1  *Painting the diversity of a world's favourite fruit: A next generation catalogue of*

2  *cultivated bananas*

3  Julie Sardos[1,2,&,*], Alberto Cenci[1,2,&], Guillaume Martin[2,3], Catherine Breton[1,2], Valentin Guignon[1,2], Ines

4  Van den Houwe[4], Yaleidis Mendez[5], Gabriel L. Sachter-Smith[1,6], Rachel Chase[1], Max Ruas[1,2], Ronan

5  Rivallan[2,3], Janet Paofa[7], William Wigmore[8], David Tilafono Hunter[9], Angélique D'Hont[2,3], Nabila

6  Yahiaoui[2,3], Christophe Jenny[2,3], Xavier Perrier[2,3], Nicolas Roux[1,2], Mathieu Rouard[1,2,*]

7

8  [1] Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier, France

9  [2] UMR AGAP Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

10  [3] CIRAD, UMR AGAP Institute, F-34398 Montpellier, France

11  [4] Bioversity International, c/o KU Leuven, Willem de Croylaan 42, P.O. Box 2455, 3001 Leuven, Belgium

12  [5] USDA–ARS Tropical Agriculture Research Station, Mayaguez, PR, 00680

13  [6] Hawaii Banana Source, Waialua, HI, USA

14  [7] NARI, Southern Regional Centre, Central Province, Port Moresby, Papua New Guinea

15  [8] Ministry of Agriculture, Rarotonga, Cook Islands

16  [9] Ministry of Agriculture and Fisheries, Samoa

17

18  &: These authors contributed equally to the study

19  *: Corresponding authors

## *Abstract*

### *Societal impact statement*

Bananas are nutritious fruits of major importance in the tropics and subtropics. Characterizing their diversity is essential to ensure their conservation and use. A catalogue showcasing cultivated bananas genomic diversity was compiled and is to be used as a tool to support the classification of banana cultivars. This research revealed that cultivated banana groups are not all made of identical clones. Materials from recent collecting missions indicated that more banana diversity is expected to be found as the exploration of the banana gene pool continues. These discoveries will drive dynamic conservation strategies for banana genetic resources and will increase their use.

### Summary

- Banana is an important food crop cultivated in many tropical and subtropical regions around the world. Due to their low fertility, banana landraces are clonally propagated. However, different factors, such as synonymy and the effects of environment, make their assignment to described sets of clones, or cultivar groups, difficult. Consequently, passport data of accessions in genebanks is often uncomplete and sometimes inaccurate.

- With the recent advances in genomics, a new powerful tool was developed enabling the fine-scale characterization of banana's ancestry along chromosomes, i.e. *in silico* chromosome painting. We applied this method to a high-throughput genotyping data set obtained from 317 banana accessions spanning most of the known cultivar groups. This set included both genebank and new uncharacterized materials.

- By comparing curated morphological assignation to the genomic patterns resulting from *in silico* chromosome painting, we were able to compile a catalogue referencing the chromosome painting patterns of most of the described cultivar groups.

- Examining the genomic patterns obtained, we discovered intra-cultivar group variability. In some cultivar groups, mitotic recombination or deletions were clonally accumulated in cultivars. In addition, we identified at least 4 cultivar groups in which cultivars likely resulting from distinct sexual events co-existed, notably Pisang Awak in which 5 distinct genomic patterns of two ploidy levels were identified. New patterns were also discovered in the newest materials of the set, showing that a wider diversity of clones still exist *on farm*.

## *Introduction*

Crop diversity is critical for maintaining the resilience and adaptability of food systems in the face of changing environmental conditions and pests and diseases (Smale and Jamora 2020; McCouch *et al.* 2020). Characterizing this diversity is therefore a much-needed effort to reach a comprehensive overview of the genetic diversity existing within a crop species and to ensure that effective conservation strategies are put in place. This knowledge serves as an essential baseline to monitor the evolution of diversity *in-situ* and to identify new material to be conserved *ex-situ*. Initially, crop characterization consisted in morphological assessments, but it later included molecular descriptions using molecular markers such as RAPD, RFLP, SSR, DArT and SNPs (Powell *et al.* 1996; Agarwal *et al.* 2008; Kilian *et al.* 2012). With the recent progresses made in the fields involving genomics, an unprecedented level of fine-scale characterization can be reached (McCouch *et al.* 2020).

Bananas are an important crop for many tropical and subtropical regions around the world. They are a staple food for millions of people and are a major source of nutrients, income, and employment for many communities. Currently, 80% of global banana production is limited to a few groups of cultivars, such as the dessert Cavendish and the cooking Plantain, but a much wider diversity exists, especially, but not only, in smallholder farms of South Asia and West Oceania, the centre of origin of the crop (Simmonds 1962). This diversity, found in smallholder fields, is conserved through *ex-situ* national, regional and international genebanks, which serve as repositories for landraces, modern varieties and wild relatives and aim at safeguarding as much diversity as possible for present and future generations (Van den houwe *et al.* 2020). Continuous efforts are necessary to properly characterize and rationalize the conserved germplasm and to identify gaps in collections.

The classification of banana cultivars is complex as cultivars are diploid or polyploid hybrids originating from crosses between different wild gene pools. To help in the assignation of cultivars, a scoring method was developed by (Simmonds and Shepherd 1955) on the assumption that most of banana cultivars were derived from the diploid wild species *Musa acuminata* Colla and *M. balbisiana* Colla. By considering the different ploidy levels existing in the crop (diploid, triploid, tetraploid) and the scored relative contribution of both wild ancestors (coded A and B, respectively), Simmonds and Shepherd (1955) introduced the concept of genome constitution groups (e.g. AA, AB, AAA, AAB, ABB). Their work also laid the foundations for the definition of an additional taxonomical level, the subgroups, aiming at refining the classification. These subgroups correspond to cultivars considered to be somatic mutants fixed through vegetative propagation of a single seedling or possibly siblings of related parents (Simmonds 1966; Stover and Simmonds 1987).

81    After more extensive exploration of banana growing regions and with the beginning of molecular
82    characterization of wild and cultivated germplasm, differential contributions of the subspecies of *M.*
83    *acuminata* were assessed (Carreel *et al.* 2002; Perrier *et al.* 2011). In addition, other wild ancestors were
84    identified. Notably, it was assessed that some banana cultivars were also hybrids with  *M. schizocarpa* (S
85    genome), and with another undefined *Musa* species of the former *Australimusa* section (T genome)
86    (Shepherd and Ferreira 1984; Jarret *et al.* 1992; Carreel *et al.* 1994).  Using this classification, a catalogue
87    of the global banana diversity, the Musalogue (Daniells *et al.* 2001), documented 14 genome groups of
88    three ploidy levels, subdivided in 37 subgroups across two botanical sections (Daniells *et al.* 2001;
89    Häkkinen 2013). All subgroups, preferentially called cultivar groups in our study, such as Cavendish,
90    Plantain, Sucrier or Pisang Awak, belong to the *Musa* section.

91    Despite being visionary for its time, and even with the wide use of published standard descriptors for
92    banana (IPGRI 1996; Taxonomy Advisory Group (TAG) 2016), classifying banana cultivars based on
93    morphology alone is challenging. First, it requires relatively controlled growing conditions since both
94    scores and descriptors were developed in and for *ex-situ* collections and do not consider variations that
95    can be due to environmental conditions (large sense). Second, the identification requires observations at
96    different stages of the plant development, including fructification, requiring space and time as well as
97    suitable climatic growing conditions. Moreover, the assignation of cultivars to cultivar groups depends
98    upon trained eyes and the experience of the observers. Consequently, cultivars are regularly inconsistently
99    classified, with a risk of negative impact on their conservation (Vogel Ely *et al.* 2017). The molecular
100   markers developed later, such as RFLP (Carreel *et al.* 2002), SSR (Hippolyte *et al.* 2012; Christelová *et*
101   *al.* 2017) or DArT (Risterucci *et al.* 2009; Sardos *et al.* 2016) enabled to characterise the genetic bases of
102   the cultivar groups and helped in the assignation process. However, the use of these technologies for such
103   complex crop do not always allow the unambiguous assignation of accessions.

104   Recent advances in genomics revealed that cultivar genomes are based on several steps of subgenome
105   combination through hybridization, complicated by homoeologous chromosome exchanges between the
106   genomes of *M. acuminata* and *M. balbisiana* (Baurens *et al.* 2019; Cenci *et al.* 2021; Higgins *et al.* 2023).
107   Moreover, the characterization of the genome diversity of *M. acuminata* subspecies and the fine-scale
108   examination of the A genomes of cultivars showed their mosaic nature. Using a technique known as
109   genome ancestry mosaics painting (or chromosome painting), contributions of *M. acuminata* subspecies
110   could be inferred along the chromosomes of diploids, triploid and tetraploid bananas. This method not
111   only allowed the allocation of specific genomic patterns to cultivars, but it also revealed unidentified

112  ancestors and the underappreciated contribution of *M. schizocarpa* to cultivated bananas (Martin *et al.*
113  2020; Martin, Cottin, *et al.* 2023).

114  In this study, we applied the genome ancestry mosaic painting methodology to a large and well curated
115  panel of accessions from the Bioversity International *Musa* Transit Center (ITC) and recent collecting
116  missions. Then, by characterizing the mosaic patterns observed within cultivar groups and for isolated
117  accessions, we developed a catalogue of genomic diversity aiming to be dynamic that will be enriched as
118  banana germplasm characterization expand. This catalogue is to be used as a baseline of reference for
119  further cultivar group assignation for existing and new genebank materials, as well as for *on-farm* projects
120  or any type of work requiring the resolution of banana cultivar classification.

## *Materials and methods*

### *Plant material*

123  A collection of 317 banana accessions was selected for this study, as detailed in **Table S1**. The sources of
124  these materials were diverse: 264 samples were obtained from the Bioversity International *Musa* Transit
125  Centre (ITC) under the form of lyophilized leaves, 44 samples were collected during recent collecting
126  missions, 3 samples were obtained from on-farm projects and 4 reference samples were sampled *in-situ.*
127  Young leaf tissues of samples collected *in-situ* were silica dried on site. A set of 2 samples from the
128  Centre de Ressources Biologiques : Plantes Tropicales (CRB-PT) were obtained as public sequence data
129  from the study of Martin, Cottin, *et al.* (2023).
130  Significant curation work was undertaken on the taxonomical classification of accessions used in this
131  study to assign or exclude accessions from cultivar groups before genomic characterization. First,
132  passport data available from the *Musa* Germplasm Information System (MGIS - www.crop-
133  diversity.org/mgis) (Ruas *et al.* 2017) was retrieved. Second, and when available, morphological
134  characteristics and taxonomic expert recommendations (Taxonomic Advisory Group members of
135  MusaNet) obtained within the field verification exercise (Chase *et al.* 2014; Van den houwe *et al.* 2020)
136  were checked to correct or refine the classification of some accessions. When no recent field observations
137  were available, collecting mission reports were checked for collector observations in the field. This was
138  notably the case for the collecting mission reports to Papua New Guinea, Cook Islands, Samoa and
139  Tanzania (Arnaud and Horry 1997; De Langhe *et al.* 2001; Byabachwezi *et al.* 2005; Irish *et al.* 2016;
140  Sardos *et al.* 2017; Sachter-Smith *et al.* 2021; Sachter-Smith and Sardos 2021a; b) (**Table S1**).

141

5

142     *DNA Sequencing and genotyping*

143     This study spanned several years and included the preparation and sequencing of accessions in separate

144     batches. As a result, distinct but comparable genotyping technologies were utilized according to the most

145     effective approach at each point in time. For all experiments, DNA from each accession was extracted

146     following a CTAB protocol (modified from Risterucci *et al.* 2000). The libraries for restriction-site-

147     associated DNA sequencing were built with the PstI, or PstI/MseI restriction enzymes, followed by the

148     addition of barcoded adapters, DNA shearing, amplification, and sequencing. The sequencing data were

149     thus generated using either RADseq, ddRADseq or GBS techniques (**Supplementary Table S1**),

150     following the respective protocols established by Davey *et al.* (2010) or Elshire *et al.* (2011). For

151     RADseq, short-insert libraries (300–500 bp) were sequenced to produce 91 bp paired-end reads on an

152     Illumina HiSeq2000 (BGI, Hong Kong, China). For GBS and ddRADseq, libraries were sequenced as

153     150 bp paired-end reads on an Illumina HiSeq2500 (Genewiz, Azenta Life Sciences, USA) and Illumina

154     NovaSeq 6000 (LGC Genomics GmbH, Germany), respectively.

155     *Single Nucleotide Polymorphism (SNP) Callings*

156     After demultiplexing with GBSX (Herten *et al.* 2015), FASTQ files (one for each sample) were examined

157     with FastQC. We then used Cutadapt to clean them by eliminating Illumina adapter sequences and

158     trimming low-quality ends with a Phred score > 20 (Martin 2011). Any reads shorter than 30 bp after

159     post-trimming were removed. These reads were subsequently mapped using BWA-MEM (Li and Durbin

160     2010) to the *Musa acuminata* DH Pahang genome v4 (D'Hont *et al.* 2012; Belser *et al.* 2021),

161     downloaded on the Banana Genome Hub (Droc *et al.* 2022). Re-alignment was done with the

162     IndelRealigner module from GATK v4.1. We then followed the GATK pipeline recommended for a non-

163     model organism by adding the recalibration step. This consisted of performing an initial round of SNP

164     calling on the original uncalibrated data, selecting the SNPs with the highest confidence, and then

165     executing a round of base recalibration on the original mapped reads files. For duplicate samples, a script

166     using Sambamba software was used to merge the recalibrated bam alignment files. The GATK module

167     HaplotypeCaller v4.1 was then used for SNPs and indels calling. Finally, a script gVCF2vcf_gz.pl was

168     written to combine the individual gVCF files obtained into a single VCF file. The GenomicDB procedure

169     from GATK was used to build the gVCF SNP database, containing all the positions, variants and non-

170     variants. The snpcluster exclusion procedure was used to process SNP clusters, set for a threshold of three

171     or more SNPs per 10 bp window. The pipeline used to perform SNP analyses is available at

172     https://github.com/CathyBreton/Genomic_Evolution.

173

174 *Genome ancestry mosaic painting*

175 From the resulting VCF files, we used the scripts provided in the VCFHunter version 2.1.2 suite

176 (https://github.com/SouthGreenPlatform/VcfHunter). For each accession, we conserved two alleles by

177 sites with more than 10 reads and less than 1000, minimal frequency >0.05 were discarded (i.e.

178 vcfFilter.1.0.py MinCov:10; MaxCov:1000; minFreq:0.05; MinAl:3; RmAlAlt 1:3:4:5:6:7:8:9:10;

179 RmType SnpCluster). In the next step, we conserved the alleles in common with the alleles identified for

180 11 ancestral gene pools in Martin, Cottin, *et al.* (2023) (see Identification of ancestry informative alleles),

181 using the vcfSelect.py script. Since this dataset was previously obtained from the whole genome scale, it

182 was possible to intersect genome position of common SNP positions inferred from any genotyping

183 method and mapped on the same reference genome. Then, the allele ratio in individuals was calculated

184 with the allele_ratio_per_acc.py script, generating one file per accessions containing counted allele ratio

185 according to allocated ancestral gene pools (statistics in **Table S1**). When necessary, these files were

186 curated to define the ancestry mosaics of unresolved chromosome segments and to infer potential

187 haplotypes. Finally, genome ancestry mosaics SNP ratios and ancestry allocation were curated and

188 refined, and graphical visualizations were drawn using GeMo (Summo *et al.* 2022). Since the mosaics

189 were inferred using non phased data, the juxtaposition of segments represents introgressions at the

190 position but may not reflect the real haplotype of a given chromosome.

191 To characterize each cultivar group at molecular level, one accession with genotyping data of good

192 quality and no ambiguous or doubtful classification was selected as a reference (as shown in **Table 1**).

193 Then, patterns of other accessions were compared against the mosaic pattern of this reference accession.

194 During this comparison, certain accessions showed signs of aneuploidy, which could result from *in vitro*

195 conservation processes, especially in the cases of deletions or duplications of chromosomes or

196 chromosome arms (Breton *et al.* 2022). However, these aspects of aneuploidy are not elaborated upon in

197 this study or represented graphically in the catalogue (**Dataset S1**). Other events, usually smaller in size

198 and repeated in several accessions, such as small duplications and deletions, were considered ancestral

199 events that accounted for the creation of different patterns, as described in Martin, Cottin, *et al.* (2023) .

## *Results*

In our study, we carefully analysed 317 accessions, through genome ancestry mosaic painting, revealing genomic mosaic patterns at the chromosome level for each cultivated banana. Comparisons of patterns, combined with the curated taxonomical assignation of each accession, enabled the identification of reference patterns for cultivar groups as morphologically defined in the Musalogue (Daniells *et al.* 2001) and in De Langhe *et al.* (2001) for the Ilalyi group. Additionally, our analysis identified accessions that did not match with any cultivar group, both morphologically and genomically. If these unclassified accessions had unique genomic patterns or matched only with accessions known to be synonyms, i.e. same cultivars with different names, we treated them as individual accessions. If the same genomic pattern appeared in several accessions that were not synonyms, we considered them as clusters of morphological variants. The cultivar group patterns, clusters of accessions, and specific genomic patterns of individual accessions were organized and compiled into a catalogue (**Dataset S1**), as exemplified in **Fig. 1**. Nine remaining patterns for which morphological characterization is still ongoing were presented separately (**Dataset S2**).

Overall, the cultivar groups were well differentiated, and their genetic backgrounds were sufficiently discriminating to assign accessions to specific cultivar groups (**Table 1**). We identified a total of 83 unique mosaic patterns. These patterns corresponded to 31 previously defined cultivar groups (46 patterns; 256 accessions) and to 61 additional accessions or clusters of accessions (37 patterns). Out of the 27 cultivar groups for which more than one accession was available, 18 were homogeneous, i.e. they were composed of accessions with strictly identical genomic mosaics. Conversely, the 9 other cultivar groups displayed heterogeneity, i.e. they were composed of accessions with several mosaic patterns.

We observed that the genomic mosaics of the described cultivar groups exhibited a range of ancestral contributions (**Table 2**). The *M. acuminata* ssp. *banksii,* extended with the accessions 'Agutay' *(*ssp. *errans)* and 'borneo' *(*ssp. *microcarpa)* (referred to as banksii) is the only ancestral genepool for which centromeres were always present, with a minimum of 2 centromeres being observed in 7 patterns. *Musa acuminata* ssp. *zebrina* (referred to as zebrina)  and *M. schizocarpa* (referred to as schizocarpa) were consistently present across cultivar groups at least under the form of introgressions, corroborating the findings reported by Martin, Cottin, *et al.* (2023). However notable exceptions were observed in the cultivar group Klue Teparod (ABB) from mainland South-East Asia and the 'Auko' clones (ABB) from Papua New Guinea.  *M. acuminata* ssp. *malaccensis* (referred to as malaccensis) was also present in 35 patterns (24 cultivar groups). Then, the presence of previously unknown genepools, referred to as m1 and proposed to be *M. acuminata* ssp. *halabanensis* (referred to as halabanensis) in Martin, Cottin, *et al.*

8

233 (2023), and m2 (referred to as unknown) were present in 14 (11 cultivar groups) and 18 patterns (15

234 cultivar groups) respectively. *Musa acuminata* ssp. *burmannica,* including ssp. s*iamea* (referred to as

235 burmannica) was found to contribute only to the Klue Teparod cultivar group, along with banksii and

236 schizocarpa for the A haplotype. The B genome contributor, *M. balbisiana,* (referred to as balbisiana) was

237 included in 2/3 of the patterns within cultivar groups, with a minimum of 10 centromeres in 3 triploid

238 cultivar groups and a maximum of 34 centromeres in the pattern Pisang Awak-4x-3. These proportions

239 were in general in line with expectations based on the genomic constitution AB, AAB, or ABB. However,

240 and as noted in Cenci *et al.* (2021), its proportion varied from strict 1:2, 1:3, 2:3 or 3:4, considering the

241 presence of homoeologous exchanges between the A and B genomes. Several of the individual accessions

242 exhibited very distinctive patterns. It included accessions with a notable contribution from burmannica, or

243 cultivars with high contribution of zebrina, as well as tetraploids that exhibited a complete haplotype of

244 Australimusa (T) genome, now included in the former Callimusa section (Häkkinen 2013).

245

### *Homogeneous cultivar groups*

247 Homogeneity was observed across cultivar groups of various ploidy levels and genomic compositions.

### *Diploid cultivar groups*

249 With an AA genomic constitution, Pisang Jari Buaya (comprising 7 accessions) and Sucrier (8

250 accessions) were homogenous in their mosaic patterns. The patterns of the two accessions classified as

251 Mchare were identical, as well as for the two accessions identified as Pisang Lilin.

### *Triploid cultivar groups*

253 In the triploid cultivar groups with an AAA genomic composition, no variation was identified within the

254 18 accessions of Cavendish analysed. Similarly,, Gros Michel (3 accessions), Red (2 accessions) and

255 Ibota (2 accessions) were homogeneous. After curation, we identified 6 accessions from Tanzania

256 wrongly assigned to the Mutika/Lujugira group that corresponded to the Ilalyi group as described by De

257 Langhe et al. 2001 and genetically validated in Perrier *et al.* (2019). We therefore revived the Ilalyi group

258 that was previously removed from the passport data. These accessions shared the same mosaic pattern,

259 with an ancestral basis composed of a combination of zebrina and banksii similar to the one observed in

260 the Mutika/Lujugira group, but with an additional important contribution of malaccensis (5 centromeres)

261 (**Table 2**). Lastly, due to the availability of only one accession for each of the cultivar groups Ambon,

262 Rio, and Orotava, it was not possible to investigate potential variations within their respective mosaic

263 patterns. The study would benefit from more samples to confirm their monoclonal status.

264   In the triploid cultivar groups with an AAB genomic composition, Laknau (4 accessions), Mysore (5

265   accessions), Pisang Raja (2 accessions) and Iholena (6 accessions) were homogenous.

266   In the triploid cultivar groups with an ABB genomic composition, the Pelipita (3 accessions) and Klue

267   Teparod (2 accessions) groups were genetically uniform while a wider sampling remains necessary to

268   validate this observation. Finally, we identified 2 patterns corresponding to 4 cultivar groups with a

269   shared A genome background. The Bluggoe group (10 accessions) and the Monthan group (5 accessions)

270   exhibited identical genetic mosaic patterns, despite differences in morphologies. For instance, Bluggoe

271   fruits are mostly straight and horizontal or slightly erect, while Monthan's curve upwards. Similarly, Ney

272   Mannan (7 accessions) and Peyan (1 accession) groups also shared the same mosaic pattern, although we

273   noted slight differences in the levels of heterozygosity within the *M. balbisiana* haplotypes. A larger

274   sample of Peyan representatives and better discrimination of allelic diversity in the B genome would be

275   necessary to provide clearer insights.

276

### *Heterogeneous cultivar groups*

278   Two types of heterogeneous cultivar groups could be distinguished. The first type, which includes cultivar

279   groups such as Plantain and Mutika/Lujugira, displayed mosaic patterns differentiated by only small

280   chromosomal region showing variations in allelic ratio. In the second type of heterogeneous cultivar

281   groups, we found two or more mosaic patterns, each exhibiting multiple differences likely resulting from

282   different mechanisms of diversification.

### *Mutika/Lujugira*

284   Mutika/Lujugira is a triploid cultivar group with a AAA genomic constitution that is typical to Burundi,

285   Uganda, Democratic Republic of Congo, Cameroon, Kenya, Rwanda and Tanzania. For this cultivar

286   group, we conducted an important curation of the passport data, notably by consulting the collecting

287   mission reports when available. In this set of 34 AAA accessions from East Africa, we identified four

288   nearly identical mosaics corresponding to 24 accessions, including well described Mutika/Lujugira

289   cultivars such as the popular 'Mbwazirume' (Shepherd 1957). These mosaics' main contributors were

290   zebrina, banksii and schizocarpa. A pattern variation was observed in 'Guineo', 'Intokatoke' and

291   'Makara' in which a small interstitial region of the second arm of chromosome 10 displayed a mitotic

292   homologous exchange between one of the zebrina haplotypes and the banksii haplotype. In addition, and

293   on the same chromosomal region, the 'Siira' accession exhibited a small deletion on the banksii

294   haplotype. Finally, the accession 'Mbwazirume' also appeared to be a variant with a switch of the allelic

295   ration resulting from a mitotic recombination on the first telomere of chromsome 10 (**Fig. S1**). No

296     correlation was found between these variations and the proposed clonesets from Karamura *et al.* (2010).

297     Finally, three additional mosaics discovered in this set were not assigned to Mutika/Lujugira but

298     corresponded to Tanzanian accessions from the homogeneous Ilalyi group described earlier.

### *Plantain*

300     The 58 Plantain accessions of the sample were remarkably homogeneous, but a small variation was

301     detected on chromosome 10 between ~8.5 Mb and 13 Mb. We interpreted this change as a diploid region

302     (balbisiana – banksii) resulting from a small deletion of one of the banksii haplotypes present in the

303     original pattern. This predominant variation was detected in 42 accessions out of 49, for which this region

304     could be characterized. Interestingly, the 4 accessions with origins in Asia, 'Bungaoisan' (a medium

305     French) and 'Daluyao' (Medium True Horn) from the Philippines, as well as 'Mantreken' from Indonesia

306     and 'Nendran' (French) from India all had the original mosaic pattern without deletion. The three other

307     accessions with a mosaic without the deletion were 'Agbagba" from Nigeria (a medium false-horn

308     Plantain widely cultivated in West Africa according to Adheka *et al.* 2013), 'Big Ebanga' (a giant false

309     horn from Cameroon, possibly synonym of Agbagba), and Maiden Plantain (a French Plantain received

310     from Honduras but of unknown origin)

### *Pome*

312     Pome cultivars are AAB triploids that originated in India and are now very popular in Brazil and Hawaii

313     under the name Prata. Ten accessions of the sample displayed a mosaic pattern associated with the Pome

314     group, within which three pattern variations were observed. The first one, identified in 6 accessions from

315     different countries, shows one deviation from a pure AAB pattern, i.e. an A/B recombination (A3:B0) on

316     the first telomere of chromosome 3. The second one, present in 3 Pome accessions received recently from

317     India, also shows an A/B recombination (A3:B0) but in the interstitial region of the first arm of

318     chromosome 9. This recombination is also present in the third pattern, in addition to another one on the

319     first telomere of chromosome 10 **(Dataset S1)**. This last pattern was identified in one accession from

320     Australia, 'Lady Finger (Nelson)', sometimes referred to as belonging to the Nadan group in other

321     collections and which is tetraploid for chromosomes 8. Except for the extra chromosome 8 of Pome 3

322     which has an *M. acuminata – M. schizocarpa* ancestry, the variations observed between the three patterns

323     are linked to A-donor introgressions in B chromosomes. Since all these introgressions correspond to

324     genepools also present in one of the two A genomes, it is difficult to assess whether the three Pome

325     patterns were derived clonally from each other or were obtained through different sexual events.

326     However, the banksii introgression observed on the B chromosome 9 of Pome-2 and Pome-3 occurs

327     frequently in cultivars of AB, AAB or ABB genomic constitutions, suggesting a common ancestry. This

328    pattern is more likely to have been inherited sexually rather than arising from a new, independent mitotic

329    recombination. The observed variations may have resulted from a combination of clonal diversification

330    and sexual events (at least two from similar parents).

331    *Maia Maoli/Popoulu (MMP)*

332    After curation of passport data and a morphological trait check, we identified 16 accessions affiliated to

333    the Maia Maoli/Popoulu group corresponding to 4 genomic patterns. The Maia Maoli/Popoulou patterns,

334    with an AAB genomic composition, are characterized by contributions from banksii, schizocarpa, and

335    zebrina for the A genome, with little to no presence of malaccensis. A striking feature of these four

336    patterns is the absence of B centromere and the presence of two S centromeres on chromosome 2 (**Table

337    2**, **Dataset S1**). Out of the 16 accessions, MMP-2 was the most frequent pattern with 12 representatives

338    from both Polynesia (Cook Islands, Samoa and Hawaii) and Melanesia (Bougainville and New Britain

339    Islands in Papua New Guinea). Three accessions from Cook Islands, Tahiti and Samoa exhibited the

340    pattern MMP-1. The differences observed between MMP-1 and MMP-2 are slight. The pattern MMP-1

341    has a small balbisiana introgression in the interstitial region of the first arm of chromosome 1, and a

342    duplication of the balbisiana first telomere on chromosome 7. The patterns MMP-3 and MMP-4 were

343    identified in one accession each, 'Mango Torotea' from Cook Islands and 'Lavugi' from New Britain

344    Island, respectively. These patterns are significantly different from MMP-1 and MMP-2 with more than

345    10 discriminating regions each. They also differ from each other by 15 events. If MMP-1 and MMP-2

346    may be derived from each other by clonal diversification, this is likely not the case of MMP-3 and MMP-

347    4, which may have resulted from independent sexual events with similar parental contribution.

348    *Silk*

349    Two closely related patterns were detected in the 15 accessions classified as Silk confirmed previous

350    findings (Sardos *et al.* 2016). Eleven accessions were displaying the pattern Silk-1 while 4 displayed the

351    pattern Silk-2. Interestingly, the Silk-2 accessions were all collected in Africa (Burundi, Tanzania,

352    Congo). The differences observed between the two Silk genomic mosaic patterns were important.

353    Notably, 5 of their 33 centromeres were of different origins. For example, on chromosome 5 both Silk

354    groups display one B chromosome, but Silk-1 displays two banksii centromeres while Silk-2 displays one

355    banksii and one zebrina centromere (**Table 2**). These differences cannot result from clonal diversification.

356    Therefore, the diversity observed within the Silk group results from two different sexual events, probably

357    from parents of similar genetic background.

358    *Kalapua*

359     The Kalapua group, characterized by an ABB genomic composition, is a popular cooking banana variety

360     in Papua New Guinea. Within our sample set of five Kalapua accessions, we identified two distinct

361     genomic mosaics. The first mosaic pattern was present in three of the samples, while the second pattern

362     was found in the remaining two. The observed differences between these mosaics consist of varying

363     proportions of A and B genomes in four specific genomic regions. The first telomere of chromosome 4

364     and the interstitial region of the first arm of chromosome 8 are A2:B1 in Kalapua-1 and A1:B2 in

365     Kalapua-2. In addition, on chromosome 9, the first telomere is A1:B2 in Kalapua-1 and A0:B3 in

366     Kalapua-2 while the second telomere is A2:B1 in Kalapua-2 and A1:B2 in Kalapua-1. Kalapua patterns

367     may have resulted from the accumulation of mitotic homoeologous chromosome exchanges. However,

368     mitotic recombination events are rare and these four cumulated events may alternatively have resulted

369     from two sexual events among similar parents.

370     *Pisang Awak*

371     The Pisang Awak group comprised two triploid and three tetraploid patterns present in our sample. For

372     the triploid patterns, the pattern Pisang-Awak-1 (PA-1), was found in 7 accessions while Pisang-Awak-2

373     (PA-2) was found in 2 accessions from India. Eight variations in the patterns of homologous exchanges

374     between the A and B genomes were observed between these two mosaic patterns. Differences in A/B

375     homoeologous exchanges consisted either in the presence or absence of these events or in variations in the

376     size of common events. This finding is not consistent with two genotypes deriving from each other

377     clonally and rather supports the idea that they were both sexually produced, probably from the same AB

378     parent who produced recombined but unreduced (2x) gamete. Furthermore, A/B homoeologous

379     exchanges enabled to hypothesize a pedigree relationship with the three tetraploids patterns identified in

380     this cultivar group (4 samples). These four accessions, with ABBB genomic composition, were

381     morphologically included in the Pisang Awak group and were produced from unreduced (3x) ABB

382     triploid Pisang Awak gametes crossed with a haploid (1x) B gamete. The A/B introgressions patterns

383     observed in the triploid and tetraploid Pisang Awak samples support pedigree relationships between PA-1

384     and 'Ramu Yawa' (Pisang-Awak-4x-1) from Papua New Guinea. Equally, direct ancestry can be inferred

385     between PA-2 and 'Pisang Awak' (Pisang-Awak-4x-2) from Sri Lanka.  The third tetraploid pattern,

386     discovered in 'Foulah 4' and 'Nzizi' (Pisang-Awak-4x-3) from Ivory Coast and Nigeria respectively, did

387     not correspond to any of the triploid Pisang Awak described, suggesting that at least a third triploid form

388     may exist or may have existed (**Fig. 2**).

389

390     *Kunnan and Ney Poovan*

391    Two cultivar groups from India and with AB genomic composition are defined, the Ney Poovan and the

392    Kunnan groups, but their morphological characteristics are not clear. In our sampling, 8 accessions with

393    an AB genomic composition could be affiliated to either cultivar group. These accessions displayed four

394    different mosaic patterns with at least one A introgression into their B genome (consistent with Cenci *et*

395    *al.* 2021), and with a malaccensis ancestry dominance as A-donor genome. Two patterns were discovered

396    in both Ney Poovan (3 accessions) and Kunnan (4 accessions) but the correspondence between the

397    morphological assignation and the patterns was incomplete. Since these cultivar groups were not

398    extensively documented and many synonyms and overlaps exist in India, the true assignation of the

399    accession confusingly named 'Kunnan' (ITC1034) but classified as Ney Poovan, was difficult to assert.

400    *Other cultivars (clusters and individual accessions)*

401    Some accessions of the set were ambiguous in classification, with morphological similarity with well-

402    known cultivar groups, but not complying enough to discriminating criteria to be considered as part of

403    these defined cultivar groups.

404    *Similar to Mutika/Lujugira*

405    Four accessions were collected in Tanzania with unclear classifications (De Langhe *et al.* 2001) and

406    comprise two distinct mosaic patterns that share similarities with the Mutika/Lujugira group. Notably,

407    they include a significant malaccensis component (12 centromeres and 6 centromeres, respectively),

408    similar to what was observed in the Ilalyi group (**Table 2**). The first pattern, Kikundi, is identified in three

409    accessions. 'Ntebwa' and 'Ntindi I', both from the Tanzania's Usambara region differ in their uses,

410    'Ntebwa' is used for cooking, and 'Ntindii I' serves both as a cooking (flour) and dessert banana. The third

411    accession 'Kikundi' differs by the pinkish colour of the pseudostem contrasting with the green observed

412    in the two others. The second pattern, Luholole, is represented by a single accession from the Morogo

413    district of Tanzania.

414    *Similar to Plantain*

415    Two accessions can be linked morphologically to the Plantain group. The accession 'Kupulik' was

416    collected in the late 1980's in Papua New Guinea (Island of New Ireland in the Bismark Archipelago) as a

417    Horn type Plantain but it is not a Plantain. Its mosaic shares similarities with both Plantain and Iholena,

418    but the malaccensis component is absent in 'Kupulik'. Two other accessions originating from Papua New

419    Guinea, 'Bubun' and 'Navente 2', exhibited the same pattern. Then, the cultivar **'**Mnalouki' from the

420    Comoros (Perrier *et al.* 2019), shares two haplotypes with Plantain cultivars and was proposed to be a

421    progeny of a Plantain (2x gamete) x Mchare (1x gamete) (Martin, Baurens, *et al.* 2023).

422     *Similar to Iholena*

423     Some level of morphological confusion exists around the Iholena group (Arnaud and Horry 1997; Kagy *et*

424     *al.* 2016; Sachter-Smith *et al.* 2021; Sachter-Smith and Sardos 2021a).   In our set, five accessions sharing

425     some, but not all, morphological features of the Iholena exhibited two different and distinct mosaic

426     patterns. The accessions 'Rukumamb Tambey', 'Tigua' and 'Balabolo 1' form the Rukumamb Tambey

427     cluster. They share the bunch shape and the colour of the flesh with Iholena, but their fruits don't turn

428     yellow when ripe and the lower surface of their new leaf is green. The accessions 'Arawa' and 'Mamae

429     Upolu' displayed a second mosaic pattern and were also different in their morphology (notably 'Arawa',

430     which had an overall more diploid look at collect). 'Mamae Upolu', collected in Samoa, differs from

431     Iholena by its slightly more upward fruits, the green lower surface of the new leaf and the more yellow

432     colour of the flesh. Despite their morphological proximity, these two mosaic patterns differ from Iholena

433     by the ancestry of 5 and 6 centromeres, respectively, and the notable presence of one malaccensis

434     centromere that is absent in Iholena (**Table 2**).

435     *Similar to Maia Maoli/Pōpōulu*

436     We observed 6 bananas accessions that morphologically resemble the Maia Maoli/Popoulu (MMP)

437     cultivars but with 3 different patterns. They were all collected in Papua New Guinea and surrounding

438     islands and are composed of three different mosaic patterns that share a common background with the

439     four MMP patterns previously identified and with other AAB cooking bananas. The first pattern, named

440     here Wan Gevi, is composed of 2 accessions. The second pattern is represented by a unique accession

441     'Buka Kiakiau'. Unlike the two other patterns, the third pattern named Ruango Block and made of 3

442     accessions, contains malaccensis as contributor (4 centromeres).

443     *Clusters of accessions with distinctive morphotypes.*

444     We noted in our set some clusters of accessions, that may correspond to morphological variants of a same

445     genomic pattern. It was notably the case of three sets of diploid accessions. The first set, composed of

446     three accessions from the Philippines and Malaysia, was called here the Bata-Bata cluster. The second set

447     was composed of four accessions from Papua New Guinea and was named here the Te'engi cluster. Then,

448     the Talasea cluster was composed of two accessions collected in Papua New Guinea outer islands, one

449     being likely the reddish variant of the other. Two morphological variants were also observed in the ABBT

450     Buka cluster, 'Buka' being a shorter variety than 'Bukayawa'.

451

### *Individual profiles*

452

453    Finally, we listed in the catalogue individual accessions which cannot be morphologically assigned to
454    described cultivar groups and which have specific genomic patterns. Several diploids of AA genomic
455    compositions presented interesting characteristics. The cultivar 'Khai Na On' is the male (1x gamete)
456    parent of the 'Gros Michel' group (Raboin *et al.* 2005; Hippolyte *et al.* 2012; Martin, Cottin, *et al.* 2023).
457    The cultivar 'Pisang Madu' is the keystone to the genome ancestry mosaic painting approach as it enabled
458    the discovery and the identification of diagnostic SNPs for two uncharacterized ancestors of cultivated
459    bananas and it is related to the Cavendish group. Its genome is indeed composed of a full haplotype of the
460    unknown ancestor m1, proposed to be halabanensis from Indonesia (Martin, Cottin, *et al.* 2023).
461    Additionally, it also contains 6 centromeres belonging to the unknown ancestor (Sardos *et al.* 2022;
462    Martin, Cottin, *et al.* 2023). 'Manang', an accession from the Philippines, has a significant contribution
463    from burmannica (4 centromeres), same as 'Matti' from India, also included in the catalogue with nearly a
464    full haplotype of burmannica.

465    For the triploids, the 'Chuoi Mit' cultivar from Vietnam exhibits an ABB genomic composition, featuring
466    an A genome that closely resembles the A genome found in the Bluggoe/Monthan pattern. The proximity
467    of these cultivar groups has already been reported based on homoeologous exchanges between A and B
468    subgenomes (Cenci *et al.* 2021) and is now confirmed analysing the A mosaic pattern. 'Pisang Slendang',
469    an Indonesian AAB cultivar, shares one of its A haplotypes with both the Bluggoe and Monthan groups,
470    as well as with 'Chuoi Mit'. 'Chuoi Xi Mon' bears an unidentified genome differing from the
471    characterized Unknown genepool.

472    The cultivar 'Auko' (synonym 'Vunapope') from Papua New Guinea, with an ABB genomic
473    composition, has a unique A mosaic pattern among all cultivated bananas (no zebrina ancestry). It is only
474    made of banksii and schizocarpa. This finding supports the hypothesis of the early domestication of
475    banana in New Guinea from where *M .a.* ssp *banksii* and *M. schizocarpa* originated (Carreel *et al.* 2002)
476    and as recently showed by chromosome painting (Martin, Cottin, *et al.* 2023). On the other side of the
477    spectrum, the cultivar 'La' from Vietnam with an AAB genomic composition has 14 zebrina centromeres
478    and only one small schizocarpa introgression. The 'Ya Ta Na Thin kha' accession collected in Myanmar
479    was provided with a poor classification (*Musa*) and was identified as a triploid ABB in our analysis. Its
480    genomic composition is rich in burmannica, like the Klue Teparod group. However, the pattern of 'Ya Ta
481    Na Thin kha' is different, notably with a small introgression of zebrina in the first arm of chromosome 9
482    that is absent in Klue Teparod which exhibits a balbisiana introgression in the same region. These two
483    patterns shared recombination breakpoints, indicating common evolutionary history, even suggesting that

16

484    'Ya Ta Na Thin kha' could be one of the genotypes at the origin of this cultivar group (**Fig. 3**). However,

485    in the absence of morphological description available, we were not able to assess whether 'Ya Ta Na Thin

486    kha' belongs to the Klue Teparod group.

487    Until recently, only one triploid cultivar with a full S haplotype was known. The cultivar 'Toitoi' was

488    collected in the island of Bougainville in Papua New Guinea. Its unrecombined schizocarpa haplotype

489    suggested it resulted from an unreduced (2x) gamete with an AA genomic composition like many diploids

490    AA cultivated in the country and a regular (1x) gamete S, probably from a wild specimen of *M.*

491    *schizocarpa* endemic to New Guinea. A second cultivar with an AAS genomic composition named

492    **'**Waga' was found since then, still in Papua New Guinea, and has a different mosaic pattern which

493    interestingly shows two *M. acuminata* ssp. *banksii* introgressions on the chromosome 4 of the *M.*

494    *schizocarpa* haplotype that probably results from a different type of cross (**Dataset S2**). Several tetraploid

495    cultivars with a T genome were discovered in Papua New Guinea. However, the SNPs assigned to T are

496    representative of the whole former Australimusa section of the *Musa* species, from which arose the Fehi

497    bananas, with no indication on the specific species or genepool involved. Among these tetraploid

498    cultivars, 'Kalmagol' which morphologically resembles the Silk group, had a genomic composition of

499    AABT with the AAB genome that could correspond to the pattern Silk-2. Among the three patterns with

500    an ABBT genomic composition, the cultivars 'Buka' and 'Bukayawa' displayed an ABB genome that

501    could have derived from the Pisang-Awak-1 pattern (**Fig. 2**) while the cultivar 'Bengani' likely derived

502    from the pattern Kalapua-2. In Cook Island, 'Rekua', a second ABBT cultivar with an ABB genome like

503    the Kalapua patterns was discovered. It is different from 'Bengani', and its ABB genome may have arisen

504    from Kalapua-1.

505

## *Discussion*

### *Clonal diversification at the genomic scale*

508    Domestication of banana was a gradual process in which the selection for edible pulp  led to today's

509    parthenocarpic and highly sterile cultivars (Simmonds 1966). In this scenario, cultivar groups were

510    expected to correspond to cultivars clonally derived from each other with the clonal accumulation of both

511    point mutations and epigenetic variations as main mechanisms of diversification (Simmonds 1966). For

512    example, despite the extremely high levels of intra-cultivar group phenotypic diversity observed in the

513    Plantain and Mutika/Lujugira groups (Tézenas Du Montcel *et al.* 1983; Karamura *et al.* 1998), they were

514    both found genetically homogenous (Noyer *et al.* 2005; Kitavi *et al.* 2016), but with significant levels of

17

515    epigenetic variations (Noyer *et al.* 2005; Kitavi *et al.* 2020). Here, we identified larger genomic variations

516    in these cultivar groups, fixed through vegetative propagation. Specifically, deletions were observed in

517    the Mutika-3 and Plantain-2 patterns, and another event interpreted as a homologous exchange due to

518    mitotic recombination, was inferred in the Mutika-2 and Mutika-4 pattern. The patterns Plantain-2 and

519    Mutika-2 are in accordance with the findings of Martin, Cottin, *et al.* (2023) for these two cultivar groups

520    but the high number of accessions analysed in our set enabled the discovery of  other variants. In the

521    Plantain group, the deletion, frequently observed in African accessions, is absent in the Asian accessions,

522    suggesting that this event occurred after the introduction of the first Plantain cultivar(s) in Africa and

523    supports previous hypotheses (Perrier *et al.* 2011; Langhe *et al.* 2015) on the origin of this cultivar group

524    in South-East Asia. For the Mutika/Lujugira cultivars studied here, the two genomic variants

525    characterized on chromosome 10 were found in only a small portion of the samples. However, most of the

526    Mutika/Lujugira analysed were introduced to the ITC from Rwanda and Burundi and may not represent

527    the entire diversity of this cultivar group. Further genomic characterization of the Mutika/Lujugira

528    germplasm across a wider geographical range is recommended for future studies.

529    Within those two cultivar groups, we did not observe obvious correlations between the small genomic

530    variations detected and the striking phenotypic features of these cultivar groups, a pattern also observed

531    with epigenetic variations (Noyer *et al.* 2005; Kitavi *et al.* 2020). However, these events constitute

532    valuable markers for tracing the evolutionary history of Plantain and Mutika/Lujugira. Additionally, these

533    genomic variations notably generate gene copy number variations, as identified in multiple crops

534    (Yakushiji *et al.* 2006; Stein *et al.* 2017; Gabur *et al.* 2019), and may be linked to interesting traits, such

535    as diseases resistance, as identified in a few somaclonal variants of Cavendish exhibiting deletions on

536    chromosome 5 (Hou *et al.* 2022). Therefore, the specific regions of chromosome 10, where structural

537    variations were identified in the Mutika/Lujugira and Plantain groups, merits further investigation.

### *Sexuality still matters in cultivated bananas*

539    The diversity of genome patterns observed within several cultivar groups, including well-known cultivar

540    groups such as Silk, Pome, Maia Maoli/Popoulu and Pisang Awak, seemed difficult to explain only by

541    clonal diversification. The variations observed, with centromeres of different origins and/or the

542    accumulation of high numbers of recombination, rather supports a meiotic origin to these differences.

543    However, it is striking that the different patterns observed in these cultivar groups are so similar that they

544    might be siblings from the same parental clones, as inferred for the triploid Pisang Awak accessions (**Fig.**

545    **2**). This important finding supports Kagy *et al.* (2016) who proposed an enlarged vision of cultivar groups

546    and considered that sets of closely related cultivars could ensue from different sexual events within

18

547  similar or closely related parents. In addition, the Pisang Awak example, with the occurrence of tetraploid

548  siblings of triploid landraces, illustrates that original clones can also be sources of sexual diversification

549  within cultivar groups (**Fig. 2**). Pisang Awak, are more prone to set seeds (Simmonds 1962) and the

550  tetraploid cultivars observed here show that farmers continue to select new banana cultivars that arise

551  accidentally from seeds, further contributing to expanding diversity.

552  Interestingly, several of the accessions not affiliated to any cultivar group brought interesting insights by

553  showing the progressive incorporation of exotic genepools. Notably, in East Africa where the Ilayi group

554  and two other patterns, Kikundi and Luholole, display a common genomic background with the

555  Mutika/Lujugira group characterized by an important contribution of banksii and zebrina, but

556  supplemented by malaccensis. Such pattern was also observed in 'Mnalouki', when compared to Plantain

557  as well as in the Ruango Block cluster when compared to Maia Maoli/Popoulu and to some extent in

558  Rukumamb Tambey and Arawa clusters when compared to Iholena. Considering that banana's

559  domestication centre was likely located in New Guinea island (Martin, Cottin, et al. 2023) where only *M.*

560  *acuminata* ssp. *banksii* and *M. schizocarpa* can be found, the genomic constitution of 'Auko', free of

561  zebrina genepool, suggests that the addition of zebrina to the genomic backgrounds of cultivars likely

562  resulted from secondary diversification events. However, the common contribution of zebrina to all other

563  cultivars suggests that the addition of zebrina precluded the insertion of malaccensis in banana cultivars.

564  This scenario is consistent with the geographic distribution of *M. acuminata* subspecies and is in line with

565  the correlation observed between the wild subspecies geographical ranges and the wild ancestors'

566  contribution to local cultivars (Sardos et al. 2022; Martin, Cottin, et al. 2023). Therefore, these accessions

567  related to known cultivar groups but with additional contribution of malaccensis may result from more

568  recent sexual diversification, such as 'Mnalouki' which may be a sibling of Plantain (Martin, Baurens, et

569  al. 2023).

570  In addition, the tetraploid accessions with a T haplotype that were collected in the Pacific showed the

571  incorporation of a supplementary genepool into cultivars from the Silk, Pisang Awak and Kalapua

572  groups. Interestingly, the Silk and Pisang Awak groups originated in India and South-East Asia

573  respectively, while wild and cultivated specimen of the former Australimusa section can be found only in

574  an area going from Sulawesi (east Indonesia) to the Pacific Islands. It shows that these hybridizations

575  occurred more recently, after the introduction of these cultivar groups in the distribution range of the ex-

576  Australimusa specimens. It illustrates the importance of conserving local genetic resources as they can

577  still be active in crop diversification.

578  ***Implications for farming system, taxonomy, breeding and conservation***

19

579 *Farming systems*

580 In clonal crops, vegetative propagation is an efficient way to preserve and multiply favourable genotypes

581 that would not be maintained through sexual reproduction (McKey et al. 2010). For banana, farmers have

582 historically selected and preserved varieties upon noticing changes in traits in the field, whether clonal

583 (Karamura et al. 2010) or resulting from residual sexual events (De Langhe et al. 2010; Martin, Baurens,

584 et al. 2023). Our hypothesis of genomic variations with two origins, clonal and sexual, co-existing in the

585 overall diversity of cultivated bananas have implications in a context where monoclonal agriculture puts

586 banana cultivation at risk in the face of biotic and abiotic stresses. These variations are valuable sources

587 of diversity that can be overlooked by farmers. As stated before, deletions and duplications are sources of

588 gene copy number variations that can result in differential phenotypes, just as homologous exchanges do.

589 The use of this intra-cultivar group diversity would be an innovative way to diversify agrosystems by

590 ensuring the co-existence of clonal and sexual variants in farmer cultivar portfolios. Since cultivar

591 adoption is affected by a combination of sensory characteristics, agronomic properties and environmental

592 and socio-cultural factors affecting production (Madalla 2021), planting different genomic patterns

593 associated to the same cultivar group could allow to overcome part of these constraints while introducing

594 diversity in farmers' fields. Equally, the cultivars sharing morphological characteristics with known

595 cultivar groups and which were found hybridized or introgressed with exotic genepools, could allow the

596 introduction of additional genetic diversity, hence with putative beneficial new traits, while enhancing the

597 likelihood of acceptance by farmers.

598 *Classification of cultivated bananas*

599 Our results show that intra-cultivar group diversification is made of a combination of sexual and clonal

600 diversification and pleads for a relaxed definition of the cultivar group concept as the set of closely

601 genetically related individuals sharing peculiar morphological characteristics. Classification criteria

602 should be revised combining morphological assessment and chromosome painting results, for example

603 using genomic determination keys (**Fig. S2**).

604 Our findings also raise new questions about the current taxonomy. For example, it may not be accurate to

605 keep Bluggoe and Monthan as two separate cultivar groups while they share seemingly identical genomic

606 backgrounds. Equally, the community should consider the creation of new cultivar groups as (clusters of)

607 new genomic patterns are discovered. In some cases, cultivar groups may also be enlarged in a way they

608 would incorporate the "grey zones" of cultivars resembling the "core" accessions of the cultivar groups

609 but differing in their genomic patterns.

610     In such revision effort, some accessions would remain alone, constituting de facto cultivar groups with
611     only one cultivar, this being relative to the current sample and possibly revisited with the addition of new
612     cultivars. The catalogue presented here (**Dataset S1**), conceived as a dynamic and evolutive document,
613     constitutes a valuable supporting tool for this task.

### *Breeding requires to maximize diversity*

614

615     The chromosome painting approach was shown to be useful to understand the formation of current
616     varieties (Martin, Baurens, *et al.* 2023), an essential point to ease successful breeding. It could also be a
617     useful tool for breeders, both to support the selection of parents and the selection of hybrids (Cenci *et al.*
618     2023). The catalogue presented here, and the association of the patterns discovered in active genebank
619     accessions, opens the door to an optimized use of banana diversity for breeding crosses. The occurrence
620     of numerous individual accessions that cannot be affiliated to existing cultivar groups and that display
621     unique and peculiar genomic background presents a fresh perspective for the use of original accessions as
622     parents. Additionally, the intra-cultivar group diversity could also be a valuable resource for breeding.
623     The selection of parents among the different clones or variants within a given cultivar group could allow
624     the incorporation of new genetic variation in existing breeding schemes and may result in the
625     incorporation of possible useful traits in the obtained progenies.

### *Let's keep characterizing and collecting*

626

627     Methods using chromosome characterization based on ancestral origin (mosaic genomes) were proven
628     efficient to support germplasm characterization (Santos *et al.* 2019; Ahmed *et al.* 2019; Martin *et al.*
629     2020; Wu *et al.* 2021). The present study aimed at providing an efficient tool to support the community of
630     banana researchers and workers in the task of classifying cultivars. This tool was also found helpful for
631     the resolution of taxonomical issues that may arise. In this study, we found that about 25% of the
632     taxonomic information displayed in the existing passport data required corrections, or clarifications when
633     details were missing (**Table S1**). This catalogue also constitutes a tool to support the routine management
634     of banana genebanks through molecular characterization. Combining high throughput genotyping with
635     this tool constitutes a much faster way to classify germplasm when compared to morphological
636     characterization. Additionally, creating a baseline by genotyping germplasm that enters collections can
637     help track *in-vitro* induced aneuploids (found here but data not shown), synonyms and potential
638     duplicates.

639     In this study we could not characterize the entire banana genepool. For example, the chromosome
640     painting results obtained for the many unique diploid cultivars from Papua New Guinea are not presented

641    here but may be subject to a separate study. Also, the characterization of the popular Saba cultivars from

642    the Philippines was not conducted due to lack of material available in the genebank. Nevertheless, this

643    catalogue constitutes a baseline that can be further enriched as prospections and genomic characterization

644    continue. An online version aims to be dynamic, continually expanded, and updated as characterizations

645    of the ancestral gene pools and sequencing technologies improve.

646    Importantly, we provided evidence for the richness of patterns identified in a small number of accessions

647    and that much more diversity exists in regions that have not been explored or have been underexplored.

648    Screening only a few new accessions from recent collecting missions was sufficient to discover new

649    genetic profiles that did not match defined cultivar groups, maybe justifying the creation of new cultivar

650    groups, or blurring the lines of the defined ones. Now that germplasm characterization has entered the

651    area of genomics, it is more than likely that further prospection of banana diversity in farmers' fields will

652    enable the discovery of new variants and new genotypes. Obviously, gaps remain in our perception of

653    banana diversity and additional collecting missions are necessary to enrich our understanding of banana

654    diversity and to fill the gaps in the collections.

## *Data availability*

656    The sequencing reads were deposited in the NCBI SRA associated with the BioProject PRJNA450532.

657    SNP datasets were recorded in a database browsable via a web application https://gigwa.cgiar.org/gigwa

658    (Sempéré *et al.* 2019; Rouard *et al.* 2022).

659

## *Acknowledgments*

672  renowned expert in banana research, significantly enriched the field through his extensive collection of
673  banana plants and laid the groundwork for the Musa Germplasm Information System. This work could
674  not have been possible without their dedication to banana germplasm collection and conservation.

## *Authors contributions*

676  Conceptualisation, JS, AC and MRo;  Provided material, IVdH, JP, WW; DTH;  Library preparation and
677  GBS genotyping: RR; Passport Data Curation, JS, CJ, RC, MRu, YM, GSS, MRo; Methodology and
678  bioinformatic Analysis, GM, CB; Mosaic analyses and curation, AC, JS; Catalogue design, MRo, JS,
679  Online catalogue development, VG; Data release in MGIS/Gigwa, CB, MRo; Writing – Original Draft
680  Preparation, JS, MRo; Writing – Review & Editing, XP, GM, NY, CJ, RC, GSS, AC, NR; Funding
681  Acquisition, AD, NR; All authors have read and agreed to the published version of the manuscript.

682 *Legends*

683 *Figures*

684 **Fig. 1.** *Example of the Red cultivar group in the catalogue. Each entry is divided into 3 sections: Passport*
685 *data, Pictures and Molecular characterization. The Passport data section includes basic information as*
686 *the biological status, ploidy, main distribution area and uses with notes on the cultivar group and*
687 *genomic features. Up to 3 pictures are intended to be representative of the cultivar group. The molecular*
688 *characterization contains from 1 to 5 mosaic patterns with the name of the reference accession(s) used*
689 *for the mosaics painting. Coloured segments show the contribution of each of the ancestral genepools*
690 (Martin, Cottin, *et al.* 2023)*.*

691 *Fig. 2.* **Genetic Diversification in the Pisang Awak Group.** *This illustration showcases the patterns of*
692 *five Pisang Awak variants and a Pisang Awak-derived intersectional hybrid for four sets of chromosomes*
693 *(chromosomes 4, 5, 6 and 7), each marked by distinctive events (red circle) inherited from one of the*
694 *parents (other differences may result from recombination events that occurred during the production of*
695 *unreduced gametes (3x). The integration of additional genomes into the triploid (3x) patterns (Pisang*
696 *Awak-1/3), denoted by specific letters (B for balbisiana, represented in black; T for Australimusa, shown*
697 *in yellow), has resulted in the formation of closely related tetraploid (4x) varieties. The triploid Pisang*
698 *Awak-3 is represented with partial transparency as it was not found in the sample but could be inferred*
699 *from its progeny. Coloured segments show the contribution of each of the ancestral genepools.*

700

701 **Fig. 3.** *Genome ancestry mosaic painting applied to the Klue Tiparod (ABB, 3x) and 'Ya Ta Na Thin kha'*
702 *(collected as Musa* spp*), revealing related pattern and a possible pedigree relationship. The colours of*
703 *segments correspond to ancestral contributions (black: M. balbisiana, green: M. acuminata banksii*
704 *genetic group, and orange: M. a. burmannica* including ssp. *siamea).*

705

706 *Tables*

707 **Table 1.** Overview of banana cultivar groups and their genomic compositions, alongside the number of
708 accessions and mosaic patterns identified for each cultivar group. It also includes reference accession
709 names with their identifiers and the total number of samples assigned to each cultivar group.

710

711

24

| Cultivars Group | Genomic composition | Nb of accessions | Nb of mosaics | Reference accessions | Nb of samples |
|---|---|---|---|---|---|
| **Sucrier (Pisang Mas)** | AA | 8 | 1 | ITC0653 Pisang Mas | 8 |
| **Pisang Jari Buaya** | AA | 8 | 1 | ITC0312 Pisang Jari Buaya | 8 |
| **M'chare (Mlali)** | AA | 3 | 1 | ITC1223 Mchare | 3 |
| **Pisang Lilin** | AA | 2 | 1 | ITC0395 Lidi | 2 |
| **Cavendish** | AAA | 19 | 1 | ITC1471 Zanzebar | 19 |
| **Gros Michel** | AAA | 3 | 1 | ITC0724 Cocos | 3 |
| **Red** | AAA | 2 | 1 | ITC1833 Shwe Ni | 2 |
| **Mutika/Lujugira** | *AAA* | *24* | *4* | ITC1630 Enjagata | *19* |
| | | | | ITC0082 Intokatoke | *3* |
| | | | | ITC1770 Siira | *1* |
| | | | | ITC0084 Mbwazirume | *1* |
| **Ilalyi** | AAA | 6 | 1 | ITC1451 Kitarasa | 6 |
| **Ambon** | AAA | 1 | 1 | DYN122 Hom Thong Mokh | 1 |
| **Orotava** | AAA | 1 | 1 | DYN121 Hom Sakhon Nakhon | 1 |
| **Rio** | AAA | 1 | 1 | ITC0277 Leite | 1 |
| **Ibota** | AAA | 2 | 1 | ITC0662 Khai Thong Ruang | 2 |
| **Kunnan** | *AB* | *5* | *2* | ITC1034 Kunnan | *3* |
| | | | | ITC1752 Poovilla Chundan | *2* |
| **Ney Poovan** | *AB* | *3* | *2* | ITC0245 Safet Velchi | *2* |
| | | | | ITC1751 Adukka Kunnan | *1* |
| **Plantain** | *AAB* | *58* | *2* | ITC0033 Bungaoisan | *7* |
| | | | | ITC0007 Asamiensa | *42* |
| | | | | not assigned† | *9* |
| **Maia Maoli/Popoulu** | *AAB* | *17* | *4* | ITC0733 Ihi U Maohi | *3* |
| | | | | ITC1135 Popoulou (CMR) | *12* |
| | | | | COOK009 Torotea | *1* |
| | | | | WNB043 Lavugi | *1* |
| **Iholena** | AAB | 6 | 1 | ITC0825 Uzakan | 6 |
| **Laknau** | AAB | 4 | 1 | ITC0332 Laknao | 4 |
| **Pome (Prata)** | *AAB* | *10* | *3* | ITC0649 Foconah | *6* |
| | | | | ITC1723 Ladies Finger | *3* |
| | | | | ITC0582 Lady Finger | 1 |
| **Mysore** | AAB | 5 | 1 | ITC1613 Karpura Chakkrakeli | 5 |
| **Silk** | *AAB* | *15* | *2* | ITC0348 Silk | *11* |

|  |  |  |  | ITC0737 Kingala n°1 | *4* |
|---|---|---|---|---|---|
| **Pisang Raja** | AAB | 2 | 1 | ITC0587 Pisang Raja | 2 |
| **Pisang Awak** | *ABB / ABBB* | *16* | *5* | ITC0659 Namwa Khom | *7* |
|  |  |  |  | ITC1719 Chinia | *5* |
|  |  |  |  | Ramu Yawa | *1* |
|  |  |  |  | ITC0213 Pisang Awak | *1* |
|  |  |  |  | ITC0334 Nzizi | *2* |
| **Bluggoe‡** | ABB | 9 | 1° | ITC0643 Cachaco | 9 |
| **Monthan‡** | ABB | 7 | 1° | ITC1483 Monthan | 7 |
| **Ney Mannan§** | ABB | 7 | 1² | ITC0361 Blue Java | 7 |
| **Peyan§** | ABB | 1 | 1² | ITC0123 Peyan | 1 |
| **Kluai Tiparod** | ABB | 2 | 1 | ITC0652 Kluai Tiparot | 2 |
| **Pelipita** | ABB | 3 | 1 | ITC0472 Pelipita | 3 |
| **Kalapua** | *ABB* | *5* | *2* | ITC2017 Kalapua | *3* |
|  |  |  |  | Dwarf Kalapua | *2* |

712     †SNP density not high enough for precise determination. ‡Identical pattern. §Identical pattern

713

714     **Table 2.** Genepool contributions to various banana cultivar groups and mosaic patterns, detailing the

715     number of centromeres contributed by each genetic source based on the reference chromosome structure.

716     The contributors are denoted as follows: Ab for *banksii*, Az for *zebrina*, Am for *malaccensis*, As for

717     *burmannica/siamea*, Ah for. *halabanensis*, S for *schizocarpa*, U for unknown, and B for *balbisiana*.

| Cultivars Group / pattern | Genome group | Ab | Az | Am | As | Ah | S | U | B | Total | Genepool contribution |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bata Bata Cluster** | AA | 13 | 4 | 4 | 0 | 1 | 0 | 0 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| **Mchare** | AA | 6 | 8 | 7 | 0 | 1 | 0 | 0 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| **Pisang Jari Buaya** | AA | 7 | 4 | 0 | 0 | 11 | 0 | 0 | 0 | 22 | Ab-Az-Am-Ah-S |
| **Pisang Lilin** | AA | 2 | 1 | 17 | 0 | 0 | 0 | 2 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| **Sucrier** | AA | 8 | 4 | 3 | 0 | 1 | 1 | 5 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| **Te'engi Cluster** | AA | 16 | 1 | 3 | 0 | 0 | 2 | 0 | 0 | 22 | Ab-Az-Am-S |
| **Ambon** | AAA | 9 | 7 | 10 | 0 | 0 | 0 | 7 | 0 | 33 | Ab-Az-Am-S-U |
| **Cavendish** | AAA | 7 | 9 | 8 | 0 | 2 | 0 | 7 | 0 | 33 | Ab-Az-Am-Ah-S-U |
| **Gros Michel** | AAA | 9 | 9 | 9 | 0 | 1 | 1 | 4 | 0 | 33 | Ab-Az-Am-Ah-S-U |
| **Ibota** | AAA | 4 | 2 | 22 | 0 | 0 | 2 | 3 | 0 | 33 | Ab-Az-Am-Ah-S-U |
| **Ilalyi** | AAA | 14 | 10 | 5 | 0 | 1 | 2 | 1 | 0 | 33 | Ab-Az-Am-Ah-S-U |
| **Mutika/Lujugira 1/2** | AAA | 12 | 18 | 0 | 0 | 1 | 2 | 0 | 0 | 33 | Ab-Az-Ah-S-U |
| **Kikundi cluster** | AAA | 8 | 11 | 12 | 0 | 0 | 0 | 2 | 0 | 33 | Ab-Az-Am-S-U |

| Name | Group | | | | | | | | | Total | Code |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Arawa | AAA | 11 | 15 | 6 | 0 | 0 | 1 | 0 | 0 | 33 | Ab-Az-Am-S-U |
| Orotava | AAA | 10 | 10 | 9 | 0 | 0 | 1 | 3 | 0 | 33 | Ab-Az-Am-S-U |
| Red | AAA | 5 | 8 | 7 | 0 | 0 | 2 | 11 | 0 | 33 | Ab-Az-Am-S-U |
| Rio | AAA | 7 | 14 | 6 | 0 | 1 | 1 | 4 | 0 | 33 | Ab-Az-Am-Ah-S-U |
| Iholena | AAB | 19 | 3 | 0 | 0 | 0 | 1 | 0 | 10 | 33 | Ab-Az-Am-S-B |
| Rukumamb Tambey Cluster | AAB | 17 | 2 | 1 | 0 | 0 | 2 | 0 | 11 | 33 | Ab-Az-Am-S-B |
| Arawa Cluster | AAB | 15 | 6 | 1 | 0 | 0 | 1 | 0 | 10 | 33 | Ab-Az-Am-S-B |
| Laknau | AAB | 18 | 2 | 0 | 0 | 0 | 2 | 0 | 11 | 33 | Ab-Az-Am-S-B |
| MMP-1/2 | AAB | 18 | 1 | 0 | 0 | 0 | 4 | 0 | 10 | 33 | Ab-Az-Am-S-B |
| MMP-3/4 | AAB | 18 | 1 | 0 | 0 | 0 | 4 | 0 | 10 | 33 | Ab-Az-S-B |
| Wan Gevi Cluster | AAB | 19 | 1 | 0 | 0 | 0 | 3 | 0 | 10 | 33 | Ab-Az-S-B |
| Buka Kiakiau | AAB | 18 | 1 | 0 | 0 | 0 | 4 | 0 | 10 | 33 | Ab-Az-S-B |
| Ruango Block CLuster | AAB | 16 | 2 | 4 | 0 | 0 | 1 | 0 | 10 | 33 | Ab-Az-Am-S-U-B |
| Mysore | AAB | 5 | 10 | 6 | 0 | 0 | 0 | 1 | 11 | 33 | Ab-Az-Am-S-U-B |
| Pisang Raja | AAB | 6 | 9 | 1 | 0 | 0 | 0 | 6 | 11 | 33 | Ab-Az-Am-S-U-B |
| Plantain-1/2/3 | AAB | 19 | 0 | 1 | 0 | 0 | 1 | 0 | 12 | 33 | Ab-Az-Am-S-B |
| Kupulik Cluster | AAB | 19 | 2 | 0 | 0 | 0 | 1 | 0 | 11 | 33 | Ab-Az-S-B |
| Mnalouki | AAB | 13 | 2 | 5 | 0 | 1 | 1 | 0 | 11 | 33 | Ab-Az-Am-Ah-S-U-B |
| Pome-1/2/3 | AAB | 6 | 8 | 7 | 0 | 1 | 0 | 0 | 11 | 33 | Ab-Az-Am-Ah-S-U-B |
| Silk-1 | AAB | 7 | 2 | 13 | 0 | 0 | 0 | 0 | 11 | 33 | Ab-Az-Am-S-B |
| Silk-2 | AAB | 5 | 2 | 16 | 0 | 0 | 0 | 0 | 10 | 33 | Ab-Az-Am-S-B |
| Kunnan-1 | AB | 2 | 1 | 8 | 0 | 0 | 0 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Kunnan-2 | AB | 4 | 1 | 6 | 0 | 0 | 0 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Ney Poovan-1 | AB | 3 | 1 | 7 | 0 | 0 | 0 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Ney Poovan-2 | AB | 3 | 0 | 7 | 0 | 0 | 1 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Bluggoe/Monthan | ABB | 8 | 0 | 0 | 0 | 0 | 2 | 0 | 23 | 33 | Ab-Az-S-B |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kalapua-1/2 | ABB | 10 | 1 | 0 | 0 | 0 | 0 | 0 | 22 | 33 | Ab-Az-S-B |
| Kluai Tiparod | ABB | 3 | 0 | 0 | 2 | 0 | 0 | 0 | 28 | 33 | Ab-As-S-B |
| Ney Mannan/Peyan | ABB | 8 | 0 | 0 | 0 | 0 | 2 | 0 | 23 | 33 | Ab-Az-S-B |
| Pelipita | ABB | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 33 | Ab-Az-Am-S-B |
| Pisang-Awak-1/2 | ABB | 2 | 1 | 7 | 0 | 0 | 1 | 0 | 22 | 33 | Ab-Az-Am-S-B |
| Pisang Awak-4x-3 | ABBB | 2 | 1 | 6 | 0 | 0 | 1 | 0 | 34 | 44 | Ab-Az-Am-S-B |
| Pisang-Awak-4x-1/2 | ABBB | 2 | 1 | 7 | 0 | 0 | 1 | 0 | 33 | 44 | Ab-Az-Am-S-B |
| La | AAB | 3 | 13 | 0 | 0 | 0 | 0 | 5 | 11 | 32 | Ab-Az-Am-Ah-S-U-B |
| Auko | ABB | 9 | 0 | 0 | 0 | 0 | 2 | 0 | 22 | 33 | Ab-S-B |
| Choi Mit | ABB | 9 | 0 | 0 | 0 | 0 | 2 | 0 | 22 | 33 | Ab-Az-S-B |
| Choi Xi Mon | ABX | 9 | 1 | 0 | 0 | 0 | 1 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Ya Ta Na Thin Kha | ABB | 3 | 0 | 0 | 8 | 0 | 0 | 0 | 22 | 33 | Ab-Az-As-S-B |
| Khai Na On | AA | 3 | 4 | 8 | 0 | 1 | 1 | 5 | 0 | 22 | Ab-Az-As-Ah-S-U |
| Manang | AA | 8 | 2 | 3 | 4 | 1 | 1 | 3 | 0 | 22 | Ab-Az-Am-As-Ah-S-U |
| Matti | AA | 4 | 0 | 7 | 11 | 0 | 0 | 0 | 0 | 22 | Ab-Az-Am-As-S |
| Pisang Madu | AA | 5 | 0 | 0 | 0 | 11 | 0 | 6 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| Pisang Pipit | AA | 5 | 7 | 1 | 0 | 1 | 0 | 8 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| Talasea Cluster | AA | 18 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 22 | Ab-Az-S |
| ToiToi | AAS | 14 | 3 | 0 | 0 | 1 | 15 | 0 | 0 | 33 | Ab-Az-Ah-S-U |
| Pisang Slendang | AAB | 10 | 2 | 1 | 0 | 1 | 3 | 4 | 12 | 33 | Ab-Az-Am-Ah-S-U-B |
| Kalmagol | AABT | 4 | 2 | 16 | 0 | 0 | 1 | 0 | 10 | 33 | Ab-Az-Am-S-B |
| Bengani | ABBT | 10 | 1 | 0 | 0 | 0 | 0 | 0 | 22 | 33 | Ab-Az-Am-S-B |
| Rekua | ABBT | 10 | 1 | 0 | 0 | 0 | 0 | 0 | 22 | 33 | Ab-Az-Am-S-B |
| Buka Cluster | ABBT | 2 | 1 | 7 | 0 | 0 | 1 | 0 | 22 | 33 | Ab-Az-Am-S-B |
| Pisang Buntal | AA | 8 | 2 | 5 | 0 | 1 | 1 | 5 | 0 | 22 | Ab-Az-Am-Ah-S-U |
| Muku Bugis | AB | 7 | 3 | 0 | 0 | 0 | 1 | 0 | 11 | 22 | Ab-Az-Am-S-B |
| Mu'u Seribu | AB | 10 | 1 | 0 | 0 | 0 | 0 | 0 | 11 | 22 | Ab-Az-Am-S-U-B |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Waga** | AAA | 18 | 3 | 0 | 0 | 0 | 12 | 0 | 0 | 33 | Ab-Az-S-U |
| **Pisang Nangka** | AAB | 11 | 4 | 10 | 0 | 0 | 2 | 2 | 4 | 33 | Ab-Az-Am-Ah-S-U-B |
| **Bagatow** | AAB | 18 | 2 | 0 | 0 | 0 | 1 | 0 | 12 | 33 | Ab-Az-S-B |
| **Muracho** | AAB | 10 | 8 | 1 | 0 | 0 | 0 | 3 | 11 | 33 | Ab-Az-Am-S-U-B |
| **Titikaveka Red** | AAB | 4 | 9 | 9 | 0 | 1 | 0 | 0 | 10 | 33 | Ab-Az-Am-AH-S-U-B |
| **Pata Tonga** | ABB | 9 | 1 | 0 | 0 | 0 | 0 | 1 | 22 | 33 | Ab-Az-S-U-B |

718

## *Supplementary material*

719

720 **Dataset S1**. Catalogue of cultivated bananas

721 **Dataset S2.** Additional mosaics for individual accessions

722 **Table S1.** List of plant materials used with consensus passport data from MGIS, morphological and
723 genomic characterization.

724 **Fig. S1.** Examples of genomic events linked to clonal diversification inferred from chromosome painting
725 raw outputs of VCFHunter. Colours correspond to *M. acuminata* ssp. *banksii* (green), ssp. *zebrina* (red)
726 and *M. schizocarpa* (light blue). Plots of diagnostic SNPs on chromosome 10 of the Mutika/Lujugira
727 group show a. a regular and most common profile, b. a switch in allelic ratio between zebrina and banksii
728 in the interstitial region of second arm, c. a deletion of a fragment of the banksii haplotype in the
729 interstitial region of second arm, and d. a switch in allelic ration between zebrina and banksii on the first
730 telomere.

731 **Fig. S2.** Tentative genomic determination keys for accessions used in the molecular catalogue. In white
732 rectangles are indicated Ancestral contributor and its ratio at the centromeric region (e.g. 7:3B for 3
733 balbisiana centromeric regions on chromosome 7), which are used to discriminate between cultivar
734 groups and individual patterns (green rectangles). The diagram must be read left to right, top to bottom, as
735 numbered.

736

## *References*

**Adheka JG, Dhed'a DB, Sivirihauma C, *et al.* 2013**. Plantain Collection and Morphological Characterization in Democratic Republic of Congo: Past and Present Activities and Prospects. *Banana Systems in the Humid Highlands of Sub-Saharan Africa*: 1.

**Agarwal M, Shrivastava N, Padh H**. **2008**. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**: 617–631.

**Ahmed D, Comte A, Curk F, *et al.* 2019**. Genotyping by sequencing can reveal the complex mosaic genomes in gene pools resulting from reticulate evolution: a case study in diploid and polyploid citrus. *Annals of Botany* **123**: 1231–1251.

**Arnaud E, Horry J-P**. **1997**. *Musalogue: a catalogue of Musa germplasm-Papua New Guinea collecting missions 1988-1989*. Rome, Italy: IPGRI.

**Baurens F-C, Martin G, Hervouet C, *et al.* 2019**. Recombination and Large Structural Variations Shape Interspecific Edible Bananas Genomes. *Molecular Biology and Evolution* **36**: 97–111.

**Belser C, Baurens F-C, Noel B, *et al.* 2021**. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Communications Biology* **4**: 1–12.

**Breton C, Cenci A, Sardos J, *et al.* 2022**. A Protocol for Detection of Large Chromosome Variations in Banana Using Next Generation Sequencing In: Jankowicz-Cieslak J, Ingelbrecht IL, eds. *Efficient Screening Techniques to Identify Mutants with TR4 Resistance in Banana: Protocols*. Berlin, Heidelberg: Springer, 129–148.

**Byabachwezi MSR, Mkulila S, Karamura DA**. **2005**. *Tanzania Musa expedition 2005: Morogoro-Mbeva region, December 2005*. INIBAP.

**Carreel F, Fauré S, De León DG, *et al.* 1994**. Evaluation de la diversité génétique chez les bananiers diploïdes (Musa sp). *Genetics Selection Evolution* **26**: 125s–136s.

**Carreel F, de Leon DG, Lagoda P, *et al.* 2002**. Ascertaining maternal and paternal lineage within Musa by chloroplast and mitochondrial DNA RFLP analyses. *Genome* **45**: 679–692.

**Cenci A, Martin G, Breton C, *et al.* 2023**. Visualizing and Inferring Chromosome Segregation in the Pedigree of an Improved Banana Cultivar (Gold Finger) with Genome Ancestry Mosaic Painting. *Horticulturae* **9**: 1330.

**Cenci A, Sardos J, Hueber Y, *et al.* 2021**. Unravelling the complex story of intergenomic recombination in ABB allotriploid bananas. *Annals of Botany* **127**: 7–20.

**Chase R, Sardos J, Ruas M, *et al.* 2014**. The field verification activity: a cooperative approach to the management of the global Musa in vitro collection at the International Transit Centre In: *XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): IX 1114*.61–66.

771    **Christelová P, Langhe ED, Hřibová E,** *et al.* **2017**. Molecular and cytological characterization of the
772    global Musa germplasm collection provides insights into the treasure of banana diversity. *Biodiversity*
773    *and Conservation* **26**: 801–824.

774    **Daniells J, Jenny C, Karamura D, Tomekpe K, Plantain IN for the I of B and**. **2001**. Musalogue (a
775    catalogue of Musa germplasm): Diversity in the genus Musa.

776    **Davey JW, Davey JL, Blaxter ML, Blaxter MW**. **2010**. RADSeq: next-generation population genetics.
777    *Briefings in functional genomics* **9**: 416–423.

778    **De Langhe E, Hřibová E, Carpentier S, Doležel J, Swennen R**. **2010**. Did backcrossing contribute to
779    the origin of hybrid edible bananas? *Annals of Botany* **106**: 849–857.

780    **De Langhe E, Karamura D, Mbwana A**. **2001**. Tanzania Musa Expedition 2001. *Rome: INIBAP/IPGRI,*
781    *Future Harvest*.

782    **D'Hont A, Denoeud F, Aury J-M,** *et al.* **2012**. The banana (Musa acuminata) genome and the evolution
783    of monocotyledonous plants. *Nature* **488**: 213.

784    **Droc G, Martin G, Guignon V,** *et al.* **2022**. The banana genome hub: a community database for
785    genomics in the Musaceae. *Horticulture Research* **9**: uhac221.

786    **Elshire RJ, Glaubitz JC, Sun Q,** *et al.* **2011**. A Robust, Simple Genotyping-by-Sequencing (GBS)
787    Approach for High Diversity Species. *PLoS ONE* **6**: e19379.

788    **Gabur I, Chawla HS, Snowdon RJ, Parkin IAP**. **2019**. Connecting genome structural variation with
789    complex traits in crop plants. *Theoretical and Applied Genetics* **132**: 733–750.

790    **Häkkinen M**. **2013**. Reappraisal of sectional taxonomy in Musa (Musaceae). *Taxon* **62**: 809–813.

791    **Herten K, Hestand MS, Vermeesch JR, Van Houdt JKJ**. **2015**. GBSX: a toolkit for experimental
792    design and demultiplexing genotyping by sequencing experiments. *BMC bioinformatics* **16**: 73.

793    **Higgins J, Osorio-Guarín JA, Olave-Achury C,** *et al.* **2023**. Characterizing subgenome recombination
794    and chromosomal imbalances in banana varietal lineages. *Annals of Botany*: mcad192.

795    **Hippolyte I, Jenny C, Gardes L,** *et al.* **2012**. Foundation characteristics of edible Musa triploids
796    revealed from allelic distribution of SSR markers. *Annals of Botany*.

797    **Hou B-H, Tsai Y-H, Chiang M-H,** *et al.* **2022**. Cultivar-specific markers, mutations, and chimerisim of
798    Cavendish banana somaclonal variants resistant to Fusarium oxysporum f. sp. cubense tropical race 4.
799    *BMC genomics* **23**: 470.

800    **IPGRI**. **1996**. *Descriptors for banana (Musa spp.)*. International Plant Genetic Resources Institute.

801    **Irish B, Rios C, Roux N, Sardos J, Goenaga R**. **2016**. *Characterization of the Musa spp. Taxonomic*
802    *Reference Collection at the USDA-ARS Tropical Agriculture Research Station*. USDA-ARS.

803    **Jarret RobertL, Gawel N, Whittemore A, Sharrock S**. **1992**. RFLP-based phylogeny of Musa species
804    in Papua New Guinea. *Theoretical and Applied Genetics* **84–84**.

805    **Kagy V, Wong M, Vandenbroucke H,** *et al.* **2016**. Traditional Banana Diversity in Oceania: An
806    Endangered Heritage. *PloS One* **11**: e0151208.

807    **Karamura E, Frison E, Karamura D, Sharrock S**. **1998**. Banana production systems in eastern and
808    southern Africa. *Bananas and food security. INIBAP, Montpellier*: 401–412.

809    **Karamura D, Karamura E, Tushemereirwe W, Rubaihayo PR, Markham R**. **2010**. Somatic
810    mutations and their implications to the conservation strategies of the east african highland bananas (musa
811    spp.). *Acta Horticulturae*: 615–622.

812    **Karamura EB, Tinzaara W, Kikulwe E, Ochola D, Ocimati W, Karamura D**. **2016**. Introduced
813    banana hybrids in Africa: seed systems, farmers' experiences and consumers' perspectives. *Acta
814    Horticulturae*: 239–244.

815    **Kilian A, Wenzl P, Huttner E,** *et al.* **2012**. Diversity arrays technology: a generic genome profiling
816    technology on open platforms. *Methods in Molecular Biology (Clifton, N.J.)* **888**: 67–89.

817    **Kitavi M, Cashell R, Ferguson M,** *et al.* **2020**. Heritable epigenetic diversity for conservation and
818    utilization of epigenetic germplasm resources of clonal East African Highland banana (EAHB)
819    accessions. *Theoretical and Applied Genetics* **133**: 2605–2625.

820    **Kitavi M, Downing T, Lorenzen J,** *et al.* **2016**. The triploid East African Highland Banana (EAHB)
821    genepool is genetically uniform arising from a single ancestral clone that underwent population expansion
822    by vegetative propagation. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*
823    **129**: 547–561.

824    **Langhe ED, Perrier X, Donohue M, Denham TP**. **2015**. The Original Banana Split: Multi-disciplinary
825    implications of the generation of African and Pacific Plantains in Island Southeast Asia. *Ethnobotany
826    Research and Applications* **14**: 299–312.

827    **Li H, Durbin R**. **2010**. Fast and accurate long-read alignment with Burrows-Wheeler transform.
828    *Bioinformatics (Oxford, England)* **26**: 589–595.

829    **Madalla N**. **2021**. Farmers traits preferences for improved banana cultivars in Tanzania and Uganda.
830    *Introductory paper at the Faculty of Landscape Architecture, Horticulture and Crop Production Science*.

831    **Martin M**. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads.
832    *EMBnet.journal* **17**: 10.

833    **Martin G, Baurens F-C, Labadie K,** *et al.* **2023**. Shared pedigree relationships and transmission of
834    unreduced gametes in cultivated banana. *Annals of Botany* **131**: 1149–1161.

835    **Martin G, Cardi C, Sarah G,** *et al.* **2020**. Genome ancestry mosaics reveal multiple and cryptic
836    contributors to cultivated banana. *The Plant Journal* **102**: 1008–1025.

837 **Martin G, Cottin A, Baurens F-C, *et al.* 2023**. Interspecific introgression patterns reveal the origins of
838 worldwide cultivated bananas in New Guinea. *The Plant Journal* **113**: 802–818.

839 **McCouch S, Navabi ZK, Abberton M, *et al.* 2020**. Mobilizing Crop Biodiversity. *Molecular Plant* **13**:
840 1341–1344.

841 **McKey D, Elias M, Pujol B, Duputié A**. **2010**. The evolutionary ecology of clonally propagated
842 domesticated plants. *New Phytologist* **186**: 318–332.

843 **Noyer JL, Causse S, Tomekpe K, Bouet A, Baurens FC**. **2005**. A new image of plantain diversity
844 assessed by SSR, AFLP and MSAP markers. *Genetica* **124**: 61–69.

845 **Perrier X, Bakry F, Carreel F, *et al.* 2009**. Combining biological approaches to shed light on the
846 evolution of edible bananas. *Ethnobotany Research & Applications* **7**: 199–216.

847 **Perrier X, Jenny C, Bakry F, *et al.* 2019**. East African diploid and triploid bananas: a genetic complex
848 transported from South-East Asia. *Annals of botany* **123**: 19–36.

849 **Perrier X, Langhe ED, Donohue M, *et al.* 2011**. Multidisciplinary perspectives on banana (Musa spp.)
850 domestication. *Proceedings of the National Academy of Sciences* **108**: 11311–11318.

851 **Powell W, Morgante M, Andre C, *et al.* 1996**. The comparison of RFLP, RAPD, AFLP and SSR
852 (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**: 225–238.

853 **Raboin L-M, Carreel F, Noyer J-L, *et al.* 2005**. Diploid Ancestors of Triploid Export Banana Cultivars:
854 Molecular Identification of 2n Restitution Gamete Donors and n Gamete Donors. *Molecular Breeding* **16**:
855 333–341.

856 **Risterucci AM, Grivet L, N'Goran J a. K, Pieretti I, Flament MH, Lanaud C**. **2000**. A high-density
857 linkage map of Theobroma cacao L. *Theoretical and Applied Genetics* **101**: 948–955.

858 **Risterucci A-M, Hippolyte I, Perrier X, *et al.* 2009**. Development and assessment of Diversity Arrays
859 Technology for high-throughput DNA analyses in Musa. *TAG. Theoretical and Applied Genetics.*
860 *Theoretische Und Angewandte Genetik* **119**: 1093–1103.

861 **Rouard M, Sardos J, Sempéré G, *et al.* 2022**. A digital catalog of high-density markers for banana
862 germplasm collections. *PLANTS, PEOPLE, PLANET* **4**: 61–67.

863 **Ruas M, Guignon V, Sempere G, *et al.* 2017**. MGIS: managing banana (Musa spp.) genetic resources
864 information and high-throughput genotyping data. *Database: The Journal of Biological Databases and*
865 *Curation* **2017**: bax046.

866 **Sachter-Smith GL, Paofa J, Sardos J**. **2021**. Bananas of West New Britain - Papua New Guinea.

867 **Sachter-Smith GL, Sardos J**. **2021a**. Bananas of Cook Islands: A catalog of banana diversity seen on
868 the islands of Rarotonga and Aitutaki.

869  **Sachter-Smith GL, Sardos J**. **2021b**. Bananas of Samoa: A catalogue of banana diversity seen on the
870  island of Upolu.

871  **Santos JD, Chebotarov D, McNally KL,** *et al.* **2019**. Fine Scale Genomic Signals of Admixture and
872  Alien Introgression among Asian Rice Landraces. *Genome Biology and Evolution* **11**: 1358–1373.

873  **Sardos J, Breton C, Perrier X,** *et al.* **2022**. Hybridization, missing wild ancestors and the domestication
874  of cultivated diploid bananas. *Frontiers in Plant Science* **13**: 969220.

875  **Sardos J, Paofa J, Janssens S, Sachter-Smith G, Rauka G, Roux N**. **2017**. *Banana collecting mission*
876  *in the Autonomous Region of Bougainville (AROB), Papua New Guinea.*

877  **Sardos J, Perrier X, Doležel J,** *et al.* **2016**. DArT whole genome profiling provides insights on the
878  evolution and taxonomy of edible Banana (Musa spp.). *Annals of Botany*: mcw170.

879  **Sempéré G, Pétel A, Rouard M,** *et al.* **2019**. Gigwa v2—Extended and improved genotype investigator.
880  *GigaScience* **8**.

881  **Shepherd K, Ferreira F**. **1984**. The PNG Biological Foundation's Banana Collection at Laloki, Port
882  Moresby, Papua New Guinea. *Newsletter-Regional Committee for Southeast Asia (IBPGR).*

883  **Simmonds NW**. **1962**. *The evolution of the bananas*. UK: Longmans.

884  **Simmonds NW**. **1966**. *Bananas*. Longmans.

885  **Simmonds NW, Shepherd K**. **1955**. The taxonomy and origins of the cultivated bananas. *Journal of the*
886  *Linnean Society of London, Botany* **55**: 302–312.

887  **Smale M, Jamora N**. **2020**. Valuing genebanks. *Food Security* **12**: 905–918.

888  **Stein A, Coriton O, Rousseau Gueutin M,** *et al.* **2017**. Mapping of homoeologous chromosome
889  exchanges influencing quantitative trait variation in Brassica napus. *Plant Biotechnology Journal* **15**:
890  1478–1489.

891  **Stover RH, Simmonds NW**. **1987**. Bananas. *Bananas.*

892  **Summo M, Comte A, Martin G,** *et al.* **2022**. GeMo: a web-based platform for the visualization and
893  curation of genome ancestry mosaics. *Database* **2022**: baac057.

894  **Taxonomy Advisory Group (TAG)**. **2016**. Minimum List of Descriptors for Musa.

895  **Tézenas Du Montcel H, De Langhe E, Swennen R**. **1983**. Essai de classification des bananiers plantains
896  (AAB). *Fruits* **38**: 461–474.

897  **Van den houwe I, Chase R, Sardos J,** *et al.* **2020**. Safeguarding and using global banana diversity: a
898  holistic approach. *CABI Agriculture and Bioscience* **1**: 15.

899    **Vogel Ely C, Bordignon SA de L, Trevisan R, Boldrini II**. **2017**. Implications of poor taxonomy in
900    conservation. *Journal for Nature Conservation* **36**: 10–13.

901    **Wu Y, Lin F, Zhou Y,** *et al.* **2021**. Genomic mosaicism due to homoeologous exchange generates
902    extensive phenotypic diversity in nascent allopolyploids. *National Science Review* **8**: nwaa277.

903    **Yakushiji H, Kobayashi S, Goto-Yamamoto N,** *et al.* **2006**. A skin color mutation of grapevine, from
904    black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional VvmybA1
905    allele. *Bioscience, Biotechnology, and Biochemistry* **70**: 1506–1508.

906

# Red group

**Classification:**  Red (AAA)

**Biological status:**  Cultivated

**Ploidy:**  Triploid ($3x$ = 33)

**Main distribution area:**  Tropics and Subtropics

**Uses:**  Dessert

**Notes:**

- Most of the plant is red
- A green variant of Red exists, it is named Green Red
- Synonym : Figue rose

**Genomic features:**

- Two schizocarpa centromeres on chromosome 2, like Ibota and Maia Maoli Popoulou
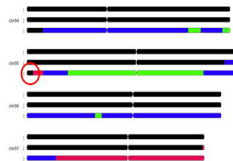- High contribution of unknown genepool

## Morphological Characterization Pictures



## Molecular Characterization



Legend:
- M. a. banksii
- M. a. malaccensis
- M. a. zebrina
- M. schizocarpa
- Unknown genepool

ITC1833 Shwe Ni

14
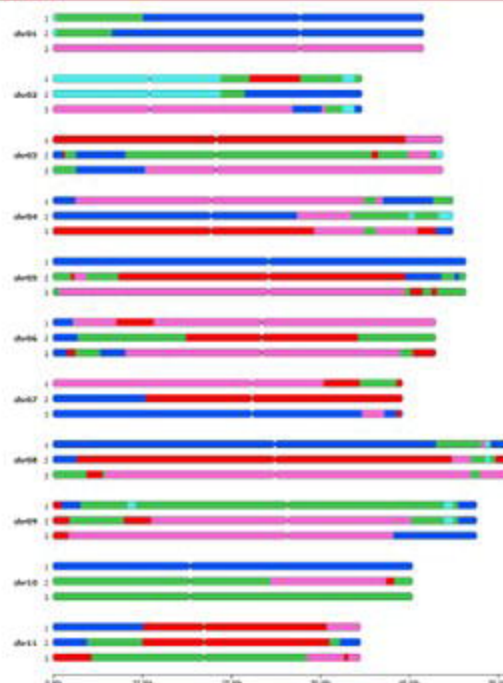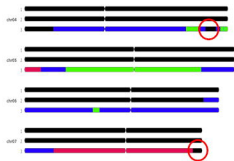
Pisang-Awak-1

Pisang-Awak-2

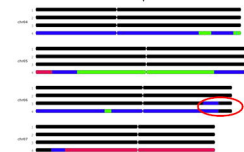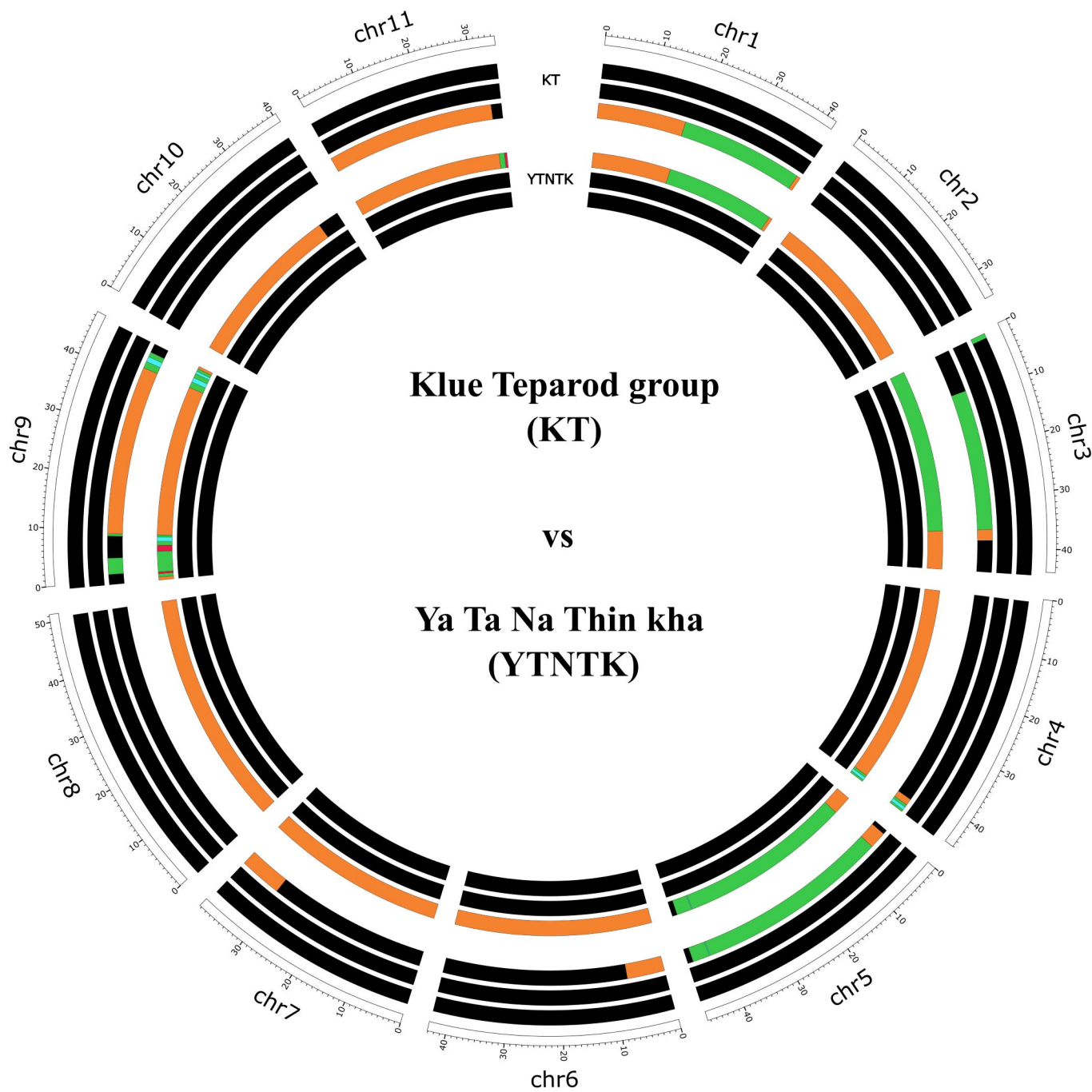Pisang-Awak-3
(not found)

3x

4x

+T

+B

+B

+B
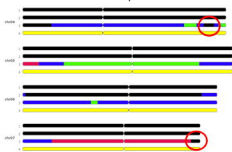
Buka
(Papua New Guinea)
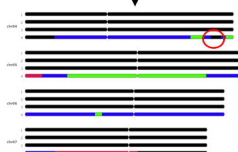
Pisang-Awak-4x-1
(Papua New Guinea)
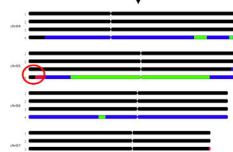
Pisang-Awak-4x-2
(Sri Lanka)

Pisang-Awak-4x-3
(Nigeria / Ivory coast)

Klue Teparod group
(KT)

vs

Ya Ta Na Thin kha
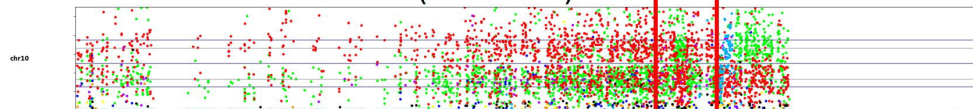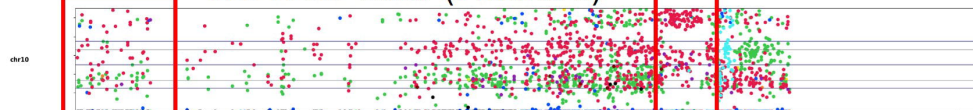(YTNTK)

B. Mutika / Lujugira (Chr. 10)

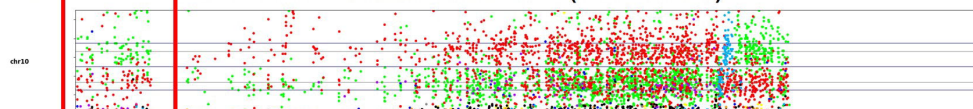a. ITC1630 'Enjagata' (Mutika 1)
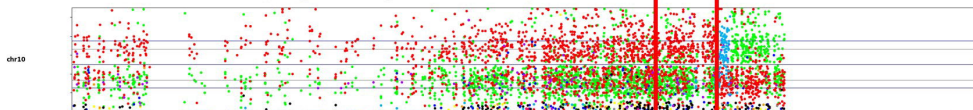
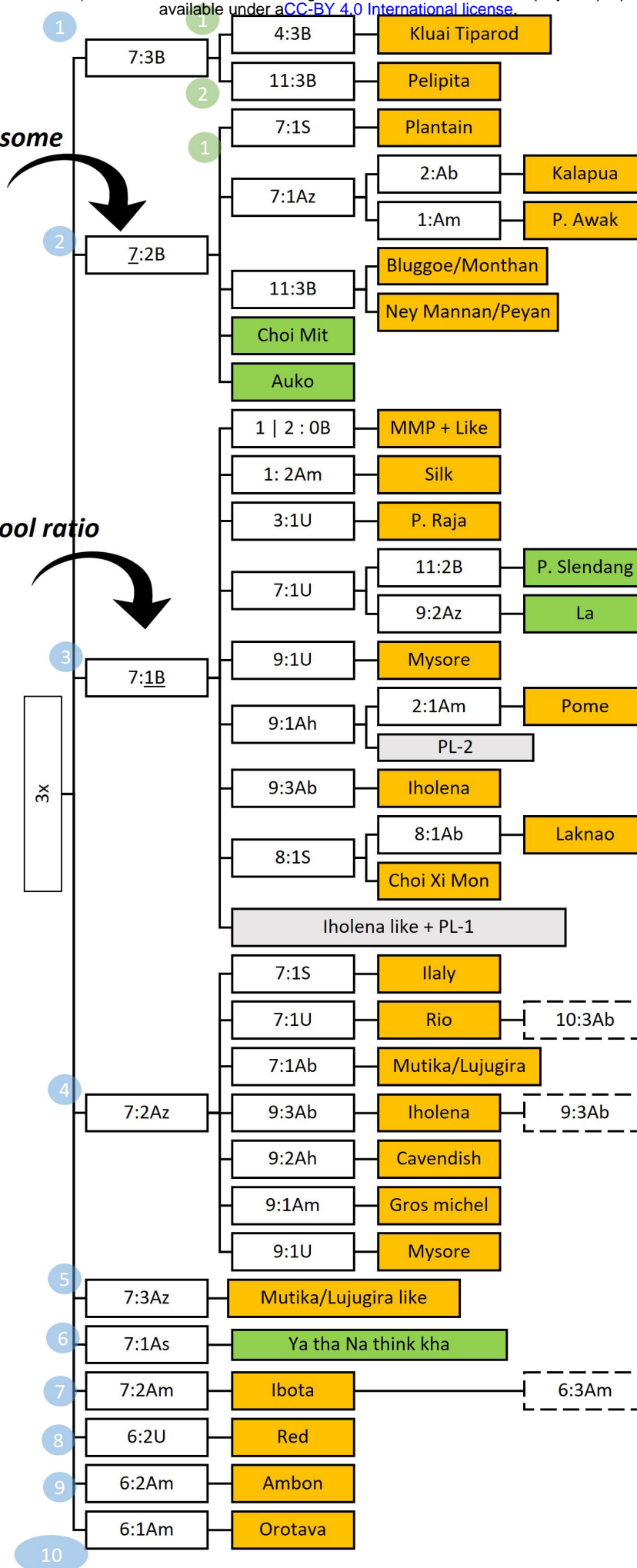b. ITC0082 'Intokatoke' (Mutika 2)

c. ITC1760 'Siira' (Mutika 3)

d. ITC0084 'Mbwazirume' (Mutika 4)

**Chromosome**

**Gene pool ratio**

- Ab *banksii*
- Az *zebrina*,
- Am *malaccensis*,
- As *burmannica/siamea*
- Ah *halabanensis*,
- S *schizocarpa*
- U unknown,
- B *balbisiana*.

① ① 4:3B — Kluai Tiparod
7:3B
② 11:3B — Pelipita
7:1S — Plantain
① 7:1Az — 2:Ab — Kalapua
1:Am — P. Awak
② 7:2B — 11:3B — Bluggoe/Monthan
Ney Mannan/Peyan
Choi Mit
Auko

1 | 2 : 0B — MMP + Like
1: 2Am — Silk
3:1U — P. Raja
7:1U — 11:2B — P. Slendang
9:2Az — La
③ 7:1B — 9:1U — Mysore
9:1Ah — 2:1Am — Pome
PL-2
9:3Ab — Iholena
8:1S — 8:1Ab — Laknao
Choi Xi Mon
Iholena like + PL-1

3x

7:1S — Ilaly
7:1U — Rio — 10:3Ab
7:1Ab — Mutika/Lujugira
④ 7:2Az — 9:3Ab — Iholena — 9:3Ab
9:2Ah — Cavendish
9:1Am — Gros michel
9:1U — Mysore
⑤ 7:3Az — Mutika/Lujugira like
⑥ 7:1As — Ya tha Na think kha
⑦ 7:2Am — Ibota — 6:3Am
⑧ 6:2U — Red
⑨ 6:2Am — Ambon
6:1Am — Orotava
⑩

**Direct identification**