



**HAL**  
open science

# Morphometric and Molecular Discrimination of the sugarcane aphid, *Melanaphis sacchari* (Zehntner, 1897) and the sorghum aphid *Melanaphis sorghi* (Theobald, 1904)

Samuel Nibouche, Laurent Costet, Raul F Medina, Jocelyn R Holt, Joëlle Sadeyen, Anne-Sophie Zoogones, Paul Brown, Roger L Blackman

## ► To cite this version:

Samuel Nibouche, Laurent Costet, Raul F Medina, Jocelyn R Holt, Joëlle Sadeyen, et al.. Morphometric and Molecular Discrimination of the sugarcane aphid, *Melanaphis sacchari* (Zehntner, 1897) and the sorghum aphid *Melanaphis sorghi* (Theobald, 1904). PLoS ONE, 2020, 16 (3), pp.e0241881. 10.1101/2020.10.23.351833 . hal-03047852

**HAL Id: hal-03047852**

**<https://hal.univ-reunion.fr/hal-03047852>**

Submitted on 9 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Morphometric and Molecular  
2 Discrimination of the sugarcane aphid,  
3 *Melanaphis sacchari*, (Zehntner, 1897)  
4 and the sorghum aphid *Melanaphis*  
5 *sorghii* (Theobald, 1904)

6 Samuel Nibouche<sup>1\*</sup>, Laurent Costet<sup>1</sup>, Raul F. Medina<sup>2</sup>, Jocelyn R. Holt<sup>2</sup>, Joëlle Sadeyen<sup>3</sup>,  
7 Anne-Sophie Zoogones<sup>1,3</sup>, Paul Brown<sup>4</sup>, Roger L. Blackman<sup>4</sup>

8

9

10 <sup>1</sup> CIRAD, UMR PVBMT, F-97410 Saint Pierre, La Réunion, France

11 <sup>2</sup> Texas A&M University, College Station, Texas, United States of America

12 <sup>3</sup> Université de La Réunion, UMR PVBMT, F-97410 Saint Pierre, La Réunion, France

13 <sup>4</sup> The Natural History Museum, London SW7 5BD, United Kingdom

14

15 \* Corresponding author

16 A-mail: [samuel.nibouche@cirad.fr](mailto:samuel.nibouche@cirad.fr)

17

18

19

## 20 Abstract

21 *Melanaphis sacchari* (Zehntner, 1897) and *Melanaphis sorghi* (Theobald, 1904) are  
22 major worldwide crop pests causing direct feeding damage on sorghum and transmitting  
23 viruses to sugarcane. It is common in the scientific literature to consider these two species as  
24 synonyms, referred to as the ‘sugarcane aphid’, although no formal study has validated this  
25 synonymy. In this study, based on the comparison of samples collected from their whole  
26 distribution area, we use both morphometric and molecular data to better characterize the  
27 discrimination between *M. sacchari* and *M. sorghi*. An unsupervised multivariate analysis of  
28 morphometric data clearly confirmed the separation of the two species. The best discriminating  
29 characters separating these species were length of the antenna processus terminalis relative  
30 to length of hind tibia, siphunculus or cauda. However, those criteria sometimes do not allow  
31 an unambiguous identification. Bayesian clustering based on microsatellite data delimited two  
32 clusters, which corresponded to the morphological species separation. The DNA sequencing  
33 of three nuclear and three mitochondrial regions revealed slight divergence between species.  
34 In particular, the COI barcode region proved to be uninformative for species separation  
35 because one haplotype is shared by both species. In contrast, one SNP located on the nuclear  
36 EF1- $\alpha$  gene was diagnostic for species separation. Based on morphological and molecular  
37 evidence, the invasive genotype damaging to sorghum in the US, Mexico and the Caribbean  
38 since 2013 is found to be *M. sorghi*.

## 39 Introduction

40 The species *Melanaphis sacchari* (Zehntner, 1897) and *Melanaphis sorghi* (Theobald,  
41 1904) were described at the turn of the 20<sup>th</sup> century by Zehntner [1] on sugarcane in Java and  
42 by Theobald [2] on sorghum in Sudan respectively. However, although these two species are  
43 commonly treated as synonyms, referred to as the ‘sugarcane aphid’, no comparative study  
44 demonstrating this synonymy has been conducted. When Remaudière & Remaudière [3]  
45 considered *M. sorghi* as a synonym of *M. sacchari* in their 1997 catalogue, following Eastop  
46 (1953) [4], they provided no reference to support this choice. Moreover, Halbert and

47 Remaudière [5] referred later to these species as the 'sorghi/sacchari group' and underlined  
48 them as '*two very variable species usually regarded as synonyms, but possibly distinct*  
49 *according to Blackman et al. (1990)*'. The communication by Blackman et al. [6] indeed  
50 supported the separation of *M. sorghi* and *M. sacchari* and provided a morphological criterion  
51 to separate the species, based on the ratio between the hind tibia length and the antennal  
52 processus terminalis length. The same separation was used in Blackman and Eastop's book  
53 in 2006 [7] and has not been challenged since.

54 Margaritopoulos et al. [8] stated that '*DNA evidence that might confirm the existence of*  
55 *two species is not yet available, but at this time it seems advisable to recognise that they are*  
56 *probably functioning as distinct taxonomic entities*'. Nibouche et al. [9] observed genetic  
57 structuring in five clonal lineages matching a geographic structure, but they could not separate  
58 the two species by using 'universal' COI barcoding.

59 Regarding host plant association, each species is observed on both sugarcane and  
60 sorghum, but *M. sorghi* is considered preferring sorghum and *M. sacchari* preferring sugarcane  
61 [6]. However, host plant association is blurred by the existence of biotypes, as shown by  
62 Nibouche et al. [10] who demonstrated the existence of a sorghum and a sugarcane biotype  
63 in Reunion populations (within the same multi locus lineage). Interestingly, *Melanaphis sorghi*  
64 is known for a long time to produce very heavy infestations on sorghum, in Africa. Early in the  
65 20<sup>th</sup> century, Vuillet & Vuillet (1914) [11] cited this aphid as responsible for famines in West  
66 Africa.

67 The objectives of this study were (1) to improve the description of the morphometric  
68 differences between *M. sorghi* and *M. sacchari*, and (2) to delimit the molecular separation of  
69 *M. sorghi* and *M. sacchari*.

## 70 Material and Methods

### 71 Material collected

72 This study is based on 199 samples, collected from 2002 to 2016 in 31 states or  
73 countries (S1 Table). No specific permissions were required when sampling aphids in the  
74 locations studied. From these 199 samples, we analyzed 2,409 apterous aphid specimens  
75 collected on cultivated sorghum (*Sorghum bicolor*) (n = 439), Johnson grass (*Sorghum*  
76 *halepense*) (n = 97), sugarcane (*Saccharum officinarum* x *S. spontaneum*) (n = 1,382),  
77 *Sorghum arundinaceum* ( $\equiv$  *Sorghum bicolor verticilliflorum  $\equiv$  *Sorghum verticilliflorum*) (n = 427),  
78 maize (*Zea mays*) (n = 2), perennial sorghum (*Sorghum* x *almum*) (n = 8) , *Sorghum* sp. (n =  
79 16), and pearl millet (*Pennisetum glaucum*) (n = 38). Most of this material was already analyzed  
80 in previous studies [9,10,12]. Using Blackman & Eastop's (2006) key, we identified these  
81 specimens as *M. sacchari* or *M. sorghi*, but the identification sometimes was ambiguous given  
82 the continuous and overlapping nature of the criteria used to separate both species. For  
83 convenience, while awaiting this taxonomy study, this material was referred to as *M. sacchari*  
84 in our previous papers.*

### 85 DNA extraction

86 DNA of individual aphids was extracted using the 'salting-out' protocol of Sunnucks and  
87 Hales [13] or using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Courtaboeuf, France). For  
88 slide-fixed specimens, a non-destructive DNA extraction was performed using the Qiagen  
89 manufacturer's protocol, but retrieving the insect body from the first elution column [12].

### 90 Microsatellites

91 Nine microsatellite (Single Sequence Repeat, SSR) markers were selected among 14  
92 previously developed markers [14]. PCR reactions were performed with labelled primers and  
93 multiplexed following previously established protocols [9]. Genotyping was carried out using

94 an ABI PRISM 3110 and alleles were identified at each locus by comparison with a size  
95 standard using Gene-Mapper version 2.5 software (Applied Biosystems). The total  
96 microsatellite genotyping dataset included 2,255 specimens: 2,175 specimens were previously  
97 analyzed in [9,10,12] and 80 additional specimens were genotyped for this study. Single  
98 combinations of alleles were characterized and arranged as distinct multilocus genotypes  
99 (MLG) and assigned to one of six multilocus lineages (MLL) [9,12]. We carried out a Bayesian  
100 clustering analysis with Structure version 2.3.4 [15], results were summarized with Structure  
101 Harvester [16], Clumpp [17] and Distruct 1.1 [18]. Parameters of the Structure analysis were:  
102 admixture, independent allele frequencies, 100,000 iterations after a 25,000 burn-in period, 10  
103 replications for each k value ranging from 1 to 8.

## 104 DNA sequencing

105 Aphids were sequenced for three mitochondrial and three nuclear DNA regions  
106 belonging to the cytochrome c oxidase subunit I (COI) [19], cytochrome c oxidase II (COII)  
107 [20], cytochrome b (CytB) [21], elongation factor-1 $\alpha$  (EF1- $\alpha$ ) [22] and the internal transcribed  
108 spacer 1 and 2 (ITS1 and ITS2) [23].

109 Three of the six sequences, COI, COII and EF1- $\alpha$ , produced informative polymorphism  
110 and were simultaneously sequenced on a large number of specimens. These sequences were  
111 concatenated and the resulting concatenated haplotypes were used to produce a minimum  
112 spanning network with PopArt [24].

## 113 Morphometry

### 114 Material

115 The morphometry dataset consisted of 89 apterous female specimens that were slide-  
116 fixed (S1 Fig) after a non-destructive DNA extraction. Among these 89 specimens, 21 were  
117 successfully genotyped with SSR and could be assigned to a multilocus lineage (MLL). The  
118 remaining 68 were not genotyped, but belonged to a sample whose multilocus lineage (MLL)

119 had been identified from other specimens. Since 186 of the 188 (98.9%) SSR genotyped  
120 samples in this study were homogeneous (i.e. contained only a single MLL), we assumed that  
121 the non-genotyped slide-fixed samples belonged to the same MLL as the other specimens in  
122 the sample that were genotyped. The distribution of the slide-fixed specimens was: 22 MLL-A,  
123 2 MLL-B, 19 MLL-C, 15 MLL-D and 31 MLL-F. Since only two specimens of MLL-B were  
124 observed, they were discarded from the statistical analysis. No MLL-E specimen was  
125 observed.

126 We also examined the Theobald type series of *Aphis sacchari* collected in Sudan in  
127 1902, which is stored in the Natural History Museum of London (NHM). In this series, only one  
128 paratype specimen (NHM-1915-81) was an apterous female and could be used for our  
129 morphometric analysis. Unfortunately, we were unable to locate the types of *Melanaphis*  
130 *sacchari*. Hollier & Hollier [25] reported that a fire destroyed the experimental station of Salatiga  
131 in Java in 1902, including Zehntner's laboratory and his collections. As Zehntner did not send  
132 his types to the Geneva museum, the type of *Melanaphis sacchari*, collected in 1897, probably  
133 was destroyed and should be considered lost.

134 We also examined some alate *M. sorghi* (n = 5) and *M. sacchari* (n = 5) specimens on  
135 slides from the MNHN (Muséum National d'Histoire Naturelle, Paris; G. Remaudière collection)  
136 and from the GBGP (Centre de Biologie et de Gestion des Populations, Montpellier; F. Leclant  
137 collection). The Theobald's type series from NHM contained one alate specimen (NHM-1915-  
138 81), but it was of insufficient quality to be included in the morphometric characterization of  
139 alatae.

## 140 Methods

141 Twenty-two characters that are used classically in aphid taxonomy [7,26] were  
142 measured on each slide-fixed specimen, using a binocular lens stereo microscope. Paired  
143 appendages (i.e. legs, antenna, siphunculi), were measured on both sides, and the mean value  
144 was used for analysis (except for the number of setae on the antenna, which was observed on  
145 one side only). Twelve ratios were computed from these characters.

146 The analysis was carried out on a subset of 11 characters (Table 1) that are known to  
147 be discriminant between species within the *Melanaphis* genus [5,7,26,27]. Because of missing  
148 data, only 50 specimens were used in the multivariate analysis. Because Theobald did not  
149 clarify his *M. sorghi* paratype specimen (NHM-1915-81), we were unable to observe *urs* and  
150 *siphBW*, causing three missing ratios in the dataset. To include the paratype in the analysis,  
151 we replaced these three missing data by zeros after the standardization step (see below).

152 **Table 1. List of the 11 morphological variables used in the discriminant analysis of**  
153 **principal components (DAPC) analysis.**

Variable name	Variable signification
<i>NsetaeCauda</i>	number of setae on the cauda
<i>pt:cauda</i>	ratio processus terminalis length / cauda length
<i>HindTibia:pt</i>	ratio hind tibia length / processus terminalis length
<i>Ant:BL</i>	ratio antenna length / body length
<i>urs:htII</i>	ratio ultimate rostral segment length / hind tarsi II length
<i>pt:VIb</i>	ratio processus terminalis length / base length of the 6th antennal segment
<i>pt:siph</i>	ratio processus terminalis length / siphunculi length
<i>cauda:urs</i>	ratio cauda length / ultimate rostral segment length
<i>siph:BL</i>	ratio siphunculi length / body length
<i>siph:siphBW</i>	ratio siphunculi length / siphunculi basal width
<i>siph:cauda</i>	ratio siphunculi length / cauda length

154

155

156 The 11 variables were standardized to obtain a mean of zero and a standard deviation  
157 of 1 prior to analysis.

158 The data were first submitted to a discriminant analysis of principal components  
159 (DAPC) with the R package ADEGENET using the *find.clusters* function [28]. This method first  
160 uses an unsupervised k-means clustering approach to determine the number of clusters  
161 without requiring any a priori clustering information. The determination of the number of



162 clusters was based on the Bayesian Information Criteria (BIC). Then in a second step, a  
163 principal component analysis (PCA) is carried out, followed by a canonical discriminant  
164 analysis (CDA) performed on the coordinates along the principal components. To verify if the  
165 clustering was congruent with the separation of *M. sorghi* and *M. sacchari*, we compared the  
166 *HindTibia:pt* ratio between the clusters delimited by ADEGENET. Indeed, *M. sorghi* specimens  
167 have a relative length of the processus terminalis shorter than *M. sacchari* specimens [6,7].  
168 According to [7,8,26], the *HindTibia:pt* ratio range for apterae is (2.0 - 3.0) for *M. sorghi* vs. (1.4  
169 - 2.2) for *M. sacchari*.

170 Secondly, the complete morphological dataset (i.e. 34 morphological traits on 88  
171 apterous specimens) was submitted to a one-way ANOVA with SAS PROC GLM [29] to detect  
172 significant differences between *M. sacchari* and *M. sorghi*. Because we carried out multiple  
173 analysis, to control the study-wise type-1 error level we used a 5% False Discovery Rate (FDR)  
174 approach [30] with SAS PROC MULTTEST [29] to detect significant differences between  
175 clusters. To carry out this analysis, slide specimens were assigned to *M. sacchari* or *M. sorghi*  
176 according to their molecular assignment using SSR and EF1- $\alpha$  data.

## 177 Results

### 178 Morphometry

179 According to the unsupervised k-means clustering, the number of inferred  
180 morphological clusters was  $k = 2$ . The membership probability of each specimen is shown by  
181 Fig 1. The *M. sorghi* paratype was assigned to the morphological cluster 1, with a 100%  
182 membership probability.

183

184 **Fig 1. DAPC analysis based on 11 morphological traits recorded on 51 slide-fixed**  
185 **specimens.** Species identification by SSR and EF1- $\alpha$ : orange = *M. sacchari*, blue = *M. sorghi*.

186

187 Both clusters differed significantly ( $P < 0.0001$ ) by their *HindTibia:pt* ratio. The blue  
188 cluster (Fig 1), which contained the *M. sorghi* paratype and exhibited the lowest *HindTibia:pt*  
189 (Fig 2 and S1 Table), was *M. sorghi*. The orange cluster (Fig 1), which exhibited the largest  
190 *HindTibia:pt* ratio (Fig 2 & S2 Table), corresponded to *M. sacchari*.

191

192 **Fig 2. Comparison between *M. sacchari* and *M. sorghi* using the three traits showing the**  
193 **highest loadings in the DAPC among 51 slide-fixed specimens.** The specimens are  
194 assigned to *M. sorghi* or *M. sacchari* according to their SSR or EF1- $\alpha$  genotype.

195

196

197 The loading plot (Fig 3) showed that the morphological characters contributing most to  
198 the DAPC were *pt:cauda*, *HindTibia:pt* and *pt:siph*. The graphical comparison of these three  
199 main contributing traits between both species is presented in Fig 2. In these plots, the  
200 assignment of the specimens to both species was made according to their SSR and EF1- $\alpha$   
201 genotype (see below). For the three traits, both species differ significantly with an uncorrected  
202 P-value  $< 0.0001$  ( $F = 100.3$ ,  $78.73$  and  $62.96$  respectively for *pt:cauda*, *HindTibia:pt* and  
203 *pt:siph*). However, some overlapping is observed for each trait.

204

205 **Fig 3. Canonical loading plot.** The horizontal line is the limit showing the variables  
206 (morphological traits) that yield a cumulated 75% contribution to the DAPC. The individual  
207 peaks show the magnitude of the influence of each variable on separation of *M. sorghi* and *M.*  
208 *sacchari*.

209

210 The comparison of all morphological traits of *M. sorghi* and *M. sacchari* apterous  
211 females is presented in S2 Table. There are significant differences between both species in  
212 16 out of 34 traits.

213 The values for alatae are given in S3 Table. Due to the small number of specimens  
214 observed, no statistical comparison was carried out and these values are only suggestive.

215 Because we only observed museum slide-fixed alate specimens, we could not genotype them.  
216 In the absence of a key for alatae, species assignment was carried out taking into account  
217 geographic origin, according to the worldwide repartition of both species (Fig 5). Specimens  
218 from West Africa (Burkina Faso, Ivory Coast and Senegal) were considered as *M. sorghi*,  
219 specimens from Reunion and Brazil were considered as *M. sacchari*. Although the number of  
220 alate specimens is insufficient to form a basis for any statistical comparison, it should be noted  
221 that, as in apterae, the *pt* of alatae assigned to *M. sacchari* tends to be longer than that of  
222 alatae assigned to *M. sorghi*, and that the ratios involving this parameter, especially *pt:siph*,  
223 may prove to be useful discriminants.

224  
225 **Fig 5. Minimum spanning network constructed using the concatenated COI, COII and**  
226 **EF1- $\alpha$  sequences.** The orange and blue boxes indicate the STRUCTURE clusters inferred  
227 from SSR data (Table 4). The eight concatenated haplotypes (cH1 to cH8) are listed in Table  
228 4. The number of hatch marks represents the number of mutations separating the  
229 concatenated haplotypes. Circle sizes are proportional to haplotype frequencies.

230

## 231 Microsatellites

232 Fifty-nine multilocus genotypes (MLG) have been identified (S6 Table). Fifty-six were  
233 published earlier [9,10,12] and three new ones (Ms25, Ms26, and Ms58) were observed during  
234 this study. The clustering with Structure, followed by the use of the Evanno et al. [31] method,  
235 leads to the conclusion that the number of inferred populations is  $k = 2$  (S2 Fig).

236 The assignment of each MLG with  $k = 2$  is presented in Fig 4. The first cluster groups  
237 MLL-A-E-F, the second cluster groups MLL-B-C-D.

238

239 **Fig 4. Assignment of each of the 59 MLG to the two clusters inferred by Structure.** MLG  
240 were defined using nine microsatellite markers.

241

242 The congruence between the DAPC morphological assignment to *M. sorghi* or *M.*  
 243 *sacchari* and the Structure Bayesian clustering based on SSR genotyping showed  
 244 unambiguously that the blue cluster is *M. sorghi* and the orange cluster is *M. sacchari* (Table  
 245 2). Among the 29 specimens belonging to the blue Structure cluster, 27 were morphologically  
 246 assigned to *M. sorghi*, one was morphologically assigned to *M. sacchari* and one was  
 247 undetermined. Among the 21 specimens belonging to the orange Structure cluster, 19 were  
 248 morphologically assigned to *M. sacchari*, one was morphologically assigned to *M. sorghi* and  
 249 one undetermined. The resulting accuracy of the congruence of Structure clustering and  
 250 morphological clustering is 95.8% (46 / 48), when excluding the two undetermined specimens.  
 251

252 **Table 2. Multi Locus Lineage (MLL) identification of the 51 slide fixed specimens**  
 253 **morphologically assigned to *M. sorghi* or *M. sacchari* by DAPC.**

Structure cluster	MLL	level of MLL identification <sup>b</sup>	morphological assignment by DAPC		
			<i>M. sacchari</i>	<i>M. sorghi</i>	undetermined <sup>a</sup>
orange	MLL-C	specimen	1		1
		sample	9	<u>1</u> <sup>c</sup>	
	MLL-D	specimen	8		
		sample	1		
blue	MLL-A	specimen			
		sample		16	
	MLL-F	specimen	<u>1</u> <sup>c</sup>	7	
		sample		4	1

254 <sup>a</sup> The specimens with a DAPC membership lower than 0.8 are considered  
 255 undetermined.

256 <sup>b</sup> MLL identification was carried out by SSR genotyping on the specimen itself or on  
 257 other specimens from the same sample.

258 <sup>c</sup> Specimens whose molecular and morphometric assignments are discordant.

259

260

## 261 DNA sequencing

262 A total of 371 aphids were sequenced for at least one of three genes, COI, COII or  
263 EF1- $\alpha$ : 340 for COI, 143 for COII, and 163 for EF1- $\alpha$ .

264 In COI (658 bp), we detected four SNPs defining four haplotypes (S4 Table). The  
265 sequences were deposited in Genbank under accession numbers KJ083108-KJ083215,  
266 KX453783-KX453784, MG838208-MG838315, MT813521-MT813656, and MT821458-  
267 MT821474.

268 In COII (763 bp), we detected four SNPs defining five haplotypes (S5 table). The  
269 sequences were deposited in Genbank under accession numbers MT847245-MT847387.

270 The EF1- $\alpha$  gene portion amplified was 1,014 bp long. The sequences were deposited  
271 in Genbank under accession numbers MT847432-MT847594. Prior to analysis, the sequences  
272 were trimmed to a 528 bp length (from position 248 bp to 775 bp), to discard low quality 5' and  
273 3' sequence ends in most specimens. In this EF1- $\alpha$  528 bp sequence portion, we detected 1  
274 unambiguous SNP and 4 ambiguous positions (heterozygous) defining seven haplotypes (S6  
275 Table). Only two haplotypes were defined when omitting the four ambiguous positions, and we  
276 considered only these two haplotypes in further analysis. Haplotype H1 was only observed in  
277 *M. sacchari* and haplotype H2 only in *M. sorghi*.

278 In CytB (745 bp), we detected one SNP defining two haplotypes. Both haplotypes were  
279 detected in both *M. sacchari* and *M. sorghi*. The sequences were deposited in Genbank under  
280 accession numbers MT847388-MT847423.

281 In the ITS1 region, we obtained 445-451 bp length sequences. These sequences  
282 included a 3-6 bp indel region and one SNP. The two haplotypes defined by the SNP were  
283 present in both *M. sacchari* and *M. sorghi* specimens. The sequences were deposited in  
284 Genbank under accession numbers MT821305-MT821342.

285 In the ITS2 region, in a 462 bp-sequence length, we detected one SNP and three  
286 indels, but the SNP was located within an indel and therefore could not be used in further

287 analysis. The sequences were deposited in Genbank under accession numbers MT821344-  
288 MT821448.

289 A total of 63 specimens were genotyped at the three genes COI, COII, EF1- $\alpha$ , and were  
290 also successfully genotyped with SSR. The relationship between MLL and haplotypes among  
291 these specimens is presented in Table 3. EF1- $\alpha$  provided a diagnostic substitution at position  
292 637 separating *M. sorghi* from *M. sacchari*. The distinctive base was T for *M. sacchari* and A  
293 for *M. sorghi*. COI provides an incomplete separation of both species: haplotypes H2 and H3  
294 are diagnostic of *M. sacchari*, H6 is diagnostic of *M. sorghi*, but haplotype H1 is present in both  
295 species.

296 **Table 3. Haplotypes defined by the concatenation of four genes and correspondence**  
297 **with the Multilocus Lineages (MLLs) defined using SSR.**

concatenated haplotype	gene / haplotype			species	MLL	Structure Cluster	n
	CO1	CO2	EF				
cH1	H1	H3	H2	<i>sorghi</i>	A	1	18
cH1	H1	H3	H2	<i>sorghi</i>	F	1	8
cH2	H1	H3	H1	<i>sacchari</i>	B	2	2
cH3	H2	H1	H1	<i>sacchari</i>	C	2	11
cH4	H2	H5	H1	<i>sacchari</i>	C	2	5
cH5	H3	H3	H1	<i>sacchari</i>	D	2	10
cH6	H1	H2	H2	<i>sorghi</i>	E	1	6
cH7	H6	H3	H2	<i>sorghi</i>	F	1	1
cH8	H1	H4	H2	<i>sorghi</i>	A	1	2

298

299

300 The number of variable sites was nine on the 1,806 bp of the three concatenated gene  
301 sequences (i.e. 0.49%), defining eight haplotypes, with a nucleotide diversity of 0.13%. The  
302 minimum spanning network among haplotypes is shown by Fig 5. *M. sorghi* specimens defined  
303 a star shaped haplogroup centered on haplotype cH1. *M. sacchari* defined a less  
304 homogeneous haplogroup, each haplotype being separated from the other by two or three  
305 substitutions.

306 The within species and between species divergences for the six sequenced gene  
 307 portions are summarized in Table 4.

308 **Table 4. Sequence divergences (pairwise uncorrected P-distances, %) between or within**  
 309 **species.**

	COI	COII	EF1 $\alpha$	CytB	ITS2	ITS1
within <i>M. sorghi</i>	0.15 (0.15 - 0.15) n = 184	0.18 (0.13-0.26) n = 70	0 - n = 88	0 - n = 11	0 - n = 67	0.16 (0 - 0.26) n = 18
within <i>M. sacchari</i>	0.30 (0.30 - 0.30) n = 136	0.18 (0.13-0.26) n = 73	0 - n = 83	0.13 (0.13 - 0.13) n = 32	0 - n = 38	0.13 (0 - 0.26) n = 20
between species	0.28 (0 - 0.45) n = 320	0.18 (0 - 0.26) n = 143	0.19 (0.19 - 0.19) n = 171	0.09 (0 - 0.13) n = 43	0 - n = 105	0.13 (0 - 0.26) n = 38

310 Values are the mean of the pairwise distances; minimum and maximum distances for these  
 311 comparisons are given in parentheses. Number of genotyped specimens = n. Species  
 312 identification was carried out by SSR genotyping on the specimens themselves or on other  
 313 specimens from the same sample.

314

315

## 316 Geographical distribution of *M. sacchari* and *M. sorghi*

317 The geographical distribution of *M. sacchari* and *M. sorghi*, is presented in Fig 6 and is  
 318 based on 2,332 genotyped specimens. Specimens belonging to MLL-A-E-F were assigned to  
 319 *M. sorghi*, while specimens belonging to MLL-B-C-D were assigned to *M. sacchari*. In the  
 320 absence of SSR genotyping, specimens bearing the EF1- $\alpha$  haplotype H1 or the COI  
 321 haplotypes H2 or H3 were assigned to *M. sacchari* while specimens bearing EF1- $\alpha$  haplotype  
 322 H2 were assigned to *M. sorghi* (Table 4). The remaining specimens (n = 24) were considered  
 323 unidentified and therefore not taken into account in Fig 6. Only one unambiguous (i.e. allowing  
 324 the species identification) data point from India was available from a public database (see Fig  
 325 6 legend). Most COI sequences stored in these public databases are either haplotypes that

326 we did not encounter in our study or are H1 haplotype, which is uninformative because it is  
327 shared by *M. sorghi* and *M. sacchari*.

328

329 **Fig 6. Molecular identification with SSRs and sequencing of COI or EF1- $\alpha$  of 2,332**  
330 **specimens:** blue = *M. sorghi*, orange = *M. sacchari*. Data from India is the EF1- $\alpha$  sequence  
331 Genbank accession KU048048.1 (exact geographical location within India not available).

332

333

334 In West and in Southern Africa, *M. sorghi* is the sole detected species. In East Africa,  
335 *M. sacchari* has been detected in Kenya and Tanzania, and *M. sorghi* in Uganda and Kenya.  
336 In Kenya, both species coexist and were collected in the same sample once. Reunion and  
337 Mauritius, in the South West Indian Ocean, are exclusively colonized by *M. sacchari*. The  
338 Neotropical zone is colonized by *M. sacchari*. The Nearctic zone is colonized by both species,  
339 as a result of the recent introduction of *M. sorghi* in the Americas. In Asia, *M. sorghi* is present  
340 in China and India, and *M. sacchari* in Cambodia. In Australia and in Hawaii, only *M. sacchari*  
341 was detected.

## 342 Discussion

343 Genetic analyses with SSRs and three gene sequences showed that two genetic  
344 clades exist, one grouping MLL-A-E-F and the other grouping MLL-C-D. The multivariate  
345 morphometric data analysis separated the specimens in two groups matching the two genetic  
346 clades. Comparison with a *M. sorghi* paratype and comparison of the *HindTibia:pt* ratio [6,7,26]  
347 confirmed that the genetic clade grouping MLL-A-F is *M. sorghi* while the clade grouping MLL-  
348 C-D is *M. sacchari*. The status of MLL-B (Australian specimens) remains to be confirmed: it is  
349 assigned to *M. sorghi* by SSRs and EF1- $\alpha$  sequence, but we did not confirm this assignment  
350 by morphometric means, due to a lack of specimens. Similarly, the lack of specimens  
351 prevented us from studying the morphology of MLL-E (all from China) and its taxonomic status



352 remains uncertain, although SSRs and EF1- $\alpha$  sequence data both suggest it belongs to *M.*  
353 *sorghii*.

354 Three morphological criteria are useful for species separation: *pt:cauda*, *pt:siph* and  
355 *HindTibia:pt*. However, as observed by Blackman and Eastop [7,26], there are no clear limits  
356 between species and values overlap largely (Fig 2). The ranges for *M. sacchari* vs. *M. sorghii*  
357 are respectively *pt:cauda* (1.72-2.35) vs. (2.19-2.8), *HindTibia:pt* (1.8-2.31) vs. (1.54-1.93) and  
358 *pt:siph* (2.71-4.28) vs. (2.26-3.19). Due to overlap, the use of these morphological criteria can  
359 lead to ambiguous results and should be applied quantitatively (i.e. at the population level)  
360 rather than qualitatively (i.e. at the individual level).

361 The molecular diagnostic methods for separation of *M. sacchari* and *M. sorghii* are  
362 summarized in Figs 7 and 8. COI can allow for the identification of *M. sacchari* through two  
363 specific SNPs in positions 263 or 294. But, because haplotype H1 (Table 3) is shared by both  
364 species, an unambiguous identification of *M. sorghii* with COI is sometimes not possible. With  
365 EF1- $\alpha$ , a specific SNP in position 637 allows the separation of both species. Genotyping with  
366 a single SSR locus also allows the separation of both species. For example, the SSR locus  
367 CIR-Ms-G01 can be used for this purpose (Fig 8). The two alleles present at this locus in *M.*  
368 *sacchari* are separated by 25 to 31 bp, whereas *M. sorghii* genotypes are mostly homozygous  
369 or exhibit two alleles separated by 4 to 8 bp only (S7 table).

370

371 **Fig 7. Molecular diagnosis for separation of *M. sacchari* and *M. sorghii* using sequencing**  
372 **of COI or EF1- $\alpha$ .**

373 **Fig 8. Molecular diagnosis for separation of *M. sacchari* and *M. sorghii* using the SSR**  
374 **locus CIR-Ms-G01.** *M. sacchari* (MLL-D) are in lanes A10 (voucher # SNIB00040\_0101) and  
375 A1 (voucher # SNIB00233\_0102). *M. sorghii* (MLL-F) are in lanes A11 (voucher #  
376 SNIB00075\_0101) and A5 (voucher # SNIB000237\_0102). The PCR was carried out  
377 according to [14]. The migration of PCR products was carried out on a Qiagen Qiaxcel  
378 electrophoresis analyzer. The image was generated by the Qiaxcel ScreenGel 1.6.0 software.

379

380

381 We observed low genetic distances between *M. sacchari* and *M. sorghi* when  
382 comparing gene sequences that are widely used in aphid taxonomy. A ‘borderline’ distance  
383 between species in the COI barcode region is sometimes used by some authors to confirm  
384 species limits [32,33]. However, it is now well recognized that there is no universal genetic  
385 distance separating aphid species and that low genetic difference between species can occur  
386 in aphids. Multiple examples of low COI, COII, CytB pairwise distance between species exist  
387 [34-36]. For example, a situation very similar to ours occurs in the genus *Megoura*, where Kim  
388 and Lee [37] observed an absence of COI, COII, CytB, ITS1, ITS2 divergence between  
389 *Megoura litoralis* Müller and *Megoura viciae* Buckton, which differ by a 0.2% P-distance on  
390 EF1- $\alpha$ .

391 According to Blackman et al. [6], *M. sorghi* is more likely observed on sorghum and *M.*  
392 *sacchari* more likely on sugarcane, although not absolutely specific to the hosts indicated by  
393 their names. In our study, the samples were distributed as follows: 14 from sugarcane vs. 94  
394 from sorghum (all *Sorghum* species together) for *M. sorghi*, and 18 samples from sorghum vs.  
395 69 from sugarcane for *M. sacchari*. Although our sampling plan was not designed to test host  
396 plant preference, the difference of preference between the two species appears obvious. The  
397 results of Boukari et al. [38] obtained in Florida confirm this preference, showing that sugarcane  
398 harbors almost only *M. sacchari* (COI haplotypes H2 and H3), which is absent from aphids on  
399 *Sorghum* spp. The recent work by Paudyal et al. [39] in the USA also supports this apparent  
400 preference. Indeed, using host transfer experiments, these authors have demonstrated that  
401 MLL-F strains collected from *Sorghum* spp. exhibited a higher fitness on sorghum than on  
402 sugarcane, and that an MLL-D strain collected from sugarcane exhibited a higher fitness on  
403 sugarcane than on sorghum. According to this host preference difference, we suggest that the  
404 common name ‘sugarcane aphid’ should be used for *M. sacchari* and ‘sorghum aphid’ for *M.*  
405 *sorghi*.

406 Our study shows that the invasive genotype responsible for outbreaks on sorghum in  
407 North and Central America and the Caribbean islands since 2013 is MLL-F [12], which, belongs

408 to *M. sorghi*, while the genotype present before 2013 (MLL-D) is *M. sacchari*. If the hypothesis  
409 of a lower fitness of *M. sacchari* on sorghum compared to *M. sorghi* is confirmed, this would  
410 explain why no damage was observed on sorghum prior to the introduction of *M. sorghi* to the  
411 Americas.

## 412 Acknowledgements

413 We are grateful to Hughes Telismart, Magali Hoarau and Antoine Franck for technical  
414 assistance. We also thank Susan E. Halbert for her helpful comments on this manuscript. We  
415 also acknowledge Armelle Coeur d'Acier (CBGP), Laurent Fauvre and Thierry Bourgoin  
416 (MNHN) for providing access to the CBGP and MNHN collections.

## 417 References

- 418 1. Zehntner L (1897) Overzicht van de Zieken van het Suikerriet op Java. Mededelingen van  
419 het Proefstation "Oost-Java" Nieuwe Serie 37: 525–575.
- 420 2. Theobald F (1904) The 'Dura' Aphis or 'Asal Fly'. Report of the Wellcome Research  
421 Laboratories at the Gordon Memorial College, Khartoum, pp. 43-45.
- 422 3. Remaudière G, Remaudière M (1997) Catalogue of the world's Aphididae. Paris, France:  
423 INRA.
- 424 4. Eastop VF (1953) Notes on East African Aphids I—Synonymy. The East African Agricultural  
425 Journal 18: 133–135.
- 426 5. Halbert SE, Remaudière G (2000) A new oriental *Melanaphis* species recently introduced  
427 in North America [Hemiptera, Aphididae]. Revue française d'entomologie 22: 109-117.
- 428 6. Blackman RL, Eastop VF, Brown PA (1990) The biology and taxonomy of the aphids  
429 transmitting barley yellow dwarf virus. In: Burnett PA, editor. World Perspectives on  
430 Barley Yellow Dwarf International Workshop. Udine (Italy): CIMMYT. pp. 197-214.
- 431 7. Blackman RL, Eastop VF (2006) Aphids on the world's herbaceous plants and shrubs.  
432 Chichester, UK: John Wiley & Sons Ltd.
- 433 8. Margaritopoulos J, Papapanagiotou A, Voudouris CC, Kati A, Blackman R (2013) Two aphid  
434 species newly introduced in Greece. Entomologia Hellenica 22: 23-28.

- 435 9. Nibouche S, Fartek B, Mississippi S, Delatte H, Reynaud B, Costet L (2014) Low genetic  
436 diversity in *Melanaphis sacchari* aphid populations at the worldwide scale. PLoS ONE  
437 9: e106067.
- 438 10. Nibouche S, Mississippi S, Fartek B, Delatte H, Reynaud B, Costet L (2015) Host Plant  
439 Specialization in the Sugarcane Aphid *Melanaphis sacchari*. PLoS ONE 10: e0143704.
- 440 11. Vuillet J, Vuillet A (1914) Les pucerons du sorgho au Soudan Français. L'agronomie  
441 coloniale: 137-143.
- 442 12. Nibouche S, Costet L, Holt JR, Jacobson A, Pekarcik A, Sadeyen J, et al. (2018) Invasion  
443 of sorghum in the Americas by a new sugarcane aphid (*Melanaphis sacchari*)  
444 superclone. PLoS ONE 13: e0196124.
- 445 13. Sunnucks P, Hales D (1996) Numerous transposed sequences of mitochondrial  
446 cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae).  
447 Molecular Biology and Evolution 13: 510-524.
- 448 14. Molecular Ecology Resources Primer Development Consortium, Andris M, Aradottir GI,  
449 Arnau G, Audzijonyte A, Bess EC, et al. (2010) Permanent Genetic Resources added  
450 to Molecular Ecology Resources Database 1 June 2010 – 31 July 2010. Molecular  
451 Ecology Resources 10: 1106-1108.
- 452 15. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using  
453 multilocus genotype data. Genetics 155: 945-959.
- 454 16. Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing  
455 STRUCTURE output and implementing the Evanno method. Conservation Genetics  
456 Resources 4: 359-361.
- 457 17. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation  
458 program for dealing with label switching and multimodality in analysis of population  
459 structure. Bioinformatics 23: 1801-1806.
- 460 18. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population  
461 structure. Molecular Ecology Notes 4: 137-138.

- 462 19. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of  
463 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.  
464 *Molecular marine biology and biotechnology* 3: 294-299.
- 465 20. Kim H, Lee S (2008) A molecular phylogeny of the tribe Aphidini (Insecta: Hemiptera:  
466 Aphididae) based on the mitochondrial tRNA/COII, 12S/16S and the nuclear EF1 $\alpha$ ;  
467 genes. *Systematic Entomology* 33: 711-721.
- 468 21. Harry M, Solignac M, Lachaise D (1998) Molecular evidence for parallel evolution of  
469 adaptive syndromes in fig-breeding *Lissocephala* (Drosophilidae). *Molecular*  
470 *Phylogenetics and Evolution* 9: 542-551.
- 471 22. von Dohlen CD, Kurosu U, Aoki S (2002) Phylogenetics and evolution of the eastern Asian–  
472 eastern North American disjunct aphid tribe, Hormaphidini (Hemiptera: Aphididae).  
473 *Molecular Phylogenetics and Evolution* 23: 257-267.
- 474 23. Ji YJ, Zhang DX, He LJ (2003) Evolutionary conservation and versatility of a new set of  
475 primers for amplifying the ribosomal internal transcribed spacer regions in insects and  
476 other invertebrates. *Molecular Ecology Notes* 3: 581-585.
- 477 24. Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network  
478 construction. *Methods in Ecology and Evolution* 6: 1110-1116.
- 479 25. Hollier J, Hollier A (2018) Leo Zehntner, Swiss pioneer of tropical applied entomology.  
480 *Antenna* 42: 56-60.
- 481 26. Blackman RL, Eastop VF (2020) Aphids on the world's plants. An online identification and  
482 information guide. Available from: <http://www.aphidsonworldsplants.info>
- 483 27. Kim H, Jang Y (2012) Taxonomic review and morphometric analysis of the genus  
484 *Melanaphis* van der Goot (Hemiptera: Aphididae) in Korea. *Animal Cells and Systems*  
485 16: 34-40.
- 486 28. Jombart T (2008) *adeigenet*: a R package for the multivariate analysis of genetic markers.  
487 *Bioinformatics* 24: 1403-1405.
- 488 29. SAS Institute (2010) SAS OnlineDoc® 9.3. Cary, NC, USA: SAS Institute Inc.

- 489 30. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and  
490 powerful approach to multiple testing. *Journal of the Royal statistical society: series B*  
491 (Methodological) 57: 289-300.
- 492 31. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals  
493 using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-  
494 2620.
- 495 32. Massimino Cocuzza GE, Cavalieri V (2014) Identification of aphids of *Aphis frangulae*-  
496 group living on Lamiaceae species through DNA barcode. *Molecular Ecology*  
497 *Resources* 14: 447-457.
- 498 33. Rakauskas R, Havelka J, Zaremba A, Bernotienė R (2014) Mitochondrial COI and  
499 morphological evidence for host specificity of the black cherry aphids *Myzus cerasi*  
500 (Fabricius, 1775) collected from different cherry tree species in Europe (Hemiptera,  
501 Aphididae). *ZooKeys* 388.
- 502 34. Lee W, Kim H, Lim J, Choi HR, Kim Y, Kim YS, et al. (2010) Barcoding aphids (Hemiptera:  
503 Aphididae) of the Korean Peninsula: updating the global data set. *Molecular Ecology*  
504 *Resources* 11: 32-37.
- 505 35. Lee W, Lee Y, Kim H, Akimoto S-I, Lee S (2014) Developing a new molecular marker for  
506 aphid species identification: Evaluation of eleven candidate genes with species-level  
507 sampling. *Journal of Asia-Pacific Entomology* 17: 617-627.
- 508 36. Chen R, Jiang L-Y, Qiao G-X (2012) The effectiveness of three regions in mitochondrial  
509 genome for aphid DNA barcoding: a case in Lachninae. *PLoS ONE* 7: e46190.
- 510 37. Kim H, Lee S (2008) Molecular systematics of the genus *Megoura* (Hemiptera: Aphididae)  
511 using mitochondrial and nuclear DNA sequences. *Molecules and Cells* 25: 510-522.
- 512 38. Boukari W, Wei C, Tang L, Hincapie M, Naranjo M, Nuessly G, et al. (2020) Lack of  
513 transmission of *Sugarcane yellow leaf virus* in Florida from Columbus grass and  
514 sugarcane to sugarcane with aphids or mites. *PLoS ONE* 15: e0230066.

- 515 39. Paudyal S, Armstrong JS, Harris-Shultz KR, Wang H, Giles KL, Rott PC, et al. (2019)  
516 Evidence of host plant specialization among the US sugarcane aphid (Hemiptera:  
517 Aphididae) genotypes. Trends in Entomology 15: 47-58.

## 518 Supporting information

519 **S1 Fig. *Melanaphis* apterous female habitus.**

520 **S2 Fig. Evanno method inferring  $k = 2$ .**

521 **S1 Table. List of samples and specimens.**

522 **S2 Table. Comparison of morphological characteristics of *M. sacchari* and *M. sorghi***  
523 **apterous females.**

524 **S3 Table. Comparison of morphological characteristics of *M. sacchari* and *M. sorghi***  
525 **viviparous alate females.**

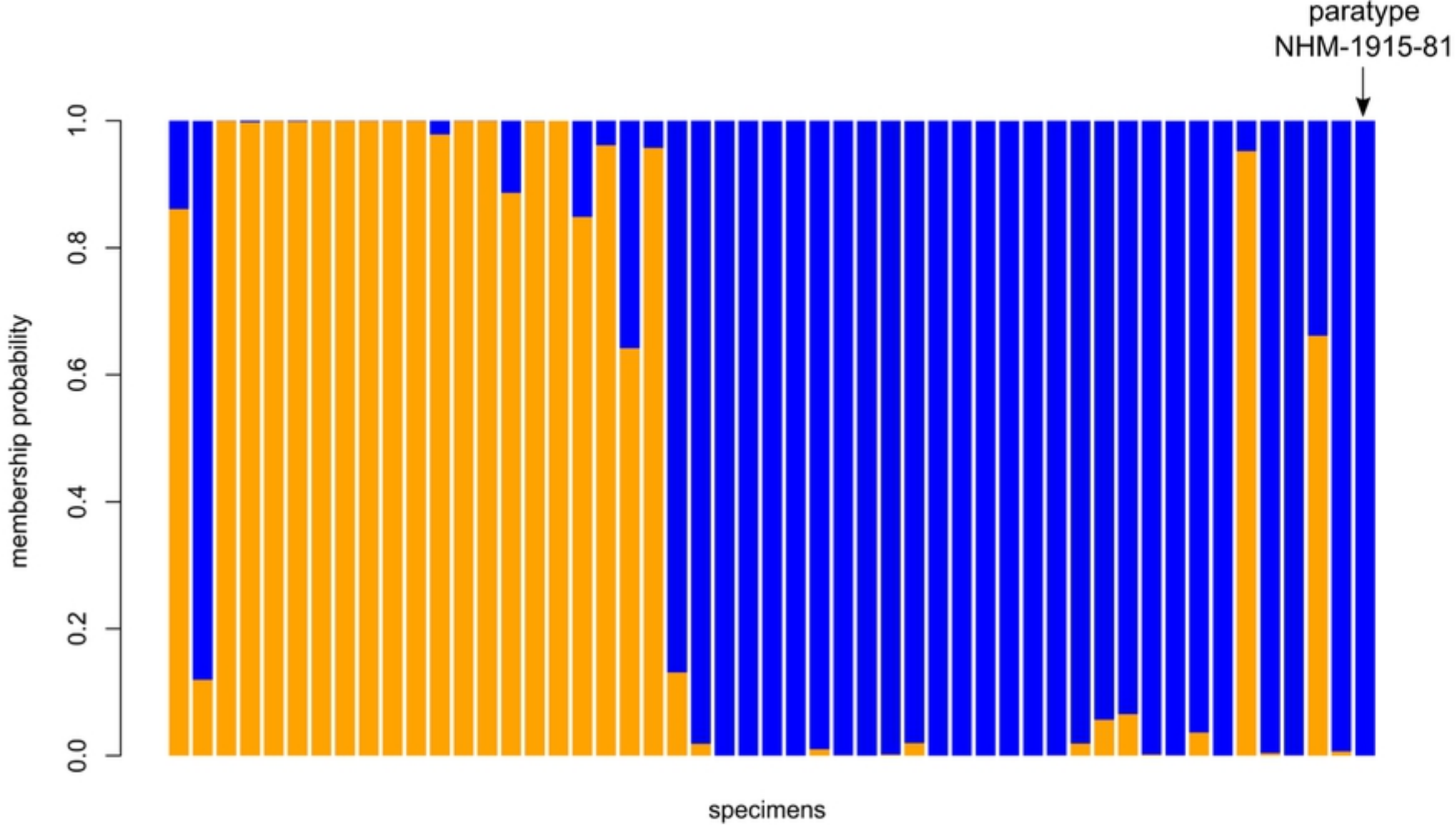
526 **S4 Table. CO1 haplotypes, position and nature of nucleotide substitutions.**

527 **S5 Table. CO2 haplotypes, position and nature of nucleotide substitutions.**

528 **S6 Table. EF1- $\alpha$  haplotypes, position and nature of nucleotide substitutions.**

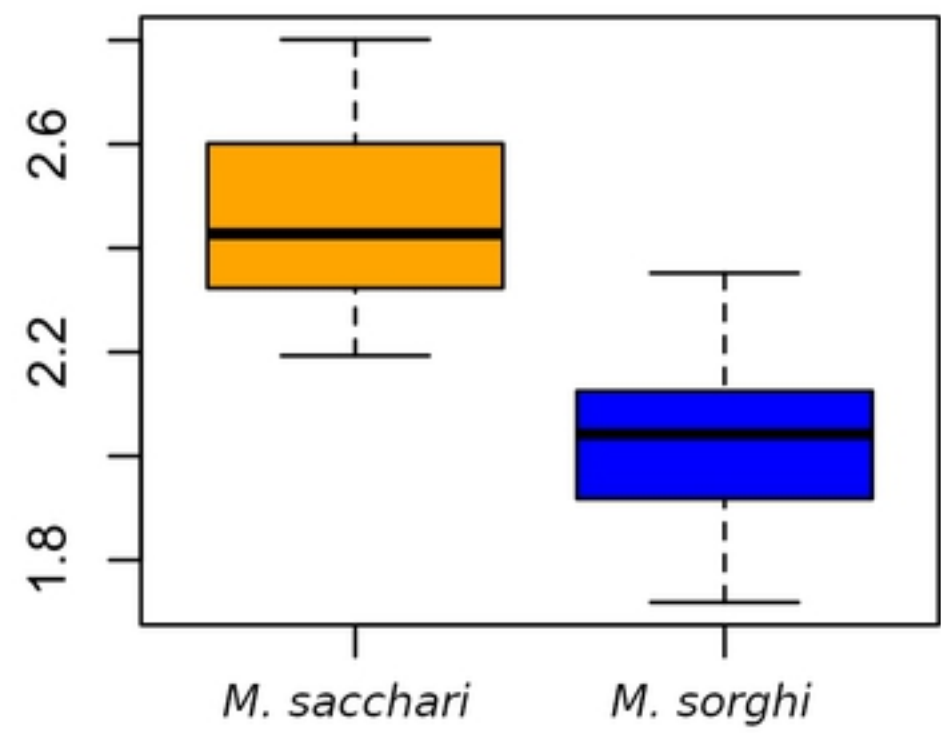
529 **S7 Table. Observed microsatellite Multi Locus Genotypes (MLG).**

530

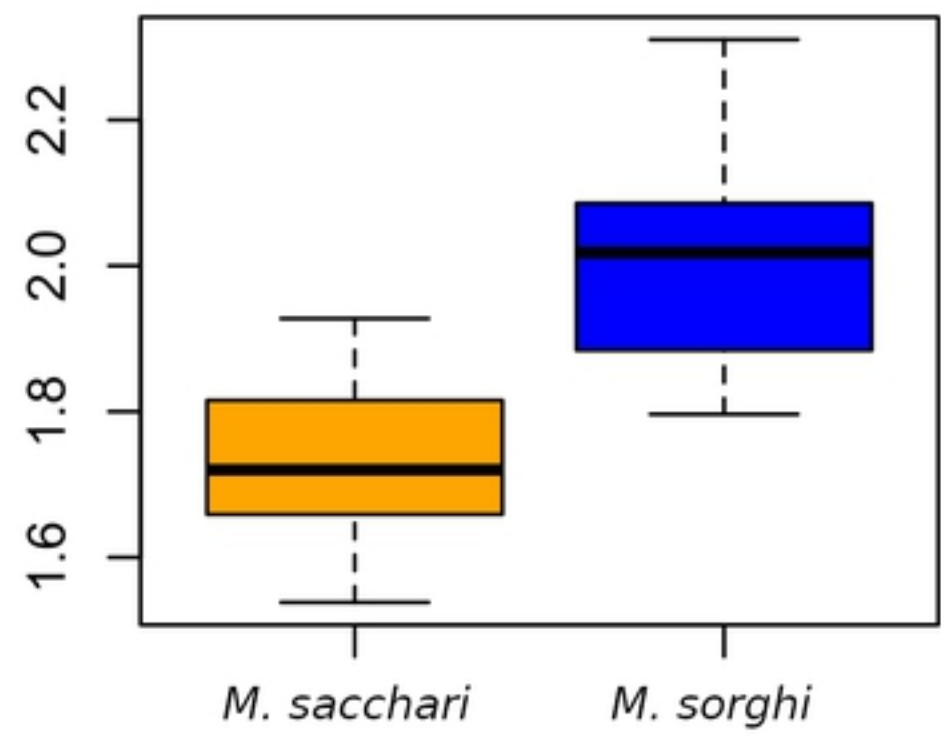




*pt:cauda*



*HindTibia:pt*



*pt:siph*

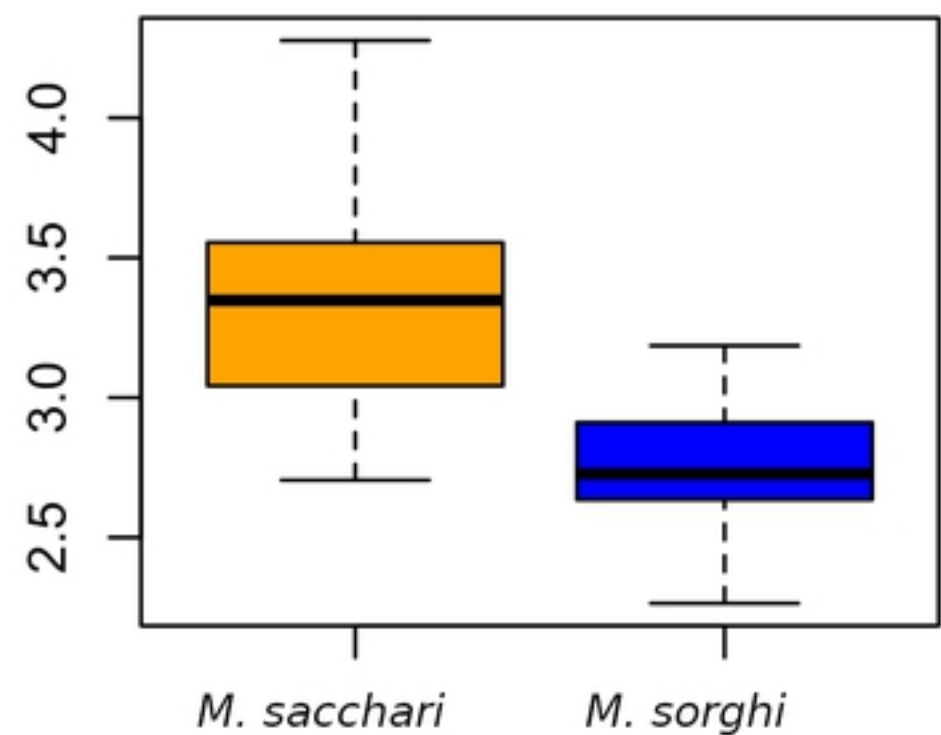


Fig 2

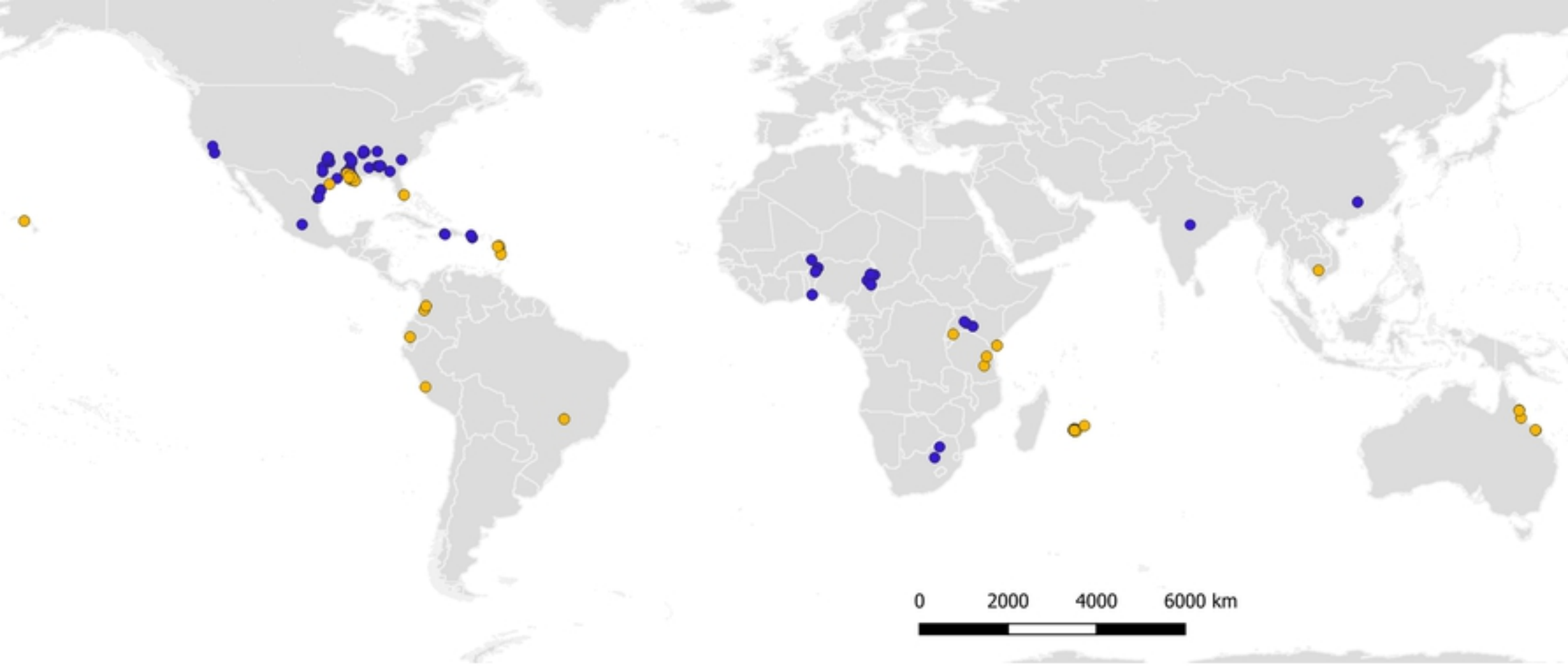


Fig 6

# COI

bioRxiv preprint doi: <https://doi.org/10.1101/2020.10.25.351833>; this version posted October 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

263 294  
ATAATATTAGATTTGA Y TTTTACCTCCATCATTAAATAATAAATTT R TAGATTTATAATTA



T = *M. sacchari*  
C = non-specific



A = *M. sacchari*  
G = non-specific

# EF1-alpha

637 663  
TTTTTATGAT W TGTAGTTTCTAATTTTAATTTATTT R TAGG

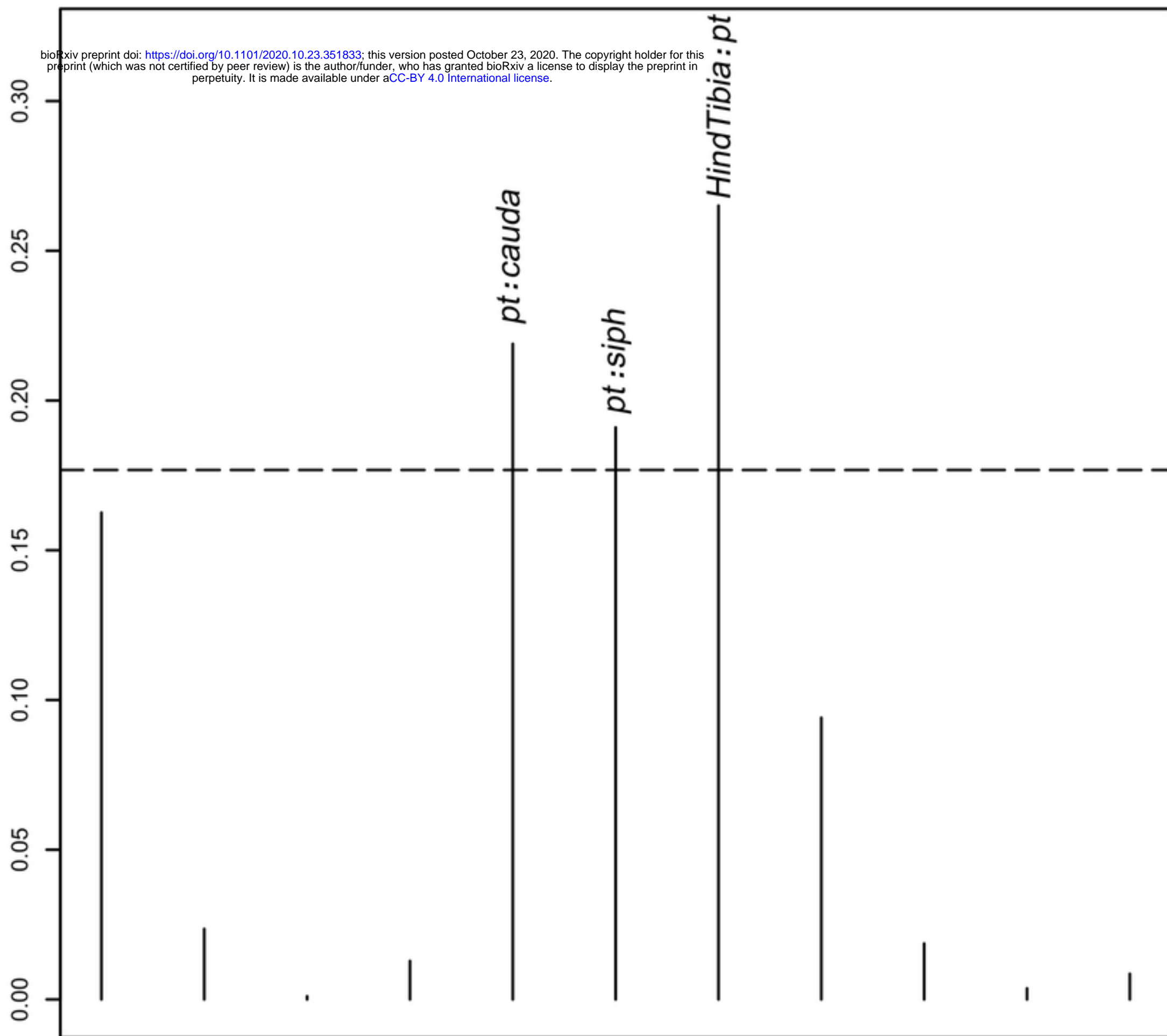


T = *M. sacchari*  
A = *M. sorghi*



A/G/R (homozygous or heterozygous)  
non-specific

Loadings



Morphological traits

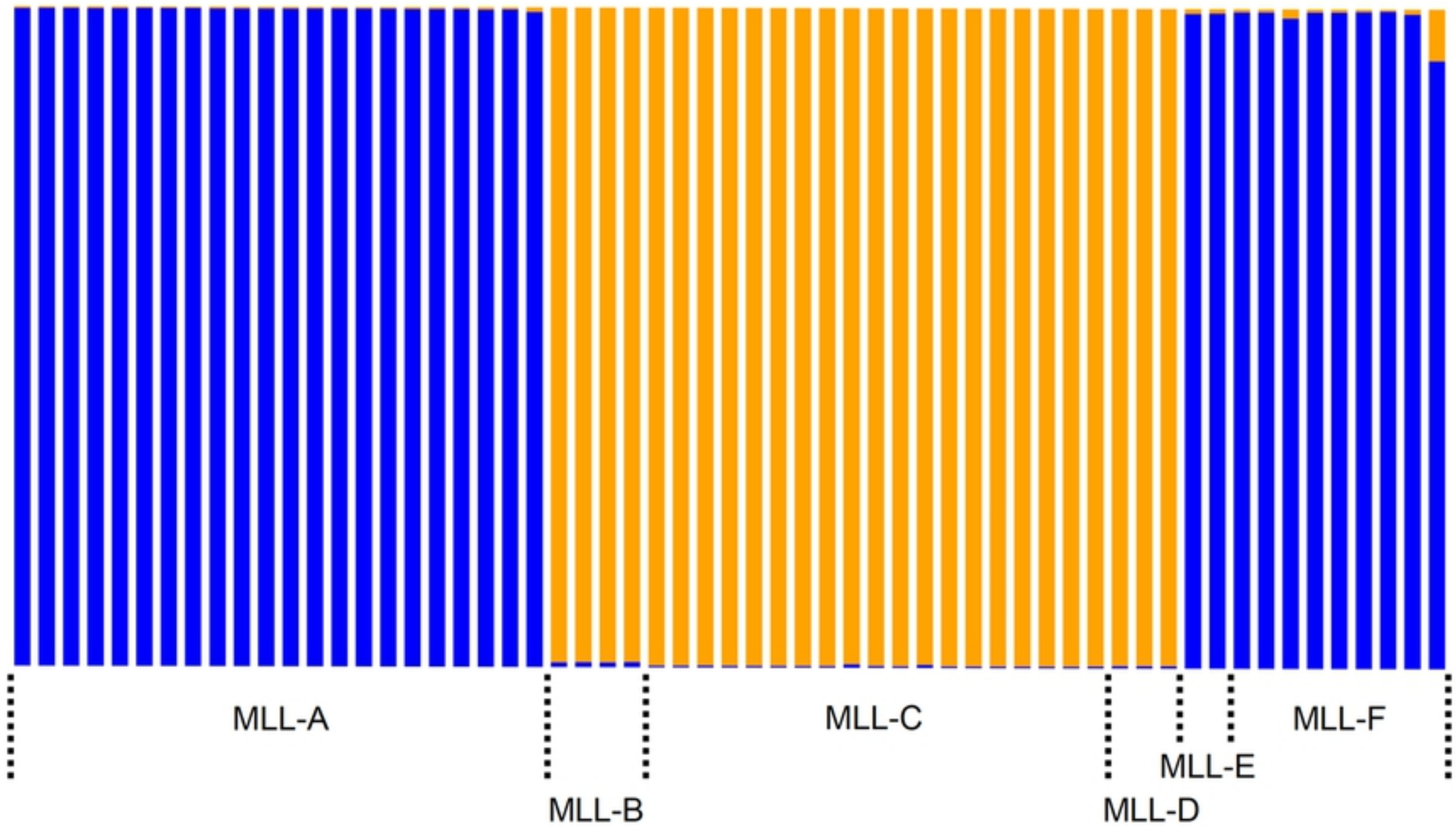


Fig 4

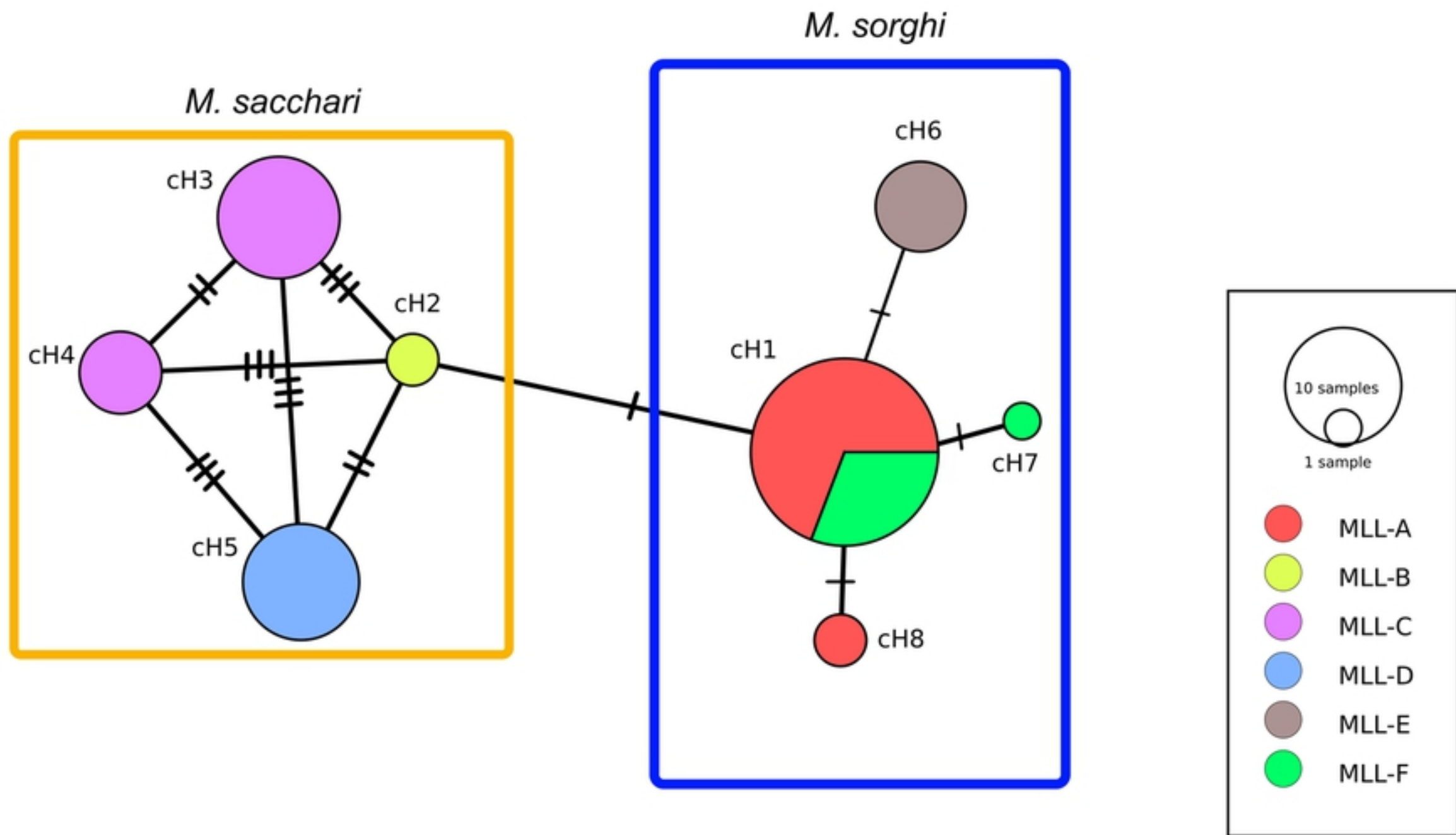


Fig 5

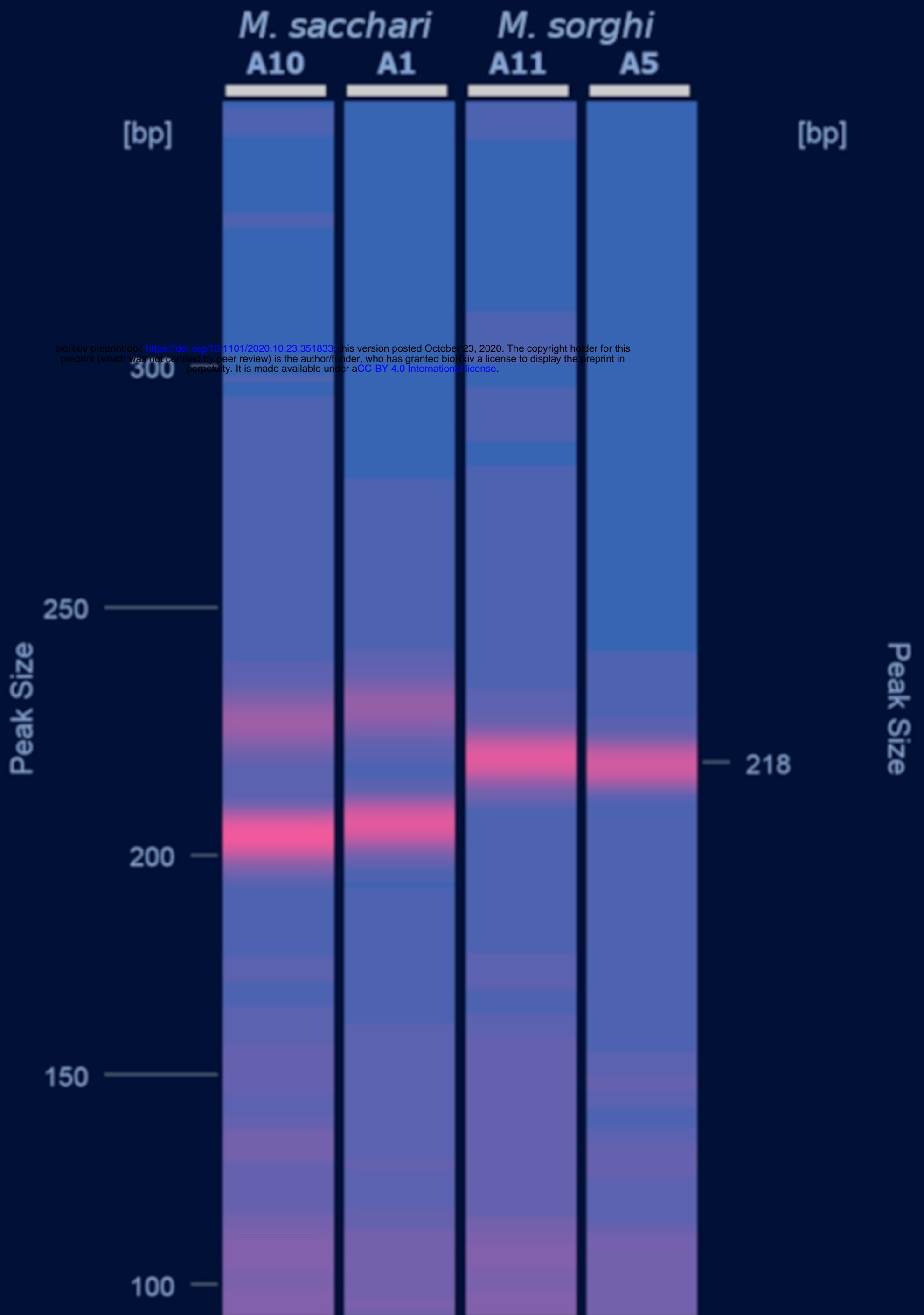


Fig 8