Identification of French Guiana anopheline mosquitoes by MALDI-TOF MS profiling using protein signatures from two body parts.

Sébastien Briolant[1,2], Monique de Melo Costa[1,2], Christophe Nguyen[1,2], Isabelle Dusfour[3], Vincent Pommier de Santi[4], Romain Girod[3,5], Lionel Almeras[1,2]

[1]Unité de Parasitologie et Entomologie, Département de Microbiologie et Maladies Infectieuses, Institut de Recherche Biomédicale des Armées, Marseille, France.

[2]Aix Marseille Université, IRD, AP-HM, SSA, UMR Vecteurs – Infections Tropicales et Méditerranéennes (VITROME), IHU - Méditerranée Infection, 19-21 bd Jean Moulin, 13385 Marseille, cedex 5, France.

[3]Unite d'Entomologie Médicale, Institut Pasteur de la Guyane, Cayenne, French Guiana.

[4]Centre d'Epidémiologie et de Santé Publique des Armées, Marseille, France.

[5]Medical Entomology Unit, Institut Pasteur de Madagascar, P.O. 1274, Ambatofotsikely – 101 Antananarivo, Madagascar

***Corresponding author:** Dr. Lionel ALMERAS. Unité de Parasitologie et Entomologie (IRBA), Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin 13385 Marseille cedex 05, France. Phone: 33 (0) 4 91 32 43 75. Fax: 33 (0) 4 91 83 03 90. E-mail address: almeras.lionel@gmail.com.

Authors' emails: SB: sbriolant@wanadoo.fr; MMC: mcosta.monique@gmail; CNG: christophenguyen2005@yahoo.fr; ID: isabelle.dusfour@pasteur.fr; VPS: v.pommierdesanti@gmail.com; RG: rgirod@pasteur.mg; LA: almeras.lionel@gmail.com.

Running title: French Guiana Anopheline identification by MALDI-TOF MS

Abstract

In French Guiana, the malaria, a parasitic infection transmitted by *Anopheline* mosquitoes, remains a disease of public health importance. To prevent malaria transmission, the main effective way remains *Anopheles* control. For an effective control, accurate *Anopheles* species identification is indispensable to distinguish malaria vectors from non-vectors. Although, morphological and molecular methods are largely used, an innovative tool, based on protein pattern comparisons, the Matrix Assisted Laser Desorption / Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) profiling, emerged this last decade for arthropod identification. However, the limited mosquito fauna diversity of reference MS spectra remains one of the main drawback for it large usage. The aim of the present study was then to create and to share reference MS spectra for the identification of French Guiana *Anopheline* species. A total of eight distinct *Anopheles* species, among which four are malaria vectors, were collected in 6 areas. To improve *Anopheles* identification, two body parts, legs and thoraxes, were independently submitted to MS for the creation of respective reference MS spectra database (DB). This study underlined that double checking by MS enhanced the *Anopheles* identification confidence and rate of reliable classification. The sharing of this reference MS spectra DB should made easier *Anopheles* species monitoring in endemic malaria area to help malaria vector control or elimination programs.

Keywords: *Anopheles*; Identification; French Guiana; MALDI-TOF MS; Innovative strategy.

Introduction

Since 2005, malaria cases declined significantly in French Guiana, an oversea territory of France located in South-America. The number of diagnosed cases has decreased from 4,479 cases in 2005 to 597 cases in 2017 [1]. However, the disease is still endemic in the inland forested areas, especially in illegal gold mining areas and so remains of public health importance [2–4]. Majority of the cases are caused by *Plasmodium vivax* (89%, p=531/597), followed by *P. falciparum* (11%, p=66/597). *Anopheles* mosquitoes are known for their role in transmitting malaria. Historically, thirty-three mosquitoes from the genus *Anopheles* have been reported in French Guiana [5]. Members in the subgenus *Anopheles* and *Nyssorhynchus* have been implicated in malaria transmission in French Guiana. *Anopheles darlingi* is the recognized primary vector in the territory [6–9]. Recently, *An. nuneztovari sl*, *An. oswaldoi sl*, *An. intermedius*, *An. marajoara* and *An. ininii* were found naturally infected with *Plasmodium* sporozoites and were suspected to be secondary vectors [2,3,10]. Malaria transmission is further complicated as some of these secondary vectors belong to species complexes, characterized by similar morphological characteristics, such as *An. oswaldoi* [11], *An. marajoara* [12] and *An. nuneztovari* [13].

A rapid and accurate identification of *Anopheles* species is then critical when designing malaria vector control strategies which should be species-specific to be effective. The most common method for mosquito species identification remains the utilization of morphological criteria [14]. However, morphological identification is skill dependent requiring entomological expertise. The correct species assignation could also be compromised for damaged specimens with the loss of determinant characters. Moreover, the description of morphological characteristic variations between intact adult conspecific specimens underlined that correct mosquito species classification could be held only by experienced mosquito taxonomists [15]. Additionally, the availability of mosquito taxonomic keys is a cornerstone for identification success. Unfortunately, morphological keys are still missing for numerous *Anopheles* species, notably for closely related-species (cryptic and complex species).

The emergence of molecular biology approaches in the 2000's has solved some long-standing taxonomic questions [16]. However, the choice of the target gene sequence for accurate mosquito identification could be complex. In South America, *Anopheles* species were classified by using different target genes: either 18S rRNA [17], either the second internal transcribed spacer (ITS2) [18] or cytochrome c oxidase I (COI) [19]. The absence of consensus for mosquito species identification complicate studies comparison. The development of the Barcode of Life Data (BOLD) system contributed to standardize gene sequencing for organism identification [20]. Nevertheless, for some cryptic mosquito species (eg, *An. gambiae* complex), the sequencing of COI could be insufficient to classify them unambiguously [21] and the sequencing of a second gene could be required [22]. For instance, some *Anopheles* and *Culex* sibling species could not be distinguished using uniquely the mitochondrial COI barcode [23,24]. Moreover, despite the large advances of this strategy this last decade, notably by shortening experiment duration and costs of reagents, gene sequencing remains time-consuming and expansive [25]. The development of a quick and low cost approach for mosquito monitoring with elevate rate of reliable identification is always in high demand.

The MALDI-TOF MS profiling have recently demonstrated its performance for reliable arthropod identification [25], including mosquitoes at adult [26] and immature stages [27,28]. At adult stage, for specimen identification by MS, different mosquito body parts were selected such as the cephalothorax [29,30] or legs, but this last compartment remains the more frequently used [26,31–33]. The recent successful identification of mosquitoes at immature stages by MALDI-TOF MS profiling validated the efficiency of this MS strategy for field monitoring of *Culicidae* [19]. The regent low costs, the rapidity and technical simplicity of protocols participated to the success of this approach. However, conversely to molecular analyses, MS protein profiles from conspecific specimens could vary according to several factors such as sample storing mode, developmental stage, homogenization mode or body part used [25,34]. To overcome these limitations, standardized protocols were established for some arthropod families [35] including mosquitoes [36]. The standardization of the protocols facilitated result comparisons and reference MS spectra sharing.

The main problem with legs is that they are breakable. The loss of one to all mosquito legs during trapping and/or storing is not infrequent, which could compromise specimen identification by MS. The selection of a second body part, the mosquito thorax, revealed that it could also generate mosquito species-specific MS but distinct from legs of conspecific specimens [37]. This last study underlined that the query of these two body parts against the reference MS spectra database (DB) improved mosquito species identification with accuracy and confidence. This pioneering study assessed to distinct 7 mosquito species from 4 genera living in sympatry in Guadeloupe Island [38].

The aim of the present study was to assess whether the submission of two body part could improve *Anopheles* identification from French Guiana. The creation of a reference MS spectra DB should made easier *Anopheles* species monitoring in endemic malaria area to help malaria vector control or elimination programs.

Methods

Mosquito collection and dissection

Anopheles adult female mosquitoes were selected from field mosquito collections done in 6 distinct sites from French Guiana, during entomological surveys, using different collection methods, over different sampling periods (Figure 1) [3,39–41]. After collections, mosquitoes were sorted by genera and *Anopheles* mosquitoes were morphologically identified under a binocular loupe at a magnification of ×56 (Leica M80, Leica, Nanterre, France) using standard taxonomic keys for the region (Floch and Abonnenc 1951, Forattini 1962, Faran and Linthicum 1981). *Anopheles* specimens were then stored individually at -20°C. According to their availability, one to 22 specimens per *Anopheles* species were selected for molecular and MS analyses (Table 1). Legs and thoraxes from mosquitoes were dissected for MALDI-TOF MS analysis as previously described [38]. The abdomens, wings and heads were kept for molecular analyses.

Molecular identification of mosquitoes

DNA was individually extracted from the head and abdomen of all mosquito specimens (n=112) using the QIAamp DNA tissue extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Molecular identification of mosquito at the species level was performed by sequencing the PCR product of a fragment of the cytochrome c oxidase I gene (*COI*) (LCO1490 (forward): 5'-GGTCAACAAATCATAAAGATATTGG-3'; HC02198 (reverse): 5'-TAAACTTCAGGGTGACCAAAAAATCA-3') as previously described [19,42]. The sequences were assembled and analyzed using the Molecular Evolutionary Genetics Analysis (MEGA) software version 7.0 and BioEdit Sequence alignment editor software version 7.2.6.0. All sequences were blasted against GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and against the Barcode of Life Data Systems (BOLD; http://www.barcodinglife.org; [20]) to assign unknown *COI* sequences to mosquito species.

133 *Sample homogenization and MALDI-TOF MS analysis*

134 Each dissected compartment (legs and thoraxes) was homogenized individually 3 x 1 minute at 30
135 Hertz using TissueLyser (Qiagen) and glass beads (#11079110, BioSpec Products, Bartlesville, OK,
136 US) in a homogenization buffer composed of a mix (50/50) of 70% (v/v) formic acid (Sigma) and
137 50% (v/v) acetonitrile (Fluka, Buchs, Switzerland) according to the standardized automated setting
138 as described previously [43]. After sample homogenization, a quick spin centrifugation at 200 g for
139 1 min was then performed and 1 µL of the supernatant of each sample was spotted on the MALDI-
140 TOF steel target plate in quadruplicate (Bruker Daltonics, Wissembourg, France). After air-drying, 1
141 µL of matrix solution composed of saturated α-cyano-4-hydroxycinnamic acid (Sigma, Lyon,
142 France), 50% (v/v) acetonitrile, 2.5% (v/v) trifluoroacetic acid (Aldrich, Dorset, UK) and HPLC-
143 grade water was added. To control matrix quality (i.e. absence of MS peaks due to matrix buffer
144 impurities) and MALDI-TOF apparatus performance, matrix solution was loaded in duplicate onto
145 each MALDI-TOF plate alone and with a bacterial test standard (Bruker Bacterial Test Standard, ref:
146 #8255343).
147
148 *MALDI-TOF MS parameters*
149 Protein mass profiles were obtained using a Microflex LT MALDI-TOF Mass Spectrometer (Bruker
150 Daltonics, Germany), with detection in the linear positive-ion mode at a laser frequency of 50 Hz
151 within a mass range of 2-20 kDa. The setting parameters of the MALDI-TOF MS apparatus were
152 identical to those previously used [44].
153
154 *MS spectra analysis*
155 MS spectra profiles were firstly controlled visually with flexAnalysis v3.3 software (Bruker
156 Daltonics). MS spectra were then exported to ClinProTools v2.2 and MALDI-Biotyper v3.0. (Bruker
157 Daltonics) for data processing (smoothing, baseline subtraction, peak picking). MS spectra
158 reproducibility was assessed by the comparison of the average spectral profiles (MSP, Main Spectrum

Profile) obtained from the four spots for each specimen according to body part with MALDI-Biotyper v3.0 software (Bruker Daltonics). MS spectra reproducibility and specificity taking into account mosquito body part were objectified using cluster analyses. Cluster analyses (MSP dendrogram) were performed based on comparison of the MSP given by MALDI-Biotyper v3.0. software and clustered according to protein mass profile (i.e. their mass signals and intensities). In addition, to visualize MS spectra distribution according to body part, principal component analysis (PCA) from ClinProTools v2.2 software were performed for each species.

The top-10 and top-5 of the most intense MS peaks per mosquito species and per body-part were analyzed with ClinProTools software to estimate their performance to discriminate the *Anopheles* species for each body-part. The parameter settings in ClinProTools software for spectrum preparation were as follows: a resolution of 300; a noise threshold of 2.00; a maximum peak shift of 800 ppm and a match to calibrating agent peaks of 10%. Peak calculation and selection were performed on individual spectra with a signal-to-noise threshold of 2.00 and an aggregation of 800 ppm. Based on the peak list obtained for each body part per species, the top-10 and top-5 of the most intense m/z peaks were selected to include them into the genetic algorithm (GA) model. The selected peaks by the operator gave a recognition capability (RC) value together with the highest cross-validation (CV) value. The presence or absence of all discriminating peak masses generated by the GA model was controlled by comparing the average spectra from each species per body-part.

Database creation and blind tests

The reference MS spectra were created using spectra from legs and thorax of two specimens per species when available using MALDI-Biotyper software v3.0. (Bruker Daltonics) [26]. MS spectra were created with an unbiased algorithm using information on the peak position, intensity and frequency. MS spectra from mosquito legs and thoraxes were tested against the in-house MS reference spectra DB, including already legs and thoraxes reference MS spectra from eight distinct mosquito species, but none from the *Anopheles* genus [38]. The reliability of species identification

was estimated using the log score values (LSVs) obtained from the MALDI Biotyper software v.3.0, which ranged from 0 to 3. According to previous studies [26,44], LSVs greater than 1.8 were considered reliable for species identification. Data were analyzed by using GraphPad Prism software version 5.01 (GraphPad, San Diego, CA, USA).

Phylogenetic analyses

After gene sequences alignment with the Clustal ω2 algorithm in the MEGA 7.0 software, a maximum likelihood tree based on the COI gene were constructed using the MEGA 7.0 software [45]. The tree with the highest log likelihood was kept. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Support for internal nodes was estimated using the nonparametric bootstrap method with 1000 replications.

Results

Morphological identification and molecular validation

Among the mosquitoes captured in the 6 distinct sites from French Guiana (Figure 1), uniquely anopheline specimens were selected. These mosquitoes were classified morphologically into eight distinct species, six from the *Nyssorhynchus* subgenus (*An. aquasalis*, *An. braziliensis*, *An. darlingi*, *An. nuneztovari sl*, *An. triannulatus sl*, *An. oswaldoi sl*) and two from the *Anopheles* subgenus (*An. intermedius*, *An. peryassui*) (Table 1). According to their availability, one to 22 specimens per species were included in the present study. A total of 111 *Anopheles* specimens were selected. The *COI* gene sequencing of all specimens was done to valid morphological identification. *COI* gene sequences were queried against GenBank (NCBI) and the Barcode of Life Data (BOLD) Systems. The query of *COI* gene sequences allowed to obtain reliable mosquito species identification for all samples with identity ranges of 98-99% against GenBank and 98.15-100% against BOLD databases (Table 1). Concordant mosquito species identification were obtained between the two molecular DBs. The *COI* gene sequencing corroborated morphological classifications, at the exception of one mosquito. It was morphologically classified as *An. peryassui* but molecular analysis revealed that it was identified as *An. intermedius*. The phylogenetic analysis was done with the *COI* gene sequences of the 15 mosquito specimens selected for MS reference creation (Additional file 1). Mosquitoes belonging to the same subgenus clustered together.

Reproducible and specific MS spectra from both mosquito body parts

MS profiles of high intensity (>2000 a.u.) were obtained for legs (Figure 2A) and thoraxes (Figure 2B) from each of the 111 mosquitoes submitted to MALDI-TOF MS. Visual reproducible MS spectra were obtained for specimens of the same species according to body part (Figure 2). To evaluate the reproducibility and specificity of MS spectra from legs and thoraxes according to species, cluster analyses were performed. Two specimens per species were used for MSP dendrogram creation, at the exception of *An. oswaldoi* for which only one specimen was available. The clustering of specimens

223 from the same species on the same branch and the absence of species intertwining underlined the
224 reproducibility and specificity of the protein profiles for each *Anopheles* species for legs (Figure 3A)
225 and thoraxes (Figure 3B). Interestingly, *Anopheles* species ordination from MSP dendrograms was
226 not similar between legs and thoraxes of paired species (Figure 3). Nevertheless, *Anopheles* species
227 from the same subgenus clustered on the same branch on both MSP dendrograms. To visualize
228 specificity of MS spectra according body part per *Anopheles* species, PCAs were performed. PCAs
229 revealed a clear separation of the dots corresponding to MS spectra from the legs and thoraxes,
230 confirming a specificity of MS profiles between these two body parts for the seven *Anopheles* species
231 tested (Additional file 2).
232 As correct specimen species classification relies mainly on the intensity of resulting MS spectra, we
233 assessed whether the most intense mass peaks from legs and thoraxes per mosquito species could be
234 enough to distinct these *Anopheles* species, at the exception of *An. oswaldoi* for which only one
235 specimen was available. The selection of the top-ten and top-five mass peak lists per species
236 conducted to a total of 41 and 27 MS peaks for legs and 39 and 24 for thoraxes, respectively
237 (Additional files S3 and S4). These MS peak lists were included in the genetic algorithm (GA) model
238 from ClinProTools 2.2 software. The combination of the presence/absence of these top-ten and top-
239 five mass peak lists per *Anopheles* species displayed, respectively, RC values of 99.6% and 97.0%
240 and CV values of 97.9% and 98.3% for MS spectra from legs. For MS spectra from thoraxes, RC
241 values of 99.4% and 97.5% and CV values of 99.8% and 99.6% were obtained for the top-ten and
242 top-five selected mass peak lists, respectively.
243
244 *MS reference spectra database creation and validation step*
245 MS spectra of legs and thoraxes from the 15 specimens used for MSP dendrogram analysis (Table
246 1), validated morphologically and molecularly, were added to our MS spectra database (DB)
247 including already legs and thoraxes reference MS spectra from eight distinct mosquito species [38].
248 The legs and thoraxes MS spectra from the 96 remaining specimens were queried against this

upgraded DB. Interestingly, 100% of the identification results were concordant between paired MS spectra from legs and thoraxes. Among them, the MS identification of 95 specimens corroborated morphological results. One specimen, classified as *An. peryassui* based on morphological criteria was identified as *An. intermedius* based on MS tool with LSVs of 2.40 and 2.49 for legs and thorax, respectively, confirming the results of *COI* gene sequencing. The LSVs ranged from 1.84 to 2.56 for legs and from 1.60 to 2.61 for thoraxes (Additional file S5). As a threshold LSV upper than 1.8 is require for reliable identification [26,43], correct classification could be considered for 100% (96/96) of MS spectra from legs and 96.9% (3/96) from thoraxes. However, if we considered the LSV results from paired-samples per specimen, 100% of the mosquitoes tested, succeeded to obtain a LSV upper than 1.8 for at least one body-part (Figure 4). Interestingly, an increase of the LSV cut-off at 2.0, which improves identification confidence, revealed that 95.8% (92/96) and 88.5% (85/96) of the specimens reached this threshold, based on their legs and thoraxes MS spectra, respectively. However, the rate of at least one body-part from paired-samples per specimen achieving this threshold (LSVs>2.0) remained at 100%.

Discussion

The correct identification of mosquito species is essential for adapted control management, allowing to index circulating species in a given region and, consequently, to estimate vector borne diseases transmission risks. Border regions such as French Guiana may be influenced by their neighboring countries. The legal and illegal flow of human and merchandises through this frontier could induce mosquito vectors migration and conducting to colonization of new areas. Solely, a rapid, accurate and low cost surveillance method of mosquitoes will succeed to improve control measures.

In French Guiana, *An. darlingi* is the main malaria vector transmitting several *Plasmodium* species, such as *P. falciparum* and *P. vivax* [46], this last one representing about 75% of human cases [47]. In addition to *An. darlingi*, other *Anopheles* species were reported to transmit *Plasmodium* pathogens in this area during their blood feeding [10,41]. Moreover these malaria vectors live in sympatry with other anopheline species non-malaria vectors [40]. Indeed, in an effective program aiming to prevent or to eliminate malaria transmission, an accurate identification of *Anopheles* species remains a key factor. The recent repetitive success of the use of MALDI-TOF MS profiling for arthropod identification, including mosquitoes, was applied in the present work to implement our home-made MS reference spectra DB with anopheline specimens from French Guiana and to palliate of the limitations of morphological and molecular analyses [25].

In the present study, we confirmed that legs and thoraxes MS spectra from paired-specimens of the same species were distinct at least for the seven species for which more than one specimen was available. The species-specificity demonstrated for each body-part, underlined that these two compartments could be used independently for mosquito identification. The independent MS submission of several body-part to MALDI-TOF MS for specimen identification, was recently reported for mosquitoes [38] but also for ticks [48,49]. The advantages to test two distinct body parts are the possibility to cross-validate the results and to improve rate of identification confidence and reliability. Effectively, the combination of the results obtained independently by the query against the MS reference DB conducted to concordant results for each body part tested. Here, based on

289 morphological criteria, a specimen was classified as *An. peryassui*. The MS submission of its legs
290 and thorax indicated a matching with *An. intermedius* MS reference spectra for both body parts with
291 high LSVs (>2.40), which was concordant with *COI* gene sequencing. Moreover, if the cut-off
292 threshold of LSV to consider identification as reliable was raised to 2.0, 100% of the specimens
293 succeeded to reach this cut-off with MS spectra from at least one body part. Finally, the concordance
294 of the MS identification results between the two body-parts more the elevate LSVs obtained are
295 complementary data which improve identification confidence. The previous study using two body
296 parts for mosquito identification was done on mosquitoes from distinct genera [38]. Here, the eight
297 mosquito species came all from the *Anopheles* genus including two subgenera, *Nyssorhynchus* and
298 *Anopheles*. The correctness identification of mosquito specimens by MALDI-TOF MS profiling for
299 these close-related *Anopheles* species comforts the accuracy of this innovative tool.
300 Among the mosquito species included in the MS reference spectra DB, 4 species are malaria vectors
301 (*An. darlingi*, *An. nuneztovari sl*, *An. intermedius, An. oswaldoi sl*), for the 4 remaining species their
302 malaria vector competence was not yet demonstrated (*An. aquasalis*, *An. braziliensis*, *An.*
303 *triannulatus sl*, *An. peryassui*) [5]. Interestingly, malaria vectors and non-vectors are present in each
304 of these two subgenera, underlining the importance to classify them correctly for disease prevention
305 and vector control of species living in sympatric.
306 The main risks of misidentification or to fail identification by MALDI-TOF MS profiling are
307 generally attributed either to the comprehensiveness of the species included in the MS reference
308 database, or more frequently either to the low intensity of MS spectra. The problem of incomplete
309 reference MS database could be easily solved by performing *COI* gene sequencing of un-matched
310 good quality MS spectra, using the remaining body part of the specimen (ie, head, abdomen or wings).
311 In case of new mosquito species, the addition of respective MS spectra, not yet included in the MS
312 reference DB, could be done. The application of this strategy could resolved step by step MS spectra
313 of good quality failing to find correspondence in the reference MS DB. Concerning the MS spectra
314 un-matched, attributed to the low intensity of protein profiles, this phenomenon is frequently

observed for legs MS spectra due to the low protein quantity contained in this compartment [26]. Moreover, as the legs are highly breakable, it is frequent that the loose of one to five legs occurred during specimen collection, transport or storing. This decrease of leg number reduces the success rate of specimen identification [43]. Moreover, for specimens which have lost all legs, their identification become not possible if the identification was based uniquely on this compartment by MS. The creation of reference MS spectra from two distinct mosquito body parts allows to succeed specimen identification by this rapid proteomic approach.

Moreover, in the present work, it was highlighted that the classification of *Anopheles* species could be correctly done using the most intense MS peaks. Effectively, the selection of the top-10 or also the top-5 of the MS peaks possessing the higher intensity appeared sufficiently discriminants to classify correctly these mosquito species. This correct classification is valid for both body-parts. These results underline that a correct identification remain possible with MS spectra of low intensity for which uniquely the most intense MS peaks could be detected.

The comparison of MSP dendrograms between legs and thoraxes revealed that, despite species from the same subgenus were clustered in the same branch, the ordination of the species inside these branches was not similar. This unreproducible classification objectifies that the MS profiles proximity were different between the *Anopheles* species for legs and thorax. These distinct ordination of species based on their MS profiles, should reduce the risk of misidentification by the submission of both body parts. Interestingly, none of these MSP dendrograms proposed a classification comparable with those obtained using *COI* gene sequences for phylogenetic tree construction. The MSP dendrograms are not adapted for phylogenetic analyses as previously reported [50,51].

Conclusion. Mosquito monitoring with fast, highly reproducible and reliable tools such as MALDI-TOF appears essential in today's globalization scenario. The specificity of MS protein profiles for mosquito legs and thoraxes confirmed that these two body parts are relevant for specimen identification. Moreover, as the most intense MS peaks were demonstrated to be sufficient for correct

341 classification, the sample which will generate MS spectra of low quality, could anyway identified.
342 The sharing of reference MS spectra is primordial to accelerate the dissemination of this innovative
343 tool for a routine use in mosquito identification contributing to adapt control of vectors.
344

List of abbreviations

sl: *sensu lato*; *ss: sensu stricto*; MALDI-TOF MS: Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry; PCR: Polymerase Chain Reaction; CCI: Composite Correlation Index; LSV: Log Score Value

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets of MS reference spectra added to the MS DB in the current study are freely available and downloadable from the additional file 6.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by the Délégation Générale pour l'Armement (DGA, MoSIS project, Grant no PDH-2-NRBC-2-B-2113).

Authors' contributions

Conceived and designed the experiments: LA, SB. Performed the experiments: LA, SB, CNG. Analyzed the data: LA, SB. Contributed reagents/materials/analysis tools: MMC, CNG, VPS, ID,

371 RG. Field collections: SB, VPS, ID, RG. Drafted the paper: LA, SB, MMC. Revised critically the
372 paper: all the authors.
373
374 **Acknowledgments**
375 We would like to acknowledge Samuel Vezenegho and Antoine Adde, from Unite d'Entomologie
376 Médicale, Institut Pasteur de la Guyane, Cayenne, French Guiana, for their involvement in sample
377 management. We also acknowledge Albin Fontaine (UPE, IRBA, Marseille) for his help in map
378 building.
379

References

1. Ardillon V, Carvalho L, Prince C, Abboud P, Djossou F. Bilans 2013 et 2014 de la situation du paludisme en Guyane. Bull Veille Sanit. 2015;1: 16–20.

2. Pommier de Santi V, Dia A, Adde A, Hyvert G, Galant J, Mazevet M, et al. Malaria in French Guiana Linked to Illegal Gold Mining. Emerg Infect Dis. 2016;22: 344–346. doi:10.3201/eid2202.151292

3. Pommier de Santi V, Girod R, Mura M, Dia A, Briolant S, Djossou F, et al. Epidemiological and entomological studies of a malaria outbreak among French armed forces deployed at illegal gold mining sites reveal new aspects of the disease's transmission in French Guiana. Malar J. 2016;15: 35. doi:10.1186/s12936-016-1088-x

4. Douine M, Sanna A, Hiwat H, Briolant S, Nacher M, Belleoud D, et al. Investigation of a possible malaria epidemic in an illegal gold mine in French Guiana: an original approach in the remote Amazonian forest. Malar J. 2019;18: 91. doi:10.1186/s12936-019-2721-2

5. Talaga S, Dejean A, Carinci R, Gaborit P, Dusfour I, Girod R. Updated Checklist of the Mosquitoes (Diptera: Culicidae) of French Guiana. J Med Entomol. 2015;52: 770–782. doi:10.1093/jme/tjv109

6. Girod R, Gaborit P, Carinci R, Issaly J, Fouque F. Anopheles darlingi bionomics and transmission of Plasmodium falciparum, Plasmodium vivax and Plasmodium malariae in Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana. Mem Inst Oswaldo Cruz. 2008;103: 702–710. doi:10.1590/s0074-02762008000700013

7. Girod R, Roux E, Berger F, Stefani A, Gaborit P, Carinci R, et al. Unravelling the relationships between Anopheles darlingi (Diptera: Culicidae) densities, environmental factors and malaria incidence: understanding the variable patterns of malarial transmission in French Guiana (South America). Ann Trop Med Parasitol. 2011;105: 107–122. doi:10.1179/136485911X12899838683322

8. Fouque F, Gaborit P, Carinci R, Issaly J, Girod R. Annual variations in the number of malaria cases related to two different patterns of Anopheles darlingi transmission potential in the Maroni area of French Guiana. Malar J. 2010;9: 80. doi:10.1186/1475-2875-9-80

9. Hiwat H, Issaly J, Gaborit P, Somai A, Samjhawan A, Sardjoe P, et al. Behavioral heterogeneity of Anopheles darlingi (Diptera: Culicidae) and malaria transmission dynamics along the Maroni River, Suriname, French Guiana. Trans R Soc Trop Med Hyg. 2010;104: 207–213. doi:10.1016/j.trstmh.2009.07.007

10. Dusfour I, Issaly J, Carinci R, Gaborit P, Girod R. Incrimination of Anopheles (Anopheles) intermedius Peryassú, An. (Nyssorhynchus) nuneztovari Gabaldón, An. (Nys.) oswaldoi Peryassú as natural vectors of Plasmodium falciparum in French Guiana. Mem Inst Oswaldo Cruz. 2012;107: 429–432. doi:10.1590/s0074-02762012000300021

11. Ruiz-Lopez F, Wilkerson RC, Ponsonby DJ, Herrera M, Sallum MAM, Velez ID, et al. Systematics of the oswaldoi complex (Anopheles, Nyssorhynchus) in South America. Parasit Vectors. 2013;6: 324. doi:10.1186/1756-3305-6-324

12. Conn JE, Wilkerson RC, Segura MNO, de Souza RTL, Schlichting CD, Wirtz RA, et al. Emergence of a new neotropical malaria vector facilitated by human migration and changes in land use. Am J Trop Med Hyg. 2002;66: 18–22. doi:10.4269/ajtmh.2002.66.18

13. Montoya-Lerma J, Solarte YA, Giraldo-Calderón GI, Quiñones ML, Ruiz-López F, Wilkerson RC, et al. Malaria vector species in Colombia: a review. Mem Inst Oswaldo Cruz. 2011;106 Suppl 1: 223–238. doi:10.1590/s0074-02762011000900028

14. Beebe NW, Cooper RD. Systematics of malaria vectors with particular reference to the Anopheles punctulatus group. Int J Parasitol. 2000;30: 1–17. doi:10.1016/s0020-7519(99)00171-x

15. Munstermann LE, Conn JE. Systematics of mosquito disease vectors (Diptera, Culicidae): impact of molecular biology and cladistic analysis. Annu Rev Entomol. 1997;42: 351–369. doi:10.1146/annurev.ento.42.1.351

16. Van Rensburg AJ, Hunt RH, Koekemoer LL, Coetzee M, Shiff CJ, Minjas J. The polymerase chain reaction method as a tool for identifying members of the Anopheles gambiae complex (Diptera:Culicidae) in northeastern Tanzania. J Am Mosq Control Assoc. 1996;12: 271–274.

17. Figueiredo MAP, Di Santi SM, Manrique WG, Gonçalves LR, André MR, Machado RZ. Molecular identification of Plasmodium spp. and blood meal sources of anophelines in environmental reserves on São Luís Island, state of Maranhão, Brazil. Parasit Vectors. 2017;10: 203. doi:10.1186/s13071-017-2133-5

18. Marrelli MT, Floeter-Winter LM, Malafronte RS, Tadei WP, Lourenço-de-Oliveira R, Flores-Mendoza C, et al. Amazonian malaria vector anopheline relationships interpreted from ITS2 rDNA sequences. Med Vet Entomol. 2005;19: 208–218. doi:10.1111/j.0269-283X.2005.00558.x

19. Nebbak A, Koumare S, Willcox AC, Berenger J-M, Raoult D, Almeras L, et al. Field application of MALDI-TOF MS on mosquito larvae identification. Parasitology. 2018;145: 677–687. doi:10.1017/S0031182017001354

20. Ratnasingham S, Hebert PDN. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes. 2007;7: 355–364. doi:10.1111/j.1471-8286.2007.01678.x

21. Versteirt V, Nagy ZT, Roelants P, Denis L, Breman FC, Damiens D, et al. Identification of Belgian mosquito species (Diptera: Culicidae) by DNA barcoding. Mol Ecol Resour. 2015;15: 449–457. doi:10.1111/1755-0998.12318

22. Norris LC, Norris DE. Phylogeny of anopheline (Diptera: Culicidae) species in southern Africa, based on nuclear and mitochondrial genes. J Vector Ecol J Soc Vector Ecol. 2015;40: 16–27. doi:10.1111/jvec.12128

23. Laurito M, Oliveira TMP de, Almirón WR, Sallum MAM. COI barcode versus morphological identification of Culex (Culex) (Diptera: Culicidae) species: a case study using samples from Argentina and Brazil. Mem Inst Oswaldo Cruz. 2013;108 Suppl 1: 110–122. doi:10.1590/0074-0276130457

455 24. Wang G, Li C, Guo X, Xing D, Dong Y, Wang Z, et al. Identifying the main mosquito species
456 in China based on DNA barcoding. PloS One. 2012;7: e47051. doi:10.1371/journal.pone.0047051
457 25. Yssouf A, Almeras L, Raoult D, Parola P. Emerging tools for identification of arthropod
458 vectors. Future Microbiol. 2016;11: 549–566. doi:10.2217/fmb.16.5
459 26. Yssouf A, Parola P, Lindström A, Lilja T, L'Ambert G, Bondesson U, et al. Identification of
460 European mosquito species by MALDI-TOF MS. Parasitol Res. 2014;113: 2375–2378.
461 doi:10.1007/s00436-014-3876-y
462 27. Dieme C, Yssouf A, Vega-Rúa A, Berenger J-M, Failloux A-B, Raoult D, et al. Accurate
463 identification of Culicidae at aquatic developmental stages by MALDI-TOF MS profiling. Parasit
464 Vectors. 2014;7: 544. doi:10.1186/s13071-014-0544-0
465 28. Schaffner F, Kaufmann C, Pflüger V, Mathis A. Rapid protein profiling facilitates
466 surveillance of invasive mosquito species. Parasit Vectors. 2014;7: 142. doi:10.1186/1756-3305-7-
467 142
468 29. Mewara A, Sharma M, Kaura T, Zaman K, Yadav R, Sehgal R. Rapid identification of
469 medically important mosquitoes by matrix-assisted laser desorption/ionization time-of-flight mass
470 spectrometry. Parasit Vectors. 2018;11: 281. doi:10.1186/s13071-018-2854-0
471 30. Müller P, Pflüger V, Wittwer M, Ziegler D, Chandre F, Simard F, et al. Identification of
472 cryptic Anopheles mosquito species by molecular protein profiling. PloS One. 2013;8: e57486.
473 doi:10.1371/journal.pone.0057486
474 31. Yssouf A, Socolovschi C, Flaudrops C, Ndiath MO, Sougoufara S, Dehecq J-S, et al. Matrix-
475 assisted laser desorption ionization--time of flight mass spectrometry: an emerging tool for the rapid
476 identification of mosquito vectors. PloS One. 2013;8: e72380. doi:10.1371/journal.pone.0072380
477 32. Raharimalala FN, Andrianinarivomanana TM, Rakotondrasoa A, Collard JM, Boyer S.
478 Usefulness and accuracy of MALDI-TOF mass spectrometry as a supplementary tool to identify
479 mosquito vector species and to invest in development of international database. Med Vet Entomol.
480 2017;31: 289–298. doi:10.1111/mve.12230

481 33. Tandina F, Almeras L, Koné AK, Doumbo OK, Raoult D, Parola P. Use of MALDI-TOF MS
482 and culturomics to identify mosquitoes and their midgut microbiota. Parasit Vectors. 2016;9: 495.
483 doi:10.1186/s13071-016-1776-y
484 34. Nebbak A, Almeras L. Identification of Aedes mosquitoes by MALDI-TOF MS biotyping
485 using protein signatures from larval and pupal exuviae. Parasit Vectors. 2020;13: 161.
486 doi:10.1186/s13071-020-04029-x
487 35. Nebbak A, El Hamzaoui B, Berenger J-M, Bitam I, Raoult D, Almeras L, et al. Comparative
488 analysis of storage conditions and homogenization methods for tick and flea species for identification
489 by MALDI-TOF MS. Med Vet Entomol. 2017;31: 438–448. doi:10.1111/mve.12250
490 36. Nebbak A, Willcox AC, Bitam I, Raoult D, Parola P, Almeras L. Standardization of sample
491 homogenization for mosquito identification using an innovative proteomic tool based on protein
492 profiling. Proteomics. 2016;16: 3148–3160. doi:10.1002/pmic.201600287
493 37. Vega-Rúa A, Pagès N, Fontaine A, Nuccio C, Hery L, Goindin D, et al. Improvement of
494 mosquito identification by MALDI-TOF MS biotyping using protein signatures from two body parts.
495 Parasit Vectors. 2018;11: 574. doi:10.1186/s13071-018-3157-1
496 38. Vega-Rúa A, Pagès N, Fontaine A, Nuccio C, Hery L, Goindin D, et al. Improvement of
497 mosquito identification by MALDI-TOF MS biotyping using protein signatures from two body parts.
498 Parasit Vectors. 2018;11: 574. doi:10.1186/s13071-018-3157-1
499 39. Adde A, Roux E, Mangeas M, Dessay N, Nacher M, Dusfour I, et al. Dynamical Mapping of
500 Anopheles darlingi Densities in a Residual Malaria Transmission Area of French Guiana by Using
501 Remote Sensing and Meteorological Data. PloS One. 2016;11: e0164685.
502 doi:10.1371/journal.pone.0164685
503 40. Adde A, Dusfour I, Roux E, Girod R, Briolant S. Anopheles fauna of coastal Cayenne, French
504 Guiana: modelling and mapping of species presence using remotely sensed land cover data. Mem Inst
505 Oswaldo Cruz. 2016;111: 750–756. doi:10.1590/0074-02760160272

506 41. Vezenegho SB, Adde A, Pommier de Santi V, Issaly J, Carinci R, Gaborit P, et al. High
507 malaria transmission in a forested malaria focus in French Guiana: How can exophagic Anopheles
508 darlingi thwart vector control and prevention measures? Mem Inst Oswaldo Cruz. 2016;111: 561–
509 569. doi:10.1590/0074-02760160150
510 42. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of
511 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol
512 Biotechnol. 1994;3: 294–299.
513 43. Nebbak A, Willcox AC, Bitam I, Raoult D, Parola P, Almeras L. Standardization of sample
514 homogenization for mosquito identification using an innovative proteomic tool based on protein
515 profiling. Proteomics. 2016;16: 3148–3160. doi:10.1002/pmic.201600287
516 44. Lafri I, Almeras L, Bitam I, Caputo A, Yssouf A, Forestier C-L, et al. Identification of
517 Algerian Field-Caught Phlebotomine Sand Fly Vectors by MALDI-TOF MS. PLoS Negl Trop Dis.
518 2016;10: e0004351. doi:10.1371/journal.pntd.0004351
519 45. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region
520 of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10: 512–526.
521 doi:10.1093/oxfordjournals.molbev.a040023
522 46. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A
523 global map of dominant malaria vectors. Parasit Vectors. 2012;5: 69. doi:10.1186/1756-3305-5-69
524 47. Musset L, Pelleau S, Girod R, Ardillon V, Carvalho L, Dusfour I, et al. Malaria on the Guiana
525 Shield: a review of the situation in French Guiana. Mem Inst Oswaldo Cruz. 2014;109: 525–533.
526 doi:10.1590/0074-0276140031
527 48. Boyer PH, Boulanger N, Nebbak A, Collin E, Jaulhac B, Almeras L. Assessment of MALDI-
528 TOF MS biotyping for Borrelia burgdorferi sl detection in Ixodes ricinus. PloS One. 2017;12:
529 e0185430. doi:10.1371/journal.pone.0185430

530 49. Boyer PH, Almeras L, Plantard O, Grillon A, Talagrand-Reboul É, McCoy K, et al.
531 Identification of closely related Ixodes species by protein profiling with MALDI-TOF mass
532 spectrometry. PloS One. 2019;14: e0223735. doi:10.1371/journal.pone.0223735
533 50. Kumsa B, Laroche M, Almeras L, Mediannikov O, Raoult D, Parola P. Morphological,
534 molecular and MALDI-TOF mass spectrometry identification of ixodid tick species collected in
535 Oromia, Ethiopia. Parasitol Res. 2016;115: 4199–4210. doi:10.1007/s00436-016-5197-9
536 51. Karger A, Kampen H, Bettin B, Dautel H, Ziller M, Hoffmann B, et al. Species determination
537 and characterization of developmental stages of ticks by whole-animal matrix-assisted laser
538 desorption/ionization mass spectrometry. Ticks Tick-Borne Dis. 2012;3: 78–89.
539 doi:10.1016/j.ttbdis.2011.11.002
540
541

542 **Figure legends**

543 **Figure 1. Map of mosquito collection sites in French Guiana.** The different sampling sites are
544 indicated by circles.

545

546 **Figure 2. Comparison of MALDI-TOF MS spectra from legs (A) and thoraxes (B) of *Anopheles***
547 **mosquitoes.** Representative MS spectra of *An. peryassui* (a, b), *An. intermedius* (c, d), *An. oswaldoi*
548 (e, f), *An. aquasalis* (g, h), *An. braziliensis* (i, j), *An. darlingi* (k, l), *An. nuneztovari* (m, n), and *An.*
549 *triannulatus* (o, p) are shown. MS spectra from two distinct specimens per species were selected,
550 excepted for *An. oswaldoi*. As only one specimen was available for this species, MS spectra from
551 biological replicates were presented. a.u., arbitrary units; m/z, mass-to-charge ratio.

552

553 **Figure 3. MSP dendrogram of MALDI-TOF MS spectra from legs (A) and thoraxes (B) of**
554 ***Anopheles* mosquitoes.** Two specimens per species were used to construct the dendrogram, at the
555 exception of *An. oswaldoi*, for which only one specimen was available. The dendrogram was created
556 using Biotyper v3.0 software and distance units correspond to the relative similarity of MS spectra.
557 The *Anopheles* and *Nyssorhynchus* subgenus were indicated at the right part.

558

559 **Figure 4. Comparison of paired body parts LSVs from MS spectra of *Anopheles* species.** Dashed
560 lines represent the threshold values (black and grey for LSV threshold of 1.8 and 2.0, respectively),
561 for relevant identification. LSV, log score value.

562 **Additional files:**

563 **Additional file 1. Unrooted Maximum-Likelihood trees based on the sequences of the COI gene**
564 **of the 15 specimens included in the MS database.**

565

566 **Additional file 2. Principal Component Analysis (PCA) from MS spectra of legs and thoraxes**
567 **from *Anopheles* mosquitoes.** PCA dimensional image from MS spectra of legs (red dots) and
568 thoraxes (green dots) from *An. intermedius* (A), *An. aquasalis* (B), *An. braziliensis* (C), *An. darlingi*
569 (D), *An. nuneztovari* (E), *An. triannulatus* (F) and *An. peryassui* (G). Respectively, 10, 22, 18, 22,
570 20, 20 and 3 specimens per species were included. Quadruplicate of each sample per body part were
571 presented.

572

573 **Additional file 3. Top-five and -ten mass peak lists per mosquito species using legs as biologic**
574 **material.**

575

576 **Additional file 4. Top-five and -ten mass peak lists per mosquito species using thoraxes as**
577 **biologic material.**

578

579 **Additional file 5. LSVs obtained following homemade MS reference database query with MS**
580 **spectra of legs (A) and thoraxes (B) from *Anopheles* mosquitoes.** Horizontal dashed lines represent
581 the threshold value for reliable identification (black and grey for LSV threshold of 1.8 and 2.0,
582 respectively). LSVs, log score values; a.u., arbitrary units.

583

584 **Additional file 6. Raw MS spectra from legs and thoraxes of mosquitoes added to the MS**
585 **reference database.** MS spectra were obtained using Microflex LT MALDI-TOF Mass Spectrometer
586 (Bruker Daltonics).

587

588 **Table 1.** Overview of *Anopheles* mosquito origins and subgroup identification by COI molecular typing

Morphological identification	Catching site (Latitude/Longitude)	Catching period	Number of specimens (Ref. MS DB§)	Species identified by NCBi (Accession Number)	COI gene sequence coverage (%) / identity (%)	Species identified by BOLD	COI gene sequence similarity (%)
An. peryassui	Cacao (4.57/52.47)	(June-2015)	3 (2)	NC_037790.1	99/99	KF698875	97.84
An. intermedius	Cacao (4.57/52.47)	(June-2015)	10 (2)	MF381700.1 / NC_037789.1	99/98	Early release	98.62-99.39
An. oswaldoi	Eau Claire (3.60/53.5)	(June-2014)	1 (1)	MG241906.1	93/99	51917708	97.85
An. aquasalis	Cayenne (4.89/52.30)	(Sept./Oct.-2014)	22 (2)	KC354822.1	99/98	KC354821	98.15-98.29
An. braziliensis	Cacao (4.57/52.47)	(June-2015)	18 (2)	NC_037791.1 / MF381732.1	99/99	Private / DQ913839 / DQ913846 / DQ913825	98.91-99.73
An. darlingi	Blondin (3.87/51.81)	(Sept.-2015)	22 (2)	MF381596.1 / MF381713.1	99/99	JF923694 / private	99.85-100
An. nuneztovari	Dorlin (3.75/53.55)	(March-2013)	20 (2)	NC_037810.1 / MF381656.1	99/99	KU865547 / KC167737	99.83-100
An. triannulatus	Saint Georges (3.89/51.81)	(May/June-2014)	20 (2)	MF381730.1 / JX205112.1	99/99	Early-Release / KC167680	99.54-99.83
Total			**111 (15)**				

589 §Ref. MS DB: Number of specimens used to create the reference MS database. *Mosquito species for whose COI gene sequences are not available
590 (N.A.) in the database (27th November 2018). #No reliable ID: Identity with top match < 97 %. BOLD: Barcode of Life Data Systems; COI: cytochrome
591 oxidase one.
592

Figure-1

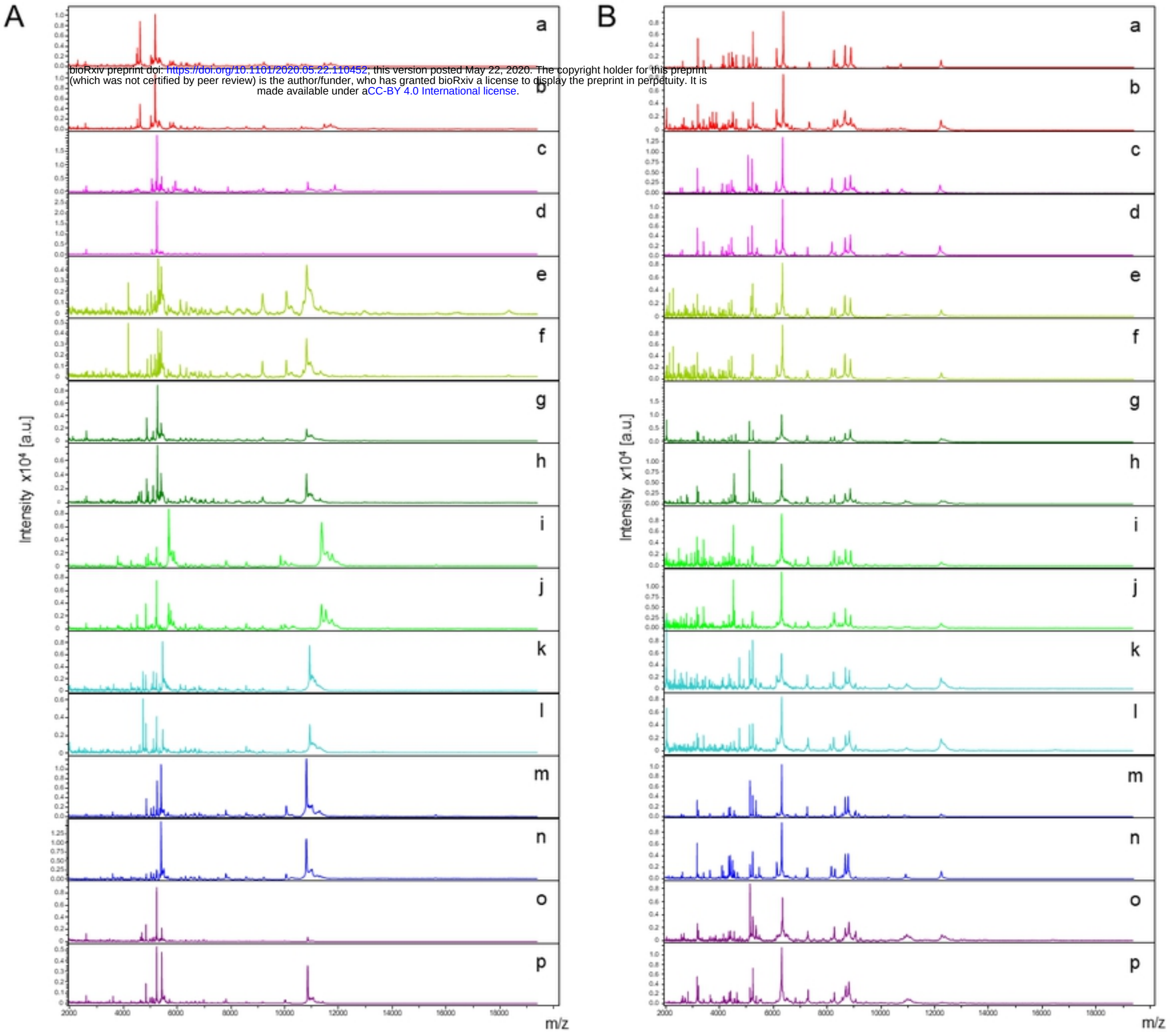
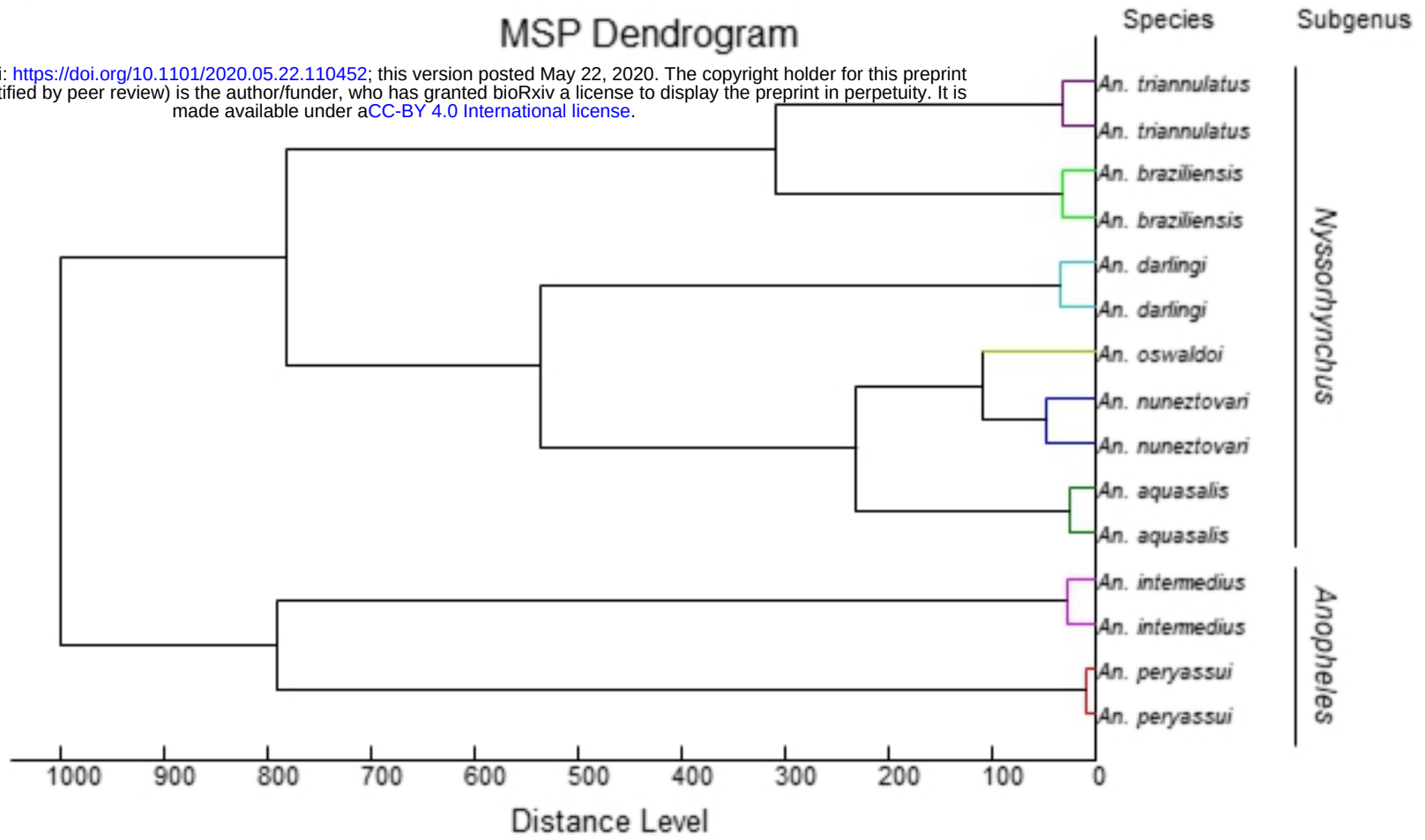


Figure-2

A



B

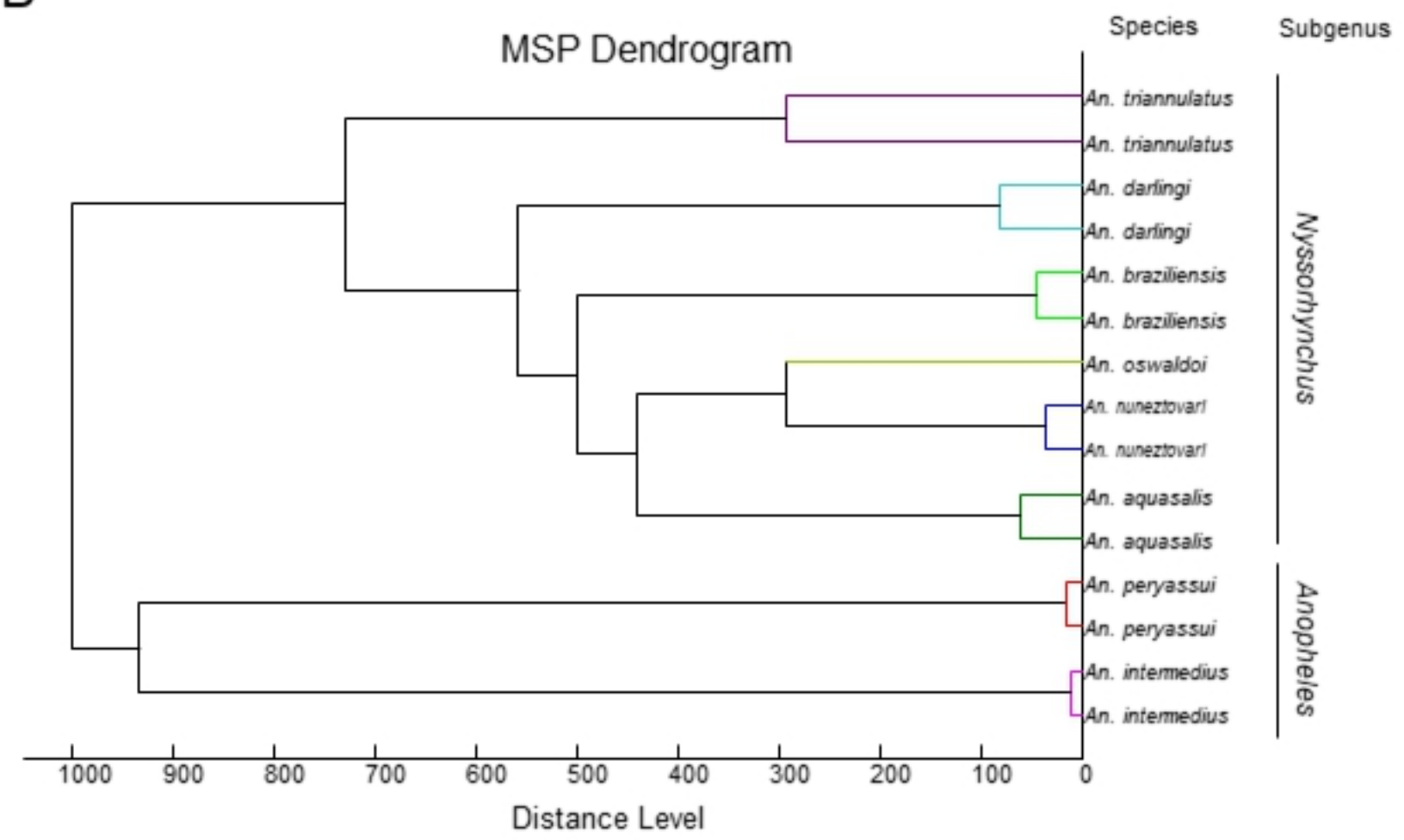


Figure-3

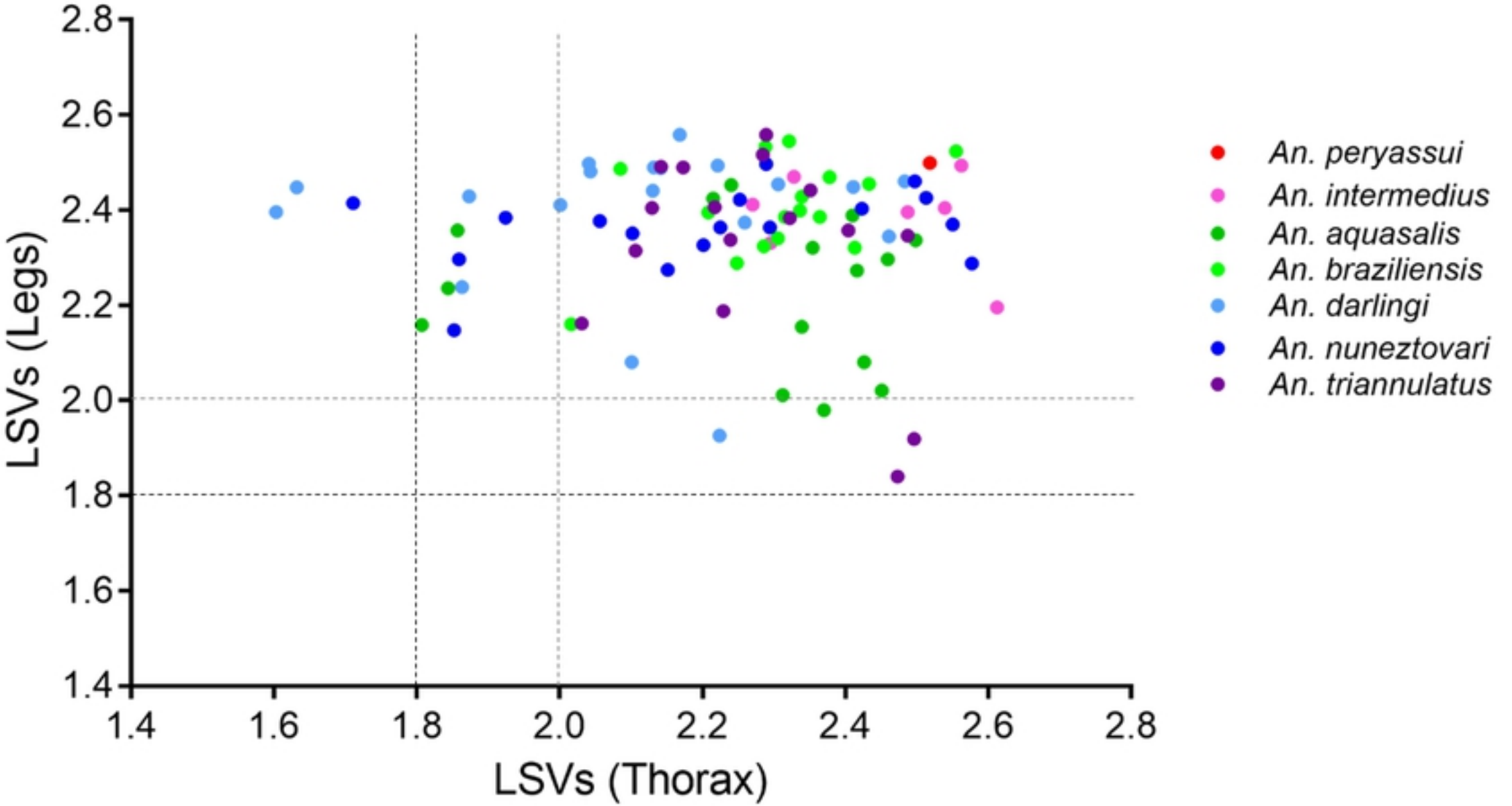

2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
LSVs (Legs)
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
LSVs (Thorax)
An. peryassui
An. intermedius
An. aquasalis
An. braziliensis
An. darlingi
An. nuneztovari
An. triannulatus

Figure-4