ISSN 1022-7954, Russian Journal of Genetics, 2016, Vol. 52, No. 8, pp. 810–820. © Pleiades Publishing, Inc., 2016.
Original Russian Text © I.A. Antonov, Yu.S. Bukin, 2016, published in Genetika, 2016, Vol. 52, No. 8, pp. 919–930.

ANIMAL GENETICS

Molecular Phylogenetic Analysis of the Ant Genus *Formica* L. (Hymenoptera: Formicidae) from Palearctic Region

I. A. Antonov[a, *] and Yu. S. Bukin[b, c, **]

[a] *Siberian Institute of Plant Physiology and Biochemistry, Siberian Branch, Russian Academy of Sciences, Irkutsk, 664033 Russia*

[b] *Limnological Institute, Siberian Branch, Russian Academy of Sciences, Irkutsk, 664033 Russia*

[c] *Department of Biotechnology and Bioinformatics, Irkutsk National Technical University, Irkutsk, 664074 Russia*

**e-mail: patologi@sifibr.irk.ru, *jfkgsr@mail.ru, **bukinyura@mail.ru*

Received July 22, 2015; in final form, February 3, 2016

Abstract—Sixty-five sequences of the mitochondrial DNA cytochrome *b* gene fragment (759 bp) and 23 sequences of the NADH dehydrogenase subunit 6 gene fragment (224 bp) were compared in ants of the genus *Formica* L. from different regions of the Palearctic and in *Polyergus rufescens* Latr. as outgroup. In total, 28 species of the genus *Formica* were examined. As a result, dated trees with a molecular clock were constructed showing the phylogenetic relationships of *Formica* ants. The topology of the obtained tree based on the *Cyt-b* sequences was found to be not consistent with the generally accepted opinion on the *Formica rufa* and *F. rufibarbis* groups. New data on the formation history of the present-day fauna of *Formica* ants of the Palearctic were obtained. It was demonstrated that a considerable fraction of the examined species (about a third) were formed in the Quaternary Period.

Keywords: strict molecular clock model, Bayesian inference, formation history, present-day fauna

DOI: 10.1134/S1022795416080020

INTRODUCTION

Ants of the genus *Formica*, owing to their relatively large size, the large number of colonies, and nests molded with plant residues, are among the most common insects in the forest ecosystems of the Palearctic. The biomass of these insects in the forests of south taiga zone is only slightly lower than that of the mass of soil animals like earthworms [1]. Most species are active predators, and their specific feature is reactivity; i.e., with the mass reproduction of any invertebrate species edible for them, they completely switch to feed on it [1, 2]. As a result, ants of the genus *Formica* are an important factor in regulating the number of phytophagous insects in forest ecosystems.

Today, there are many controversial issues in the taxonomy of the genus *Formica*, which includes 175 valid species [3]. It is clear that, without proper definition of ant species, it is impossible to study their ecology and biology and to organize an effective system of forest protection in which ants of the genus *Formica* are the central core. The reason for this situation is the existence of sibling species (for example, *F. rufa* and *F. polyctena*), allopatric species (for example, *F. cunicularia* and *F. rufibarbls*, *F. lemani* and *F. fusca*, etc.), and geographical variation and hybridization of these species [1, 4, 5]. One of the solutions is to reveal the relationships between the species. It is now conventional to use the methods of phylogenetic analysis based on the determined sequences of specific molecular marker genes to address various issues of systematics, biogeography, evolutionary biology, ethology, ecology, phylogeography, and population biology. Commonly, phylogenetic analysis makes it possible to gain a deeper insight into the evolutionary relationships of extinct and modern ant taxa [6, 7]. Therefore, in recent years, there were many phylogenetic and population studies of ants using different molecular markers [6–17, etc.]. At the same time, the reconstruction of phylogenetic relationships between the species is a much easier task in the case of existence of a molecular clock in their evolution, and the very possibility of the existence of a molecular clock principle is not in question for most researchers [18].

A phylogenetic study of the ant genus *Formica* was already performed earlier [19]. However, this study was focused on phylogeography and genetic diversity of 20 ant species of the genus *Formica* from Eurasia, taking into consideration the history of the biotic responses to environmental changes in the Quaternary Period [19]. The objective of the present study was to carry out a molecular phylogenetic analysis of *Formica* ants from the Palearctic using the molecular clock for dating the evolutionary events in this genus and to compare the results of phylogenetic analysis with tra-

ditional ideas of systematics based on morphological characters.

The GenBank sequence database was used as a source of information on the determined DNA sequences of ants of the genus *Formica*. It was decided to perform the study using the sequences of the mitochondrial DNA cytochrome *b* (*Cyt-b*) and NADH-dehydrogenase 6 subunit (*NADH-6*) genes, since these sequences are available for the greatest number of species of the investigated genus. The selected genetic markers are often used for species identification and the reconstruction of phylogenetic relationships between different ant taxa [6–11]. In addition, the values of genetic distance (*p* distance) provide insights into the taxonomic status of individuals [5]. Mitochondrial DNA is often used in phylogenetic studies of animals owing to a number of advantages (maternal inheritance, lack of recombination, and simple organization) [19].

MATERIALS AND METHODS

All the sequences used in the analysis were taken from the GenBank, NCBI (http://www.ncbl.nlm.nih.gov/) [20]. The search for genus *Formica* DNA sequences was carried out in the GenBank taxonomic browser. This made it possible to cover the entire spectrum of species of the genus *Formica* presented in the GenBank database. A member of the genus *Polyergus* (*Polyergus rufescens*), as one of the most similar in morphology to the genus *Formica*, was used as an outgroup.

To visualize the alignment and trimming of sequences, the BioEdit software program (version 7.2.5) was used [21]. The alignment itself was carried out using the Clustal W built-in utility of multiple alignment; all parameters were set as the default options. The control of alignment of sequences was carried out by translating them into amino acid sequences using an online service (http://www.ebl.ac.uk/Tools/st/ emboss_transeq/) [22].

The search for unique haplotypes for phylogenetic analysis was conducted with the help of the online application FaBox (utility DNA to haplotype collapser and converter in insert of work with format Fasta) [23].

Testing the hypothesis of strict molecular clock for ants of the genus *Formica* with respect to the outgroup was carried out with the help of the Tajima Relative Rate Test of Molecular Clock [24]. All calculations were performed using the APE 3.0-8 (Analysis of Phylogenetics and Evolution) [25] and PEGAS software packages [26] in the R programming environment [27]. In addition, using the APE 3.0-8 software package, genetic distances between the sequences were calculated.

Selection of the best-fit model of nucleotide substitution for further construction of gene trees was carried out in the jModelTest 2.1.7 software program [28]. Phylogenetic analysis and the reconstruction of phylogenetic trees with the molecular clock model and without it were performed using Bayesian inference as implemented in the MrBayes 3.2 software program [29]. In this case, the models suggested by the jModelTest program served as the substitution accumulation models. To construct the phylogenetic trees, the stability of the resulting reconstruction with respect to the noise data was assessed by estimation of the a posteriori probabilities of nodes on the phylogenetic tree calculated by the MrBayes 3.2 program.

Calibration of the molecular clock was carried out on the basis of the published data [6, 7, 30], according to which the genus *Formica* and the genus *Polyergus* diverged from a common ancestor about 44.1 Ma. The substitution rate in the *Cyt-b* and *NADH-6* markers was calculated on the basis of genetic distances defined using a rooted nonclocklike phylogenetic tree. We calculated the mean length of the branch connecting the *Formica* taxon with a node separating the outgroup from the rest of the tree. The value obtained was divided by the calibration of 44.1 million years, which made it possible to estimate the mean substitution rate per million years. Calculations were carried out in the APE 3.0-8 and PHYLOCH software packages [31] for the R programming language. The 95% confidence interval for the mean substitution rate was evaluated with the help of bootstrap analysis implemented in the BOOT package [32] of the R programming language. Visualization of phylogenetic trees and export of the resulting images in vector graphics format was performed using the FigTree 1.4.0 software program.

RESULTS

The search for partial *Cyt-b* and *NADH-6* sequences of ants in the GenBank database yielded 224 such sequences for the *Cyt-b* fragments and 126 sequences for the *NADH-6* fragments belonging to 31 *Formica* species. After the alignment of the set of *Cyt-b* gene fragments, the overlapping region constituted 759 bp for all sequences. It is this fragment of the *Cyt-b* gene that was used for further analysis. For the *NADH-6* gene fragments, the overlapping region after the alignment constituted 224 bp. This *NADH-6* gene fragment was used for further analysis. Translation of the nucleotide sequences in the amino acid sequences indicated that both *Cyt-b*, and *NADH-6* gene fragments encoded amino acid sequences that matched the fragments of the Cyt-b and NADH-6 proteins.

At the next stage, from the whole dataset of the *Cyt-b* and *NADH-6* sequences, the redundant sequences were removed using the FaBox online service. Thus, for further analysis, only the unique haplotypes were used, which reduced the amount of computations.

With the script written in the R programming language and using the PEGAS software package, for the whole sequence dataset of the *Cyt-b* and *NADH-6* fragments, the conformity to the strict molecular

clock model was tested using the Tajima Relative Rate Test of Molecular Clock. In the script, the hypothesis was tested for each combination of sequences with the *Polyergus rufescens* outgroup. The null hypothesis of conformity to the molecular clock model of a pair of compared sequences was accepted upon its calculated probability of $P_{val} > 0.05$. Thus, it was demonstrated that, out of the whole set of sequences containing 31 species of the genus *Formica*, only 28 species were consistent with the hypothesis of strict molecular clock and could be used for phylogenetic reconstructions considering this hypothesis. A list of these species with the accession numbers of nucleotide sequences of the *Cyt-b* and *NADH-6* fragments is presented in the table.

Testing of the *Cyt-b* and *NADH-6* sequence datasets in the jModelTest 2.1.7 software program showed that, taking into account the AIC (Akaike information criterion) values, the best-fit substitution model for the *Cyt-b* dataset was the GTR+I+G model, and for the *NADH-6* dataset, the GTR+G model [33].

Taking into the account the selected nucleotide substitution models, the phylogenetic reconstruction of the evolutionary history with application of the strict molecular clock model was performed in the MrBayes 3.2 software program. The calculations for both the *Cyt-b* and *NADH-6* sequence datasets were performed with the construction of 50000000 Markov chain generations. The analysis was performed in two parallel independent runs. To build a consensus phylogenetic tree, trees were saved every 1000th generation. The process of Markov chain generation was stopped automatically when the level of differences between the trees of two independent parallel trials reached the level of <0.008. In the process of calculation, the Bayesian inference was automatically stopped for the *Cyt-b* marker after 8940000 Markov chain generations and for the *NADH-6* marker after 16890000 generations. The convergence in Bayesian phylogenetic inferences was additionally controlled by the analysis of minimum values of ESS statistics for the substitution model parameters in two independent trials. During the phylogenetic analysis, minimum ESS values for all parameters of the substitution models for both the *Cyt-b* and *NADH-6* gene fragments were higher than 200, as recommended by the manual to the MrBayes 3.2 software program. The phylogenetic trees with calibrated molecular clock for the *Cyt-b* and *NADH-6* gene fragments are shown in Figs. 1 and 2.

Comparison of the phylogenetic trees revealed considerable differences in their topologies. For example, on the tree based on the *Cyt-b* sequences, the subgenus *Formica* s. str. formed an independent cluster at the dating point of 17 Ma, while other subgenera formed a common cluster. On the tree formed on the basis of the *NADH-6* sequences, at the dating point of 17 Ma, most species of the subgenus *Serviformica* formed an independent cluster, while other subgenera and the species *F. uralensis* from the subgenus *Serviformica* formed a common cluster. Interestingly, on the tree formed on the basis of the *Cyt-b* sequences, at the dating point of 14 Ma, almost all subgenera formed independent clusters. The exception was the subgenus *Serviformica*, which formed two clusters. Therefore, at this time, the modern subgeneric structure of the genus *Formica* was formed, if we take into account only the *Cyt-b* marker. At the same time, the tree topology based on the *NADH-6* sequences at the dating point of 14 Ma was quite different. Specifically, the subgenera *Formica* s. str. and *Coptoformica* formed a common cluster, while the main part of the species of the subgenus S*erviformica* divided into two clusters, and the subgenus *Raptiformica* (*F. sanguinea*) formed a common cluster with *F. uralensis* from the genus *Serviformica*. It should be noted that statistical support for most nodes was low (below 70%).

Interestingly, the obtained tree topologies mainly coincided with the taxonomic structure of the genus *Formica* based on the study of morphological characters of ants [1]. On the trees, all species are located in their subgeneric clusters, except for *F. uralensis* (Fig. 2). However, not all branches of the species are statistically well supported (more than 80%). This mainly refers to *F. lugubris* and *F. pratensis* (Figs. 1 and 2).

The trees with calibrated molecular clock show the approximate time of the appearance of the first representatives of the subgenera, species, and populations (Figs. 1, 2). Thus, the first member of the subgenus *Serviformica* probably appeared at 16.1–20.7 Ma; of the subgenus *Formica* s. str., at 13.9–17.5 Ma; of the subgenus *Coptoformica*, at 13.9–14.5 Ma; and of the subgenus *Raptiformica*, at 11.2–14.5 Ma. However, because of weak statistical support of the majority of branch nodes (below 70%), many of these dates can be used only as a guide for further investigations.

In order to estimate the substitution rate in the *Cyt-b* and *NADN-6* markers, the phylogenetic trees were reconstructed in the MrBayes 3.2 software program without enforcing the molecular clock and using the substitution models suggested by the jModelTest 2.1.7 software program (GTR+I+G for the *Cyt-b* and GTR+G for the *NADH-6*). Setting the parameters for the Markov chain modeling and evaluating the convergence of the results of phylogenetic analysis were performed as in the reconstruction of the tree with molecular clock. Each tree after reconstruction was rooted at the *Polyergus rufescens* outgroup. Then, for the phylogenetic trees reconstructed on the basis of the *Cyt-b* and *NADH-6* gene fragments, the sample of the distances from the tree root to its leaf was determined. These samples define the variability of the genetic distances accumulated by the *Cyt-b* and *NADH-6* markers of *Formica* ants since the separation from a common ancestor with the genus *Polyergus*. The genetic distances divided by the calibration of 44.1 million years give the samples of variability of the substitution rates per million years in the *Cyt-b* and

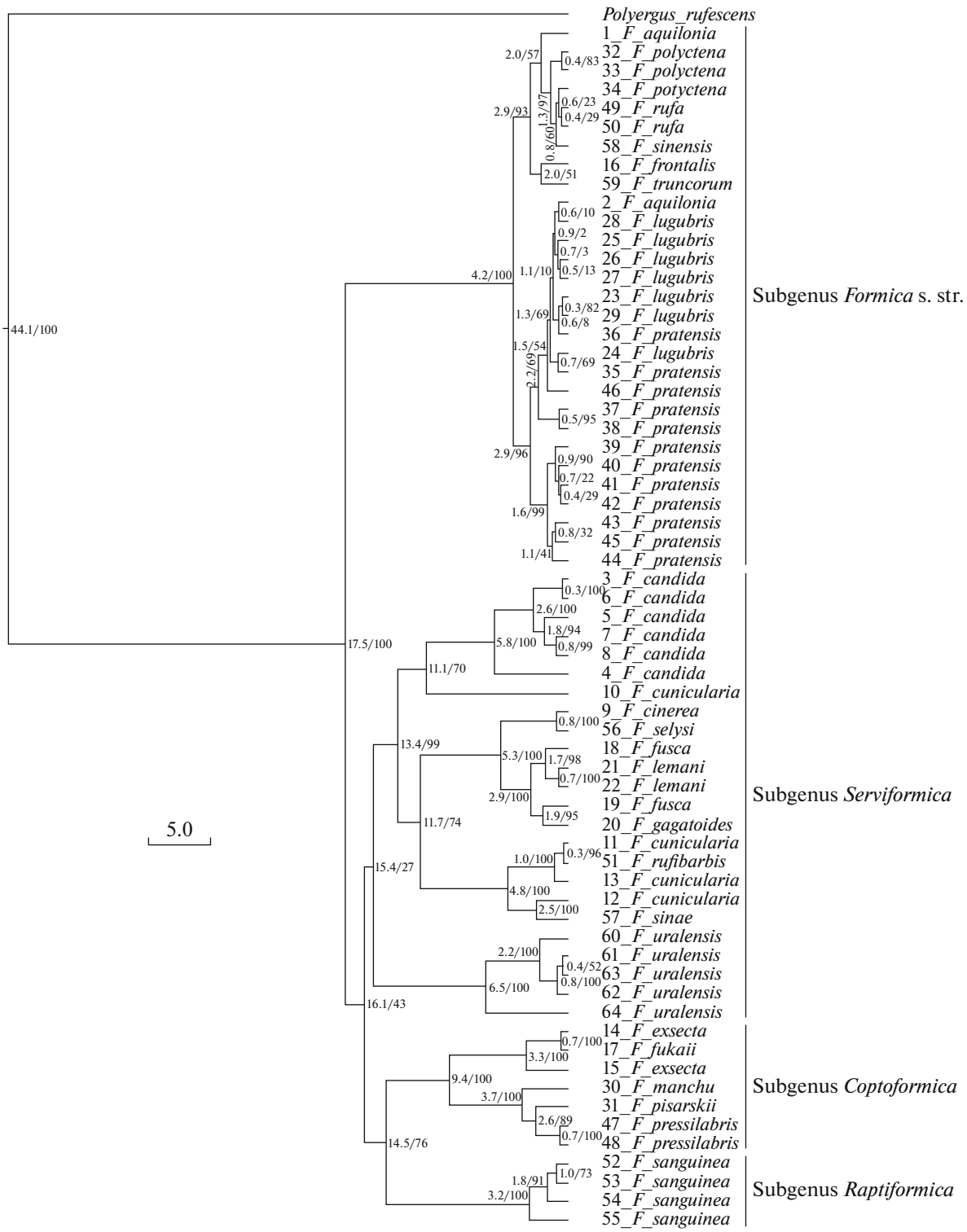

Fig. 1. Rooted tree with calibrated molecular clock showing the phylogenetic relationships of *Formic*a ants based on the 65 *Cyt-b* gene sequences. At the nodes are sequentially shown the dating points in million years and statistical support in percent. The scale interval corresponds to five million years.

NADH-6 markers of *Formica* ants. The calculations showed that the mean substitution rate in the *Cyt-b* marker was 0.0184 substitutions per million years with the 95% confidence interval from 0.0182 to 0.0185; in the *NADH-6* marker, it was 0.0087 substitutions per million years with the 95% confidence interval from 0.0084 to 0.0090. The maximum genetic distance (*p*) for the *Cyt-b* gene fragment was 0.198; the estimate

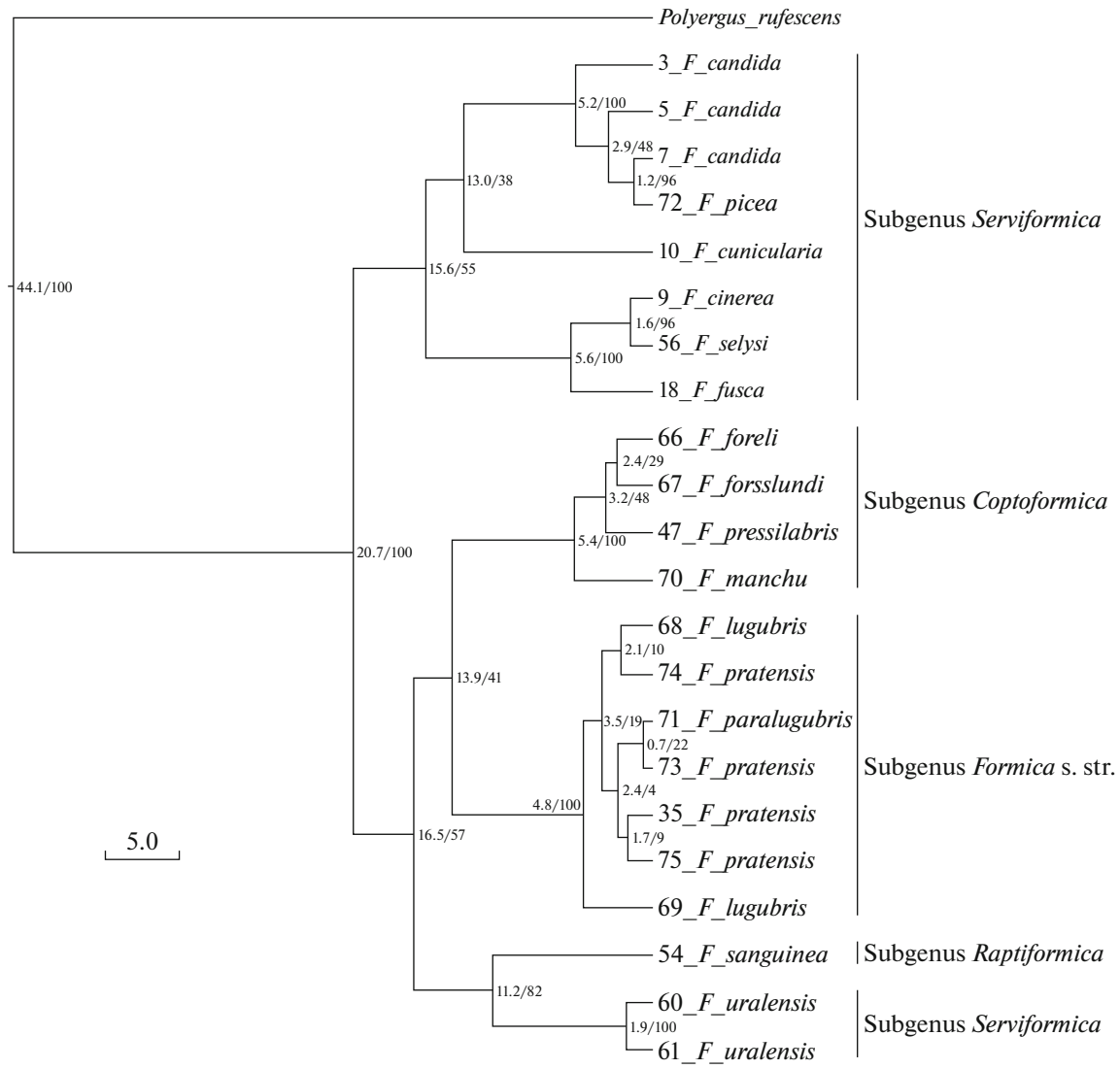

Fig. 2. Rooted tree with calibrated molecular clock showing the phylogenetic relationships of *Formica* ants based on the 23 *NADH-6* gene sequences. At the nodes are sequentially shown the dating points in million years and statistical support in percent. The scale interval corresponds to five million years.

using Kimura's two parameter model [34] was 0.232. For the *NADH-6* gene fragment, the maximum genetic distance was 0.175; the estimate using Kimura's two parameter model was 0.2.

DISCUSSION

From the published data, it follows that mitochondrial genetic markers are characterized by a certain limit of genetic distances at which substitution saturation of the compared sequences occurs. Analysis of the *Cyt-b* genetic marker in cichlid fishes [35] showed that the limit of genetic distances at which there was substitution saturation was 0.468 substitutions. In our case, the maximum distance (*p*) and the distance estimated by Kimura's two parameter model are less than this limit. This means that the dating estimates of the phylogenetic events and the tree topology are not distorted by the effect of substitution saturation of the compared sequences.

The amino acid sequence of the protein encoded by the *NADH 6* gene is less conservative than the amino acid sequence of the Cyt-b protein [36]. However, our results on the substitution rate (0.0184 substitutions per million years for the *Cyt-b* and 0.0087 substitutions per million years for the *NADN-6*) seem to contradict this fact. The size of the investigated *NADH-6* gene fragment constituted 224 bp, which was less than 50% of the full-length sequence (552 bp) of this gene in the mitochondrial genome of the *Formica* ants [37]. According to the literature [37], as well as our data obtained with the help of the jModelTest software pro-

List of species used in the study with the GenBank accession numbers, molecular markers, and the collection localities

No.	Species name	GenBank accession number	Molecular marker	Collection locality
1	*Formica aquilonia* Yarr.	HQ651085.1	*Cyt-b*	China: Hebei Province
2	*F. aquilonia* Yarr.	AY488781.1	*Cyt-b*	Finland: Riisitunturi
3	*F. candida* Smith	JX170887.1	*Cyt-b* and *NADH-6*	Kyrgyzstan: Alay Valley
4	*F. candida* Smith	HQ651081.1	*Cyt-b*	China: Hebei Province
5	*F. candida* Smith	AY786157.1	*Cyt-b* and *NADH-6*	China: Qinghai, Tibet
6	*F. candida* Smith	AY786153.1	*Cyt-b*	Kyrgyzstan: Alay Valley
7	*F. candida* Smith	AY786149.1	*Cyt-b* and *NADH-6*	Russia: Moscow oblast
8	*F. candida* Smith	AY786146.1	*Cyt-b*	Finland: outskirts of Helsinki
9	*F. cinerea* Mayr	JX170884.1	*Cyt-b* and *NADH-6*	Sweden: Skane County
10	*F. cunicularia* Latr.	JX170885.1	*Cyt-b* and *NADH-6*	Russia: outskirts of Novosibirsk
11	*F. cunicularia* Latr.	HQ651084.1	*Cyt-b*	China: Hebei Province
12	*F. cunicularia* Latr.	HQ651079.1	*Cyt-b*	China: Shandong Province
13	*F. cunicularia* Latr.	HQ651075.1	*Cyt-b*	China: Shandong Province
14	*F. exsecta* Nyl.	JX170868.1	*Cyt-b*	China: Qinghai, Tibet
15	*F. exsecta* Nyl.	JX170867.1	*Cyt-b*	Germany
16	*F. frontalis* Sant.	AY488791.1	*Cyt-b*	Spain: outskirts of Leon
17	*F. fukaii* Wheel.	HQ651088.1	*Cyt-b*	China: Ningxia Province
18	*F. fusca* L.	JX170888.1	*Cyt-b* and *NADH-6*	Sweden: outskirts of Tierp
19	*F. fusca* L.	HQ651077.1	*Cyt-b*	China: Ningxia Province
20	*F. gagatoides* Ruzs.	HQ651073.1	*Cyt-b*	China: Ningxia Province
21	*F. lemani* Bondr.	HQ651086.1	*Cyt-b*	China
22	*F. lemani* Bondr.	HQ651082.1	*Cyt-b*	China: Hebei Province
23	*F. lugubris* Zett.	AY573885.1	*Cyt-b*	Switzerland
24	*F. lugubris* Zett.	AY573872.1	*Cyt-b*	Russia: Kamchatka
25	*F. lugubris* Zett.	AY573863.1	*Cyt-b*	Russia: Altai
26	*F. lugubris* Zett.	AY573861.1	*Cyt-b*	Sweden: outskirts of Ahus
27	*F. lugubris* Zett.	AY573859.1	*Cyt-b*	Sweden: outskirts Uppsala
28	*F. lugubris* Zett.	AY573857.1	*Cyt-b*	Pyrenees
29	*F. lugubris* Zett.	DQ836181.1	*Cyt-b*	Scotland
30	*F. manchu* Wheel.	JX170874.1	*Cyt-b*	Russia: West Transbaikal region
31	*F. pisarskii* Dluss	JX170876.1	*Cyt-b*	Russia: West Transbaikal region
32	*F. polyctena* Foerst.	AY517508.1	*Cyt-b*	Russia: outskirts of Novosibirsk
33	*F. polyctena* Foerst.	AY488763.1	*Cyt-b*	Russia: outskirts of Severobaikalsk
34	*F. polyctena* Foerst.	AY488760.1	*Cyt-b*	Denmark: Jutland
35	*F. pratensis* Retz.	AY584232.1	*Cyt-b* and *NADH-6*	Pyrenees
36	*F. pratensis* Retz.	AY584227.1	*Cyt-b*	Russia: outskirts of Snezhinsk, Chelyabinsk oblast
37	*F. pratensis* Retz.	AY584220.1	*Cyt-b*	Russia: outskirts of Novosibirsk
38	*F. pratensis* Retz.	AY584219.1	*Cyt-b*	Sweden: outskirts of Uppsala
39	*F. pratensis* Retz.	AY584210.1	*Cyt-b*	Russia: outskirts of Krasnoyarsk
40	*F. pratensis* Retz.	AY584208.1	*Cyt-b*	Russia: outskirts of Tyumen
41	*F. pratensis* Retz.	AY584207.1	*Cyt-b*	Russia: outskirts of Tyumen
42	*F. pratensis* Retz.	AY584206.1	*Cyt-b*	Russia: Sverdlovsk oblast

Table. (Contd.)

No.	Species name	GenBank accession number	Molecular marker	Collection locality
43	*F. pratensis* Retz.	AY584205.1	*Cyt-b*	Russia: Altai
44	*F. pratensis* Retz.	AY584198.1	*Cyt-b*	Russia: Altai
45	*F. pratensis* Retz.	AY584197.1	*Cyt-b*	Finland: outskirts of Hanko
46	*F. pratensis* Retz.	AY604525.1	*Cyt-b*	Russia: Sverdlovsk oblast
47	*F. pressilabris* Nyl.	JX170872.1	*Cyt-b* and *NADH-6*	Russia: Perm oblast
48	*F. pressilabris* Nyl.	JX170871.1	*Cyt-b*	Russia: outskirts of Yekaterinburg
49	*F. rufa* L.	AY488766.1	*Cyt-b*	Russia: outskirts of Novosibirsk
50	*F. rufa* L.	AY488765.1	*Cyt-b*	Spain
51	*F. rufibarbis* Fabr.	JX170889.1	*Cyt-b*	Sweden
52	*F. sanguinea* Latr.	JX170892.1	*Cyt-b*	Russia: outskirts of Magadan
53	*F. sanguinea* Latr.	JX170891.1	*Cyt-b*	Sweden: outskirts of Uppsala
54	*F. sanguinea* Latr.	JX170890.1	*Cyt-b* and *NADH-6*	Spain: outskirts of Leon
55	*F. sanguinea* Latr.	HQ651087.1	*Cyt-b*	China: Ningxia Province
56	*F. selysi* Bondr.	JX170883.1	*Cyt-b* and *NADH-6*	Switzerland
57	*F. sinae* Em.*	HQ651071.1	*Cyt-b*	China: Shandong Province
58	*F. sinensis* Wheel.	HQ651083.1	*Cyt-b*	China: Hebei Province
59	*F. truncorum* Fabr.	AY488789.1	*Cyt-b*	Sweden: outskirts of Uppsala
60	*F. uralensis* Ruzs.	JX170881.1	*Cyt-b* and *NADH-6*	South of Mongolia
61	*F. uralensis* Ruzs.	JX170880.1	*Cyt-b* and *NADH-6*	Russia: outskirts of Yekaterinburg
62	*F. uralensis* Ruzs.	JX170879.1	*Cyt-b*	Germany
63	*F. uralensis* Ruzs.	JX170878.1	*Cyt-b*	Finland
64	*F. uralensis* Ruz.	HQ651080.1	*Cyt-b*	China: Hebei Province
65	*Polyergus rufescens* Latr.	JX170870.1	*Cyt-b* and *NADH-6*	Germany
66	*Formica foreli* Bondr.	JX170873.1	*NADH-6*	Sweden: Oland
67	*F. forsslundi* Lohm.	JX170877.1	*NADH-6*	South of Mongolia
68	*F. lugubris* Zett.	AY573869.1	*NADH-6*	United Kingdom: District Sheffield
69	*F. lugubris* Zett.	AY573868.1	*NADH-6*	United Kingdom: District Sheffield
70	*F. manchu* Wheel.	JX170875.1	*NADH-6*	Russia: Zabaykalsky krai
71	*F. paralugubris* Seifert	EU600792.1	*NADH-6*	Alps
72	*F. picea* Nyl.	JX170886.1	*NADH-6*	Sweden: Värmland County
73	*F. pratensis* Retz.	AY584221.1	*NADH-6*	Russia: Novosibirsk oblast
74	*F. pratensis* Retz.	AY584218.1	*NADH-6*	Russia: Sverdlovsk oblast
75	*F. pratensis* Retz.	AY584214.1	*NADH-6*	Finland: outskirts of Hanko

* Became a subspecies (*Formica clara sinae* Em.) [3].

gram, the examined fragment of the *NADN-6* gene is characterized by extremely depleted nucleotide composition. According to the data obtained with the jModelTest software program, the analyzed sequences were characterized by the highest occurrence of thymine nucleotide (44.7%), followed by adenine (35.8%), cytosine (15.8%), and guanine (≈3.7%). The nucleotide composition of the tested *NADH-6* fragment is determined by the codon composition of the sequence that determines the amino acid composition of the encoded polypeptide. For the sequences with uniform representation of A, T, G, or C nucleotides, the *p* distance substitution saturation occurs at its value of about 0.75; at a given threshold of *p* values, all further substitutions will be reversible. For the coding markers, the *p* distance saturation effect occurs before the threshold of 0.75 (for example, the aforementioned threshold of 0.468 for the *Cyt-b*). In the case of

depletion of the nucleotide composition, the substitution saturation effect for the noncoding sequence will occur much earlier than the threshold of 0.75. For example, when comparing two sequences consisting of two types of nucleotides, the limiting *p* distance is equal to 0.5. For the coding sequence, the combination of the nucleotide composition depletion with the codon structure can lead to a sharp decrease in the threshold *p* value for the substitution saturation limit. Most probably, this situation is observed for the examined *NADH-6* gene fragment within a time interval encompassing the evolutionary history of the *Formica* ants (≈40 million years). Within the investigated period, the substitution saturation of the *NADH-6* gene fragment occurred, resulting in the downward bias of the calculated substitution rate parameters. The substitution saturation of the examined *NADH-6* gene fragment occurred in a time interval of less than 40 million years. Later on, the substitutions occurred, but they were already reversal and did not increase the genetic distance. As the experience shows [35], even the use of complex substitution saturation models cannot prevent the biases introduced by the substitution saturation effect in the calculated genetic distances. It is possible that, at shorter time intervals (≪40 million years), owing to a higher rate of the NADH-6 protein evolution, the analyzed nucleotide fragment will give adequate values of linearized genetic distances calculated from one of the substitution saturation models.

The substitution saturation effect of the investigated *NADH-6* gene fragment is also observed in the analysis of phylogenetic trees (Figs. 1, 2). On the tree reconstructed on the basis of the *NADH-6* marker, division of the genus *Formica* into species is observed at the dating point of 20.7 Ma, and on the tree reconstructed on the basis of the *Cyt-b* marker, it is at the dating point of 17.5 Ma. The increase in the time of the species separation on the *NADH-6* tree results from the effect of the shortening of the branch connecting the root of the tree to the point of the species separation owing to the influence of the substitution saturation effect on the calculated genetic distances.

In insects, the mean substitution rate in the mitochondrial genome is 0.012 substitutions per million years. The most slowly evolving marker of insect mitochondrial DNA is the 16S ribosomal RNA [38] with the substitution rate of 0.005 per million years. The closest with respect to evolutionary properties to the *Cyt-b* mitochondrial marker is the gene for the cytochrome c-oxidase subunit 1 with the mean substitution rate in insects of 0.016 substitutions per million years. For the *Formica* ants, the calculated mean substitution rate in the *Cyt-b* marker was 0.0184 substitutions per million years, which is consistent with the published data for insects.

As was already mentioned, the topologies of the obtained trees mostly coincided with the taxonomic structure of the genus *Formica* based on the analysis of the morphological characters of ants, excluding *F. uralensis*. These results are also supported by the published data [19]. However, the tree topology based on the *NADH-6* sequences points to insufficient informativeness of the used 224-bp sequences for the resolution of all branches and thus, for the determination of the relationships of many taxonomic groups studied. Many nodes of subgeneric and specific branches are weakly statistically supported (less than 70%). Grouping of *F. uralensis* and *F. sanguinea* in one cluster at the dating point of 12 Ma should be considered as an unsatisfactory result of the *NADH-6* sequence analysis (Fig. 2). In our opinion, the reasons for this deviation are as follows. Previously, myrmecologists assigned *F. uralensis* to the subgenus *Formica* s. str. on the basis of external similarity of worker ants of this species to the worker ants of *Formica* s. str. and the nest construction similar to *F. pratensis*. However, comparison of the males of *F. uralensis* with the males of *Formica* s. str. and *Serviformica* showed that they were considerably distant from the males of *Formica* s. str., but at the same time very similar to the males of some *Serviformica* species. In this case, the *Formica* males are the most conservative and slow changing cast [1]. Therefore, this species stands out in the subgenus *Serviformica*, as evidenced by the data obtained. The subgenus *Serviformica* itself is central in the genus *Formica*, because one cannot find a single character at which it is completely different from the other subgenera [1]. On the contrary, many of the results of the analysis of 65 sequences of the 759-bp *Cyt-b* fragment are supported by the published data (for example, *F. uralensis*) based on paleontological research and the studies of morphological characters of ants. It should be noted in this respect that the inclusion in the analysis of a number of molecular markers increases the number of species studied and complements the data on the relationships of some taxa.

The results of the present study are not consistent with the data of morphological investigations. First, the attention is drawn to *F. aquilonia*. On the tree based on the *Cyt-b* sequences, the individuals of this species are found in two different clusters (one with *F. rufa*, *F. polyctena*, *F. sinensis*, *F. frontalis*, and *F. truncorum*, the other with *F. lugubris* and *F. pratensis*) at the dating point of 4 Ma. In our opinion, there are two possible explanations for this situation. First, the study was carried out with participation of the hybrids of different ant species of the *Formica rufa* group, which are often found in mixed colonies [4, 30]. Second, the specimens were taken from the limits of the species range (Riisitunturi, Finland, and Hebei Province, China), which may indicate great interpopulation differences with respect to the *Cyt-b* gene fragment in *F. aquilonia*. Moreover, in *F. aquilonia*, there are two forms (possibly, two different species). One of the forms with respect to the chaetotaxy patterns is more similar to *F. rufa*, while the other form is similar to *F. polyctena* [1]. The situation with the *Formica rufa* group deserves

special attention. First, the meaning of the term species group should be considered. The species group (superspecies) is the monophyletic group of very close and largely or completely allopatric species [39]. G.M. Dlussky used this term for convenience in the description, because the species belonging to the same group are very similar, and there are no special group names; the groups are named by the main species (*Formica rufa* group, *Formica exsecta* group, etc.) [1]. Thus, the *Formica rufa* group includes the following species investigated in the present study (*Cyt-b* molecular marker): *F. aquilonia*, *F. lugubris*, *F polyctena*, and *F. rufa*. However, on the *Cyt-b* tree, the members of these four species, instead of forming a single cluster, are found in a number of clusters with the other species of the subgenus *Formica* s. str. Therefore, the obtained *Cyt-b* tree topology does not conform to the generally accepted views of the *Formica rufa* group. The *F. cunicularia* ants also deserve special interest. It is well known that this species belongs to the *Formica rufibarbis* group [40]. At the same time, on the *Cyt-b* tree, the sequences of the representatives of this species are found not only in the cluster of the *Formica rufibarbis* group (*F. rufibarbis* and *F. clara sinae* (*F. sinae*)) but also in the cluster with *F. candida*. Analysis of the geographic distribution of the *F. cunicularia* individuals showed that this uncertainty is associated with the specimen from Russia (outskirts of the city of Novosibirsk). It seems likely that, in this case, either there are two different species, or the informativeness of the *Cyt-b* fragment is insufficient to make an adequate conclusion. In summary, the resulting *Cyt-b* tree topology does not conform to the generally accepted views on the *Formica rufibarbis* group.

The interpopulation genetic variability of *F. candida* and *F. uralensis* deserves special interest, since the *Cyt-b* tree shows that the specimens from Chinese populations of these species are considerably distant from other representatives of the same species. These differences are considerably greater than the intraspecific variability of other species investigated. Let us consider each species separately. Analysis of the geographical distribution of the *F. candida* individuals shows that, in addition to the ant specimen from Hebei Province (east of China), an individual from Qinghai Province (west of the central part of China) was also included in the study. Moreover, the sequence of the individual from Qinghai Province forms a tight cluster with the specimens from the populations of Russia and Finland, while the sequence of the ant from Hebei Province is far distant from them (the point of divergence at 5.8 Ma). Apparently, these are ants from two different but morphologically close species (for instance, *F. candida* and *F. picea*). Distinguishing between these species requires multiple measurements and statistical treatment of a relatively large number of ant individuals. In turn, the analysis of the collection localities of *F. uralensis* shows that the sequence of the ant from Hebei Province (east of China) is very different from others and even from the ant from the population from the south of Mongolia. However, *F. uralensis* is a good species, which differs from the others in a set of characters [1]. Most likely, the Chinese populations of *F. uralensis* were isolated in the Late Miocene (the point of divergence at 6.5 Ma) because of the landscape transformations caused by climate changes (temperature fall and drying out) at that time [41]. Although the sequence of the ant from China is considerably different from the others, morphologically this individual is similar to them. Moreover, *F. uralensis* is a remnant of the ancient Palearctic fauna [1], which is also supported by our data (the point of divergence point at 15.4 Ma) (Fig. 1).

The new data on the formation history of the present-day fauna of *Formica* ants of the Palearctic were obtained. This is especially true concerning the Quaternary period, since there are practically no ant collections from Pleistocene and Quaternary sediments [42]. At the same time, because of the peculiarities of their biology, *Formica* ants are rarely found in deposits of different geological periods, compared to other genera (for example, *Camponotus*) [42]. Therefore, it is currently impossible to trace in detail the history of the modern fauna of the genus *Formica* using only fossil evidence.

From the literature, it is known that, by the Pliocene (from 5.1 to 1.6 Ma, [43]), the *Formica* fauna was formed, which was not very different from the present, and during the Quaternary period, there was probably only the segregation of several species caused by the isolation of populations [1]. However, our data show that a considerable proportion of species (about a third) was formed in the Quaternary period. We begin with the structure of the *Cyt-b* tree:

(1) The cluster of *F. polyctena*, *F. rufa*, and *F. sinensis*; moreover, *F. sinensis* is morphologically closer to *F. truncorum* and inhabits the high mountain coniferous and mixed forests of China [44], while *F. polyctena* and *F. rufa* are sibling species [1].

(2) The cluster of *F. aquilonia*, *F. lugubris*, and *F. pratensis*; moreover, worker ants of *F. lugubris* differ from the worker ants of *F. pratensis* in either presence or absence of a dark spot with diffuse borders on the thorax [1].

(3) The cluster of *F. cinerea* and *F. selysi*, which belong to the *Formica cinerea* group. The specific feature of *F. selysi* is extreme pilosity [45].

(4) The cluster of *F. cunicularia* and *F. rufibarbis*. Confident distinguishing of *F. rufibarbis* from other species of the group is possible only with the use of morphometry and discriminant analysis [40].

(5) The cluster of *F. exsecta* and *F. fukaii*. The distinctive feature of *F. fukaii* is constant absence of setae from first gaster tergite of worker ants [46].

Now we proceed to the analysis of the *NADH-6* tree:

(1) The cluster of *F. candida* and *F. picea*. At present, with these species, a dual situation has developed. On one hand, outside of Russia, they are recognized as two different species [3]; on the other hand, Russian myrmecologists consider *F. picea* and *F. candida* as synonyms, because Seifert [47] made a conclusion based on limited material, and the distinctive characters suggested by him do not allow a satisfactory separation of these species [48]. Our data show that the genetic differences between these species are small, and they are smaller than the differences between the populations of *F. candida* (Fig. 2). However, because of a small sample size, making an unambiguous conclusion is difficult.

(2) The cluster of *F. paralugubris* and *F. pratensis*. In this cluster, the sympatric sibling species of *F. lugubris* is *F. paralugubris*, which clearly differs from other species in the female traits (especially chaetotaxy and size) [49].

Formation of the modern fauna of the Palearctic genus *Formica* took place during climate cooling that began in the second half of the Oligocene [41]. The main feature of the climate development in Eurasia in the Late Oligocene–Neogene was the strengthening of its differentiation in relation to the increasing interlatitudinal thermal contrasts and owing to the complication of the continental relief that influenced the distribution of atmospheric precipitation [41]. As a result, complication of natural zoning occurred. During this period, the modern subgeneric structure of the genus *Formica* was formed, as well as many of today's species of this genus. One of the important factors affecting the speciation was the appearance in the Pliocene of many orographic barriers [41]. The Quaternary period is characterized by high speed, contrast of all natural processes, and extremely short duration. During this period, there were repeated sharp fluctuations in temperature and humidity, which led to the alternation of glacial and interglacial periods in the high latitudes of Eurasia [41]. This caused intense speciation. However, the species formed during this period had no considerable morphological differences.

Thus, it is noted that the topology of the tree constructed on the basis of the *Cyt-b* fragment sequences does not conform with the generally accepted views on the *Formica rufa* and *Formica rufibarbis* groups. New data on the history of the modern fauna of the Palearctic genus *Formica* were obtained. Furthermore, it was demonstrated that a considerable fraction of the studied species (about a third) were formed in the Quaternary period.

REFERENCES

1. Dlussky, G.M., *Murav'i roda Formica* (Ants of the Genus Formica), Moscow: Nauka, 1967.
2. Dlussky, G.M., *Murav'i pustyn'* (Ants of Deserts), Moscow: Nauka, 1981.
3. An Online Catalog of the Ants of the World. http://www.antcat.org/
4. Zakharov, A.A. and Zakharov, R.A., The phenomenon of mixed families in red wood ants, in *Murav'i i zashchita lesa* (Ants and the Forest Conservation) (Proc. 13th All-Russ. Mirmecol. Symp.), Nizhny Novgorod: Nizhegorodsrii Gos. Univ., 2009, pp. 160–165.
5. Seifert, B. and Goropashnaya, A.V., Ideal phenotypes and mismatching haplotypes—errors of mtDNA treeing in ants (Hymenoptera: Formicidae) detected by standardized morphometry, *Org. Diversity Evol.,* 2004, vol. 4, pp. 295–305. doi 10.1016/j.ode.2004.04.005
6. Moreau, C.S., Inferring ant evolution in the age of molecular data (Hymenoptera: Formicidae), *Myrmecol. News,* 2009, vol. 12, pp, 201–210.
7. Moreau, C.S. and Bell, C.D., Fossil cross-validation of the dated ant phylogeny (Hymenoptera: Formicidae), *Entomol. Am.,* 2011, vol. 117, pp. 22–27. doi 10.1664/10-RA-012.1
8. Goropashnaya, A.V., Fedorov, V.B., Seifert, B., et al., Limited phylogeographical structure across Eurasia in two red wood ant species *Formica pratensis* and *F. lugubris* (Hymenoptera, Formicidae), *Mol. Ecol.,* 2004, vol. 13, no. 7, pp. 1849–1858. doi 10.1111/j.1365-294X.2004.02189.x
9. Goropashnaya, A.V., Fedorov, V.B., and Pamilo, P., Recent speciation in the *Formica rufa* group ants (Hymenoptera, Formicidae): inference from mitochondrial DNA phylogeny, *Mol. Phylogenet. Evol.,* 2004, vol. 32, no. 1, pp. 198—206. doi 10.1016/j.ympev.2003.11.016
10. Goropashnaya, A.V., Fedorov, V.B., Seifert, B., et al., Phylogeography and population structure in the ant *Formica exsecta* (Hymenoptera, Formicidae) across Eurasia reflected by mitochondrial DNA variation and microsatellites, *Ann. Zool. Fennici,* 2007, vol. 44, no. 6, pp. 462–474.
11. Goropashnaya, A.V., Fedorov, V.B., Seifert, B., et al., Phylogenetic relationships of Palearctic *Formica* species (Hymenoptera, Formicidae) based on mitochondrial cytochrome B sequences, *PLoS One,* 2012, vol. 7, no. 7, E41697. doi 10.1371/journal.pone.0041697
12. Gyllenstrand, N. and Seppa, P., Conservation genetics of the wood ant, *Formica lugubris,* in a fragmented landscape, *Mol. Ecol.,* 2003, vol. 12, no. 11, pp. 2931–2940. doi 10.1046/j.1365-294X.2003.01975.x
13. Gyllenstrand, N., Seppa, P., and Pamilo, P., Restricted gene flow between two social forms in the ant *Formica truncorum, J. Evol. Biol.,* 2005, vol. 18, no. 4, pp. 978–984. doi 10.1111/j.1420-9101.2005.00908.x
14. Holzer, B., Keller, L., and Chapuisat, M., Genetic clusters and sex-biased gene flow in a unicolonial *Formica* ant, *BMC Evol. Biol.,* 2009, vol. 9, no. 1, p. 69. doi 10.1186/1471-2148-9-69
15. Jansen, G., Savolainen, R., and Vepsalainen, K., Phylogeny, divergence—time estimation, biogeography and social parasite-host relationships of the Holarctic ant genus *Myrmica* (Hymenoptera: Formicidae), *Mol. Phylogenet. Evol.,* 2010, vol. 56, no. 1, pp. 294–304. doi 10.1016/j.ympev.2010.01.029
16. Johnson, R.N., Agapow, P.M., and Crozier, R.H., A tree island approach to inferring phylogeny in the ant

subfamily Formicinae, with especial reference to the evolution of weaving, *Mol. Phylogenet. Evol.,* 2003, vol. 29, no. 2, pp. 317–330. doi 10.1016/S1055-7903 (803)00114-3

17. Maki-Petays, H. and Breen, J., Genetic vulnerability of a remnant ant population, *Conserv. Genet.,* 2007, vol. 8, no. 2, pp. 427–435. doi 10.1007/s10592-006-9182-1
18. Lukashov, V.V., *Molekulyarnaya evolyutsiya i filogenetic heskii analiz* (Molecular Evolution and Phylogenetic Analysis), Moscow: BINOM, Laboratoriya znanii, 2009.
19. Goropashnaya, A.V., *Phylogeographic Structure and Genetic Variation in Formica Ants*, PhD Thesis, Uppsala: Uppsala Univ., 2003.
20. GenBank, NCBI. http:// www.ncbi.nlm.nih.gov/
21. Hall, T.A., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp. Ser.,* 1999, vol. 41, pp. 95–98.
22. Goujon, M., McWilliam, H., Li, W., et al., A new bioinformatics analysis tools framework at EMBL-EBI, *Nucleic Acids Res.,* 2010, vol. 38, no. 2, pp. W695–W699. doi 10.1093/nar/gkq313
23. FaBox. http://users-birc.au.dk/biopv/php/fabox/
24. Tajima, F., Simple methods for testing molecular clock hypothesis, *Genetics,* 1993, vol. 135, no. 2, pp. 599–607.
25. Paradis, E., Claude, J., and Strimmer, K., APE: analyses of phylogenetics and evolution in R language, *Bioinformatics,* 2004, vol. 20, no. 2, pp. 289—290. doi 10.1093/bioinformatics/btg412
26. Paradis, E., Pegas: an R package for population genetics with an integrated–modular approach, *Bioinformatics,* 2010, vol. 26, no. 3, pp. 419–420. doi 10.1093/bioinformatics/btp696
27. Kabacoff, R.I., *R in Action: Data Analysis and Graphics with R,* New York: Manning, 2011.
28. Darriba, D., Taboada, G.L., Doallo, R., et al., jModelTest 2: more models, new heuristics and parallel computing, *Nat. Methods*, 2012, vol. 9, no. 8, p. 772. doi 10.1038/nmeth.2109
29. Ronquist, F., Teslenko, M., van der Mark, P., et al., MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Syst. Biol.,* 2012, vol. 61, no. 3, pp. 539–542. doi 10.1093/sysbio/sys029
30. Dlussky, G.M., Ants of the tribe *Formicini* (Hymenoptera, Formicidae) from Late Eocene amber of Europe, *Paleontol. J.,* 2008, vol. 42, no. 5, pp. 500–513.
31. PHYLOCH. http://www.christophheibl.de/Rpackages.html/
32. BOOT. http://cran.r-project.org/web/packages/simpleboot/index.html/
33. Waddell, P.J. and Steel, M.A., General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites, *Mol. Phylogenet. Evol.,* 1997, vol. 8, no. 3, pp. 398—414. doi 10.1006/mpev.1997.0452
34. Kimura, M., A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.,* 1980, vol. 16, no. 2, pp. 111–120.
35. Farias, I.P., Ortí, G., Sampaio, I., et al., The cytochrome b gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes, *J. Mol. Evol.,* 2001, vol. 53, no. 2, pp. 89–103. doi 10.1007/s002390010197
36. Minchenko, A.T. and Dudareva, N.A., *Mitochondrial Genom,* Novosibirsk: Nauka, 1990.
37. Yang, S., Li, X., Cai, L.G., and Qian, Z.Q., Characterization of the complete mitochondrial genome of *Formica selysi* (Insecta: Hymenoptera: Formicidae: Formicinae), *Mitochondrial DNA,* 2015, vol. 26, pp. 1–3. doi 10.3109/19401736.2015.1018229
38. Papadopoulou, A., Anastasiou, I., and Vogler, A.P., Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration, *Mol. Biol. Evol.,* 2010, vol. 27, no. 7, pp. 1659–1672. doi 10.1093/molbev/msq051
39. Mayr, E., Linsley, E., and Usinger, R., *Methods and Principles of Systematic Zoology,* New York: McGraw-Hill, 1953.
40. Seifert, B. and Schultz, R.B., A taxonomic revision of the *Formica rufibarbis* Fabricius, 1793 group (Hymenoptera: Formicidae), *Myrmecol. News,* 2009, vol. 12, pp. 255–272.
41. Sinitsyn, V.M., *Vvedenie v paleoklimatologiyu* (Introduction to Paleoclimatology), Leningrad: Nedra, 1980.
42. Dlusskii, G.M. and Rasnitsyn, A.P., Paleontological record and stages of ant evolution, *Usp. Sovrem. Biol.,* 2007, vol. 127, no. 2, pp. 118–134.
43. Zhamoida, A.I., *Obshchaya stratigraficheskaya shkala, prinyataya v SSSR—Rossii: ee znachenie, naznachenie i sovershenstvovanie* (General Stratigraphic Scale Adopted in the Soviet Union—Russia: Its Value, Purpose and Improvement), St. Petersburg: Vseross. Nauchno-Issled. Geol. Inst., 2013.
44. Wu, J., Taxonomic studies on the genus *Formica* L. of China (Hymenoptera: Formicidae), *For. Res.,* 1990, vol. 3, pp. 1–8.
45. Seifert, B., A taxonomic revision of the *Formica cinerea* group (Hymenoptera: Formicidae), *Abh. Ber. Naturkd. Mus. Görlitz,* 2002, vol. 74, no. 2, pp. 245–272.
46. Seifert, B., A taxonomic revision of the ant subgenus *Coptoformica* Mueller, 1923 (Hymenoptera, Formicidae), *Zoosystema,* 2000, vol. 22, no. 3, pp. 517–568.
47. Seifert, B., The "Black Bog Ant" *Formica picea* Nylander, 1846—a species different from *Formica candida* Smith, 1878 (Hymenoptera: Formicidae), *Myrmecol. Nachr.,* 2004, vol. 6, pp. 29–38.
48. Zakharov, A.A., Dlusskii, G.M., Goryunov, D.N., et al., *Monitoring murav'ev Formica* (Monitoring of *Formica* Ants), Moscow: KMK, 2013.
49. Seifert, B., *Formica paralugubris* nov. spec.—a sympatric sibling species of *Formica lugubris* from the western Alps (Insecta: Hymenoptera: Formicoidea: Formicidae), *Reichenbachia,* 1996, vol. 31, no. 35, pp. 193–201.

Translated by N. Maleeva