Genomic, chromosomal, and ecological differentiation between the cryptic Eurasian malaria vector-species Anopheles messeae and Anopheles daciae

Anastasia N. Naumenko
Virginia Polytechnic and State University
Dmitriy A. Karagodin
The Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science
Andrey A. Yurchenko
Virginia Polytechnic and State University
Anton V. Moskaev
Moscow Regional State University
Olga I. Martin
Virginia Polytechnic and State University
Elina M. Baricheva
The Federal Reserch Center Institute of Cytology and Genetics, Siberian Branch of the Russia Academy of Science
Igor V. Sharakhov
Virginia Polytechnic and State University
Mikhail I. Gordeev
Moscow Regional State University
Maria Sharakhova (msharakh@vt.edu)
Virginia Polytechnic Institute and State University https://orcid.org/0000-0002-5790-3548

Research article

Keywords: Malaria mosquitoes, internal transcribed spacer 2, genomics, polymorphic inversions

Posted Date: July 1st, 2019

DOI: https://doi.org/10.21203/rs.2.10780/v1

Abstract

Background A dominant malaria vector, Anopheles messeae, is a highly polymorphic species with wide distribution throughout Eurasia. Five highly polymorphic inversions associated with the geographical distribution of the species have been reported. A sister species, An. daciae, was described and discriminated from An. messeae based on five fixed nucleotide substitutions in the internal transcribed spacer 2 (ITS2) of ribosomal DNA. However, the levels of genomic divergence, chromosomal variation, and ecological differentiation between these two cryptic species remain unexplored. Results In this study, we sequenced ITS2 and analyzed the inversion frequencies of 289 Anopheles larvae specimens collected from three locations in the Moscow region. We identified a high abundance of both An. messeae and An. daciae in all three locations. Five individual genomes for each species of An. messeae and An. daciae from one location were sequenced. Our study confirmed five previously described nucleotide substitutions in the ITS2 of An. messeae. However, we found that the ITS2 sequence in An. daciae was heterogenic in three of the five positions. Fixed nucleotide differences between An. messeae and An. daciae were found only in the last two positions. One mosquito was identified as a hybrid between An. messeae and An. daciae based on heterogeneous substitutions in all five positions. Although, the genomic sequence comparison demonstrated genome-wide divergence between the two species, which is especially pronounced on the X chromosome, an ADMIXTURE cluster analysis demonstrated the presence of two admixed individuals suggesting ongoing hybridization. Cytogenetic analysis demonstrated that An. messeae and An. daciae significantly differ from each other by their frequency of polymorphic inversions. Inversion X1 was fixed in An. messeae but was polymorphic in all An. daciae populations. The frequency of polymorphic autosomal inversions was higher in An. messeae than in An. daciae. The species composition was different among the studied locations suggesting species-specific ecological preferences. Conclusions Our study demonstrated that An. messeae and An. daciae represent closely related cryptic species with incomplete reproductive isolation that are able to maintain genomic differentiation in sympatry despite ongoing genetic introgression. The X chromosome plays an important role in the reproductive isolation between the species.

Background

The malaria mosquito *An. messeae* Falleroni was originally described as a subspecies of *An. maculipennis* within the *Anopheles maculipennis* complex [1]. This species was the primary malaria vector in Russia [2] and now represents a threat to malaria re-emergence in the Northern territories of Eurasia because of global climate change [3, 4]. *An. messeae* is susceptible to *Plasmodium vivax* but not to the tropical *P. falciparum* malaria parasite [5]. According to current systematics, the Maculipennis group of malaria mosquitoes comprises eleven Palearctic species: *An. artemievi, An. atroparvus*, *An. beklemishevi*, *An. daciae*, *An. labranchiae*, *An. maculipennis*, *An. martinius*, *An. melanoon*, *An. messeae*, *An. persiensis, and An. sacharovi [6]. Species in this group have traditionally been recognized based on the chorion patterns of their eggs [7, 8] and the banding patterns of their polytene chromosomes [9-13]. Three species – An. artemievi* [14], *An. persiensis* [15], and *An. daciae* [16] – were

identified more recently using nucleotide sequence substitutions in the Internal Transcribed Spacer 2 (ITS2) of the ribosomal DNA (rDNA). *An. daciae*,collected near the Danube river in Romania, was discriminated from *An. messeae* based on five fixed nucleotide substitutions in ITS2 and by egg morphology; the eggs of *An. daciae* are generally darker and smaller and have tubercles that are organized in patches of a slightly different shape [16]. Using the molecular approach, *An. daciae* was later discovered in Germany [17-19], England and Wales [20], and more recently in Poland [21], the Czech republic, Slovakia [22], and Serbia [23]. In most locations*, An. daciae* was found to be in sympatry with its cryptic species *An. messeae*. However, the species status of *An. daciae* was questioned by the study conducted in Russia [24] where *An. messeae* isthe most widespread malaria vector [2]. Sequencing of ITS2 in 22 individual mosquitoes from 11 distant populations in Russia enabled the discovery of a high level of intra-species and intra-individual polymorphism in nucleotide substitution, suggesting the simultaneous presence of both *An. messeae* and *An. daciae* ITS2 variants in each individual [24]. Thus, the taxonomic status of *An. daciae* in Russia is still under debate and requires further validation by genomic, molecular, and cytogenetic approaches.

Among other Palearctic members of the Maculipennis group, *An. messeae* is one of the most geographically widespread [5] and genetically diverse [13] species of malaria mosquitoes in Eurasia. Distribution of the species extends from Ireland in the West to the Amur river region in the East and from Scandinavia and Yakutia in the North to Iran and Northern China in the South [25]. Five highly polymorphic chromosomal inversions have been described in this species: X1, X2, and X4 on chromosome X; 2R1 on chromosome 2; 3R1 and 3L1 on different arms of chromosome 3 [26]. A longitude cline was described for the 2R1 inversion where the inverted variant was more abundant in northern populations, suggesting that this inversion could be involved in adaptation to cold temperatures and success in overwintering [27]. The frequencies of the X1 and 3R1 inversions displayed a West-East latitude cline with higher frequencies of the inverted variants found in Eastern populations. The X2 inversion is endemic in Western Siberian populations [26]. Comparison of the inversion polymorphisms among populations also suggested that area-specific biological or behavioral adaptations likely have occurred [28, 29]. Inversion frequencies have been shown to be variable in different water reservoirs located next to each other, suggesting involvement of the inversions in local adaptations at the subpopulation level [30]. Inversion frequencies also varied during the summer period with standard variants more abundant in the middle of the summer [31]. Although, inversion frequencies did not significantly change during the 10-year period from 1972 to 1982 [32], a more recent study conducted over a 40-year period from 1974 to 2014 demonstrated a significant increase in the standard inversion arrangements and a decrease of the 2R1 variant in populations across the *An. messeae* range in Russia; these changes correlate with the increase in year-round temperatures [33].

Y. Novikov and V. Kabanova introduced the idea that combinations of inversions in natural populations of *An. messeae* represent two distinct chromosomal complexes [34]. The variant X0 is associated with 2R0, 3R0, and 3L0, and, alternatively, inverted X1 and X2 tends to associate with 2R1, 3R1, and 3L1 [34-36]. These chromosomal complexes confer differences in female fecundity, viability of imago and larvae, food preferences, rate of development, relationship with predators and parasites, and sensitivity to the

toxins of *Bacillus thuringiensis subsp. israelensis* [29, 35, 37-41]. Later, Y. Novikov referred to these chromosomal complexes as cryptic genetically isolated forms, named “A” and “B” [42]. However, according to this study, these two cryptic forms have overlapping inversion polymorphisms and cannot be distinguished by any fixed inversion differences. Recent ITS2 sequencing studies demonstrated that form “A” of *An. messeae* is synonymous with *An. daciae* [43, 44].

In this study, we characterized molecular and chromosomal differentiation of *An. messeae* and *An. daciae* in three populations from the Moscow region (Novokosino, Noginsk and Yegoryevsk) in Russia based on ITS2 sequencing and karyotyping of 289 individual mosquitoes. Additionally, whole-genome sequencing of five specimens from each species from one population was also performed to characterize genomic diversity and divergence.

Results

Molecular structure of ITS2 in *An. messeae* and *An. daciae* from the Moscow region

As was reported earlier, *An. messeae* and *An. daciae* differ by five nucleotide substitutions in the ITS2 region of their rDNA sequence [16]. These ITS2 sequences contained the following nucleotides in *An. messeae*: T in position 211, T in position 215, C in position 217, G in position 412, and G in position 432, whereas *An. daciae* had A, A, T, A, and C in the same positions, respectively [16]. Here and below, the numbers of the nucleotide positions are given in correspondence to the original sequence submitted under GenBank number AY648982 [16]. We found that ITS2 sequences in 160 mosquitoes of our specimens corresponded to the previously described *An. messeae* sequence in all five nucleotide positions. We found the same arrangement of T, T, C, G, and G in positions 211, 215, 217, 412, and 432 in all *An. messeae* samples from the three Moscow region populations (Fig. 1A, Table 1). Contrary to the previous observation, 119 specimens that we identified as *An. daciae* had ITS2 sequences with heterogeneous substitutions in the first three positions. We found the following double peaks in sequence chromatograms: A+T (W), A+T (W) and T+C (Y) in positions 211, 215, and 217 and single picks A and C in positions 412 and 432 (Fig. 1B, Table 1). Thus, in all our specimens, we found a mix of two or more ITS2 variants, possibly containing *An. messeae* (TTC) and *An. daciae* (AAT) nucleotides in the first three positions but not in the last two positions. In two cases only, the original *An. daciae* nucleotide T was found in the third position (position 217). Similar sequencing results were obtained for *An. messeae* and *An. daciae* from samples in the Novosibirsk region of Russia [44].

In addition to the originally described five substitutions [16], we found another nucleotide in position 150 of the ITS2 sequence that can be used for speciesdiagnostics (Fig. 1, Table 1). In all our *An. messeae* specimens, we found a double peak A+C (M) in this position, whereas, in the previous report, only nucleotide C was found in this position in *An. messeae* [16]. In all our *An. daciae* specimens, we determined the C nucleotide in position 150 with only two exclusions where a double peak A+C (M) were observed. It should be noted that the nucleotide in position 150 can be used to identify *An. messeae*

using Restriction Fragment Length Polymorphism (RFLP) analysis, since *An. messeae* present the unique AATT site in one of the ITS2 forms, which can be cleaved by Sse9 I. Based on the sequencing data from 280 samples, we were able to identify only one individual mosquito that we considered a hybrid between *An. daciae* and *An. messeae,* where double peaks were presented in all six diagnostic positions (Fig. 1C). Thus, we conclude that unique substitutions in positions 412 and 432 – G, G for *An. messeae* and A, C for *An. daciae* – are the most reliable for the species and hybrid diagnostics.

In addition to the differences in the 6 mentioned positions above, we also observed overlapping double peaks in some samples in other positions of the ITS2 sequences in both *An. messeae* and *An. daciae*. However, the intensity of the minor peak was usually lower than the intensity of the major peak. It is likely that these replacements also reflect heterogeneity in the ITS2 sequences that are not species-specific.

Genome sequence analysis of *An. messeae* and *An. daciae*

To compare genomic differentiation between the species, we sequenced 10 individual mosquito genomes from the Yegoryevsk population: five individuals each of *An. messeae* and *An. daciae*. For each sequenced library, 88.24% to 90.69% of the reads were aligned to the *An. atroparvus* genome as a reference [45]. The average coverage of the sequenced genomes was 20X and 1,858,004 high-quality Single Nucleotide Variants (SNVs) were identified in total after the SNV calling and the filtering step. Analysis of the fixation index (Fst) distribution along the 5 kb genomic windows revealed small or modest genetic differentiation between the species on autosomes (mean Fst = 0.027), and dramatic increases in the level of genetic differentiation on the X chromosome (mean Fst = 0.331, Fig. 2). The large increase in genomic differentiation is seen in the telomeres, centromeres, and in the middle part of the X chromosome. The increase in genomic differentiation in the middle part of the chromosomes likely overlaps with the location of the highly polymorphic inversion X1 and is caused by the low recommendation inside the inversions. Some increases in differentiation between the species was also observed in the centromeres of the autosomes, which are regions of low recombination. Principal Component Analysis (PCA) based on autosomal polymorphisms reliably separated the two species by the PC1 (Fig. 3A), accounting for 14.69% of the total variance. The ADMIXTURE analysis allowed us to separate individuals into two clusters (K=2) based on their species status and on autosomal SNVs, but it also indicated a considerable level of admixture between the species (Fig. 3B). Two out of 10 individual mosquitoes were admixed and may represent second or third generation hybrids between the two species. This result suggests that hybridization between the two species is still ongoing in the population although a larger sample size is needed for a precise estimation of the current gene flow amount.

Chromosomal inversions in *An. messeae* and *An. daciae*

Five inversions in four chromosomal arms, X1, X4, 2R1, 3R1, and 3L1, have been described in the Moscow region [14, 46, 47]. The most common inversions in *An. messeae* and *An. daciae* polytene chromosomes are shown in Fig. 4. Inversion X4 is rare and was only found in the Moscow region at a low frequency. The

other inversions are common in the populations of *An. messeae* across the range of its distribution [13]. In our study, each genotyped sample was assigned to a corresponding karyotype (Supplementary Table 1). According to the common nomenclature of inversions in *An. messeae,* standard arrangements are referred to as 00 variants, inverted as 11 variants, and heterozygotes as 01 or 14 variants. Inversions on the X chromosome in hemizygous males are referred as 0, 1, and 4 variants.

The frequencies of the chromosomal inversion variants were calculated for both *An. messeae* and *An. daciae* in three populations (Fig. 5). Our analysis demonstrated that frequencies of all chromosomal inversions were significantly different between *An. messeae* and *An. daciae*. The most dramatic difference was observed in the X sex chromosome (Fig. 5A). Inversion X1 was fixed in *An. messeae.* We found no standard arrangement X0 in all three populations of *An. messeae*. In contrast, this inversion was highly polymorphic in all three *An. daciae* populations with almost equal frequencies of standard and inverted variants. Moreover, the endemic inversion X4 was found only in *An. messeae* populations in all three locations, although in low frequencies. The X4 arrangement was observed only in X14 heterozygote females and X4 hemizygote males. Although the inverted autosomal variants 2R1 and 3R1 were present in *An. daciae*, we found significantly higher frequencies of the inverted variants in *An. messeae* populations (Fig. 5B-C). Moreover, inverted variants of the 3L1 inversion were only found in *An. messeae* in low frequency (Fig. 5D). In addition to the previously described chromosomal inversions, we discovered a new inversion, 2R4, that was found as the heterozygote 2R04 in *An. daciae* from the Yegoryevsk population (Fig. 4B, 5B). Interestingly, a hybrid between *An. daciae* and *An. messeae* carried an unusual karyotype of an X04 heterozygote of the standard variant X0, which is typical for *An. daciae* and the inverted variantX4, which was only observed in *An. messeae* (ME21 sample in Supplementary Table S1). Overall, our study demonstrated that inversion frequencies were higher on the X chromosome than on the autosomes in *An. daciae*, whereas inversion frequencies were higher on autosomes in *An. messeae*.

Our statistical analysis demonstrated that the frequencies of homozygotes and heterozygotes within *An. messeae* and *An. daciae* samples were in Hardy-Weinberg equilibrium for autosomal inversions but deviated from Hardy-Winberg equilibrium for sex-linked X-chromosomal inversions in the Noginsk population of *An. daciae* and in the Yegoryevsk population of both species (Table 2). Pairwise analysis of population differentiation revealed no significant population differentiation within the species (Fst= 0.003 – 0.076 within *An. messeae*, Fst = -0.017 – 0.01 within *An. daciae*, nonsignificant,Table 3), but highly significant Fst values between the two studied species. The interspecies Fstvalues ranged from 0.076 to 0.28 and were highly significant for a majority of the pairwise comparisons after the Bonferroni correction for multiple testing was applied. The PCA based on the frequency of autosomal inversions only reliably separated the two species by the PC1, accounting for 96.2% of the total variance (Fig. 6).

Species compositions in Moscow populations

We collected the larvae of *Anopheles* mosquitoes from three locations in the Moscow region (Fig. 7). The Novokosino larval breeding site was located within the border of the city of Moscow, the Noginsk breeding site was located 36 km east of Moscow, and the Yegoryevsk breeding site was located in the village of Parykino, which is located 99 km to the southeast of Moscow. All three populations were located near human settlements in the same landscape zone of the Meshcherskaya Lowland territory. The ponds in Novokosino and Noginsk represented typical breeding sites of *Anopheles* mosquitoes (Fig. 8A, B). The open water surface was exposed to the sun and the abundance of floating vegetation created favorable conditions for the development of mosquito larvae. On the contrary, the larval pond in Yegoryevsk was shaded by trees and contaminated with tree litter (Fig. 8C). The three locations differed in the predominant aquatic and near-water plants (see the Materials and Methods section).

Species composition in the three locations was determined by sequencing ITS2. We found a higher proportion of *An. messeae* in Novokosino where the percentages of *An. messeae* to *An. daciae* were 73 % to 26 %, respectively, and in Noginsk where the percentages of *An. messeae* to *An. daciae* were 66.3 % and 32.6 %, respectively (Fig. 7). A different species composition pattern was found in Yegorevsk, with a ratio of 25.5 % to 66 % of *An. messeae* to *An. daciae*, respectively. One hybrid mosquito (1%) was found in Yegorevsk. In addition, four (4.3 %) *An. beklemishevi* and three (3.2 %) *An. maculipennis* larvae were identified based on karyotyping [9, 48] and ITS2 PCR product length [49] in Yegorevsk and one (1.1 %) *An. maculipennis* in Noginsk and Novokosino each.

Thus, our study demonstrated that *An. messeae* dominated in the typical malaria mosquito open, sunny breeding sites of Novokosino and Noginsk, with an increased oxygen content of 1.8–4.0 mg/L (Fig. 8A, B). The density of larvae in these breeding sites was 93 and 38 per sq. m., respectively. *An. daciae* had the advantage in the unusually shaded breeding site of Yegoryevsk, with high water saprobity and a low oxygen content of 0.8 mg/L (Fig. 8C). The density of larvae in this pool was only 25 per sq. m. It is unclear, if such differentiation is related to the species-specific female oviposition preferences or *An. daciae* larvae being better adapted to tolerate unfavorable developmental conditions.

Discussion

ITS2 sequences represent a reliable tool for discriminating *An. messeae* and *An. daciae*

The ITS2 region located between the 5.8S and 28S rDNA genes was used to discriminate *Anopheles* species from the Moscow region populations in this study [50]. rDNA genes are organized in arrays of multiple copies in eukaryotic genomes and undergo so called “concert evolution,” which leads to unification of the genes and intergenic sequences [51]. Unlike Intergenic Spacer (IGS) and Intergenic Transcribed Spacer 1 (ITS1), ITS2 is a proven reliable tool for species diagnostics in the *Anopheles* genus [50]. ITS2 sequences have been successfully used for the identification of North American species of the Maculipennis group such as *An. hermsi* in the Freeborni subgroup [52] and four new species in the Quadrimaculatus complex [53]. ITS2 sequences have also been utilized for species diagnostics in the

African *An. gambiae* complex [54] and in the *An. funestus* group [55]. One nucleotide difference in ITS2 led to the discovery of the M and S forms of *An. gambiae*, which have now been elevated to species status as *An. gambiae* (formerS form) and *An. coluzzii* (formerM form) [56]. Eurasian *Anopheles* mosquitoes became a subject of ITS2 studies in 1999 when sequences for seven species from the Maculipennis group were obtained [57]. *Three new Palearctic species in the Maculipennis group – An. artemievi* [14], *An. persiensis* [15], and *An. daciae* [16] – were described based on the differences in their ITS2 sequences.

In this study, we sequenced ITS2 from 289 individual *Anopheles* mosquitoes from three populations in the Moscow region: Novokosino, Noginsk, and Yegoryevsk. Our study confirmed the previously reported five SNP differences between *An. messeae* and *An. dacie* in positions 211, 215, 217, 412, and 432 [16]. In addition, our study found a diagnostic SNP in position 150 between the species that had not been found before (Table 1). However, in contrast to the majority of previously published data, our results revealed heterogeneity in the first three nucleotide positions 211, 215, 217 – W(A+T), W(A+T), and Y(C+T) – in all specimens of *An. daciae* (Table 1). Two nucleotides, M(A+C), in position 150 were present simultaneously in all *An. messeae* individuals and in one *An. daciae* individual. When multiple sequences simultaneously occur in the ITS2, it may suggest the presence of multiple rDNA loci in different regions of chromosome X or different chromosomes. The presence of rDNA in several loci that are not subject to inter-locus recombination could make the homogenizing processes less efficient and can lead to the development of locus-specific variants [50]. For example, the presence of an additional rDNA locus on the Y chromosome was found in the ASEMBO1 strain of *An. gambiae* [58]. The presence of two rDNA loci, in the centromere and in the middle of the q arm of chromosome 1, was shown in another mosquito, *Aedes aegypti* [59]. However, localization of the rDNA loci requires further investigation in *An. messeae* and *An. daciae* using fluorescence *in situ* hybridization of rDNA.

Interestingly, the heterogeneity of *An. daciae* ITS2 sequences in positions 211, 215, and 217 was reported in studies of different populations in Russia [24, 44]. However, such heterogeneity was misinterpreted as polymorphism within *An. messeae* populations and the species status of *An. daciae* was doubted by some authors [24]. Our data clearly demonstrate that ITS2 sequences can reliably identify *An. messeae* and *An. daciae* in the Moscow region of Russia. A situation when two nucleotides simultaneously present in chromatograms of ITS2 sequences was shown for malaria mosquitoes from the *An. bancroftii* group in Australia and New Guinea [60] and for the *An. crusians* complex in North America [61]. However, because *An. messeae* is a species with a wide geographic distribution in Eurasia, additional continent-wide studies of the ITS2 sequences are required.

A whole-genome analysis determines the genome-wide divergence between *An. messeae* and *An. daciae*

In addition to ITS2 sequencing, we performed whole-genome sequencing of five specimens of *An. messeae* and *An. daciae* from the Yegoryevsk population. Our study identified the whole-genome

divergence between the two species. Genomic divergence was most pronounced on the X chromosome and in the pericentromeric areas of the autosomes, suggesting a role of the X chromosome and heterochromatin in species differentiation. At the same time, the level of genomic divergence was quite low along the autosomal chromosomal arms indicating that the gene flow is still ongoing in most of the genome. Although PCA and ADMIXTURE analyses based on autosomal SNVs clearly separated the two species into two clusters, we identified two admixed individuals in the sample of 10. Together with a single first-generation hybrid identified by ITS2 sequencing, we can conclude that hybridization occurs between *An. messeae* and *An. daciae* in the Moscow region. Further genomic studies from different geographic locations are required to estimate hybridization frequencies and the level of genomic differentiation between the two species.

Similar genomic divergence was discovered between *An. gambiae* and *An. coluzzi*, formerly the molecular forms S and M, respectively, of the *An. gambiae* complex [62]. As we mentioned above, sequencing of the ITS2 in *An. gambiae* revealed an intriguing one nucleotide difference in the natural populations suggesting incipient speciation within *An. gambiae* [62, 63]. Despite the ongoing gene flow between the M and S forms [64], the whole-genome comparison based on Affymetrix GeneChip microarrays identified so called “speciation islands” [65], which reside in heterochromatic regions [66] of the genome of *An. gambiae*. Further whole-genome sequencing of the two forms demonstrated a high level of genomic differentiation across the entire genome [67], which was especially high near the centromeres and inside the polymorphic inversions. As a result of these studies, the M and S forms were elevated to species status and named *An. coluzzi* and *An. gambiae*, respectively [56]. Interestingly, the level of the genomic divergence between *An. gambiae* and *An. coluzzii* was slightly higher on chromosome X but not as dramatic as between *An. messeae* and *An. daciae*. Such a high level of the genomic divergence on chromosome X between *An. messeae* and *An. daciae* could be explained by the inversions located on this chromosome. Overall, the high level of genomic differentiation between *An. messeae* and *An. daciae* found in our study suggests a long history of reproductive isolation between the species.

Chromosomal inversions differentiate *An. messeae* and *An. daciae*

Our study demonstrated chromosomal differentiation between *An. messeae* and *An. daciae* that is more pronounced on the X sex chromosome. Inversion X1 was found only as a fixed arrangement in *An. messeae,* whereas it was polymorphic in *An. daciae*. Another inversion, X4, was only found as a heterozygote in *An. messeae*. Although the 2R1 and 3R1 autosomal inversions were found in both species, the frequencies of these inversion were much higher in *An. messeae* than in *An. daciae*. Inversion 3L1 was found at low frequency only in *An. messeae.* Rare inversions, X4 and 2R4*, were found as heterozygotes only in *An. messeae* and *An. daciae*, respectively. These data indicate that *An. messeae* is a more polymorphic species at the chromosomal level than *An. daciae.* PCA analyses conducted based on autosomal inversions clearly separated the two species, suggesting the presence of a reproductive barrier between the species.

Chromosomal analysis is considered to be a powerful tool for determining the taxonomic status of a species and for population analyses. Studies conducted in the past using fruit flies and malaria mosquitoes demonstrated that the species often have fixed inversion differences in their chromosomes with rare occurrences of homosequential species. For example, fixed chromosomal differences discriminated *An. maculipennis* and *An. messeae* [9], and *An. sacharovi* and *An. martinius* [10]. Another species from the Maculipennis group, *An. beklemishevi*, was identified based on the fixed chromosome differences [11, 12]. Based on the banding pattern of polytene chromosomes, species in the *An. gambiae* complex have been differentiated by 10 fixed inversions [68]. Five of these inversions are located on the X chromosome separating species into three groups possessing the compound Xag inversion, the standard X arrangement, and the compound Xbcd inversion [69]. Intensive studies of inversion polymorphism within the major malaria vector *An. gambiae* identified five chromosomal forms: Bamako, Savanna, Mopti, Forest, and Bissau [70, 71]. Interestingly, some of the combinations of chromosomal inversions have never been observed in nature, suggesting the presence of reproductive barriers within the populations. Originally, it was shown that the M and S molecular forms of *An. gambiae* correlate with the previously identified chromosomal forms, Mopti and Savanna. However, later studies determined that these inversion patterns did not always correspond to *An. gambiae* and *An. coluzzi* across the African continent [72, 73]. Interestingly, all chromosomal inversions located in the X sex chromosome in the *An. gambiae* complex are fixed between the species, suggesting that this chromosome may play an important role in speciation. In contrast, 2 polymorphic inversions were found on the X chromosome between *An. daciae* and *An. messeae* in our study.

An. messeae and *An. daciae* differ in their ecological preferences and behavior

Understanding species composition and ecological preferences is important for the development and application of any strategy for vector control [5]. The malaria mosquitoes *An. daciae* and *An. messeae* occupy regions with temperate oceanic and humid continental climates according to the Eurasian Koppen–Geiger climate classification system [74]. These two species often occur in sympatry [17, 19-23, 75, 76] but their ecological niches do not completely overlap. A recent study conducted in Germany demonstrated that the probability of occurrence of *An. messeae* had a negative correlation with the maximum temperature of the warmest month and altitude, whereas the probability of occurrence of *An. daciae* had a positive correlation [75]. In addition, *An. messeae* negatively correlated with the proportion of variable agricultural land cover and *An. daciae* was negatively correlated with pasture land cover. As a result, *An. messeae* demonstrates a decrease in the probability of occurrence from the north to the south of Germany. In contrast, *An. daciae* has a low probability of occurrence in northwestern and southern Germany. The latter species is also absent in coastal areas of Germany, where *An. messeae* occurs [19]. Two species demonstrate differences in seasonal dynamics as well: *An. messeae* is more abundant early and late in the season whereas *An. daciae* is prevalent in the warmest mid-summer months [19, 23, 44] . Both species have differences in blood feeding behavior: females of *An. messeae* prefer feeding on

animals whereas females of *An. daciae* have opportunistic blood choice behavior and feed on birds, humans, deer, and livestock [20]. This finding suggests that *An. daciae* may serve as a bridge vector for arboviruses to humans. Moreover, a female of *An. daciae* infected with the nematode *Dirofilaria repens* was detected in Germany [18].

Our study provides further insights into possible ecological differentiation of the two species with respect to their preference for larval breeding places. Previous studies demonstrated that *An. messeae* prefers large clean water reservoirs with a relatively low dissolved ion content [17, 23]. In our study, *An. messeae* alsodominated in the typical oxygen-rich (1.8–4.0 mg /l) anophelogenic breeding sites of Novokosino and Noginsk, with the density of *Anopheles* larvae being 93 and 38 per sq. m., respectively (Fig. 8A, B). *An. daciae* had an advantage in the atypical breeding place in Yegoryevsk, with high water saprobity, low oxygen content (0.8 mg /l), and low density of *Anopheles* larvae (25 per sq. m.). We noted a low proportion of younger larvae in Yegoryevsk. This finding suggests that *An. daciae* could better tolerate unfavorable development conditions than *An. messeae*. We conclude that the survival optimum for both species, *An. daciae* and *An. messeae*, is in the zone of temperate deciduous and mixed forests. There is a sufficient number of favorable biotopes with abundant aquatic vegetation for joint development of both species in the temperate zone. Ecological specialization of sibling species becomes obvious with deviations from the optimum, up to the competitive exclusion of one of the species. Such competitive displacement can be observed at the edges of the ranges, in the steppe and southern taiga zones. However, it is obvious that ecological preferences and behavioral differences of *An. messeae* and *An. daciae* and the details of their geographical distribution in Russia and other Eurasian countries requires further investigation.

Conclusions

In this study, we performed ITS2 sequencing and cytogenetic analysis of 289 specimens of malaria mosquitoes collected in three locations in the Moscow region of Russia. In addition, we conducted whole-genome sequencing on five individual mosquitoes of *An. messeae* and *An. daciae* from each location. Our study determined that the ITS2 sequence represents a reliable tool for distinguishing *An. messeae* and *An. daciae.* We demonstrated that ITS2 sequences of *An. messeae* in the Moscow region have the same structure as one described in European populations, whereas ITS2 of *An. daciae* displayed heterogeneity in the first three nucleotide substitutions but not in the last two positions. In addition to the previously described substitutions, we identified a diagnostic substitution in position 150. Genomic analyses of 10 individual mosquitoes demonstrated a genome-wide differentiation along the chromosomes. The most dramatic genomic differences were found on the sex chromosome X, suggesting a role of this chromosome in reproductive isolation between the species. The ADMIXTURE cluster analysis of 10 individual genomes identified two clusters corresponding to *An. messeae* and *An. daciae*. The presence of two admixed individuals suggests that hybridization is ongoing between the populations of *An. messeae* and *An. daciae*. A cytogenetic analysis demonstrated strong differentiation between *An. messeae* and *An. daciae* by polymorphic inversions especially on the X chromosome.

Inversion X1 was completely fixed in all three populations of *An. messeae,* whereas frequencies of the inverted and standard arrangements in *An. daciae* were almost equal in all populations. Although the two most abundant polymorphic autosomal inversions were found in both species, their frequencies were much higher in *An. messeae* than in *An. daciae.* Overall, our study demonstrated that *An. messeae* and *An. daciae* represent closely related cryptic species with incomplete reproductive isolation. These species have been able to maintain their genome integrity in sympatry despite genetic introgression. Better understanding of the species composition, their population structure, and ecological preferences will help to develop adequate and efficient mosquito control.

Methods

Field collection and material preservation

300 mosquito larvae were collected in the summer of 2016 in three locations of the Moscow region, Russia: Novokosino (55°44'02.6"N, 37°50'33.5"E), Kolyshkino boloto in Noginsk (55°53'57.4"N, 38°26'17.9"E), and the village of Parykino near Yegoryevsk (55°17'17.9"N, 39°22'03.5"E). Positions of the mosquito breeding sites are shown at the map that was developed using OpenStreetMap [77] software (Fig. 7). The Novokosino population is located within the boundaries of Moscow. The ponds were characterized by an increased oxygen level (1.8–4.0 mg /l) and high density of larvae (93 and 38 per sq. m., respectively) in Novokosino and Noginsk and by a low oxygen level (0.8 mg /l) and low density of larvae (25 per sq. m.) in Yegoryevsk. The main species of aquatic and near-aquatic plants were *Lemna minor* L., *Spirodela polyrhiza* Schleid, *Alisma plantago-aquatica* L., *Carex pseudocyperus* L., *Pharagmites australis* Trin ex Steud, *Typha latifolia* L.in Novokosino; *Cladophora sp., Alisma plantago-aquatica* L., *Carex hirta* L., *Typha latifolia* L.in Noginsk; *Lemna trisulca* L., *Rhizoclonium hieroglyphicum Kutz, Scirpus radicans* Schkuhr, *Oenanthe aquatica* Poir *in Yegoryevsk. Anopheles* larvae were collected by the dipping method and then fixed in Carnoy's solution (1:3 acetic acid: ethanol). Larvae were dissected and each individual mosquito was numbered (Supplementary Table S1). Each larva was divided into two parts: head with thorax and abdomen. These dissected parts were placed into separate tubes for further analysis. Heads with thoraxes were kept in Carnoy's solution and abdomens were placed in 70% ethanol.

Genotyping

The abdomens from each of the 300 specimens were used for DNA extraction. For individual homogenization, sterile 1.5 ml tubes were used to prevent the risk of contamination. Genomic DNA was extracted from the specimens using a standard protocol for the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD, USA) with slight modifications. Samples were homogenized in 45 µl of extraction buffer with 5 µl added Proteinase K and incubated at 56 °C overnight. DNA elution was performed in 100 µl of water.

ITS2 from rDNA was amplified using the forward universal primer designed by J. Proft [49] 5'-ATCACTCGGCTCTCGTGGATCG-3'(Tm=64.5 °C) and the reverse primer used by Y. Novikov [42]: 5'-ATGCTTAAATTTAGGGGGTA-3' (Tm=54.2 °C). This primer combination was chosen based on the longer amplicon length (613 bp) and high rate of PCR success (100%). Hot start ImmoMix^{TM} polymerase reaction mix (Bioline, Taunton, MA, USA) was used for the PCR reaction. The PCR mixture contained 1-2 µl of DNA template (depending on concentration), 1 µl of both forward and reverse primers at 10 µM concentration, and 10 µl of 2×Immomix reaction-mix. Water was added to the mixture to make 20 µl of total volume. Amplification was performed using a thermal cycler (Eppendorf, Hauppauge, NY, USA) with the following programmed parameters: initial denaturation at 95 °C for 10 min, followed by 35 cycles of 95 °C for 15 s, 55 °C for 30 s, and 72 °C for 30s, and a final extension step at 72 °C for 5 min. The reaction was then placed on hold at 4 °C.

For DNA sequencing, amplicons were visualized using gel electrophoresis and then purified with a Wizard^{TM} PCR Clean Up kit (Promega, Fitchburg, WI, USA). Concentrations of purified PCR products were measured using a Nanodrop (Thermo Fisher Scientific, Haverhill, MA, USA). PCR products were mixed with either forward or reverse primers and sequenced on the Sanger sequencing platform at the Virginia Biocomplexity Institute. The majority of samples were sequenced with both forward and reverse primers to eliminate the possibility of contamination and to confirm the presence of single-nucleotide polymorphisms (SNPs) on both DNA strands. Samples from Noginsk were sequenced with forward primers only.

A total of 289 specimens were successfully genotyped and further analyzed. ITS2 sequences were analyzed using the SeqScape v 2.5 software (Thermo Fisher Scientific, Haverhill, MA, USA). The ITS2 sequence of *An. messeae* (Genbank: AY648982) [16] was used as a reference. Base calling was performed using KB basecaller with mixed base identification enabled. The previously described SNPs in positions 211, 215, 217, 412, and 432 distinguish ITS2 sequences of *An. measseae* from *An. daciae* and the new diagnostic SNP in position 150 were investigated in detail. To verify the appropriateness of the assignment of a nucleotide base in these positions, all diagnostic SNPs were analyzed manually. All ITS2 sequences were submitted to Genbank to obtain accession numbers (Supplementary Table S1).

Sequencing and analysis of individual mosquito genomes

The individual mosquito genomes of *An. daciae* (ME36, ME71, ME72, ME78, ME89) and *An. messeae* (ME3, ME14, ME24, ME25, ME71) were sequenced using Illumina HiSeq 4000 platform in the Fasteris, SA sequencing center. Standard genomic Nano libraries for 2x 150 bp paired-read sequencing were prepared for each sample to obtain 10 million reads aiming at 10x genome coverage. The raw reads were checked with FastQC software [78] and then trimmed using Trimmomatic v0.36 [79] to remove bases with poor base quality and the remains of adapters with the following flags: LEADING:24 TRAILING:24 SLIDINGWINDOW:4:24 MINLEN:50. The trimmed reads were aligned to the *An. atroparvus* genome AatrE2 [45, 80] using the BWA mem algorithm [81] with -M flag and sorted with the *samtools sort* command of

the SAMtools package [82]. The SNVs were called using *samtools mpileup* | *bcftools call* command [83] with the following settings: minimum mapping quality = 55 (-q 55), minimum base quality = 26 (-Q 26), calling only SNVs (-I), and multiallelic calling model (-m). The resulting VCF file was filtered out with VCFtools [84] allowing a minimal phred-scaled SNV quality = 400 (–minQ 400), minimal genotype quality = 10 (–minGQ=10), no more than 1 missing genotype per SNV (–max-missing-count 1), minimal distance between the SNVs=10Kbp (–thin 10000) to reduce the number of linked loci, and allow only biallelic SNVs. The scaffold-based coordinates of the SNVs were converted to the chromosome-based coordinates with the Chromosomer tool [85] based on comparison of the cytogenetic maps of *An. measseae* [26] and *An. atroparvus* [86].

We calculated the index of population differentiation, Fst [87] between *An. messeae* and *An. daciae* using the VCFtools function –weir-fst-pop in the sliding nonoverlapping 5 Kbp windows along the genome. To infer genetic ancestry of the individual genomes, we used ADMIXTURE software [88] to assign fractions of the individual genomes to putative ancestral genetic clusters with K=2 and only autosomal sequences. Principal Component Analysis based on the whole genome sequences was carried out with SNPRelate software [89].

Karyotyping

Salivary glands were dissected from the larval thorax for preparations of polytene chromosomes. Chromosome preparations were made by the standard lacto-aceto-orcein method [90]. Polytene chromosomes were visualized using an Eclipse E200 light microscope (Nikon, BioVitrum, Moscow, Russia). Specimens were karyotyped using a standard chromosome map for the salivary glands of *An. messeae* [26, 33, 90]. All inversions—X1, X2, X4, 2R1, 3R1, and 3L1—were considered and the karyotype of each specimen was described for the whole chromosomal complement (Supplementary Table S1). We attempted to karyotype a total of 300 specimens (100 samples from each location) but only 289 samples were successfully karyotyped.

Statistical analysis of chromosomal inversions

We used the software GENODIVE [91] to statistically analyze the population architecture of the chromosomal polymorphism. We exploited the AMOVA framework [92, 93] to test for significance of the deviations from the Hardy-Weinberg equilibrium with 10k permutations and applied the Bonferroni correction for multiple testing. The pairwise Fst values [87, 92, 93] between the studied populations were calculated based on 10k permutations using only autosomal inversions and the Bonferroni correction was applied. Finally, we performed the Principal Component Analysis with GENODIVE software using the frequencies of autosomal inversions on the population level to infer and visualize the relationship between the species and its populations.

List Of Abbreviations

ITS2: internal transcribed spacer 2; FISH: fluorescent *in situ* hybridization; Kbp: kilo base pairs; PCR: polymerase chain reaction; SNV: single nucleotide variant; SNP: single nucleotide polymorphism; Fst: fixation index; PCA: principal component analysis; RFLP: restriction fragment length polymorphism.

Declarations

Ethics and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

ITS2 sequences are available in NCBI [94] under accession numbers shown in Supplementary Table S1. Genomic sequencing data will be available via NCBI upon the publication of the manuscript.

Competing interests

Authors declare no competing interests.

Funding

DNA sequencing, bioinformatic and statistical analyses were supported by the Russian Science Foundation grant № 19-14-00130 to MVS. Mosquito collections as well as cytogenetic analyses were supported by the Russian Foundation of Basic Research grant № 18-04-01117 A to MIG. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author contribution

MVS, MIG, and IVS designed experiments, ANN and OIM performed genotyping, DAK and EMB conducted sequence analysis, AAY performed statistical and bioinformatics analysis, MIG and AVM conducted cytogenetic analysis, EMB provided resources, MVS, ANN, DAK, AAY and MIG wrote the manuscript. All authors read and approved the manuscript.

Acknowledgments

We thank Lyudmila Tega for her help in collecting mosquito larvae and Janet Webster for proofreading the text.

Referencies

1. Falleroni D: Fauna anofelicaitaliana e suo 'habitat' (paludi, risaie, canali). Metodi di lottacontro la malaria. *Riv Malar* 1926, 5(5-6):553-593.

2. Sokolova MI, Snow KR: Malaria vectors in European Russia. *European Mosquito Bulletin Journal of the European Mosquito Control Association* 2002, 12:1-6.

3. Novikov YM, Vaulin OV: Expansion of *Anopheles maculipennis s.s.* (Diptera: Culicidae) to northeastern Europe and northwestern Asia: causes and consequences. *Parasit Vectors* 2014, 7:389.

4. Githeko AK, Lindsay SW, Confalonieri UE, Patz JA: Climate change and vector-borne diseases: a regional analysis. *Bull World Health Organ* 2000, 78(9):1136-1147.

5. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW *et al*: The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis. *Parasit Vectors* 2010, 3:117.

6. Harbach RE: The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull Entomol Res* 2004, 94(6):537-553.

7. Hackett LW, Missiroli A: The varieties of *Anopheles maculipennis* and their relation to the distribution of malaria in Europe. *Riv Malariol* 1935, 14:45-109.

8. Gutsevich AV, Monchadskii AS, Shtakelberg AA: Fauna of the USSR. Diptera. Mosquitoes. Leningard: Zoological Institute, USSR Academy of Science; 1970.

9. Stegnii VN, T.S. P, Kabanova VM: Cytological identification of sibling species of the malaria mosquitoes *Anopheles maculipennis* and *An. messeae*. *Zoologicheskiy Zhurnal* 1973, 52:1971-1976.

10. Stegnii VN: Revealing of chromosme races in the malaria mosquito *Anopheles sacharovi* (Diptera, Culicidae). *Tsitologia* 1976, 18.

11. Stegnii VN, Kabanova VM: [Cytoecological study of natural populations of malaria mosquitoes on the USSR territory. 1. Isolation of a new species of *Anopheles* in Maculipennis complex by the cytodiagnostic method]. *Med Parazitol (Mosk)* 1976, 45(2):192-198.

12. Stegniy VN, Kabanova VM: Cytoecological study of indigenous populanions of the malaria mosquito in the territory of the USSR. 1. Identification new species of *Anopheles* in the maculipennis complex by

the cytodiagnostic method. *Mosquito Systematics* 1978, 10(1):1-12.

13. Stegniy VN: Population genetics and evolution of malaria mosquitoes: Tomsk State University Publisher; 1991.

14. Gordeev MI, Zvantsov AB, Goriacheva, II, Shaikevich EV, Ezhov MN: [Description of the new species *Anopheles artemievi* sp.n. (Diptera, Culicidae)]. *Med Parazitol (Mosk)* 2005(2):4-5.

15. Sedaghat MM, Linton YM, Oshaghi MA, Vatandoost H, Harbach RE: The *Anopheles maculipennis* complex (Diptera: Culicidae) in Iran: molecular characterization and recognition of a new species. *Bull Entomol Res* 2003, 93(6):527-535.

16. Nicolescu G, Linton YM, Vladimirescu A, Howard TM, Harbach RE: Mosquitoes of the *Anopheles maculipennis* group (Diptera: Culicidae) in Romania, with the discovery and formal recognition of a new species based on molecular and morphological evidence. *Bull Entomol Res* 2004, 94(6):525-535.

17. Weitzel T, Gauch C, Becker N: Identification of *Anopheles daciae* in Germany through ITS2 sequencing. *Parasitol Res* 2012, 111(6):2431-2438.

18. Kronefeld M, Werner D, Kampen H: PCR identification and distribution of *Anopheles daciae* (Diptera, Culicidae) in Germany. *Parasitol Res* 2014, 113(6):2079-2086.

19. Kampen H, Schafer M, Zielke DE, Walther D: The *Anopheles maculipennis* complex (Diptera: Culicidae) in Germany: an update following recent monitoring activities. *Parasitol Res* 2016, 115(9):3281-3294.

20. Danabalan R, Monaghan MT, Ponsonby DJ, Linton YM: Occurrence and host preferences of *Anopheles maculipennis* group mosquitoes in England and Wales. *Med Vet Entomol* 2014, 28(2):169-178.

21. Rydzanicz K, Czulowska A, Manz C, Jawien P: First record of *Anopheles daciae* (Linton, Nicolescu & Harbach, 2004) in Poland. *J Vector Ecol* 2017, 42(1):196-199.

22. Blazejova H, Sebesta O, Rettich F, Mendel J, Cabanova V, Miterpakova M, Betasova L, Pesko J, Hubalek Z, Kampen H *et al*: Cryptic species *Anopheles daciae* (Diptera: Culicidae) found in the Czech Republic and Slovakia. *Parasitol Res* 2018, 117(1):315-321.

23. Kavran M, Zgomba M, Weitzel T, Petric D, Manz C, Becker N: Distribution of *Anopheles daciae* and other *Anopheles maculipennis* complex species in Serbia. *Parasitol Res* 2018, 117(10):3277-3287.

24. Bezzhonova OV, Goryacheva, II: Intragenomic heterogeneity of rDNA internal transcribed spacer 2 in *Anopheles messeae* (Diptera: Culicidae). *J Med Entomol* 2008, 45(3):337-341.

25. Gornostaeva RM, Danilov AV: On distribution of malaria mosquitoes (Diptera, Culicidae: *Anopheles*) from Maculipennis Complex in Russian territory. *Parsitologia* 2002, 26(1):33-47.

26. Stegnii VN, Kabanova VM, Novikov Iu M: [Study of the karyotype of the malaria mosquito]. *Tsitologiia* 1976, 18(6):760-766.

27. Stegnii VN, Kabanova VM, Novikov YM, Pleshkova GN: Inversion polymorphism in malaria mosquito *Anopheles messeae*. I. Distribution of the inversions in the species areal. *Genetika* 1976, 12(4):47-55.

28. Gordeev MI, Sibataev AK: [Cytogenetic and phenotypic variation in central and peripheral populations of the malaria mosquito, *Anopheles messeae* Fall. (Diptera, Culicidae)]. *Genetika* 1996, 32(9):1199-1205.

29. Gordeev MI, Bezzhonova OV, Moskaev AV: [Chromosomal polymorphism in the populations of malaria mosquito *Anopheles messeae* (Diptera, Culicidae) at the south of Russian Plain]. *Genetika* 2012, 48(9):1124-1128.

30. Perevozkin VP, Gordeev MI, Bondarchuk SS: [Chromosome polymorphism and regularities of the subpopulation organization of malaria mosquitoes *Anopheles* (Diptera, Culicidae) in biotopes of the Tomsk oblast]. *Genetika* 2009, 45(4):478-487.

31. Kabanova VM, Stegnii VN, Luzhkova AG: [Seasonal dynamics of inversion polymorphism in a natural population of the malarial mosquito *Anopheles messeae* (Diptera: Culicidae)]. *Genetika* 1973, 9(10):78-82.

32. Stegnii VN: [Inversion polymorphism of the malarial mosquito *Anopheles messeae*. IV. The stability of the frequency distribution of the inversions by species area]. *Genetika* 1983, 19(3):466-473.

33. Stegniy VN, Pishchelko AO, Sibataev AK, Abylkassymova G: [Spatial and temporal variations of the chromosomal inversion frequencies across the range of malaria mosquito *Anopheles messeae* Fall. (Culicidae) during the 40-year monitoring period]. *Genetika* 2016, 52(6):664-671.

34. Novikov Iu M, Kabanova VM: [Adaptive association of inversions in a natural population of the malaria mosquito *Anopheles messeae* Fall]. *Genetika* 1979, 15(6):1033-1045.

35. Gordeev MI, Stegnii VN: [Inversion polymorphism in the malaria mosquito *Anopheles messeae*. VII. Fertility and the population genetics structure of the species]. *Genetika* 1987, 23(12):2169-2174.

36. Stegnii VN: [Inversion polymorphism of the malarial mosquito *Anopheles messeae*. V. The interaction of different chromosomal inversions in the spatial area]. *Genetika* 1983, 19(3):474-482.

37. Gordeev MI, Troshkov N: [Inversion polymorphism of the malaria mosquito *Anopheles messeae.* IX. Cannibalism in larvae as a selection factor]. *Genetika* 1990, 26(9):1597-1603.

38. Burlak VA, Gordeev MI: [The effect of infection by the entomopathogenic bacterium Bacillus thuringiensis on the spread of microsporidia in an inversion-polymorphic population of the malarial mosquito *Anopheles messeae* (Diptera: Culicidae)]. *Parazitologiia* 1998, 32(3):264-267.

39. Gordeev MI, Perevozkin VP: [Strategies for selection and stability to asphyxia in larvae of the malaria mosquito *Anopheles messeae* with various karyotypes]. *Genetika* 1995, 31(2):180-184.

40. Gordeev MI, Burlak VA: [Inversion polymorphism in malaria mosquito *Anopheles messeae*. Part X. Resistance of larvae with different genotypes to toxins of crystal-forming bacteria *Bacillus thuringiensis subsp. israelensis* (serovar H14)]. *Genetika* 1991, 27(2):238-246.

41. Gordeev MI, Burlak VA: [Inversion polymorphism in the malaria mosquito *Anopheles messeae*. XI. The group effect of larval infection with *Bacillus thuringiensis subsp. israeliensis* bacteria]. *Genetika* 1992, 28(7):82-88.

42. Novikov Iu M, Shevchenko AI: [Inversion polymorphism and the divergence of two cryptic forms of *Anopheles messeae* (Diptera, Culicidae) at the level of genomic DNA repeats]. *Genetika* 2001, 37(7):915-925.

43. Vaulin OV, Novikov YM: Geographic variability of ITS2 rDNA and COI mtDNA and cryptic species of malaria mosquito *Anopheles messeae* Fall. (Diptera: Culicidae). *Vestnik VOGIS* 2010, 14(3):546-555.

44. Vaulin OV, Karagodin DA, Zakharov IK, Baricheva EM: The dynamics of the species composition of malaria mosquitoes in Siberian populations, detected using restriction analysis. *Genetika* 2018, 54(7):832-842.

45. Artemov GN, Bondarenko SM, Naumenko AN, Stegniy VN, Sharakhova MV, Sharakhov IV: Partial-arm translocations in evolution of malaria mosquitoes revealed by high-coverage physical mapping of the *Anopheles atroparvus* genome. *BMC Genomics* 2018, 19(1):278.

46. Tanygina E, Gordeev MI, Moskaev AV, Ganushkina LA: [The species and karyotype composition of malaria mosquito larvae in different water reservoirs of the city of Moscow]. *Med Parazitol (Mosk)* 2014(2):29-33.

47. Gordeev MI, Moskaev AV: [Chromosomal polymorphism in the populations of malaria mosquito *Anopheles messeae* (Diptera, Culicidae) in the Volga region]. *Genetika* 2016, 52(6):685-690.

48. Stegniy VN, Novikov Iu M, Kabanova VM: Cytogenetic analysis and distribution of *Anopheles beklemishevi*. *Zoological Journal* 1978, LVII(6):873-876.

49. Proft J, Maier WA, Kampen H: Identification of six sibling species of the *Anopheles maculipennis* complex (Diptera: Culicidae) by a polymerase chain reaction assay. *Parasitol Res* 1999, 85(10):837-843.

50. Collins FH, Paskewitz SM: A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Mol Biol* 1996, 5(1):1-9.

51. Dover GA, Flavell RB: Molecular coevolution: DNA divergence and the maintenance of function. *Cell* 1984, 38(3):622-623.

52. Collins FH, Porter CH, Cope SE: Comparison of rDNA and mtDNA in the sibling species *Anopheles freeborni* and *A. hermsi*. *Am J Trop Med Hyg* 1990, 42(5):417-423.

53. Cornel AJ, Porter CH, Collins FH: Polymerase chain reaction species diagnostic assay for *Anopheles quadrimaculatus* cryptic species (Diptera: Culicidae) based on ribosomal DNA ITS2 sequences. *J Med Entomol* 1996, 33(1):109-116.

54. Paskewitz SM, Wesson DM, Collins FH: The internal transcribed spacers of ribosomal DNA in five members of the *Anopheles gambiae* species complex. *Insect Mol Biol* 1993, 2(4):247-257.

55. Hackett BJ, Gimnig J, Guelbeogo W, Costantini C, Koekemoer LL, Coetzee M, Collins FH, Besansky NJ: Ribosomal DNA internal transcribed spacer (ITS2) sequences differentiate *Anopheles funestus* and *An. rivulorum*, and uncover a cryptic taxon. *Insect Mol Biol* 2000, 9(4):369-374.

56. Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ: *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* 2013, 3619:246-274.

57. Marinucci M, Romi R, Mancini P, Di Luca M, Severini C: Phylogenetic relationships of seven palearctic members of the maculipennis complex inferred from ITS2 sequence analysis. *Insect Mol Biol* 1999, 8(4):469-480.

58. Wilkins EE, Howell PI, Benedict MQ: X and Y chromosome inheritance and mixtures of rDNA intergenic spacer regions in *Anopheles gambiae*. *Insect Mol Biol* 2007, 16(6):735-741.

59. Dickson LB, Sharakhova MV, Timoshevskiy VA, Fleming KL, Caspary A, Sylla M, Black WCt: Reproductive Incompatibility Involving Senegalese *Aedes aegypti* (L) Is Associated with Chromosome Rearrangements. *PLoS Negl Trop Dis* 2016, 10(4):e0004626.

60. Beebe NW, Maung J, van den Hurk AF, Ellis JT, Cooper RD: Ribosomal DNA spacer genotypes of the *Anopheles bancroftii* group (Diptera: Culicidae) from Australia and Papua New Guinea. *Insect Mol Biol* 2001, 10(5):407-413.

61. Wilkerson RC, Reinert JF, Li C: Ribosomal DNA ITS2 sequences differentiate six species in the *Anopheles crucians* complex (Diptera: Culicidae). *J Med Entomol* 2004, 41(3):392-401.

62. della Torre A, Costantini C, Besansky NJ, Caccone A, Petrarca V, Powell JR, Coluzzi M: Speciation within *Anopheles gambiae*--the glass is half full. *Science* 2002, 298(5591):115-117.

63. Favia G, della Torre A, Bagayoko M, Lanfrancotti A, Sagnon N, Toure YT, Coluzzi M: Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Mol Biol* 1997, 6(4):377-383.

64. Lanzaro GC, Toure YT, Carnahan J, Zheng L, Dolo G, Traore S, Petrarca V, Vernick KD, Taylor CE: Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by

microsatellite DNA analysis. *Proc Natl Acad Sci U S A* 1998, 95(24):14260-14265.

65. Turner TL, Hahn MW, Nuzhdin SV: Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 2005, 3(9):e285.

66. Sharakhova MV, George P, Brusentsova IV, Leman SC, Bailey JA, Smith CD, Sharakhov IV: Genome mapping and characterization of the *Anopheles gambiae* heterochromatin. *BMC Genomics* 2010, 11:459.

67. Lawniczak MK, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J *et al*: Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 2010, 330(6003):512-514.

68. Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V: A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 2002, 298(5597):1415-1418.

69. Kamali M, Xia A, Tu Z, Sharakhov IV: A New Chromosomal Phylogeny Supports the Repeated Origin of Vectorial Capacity in Malaria Mosquitoes of the *Anopheles gambiae* Complex. *PLoS Pathog* 2012, 8(10):e1002960.

70. Toure YT: The current state of studies of malaria vectors and the antivectorial campaign in west Africa. *Trans R Soc Trop Med Hyg* 1989, 83 Suppl:39-41.

71. Toure YT, Petrarca V, Traore SF, Coulibaly A, Maiga HM, Sankare O, Sow M, Di Deco MA, Coluzzi M: Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s.str.* in Mali, west Africa. *Genetica* 1994, 94(2-3):213-223.

72. della Torre A, Tu Z, Petrarca V: On the distribution and genetic differentiation of *Anopheles gambiae s.s. molecular forms. Insect Biochem Mol Biol* 2005, 35(7):755-769.

73. Wondji C, Frederic S, Petrarca V, Etang J, Santolamazza F, Della Torre A, Fontenille D: Species and populations of the *Anopheles gambiae* complex in Cameroon with special emphasis on chromosomal and molecular forms of *Anopheles gambiae s.s. J Med Entomol* 2005, 42(6):998-1005.

74. Geiger R: Classification of climates after W. Koppen, vol. 3. Berlin: Springer; 1954.

75. Luhken R, Czajka C, Steinke S, Jost H, Schmidt-Chanasit J, Pfitzner W, Becker N, Kiel E, Kruger A, Tannich E: Distribution of individual members of the mosquito *Anopheles maculipennis* complex in Germany identified by newly developed real-time PCR assays. *Med Vet Entomol* 2016, 30(2):144-154.

76. Kronefeld M, Dittmann M, Zielke D, Werner D, Kampen H: Molecular confirmation of the occurrence in Germany of *Anopheles daciae* (Diptera, Culicidae). *Parasit Vectors* 2012, 5:250.

77. OpenStreetMap. Avalible at: https://www.openstreetmap.org/

78. Andrews S: FastQC: a quality control tools for high throughput sequnce data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010.

79. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.

80. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arca B, Arensburger P, Artemov G *et al*: Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* 2015, 347(6217):1258522.

81. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25(14):1754-1760.

82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25(16):2078-2079.

83. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011, 27(21):2987-2993.

84. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: The variant call format and VCFtools. *Bioinformatics* 2011, 27(15):2156-2158.

85. Tamazian G, Dobrynin P, Krasheninnikova K, Komissarov A, Koepfli KP, O'Brien SJ: Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* 2016, 5(1):38.

86. Artemov GN, Sharakhova MV, Naumenko AN, Karagodin DA, Baricheva EM, Stegniy VN, Sharakhov IV: A standard photomap of ovarian nurse cell chromosomes in the European malaria vector *Anopheles atroparvus*. *Med Vet Entomol* 2015, 29(3):230-237.

87. Weir BS, Cockerham CC: Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 1984, 38(6):1358-1370.

88. Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009, 19(9):1655-1664.

89. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS: A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012, 28(24):3326-3328.

90. Kabanova VM, Kartashova NN, Stegnii VN: [Karyological study of natural populations of malarial mosquitoes in the Middle Ob river. I. Characteristics of the karyotype of *Anopheles maculipennis messeae*]. *Tsitologiia* 1972, 14(5):630-636.

91. Meirmans PG, Van Tienderen PH: GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 2004, 4(4):792-794.

92. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992, 131(2):479-491.

93. Michalakis Y, Excoffier L: A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 1996, 142(3):1061-1064.

94. NCBI. Avaliable at: https://www.ncbi.nlm.nih.gov/

Tables

Table 1. Variations in ITS2 genotypes of *An. messeae* and *An. daciae* from three populations in the Moscow region.

Species	Genotype						Population		
	150	211	215	217	412	432	Novokosino	Noginsk	Yegoryevsk
An. daciae	C	W(A+T)	W(A+T)	Y(C+T)	A	C	23	32	60
	C	W	W	T	A	C	2	0	0
	M(A+C)	W	W	Y	A	C	0	0	2
An. messeae	M(A+C)	T	T	C	G	G	72	64	24
Hybrid	M	W	W	Y	R(A+G)	S(C+G)	0	0	1
Total							97	96	87

Table 2. Statistical significance (P-values) of the deviations from Hardy-Weinberg equilibrium in the studied populations by inversion polymorphism. * - results are significant with α < 0.05 after applied Bonferroni correction.

Population, species	X	2R	3R	3L
Novokosino, *An. messeae*	0.817	0.071	0.433	0.958
Novokosino, *An. daciae*	0.046	Monomorphic	0.878	Monomorphic
Noginsk, *An. messeae*	0.051	0.111	0.302	0.838
Noginsk, *An. daciae*	0.002*	1.000	0.242	Monomorphic
Yegoryevsk, *An. messeae*	0.022*	0.595	0.601	0.791
Yegoryevsk, *An. daciae*	0.001*	0.992	0.922	Monomorphic

Table 3. Pairwise Fst values between the studied populations of *An. messeae* and *An. daciae* based on the frequency of autosomal inversions. * - results are significant with α < 0.05 after applied Bonferroni

correction.

Population species	Novokosino *An. messeae*	Novokosino *An. daciae*	Noginsk *An. messeae*	Noginsk *An. daciae*	Yegoryevsk *An. messeae*	Yegoryevsk *An. daciae*
Novokosino *An. messeae*	--	0.159*	0.003	0.143*	0.021	0.218*
Novokosino *An. daciae*	0.159*	--	0.211*	-0.017	0.095*	0.000
Noginsk *An. messeae*	0.003	0.211*	--	0.197*	0.046	0.280*
Noginsk *An. daciae*	0.143*	-0.017	0.197*	--	0.076	0.010
Yegoryevsk *An. messeae*	0.021	0.095*	0.046	0.076	--	0.164*
Yegoryevsk *An. daciae*	0.218*	0.000	0.280*	0.010	0.164*	--

Figures

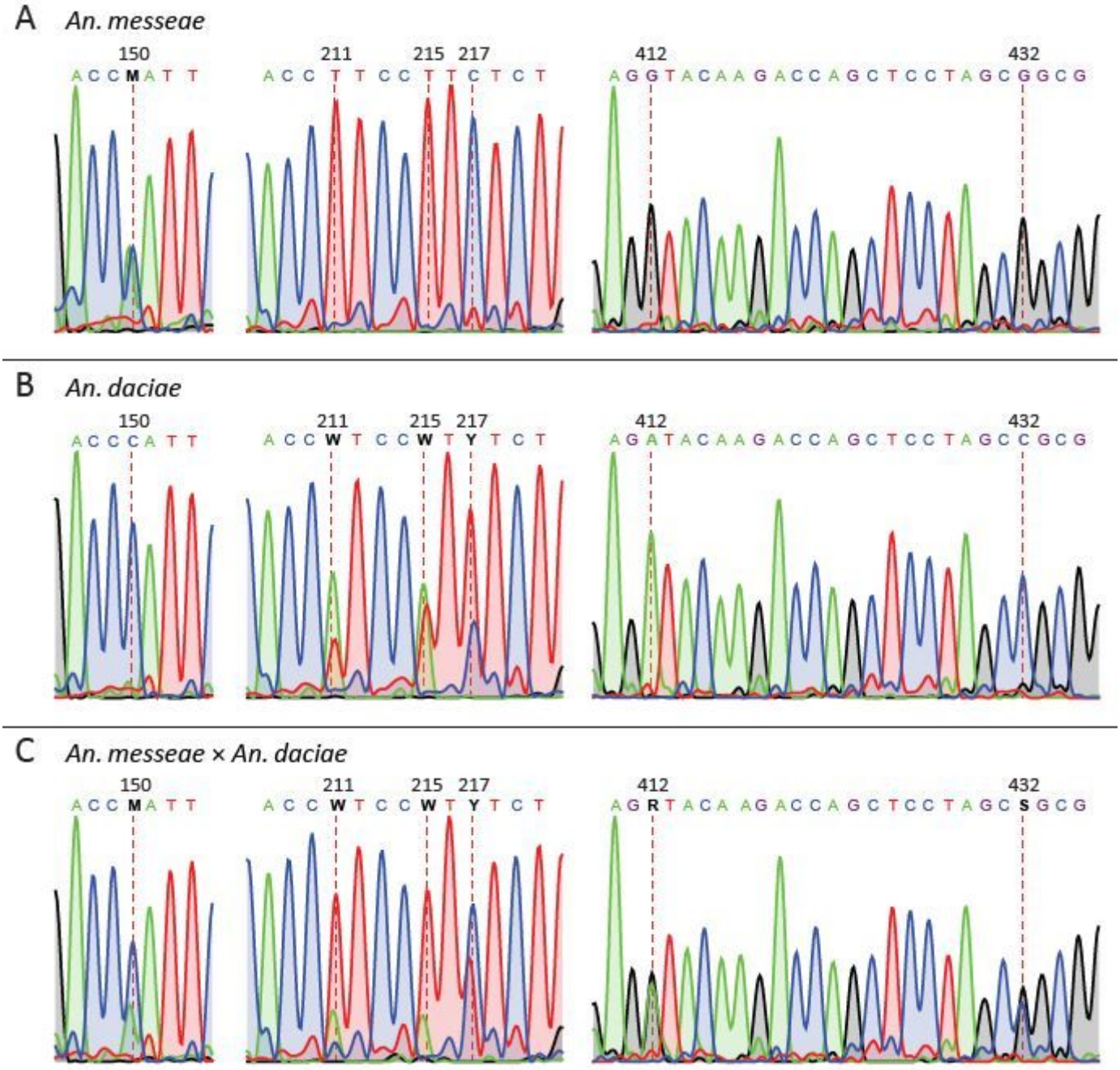

Figure 1

Examples of ITS2 sequence chromatograms for An. messeae (A), An. daciae (B), and their hybrid (C). Dash lines indicate positions of SNPs that distinguish the two species. Chromatograms demonstrate the presence of double picks in position 150 of An. messeae and in positions 211, 215, and 217 of An. daciae. The specimen with double picks in position 150, 211, 215, 217, 412, and 432 was identified as a hybrid between An. messeae and An. daciae.

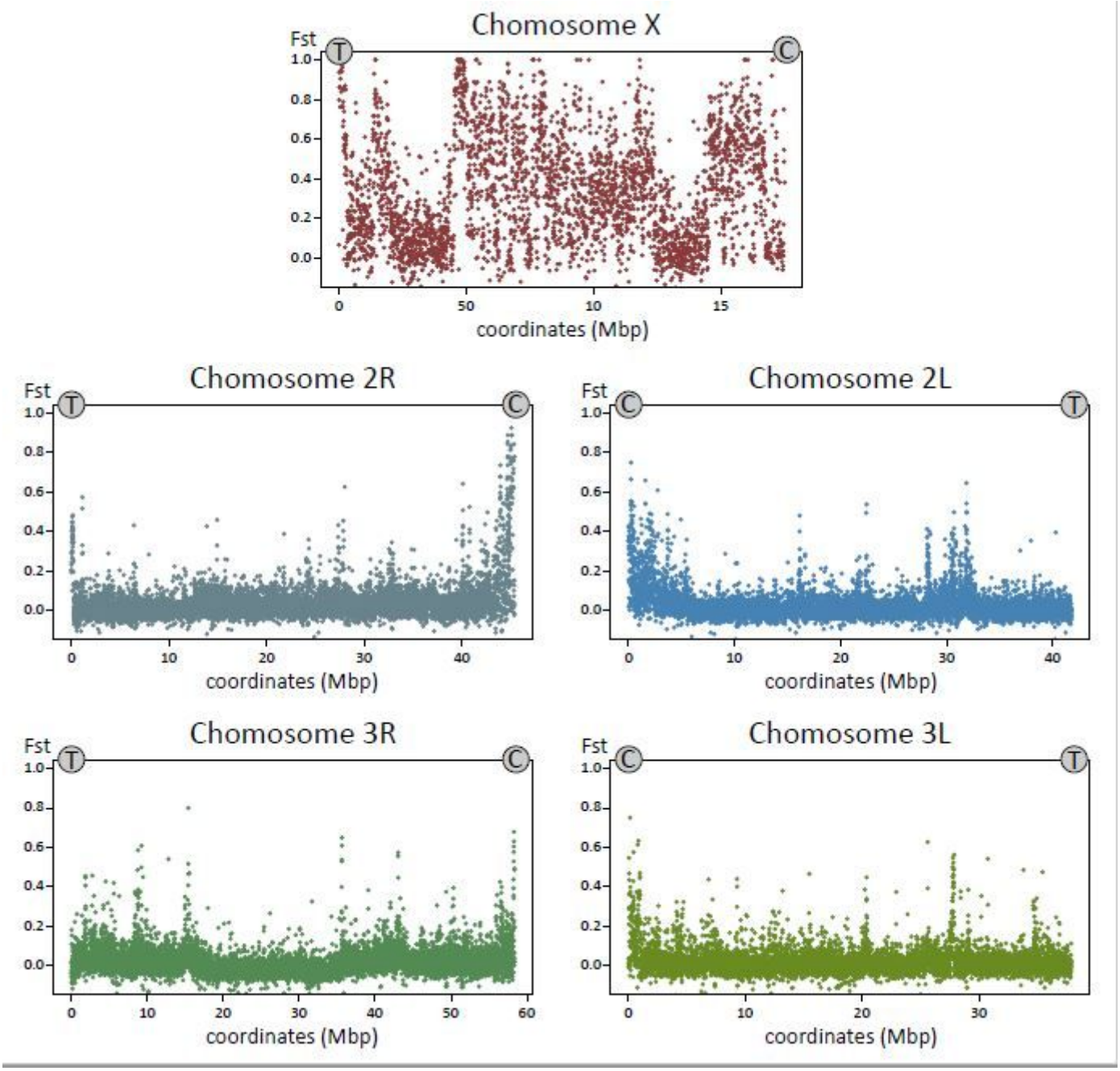

Figure 2

Level of genetic differentiation (Fst) between An. messeae (n=5) and An. daciae (n=5) along the chromosomal arms. Each dot represents a 5 kb window. The x-axis represents Fst values and the y-axis represents the genomic coordinates (bp). The X chromosome demonstrates the highest Fst values while the autosomal arms have a low overall level of differentiation, which is elevated in the centromeric regions. T and C stands for telomeres and centromeres, respectively.

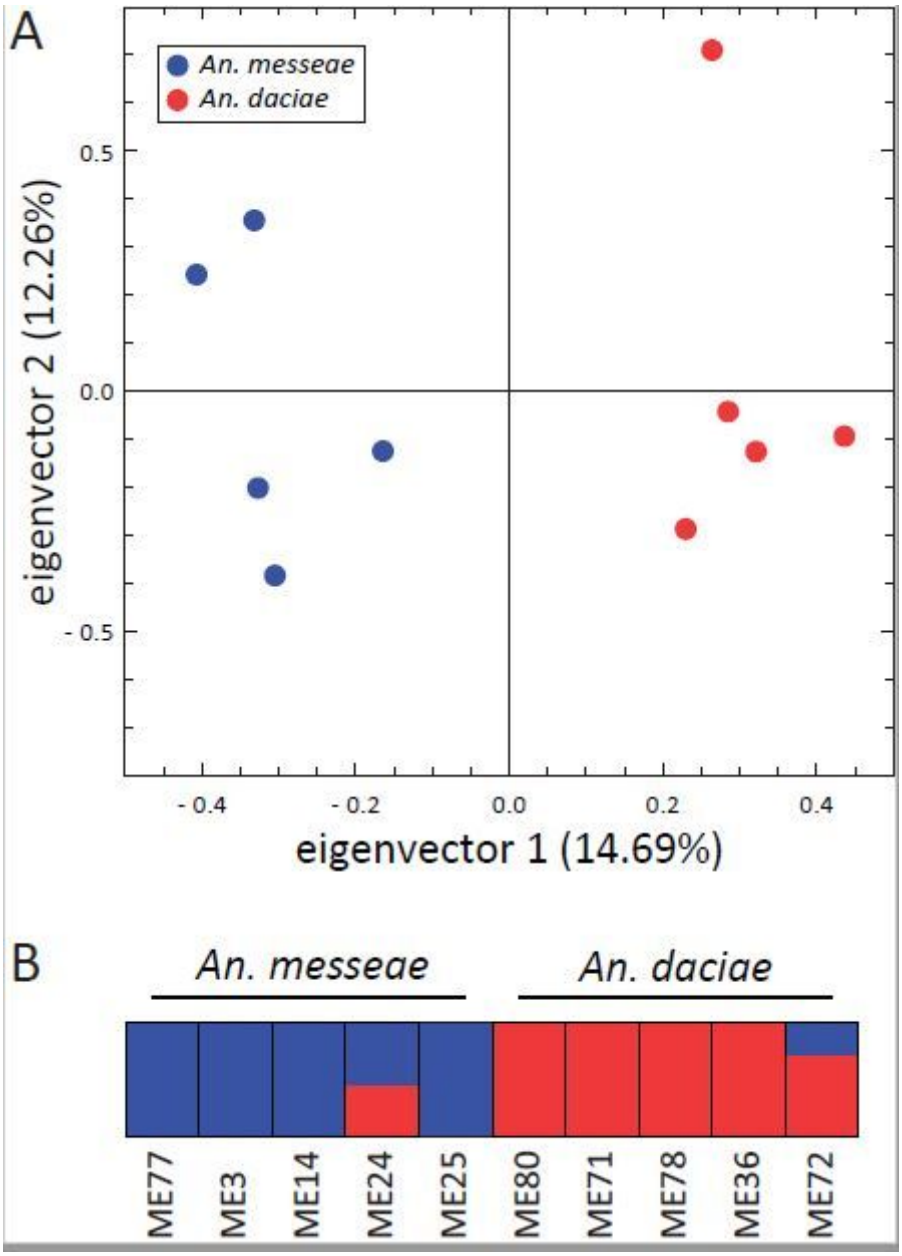

Figure 3

Principal Component Analysis (A) and ADMIXTURE (B) plots based on the autosomal SNVs of the whole genomes of An. messeae and An. daciae. Species are shown by different colors. PCA (A) reliably differentiates two species by the PC1. Each bar (B) represents the proportion of ancestral species in a given individual. Two admixed individuals were identified.

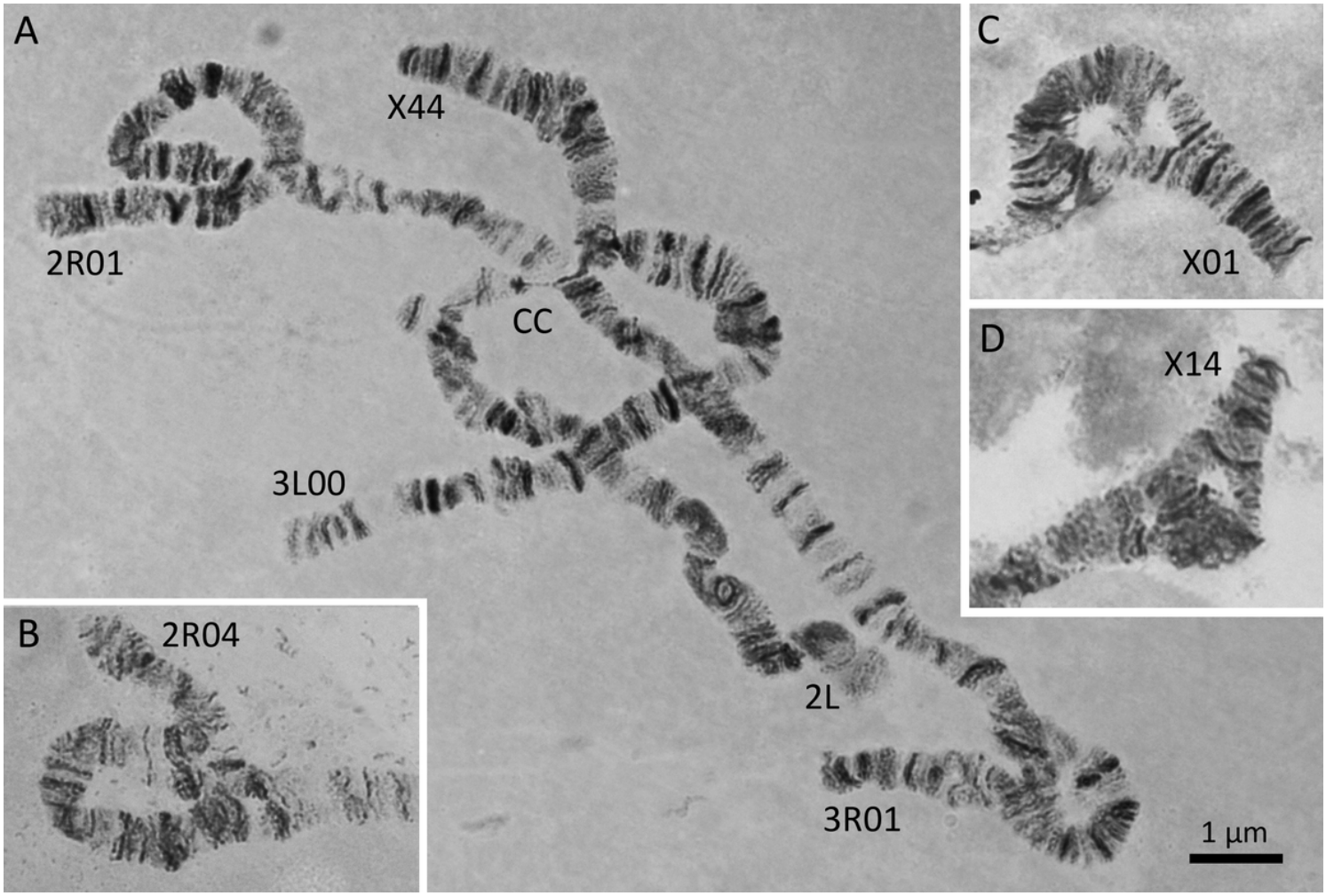

Figure 4

Inversions in polytene chromosomes of An. messeae and An. daciae. The specimens in the Moscow region is characterized by the presence of 4 highly polymorphic inversions X1, 2R1, 3R1, and 3L1, and 2 rare endemic inversions X4 and 2R4*. A rare karyotype X44, 2R01, 3R01, and 3L00 in An. messeae is shown on panel A. Another rare inversion heterozygote, 2R04*, in An. daciae is demonstrated on panels B. Inversion heterozygotes X01 in An. daciae and X14 in An. messeae are indicated on panels C and D, respectively.

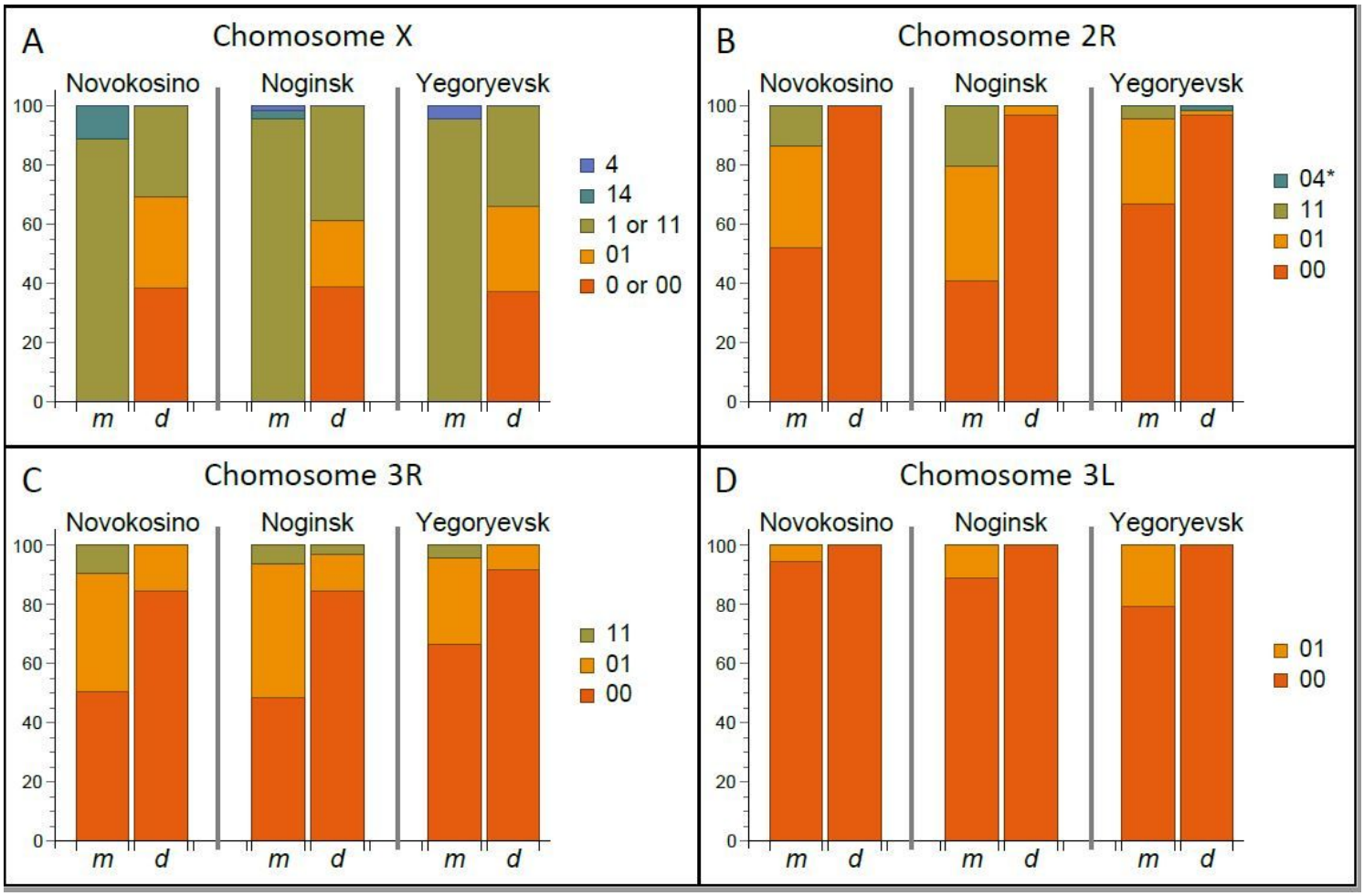

Figure 5

Inversion frequencies in An. messeae and An. daciae from the Novokosino, Noginsk, and Yegoryevsk populations. Frequencies of inversions: X0, X1, and X4 (A); 2R1 and 2R4* (B); 3R1 (C); and 3L1 (D) are shown by charts. Proportions of standard, inverted, and heterozygote arrangements are shown by different colors. Inversion X1 is fixed in An. messeae but is highly polymorphic in An. daciae in the three Moscow populations. Although all autosomal inversions are present in both species, polymorphism is higher in An. messeae than in An. daciae. Rare inversions, X4 and 2R4*, were found in low frequencies in An. messeae and An. daciae, respectively.

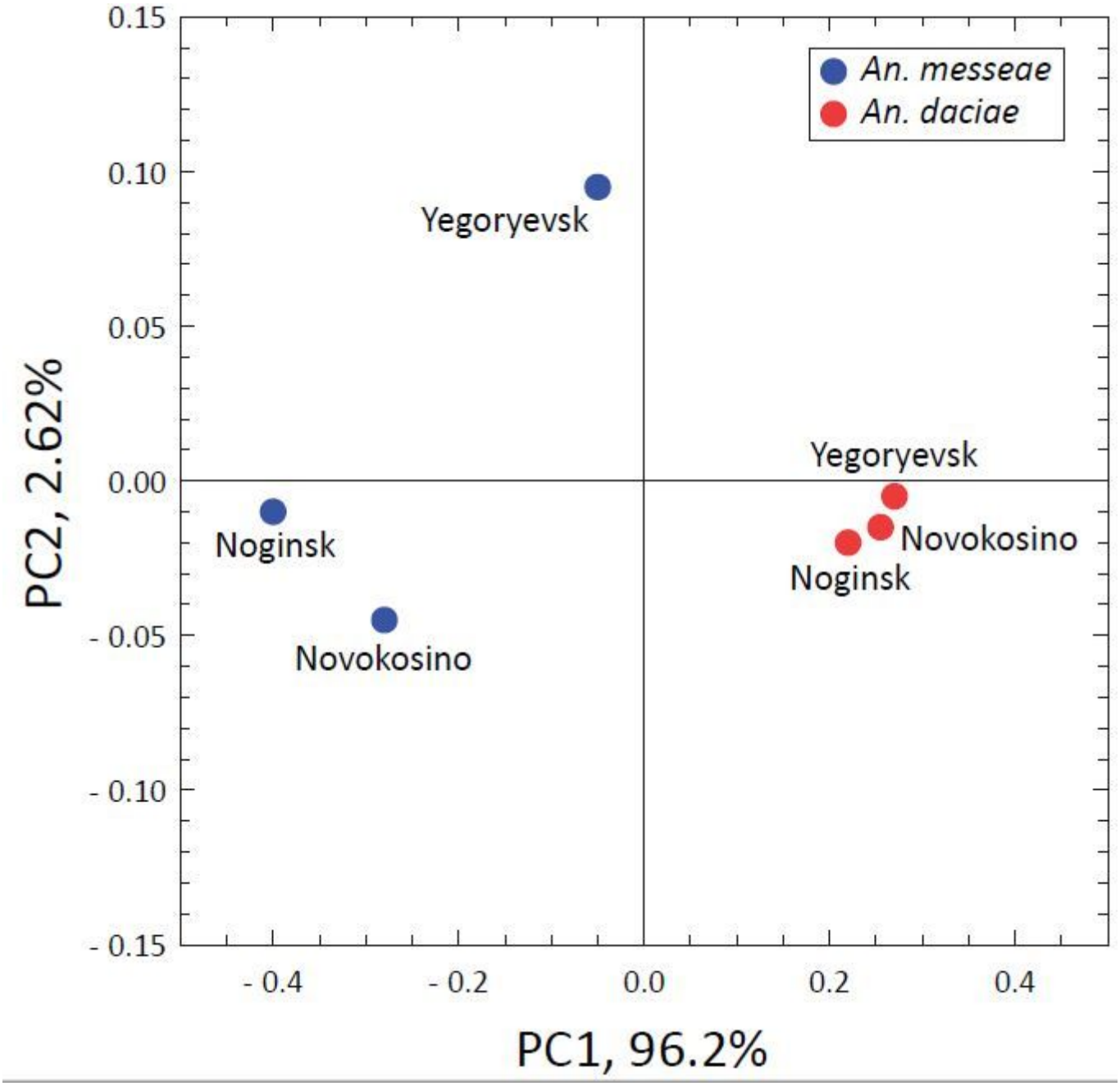

Figure 6

Interpopulation PCA plot based on the frequencies of the autosomal chromosomal inversions in three populations of An. messeae and An. daciae. Species are indicated by different colors. PCA analysis separates An. messeae and An. daciae along the PC1.

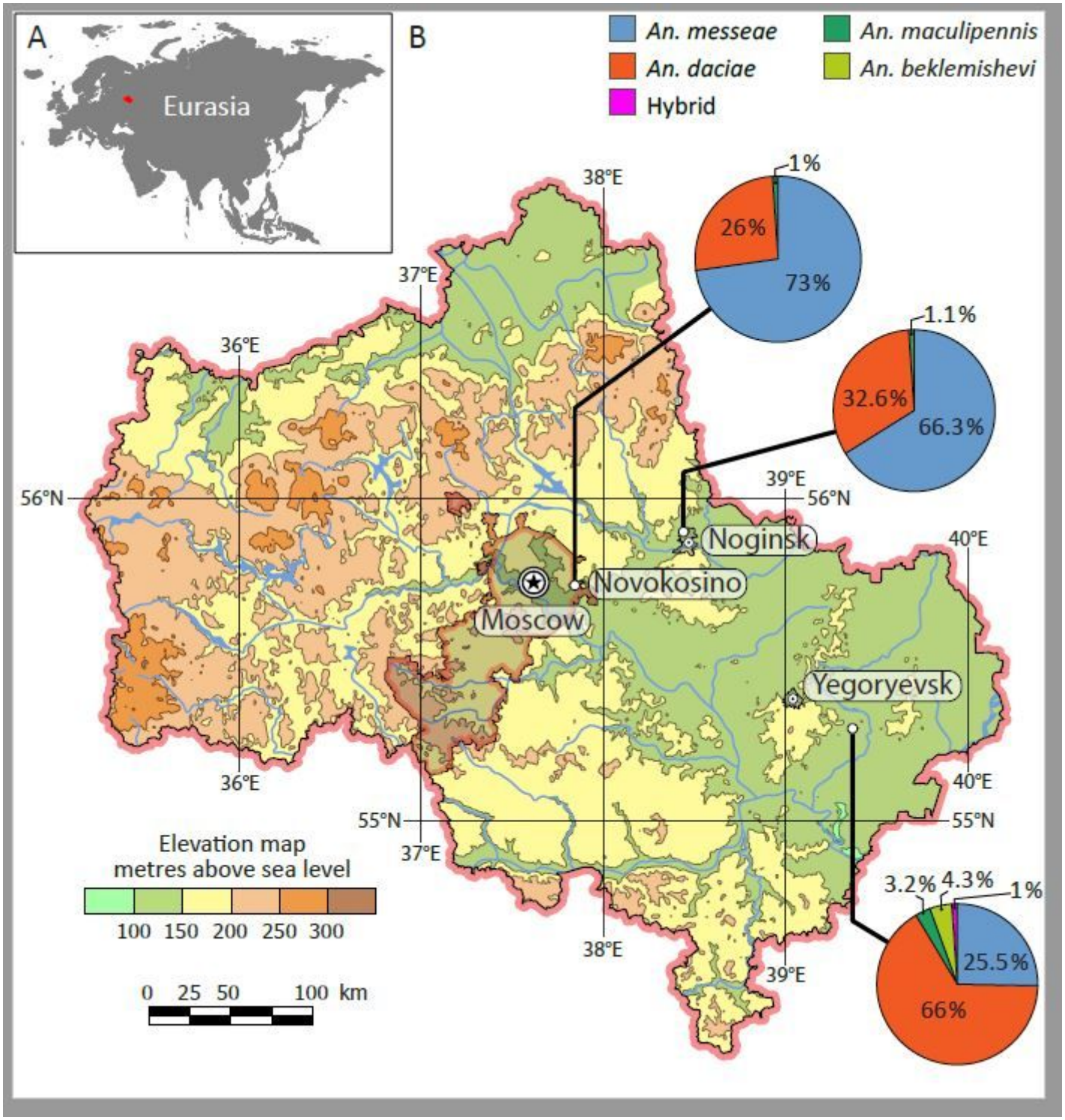

Figure 7

Location of the Moscow region in Eurasia (A) and collection sites of mosquito larvae (B). The ratios of An. messeae, An. daciae, their hybrids, An. maculipennis, and An. beklemishevi are shown as pie charts for each population. The charts demonstrate different proportions of the species in three compared populations. The map was developed using OpenStreetMap software [77].

Figure 8

Mosquito breeding sites in Novoskosino (A), Noginsk (B), and Yegoryevsk (C). Ponds in Novokosino and Noginsk are preferred by An. messeae and represent typical for Anopheles sunny larval breeding sites with open water and abundant vegetation. The water reservoir in Yegorevsk, preferred by An. daciae, is shaded and characterized by high water saprobity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarytableS1.docx