

Opening the Pandora Box: DNA-barcoding evidence limitations of morphology to identify Spanish mosquitoes

DELGADO-SERRA, SOFÍA¹; VIADER, MIRIAM¹; RUIZ-ARRONDO, IGNACIO²; MIRANDA, MIGUEL ÁNGEL¹; BARCELÓ, CARLOS¹; BUENO-MARI³, RUBÉN; HERNÁNDEZ-TRIANA, LUIS M.⁴; MIQUEL, MARGA¹ AND PAREDES-ESQUIVEL, CLAUDIA^{1*}

1. Laboratory of Zoology. University of the Balearic Islands. Palma de Mallorca. Spain
2. Center for Rickettsiosis and Arthropod-Borne Diseases, CIBIR, Logroño, La Rioja, Spain.
3. Spain.Research and Development (R+D) Department. Laboratorios Lokímica. Valencia. Spain.
4. Wildlife Zoonoses and Vector-borne Diseases Research Group, Virology Department, Animal and Plant Health Agency, Addlestone, United Kingdom.

* Corresponding author

Abstract

Cryptic speciation is frequent in the medically important mosquitoes. While most findings have been reported in tropical regions, it is an unexplored topic in countries where mosquito-borne diseases are not endemic, like Spain. The occurrence of recent outbreaks in Europe has increased the awareness on the native and invasive mosquito fauna present in the continent. Therefore, the central question of this study is whether the typological approach is sufficient to identify Spanish mosquitoes. To address this problem, we confronted the results of the morphological identification of 62 adult specimens collected from four different regions of Spain (La Rioja, Navarra, Castellón and the Island of Majorca) with the results obtained through DNA-barcoding. We conducted a comprehensive phylogenetic analysis of the *COI* gene region and compared this with the results of four species delimitation algorithms (ABGD initial partition, ABGD P=0.46%, bPTP and TCS). We report strong evidence for cryptic speciation in *Anopheles algeriensis* and *Aedes vexans* and reproductive isolation of the rock pool mosquito *Aedes mariae*. In addition, we report that the character present in the wings is not efficient to distinguish species *Culiseta annulata* from *Culiseta subochrea*, which distribution in the country may be different than previously described.

1. Introduction

Mosquitoes are responsible for the death of millions of people each year due to its role as vectors of some of the most devastating diseases in the world such as malaria, dengue and yellow fever. Although the majority of the cases are reported from tropical countries, several outbreaks have also taken place in the Northern hemisphere showing that the landscape of mosquito-borne diseases is changing in the world. In Italy, the rapidly-spreading invasive species *Aedes albopictus* have been implicated in outbreak of Chikungunya fever that affected 205 people in 2007 (Rezza *et al.*, 2007) and it was also identified as the vector of two Dengue fever cases in France (Gould *et al.*, 2010). Although it is unquestionable that invasive species represent a serious Public Health threat, the role of native mosquitoes should not be underestimated. Recent outbreaks of West Nile Virus (WNV) in Romania (Campbell *et al.*, 2001) and in the United States (Nash *et al.*, 2001), were attributed to local populations of *Culex* mosquitoes. Mosquito species identification has been traditionally carried out using dichotomous keys that describe the morphological characters of a particular life stage. Although this technique has proved valuable and is still widely used to distinguish many mosquito species, it has important limitations like the need of taxonomic experts to perform accurate identifications and the loss of morphological characters during collection and preservation of specimens. Furthermore, morphology-based taxonomy cannot differentiate indistinguishable members of species complexes, which frequently differ on their ecology, host preferences, behavioral patterns, therefore in their role as disease vectors, greatly impairing the design of effective control strategies (Paredes-Esquivel, Del Rio, *et al.*, 2009).

The field of mosquito taxonomy has being renewed with the development of new tools for accurate identification (Cywinska *et al.*, 2006). Molecular-based techniques are now widely used not only to identify sibling species but also to solve phylogenetic relationships (Bourke *et al.*, 2013). The *Cytochrome Oxidase I* gene region of the mitochondrial genome is the gold standard barcoding identification of species (Hebert *et al.*, 2003) and has proved valuable to distinguish many mosquito species (Cywinska *et al.*, 2006; Kumar *et al.*, 2007; Wang *et al.*, 2012; Ashfaq *et al.*, 2014). This approach is particularly important to establish species delimitations in unexplored groups (Puillandre *et al.*, 2012). Although, there have been reports that *COI* barcoding has failed in differencing some species of mosquitoes of *Anopheles* and *Culex* genus (Cywinska *et al.*, 2006; Kumar *et al.*, 2007; Wang *et al.*, 2012; Bourke *et al.*, 2013; Magdalena Laurito *et al.*, 2013), there is no doubt that DNA barcoding erects as a valuable alternative for mosquito identification in most areas of the world. However, the use DNA barcoding as a surveillance tool would require a database containing mosquito barcodes of correctly identified mosquitoes.

From the 3,552 formally recognized mosquito species in the world (Harbach, 2016), 100 have been recorded in Europe and this number might be even higher since six invasive *Aedes* species have recently colonized the continent (Versteirt *et al.*, 2015). Spain reports a similar

increasing pattern, with 59 species reported at the beginning of this decade and 64 described in 2012 according to the latest checklist (R. Bueno Marí *et al.*, 2012), including competent vectors of human diseases. To the best of our knowledge, this would be the first comprehensive faunistic study that includes molecular-based techniques in Spain where identifications have mainly relied on the use of dichotomous keys. We are hoping to contribute to the knowledge of mosquito fauna and demonstrate that molecular based studies can uncover hidden diversity among well-known mosquito species.

2. Material and methods

2.1 Mosquito collection and morphological identification

The mosquito specimens were collected from 17 sites located in four different regions of Spain, including the Island of Majorca in the Balearic archipelago, Castellón, La Rioja and Navarra (Figure 1). The majority of the specimens (42) were collected from the island of Majorca, the largest at the Balearic archipelago, located in the Western region of the Mediterranean Basin. Additionally, for comparison purposes, three specimens from *Ae. mariaae* were obtained from Peñiscola (Castellón, Spain), a coastal region with high abundance of salt marshes and rocky shores. Finally, 17 specimens from *Cs. annulata*, *Cs. subochrea* and *An. algeriensis* were collected from the wetlands of La Grajera and Hervías and in the River Iregua in La Rioja and Navarra, respectively. These sampling sites are located in the Ebro Valley, in the Northern region of Spain; with a transitional climate between the oceanic one in the North and a the semiarid found in the central Ebro Valley.

Adult mosquitoes were collected from outdoor using BG and CDC light traps containing CO₂. The samplings were carried out weekly and each trap was set at sunset. On the following morning, collected specimens were transported *in vivo* to the laboratory and then frozen at -20 °C for molecular studies. Larvae were also collected from breeding sites, transported to the laboratory and kept alive at 25°C until adults emerged. Female mosquitoes were identified to species level using keys by Becker et al. (2010). Scientific names of mosquitoes were verified using the Mosquito taxonomy inventory (Harbach, 2014).

2.2. DNA extraction, PCR amplification and sequencing

Genomic DNA was extracted from individual mosquitoes, using DNeasy Blood & Tissue Kit (QIAGEN, GmbH, Hilden, Germany). To increase DNA yield we used 100 µl elution buffer instead of 200 µl in the final step. DNA concentration was measured with a NanoDrop 1000 spectrophotometer Thermo Scientific (Saveen Werner ApS, Denmark) and then DNA extractions were kept at -20 °C (working solutions) and -80 °C (Stock solutions). PCR was performed using the thermal cycler GeneAmp PCR System 2400 (Perkin Elmer) using primers C1-J-1718 forward 5' - GGAGGATTTGGAAATTGATTAGTGCC- 3' and C1-N-2191 reverse 5'- CCCGGTAAAATTTAAAATATAAACTTC- 3'. Folmer primers (LCO1490 5'- GGTCACAAATCATAAAGATATTGG-3' and HCO2198 5'- TAAACTTCAGGGTGACCAAAAAATCA -3') were also tested but failed to amplify all specimens. The 50 µl PCR reaction contained of 2 µl of DNA template, QIAGEN 1X reaction buffer (Tris-Cl, KCl) and (NH₄)₂SO₄ 1.5–3.0 mM of MgCl₂, 0.2 mM of dNTPs, 1 unit of QIAGEN DNA polymerase and 125 nM of each primer. The PCR reaction conditions were as follow: 95 °C for 5 min followed by 92 °C for 2 min and 15 s; 30 cycles of 92 °C x 15 s, annealing temperature of 50 °C x 15 s and an extension temperature of 72 °C x 30 s, followed by a final extension at 72 °C for 1 min. PCR products were stained with Pronasafe nucleic acid staining solution (Conda Laboratories) and observed in a 2% agarose - TBE buffer gel. Samples were purified using the QIAquick PCR Purification Kit (QIAGEN), according to manufacturer's specifications, but purified DNA was eluted in 30 µl MilliQ water instead of 50 µl elution buffer and the template was sequenced in both directions in a ABI3730XL automatic sequencer (Macrogen Inc., Spain) with the same primers used in first amplification.

2.3. Sequence analysis

Sequences obtained were aligned using Clustal W and edited manually by using software BioEdit v7.0.8.0 (Hall, 1999). When point substitutions were detected, the reliability of the sequence was confirmed by looking at nucleotide peaks. All sequences were translated using the *Drosophila* genetic code in DnaSP version 5.10.1 (Rozas & Rozas, 1999) to ensure these were not pseudogenes using the mDNA. This program was also used to calculate haplotype diversity within species. The transition/transversion ratio and A+T content was calculated with Mega 6. For comparative purposes, we used BLASTn and sequences were retrieved from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

2.4. Phylogenetic analysis

The best evolutionary model to construct phylogenetic trees was selected according to their BIC score (Bayesian Information Criterion) obtained in Mega6 (Tamura *et al.*, 2007). With this program we also constructed the Maximum Likelihood (ML) tree and the robustness of the clustering was determined by the bootstrap analysis with 500 replicates. The Bayesian inferences were constructed with MRBAYES 3.2.6 (Ronquist & Huelsenbeck, 2003); using four Metropolis coupled Markov chain Monte Carlo (MCMC) chains were run for 500,000 generations (sampling every 10 generations) to allow adequate time for convergence. The analysis was finished when the standard deviation of split frequencies was less than 0.01. To assure we obtained the correct topology, analyses were run twice. All analyzes were performed with random starting trees. Branch support is indicated by the posterior probability values. Consensus trees were visualized with FigTree v1.4.3.

2.5. Species boundaries

We used the results obtained from the phylogenetic analysis as a first approach to distinguish species. Species were defined following the Phylogenetic species concept, that establishes that these are “irreducible cluster of organisms, within which there is a parental pattern of ancestry and descent, and which is diagnosably distinct from other such clusters” (Cracraft, 1987). The second approach we used to delimit species within our dataset was the Automatic Barcode Gap Discovery algorithm (ABGD), available at <http://www.wabi.snv.jussieu.fr/public/abgd/>. This method is independent of tree typology and split dataset into candidate species, based on the existence of barcode gap (interspecific divergence higher than intraspecific ones) and a prior intraspecific divergence (p) (Puillandre *et al.*, 2012). ABGD was carried out applying the K2P model, using default parameters, with the exception of the relative gap width, which was defined as 1.4. Results of a recursive and initial partition analysis were compared. The third approach to identify species was the Bayesian approach to Poisson Tree Processes model (bPTP) (Zhang *et al.*, 2013), which is based on the number of mutations in the branches of a phylogenetic tree. We run 100 000 trees, a thinning of 100 using a burn-in value of 0.1. A statistical parsimony networks were constructed using TCS (Clement *et al.*, 2000), which collapses DNA sequences into haplotypes and determines the frequency of those haplotypes within the dataset. Gaps were treated as 5th state or missing data. Size of circles represents the number of sequences within a haplotype. Although this method has been mainly used to construct genealogies, it has been also used in diversity studies in other taxa. Molecularly Defined Operational Taxonomic Units (MOTUs) were defined according to the results obtained from the different species delimitation approaches. If these methods resulted in a different species composition, we defined MOTUs as those defined by the majority of such methods. We constructed a pairwise distance matrix using the Kimura 2-parameter model (K-2P) of substitutions using the aligned *COI* sequences in Mega 6 (Tamura *et al.*, 2007). We

tested in the 10x threshold by Hebert et al (2004) to confirm morphological identification. This rule establishes that a new species is suspected if the average interspecific distance is at least 10 times higher than the average intraspecific divergence of the putative sister species.

3. Results

3.1. Sequence analysis

After alignment of the dataset, the matrix contained 406 bp long sequences. Neither insertion nor deletion events were observed in the dataset. All sequences were A+T rich, with an average content of 67.3% ranging 62.1% in *Coquillettidia richiardii* up to 69.2% in *Ae. mariae*, which is a similar value observed in other species of Culicidae (Paredes-Esquivel, Donnelly, *et al.*, 2009). From the 406 bp fragment, we found 147 parsimony informative sites and 10 singleton sites. Most substitutions were found at the third codon position, with a ratio of transitions/transversions of 1.13. Thirty-eight haplotypes were found in mosquitoes analyzed and their sequences were deposited in GenBank under accession numbers MH348258-MH348270 and MH370023-MH370049 (table 1). Haplotype diversity varied from 0 in *Ae. albopictus* up to 1 in *Ae. mariae* C, *Ur. unguiculata*, *Cq. richiardii*, *Ae. vexans* B and *An. algeriensis* B (Table 2). No evidence of pseudogenes was found as stop codons were not observed after translation, using the mitochondrial DNA genetic code for *Drosophila* sp.

Twenty-five sequences were retrieved from GenBank for comparative purposes (Table 2). Sequences belonged to the same species included in this study, with the exception of the *Mariae* Complex members *Ae. phoeniciae* and *Ae. zammitii* which were included to provide additional information on the *Ae. mariae* specimens from Majorca. A Kimura-2-parameter distance matrix was built for the entire dataset. The intraspecific divergence in the 406 bp fragment ranged 0.00-2.79 %, with an average 0.86%. On the other hand, distance among co-generic species ranged 2.13-15.74%. There was an overlap between the intraspecific and interspecific divergence (Table 3).

3.2. Phylogenetic analysis

Under the Bayesian Information criterion, the best evolutionary model obtained for our dataset was GTR+R. We used this model to the Bayesian analysis and substitution model by Tamura and Nei (Tamura & Nei, 1993) to construct the Maximum Likelihood phylogenetic tree. Both trees show clearly defined highly supported clusters (>87% and >96% in the Maximum

likelihood and Bayesian trees, respectively) in all MOTUs identified (Figures 2 and 3). In both trees, members of the Mariae complex clustered in the same clade with *Oc. caspius*. *Aedes mariae* appears polyphyletic, with the Majorcan clade being more closely related to *Ochlerotatus caspius* than to *Ae. mariae* specimens from Continental Europe. Nevertheless, the relationship between these clusters was poorly supported by the bootstrap value. Specimens identified as *Ae. vexans* were separated in two highly supported clusters (>99%), herein *Ae. vexans* A, which included specimens from Majorca and sequences retrieved from GenBank from Sweden and Hungary. *Ae. vexans* B was formed by sequences from Sweden. These two groups have a 6% sequence divergence. Two monophyletic clades with high bootstrap support were also found in *An. algeriensis*. Those specimens from Majorca and La Rioja (*An. algeriensis* A) were clearly separated (bootstrap > 98%) from those sequences retrieved from Sweden and Romania (*An. algeriensis* B). These groups were separated by a 7% divergence (K2P). In addition all specimens from Majorca initially identified as *Cs. annulata* by field staff were placed in the *Cs. subochrea* clade (Figure 3). This resulted into a re-examination of the morphology of these specimens (see below). The remaining species *Aedes albopictus*, *Culiseta longiareolata*, *Culex theileri*, *Ochlerotatus detritus* and *Culex pipiens* formed highly supported monophyletic clades, although the last two showed some level of differentiation (Figure 2). Even when a single specimen of species *Uranotaenia unguiculata* and *Coquillettidia richiardii* were obtained, these were placed in a well-supported clade with those sequences from the same species retrieved from GenBank.

3.3. Species delimitation analysis

Using standard settings initial partition of the ABGD analysis resulted in 12 molecularly defined operational taxonomic units (MOTUs) for our 406 fragment of the *Cytochrome Oxidase I* gene region. This first result clustered *Oc. caspius* and all members of the *Aedes mariae* complex (*Ae. mariae*, *Ae. zammitii* and *Ae. phoeniciae*) in a single MOTU and also failed to separate *Cs. annulata* from *Cs. subochrea*, while the remaining MOTUs corresponded to morphologically-identified species. On the other hand, this conservative approach split *An. algeriensis* in two species each. We also selected a recursive partition with a P value=4.64e-03, since this was the first to separate the three species of the Mariae Complex. However, when these were divided, the cluster containing *Ae. mariae* specimens from Majorca were also separated as a distinctive species (Figure 2). No significant differences were observed when using Jukes Cantor or Kimura K80. The bPTP analysis based on the Bayesian phylogenetic

tree identified 23 MOTUs. The results of bPTP are mostly consistent with the ABGDr analysis; however, it is not as accurate as it separates some individuals as different MOTUs.

The last approach to separate species using the *COI* gene region was the genealogical haplotype network analysis using the statistical parsimony method implemented in TCS (Clement *et al.*, 2000). This method split our dataset in 19 networks separated by 11 mutational steps. The three species from the Mariae Complex were separated as different groups; however, *Oc. caspius* was included in the same network than *Ae. mariae*. The remaining specimens showed a clear agreement among haplotype networks and the morphological identification (Figure 3). There were clear disagreements in the number of species identified by each approach. However, most algorithms agreed in the separation of the MOTUs detailed in Figure 3. The Kimura 2-parameter showed that the majority of MOTUs identified by the species boundaries approaches could be identified using Hebert's 10 X threshold (Table 3), with the exception of the members of the Mariae Complex and *Oc. caspius* from Majorca.

3.4. Agreement between molecular and morphologically-based identification

Using morphology-based keys we identified 13 mosquito species that belong to 7 different genera. Morphological identification using available keys were efficient to distinguish 11 of the species, with the exception of *Cs. annulata* and *Cs. subochrea*. Since the molecular identification resulted in conflictive arrangement of specimens, we re-examined the morphology of these and confirmed that all Balearic specimens showed morphological characters in the wings similar to *Cs. annulata* ("Cross veins r-m and m-cu aligned"), but abdominal scales like *Cs. subochrea* "Terga with indistinct pale basal bands formed by yellowish scales (not white), pale scales are also present among the dark scales in the apical half of the terga". A somehow different situation was observed in La Rioja, with most specimens falling into the key's description but some of them (*Cs.a_93LR*) with wings like *Cu. annulata* and terga like *Cs. subochrea* (Fig. 2). When characters located in the wings were ignored, the morphological identification was in agreement with the DNA-barcoding analysis.

4. Discussion

Establishing a genetic distance value in *COI* for species delimitation is a controversial topic that has been largely under scrutiny (Janzen *et al.*, 2017). Although a general value of 3% split for invertebrates species has been proposed, for mosquito identification the Barcoding

community generally agrees with a Kimura-2-parameter value higher than 0.02 (>2%) to separate species (Kumar *et al.*, 2007; Ruiz-Lopez *et al.*, 2012). A low genetic divergence value (2,5%) has been used to differentiate sibling species with different vectorial importance such as *Anopheles nuneztovari* from *An. dunhami*, a less important vector, in Colombia (Ruiz *et al.*, 2010). Closely related species within the *Anopheles albitarsis* group could also be differentiated using this value (Ruiz-Lopez *et al.*, 2012). Although this value may seem insufficient, a large-scale barcoding study conducted in butterflies, combining genomic and ecological data showed that a difference as shallow as 1 to 2% in the *COI* studied could be enough to differentiate species (Janzen *et al.*, 2017).

Using Bayesian and distance-based phylogenetic approaches and a combination of four additional species delimitation algorithms, we confirmed the identity of 11 out of 13 morphologically-identified mosquitoes (85%), demonstrating that DNA barcoding is an effective strategy to identify mosquitoes in Spain. These results are consistent with reports in other regions of the world, where congruence between these approaches ranged from 82% in Colombia (Rozo-Lopez & Mengual, 2015) up to more than 98% in most countries (Kumar *et al.*, 2007; Wang *et al.*, 2012; Ashfaq *et al.*, 2014; Chan *et al.*, 2014; Versteirt *et al.*, 2015). These differences may be related to the limitation of this technique in some taxa (M. Laurito *et al.*, 2017) or to errors in the initial morphological identification leading to incorrectly naming the sequences in the public databases used for sequence comparisons (Krzywinski & Besansky, 2003). We found disagreement between the barcoding and morphological approaches to separate *Cs. annulata* from *Cs. subochrea*. Furthermore, the sequence analysis also showed preliminary evidence of cryptic speciation in *An. algeriensis*, *Ae. vexans* and in *Ae. mariaae* s.s. It is important to emphasize that the results obtained in this study are not definitive, as this only included one molecular marker. An inflation of the number of species due to the sole use of molecular methods is a risk to take into consideration (Harbach, 2004). Therefore, a combination of several markers is compulsory in order to arrive to firm conclusions on these potential cryptic species. Nevertheless, the DNA-barcoding approach has often served as a first step to identify cryptic species within Culicidae (Ruiz-Lopez *et al.*, 2012) and we hope that this will open a new window of research in this field in Spain, where identifications rely almost entirely in morphology. Results are discussed in depth below.

4.1. Preliminary evidence for cryptic speciation

All species delimitation approaches, including Hebert's 10 x thresholds, agreed in the existence of at least two members within *An. algeriensis*. This is formed by two highly supported monophyletic clades separated by a high genetic divergence ($K-2P = 6\%$). One of them, herein putative *An. algeriensis* A would be formed only by the Spanish specimens from Navarra and Majorca, while *An. algeriensis* B would be formed by specimens from Romania and Sweden

and probably others not included in this study. *Anopheles algeriensis* is widely distributed in the Palearctic region, with Yemen as the easternmost limit (Al-Eryani *et al.*, 2016). This species is scarce in most areas where it has been reported (Krüger & Tannich, 2013), including the Balearic Islands. This scarcity led to the conclusion that *An. algeriensis* is a malaria vector of secondary importance (Becker *et al.*, 2010). However, this would not be the case of La Rioja, where it has been recorded as the third most abundant species (23,3% of all collected mosquitoes) in the wetland of La Grajera (Ruiz-Arrondo *et al.*, 2017). Furthermore, it has recently been found naturally infected with *Plasmodium vivax* in Yemen (Al-Eryani *et al.*, 2016), which should open new questions on their role in malaria transmission in endemic areas. In Spain, this species has been reported showing a “confusing” behavior, with variations in its host and resting preferences (R Bueno Marí *et al.*, 2011). Different patterns in mosquito behavior have been the first evidence of the possibility of cryptic speciation in other *Anopheles* mosquitoes (Paredes-Esquivel, Donnelly, *et al.*, 2009) and may be related to our findings. However these would not be surprising considering that half of the *Anopheles* species are part of species complexes (Harbach, 2004).

The use of nuclear and mitochondrial markers has recently allowed the detection of two different lineages within *Ae. vexans* (Lilja *et al.*, 2018). Our results are consistent with these findings, with all species delimitation analysis confirming the separation of at least two groups of *Ae. vexans* in the dataset. Mosquitoes collected from Majorca (*Ae. vexans* A) would belong to the authors’ Group 2, which has also been reported in Hungary and Sweden (Lilja *et al.*, 2018). *Ae. vexans* is a competent vector to more than 30 arboviruses, including West Nile virus, Zika virus and the dog heartworm *Dirofilaria immitis* (Gendernalik *et al.*, 2017). In Spain it has been reported in 22 provinces, including the Balearic Islands (R. Bueno Marí *et al.*, 2012). *Ae. vexans* is a floodwater mosquito, with highly resistant eggs which can remain viable for several years. This species can travel long distances, which has direct relationship with its invasive capacity.

4.2. Incipient speciation in the Balearic rock pool mosquitoes?

The sea water contained in the rock pools of the Mediterranean coasts is habitat for mosquitoes of the *Aedes mariaae* complex, which include three sibling species *Ae. mariaae* s.s., *Ae. phoeniciae* and *Ae. zammitii* (Coluzzi & Sabatini, 1968). The phylogenetic analysis shows that *Ae. mariaae* is not reciprocally monophyletic, with the Majorcan clade separated by >2% from the other members of the Mariae Complex, including the *Ae. mariaae* specimens from mainland Spain and Italy. The Majorcan clade is also 2% distant to *Oc. caspius* and it appears more closely related to this species in the phylogenetic tree. However, the low bootstrap values obtained will not allow us to arrive to further conclusions. Contradictory results were obtained with the 10x threshold proposed by Hebert (2004) and the initial partition in ABGD. These

approaches not only failed in separating the Majorcan clade, but also placed all members of the Mariae Complex plus *Oc. caspius* as a single species. These results are not consistent with the current knowledge of these species. The Mariae Complex has been subjected to a broad analysis, which included the morphological, genetic and reproduction studies of this group (Coluzzi & Sabatini, 1968; Mastrantonio *et al.*, 2016). There is no discussion on the species status of its members. However, it is known that they have a parapatric distribution with incomplete reproductive isolation (Mastrantonio *et al.*, 2016). Furthermore, including *Oc. caspius* and the Mariae Complex members in the same MOTU does not seem reasonable if we consider that these have a very distinctive ecology. While the former is a salt-marsh mosquito, the latter are rock pool mosquitoes with an extreme capacity to tolerate high salinity values. Further analysis using additional markers and running some experimental crosses in the laboratory would be necessary to establish the species status of the Majorcan *Ae. mariae* and its role as disease vector.

4.3. Current morphological keys cannot always distinguish *Culiseta annulata* from *Cs. subochrea* in Spain

Species *Cs. annulata* and *Cs. subochrea* are morphologically similar and can be differentiated by the presence of white (*Cs. annulata*) or yellow (*Cs. subochrea*) basal bands and the absence or presence of pale scales in the apical half of terga, respectively. In addition the alignment of veins r-m and m-c is effective to distinguish both species, being totally aligned in *Cs. annulata* and slightly separated in the case of *Cs. subochrea* (Becker *et al.*, 2010). The phylogenetic analysis and the species delimitation methods agreed in the clustering of two distinctive groups. As we obtained conflictive results with members of both species placed indistinctively in both clades (Figure 3), we re-checked the morphology and found that the character in the wings could not always separate both species, while the “the absence or presence of pale scales in the apical half of terga” was in clear agreement with DNA-barcoding identification. Conspicuous characters are frequently used by field workers as a careful examination of specimens can be time consuming. This may lead to erroneous delimitation of the distribution of these vectors. It is crucial to include other molecular markers to confirm these findings and re-design the morphological keys for these species in Spain. *Culiseta annulata* is a competent vector of the Tahyna virus vector (Bardos *et al.*, 1975) while we have found no reports on the vectorial capacity of *Cs. subochrea*. Therefore, its correct identification is crucial in case of outbreaks in the country. This would be the first record of *Cs. subochrea* in the Balearic Islands.

4.4. Disagreement among species delimitation methods to separate Spanish mosquitoes

We found disagreement among all species delimitation approaches used in this study. Methods such as ABGD and TCS allow the user to set the initial parameters, then results may be considered subjective. According to Puillandre *et al.* (2012) a P value = 0,01 is the closest to the species defined by authors when testing ABGD in four datasets. However, this value did not allow us to separate members of the *Aedes mariaae* complex and *Oc. caspius* that appeared together in the same MOTU. On the other hand, bPTP, ABGD using a P value=4.64e-03 and the combination of the Phylogenetic Species Concept plus the 2% K2P approach were effective methods to separate these species. The fact that ABGD at initial partition and TCS were not able to separate members of the Mariae Complex, may be related to the incipient speciation of the group, which were more easily detected by the remaining species delimitation approaches. largely studied groups such as those from the Mariae complex is important as they may also be used as a standardization strategy to determine which species delimitation method is more appropriate for our dataset. In the case of *Oc. caspius* this would require a more careful analysis as this species seem close to *Ae. mariaae* even when these vary in their habitats. Hybridization between these species cannot be ruled out.

A multilocus approach and a morphological revision is crucial to confirm the identification of these and other putative cryptic species and to confirm the validity of the morphological characters in *Culiseta annulata* and *Cs. subochrea*. As the identification approaches may likely disagree on the species composition, data on the ecological distribution the species should be incorporated as this may help clarify the species status as with *Oc. caspius*, which lives in salt-marsh mosquitoes, differing from the *Aedes mariaae* complex species, which live in the rock pools of the Mediterranean coast.

4.5. Implications on disease transmission

Nine of the eleven morphologically-identified species have been recorded as vectors of human diseases. The remaining two are *Culiseta longiareolata* which rarely feeds on humans but it is considered a vector of veterinary importance and *Uranotaenia unguiculata* an autogenous species with no importance as a disease vector (Becker *et al.*, 2010). During the last years, there has been an increased awareness of mosquitoes and the diseases they transmit in Europe. Changes in climatic conditions together with globalization may result in active transmission of pathogens as it has already occurred in some areas. Their relative importance of such vectors depends on their feeding behavior, longevity, and other aspects that may vary in species complexes (Gendernalik *et al.*, 2017). Therefore, it is crucial to know which

vector species are currently circulating in the continent. Identifying lineages within mosquito species has important epidemiological implications. For instance, the geographic limits of the lineages of the species *Culex annulirostris* has been associated to the transmission of Japanese Encephalitis Virus in Australasia (Hemmerter *et al.*, 2007). Cryptic speciation in mosquitoes is an underexplored topic in Europe and this study demonstrates that it may be present in many taxa. We hope this study will open the door to new studies in this topic.

Acknowledgements

This work was funded by various sources: Mosquito collection was supported by the *Conselleria de Medi Ambient Agricultura y Pesca* of the *Govern of the Illes Balears*, DNA sequencing was possible thanks to *Ajuts a la recerca* for unfunded projects from the University of the Balearic Islands. We would also like to thank the *SOIB JOVE* program for funding the contract of the first author of the study and Mr. David Borràs for providing mosquito samples.

Table 1. List of specimens included in this study

<i>Haplotype</i>	<i>Sample ID</i>	<i>Specie</i>	<i>Sampling site</i>	<i>Genbank Accession</i>
H1	Ae.m_54M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H1	Ae.m_83M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H1	Ae.m_84M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H2	Ae.m_37M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370031
H3	Ae.m_25M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H3	Ae.m_55M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H3	Ae.m_79M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H3	Ae.m_82M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H4	Oc.c_KM243949	<i>Ochlerotatus caspius</i>	Germany	KM243949
H4	Oc.c_85M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH370035
H5	Oc.c_4M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH370036
H6	Oc.c_86M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH348267
H7	Ae.m_76V	<i>Aedes mariaae</i>	Castellón (Spain)	MH348265
H8	Ae.m_77V	<i>Aedes mariaae</i>	Castellón (Spain)	MH370033
H9	Ae.m_78V	<i>Aedes mariaae</i>	Castellón (Spain)	MH370034
H10	Ae.m_KM592039	<i>Aedes mariaae</i>	Scilla (Italy)	KM592039
H11	Ae.z_KM592049	<i>Aedes zammitii</i>	Crete (Greece)	KM592049
H12	Ae.p_KM592051	<i>Aedes phoeniciae</i>	Cyprus (Greece)	KM592051
H13	Oc.d_KM258326	<i>Ochlerotatus detritus</i>	Westkapelle (Belgium)	KM258326
H14	Oc.d_KU877017	<i>Ochlerotatus detritus</i>	Hampshire (U.K)	KU877017
H14	Oc.d_29M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH348269
H14	Oc.d_42M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH348269
H15	Oc.d_2M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH370048
H16	Oc.d_23M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH370049
H17	Cs.a_KX675395	<i>Culiseta annulata</i>	Bonn (Germany)	KX675395
H17	Cs.a_KU877021	<i>Culiseta annulata</i>	Surrey (U.K)	KU877021
H17	Cs.a_19LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370038
H17	Cs.a_20LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370038
H18	Cs.a_18LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370039
H19	Cs.a_52LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370042
H20	Cs.a_21LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370041
H20	Cs.s/a_093LR	<i>Culiseta subochrea/annulata</i>	La Rioja (Spain)	MH370041
H20	Cs.s/a_93LR	<i>Culiseta subochrea/annulata</i>	La Rioja (Spain)	MH370040
H21	Cs.s/a_5M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370043
H21	Cs.s_36LR	<i>Culiseta subochrea</i>	La Rioja (Spain)	MH370044
H21	Cs.s_42N	<i>Culiseta subochrea</i>	Navarra (Spain)	MH370044
H21	Cs.s_43N	<i>Culiseta subochrea</i>	Navarra (Spain)	MH370044

H22	Cs.s_37LR	<i>Culiseta subochrea</i>	La Rioja (Spain)	MH370047
H23	Cs.s/a_30M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370045
H23	Cs.s/a_58M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370046
H24	Cs.l_HG931139	<i>Culiseta longiareolata</i>	Kabul (Afghanistan)	HG931139
H24	Cs.l_JQ388785	<i>Culiseta longiareolata</i>	Freiburg (Germany)	JQ388785
H24	Cs.l_7M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H24	Cs.l_27M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H24	Cs.l_44M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H25	Cs.l_31M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH370037
H26	Cx.t_KJ012234	<i>Culex theileri</i>	Kars Ili (Turkey)	KJ012234
H27	Cx.t_6M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H27	Cx.t_19M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H27	Cx.t_28M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H28	Cx.t_36M	<i>Culex theileri</i>	Majorca (Spain)	MH370030
H29	Cx.p_10M	<i>Culex pipiens</i>	Majorca (Spain)	MH370028
H30	Cx.p_KF919189	<i>Culex pipiens</i>	Cordoba (Argentina)	KF919189
H31	Cx.p_KP293425	<i>Culex pipiens</i>	Stinson Beach (EEUU)	KP293425
H31	Cx.q_KF919190	<i>Culex quinquefasciatus</i>	Cordoba (Argentina)	KF919190
H31	Cx.p_18M	<i>Culex pipiens</i>	Majorca (Spain)	MH348262
H31	Cx.p_67M	<i>Culex pipiens</i>	Majorca (Spain)	MH348262
H32	Cx.p_11M	<i>Culex pipiens</i>	Majorca (Spain)	MH370029
H33	Ur.u_KM280589	<i>Uranotaenia unguiculata</i>	Austria	KM280589
H34	Ur.u_13M	<i>Uranotaenia unguiculata</i>	Majorca (Spain)	MH348263
H35	Cq.r_JX040513	<i>Coquillettidia richiardii</i>	Vasterbotten (Sweden)	JX040513
H36	Cq.r_91M	<i>Coquillettidia richiardii</i>	Majorca (Spain)	MH348270
H37	Ae.v_KM452935	<i>Aedes vexans</i>	Szeged (Hungary)	KM452935
H38	Ae.v_12M	<i>Aedes vexans</i>	Majorca (Spain)	MH348259
H39	Ae.v_KY609213	<i>Aedes vexans</i>	Kävlinge (Sweden)	KY609213
H39	Ae.v_KY609201	<i>Aedes vexans</i>	Kävlinge (Sweden)	KY609201
H39	Ae.v_21M	<i>Aedes vexans</i>	Majorca (Spain)	MH370023
H40	Ae.v_KP942689	<i>Aedes vexans</i>	Norralla (Sweden)	KP942689
H41	Ae.v_KP942682	<i>Aedes vexans</i>	Deherhamn (Sweden)	KP942682
H42	Ae.a_KU738431	<i>Aedes albopictus</i>	China	KU738431
H42	Ae.a_16M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_17M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_43M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_70M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H43	An.a_15N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370027
H44	An.a_11N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370025
H44	An.a_18N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370025
H45	An.a_3M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH370024
H46	An.a_12N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH348261
H46	An.a_14N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH348261
H47	An.a_8M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260
H47	An.a_14M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260
H47	An.a_15M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260

H48	An.a_13N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370026
H49	An.a_KU214675	<i>Anopheles algeriensis</i>	Romania	KU214675
H50	An.a_KP942711	<i>Anopheles algeriensis</i>	Gotland (Sweden)	KP942711
H51	Outgroup_JF923694	<i>Anopheles darlingi</i>	Brazil	JF923694

Sequences in bold correspond to those retrieved from GenBank for comparison purposes

Table 2. Summary of genetic diversity parameters per mosquito species included in this study

Species	N	SS	h	Hd (SD)	π
<i>Aedes mariaae</i> M	8	3	3	0.679 (0.122)	0.00405
<i>Ochlerotatus caspius</i>	4	5	3	0.833 (0.222)	0.00616
<i>Aedes mariaae</i> C	4	4	4	1.000 (0.177)	0.00493
<i>Ochlerotatus detritus</i>	6	10	4	0.800 (0.172)	0.00870
<i>Culiseta annulata</i>	9	3	4	0.750 (0.112)	0.00233
<i>Culiseta subochrea</i>	7	2	3	0.667 (0.160)	0.00188
<i>Culiseta longiareolata</i>	6	2	2	0.333 (0.215)	0.00164
<i>Culex theileri</i>	5	2	3	0.700 (0.218)	0.00197
<i>Culex pipiens</i>	7	12	4	0.714 (0.181)	0.00891
<i>Uranotaenia unguiculata</i>	2	9	2	1.000 (0.500)	0.02217
<i>Coquillettidia richiardii</i>	2	4	2	1.000 (0.500)	0.00985
<i>Aedes vexans</i> A	5	4	3	0.700 (0.218)	0.00443
<i>Aedes vexans</i> B	2	11	2	1.000 (0.500)	0.02709
<i>Aedes albopictus</i>	5	0	1	0.000 (0.000)	0.00000
<i>Anopheles algeriensis</i> A	10	4	6	0.889 (0.075)	0.00421
<i>Anopheles algeriensis</i> B	2	11	2	1.000 (0.500)	0.02709

N=number of samples, SS= Number of segregating sites, h=number of haplotypes, Hd=Haplotype diversity (Standard Deviation), π = nucleotide diversity

Table 3. Kimura 2-parameter of Molecular Operational Taxonomic Units identified in the phylogenetic analysis

MOTUs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 <i>Ae.mariae</i> M	0,004																	
2 <i>Oc.caspicus</i>	0,02	0,006																
3 <i>Ae.mariae</i> C	0,02	0,03	0,005															
4 <i>Ae.zammitii</i>	0,04	0,04	0,03	-														
5 <i>Ae.phoeniciae</i>	0,04	0,04	0,03	0,03	-													
6 <i>Oc.detritus</i>	0,10	0,11	0,10	0,09	0,09	0,009												
7 <i>Cs.annulata</i>	0,14	0,13	0,14	0,15	0,14	0,14	0,002											
8 <i>Cs.subochrea</i>	0,13	0,12	0,13	0,13	0,13	0,14	0,05	0,002										
9 <i>Cs.longiareolata</i>	0,12	0,12	0,12	0,13	0,12	0,14	0,13	0,14	0,002									
10 <i>Cx.theileri</i>	0,12	0,13	0,12	0,13	0,13	0,14	0,15	0,14	0,12	0,002								
11 <i>Cx.pipiens</i>	0,11	0,11	0,11	0,12	0,11	0,14	0,15	0,16	0,12	0,07	0,009							
12 <i>Ur.unguiculata</i>	0,18	0,17	0,18	0,16	0,18	0,18	0,17	0,17	0,18	0,18	0,17	0,023						
13 <i>Cq.richiardii</i>	0,21	0,20	0,21	0,22	0,22	0,22	0,21	0,21	0,20	0,19	0,19	0,23	0,010					
14 <i>Ae.vexans</i> A	0,12	0,12	0,12	0,12	0,11	0,13	0,16	0,14	0,14	0,15	0,12	0,18	0,20	0,004				
15 <i>Ae.vexans</i> B	0,12	0,12	0,12	0,13	0,12	0,14	0,18	0,15	0,15	0,15	0,14	0,19	0,21	0,06	0,028			
16 <i>Ae.albopictus</i>	0,16	0,16	0,15	0,13	0,13	0,14	0,16	0,17	0,16	0,14	0,13	0,18	0,21	0,12	0,12	0,00		
17 <i>An.algeriensis</i> A	0,17	0,17	0,16	0,15	0,15	0,21	0,21	0,21	0,17	0,17	0,17	0,21	0,24	0,18	0,20	0,18	0,004	
18 <i>An.algeriensis</i> B	0,18	0,19	0,18	0,17	0,18	0,23	0,23	0,22	0,19	0,18	0,19	0,22	0,24	0,19	0,22	0,19	0,07	0,028

References

- Al-Eryani, S. M. A., Kelly-Hope, L., Harbach, R. E., Briscoe, A. G., Barnish, G., Azazy, A., *et al.* (2016) Entomological aspects and the role of human behaviour in malaria transmission in a highland region of the Republic of Yemen. *Malaria Journal*, **15**.
- Ashfaq, M., Hebert, P. D. N., Mirza, J. H., Khan, A. M., Zafar, Y. & Mirza, M. S. (2014) Analyzing Mosquito (Diptera: Culicidae) Diversity in Pakistan by DNA Barcoding. *Plos One*, **9**.
- Bardos, V., Ryba, J. & Hubalek, Z. (1975) Isolation of Tahyna virus from field collected *Culiseta annulata* (Schrk) larvae *Acta Virologica*, **19**, 446-446.
- Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., *et al.* (2010) *Mosquitoes and their control*. Springer-Verlag Berlin Heidelberg, New York.
- Bourke, B. P., Oliveira, T. P., Suesdek, L., Bergo, E. S. & Mureb Sallum, M. A. (2013) A multi-locus approach to barcoding in the Anopheles strodei subgroup (Diptera: Culicidae). *Parasites & Vectors*, **6**.
- Bueno Marí, R., Bernués Bañeres, A., Chordá Olmos, F. A. & Jiménez Peydró, R. (2011) Contributions to the knowledge of the distribution and biology of Anopheles algeriensis Theobald, 1903 in Spain. *Boletín de malariología y Salud Ambiental*, **51**, 93-96.
- Bueno Marí, R., Bernués Bañeres, A. & Jiménez Peydró, R. (2012) Updated checklist and distribution maps of mosquitoes (Diptera: Culicidae) of Spain., pp. 91 - 127. European Mosquito Bulletin.
- Campbell, G. L., Ceianu, C. S. & Savage, H. M. (2001) Epidemic West Nile encephalitis in Romania - Waiting for history to repeat itself. In *West Nile Virus: Detection, Surveillance, and Control* (ed. by D. J. White & D. L. Morse), pp. 94-101.
- Chan, A., Chiang, L.-P., Hapuarachchi, H. C., Tan, C.-H., Pang, S.-C., Lee, R., *et al.* (2014) DNA barcoding: complementing morphological identification of mosquito species in Singapore. *Parasites & Vectors*, **7**.
- Clement, M., Posada, D. & Crandall, K. A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.
- Coluzzi, M. & Sabatini, A. (1968) Divergenze morfologiche e barriere di sterilità nel complesso *Aedes mariaae* (Diptera: Culicidae). *Riv. Parasit*, **29**, 49-70.
- Cracraft, J. (1987) Species concepts and the ontology of evolution. *Biology and philosophy*, **2**, 329-346.
- Cywinska, A., Hunter, F. F. & Hebert, P. D. N. (2006) Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology*, **20**, 413-424.
- Gendernalik, A., Weger-Lucarelli, J., Luna, S. M. G., Fauver, J. R., Ruckert, C., Murrieta, R. A., *et al.* (2017) American *Aedes vexans* Mosquitoes are Competent Vectors of Zika Virus. *American Journal of Tropical Medicine and Hygiene*, **96**, 1338-1340.

Gould, E. A., Gallian, P., de Lamballerie, X. & Charrel, R. N. (2010) First cases of autochthonous dengue fever and chikungunya fever in France: from bad dream to reality! *Clinical Microbiology and Infection*, **16**, 1702-1704.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Harbach, R. E. (2004) The classification of genus *Anopheles* (Diptera : Culicidae): a working hypothesis of phylogenetic relationships. *Bulletin of Entomological Research*, **94**, 537-553.

Harbach, R. E. (2016) Mosquito Taxonomic Inventory. Valid Species List.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 313-321.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.

Hemmerter, S., Slapeta, J., van den Hurk, A. F., Cooper, R. D., Whelan, P. I., Russell, R. C., *et al.* (2007) A curious coincidence: mosquito biodiversity and the limits of the Japanese encephalitis virus in Australasia. *BMC Evolutionary Biology*, **7**.

Janzen, D. H., Burns, J. M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., *et al.* (2017) Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 8313-8318.

Krüger, A. & Tannich, E. (2013) Rediscovery of *Anopheles algeriensis* Theob. (Diptera: Culicidae) in Germany after half a century. *Journal of the European Mosquito Control Association*, **31**, 14-16.

Krzywinski, J. & Besansky, N. J. (2003) Molecular systematics of *Anopheles*: From subgenera to subpopulations. *Annual Review of Entomology*, **48**, 111-139.

Kumar, N. P., Rajavel, A. R., Natarajan, R. & Jambulingam, P. (2007) DNA barcodes can distinguish species of Indian mosquitoes (Diptera : Culicidae). *Journal of Medical Entomology*, **44**, 1-7.

Laurito, M., Ayala, A. M., Almiron, W. R. & Gardenal, C. N. (2017) Molecular identification of two *Culex* (*Culex*) species of the neotropical region (Diptera: Culicidae). *Plos One*, **12**.

Laurito, M., de Oliveira, T. M. P., Almiron, W. R. & Mureb Sallum, M. A. (2013) *COI* barcode versus morphological identification of *Culex* (*Culex*) (Diptera: Culicidae) species: a case study using samples from Argentina and Brazil. *Memorias Do Instituto Oswaldo Cruz*, **108**, 110-U152.

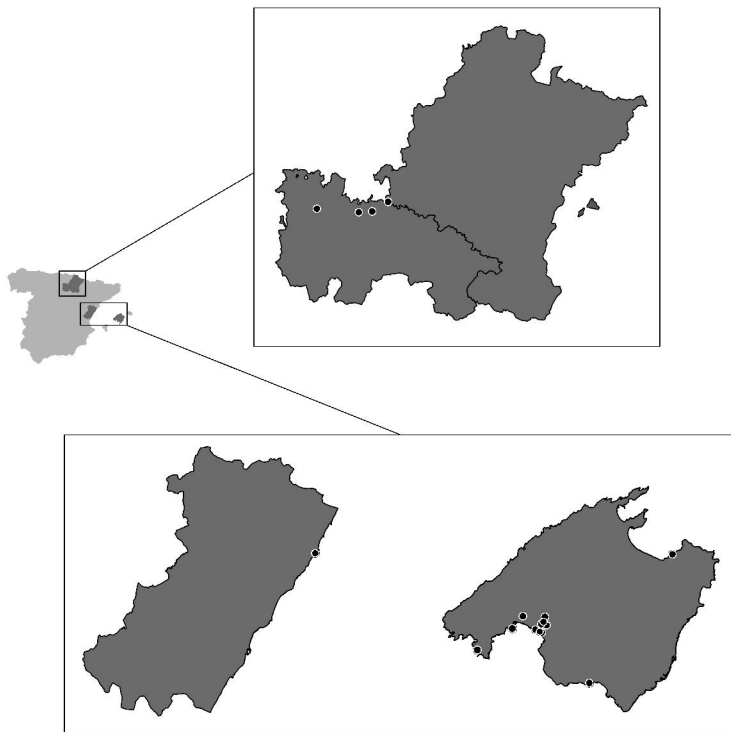
Lilja, T., Troell, K., Kirik, H. & Lindström, A. (2018) A distinct group of north European *Aedes vexans* as determined by mitochondrial and nuclear markers. *Medical and Veterinary Entomology*.

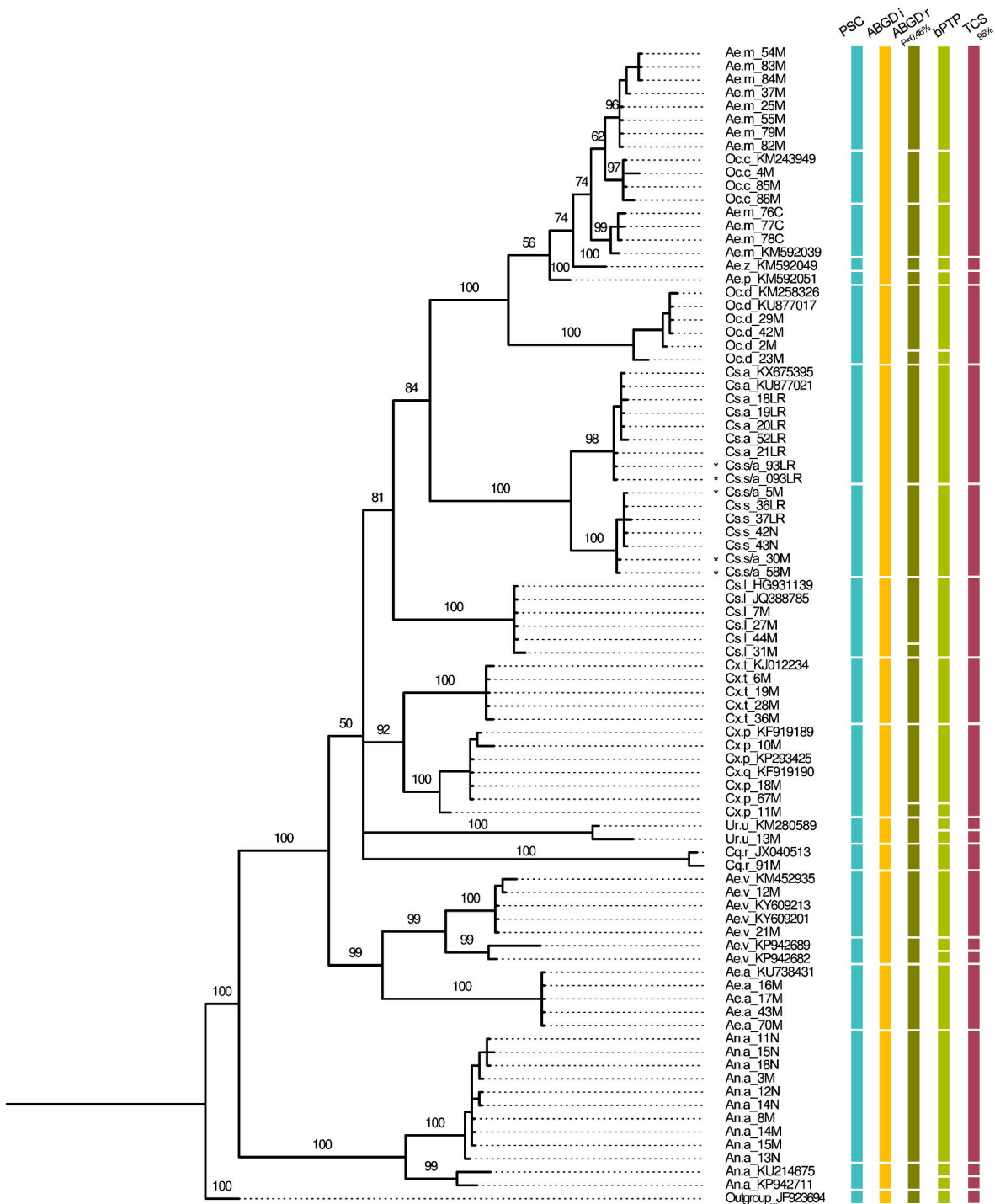
- Mastrantonio, V., Porretta, D., Urbanelli, S., Crasta, G. & Nascetti, G. (2016) Dynamics of mtDNA introgression during species range expansion: insights from an experimental longitudinal study. *Scientific Reports*, **6**.
- Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., *et al.* (2001) The outbreak of West Nile virus infection in the New York City area in 1999. *New England Journal of Medicine*, **344**, 1807-1814.
- Paredes-Esquivel, C., Del Rio, R., Monerris, M., Marti, T., Borrás, D. & Miranda, M. A. (2009) High Prevalence of Myiasis by *Oestrus ovis* in the Balearic Islands *Parasite*, **16**.
- Paredes-Esquivel, C., Donnelly, M. J., Harbach, R. E. & Townson, H. (2009) A molecular phylogeny of mosquitoes in the *Anopheles barbirostris* Subgroup reveals cryptic species: Implications for identification of disease vectors *Molecular Phylogenetics and Evolution*, **50**, 141-151.
- Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, **21**, 1864-1877.
- Rezza, G., Nicoletti, L., Angelini, R., Romi, R., Finarelli, A. C., Panning, M., *et al.* (2007) Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet*, **370**, 1840-1846.
- Ronquist, F. & Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.
- Rozas, J. & Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174-175.
- Rozo-Lopez, P. & Mengual, X. (2015) Mosquito species (Diptera, Culicidae) in three ecosystems from the Colombian Andes: identification through DNA barcoding and adult morphology. *Zookeys*, 39-64.
- Ruiz-Arrondo, I., Hernández Triana, L. & Oteo Revuelta, J. A. (2017) Mosquito fauna (Diptera, Culicidae) of the wetland of La Grajera (Logroño) and its implications for public health. *Zubía* **35**, 123-140.
- Ruiz-Lopez, F., Wilkerson, R. C., Conn, J. E., McKeon, S. N., Levin, D. M., Quinones, M. L., *et al.* (2012) DNA barcoding reveals both known and novel taxa in the Albitarsis Group (*Anopheles*: Nyssorhynchus) of Neotropical malaria vectors. *Parasites & Vectors*, **5**.
- Ruiz, F., Linton, Y.-M., Ponsonby, D. J., Conn, J. E., Herrera, M., Quinones, M. L., *et al.* (2010) Molecular comparison of topotypic specimens confirms *Anopheles* (Nyssorhynchus) *dunhami* Causey (Diptera: Culicidae) in the Colombian Amazon. *Memorias Do Instituto Oswaldo Cruz*, **105**, 899-903.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.
- Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of the Mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

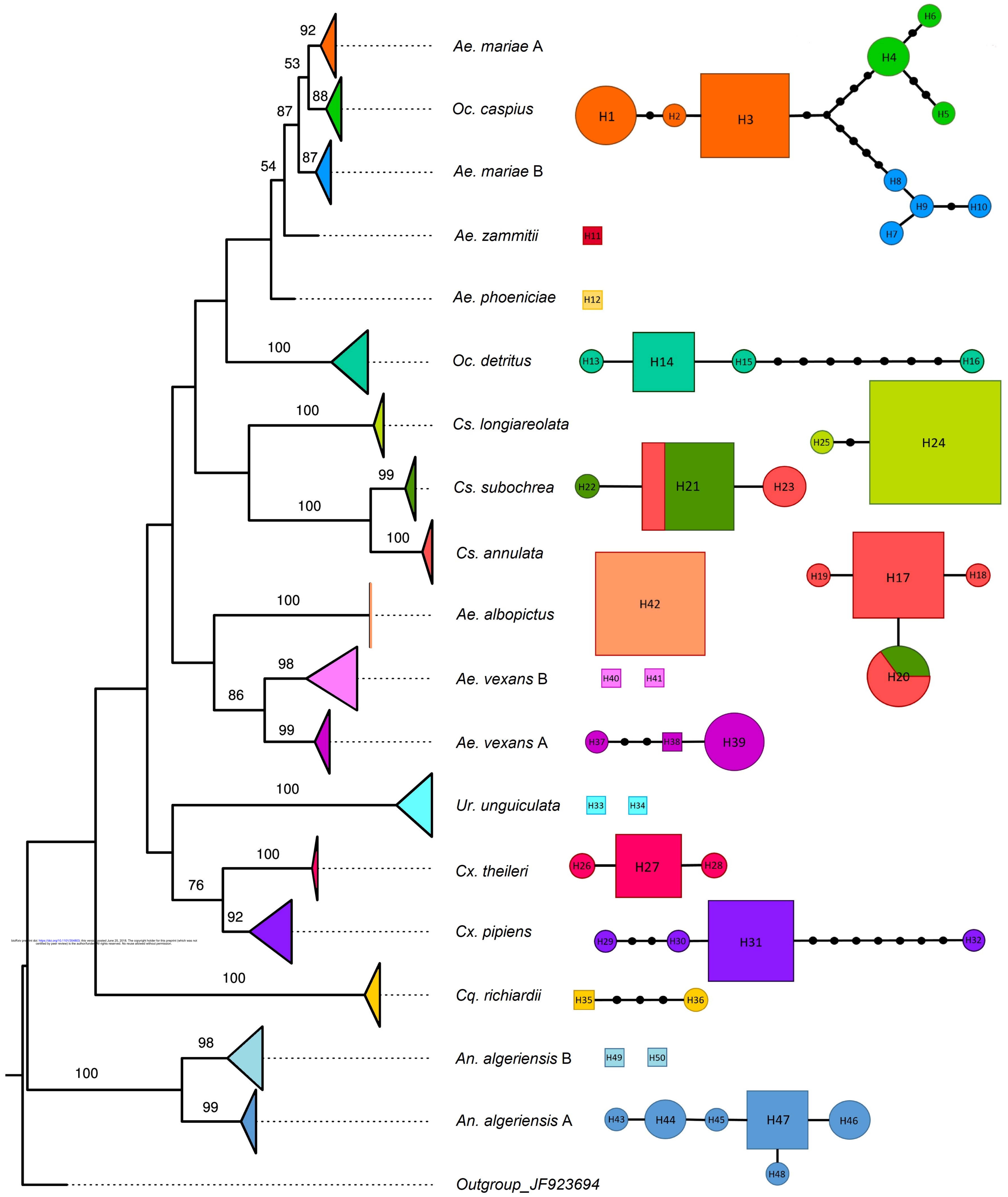
Versteirt, V., Nagy, Z. T., Roelants, P., Denis, L., Breman, F. C., Damiens, D., *et al.* (2015) Identification of Belgian mosquito species (Diptera: Culicidae) by DNA barcoding. *Molecular Ecology Resources*, **15**, 449-457.

Wang, G., Li, C., Guo, X., Xing, D., Dong, Y., Wang, Z., *et al.* (2012) Identifying the Main Mosquito Species in China Based on DNA Barcoding. *Plos One*, **7**.

Zhang, J. J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, **29**, 2869-2876.







bioRxiv preprint doi: <https://doi.org/10.1101/354803>; this version posted June 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Opening the Pandora Box: DNA-barcoding evidence limitations of morphology to identify Spanish mosquitoes

DELGADO-SERRA, SOFÍA¹; VIADER, MIRIAM¹; RUIZ-ARRONDO, IGNACIO²; MIRANDA, MIGUEL ÁNGEL¹; BARCELÓ, CARLOS¹; BUENO-MARI³, RUBÉN; HERNÁNDEZ-TRIANA, LUIS M.⁴; MIQUEL, MARGA¹ AND PAREDES-ESQUIVEL, CLAUDIA^{1*}

1. Laboratory of Zoology. University of the Balearic Islands. Palma de Mallorca. Spain
2. Center for Rickettsiosis and Arthropod-Borne Diseases, CIBIR, Logroño, La Rioja, Spain.
3. Spain.Research and Development (R+D) Department. Laboratorios Lokímica. Valencia. Spain.
4. Wildlife Zoonoses and Vector-borne Diseases Research Group, Virology Department, Animal and Plant Health Agency, Addlestone, United Kingdom.

* Corresponding author

Abstract

Cryptic speciation is frequent in the medically important mosquitoes. While most findings have been reported in tropical regions, it is an unexplored topic in countries where mosquito-borne diseases are not endemic, like Spain. The occurrence of recent outbreaks in Europe has increased the awareness on the native and invasive mosquito fauna present in the continent. Therefore, the central question of this study is whether the typological approach is sufficient to identify Spanish mosquitoes. To address this problem, we confronted the results of the morphological identification of 62 adult specimens collected from four different regions of Spain (La Rioja, Navarra, Castellón and the Island of Majorca) with the results obtained through DNA-barcoding. We conducted a comprehensive phylogenetic analysis of the *COI* gene region and compared this with the results of four species delimitation algorithms (ABGD initial partition, ABGD P=0.46%, bPTP and TCS). We report strong evidence for cryptic speciation in *Anopheles algeriensis* and *Aedes vexans* and reproductive isolation of the rock pool mosquito *Aedes mariae*. In addition, we report that the character present in the wings is not efficient to distinguish species *Culiseta annulata* from *Culiseta subochrea*, which distribution in the country may be different than previously described.

1. Introduction

Mosquitoes are responsible for the death of millions of people each year due to its role as vectors of some of the most devastating diseases in the world such as malaria, dengue and yellow fever. Although the majority of the cases are reported from tropical countries, several outbreaks have also taken place in the Northern hemisphere showing that the landscape of mosquito-borne diseases is changing in the world. In Italy, the rapidly-spreading invasive species *Aedes albopictus* have been implicated in outbreak of Chikungunya fever that affected 205 people in 2007 (Rezza *et al.*, 2007) and it was also identified as the vector of two Dengue fever cases in France (Gould *et al.*, 2010). Although it is unquestionable that invasive species represent a serious Public Health threat, the role of native mosquitoes should not be underestimated. Recent outbreaks of West Nile Virus (WNV) in Romania (Campbell *et al.*, 2001) and in the United States (Nash *et al.*, 2001), were attributed to local populations of *Culex* mosquitoes. Mosquito species identification has been traditionally carried out using dichotomous keys that describe the morphological characters of a particular life stage. Although this technique has proved valuable and is still widely used to distinguish many mosquito species, it has important limitations like the need of taxonomic experts to perform accurate identifications and the loss of morphological characters during collection and preservation of specimens. Furthermore, morphology-based taxonomy cannot differentiate indistinguishable members of species complexes, which frequently differ on their ecology, host preferences, behavioral patterns, therefore in their role as disease vectors, greatly impairing the design of effective control strategies (Paredes-Esquivel, Del Rio, *et al.*, 2009).

The field of mosquito taxonomy has being renewed with the development of new tools for accurate identification (Cywinska *et al.*, 2006). Molecular-based techniques are now widely used not only to identify sibling species but also to solve phylogenetic relationships (Bourke *et al.*, 2013). The *Cytochrome Oxidase I* gene region of the mitochondrial genome is the gold standard barcoding identification of species (Hebert *et al.*, 2003) and has proved valuable to distinguish many mosquito species (Cywinska *et al.*, 2006; Kumar *et al.*, 2007; Wang *et al.*, 2012; Ashfaq *et al.*, 2014). This approach is particularly important to establish species delimitations in unexplored groups (Puillandre *et al.*, 2012). Although, there have been reports that *COI* barcoding has failed in differencing some species of mosquitoes of *Anopheles* and *Culex* genus (Cywinska *et al.*, 2006; Kumar *et al.*, 2007; Wang *et al.*, 2012; Bourke *et al.*, 2013; Magdalena Laurito *et al.*, 2013), there is no doubt that DNA barcoding erects as a valuable alternative for mosquito identification in most areas of the world. However, the use DNA barcoding as a surveillance tool would require a database containing mosquito barcodes of correctly identified mosquitoes.

From the 3,552 formally recognized mosquito species in the world (Harbach, 2016), 100 have been recorded in Europe and this number might be even higher since six invasive *Aedes* species have recently colonized the continent (Versteirt *et al.*, 2015). Spain reports a similar

increasing pattern, with 59 species reported at the beginning of this decade and 64 described in 2012 according to the latest checklist (R. Bueno Marí *et al.*, 2012), including competent vectors of human diseases. To the best of our knowledge, this would be the first comprehensive faunistic study that includes molecular-based techniques in Spain where identifications have mainly relied on the use of dichotomous keys. We are hoping to contribute to the knowledge of mosquito fauna and demonstrate that molecular based studies can uncover hidden diversity among well-known mosquito species.

2. Material and methods

2.1 Mosquito collection and morphological identification

The mosquito specimens were collected from 17 sites located in four different regions of Spain, including the Island of Majorca in the Balearic archipelago, Castellón, La Rioja and Navarra (Figure 1). The majority of the specimens (42) were collected from the island of Majorca, the largest at the Balearic archipelago, located in the Western region of the Mediterranean Basin. Additionally, for comparison purposes, three specimens from *Ae. mariaae* were obtained from Peñiscola (Castellón, Spain), a coastal region with high abundance of salt marshes and rocky shores. Finally, 17 specimens from *Cs. annulata*, *Cs. subochrea* and *An. algeriensis* were collected from the wetlands of La Grajera and Hervías and in the River Iregua in La Rioja and Navarra, respectively. These sampling sites are located in the Ebro Valley, in the Northern region of Spain; with a transitional climate between the oceanic one in the North and a the semiarid found in the central Ebro Valley.

Adult mosquitoes were collected from outdoor using BG and CDC light traps containing CO₂. The samplings were carried out weekly and each trap was set at sunset. On the following morning, collected specimens were transported *in vivo* to the laboratory and then frozen at -20 °C for molecular studies. Larvae were also collected from breeding sites, transported to the laboratory and kept alive at 25°C until adults emerged. Female mosquitoes were identified to species level using keys by Becker et al. (2010). Scientific names of mosquitoes were verified using the Mosquito taxonomy inventory (Harbach, 2014).

2.2. DNA extraction, PCR amplification and sequencing

Genomic DNA was extracted from individual mosquitoes, using DNeasy Blood & Tissue Kit (QIAGEN, GmbH, Hilden, Germany). To increase DNA yield we used 100 µl elution buffer instead of 200 µl in the final step. DNA concentration was measured with a NanoDrop 1000 spectrophotometer Thermo Scientific (Saveen Werner ApS, Denmark) and then DNA extractions were kept at -20 °C (working solutions) and -80 °C (Stock solutions). PCR was performed using the thermal cycler GeneAmp PCR System 2400 (Perkin Elmer) using primers C1-J-1718 forward 5' - GGAGGATTTGGAAATTGATTAGTGCC- 3' and C1-N-2191 reverse 5'- CCCGGTAAAATTTAAAATATAAACTTC- 3'. Folmer primers (LCO1490 5'- GGTCACAAATCATAAAGATATTGG-3' and HCO2198 5'- TAAACTTCAGGGTGACCAAAAAATCA -3') were also tested but failed to amplify all specimens. The 50 µl PCR reaction contained of 2 µl of DNA template, QIAGEN 1X reaction buffer (Tris-Cl, KCl) and (NH₄)₂SO₄ 1.5–3.0 mM of MgCl₂, 0.2 mM of dNTPs, 1 unit of QIAGEN DNA polymerase and 125 nM of each primer. The PCR reaction conditions were as follow: 95 °C for 5 min followed by 92 °C for 2 min and 15 s; 30 cycles of 92 °C x 15 s, annealing temperature of 50 °C x 15 s and an extension temperature of 72 °C x 30 s, followed by a final extension at 72 °C for 1 min. PCR products were stained with Pronasafe nucleic acid staining solution (Conda Laboratories) and observed in a 2% agarose - TBE buffer gel. Samples were purified using the QIAquick PCR Purification Kit (QIAGEN), according to manufacturer's specifications, but purified DNA was eluted in 30 µl MilliQ water instead of 50 µl elution buffer and the template was sequenced in both directions in a ABI3730XL automatic sequencer (Macrogen Inc., Spain) with the same primers used in first amplification.

2.3. Sequence analysis

Sequences obtained were aligned using Clustal W and edited manually by using software BioEdit v7.0.8.0 (Hall, 1999). When point substitutions were detected, the reliability of the sequence was confirmed by looking at nucleotide peaks. All sequences were translated using the *Drosophila* genetic code in DnaSP version 5.10.1 (Rozas & Rozas, 1999) to ensure these were not pseudogenes using the mDNA. This program was also used to calculate haplotype diversity within species. The transition/transversion ratio and A+T content was calculated with Mega 6. For comparative purposes, we used BLASTn and sequences were retrieved from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

2.4. Phylogenetic analysis

The best evolutionary model to construct phylogenetic trees was selected according to their BIC score (Bayesian Information Criterion) obtained in Mega6 (Tamura *et al.*, 2007). With this program we also constructed the Maximum Likelihood (ML) tree and the robustness of the clustering was determined by the bootstrap analysis with 500 replicates. The Bayesian inferences were constructed with MRBAYES 3.2.6 (Ronquist & Huelsenbeck, 2003); using four Metropolis coupled Markov chain Monte Carlo (MCMC) chains were run for 500,000 generations (sampling every 10 generations) to allow adequate time for convergence. The analysis was finished when the standard deviation of split frequencies was less than 0.01. To assure we obtained the correct topology, analyses were run twice. All analyzes were performed with random starting trees. Branch support is indicated by the posterior probability values. Consensus trees were visualized with FigTree v1.4.3.

2.5. Species boundaries

We used the results obtained from the phylogenetic analysis as a first approach to distinguish species. Species were defined following the Phylogenetic species concept, that establishes that these are “irreducible cluster of organisms, within which there is a parental pattern of ancestry and descent, and which is diagnosably distinct from other such clusters” (Cracraft, 1987). The second approach we used to delimit species within our dataset was the Automatic Barcode Gap Discovery algorithm (ABGD), available at <http://wwwabi.snv.jussieu.fr/public/abgd/>. This method is independent of tree typology and split dataset into candidate species, based on the existence of barcode gap (interspecific divergence higher than intraspecific ones) and a prior intraspecific divergence (p) (Puillandre *et al.*, 2012). ABGD was carried out applying the K2P model, using default parameters, with the exception of the relative gap width, which was defined as 1.4. Results of a recursive and initial partition analysis were compared. The third approach to identify species was the Bayesian approach to Poisson Tree Processes model (bPTP) (Zhang *et al.*, 2013), which is based on the number of mutations in the branches of a phylogenetic tree. We run 100 000 trees, a thinning of 100 using a burn-in value of 0.1. A statistical parsimony networks were constructed using TCS (Clement *et al.*, 2000), which collapses DNA sequences into haplotypes and determines the frequency of those haplotypes within the dataset. Gaps were treated as 5th state or missing data. Size of circles represents the number of sequences within a haplotype. Although this method has been mainly used to construct genealogies, it has been also used in diversity studies in other taxa. Molecularly Defined Operational Taxonomic Units (MOTUs) were defined according to the results obtained from the different species delimitation approaches. If these methods resulted in a different species composition, we defined MOTUs as those defined by the majority of such methods. We constructed a pairwise distance matrix using the Kimura 2-parameter model (K-2P) of substitutions using the aligned *COI* sequences in Mega 6 (Tamura *et al.*, 2007). We

tested in the 10x threshold by Hebert et al (2004) to confirm morphological identification. This rule establishes that a new species is suspected if the average interspecific distance is at least 10 times higher than the average intraspecific divergence of the putative sister species.

3. Results

3.1. Sequence analysis

After alignment of the dataset, the matrix contained 406 bp long sequences. Neither insertion nor deletion events were observed in the dataset. All sequences were A+T rich, with an average content of 67.3% ranging 62.1% in *Coquillettidia richiardii* up to 69.2% in *Ae. mariae*, which is a similar value observed in other species of Culicidae (Paredes-Esquivel, Donnelly, *et al.*, 2009). From the 406 bp fragment, we found 147 parsimony informative sites and 10 singleton sites. Most substitutions were found at the third codon position, with a ratio of transitions/transversions of 1.13. Thirty-eight haplotypes were found in mosquitoes analyzed and their sequences were deposited in GenBank under accession numbers MH348258-MH348270 and MH370023-MH370049 (table 1). Haplotype diversity varied from 0 in *Ae. albopictus* up to 1 in *Ae. mariae* C, *Ur. unguiculata*, *Cq. richiardii*, *Ae. vexans* B and *An. algeriensis* B (Table 2). No evidence of pseudogenes was found as stop codons were not observed after translation, using the mitochondrial DNA genetic code for *Drosophila* sp.

Twenty-five sequences were retrieved from GenBank for comparative purposes (Table 2). Sequences belonged to the same species included in this study, with the exception of the *Mariae* Complex members *Ae. phoeniciae* and *Ae. zammitii* which were included to provide additional information on the *Ae. mariae* specimens from Majorca. A Kimura-2-parameter distance matrix was built for the entire dataset. The intraspecific divergence in the 406 bp fragment ranged 0.00-2.79 %, with an average 0.86%. On the other hand, distance among co-generic species ranged 2.13-15.74%. There was an overlap between the intraspecific and interspecific divergence (Table 3).

3.2. Phylogenetic analysis

Under the Bayesian Information criterion, the best evolutionary model obtained for our dataset was GTR+R. We used this model to the Bayesian analysis and substitution model by Tamura and Nei (Tamura & Nei, 1993) to construct the Maximum Likelihood phylogenetic tree. Both trees show clearly defined highly supported clusters (>87% and >96% in the Maximum

likelihood and Bayesian trees, respectively) in all MOTUs identified (Figures 2 and 3). In both trees, members of the Mariae complex clustered in the same clade with *Oc. caspius*. *Aedes mariae* appears polyphyletic, with the Majorcan clade being more closely related to *Ochlerotatus caspius* than to *Ae. mariae* specimens from Continental Europe. Nevertheless, the relationship between these clusters was poorly supported by the bootstrap value. Specimens identified as *Ae. vexans* were separated in two highly supported clusters (>99%), herein *Ae. vexans* A, which included specimens from Majorca and sequences retrieved from GenBank from Sweden and Hungary. *Ae. vexans* B was formed by sequences from Sweden. These two groups have a 6% sequence divergence. Two monophyletic clades with high bootstrap support were also found in *An. algeriensis*. Those specimens from Majorca and La Rioja (*An. algeriensis* A) were clearly separated (bootstrap > 98%) from those sequences retrieved from Sweden and Romania (*An. algeriensis* B). These groups were separated by a 7% divergence (K2P). In addition all specimens from Majorca initially identified as *Cs. annulata* by field staff were placed in the *Cs. subochrea* clade (Figure 3). This resulted into a re-examination of the morphology of these specimens (see below). The remaining species *Aedes albopictus*, *Culiseta longiareolata*, *Culex theileri*, *Ochlerotatus detritus* and *Culex pipiens* formed highly supported monophyletic clades, although the last two showed some level of differentiation (Figure 2). Even when a single specimen of species *Uranotaenia unguiculata* and *Coquillettidia richiardii* were obtained, these were placed in a well-supported clade with those sequences from the same species retrieved from GenBank.

3.3. Species delimitation analysis

Using standard settings initial partition of the ABGD analysis resulted in 12 molecularly defined operational taxonomic units (MOTUs) for our 406 fragment of the *Cytochrome Oxidase I* gene region. This first result clustered *Oc. caspius* and all members of the *Aedes mariae* complex (*Ae. mariae*, *Ae. zammitii* and *Ae. phoeniciae*) in a single MOTU and also failed to separate *Cs. annulata* from *Cs. subochrea*, while the remaining MOTUs corresponded to morphologically-identified species. On the other hand, this conservative approach split *An. algeriensis* in two species each. We also selected a recursive partition with a P value=4.64e-03, since this was the first to separate the three species of the Mariae Complex. However, when these were divided, the cluster containing *Ae. mariae* specimens from Majorca were also separated as a distinctive species (Figure 2). No significant differences were observed when using Jukes Cantor or Kimura K80. The bPTP analysis based on the Bayesian phylogenetic

tree identified 23 MOTUs. The results of bPTP are mostly consistent with the ABGDr analysis; however, it is not as accurate as it separates some individuals as different MOTUs.

The last approach to separate species using the *COI* gene region was the genealogical haplotype network analysis using the statistical parsimony method implemented in TCS (Clement *et al.*, 2000). This method split our dataset in 19 networks separated by 11 mutational steps. The three species from the Mariae Complex were separated as different groups; however, *Oc. caspius* was included in the same network than *Ae. mariae*. The remaining specimens showed a clear agreement among haplotype networks and the morphological identification (Figure 3). There were clear disagreements in the number of species identified by each approach. However, most algorithms agreed in the separation of the MOTUs detailed in Figure 3. The Kimura 2-parameter showed that the majority of MOTUs identified by the species boundaries approaches could be identified using Hebert's 10 X threshold (Table 3), with the exception of the members of the Mariae Complex and *Oc. caspius* from Majorca.

3.4. Agreement between molecular and morphologically-based identification

Using morphology-based keys we identified 13 mosquito species that belong to 7 different genera. Morphological identification using available keys were efficient to distinguish 11 of the species, with the exception of *Cs. annulata* and *Cs. subochrea*. Since the molecular identification resulted in conflictive arrangement of specimens, we re-examined the morphology of these and confirmed that all Balearic specimens showed morphological characters in the wings similar to *Cs. annulata* ("Cross veins r-m and m-cu aligned"), but abdominal scales like *Cs. subochrea* "Terga with indistinct pale basal bands formed by yellowish scales (not white), pale scales are also present among the dark scales in the apical half of the terga". A somehow different situation was observed in La Rioja, with most specimens falling into the key's description but some of them (*Cs.a_93LR*) with wings like *Cu. annulata* and terga like *Cs. subochrea* (Fig. 2). When characters located in the wings were ignored, the morphological identification was in agreement with the DNA-barcoding analysis.

4. Discussion

Establishing a genetic distance value in *COI* for species delimitation is a controversial topic that has been largely under scrutiny (Janzen *et al.*, 2017). Although a general value of 3% split for invertebrates species has been proposed, for mosquito identification the Barcoding

community generally agrees with a Kimura-2-parameter value higher than 0.02 (>2%) to separate species (Kumar *et al.*, 2007; Ruiz-Lopez *et al.*, 2012). A low genetic divergence value (2,5%) has been used differentiate sibling species with different vectorial importance such as *Anopheles nuneztovari* from *An. dunhami*, a less important vector, in Colombia (Ruiz *et al.*, 2010). Closely related species within the *Anopheles albitarsis* group could also be differentiated using this value (Ruiz-Lopez *et al.*, 2012). Although this value may seem insufficient, a large-scale barcoding study conducted in butterflies, combining genomic and ecological data showed that a difference as shallow as 1 to 2% in the *COI* studied could be enough to differentiate species (Janzen *et al.*, 2017).

Using Bayesian and distance-based phylogenetic approaches and a combination of four additional species delimitation algorithms, we confirmed the identity of 11 out of 13 morphologically-identified mosquitoes (85%), demonstrating that DNA barcoding is an effective strategy to identify mosquitoes in Spain. These results are consistent with reports in other regions of the world, where congruence between these approaches ranged from 82% in Colombia (Rozo-Lopez & Mengual, 2015) up to more than 98% in most countries (Kumar *et al.*, 2007; Wang *et al.*, 2012; Ashfaq *et al.*, 2014; Chan *et al.*, 2014; Versteirt *et al.*, 2015). These differences may be related to the limitation of this technique in some taxa (M. Laurito *et al.*, 2017) or to errors in the initial morphological identification leading to incorrectly naming the sequences in the public databases used for sequence comparisons (Krzywinski & Besansky, 2003). We found disagreement between the barcoding and morphological approaches to separate *Cs. annulata* from *Cs. subochrea*. Furthermore, the sequence analysis also showed preliminary evidence of cryptic speciation in *An. algeriensis*, *Ae. vexans* and in *Ae. mariaae* s.s. It is important to emphasize that the results obtained in this study are not definitive, as this only included one molecular marker. An inflation of the number of species due to the sole use of molecular methods is a risk to take into consideration (Harbach, 2004). Therefore, a combination of several markers is compulsory in order to arrive to firm conclusions on these potential cryptic species. Nevertheless, the DNA-barcoding approach has often served as a first step to identify cryptic species within Culicidae (Ruiz-Lopez *et al.*, 2012) and we hope that this will open a new window of research in this field in Spain, where identifications rely almost entirely in morphology. Results are discussed in depth below.

4.1. Preliminary evidence for cryptic speciation

All species delimitation approaches, including Hebert's 10 x thresholds, agreed in the existence of at least two members within *An. algeriensis*. This is formed by two highly supported monophyletic clades separated by a high genetic divergence (K-2P = 6%). One of them, herein putative *An. algeriensis* A would be formed only by the Spanish specimens from Navarra and Majorca, while *An. algeriensis* B would be formed by specimens from Romania and Sweden

and probably others not included in this study. *Anopheles algeriensis* is widely distributed in the Palearctic region, with Yemen as the easternmost limit (Al-Eryani *et al.*, 2016). This species is scarce in most areas where it has been reported (Krüger & Tannich, 2013), including the Balearic Islands. This scarcity led to the conclusion that *An. algeriensis* is a malaria vector of secondary importance (Becker *et al.*, 2010). However, this would not be the case of La Rioja, where it has been recorded as the third most abundant species (23,3% of all collected mosquitoes) in the wetland of La Grajera (Ruiz-Arrondo *et al.*, 2017). Furthermore, it has recently been found naturally infected with *Plasmodium vivax* in Yemen (Al-Eryani *et al.*, 2016), which should open new questions on their role in malaria transmission in endemic areas. In Spain, this species has been reported showing a “confusing” behavior, with variations in its host and resting preferences (R Bueno Marí *et al.*, 2011). Different patterns in mosquito behavior have been the first evidence of the possibility of cryptic speciation in other *Anopheles* mosquitoes (Paredes-Esquivel, Donnelly, *et al.*, 2009) and may be related to our findings. However these would not be surprising considering that half of the *Anopheles* species are part of species complexes (Harbach, 2004).

The use of nuclear and mitochondrial markers has recently allowed the detection of two different lineages within *Ae. vexans* (Lilja *et al.*, 2018). Our results are consistent with these findings, with all species delimitation analysis confirming the separation of at least two groups of *Ae. vexans* in the dataset. Mosquitoes collected from Majorca (*Ae. vexans* A) would belong to the authors’ Group 2, which has also been reported in Hungary and Sweden (Lilja *et al.*, 2018). *Ae. vexans* is a competent vector to more than 30 arboviruses, including West Nile virus, Zika virus and the dog heartworm *Dirofilaria immitis* (Gendernalik *et al.*, 2017). In Spain it has been reported in 22 provinces, including the Balearic Islands (R. Bueno Marí *et al.*, 2012). *Ae. vexans* is a floodwater mosquito, with highly resistant eggs which can remain viable for several years. This species can travel long distances, which has direct relationship with its invasive capacity.

4.2. Incipient speciation in the Balearic rock pool mosquitoes?

The sea water contained in the rock pools of the Mediterranean coasts is habitat for mosquitoes of the *Aedes mariae* complex, which include three sibling species *Ae. mariae* s.s., *Ae. phoeniciae* and *Ae. zammitii* (Coluzzi & Sabatini, 1968). The phylogenetic analysis shows that *Ae. mariae* is not reciprocally monophyletic, with the Majorcan clade separated by >2% from the other members of the *Mariae* Complex, including the *Ae. mariae* specimens from mainland Spain and Italy. The Majorcan clade is also 2% distant to *Oc. caspius* and it appears more closely related to this species in the phylogenetic tree. However, the low bootstrap values obtained will not allow us to arrive to further conclusions. Contradictory results were obtained with the 10x threshold proposed by Hebert (2004) and the initial partition in ABGD. These

approaches not only failed in separating the Majorcan clade, but also placed all members of the Mariae Complex plus *Oc. caspius* as a single species. These results are not consistent with the current knowledge of these species. The Mariae Complex has been subjected to a broad analysis, which included the morphological, genetic and reproduction studies of this group (Coluzzi & Sabatini, 1968; Mastrantonio *et al.*, 2016). There is no discussion on the species status of its members. However, it is known that they have a parapatric distribution with incomplete reproductive isolation (Mastrantonio *et al.*, 2016). Furthermore, including *Oc. caspius* and the Mariae Complex members in the same MOTU does not seem reasonable if we consider that these have a very distinctive ecology. While the former is a salt-marsh mosquito, the latter are rock pool mosquitoes with an extreme capacity to tolerate high salinity values. Further analysis using additional markers and running some experimental crosses in the laboratory would be necessary to establish the species status of the Majorcan *Ae. mariae* and its role as disease vector.

4.3. Current morphological keys cannot always distinguish *Culiseta annulata* from *Cs. subochrea* in Spain

Species *Cs. annulata* and *Cs. subochrea* are morphologically similar and can be differentiated by the presence of white (*Cs. annulata*) or yellow (*Cs. subochrea*) basal bands and the absence or presence of pale scales in the apical half of terga, respectively. In addition the alignment of veins r-m and m-c is effective to distinguish both species, being totally aligned in *Cs. annulata* and slightly separated in the case of *Cs. subochrea* (Becker *et al.*, 2010). The phylogenetic analysis and the species delimitation methods agreed in the clustering of two distinctive groups. As we obtained conflictive results with members of both species placed indistinctively in both clades (Figure 3), we re-checked the morphology and found that the character in the wings could not always separate both species, while the “the absence or presence of pale scales in the apical half of terga” was in clear agreement with DNA-barcoding identification. Conspicuous characters are frequently used by field workers as a careful examination of specimens can be time consuming. This may lead to erroneous delimitation of the distribution of these vectors. It is crucial to include other molecular markers to confirm these findings and re-design the morphological keys for these species in Spain. *Culiseta annulata* is a competent vector of the Tahyna virus vector (Bardos *et al.*, 1975) while we have found no reports on the vectorial capacity of *Cs. subochrea*. Therefore, its correct identification is crucial in case of outbreaks in the country. This would be the first record of *Cs. subochrea* in the Balearic Islands.

4.4. Disagreement among species delimitation methods to separate Spanish mosquitoes

We found disagreement among all species delimitation approaches used in this study. Methods such as ABGD and TCS allow the user to set the initial parameters, then results may be considered subjective. According to Puillandre *et al.* (2012) a P value = 0,01 is the closest to the species defined by authors when testing ABGD in four datasets. However, this value did not allow us to separate members of the *Aedes mariaae* complex and *Oc. caspius* that appeared together in the same MOTU. On the other hand, bPTP, ABGD using a P value=4.64e-03 and the combination of the Phylogenetic Species Concept plus the 2% K2P approach were effective methods to separate these species. The fact that ABGD at initial partition and TCS were not able to separate members of the Mariae Complex, may be related to the incipient speciation of the group, which were more easily detected by the remaining species delimitation approaches. largely studied groups such as those from the Mariae complex is important as they may also be used as a standardization strategy to determine which species delimitation method is more appropriate for our dataset. In the case of *Oc. caspius* this would require a more careful analysis as this species seem close to *Ae. mariaae* even when these vary in their habitats. Hybridization between these species cannot be ruled out.

A multilocus approach and a morphological revision is crucial to confirm the identification of these and other putative cryptic species and to confirm the validity of the morphological characters in *Culiseta annulata* and *Cs. subochrea*. As the identification approaches may likely disagree on the species composition, data on the ecological distribution the species should be incorporated as this may help clarify the species status as with *Oc. caspius*, which lives in salt-marsh mosquitoes, differing from the *Aedes mariaae* complex species, which live in the rock pools of the Mediterranean coast.

4.5. Implications on disease transmission

Nine of the eleven morphologically-identified species have been recorded as vectors of human diseases. The remaining two are *Culiseta longiareolata* which rarely feeds on humans but it is considered a vector of veterinary importance and *Uranotaenia unguiculata* an autogenous species with no importance as a disease vector (Becker *et al.*, 2010). During the last years, there has been an increased awareness of mosquitoes and the diseases they transmit in Europe. Changes in climatic conditions together with globalization may result in active transmission of pathogens as it has already occurred in some areas. Their relative importance of such vectors depends on their feeding behavior, longevity, and other aspects that may vary in species complexes (Gendernalik *et al.*, 2017). Therefore, it is crucial to know which

vector species are currently circulating in the continent. Identifying lineages within mosquito species has important epidemiological implications. For instance, the geographic limits of the lineages of the species *Culex annulirostris* has been associated to the transmission of Japanese Encephalitis Virus in Australasia (Hemmerter *et al.*, 2007). Cryptic speciation in mosquitoes is an underexplored topic in Europe and this study demonstrates that it may be present in many taxa. We hope this study will open the door to new studies in this topic.

Acknowledgements

This work was funded by various sources: Mosquito collection was supported by the *Conselleria de Medi Ambient Agricultura y Pesca* of the *Govern of the Illes Balears*, DNA sequencing was possible thanks to *Ajuts a la recerca* for unfunded projects from the University of the Balearic Islands. We would also like to thank the *SOIB JOVE* program for funding the contract of the first author of the study and Mr. David Borràs for providing mosquito samples.

<i>Haplotype</i>	<i>Sample ID</i>	<i>Specie</i>	<i>Sampling site</i>	<i>Genbank Accesion</i>
H1	Ae.m_54M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H1	Ae.m_83M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H1	Ae.m_84M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370032
H2	Ae.m_37M	<i>Aedes mariaae</i>	Majorca (Spain)	MH370031
H3	Ae.m_25M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H3	Ae.m_55M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266

H3	Ae.m_79M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H3	Ae.m_82M	<i>Aedes mariaae</i>	Majorca (Spain)	MH348266
H4	Oc.c_KM243949	<i>Ochlerotatus caspius</i>	Germany	KM243949
H4	Oc.c_85M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH370035
H5	Oc.c_4M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH370036
H6	Oc.c_86M	<i>Ochlerotatus caspius</i>	Majorca (Spain)	MH348267
H7	Ae.m_76V	<i>Aedes mariaae</i>	Castellón (Spain)	MH348265
H8	Ae.m_77V	<i>Aedes mariaae</i>	Castellón (Spain)	MH370033
H9	Ae.m_78V	<i>Aedes mariaae</i>	Castellón (Spain)	MH370034
H10	Ae.m_KM592039	<i>Aedes mariaae</i>	Scilla (Italy)	KM592039
H11	Ae.z_KM592049	<i>Aedes zammitii</i>	Crete (Greece)	KM592049
H12	Ae.p_KM592051	<i>Aedes phoeniciae</i>	Cyprus (Greece)	KM592051
H13	Oc.d_KM258326	<i>Ochlerotatus detritus</i>	Westkapelle (Belgium)	KM258326
H14	Oc.d_KU877017	<i>Ochlerotatus detritus</i>	Hampshire (U.K)	KU877017
H14	Oc.d_29M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH348269
H14	Oc.d_42M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH348269
H15	Oc.d_2M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH370048
H16	Oc.d_23M	<i>Ochlerotatus detritus</i>	Majorca (Spain)	MH370049
H17	Cs.a_KX675395	<i>Culiseta annulata</i>	Bonn (Germany)	KX675395
H17	Cs.a_KU877021	<i>Culiseta annulata</i>	Surrey (U.K)	KU877021
H17	Cs.a_19LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370038
H17	Cs.a_20LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370038
H18	Cs.a_18LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370039
H19	Cs.a_52LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370042
H20	Cs.a_21LR	<i>Culiseta annulata</i>	La Rioja (Spain)	MH370041
H20	Cs.s/a_093LR	<i>Culiseta subochrea/annulata</i>	La Rioja (Spain)	MH370041
H20	Cs.s/a_93LR	<i>Culiseta subochrea/annulata</i>	La Rioja (Spain)	MH370040
H21	Cs.s/a_5M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370043
H21	Cs.s_36LR	<i>Culiseta subochrea</i>	La Rioja (Spain)	MH370044
H21	Cs.s_42N	<i>Culiseta subochrea</i>	Navarra (Spain)	MH370044
H21	Cs.s_43N	<i>Culiseta subochrea</i>	Navarra (Spain)	MH370044
H22	Cs.s_37LR	<i>Culiseta subochrea</i>	La Rioja (Spain)	MH370047
H23	Cs.s/a_30M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370045
H23	Cs.s/a_58M	<i>Culiseta subochrea/annulata</i>	Majorca (Spain)	MH370046
H24	Cs.l_HG931139	<i>Culiseta longiareolata</i>	Kabul (Afghanistan)	HG931139
H24	Cs.l_JQ388785	<i>Culiseta longiareolata</i>	Freiburg (Germany)	JQ388785
H24	Cs.l_7M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H24	Cs.l_27M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H24	Cs.l_44M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH348268
H25	Cs.l_31M	<i>Culiseta longiareolata</i>	Majorca (Spain)	MH370037
H26	Cx.t_KJ012234	<i>Culex theileri</i>	Kars Ili (Turkey)	KJ012234
H27	Cx.t_6M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H27	Cx.t_19M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H27	Cx.t_28M	<i>Culex theileri</i>	Majorca (Spain)	MH348264
H28	Cx.t_36M	<i>Culex theileri</i>	Majorca (Spain)	MH370030
H29	Cx.p_10M	<i>Culex pipiens</i>	Majorca (Spain)	MH370028

H30	Cx.p_KF919189	<i>Culex pipiens</i>	Cordoba (Argentina)	KF919189
H31	Cx.p_KP293425	<i>Culex pipiens</i>	Stinson Beach (EEUU)	KP293425
H31	Cx.q_KF919190	<i>Culex quinquefasciatus</i>	Cordoba (Argentina)	KF919190
H31	Cx.p_18M	<i>Culex pipiens</i>	Majorca (Spain)	MH348262
H31	Cx.p_67M	<i>Culex pipiens</i>	Majorca (Spain)	MH348262
H32	Cx.p_11M	<i>Culex pipiens</i>	Majorca (Spain)	MH370029
H33	Ur.u_KM280589	<i>Uranotaenia unguiculata</i>	Austria	KM280589
H34	Ur.u_13M	<i>Uranotaenia unguiculata</i>	Majorca (Spain)	MH348263
H35	Cq.r_JX040513	<i>Coquillettidia richiardii</i>	Vasterbotten (Sweden)	JX040513
H36	Cq.r_91M	<i>Coquillettidia richiardii</i>	Majorca (Spain)	MH348270
H37	Ae.v_KM452935	<i>Aedes vexans</i>	Szeged (Hungary)	KM452935
H38	Ae.v_12M	<i>Aedes vexans</i>	Majorca (Spain)	MH348259
H39	Ae.v_KY609213	<i>Aedes vexans</i>	Kävlinge (Sweden)	KY609213
H39	Ae.v_KY609201	<i>Aedes vexans</i>	Kävlinge (Sweden)	KY609201
H39	Ae.v_21M	<i>Aedes vexans</i>	Majorca (Spain)	MH370023
H40	Ae.v_KP942689	<i>Aedes vexans</i>	Norråla (Sweden)	KP942689
H41	Ae.v_KP942682	<i>Aedes vexans</i>	Deherhamn (Sweden)	KP942682
H42	Ae.a_KU738431	<i>Aedes albopictus</i>	China	KU738431
H42	Ae.a_16M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_17M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_43M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H42	Ae.a_70M	<i>Aedes albopictus</i>	Majorca (Spain)	MH348258
H43	An.a_15N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370027
H44	An.a_11N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370025
H44	An.a_18N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370025
H45	An.a_3M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH370024
H46	An.a_12N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH348261
H46	An.a_14N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH348261
H47	An.a_8M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260
H47	An.a_14M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260
H47	An.a_15M	<i>Anopheles algeriensis</i>	Majorca (Spain)	MH348260
H48	An.a_13N	<i>Anopheles algeriensis</i>	Navarra (Spain)	MH370026
H49	An.a_KU214675	<i>Anopheles algeriensis</i>	Romania	KU214675
H50	An.a_KP942711	<i>Anopheles algeriensis</i>	Gotland (Sweden)	KP942711
H51	Outgroup_JF923694	<i>Anopheles darlingi</i>	Brazil	JF923694