# 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

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

Samuel Nibouche[1]*, Laurent Costet[1], Raul F. Medina[2], Jocelyn R. Holt[2], Joëlle Sadeyen[3], Anne-Sophie Zoogones[1,3], Paul Brown[4], Roger L. Blackman[4]

[1] CIRAD, UMR PVBMT, F-97410 Saint Pierre, La Réunion, France

[2] Texas A&M University, College Station, Texas, United States of America

[3] Université de La Réunion, UMR PVBMT, F-97410 Saint Pierre, La Réunion, France

[4] The Natural History Museum, London SW7 5BD, United Kingdom

* Corresponding author

A-mail: samuel.nibouche@cirad.fr

# Abstract

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


# Introduction

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

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

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

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

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

# Material and Methods

## Material collected

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

## DNA extraction

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

## Microsatellites

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

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

# DNA sequencing

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

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

# Morphometry

## Material

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

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

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

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

## Methods

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

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

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

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

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

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

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

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

# Results

## Morphometry

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

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

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

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

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

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

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

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

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

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

## Microsatellites

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

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

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

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

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

| Structure cluster | MLL | level of MLL identification [b] | morphological assignment by DAPC | | |
|---|---|---|---|---|---|
| | | | *M. sacchari* | *M. sorghi* | undetermined [a] |
| orange | MLL-C | specimen | 1 | | 1 |
| | | sample | 9 | 1[c] | |
| | MLL-D | specimen | 8 | | |
| | | sample | 1 | | |
| blue | MLL-A | specimen | | | |
| | | sample | | 16 | |
| | MLL-F | specimen | 1[c] | 7 | |
| | | sample | | 4 | 1 |

[a] The specimens with a DAPC membership lower than 0.8 are considered undetermined.

[b] MLL identification was carried out by SSR genotyping on the specimen itself or on other specimens from the same sample.

[c] Specimens whose molecular and morphometric assignments are discordant.

# DNA sequencing

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

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

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

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

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

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

In the ITS2 region, in a 462 bp-sequence length, we detected one SNP and three 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-α, and were
290 also successfully genotyped with SSR. The relationship between MLL and haplotypes among
291 these specimens is presented in Table 3. EF1-α 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 | *sorghi* | A | 1 | 18 |
| cH1 | H1 | H3 | H2 | *sorghi* | F | 1 | 8 |
| cH2 | H1 | H3 | H1 | *sacchari* | B | 2 | 2 |
| cH3 | H2 | H1 | H1 | *sacchari* | C | 2 | 11 |
| cH4 | H2 | H5 | H1 | *sacchari* | C | 2 | 5 |
| cH5 | H3 | H3 | H1 | *sacchari* | D | 2 | 10 |
| cH6 | H1 | H2 | H2 | *sorghi* | E | 1 | 6 |
| cH7 | H6 | H3 | H2 | *sorghi* | F | 1 | 1 |
| cH8 | H1 | H4 | H2 | *sorghi* | 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 | EF1α | CytB | ITS2 | ITS1 |
|---|---|---|---|---|---|---|
| within *M. sorghi* | 0.15<br>(0.15 - 0.15)<br>n = 184 | 0.18<br>(0.13-0.26)<br>n = 70 | 0<br>-<br>n = 88 | 0<br>-<br>n = 11 | 0<br>-<br>n = 67 | 0.16<br>(0 - 0.26)<br>n = 18 |
| within *M. sacchari* | 0.30<br>(0.30 - 0.30)<br>n = 136 | 0.18<br>(0.13-0.26)<br>n = 73 | 0<br>-<br>n = 83 | 0.13<br>(0.13 - 0.13)<br>n = 32 | 0<br>-<br>n = 38 | 0.13<br>(0 - 0.26)<br>n = 20 |
| between species | 0.28<br>(0 - 0.45)<br>n = 320 | 0.18<br>(0 - 0.26)<br>n = 143 | 0.19<br>(0.19 - 0.19)<br>n = 171 | 0.09<br>(0 - 0.13)<br>n = 43 | 0<br>-<br>n = 105 | 0.13<br>(0 - 0.26)<br>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-α haplotype H1 or the COI
321 haplotypes H2 or H3 were assigned to *M. sacchari* while specimens bearing EF1-α 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

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

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

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

# Discussion

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

352 remains uncertain, although SSRs and EF1-α sequence data both suggest it belongs to *M.*
353 *sorghi*.

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. sorghi*
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. sorghi* 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. sorghi* with COI is sometimes not possible. With
365 EF1-α, 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. sorghi* 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. sorghi* using sequencing**
372 **of COI or EF1-α.**

373 **Fig 8. Molecular diagnosis for separation of *M. sacchari* and *M. sorghi* 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. sorghi* (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

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

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

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

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

# Acknowledgements

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

# References

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

435 9. Nibouche S, Fartek B, Mississipi 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, Mississipi 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 EF1α;
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 pioner 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) *adegenet*: 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.

30. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57: 289-300.

31. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

32. Massimino Cocuzza GE, Cavalieri V (2014) Identification of aphids of *Aphis frangulae*-group living on Lamiaceae species through DNA barcode. Molecular Ecology Resources 14: 447-457.

33. Rakauskas R, Havelka J, Zaremba A, Bernotienė R (2014) Mitochondrial COI and morphological evidence for host specificity of the black cherry aphids *Myzus cerasi* (Fabricius, 1775) collected from different cherry tree species in Europe (Hemiptera, Aphididae). ZooKeys 388.

34. Lee W, Kim H, Lim J, Choi HR, Kim Y, Kim YS, et al. (2010) Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. Molecular Ecology Resources 11: 32-37.

35. Lee W, Lee Y, Kim H, Akimoto S-I, Lee S (2014) Developing a new molecular marker for aphid species identification: Evaluation of eleven candidate genes with species-level sampling. Journal of Asia-Pacific Entomology 17: 617-627.

36. Chen R, Jiang L-Y, Qiao G-X (2012) The effectiveness of three regions in mitochondrial genome for aphid DNA barcoding: a case in Lachininae. PLoS ONE 7: e46190.

37. Kim H, Lee S (2008) Molecular systematics of the genus *Megoura* (Hemiptera: Aphididae) using mitochondrial and nuclear DNA sequences. Molecules and Cells 25: 510-522.

38. Boukari W, Wei C, Tang L, Hincapie M, Naranjo M, Nuessly G, et al. (2020) Lack of transmission of *Sugarcane yellow leaf virus* in Florida from Columbus grass and 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-α haplotypes, position and nature of nucleotide substitutions.**

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

530

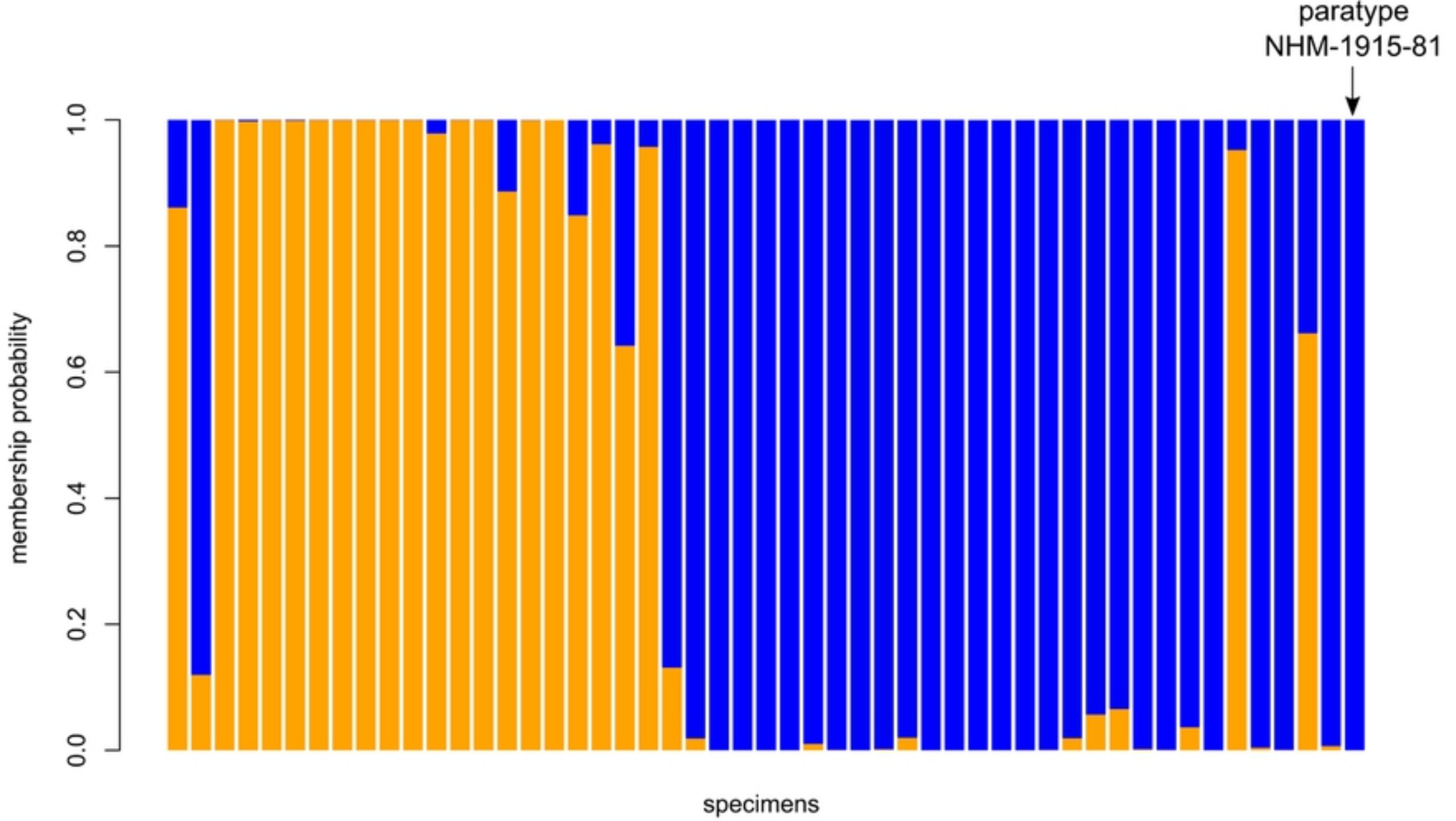
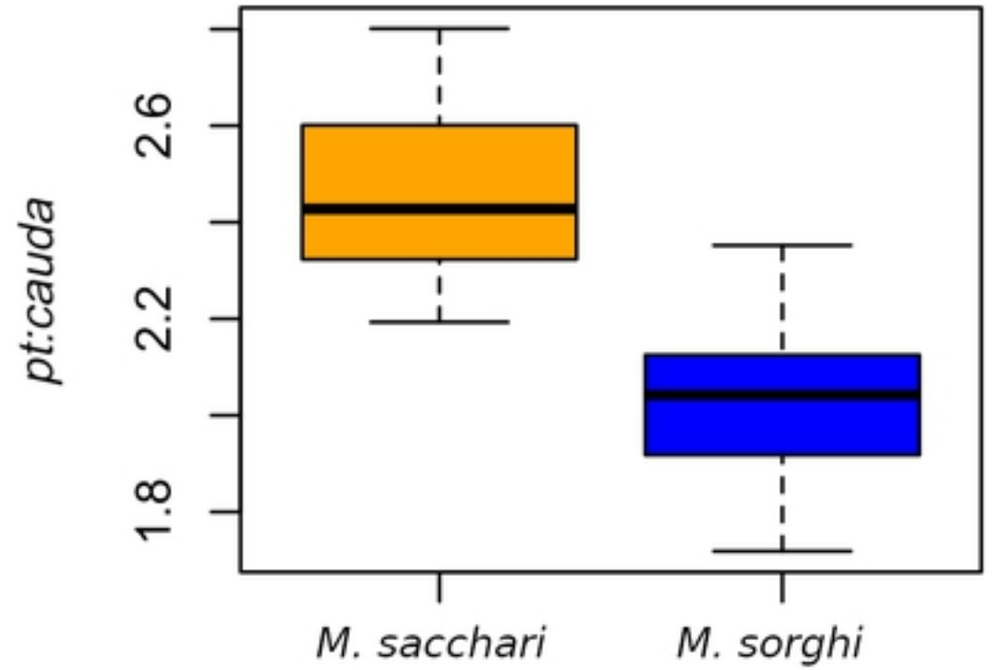
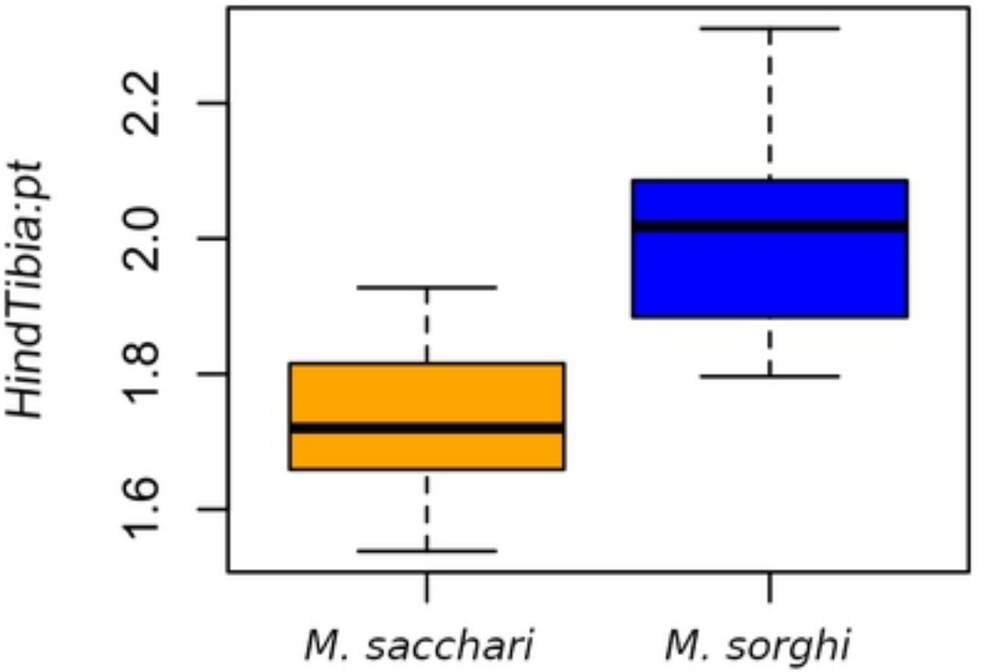

paratype
NHM-1915-81
membership probability
0.0
0.2
0.4
0.6
0.8
1.0
specimens

Fig 1


pt:cauda
1.8
2.2
2.6
M. sacchari
M. sorghi


HindTibia:pt
1.6
1.8
2.0
2.2
M. sacchari
M. sorghi

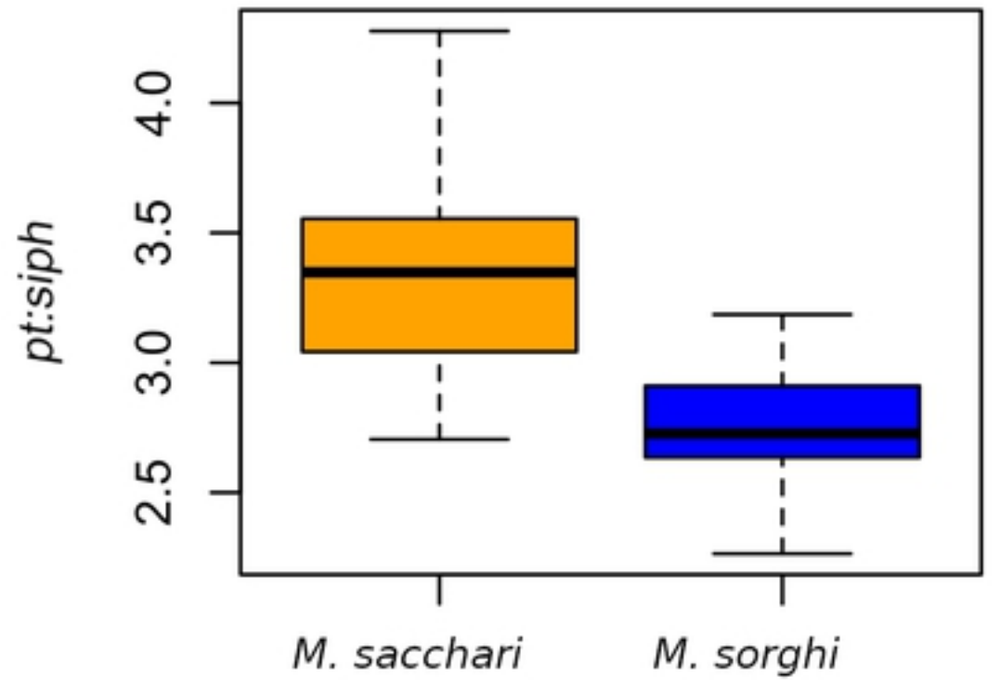

pt:siph
2.5
3.0
3.5
4.0
M. sacchari
M. sorghi

Fig 2

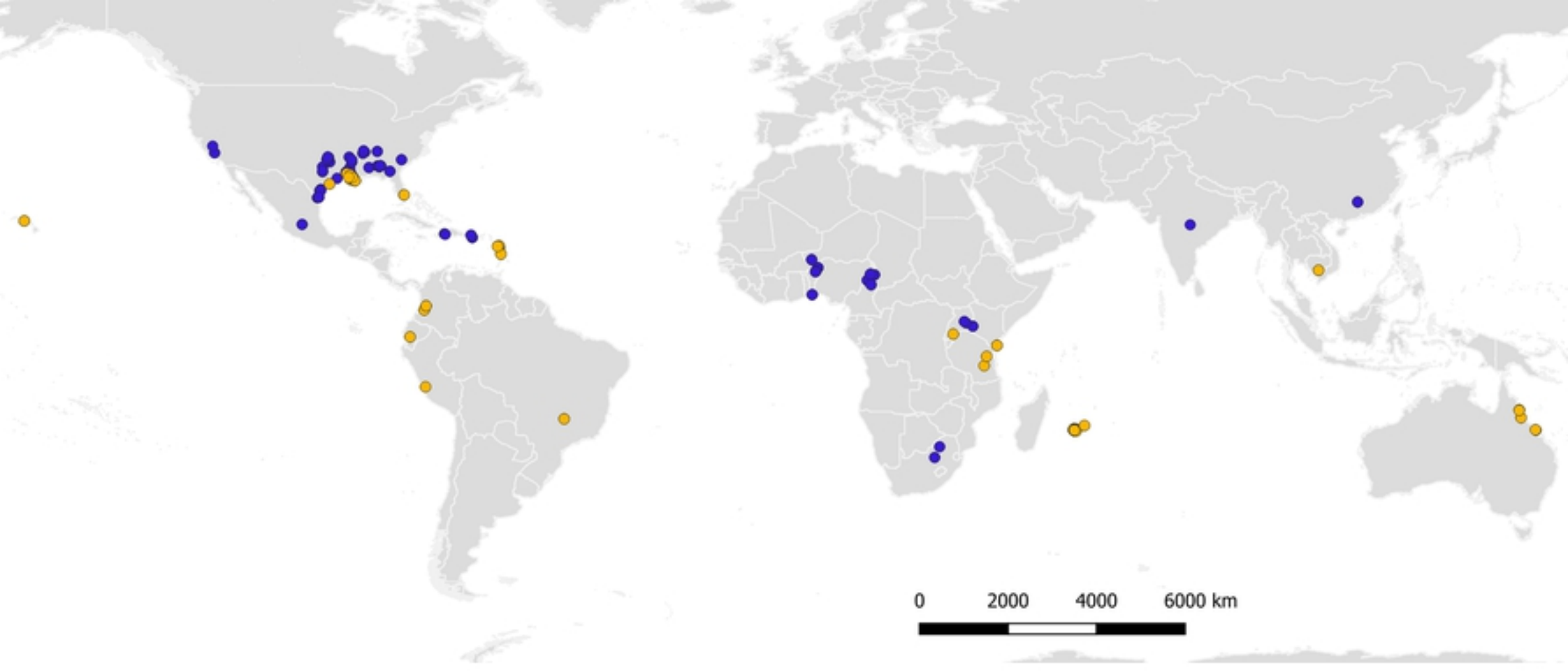
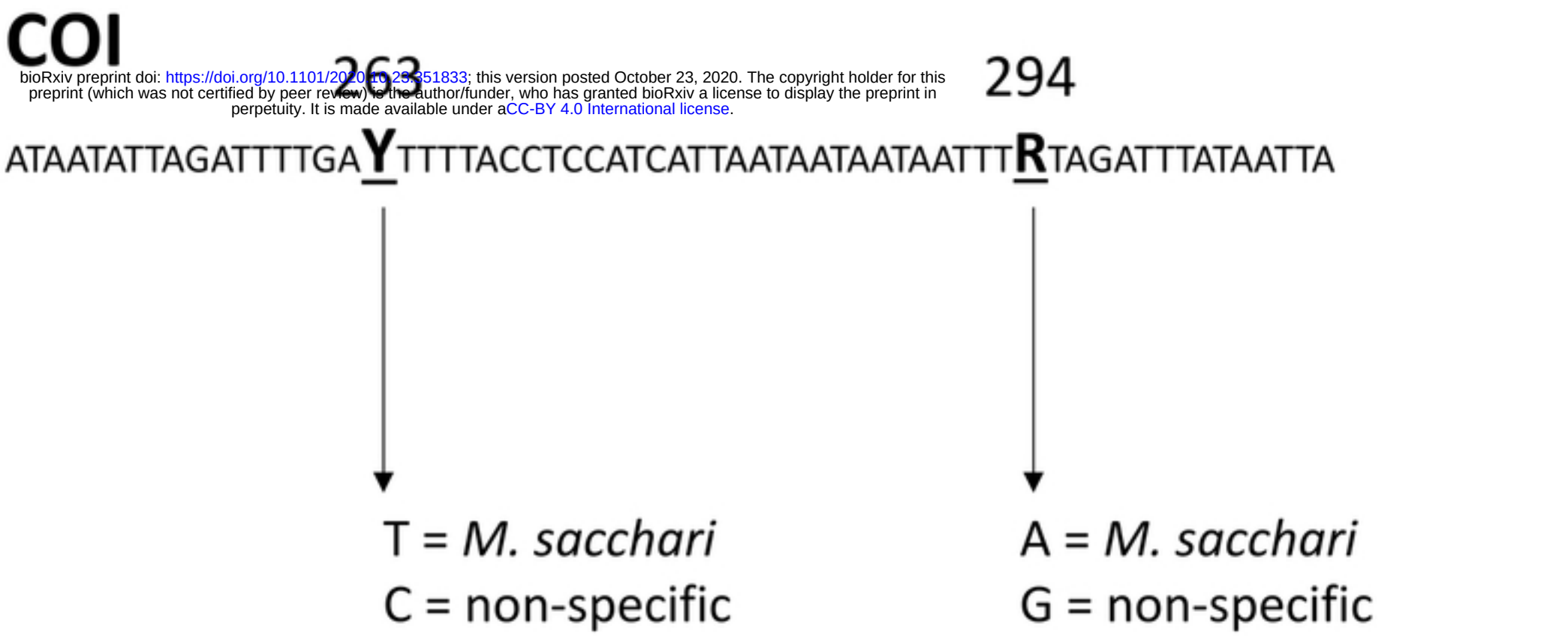



Fig 6



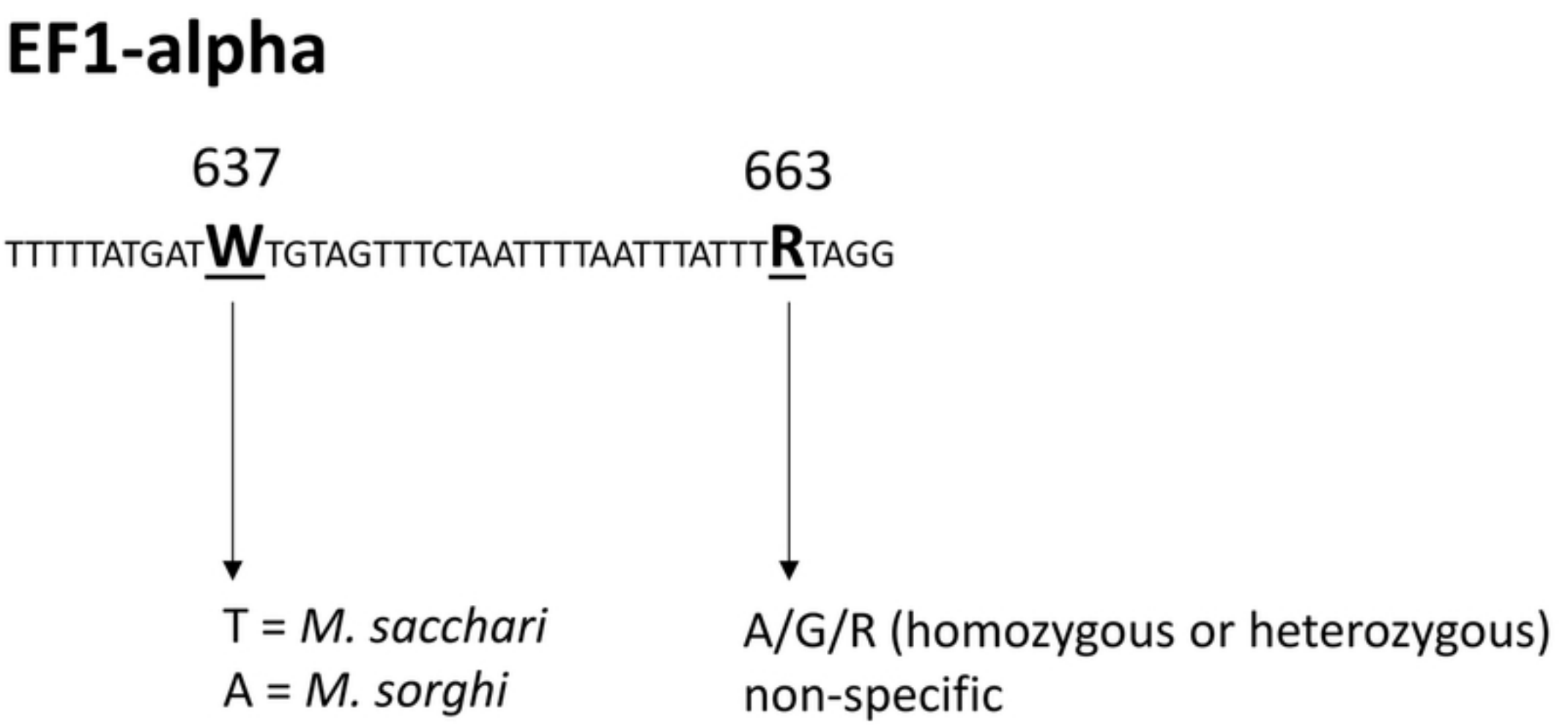




Fig 7

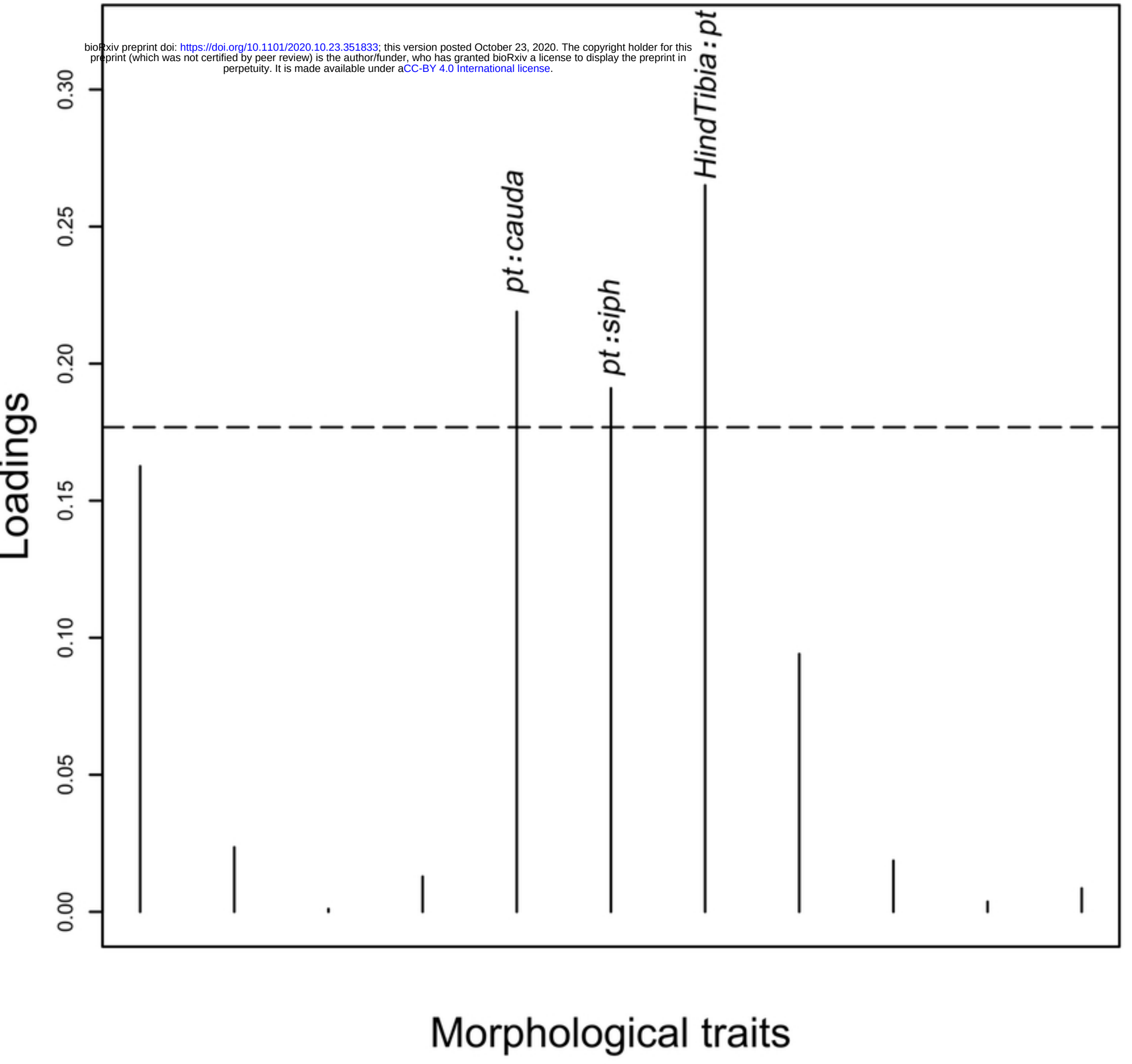



Fig 3

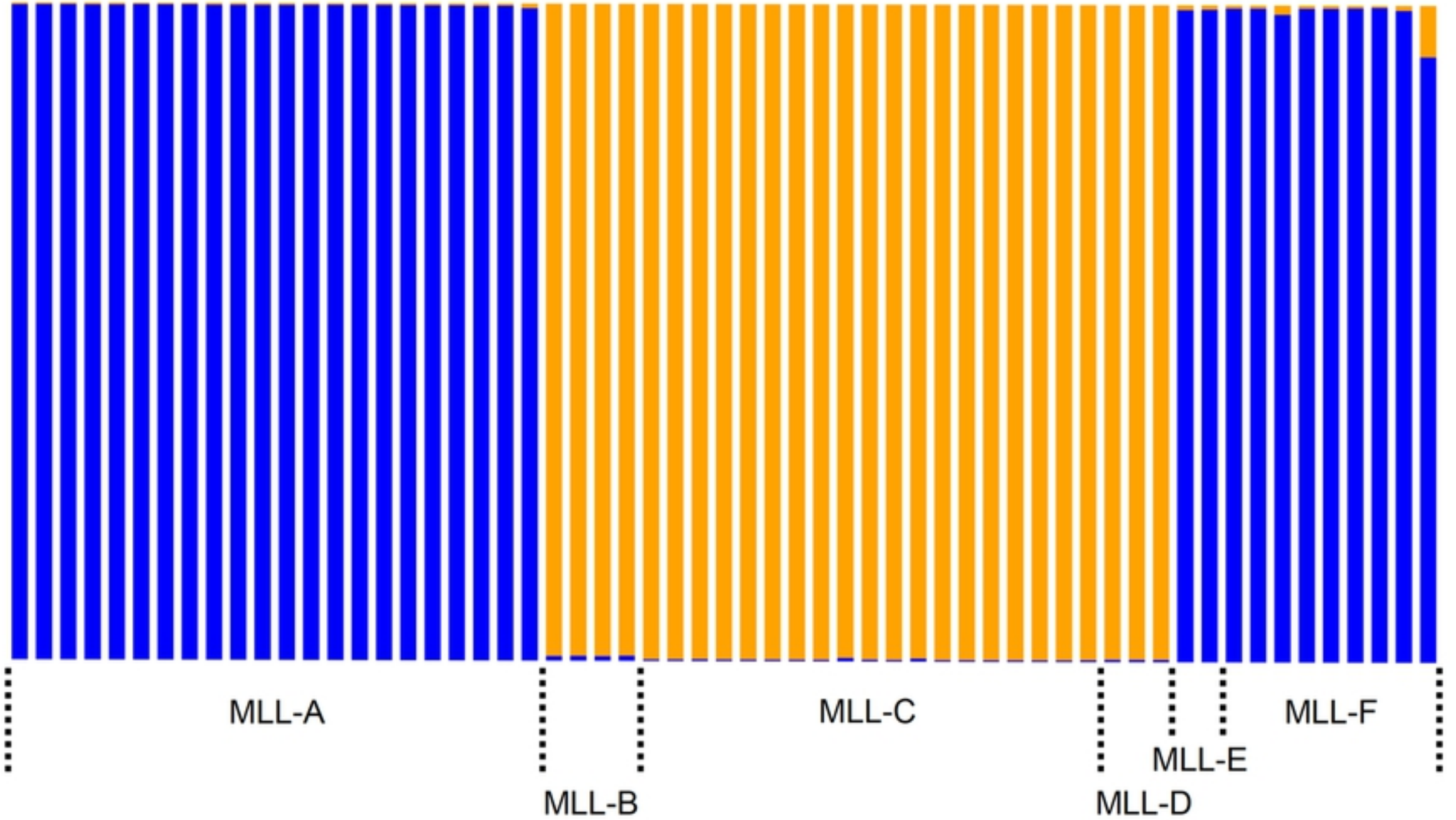



Fig 4

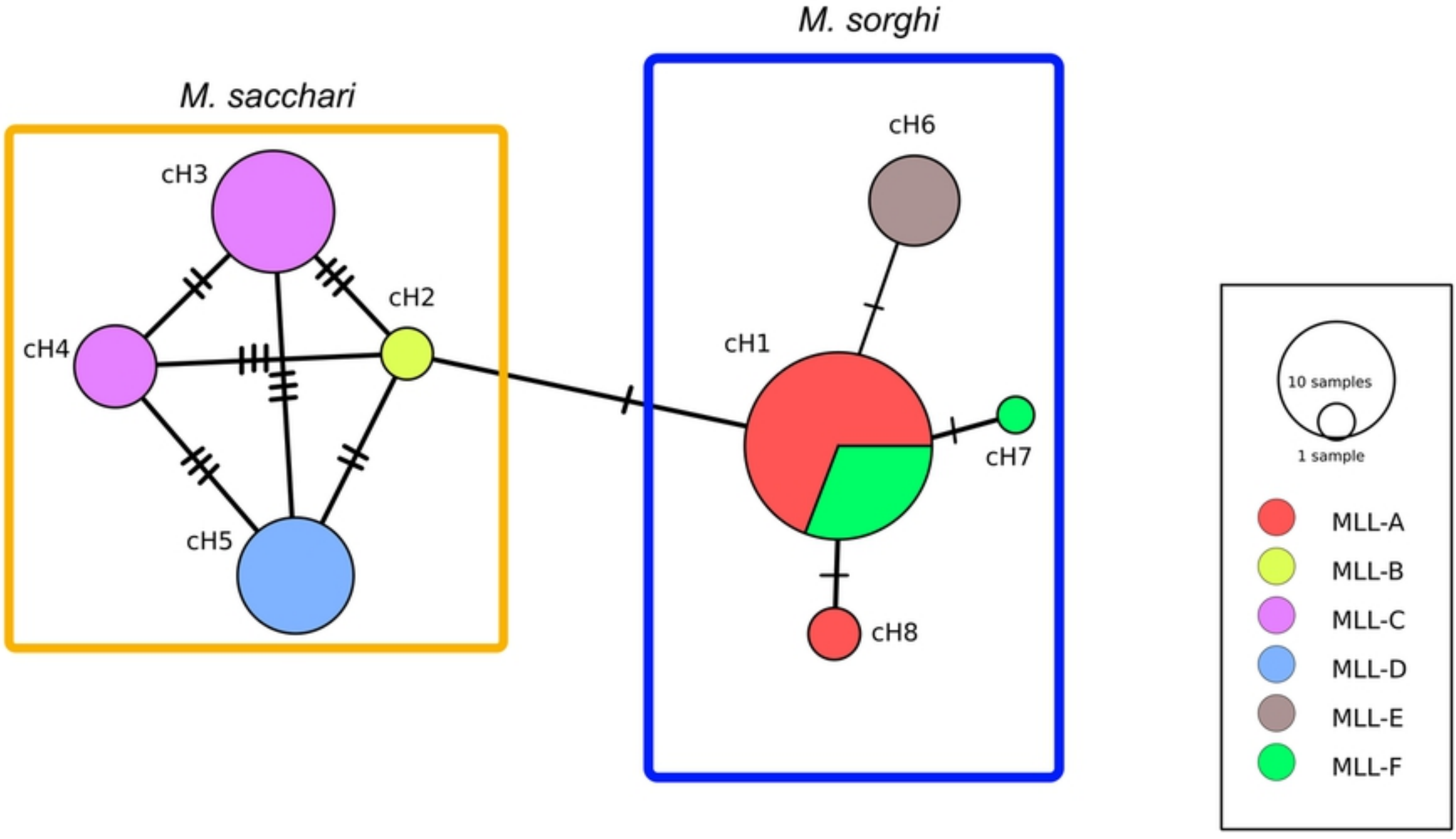

M. sacchari
M. sorghi
cH3
cH2
cH4
cH5
cH6
cH1
cH7
cH8
10 samples
1 sample
MLL-A
MLL-B
MLL-C
MLL-D
MLL-E
MLL-F

Fig 5

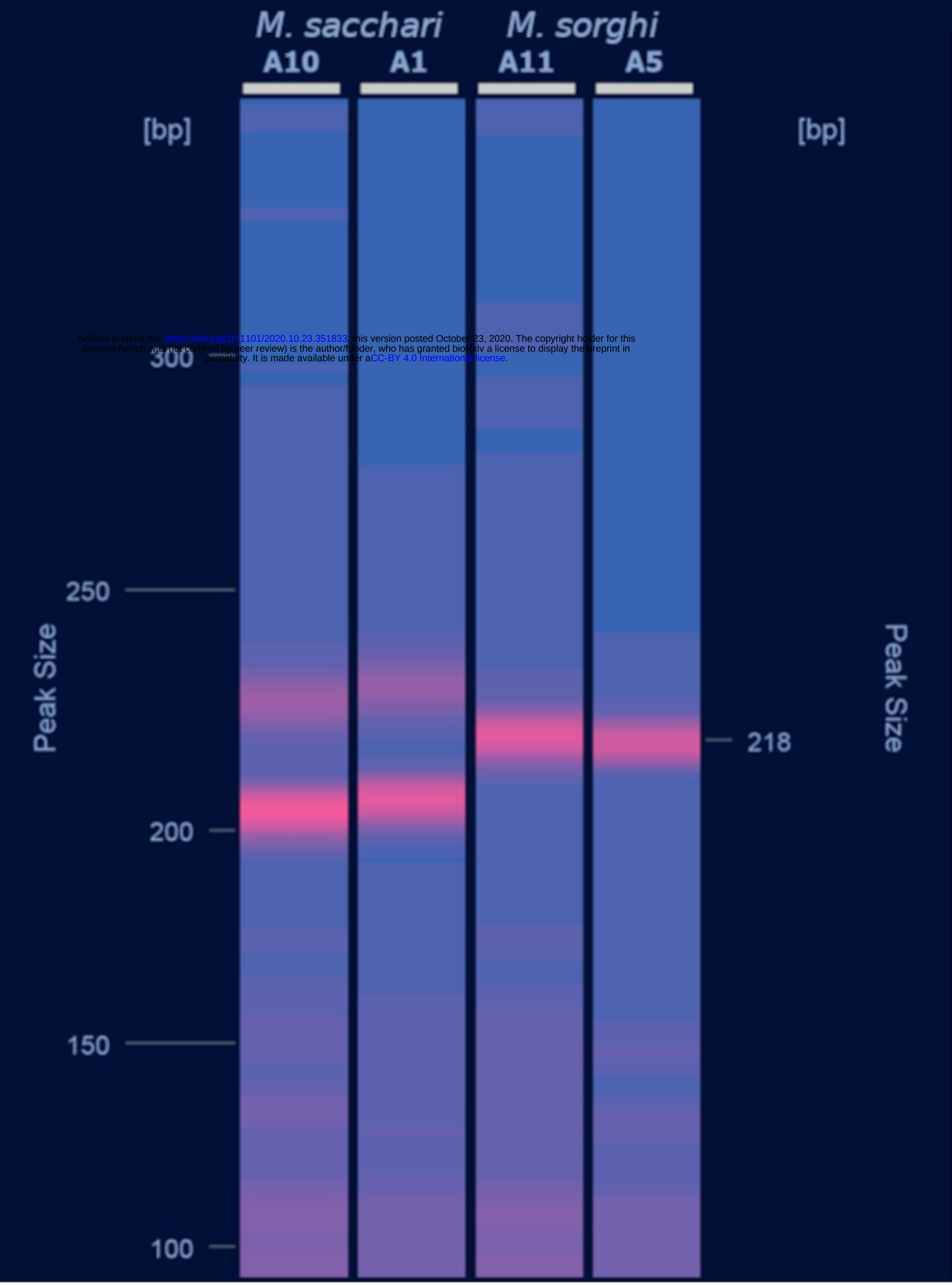



Fig 8