.

- [2] F.-Y. Zhu, M.-X. Chen, N.-H. Ye, W.-M. Qiao, B. Gao, W.-K. Law, Y. Tian, D. Zhang, D. Zhang, T.-Y. Liu, Q.-J. Hu, Y.-Y. Cao, Z.-Z. Su, J. Zhang, Y.-G. Liu, Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants, *Plant Methods* 14 (2018) 69.
- [3] R. Leinonen, H. Sugawara, M. Shumway, The sequence read archive, *Nucleic Acids Res.* 39 (2011) D19–D21.
- [4] M. Pillay, E. Ogundiwin, A. Tenkouano, J. Dolezel, Ploidy and genome composition of *Musa* germplasm at the International Institute of Tropical Agriculture (IITA), *Afr. J. Biotechnol.* 5 (2006) 1224–1232.
- [5] N.J. Gawel, R.L. Jarret, A modified CTAB DNA extraction procedure for *Musa* and *Impatiens*, *Plant Mol. Biol. Rep.* 9 (1991) 262–266.
- [6] S.R. Head, H.K. Komori, S.A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D.R. Salomon, P. Ordoukhanian, Library construction for next-generation sequencing: Overviews and challenges, *Biotechniques* 56 (2014).
- [7] S. Andrews, FastQC: a quality control tool for high throughput sequence data, (2019) Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- [8] F. Krueger, Babraham Bioinformatics – Trim Galore!, (2019) Available online at: [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore).
- [9] A. D'Hont, F. Denoeud, J.-M.J. Aury, F.-C.F. Baurens, A. D'Hont, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lengellé, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M.R. Mckain, J. Leebens-Mack, D. Burgess, M. Freeling, D. Mbéguié-A-Mbéguié, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poiran, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievart, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A.-M. Risterucci, J. Weissenbach, M. Ruiz, J.-C. Glaszmann, F. Quétier, N. Yahiaoui, P. Wincker, The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants, *Nature* 488 (2012) 213–217.
- [10] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (2009) 1754–1760.
- [11] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, Subgroup, 1000 genome project data processing the sequence alignment/map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079.
- [12] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics* 27 (2011) 2987–2993.
- [13] Z. Wang, H. Miao, J. Liu, B. Xu, X. Yao, C. Xu, S. Zhao, X. Fang, C. Jia, J. Wang, J. Zhang, J. Li, Y. Xu, J. Wang, W. Ma, Z. Wu, L. Yu, Y. Yang, C. Liu, Y. Guo, S. Sun, F. Baurens, G. Martin, F. Salmon, O. Garsmeur, N. Yahiaoui, C. Hervouet, M. Rouard, N. Laboureau, R. Habas, S. Ricci, M. Peng, A. Guo, J. Xie, Y. Li, Z. Ding, Y. Yan, W. Tie, A. D'Hont, W. Hu, Z. Jin, *Musa balbisiana* genome reveals subgenome evolution and functional divergence, *Nat. Plants* (2019).
- [14] G. Martin, F.-C. Baurens, G. Droc, M. Rouard, A. Cenci, A. Kilian, A. Hastie, J. Doležel, J.-M. Aury, A. Alberti, F. Carreel, A. D'Hont, Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods, *BMC Genom.* 17 (2016) 243.
- [15] W.S. Cleveland, E. Grosse, W.M. Shyu, Local regression models, in: *Statistical Models in S*, Routledge, 2018, pp. 309–376.
- [16] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, *Genome Res.* 19 (2009) 1639–1645.
- [17] C.L. Weiß, M. Pais, L.M. Cano, S. Kamoun, H.A. Burbano, nQuire: a statistical framework for ploidy estimation using next generation sequencing, *BMC Bioinform.* 19 (2018) 122.