Вавиловский журнал генетики и селекции. 2024;28(1):98-107
DOI 10.18699/vjgb-24-12

Original Russian text https://vavilovj-icg.ru/

A universal panel of STR loci for the study of polymorphism of the species *Canis lupus* and forensic identification of dog and wolf

A.E. Hrebianchuk[1] ✉, I.S. Tsybovsky[2]

[1] Forensic Examination Committee of the Republic of Belarus, Minsk, Belarus
[2] BelJurZabespjachjenne, Minsk, Belarus
✉ iamsanya94@mail.ru

Abstract. Commercial panels of microsatellite (STR) loci are intended for DNA analysis of the domestic dog (*Canis lupus familiaris*) and, therefore, when genotyping the Grey wolf (*Canis lupus lupus*), most markers reveal significant deviations from the Hardy–Weinberg equilibrium and have a low informative value, which complicates their use in a forensic examination. The aim of this study was to select STR markers that equally effectively reflect population polymorphism in the wolf and the dog, and to create a universal panel for the identification of individuals in forensic science. Based on the study of polymorphisms of 34 STR loci, a CPlex panel of 15 autosomal loci and two sex loci was developed, which is equally suitable for identifying wolfs and dogs. Analysis of molecular variance (AMOVA) between samples revealed significant differentiation values ($F_{ST} = 0.0828$, $p < 0.05$), which allows the panel to be used for differentiating between wolf and dog samples. For the first time in the forensic examination of objects of animal origin in the Republic of Belarus, population subdivision coefficients (θ-values) were calculated for each of the 15 STR loci of the test system being reported. It was shown that the values of the genotype frequency, when averaged over all studied animals without and with considering the θ-value, differ by three orders of magnitude ($3.39 \cdot 10^{-17}$ and $4.71 \cdot 10^{-14}$, respectively). The use of population subdivision coefficients will provide the researcher with the most relevant results of an expert identification study. The test system was validated in accordance with the protocol of the Scientific Working Group on DNA Analysis Methods. A computational tool was developed to automate the analysis of genetic data on the wolf and dog in the forensic examination; two guides were approved for practicing forensic experts. This methodology is being successfully used in expert practice in investigating cases of illegal hunting, animal abuse and other offenses in the Republic of Belarus.
Key words: microsatellites; polymorphism; differentiation; identification; *Canis lupus familiaris*; *Canis lupus lupus*; wildlife forensic science.

For citation: Hrebianchuk A.E., Tsybovsky I.S. A universal panel of STR loci for the study of polymorphism of the species *Canis lupus* and forensic identification of dog and wolf. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):98-107. DOI 10.18699/vjgb-24-12

Универсальная панель STR-локусов для исследования полиморфизма вида *Canis lupus* и криминалистической идентификации собаки и волка

А.Е. Гребенчук[1] ✉, И.С. Цыбовский[2]

[1] Государственный комитет судебных экспертиз Республики Беларусь, Минск, Беларусь
[2] Республиканское унитарное предприятие «БелЮрОбеспечение», Минск, Беларусь
✉ iamsanya94@mail.ru

Аннотация. Коммерческие панели микросателлитных (STR) локусов предназначены для работы с ДНК собаки домашней (*Canis lupus familiaris*), в связи с чем при генотипировании волка обыкновенного (*Canis lupus lupus*) большинство маркеров показывают существенные отклонения от равновесия Харди–Вайнберга и имеют низкий показатель информативной ценности, что осложняет их использование в судебной экспертизе. Целью настоящего исследования стал подбор STR-маркеров, которые одинаково эффективно отображают популяционный полиморфизм волка и собаки, с последующим созданием универсальной панели для дифференциации и идентификации особей волка и собаки в криминалистике. На основе исследования полиморфизма 34 STR-локусов сконструирована панель CPlex из 15 аутосомных локусов и двух локусов половой принадлежности, которая одинаково применима для идентификации волка и собаки. Анализ молекулярной дисперсии (AMOVA) между выборками выявил достоверные значения дифференциации ($F_{ST} = 0.0828$, $p < 0.05$), что позволяет использовать панель для дифференциации образцов волка обыкновенного и собаки домашней. Впервые в судебной экспертизе объектов животного происхождения в Республике Беларусь рассчитаны коэффициенты подразделенности популяции (θ-value) для каждого из 15 STR-локусов разработанной тест-системы. Показано, что значения частот генотипов, усреднен-

ные по всем исследованным животным без учета и с учетом θ-value, различаются на три порядка ($3.39 \cdot 10^{-17}$ и $4.71 \cdot 10^{-14}$ соответственно). Применение коэффициентов подразделенности популяции позволит оперировать наиболее достоверными результатами экспертного идентификационного исследования. Предложенная тест-система валидирована в соответствии с протоколом Scientific Working Group on DNA Analysis Methods. Создан информационно-статистический комплекс для автоматизации обсчета генетических данных волка обыкновенного и собаки домашней в судебной экспертизе, утверждены две методики для практикующих судебных экспертов. Методические разработки успешно применяются в экспертной практике при расследовании фактов незаконной охоты, жестокого обращения с животными и других правонарушений в Республике Беларусь.
Ключевые слова: микросателлиты; полиморфизм; дифференциация; идентификация; собака домашняя; волк обыкновенный; судебная экспертиза объектов животного происхождения.

Introduction

According to the statistics of the Ministry of Forestry, in the Republic of Belarus, the Grey wolf (*Canis lupus lupus*) population has stabilized over the past five years in the range of 1,530–1,630 individuals, which is one of the leading indicators among European countries (Ministry of Forestry of the Republic of Belarus, 2021). Notably, wolf hunting in Belarus is allowed all year long with no sex or age restrictions. At the same time, hunting in protected natural areas and hunting without a permit results in criminal cases and, consequently, the need for forensic examination.

According to the Ministry of Housing and Communal Services, about 80,000 stray cats and dogs are exterminated in Belarus every year, and this number is growing, while the exact number of dogs is unknown. Being one of the most common companion animals, the domestic dog (*Canis lupus familiaris*) has a special status among farm and domestic animals. The active use of dogs by humans in various roles is also reflected in the criminal aspects that accompany the development of society.

The natural history of European wolf (*C. lupus lupus*) populations has been characterized by a strong reduction in the number of individuals over the past few hundred years (Boitani, 2003). The decline of the population, its fragmentation, and disruption of the gene flow are well-known triggers of genetic impoverishment and increased inbreeding in natural populations, which increase the risk of extinction for wolves as well as for many other species. An example of such a situation was documented for wolves in Italy, where the values of genetic diversity determined by the level of heterozygosity were clearly lower than those in populations from Russia, Alaska, and Canada (Godinho et al., 2011).

Intensive extermination of the wolf can lead to the replacement of the wolf with wolf-dog hybrids. Recently, the problem of hybridization between wolves and free-living dogs in Europe has become a major topic in many research programs (Stronen et al., 2022).

The main difficulty in the genetic differentiation between the wolf and the dog is that DNA markers unique to both the wolf and the dog have not been found. A comparison of dog and wolf genomes showed a similarity of 99 %, which once again confirms their common origin (Freedman, Wayne, 2017).

The study of wolf populations is usually designed according to a typical approach that includes the use of loci recommended by the International Society of Animal Genetics (ISAG) with calculation of statistical indexes of distribution of alleles of the studied loci and assessment of the representation of alleles in the population. Due to the high level of identity of wolf and dog genomes, the ranges of alleles of the loci are very similar. Therefore, differentiation of individuals using selected loci becomes possible if genetic differentiation of the studied samples is revealed by statistical processing of genotyping results (Halverson, Basten, 2005; Fan et al., 2016).

Most panels for canine DNA analysis are unsuitable for the study of wolf DNA due to deviation from the Hardy–Weinberg equilibrium and the presence of null alleles in DNA markers. On the other hand, when selecting markers for DNA analysis of the wolf, researchers generally do not take into account the possibility of using the selected markers on an inbred dog population, which precludes the use of these markers to identify a wolf and a dog simultaneously in forensic studies.

The aim of this study was to select STR markers that equally effectively reflect the population structure and the polymorphism of the Grey wolf and the domestic dog, followed by the creation of a universal panel for the identification and differentiation of individuals in forensic science.

Materials and methods

Biological objects. Biological samples of the Grey wolf and the domestic dog were obtained legally and are represented by fragments of muscle and cartilage tissues of wolves ($n = 103$) and samples of blood, hair and saliva of purebred dogs, mestizo dogs and outbred dogs ($n = 198$). The list of the most represented breeds is reflected in Supplementary Material 1[1].

DNA extraction, amplification and genotyping of microsatellite loci. DNA from muscle and cartilage tissues, blood, saliva and hair of wolves and dogs was extracted according to a method based on the release of DNA during the incubation of biological material in a lysis buffer with proteinase K and 0.01 mM dithiothreitol at 37–56 °C. The lysate was subjected to the purification procedure on silica (Boom et al., 1990).

All selected loci were grouped into two test systems: a test system that includes mainly dinucleotide loci recommended by ISAG: AHTk211, FH2054, CXX279, Ren169O18, INU055, AHTh260, INU030, FH2079, FH2848, AHT121, AHTh171, Ren247M23, AHTh130, INRA21, AHTk253, AHT137, Ren54P11, INU005, Ren105L03, Ren64E19, Ren162C04 and Amelogenin sex locus (Radco, Podbielska, 2021); and a test system that includes mainly tetranucleotide loci – FH2096, CPH12, CPH4, FH2004 (Caniglia et al., 2010), FH2016 (Fan et al., 2016), FH2361, PEZ17, FH2328 (van Asch et al., 2010), PEZ16, vWF.x (DeNise et al., 2004), FH2010 (Eichmann et at., 2004), FH2001 (Verardi et al., 2006), VGL3438 (Magory Cohen et al., 2013) and sex loci – DBX and DBY (Seddon, 2005).

[1] Supplementary Materials 1–6 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Hrebian_Engl_28_1.pdf

Accordingly, the 10 μl PCR reaction volume contained 10 mM tris-HCl, pH 8.6; 25 mM KCl; 2.0 mM $MgCl_2$; 0.2 mM of each dNTP; 0.2–1.0 μM of each of the pair of primers; 0.15 u. DNA polymerase; 1.5 ng/μl BSA; 0.1 % Triton X-100 and 1–20 ng DNA to be analyzed. The polymerase chain reaction was conducted in a thermocycler C1000 (BioRad, USA) using the following program: (1) an initial denaturation step at 95 °C for 5 min; (2) 30 cycles of denaturation at 95 °C (for 30 s), annealing at 60 °C (for 40 s) and elongation at 72 °C (for 1 min); (3) a final elongation at 72 °C for 30 min.

The combination of alleles of each of the samples was detected by electrophoretic separation of PCR products on a 3500 Genetic Analyzer (ThermoFisher Scientific, USA). The size of the detected alleles (in bp) in the studied loci was determined using the Orange 500 bp internal size standards (NimaGen®, The Netherlands) and GeneScan-600 LIZ™ SizeStandard v2.0 (ThermoFisher Scientific, USA). Genotyping was evaluated with GeneMapper ID-X v1.6 software package (ThermoFisher Scientific, USA).

Statistical analysis of the data. Since incorrect species identification distorts calculations based on the analysis of genetic diversity indexes (Galinskayaa et al., 2019), a cluster analysis of genotyping data for wolves and dogs was first carried out for the studied loci. The population structure was determined using the Monte Carlo algorithm according to the Markov chain method using STRUCTURE v.2.3.4 software and Admixture model (Pritchard et al., 2000) with further determination of the true number of clusters by the method of Evanno et al. (Evanno et al., 2005). The burn-in period included 500,000 iterations, followed by the construction of the Markov chain for 1,000,000 iterations for the expected number of groups in the sample, *K*, from 1 to 10, with six repeats for each value of *K*. The analysis of clustering in the combined pool of wolves and dogs was performed using the unweighted pair group method with arithmetic mean (UPGMA) and the nearest neighbors joining method (NJ), and the construction of the corresponding dendrograms was carried out using MEGA v.11.0.10 software. To visualize the genetic structure, a multidimensional analysis was performed using the genetic distance matrix following the PCoA method in GenAlEx v.6.5 (Peakall, Smouse, 2006, 2012).

The calculation of frequencies of alleles of STR loci and values of the observed (H_O) and expected heterozygosity (H_E), as well as evaluation of deviation from Hardy–Weinberg equilibrium was performed using Cervus v.3.0.7 software package (Kalinowski et al., 2007). To identify possible errors in the interpretation of genetic profiles, null alleles, and PCR artifacts, an analysis was carried out using Micro-Checker v.2.2.1 software (van Oosterhout et al., 2004).

Analysis of molecular dispersion and estimation of inbreeding coefficients were performed using Arlequin v.3.5.1.3 software (Excoffier et al., 2005). Analysis of assignment of the individual to the sample pool (Assignment Test) based on the selected loci was performed using GenAlEx v.6.5. Polymorphism information content (PIC) of the selected STR loci was calculated using Cervus software v.3.0.7.

Alleles sequence determination. To identify possible microvariants of the sequence, as well as to perform tandem determination of alleles, which is a common practice in forensic science, the primary structure of alleles was determined by Sanger dideoxy sequencing (Sanger et al., 1977). Nucleotide sequences of alleles of each STR locus and sex loci were determined in the forward and reverse directions. Sequencing was performed on a 3500 Genetic Analyzer and was conducted with the BigDye® Terminator v.3.1 Cycle Sequencing Kit (ThermoFisher Scientific). Comparative analysis of allele sequences of the studied loci was carried out in BioEdit v.7.0.5.3 (Hall et al., 2011). The sequences of each locus with the minimum and maximum allele molecular sizes were deposited in the GenBank database (Benson et al., 2005) with assignment of corresponding access numbers.

Results and discussion

A posteriori analysis of the STRUCTURE results for the combined pool of the wolf and dog genotypes revealed the maximum value of the test statistics ΔK at $K = 2$, which indicates the presence of two genetic clusters in the analyzed sample of animals: the Grey wolf (green cluster) and the domestic dog (red cluster) (Fig. 1, Supplementary Material 2). When determining the population structure of samples of wolves and dogs separately from each other, the cluster formed by the wolf samples remained homogeneous; the absence of clustering within the wolf population using STR loci was previously shown by researchers in Europe (Aspi et al., 2006; Sastre et al., 2011; Đan et al., 2016).

For a separate analysis of STRUCTURE in the samples of dogs, four groups were formed: purebred dogs ($n = 78$) and three breed groups: Molossians ($n = 32$), Samoyeds ($n = 66$) and Shepherd dogs ($n = 22$). The maximum value of ΔK at $K = 2$ indicated the formation of two clusters (Supplementary Materials 3 and 4), one of which corresponds to Shepherd dogs, whereas the breed groups of Samoyeds and Molossians were not separated and, moreover, did not differ from the group of outbred dogs.

At the same time, the analysis of clustering in the combined pool of wolves and dogs by the methods of UPGMA and NJ with construction of corresponding dendrograms revealed four clusters of different hierarchical levels in the sample of dogs. The greatest similarity was observed between the breed group of Molossians and outbred dogs. They are adjacent to the breed group of Samoyeds, and this entire group is separated from the group of Shepherd dogs. Since both dendrograms showed a similar structure, Supplementary Material 5 shows only the dendrogram built according to the UPGMA method (with bootstrapping for 10,000 permutations). It should be noted that a similar pattern of clustering of dog breeds in the study of SNP markers was obtained by other authors (Parker et al., 2017).

Analysis of the population structure of wolves and dogs showed a strong genetic differentiation between them with the average values of the cluster membership coefficient *Q* of 0.984 and 0.981, respectively. The data obtained on differentiation of the wolf and the dog by 34 STRs are in good agreement with the data of Korablev et al. (2021). The level of differentiation between dog breeds was much lower than between the wolf and the dog. *Q* values varied from 0.457 to 0.495 for Molossians, from 0.451 to 0.476 for Samoyeds, and from 0.740 to 0.757 for Shepherd dogs.

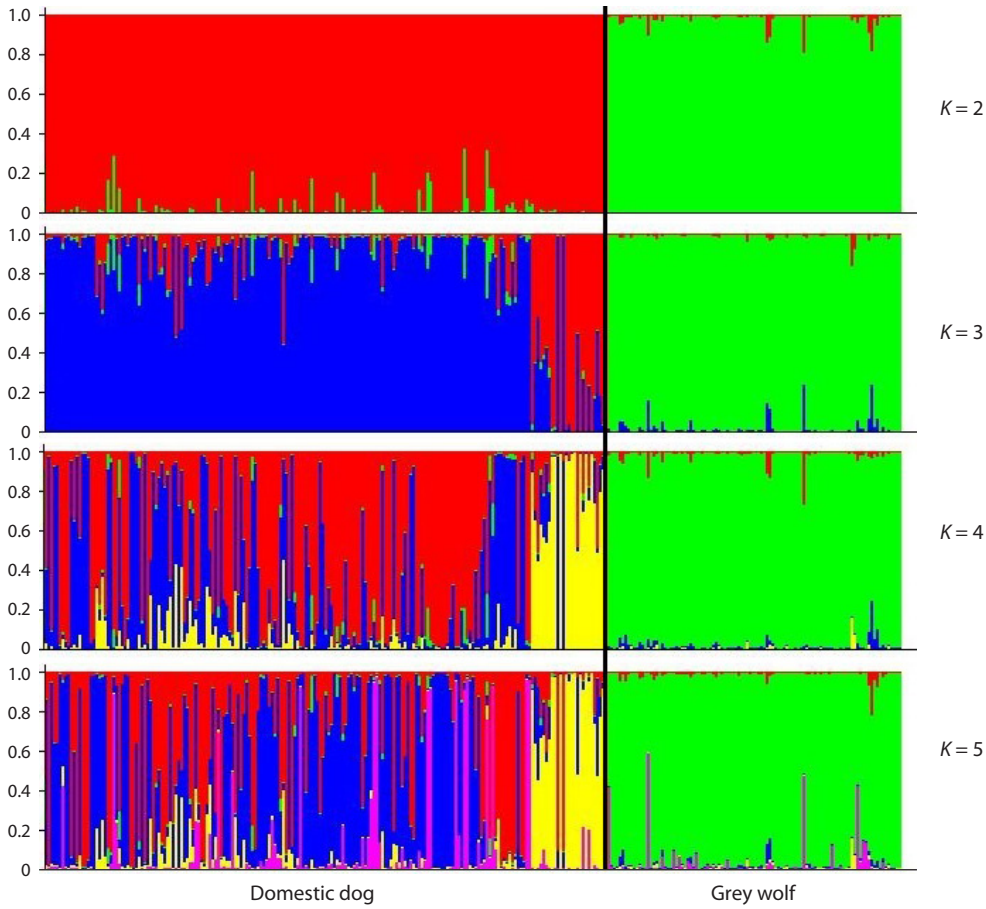

Fig. 1. Results of cluster analysis of wolf and dog samples, performed in STRUCTURE software, for the most probable value of the number of genetic clusters $K = 2–5$, sorted by samples.

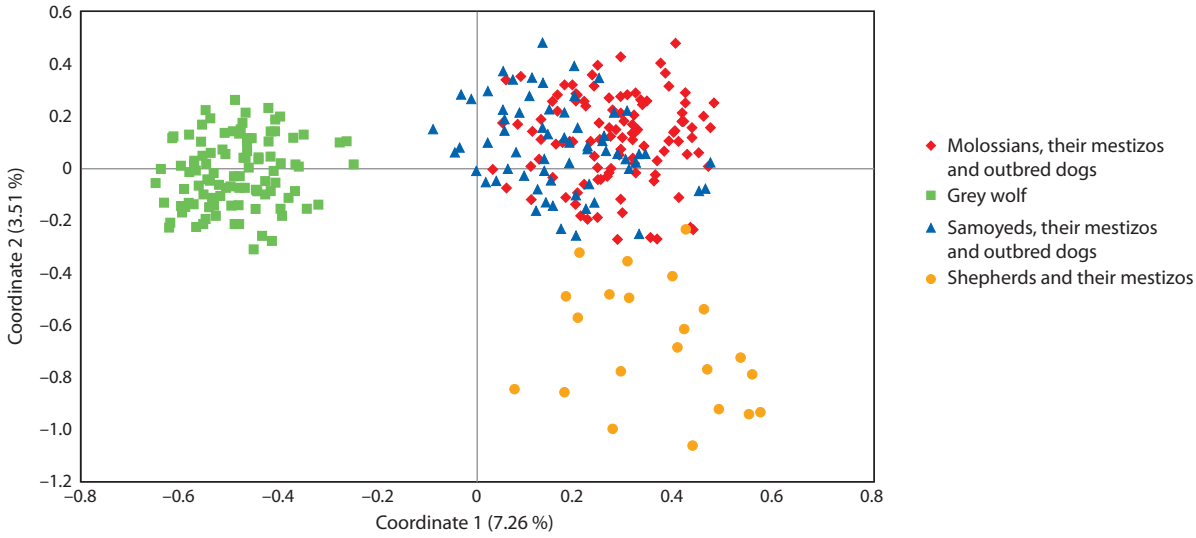

Fig. 2. The diagram of the results of PCoA analysis based on the matrix of paired F_{ST} values for samples of the Grey wolf and domestic dog.

The results of the cluster analysis are consistent with the multivariate genetic distance matrix analysis (PCoA method), which also shows a strong differentiation between the wolf and dog samples (Fig. 2). Similarly, Shepherd dogs form a separate group. The latter outcome may be explained by selection differences or may be a consequence of the high heterogeneity of the samples of Molossians and Samoyeds, which have a common historic origin (over 20 different breeds), while the sample of Shepherd dogs included only purebred German Shepherds.

Table 1. Average values of the level of polymorphism of the studied loci in samples of wolves and three historical breed groups of dogs

Samples	N_a	N_e	H_O	H_E
Wolves	9.265 ± 0.476	5.258 ± 0.306	0.730 ± 0.017	0.786 ± 0.013
Molossians, their mestizos and outbred dogs	9.706 ± 0.575	4.916 ± 0.293	0.656 ± 0.017	0.768 ± 0.016
Samoyeds, their mestizos and outbred dogs	9.235 ± 0.560	5.212 ± 0.312	0.688 ± 0.018	0.782 ± 0.015
Shepherds and their mestizos	5.824 ± 0.328	2.992 ± 0.197	0.556 ± 0.029	0.617 ± 0.025

Note. N_a – number of alleles per locus; N_e – effective number of alleles per locus; H_O – observed heterozygosity; H_E – expected heterozygosity.

In total, 405 alleles were identified in the sample of wolves and domestic dogs using the two test systems. All the loci were polymorphic and had from 5 (FH2096) to 26 (FH2361) alleles per locus. The average number of alleles per locus in all samples was similar, amounting to 9.402 ± 0.617 (Table 1). The Shepherd dog was an exception, with the mean value per locus in the range of 5.824 ± 0.328, which could be a consequence of the small sample size (n = 22).

The highest observed and expected heterozygosity rates were obtained for the wolf sample, at 0.730 and 0.786, respectively (see Table 1). In dog samples, lower values of expected heterozygosity compared to observed heterozygosity are a potent indicator of the presence of inbreeding resulting in synthetic selection and genetic drift, which can irreversibly remove alleles from the population, leading to significantly reduced diversity (Galinskayaa et al., 2019). Although mutations counteract the genetic drift, it is difficult to achieve a balance of genetic processes in dog breeding due to exclusion of individuals with identified mutations in a particular trait. The highest values of heterozygosity and the highest effective number of alleles were found in the wolf sample, which indicates natural development and presence of mutation-drift balance in the natural population.

The analysis of the profiles of dinucleotide loci requires special attention since fragments that do not belong to true alleles and are stutter products can be present on electrophoregrams. The percentage of stutter products usually increases with the length of the allele. Based on the analysis of the genotypes using Micro-Checker software, three of the 20 dinucleotide markers (INU055, Ren169O18, and Ren64E19) showed a high probability of genotyping errors and were excluded from further analysis.

The analysis of the distribution of alleles for two loci (AHT121 and AHTk211) showed the presence of a large number of null alleles in three samples, including the sample of the wolf. AHT137 and INRA21 loci showed a high null alleles content in two samples of dogs, 8 loci (AHTh130, AHTh260, CXX279, FH2848, Ren105L03, Ren162C04, Ren247M23, and Ren54P11) showed a rather high null alleles content (from 5.7 % in the sample of Molossians at the FH2848 locus to 12.6 % at the CXX279 locus in the sample of Shepherd dogs). These loci were also excluded from further work on the design of a universal forensic panel.

When analyzing the pattern of frequency distribution of alleles in specific loci, special attention was paid to loci with a significant predominance of major alleles. Differences in the frequencies of major alleles can be very instrumental for differentiating wolves and dogs by microsatellite analysis; however, a significant predominance of one allele can affect the level of identification confidence. Due to the pronounced dominance of major alleles, INU030, INU005, AHTk171 and AHTk253 loci were excluded from further analysis.

Most of the studied loci in the samples of wolves and dogs conformed to the Hardy–Weinberg distribution ($p > 0.05$), including AHTh171, AHTk253, CPH12, CPH4, FH2001, FH2004, FH2010, FH2016, FH2096, FH2328, FH2361, INU005, INU030, INU055, PEZ16, PEZ17, Ren169O18, Ren64E19, VGL3438 and vWF.x. Two loci, FH2054 and FH2079, deviated from the equilibrium ($p = 0.005$) in all studied samples; however, when using the Bonferroni correction, p-values ceased to be statistically significant. In 12 loci, a statistically significant deviation from the Hardy–Weinberg equilibrium was revealed in at least one sample, which may reflect the manifestation of loci in the studied samples and can be explained by the presence of null alleles.

The allele fixation index (F_{ST}) ranged from 0.025 (FH2001) to 0.158 (CPH4), with an average value of 0.077 for all loci. The highest F_{IS} values were obtained for the FH2079, FH2096, and CPH12 loci (0.221, 0.174, and 0.162, respectively). Overall, for a panel of 15 selected loci, inbreeding indexes (F_{IS} and F_{IT}) were 0.103 and 0.172, respectively (Table 2).

F_{IS} values in the sample of wolves in most of the studied loci showed values approaching zero, which, together with high heterozygosity values, can indicate the presence of panmixia in the population (Galinskayaa et al., 2019).

A correct assignment of a sample to a Grey wolf or a domestic dog can be of crucial importance in a forensic investigation. Analysis of molecular variance (AMOVA) was performed to assess the possibility of differentiation between a wolf and a dog using selected microsatellite loci. The AMOVA results showed that the percentage of variation between wolf and dog samples was 8.28 %, while within samples it was 91.72 %. Variance components in the population were significant for all studied loci (Supplementary Material 6), which indicates differentiation of the wolf and dog samples. The vWF.x, FH2096, and CPH4 loci accounted for 21.11, 19.34, and 14.40 % of inter-sample genetic variability, respectively, while FH2079 and FH2016 showed the lowest inter-population variability (2.45 and 2.20 %, respectively).

According to Wright's interpretation (Wright, 1978), the range of F_{ST} values from 0.15 to 0.25 indicates moderate differentiation. At the same time, values in the range of 0.00–0.05 indicate a weak but noteworthy difference between the samples. Since hypervariable markers with a large number of

Table 2. Mean values of heterozygosity for the wolf and dog and Wright F-statistic values for the total sample of *Canis lupus*

Loci	H_O	H_E	H_O	H_E	F_{IS}	F_{ST}	F_{IT}
	Domestic dog		Grey wolf				
FH2016	0.783	0.894	0.832	0.886	0.129	0.033	0.158
FH2079	0.555	0.664	0.612	0.656	0.221	0.093	0.293
VGL3438	0.751	0.858	0.709	0.784	0.130	0.029	0.155
FH2361	0.793	0.852	0.861	0.821	0.003	0.032	0.035
FH2054	0.792	0.822	0.760	0.781	0.098	0.077	0.167
FH2001	0.737	0.804	0.612	0.744	0.054	0.025	0.078
FH2328	0.763	0.865	0.718	0.743	0.065	0.058	0.119
FH2004	0.767	0.803	0.784	0.885	0.081	0.067	0.142
CPH12	0.571	0.707	0.553	0.650	0.162	0.060	0.212
FH2010	0.641	0.713	0.767	0.777	0.022	0.070	0.090
PEZ17	0.675	0.797	0.777	0.776	0.109	0.068	0.170
PEZ16	0.611	0.793	0.786	0.825	0.145	0.120	0.247
CPH4	0.595	0.611	0.767	0.837	0.103	0.158	0.245
FH2096	0.414	0.524	0.485	0.603	0.174	0.133	0.284
vWF.X	0.545	0.613	0.786	0.838	0.052	0.136	0.182
Average	0.666	0.755	0.721	0.774	0.103	0.077	0.172

Note. H_O – observed heterozygosity; H_E – expected heterozygosity; F_{IS} – inbreeding index of individuals within the sample; F_{ST} – allele fixation index; F_{IT} – inbreeding index of individuals in the total sample.

alleles can have significantly lower F_{ST} values than markers with a small number of alleles, it is more important to detect significant genetic differentiation between wolves and dogs in the totality of the selected STR loci (Hedrick, 2000). Between-sample AMOVA analysis revealed significant differentiation between the wolf and the dog ($F_{ST} = 0.0828$, $p < 0.05$).

The significance of the differentiation of the wolf and the dog using the selected loci can be illustrated by the analysis of assignment of a definite sample (Assignment test). The analysis is based on the calculation of the probability value of the presence of the genotype of a certain individual in the sample from which it was selected, and its comparison with the probability value of the same genotype in another sample. Based on these calculations, an individual belongs to the sample for which it has the highest probability (Fig. 3).

The calculation of true genetic affiliation to the sample showed a high consolidation of wolves and dogs, with 100 % of all studied animals genetically assigned to their own cluster. At the same time, we observed a large difference in the moduli of natural logarithms of expected frequencies of the genotypes when they belonged to their own versus an alternative sample; for wolves, the difference was on average 8.722, and for the domestic dog, it was 8.584. The high values of the difference between the moduli of the logarithms of the expected genotype frequencies confirm successful differentiation of the Grey wolf and the domestic dog using the proposed loci.

An important criterion in forensics is adequate interpretation of the value of the reliability level of an identification study. To calculate the probability of accidental matching of genotypes, an appropriate reference database and correct application of a conservative calculation procedure are required (Buckletone et al., 2016). The study of genotype pools of samples of wolves and dogs made it possible to calculate the frequencies of occurrence of alleles and corrections for the genetic subdivision of the populations of wolves and dogs living in Belarus.

A complex social organization of the wolf pack, along with targeted dog breeding, lead to formation of structured populations of these animals. Therefore, in order to obtain the most reliable conclusions, using frequency characteristics of the loci of specific subpopulations would be the preferred approach in an identification study, but this is hardly possible in practice in forensic examination of objects of animal origin. An alternative solution is to introduce a subdivision coefficient (θ-value) into the calculation of the frequency of the genotype, which takes into account the presence of structured populations (The Evaluation…, 1996; Buckleton et al., 2006).

As can be seen from Table 3, the values of genotype frequencies averaged over all studied animals and calculated with the inclusion of subsequent studied loci of the test system without and with taking into account the θ-value differ by three orders of magnitude ($3.39 \cdot 10^{-17}$ and $4.71 \cdot 10^{-14}$, respectively). This suggests that not factoring in the θ-value in the identification study will lead to overestimation (underestimation) of the genotype frequency, which in forensic examination will lead to an unintended overestimation of the reliability level of the study.

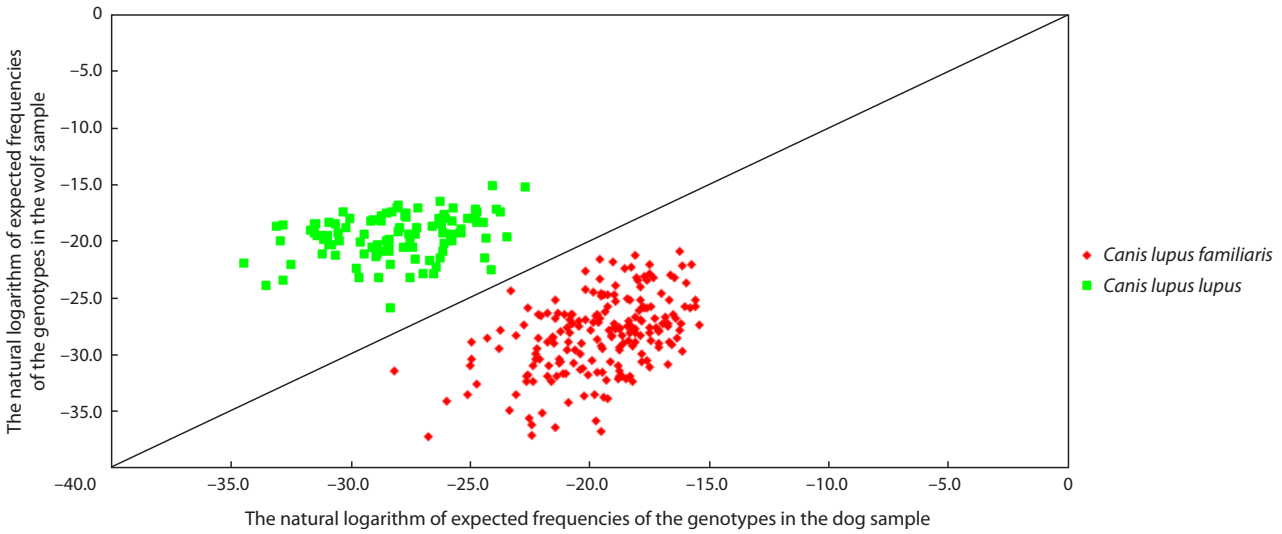

Fig. 3. Graph of the genetic assignment of the genotype to the study sample (graphical interpretation of the Assignment test).

Table 3. Genotype frequencies averaged over all animals, without and with using the θ-value

Loci	θ-value	Genotype frequency excluding θ-value and with the inclusion of the subsequent locus	Genotype frequencies with θ-value and with the inclusion of the subsequent locus**
FH2016	0.039	$2.08 \cdot 10^{-2}$	$3.62 \cdot 10^{-2}$
FH2079	0.145	$3.31 \cdot 10^{-3}$*	$9.58 \cdot 10^{-3}$**
VGL3438	0.020	$1.55 \cdot 10^{-4}$	$5.44 \cdot 10^{-4}$
FH2361	0.035	$7.19 \cdot 10^{-6}$	$3.39 \cdot 10^{-5}$
FH2054	0.081	$4.26 \cdot 10^{-7}$	$3.61 \cdot 10^{-6}$
FH2001	0.031	$3.23 \cdot 10^{-8}$	$3.38 \cdot 10^{-7}$
FH2328	0.065	$1.81 \cdot 10^{-9}$	$3.04 \cdot 10^{-8}$
FH2004	0.090	$8.97 \cdot 10^{-11}$	$2.96 \cdot 10^{-9}$
CPH12	0.102	$1.26 \cdot 10^{-11}$	$6.27 \cdot 10^{-10}$
FH2010	0.070	$1.46 \cdot 10^{-12}$	$9.84 \cdot 10^{-11}$
PEZ17	0.080	$1.05 \cdot 10^{-13}$	$1.16 \cdot 10^{-11}$
PEZ16	0.170	$6.07 \cdot 10^{-15}$	$1.91 \cdot 10^{-12}$
CPH4	0.173	$8.54 \cdot 10^{-16}$	$4.92 \cdot 10^{-13}$
FH2096	0.131	$2.25 \cdot 10^{-16}$	$1.76 \cdot 10^{-13}$
vWF.X	0.192	$3.39 \cdot 10^{-17}$	$4.70 \cdot 10^{-14}$

Note. The loci are listed in the order of increasing F_{ST}. * Here and below the product of the genotype frequency of the previous and current loci; ** here and below the product of the genotype frequency of the previous and current loci.

The analysis of PIC measures by Botstein et al. (1980) revealed that all selected loci in the total sample of dogs are highly informative for the study of the DNA of both the domestic dog and the Grey wolf (Table 4).

In the combined sample of dogs, the minimum PIC value (0.472) was found at locus FH2096. The maximum PIC values were observed at FH2016 for the wolf (0.882) and at FH2016 for the dog (0.885). The average PIC values were 0.720 for the dog and 0.742 for the wolf, which can be considered significant for interpretation of results in a forensic genetic investigation.

While the allele sequences of the wolf and dog were identical, allele sequencing revealed the presence of simple repeats. Incomplete tandem repeats were identified in the alleles of the FH2016, FH2361, and FH2328 loci. Specifically, for the FH2016 and FH2328 loci, incomplete tandems were detected both in the sample of wolves and in the sample of dogs. For the FH2361 locus, microvariants were identified only in the sample of dogs.

Sequencing of the alleles of the FH2001 locus produced an unexpected result (Fig. 4). A 6 bp insertion located in the non-tandem region of the locus was found in the combined

Table 4. Values of the polymorphism information content of the selected loci for the samples of the Grey wolf and the domestic dog

Loci	PIC	
	Domestic dog	Grey wolf
FH2096	0.472	0.552
vWF.X	0.541	0.813
CPH4	0.566	0.817
FH2079	0.620	0.579
CPH12	0.651	0.580
FH2010	0.657	0.747
PEZ17	0.766	0.746
PEZ16	0.769	0.827
FH2001	0.774	0.701
FH2004	0.778	0.876
FH2054	0.801	0.759
FH2361	0.830	0.788
VGL3438	0.845	0.755
FH2328	0.848	0.713
FH2016	0.885	0.882
Average	0.720	0.742

Note. PIC – polymorphism information content of a locus.

sample of wolves and dogs. This insertion was observed only in long alleles (with 10 or more tandem repeats), and there were no alleles with 11 or more repeats that did not contain insertions. The FH2001 locus was not excluded from the final forensic panel, and special names were assigned to alleles with insertion ("10in"–"14in" alleles).

Based on the results, we created the CPlex test system, which contains 15 STR loci and two sex loci. We obtained the nucleotide sequences of all the identified alleles; these alleles were identified in the tandem format for compatibility of the panel with the instrumentation. The sequences were deposited in the GenBank database (Table 5).

The developed test system was validated in accordance with the protocol of the Scientific Working Group on DNA Analysis Methods (R.V. Guideline, 2004), and it was tested on benchmark samples and on real forensic objects. These methodological developments are being successfully used in expert practice when investigating the facts of illegal hunting, cruelty to animals and other offenses in the Republic of Belarus.

Conclusion

In this study, we selected 15 microsatellite loci and two sex loci that are suitable for forensic DNA examination of both the Grey wolf and the domestic dog. Furthermore, we designed a universal CPlex test system for the identification of individuals of the *C. lupus* species. Finally, we carried out an analysis of polymorphism and forensic parameters of loci and studied the genetic structure of the Belarusian populations of the *C. lupus* species. According to the results of statistical analysis of the dog and wolf genotype pools, the selected loci conform to the Hardy–Weinberg equilibrium. The coefficients of subdivision of the population for each STR locus of the test system were calculated, and the effectiveness of their use was proven. The developed CPlex test system was validated in accordance with the international standard and used in forensic research of cases of illegal hunting, animal attacks of people and livestock, as well as cases of cruelty to animals.

Based on the developed test system, we created two guides for practicing forensic experts, "The method of DNA-based identification of biological samples of the Grey wolf (*Canis lupus lupus*) and domestic dog (*Canis lupus familiaris*)" and "The method of using the computational tool for the analysis of genetic data of animals of the biological species *Canis lupus* – the Grey wolf (*Canis lupus lupus*) and the domestic dog (*Canis lupus familiaris*)". The computational tool contains arrays of genotypes and a mathematical apparatus that allows one to automate the analysis of data when identifying species of the Grey wolf or the domestic dog, calculating the reliability of an expert conclusion in an identification study or establishing biological relationship; it also allows conducting DNA fingerprinting registration of biological samples of

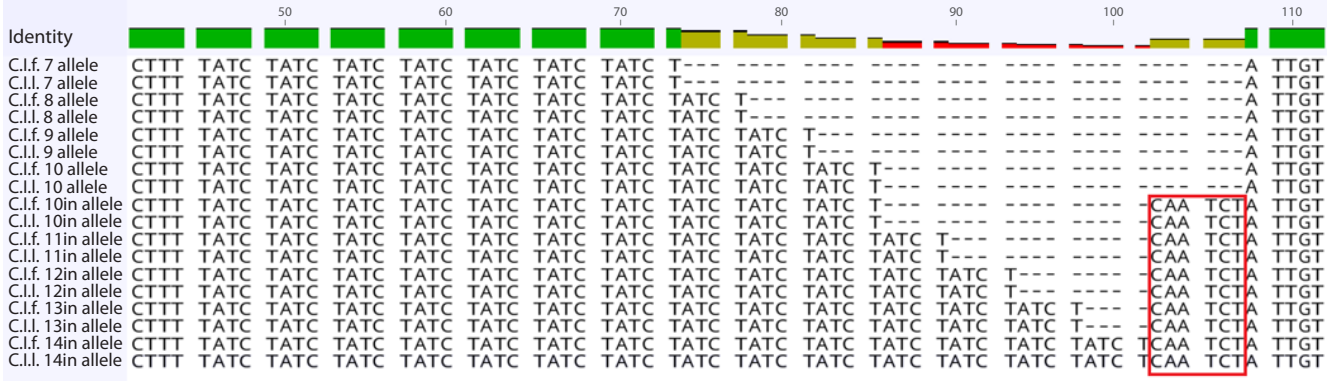

Fig. 4. Part of the sequence of the FH2001 locus in DNA of the dog and the wolf.

C.l.f. – domestic dog, C.l.l. – grey wolf. Red rectangle – 6 bp insertion in the non-tandem region of the locus.

Table 5. Characterization of microsatellite loci of the CPlex test system

Loci	Revealed range in bp	Tandem type and range	GenBank access number	
			Minimum length allele	Maximum length allele
DBX	246	–	OQ216490	
DBY	117	–	OQ216491	
FH2096	88–108	$(AATG)_{5–10}$	OQ216492	OQ216493
vWF.X	133–199	$(AGGAAT)_{5–16}$	OQ216494	OQ216495
CPH4	138–152	$(CA)_{14–22}$	OQ216496	OQ216497
FH2079	260–292	$(TGGA)_{6–14}$	OQ216498	OQ216499
CPH12	180–200	$(AC)_{5–19}$	OQ216500	OQ216501
FH2010	215–239	$(GAAT)_{7–13}$	OQ216502	OQ216503
PEZ17	190–222	$(TTTC)_{11–19}$	OQ216504	OQ216505
PEZ16	269–337	$(GAAA)_{5–22}$	OQ216506	OQ216507
FH2004	229–337	$(TTCT)_{11–38}$	OQ216508	OQ216509
FH2054	140–180	$(ATCT)_{13–23}$	OQ216510	OQ216511
VGL3438	101–145	$(AAAG)_{10–20}$	OQ216512	OQ216513
FH2001	124–158	$(ATCT)_{7–10}$	OQ216514	OQ216515
		$(ATCT)_{10–14}CAACTC$	OQ216516	OQ216517
FH2361	329–425	$(TCTT)_{11–35}$	OQ216518	OQ216519
		$(TCTT)_{11–18}TC$	OQ216520	OQ216521
FH2328	181–219	$(AAAG)_{7–16}$	OQ216522	OQ216523
		$(AAAG)_{15}AA(AAAG)_{1}$	OQ216524	
FH2016	276–340	$(CTTT)_{15–31}$	OQ216525	OQ216526
		$(CTTT)_{19–21}CT$	OQ216527	OQ216528

animals of *C. lupus* species in a forensic examination. The developed methods are included in the Register of forensic methods and other methodological materials of the State Forensic Examination Committee of the Republic of Belarus, which constitutes the implementation of the development in the national legal system.

References

Aspi J., Roininen E., Ruokonen M., Kojola I., Vila C. Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population. *Mol. Ecol*. 2006;15(6):1561-1576. DOI 10.1111/j.1365-294X.2006.02877.x

Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. GenBank. *Nucleic Acids Res*. 2005;34:D16-D20. DOI 10.1093/nar/gkj157

Boitani L. Wolf Conservation and Recovery. Chicago: Univ. of Chicago Press, 2003;317-340

Boom R.C.J.A., Sol C.J., Salimans M.M., Jansen C.L., Wertheim-van Dillen P.M., Van der Noordaa J.P.M.E. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol*. 1990;28(3):495-503. DOI 10.1128/jcm.28.3.495-503.1990

Botstein D., White R.L., Skolnick M., Davis R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet*. 1980;32(3):314

Buckleton J.S., Curran J.M., Walsh S.J. How reliable is the sub-population model in DNA testimony? *Forensic Sci. Int*. 2006;157(2-3):144-148. DOI 10.1016/j.forsciint.2005.04.004

Buckleton J.S., Bright J.A., Taylor D. Forensic DNA Evidence Interpretation. CRC Press, 2016. DOI 10.4324/9781315371115

Caniglia R., Fabbri E., Greco C., Galaverni M., Randi E. Forensic DNA against wildlife poaching: identification of a serial wolf killing in Italy. *Forensic Sci. Int. Genet*. 2010;4(5):334-338. DOI 10.1016/j.fsigen.2009.10.012

Đan M., Šnjegota D., Veličković N., Stefanović M., Obreht Vidaković D., Ćirović D. Genetic variability and population structure of grey wolf (*Canis lupus*) in Serbia. *Russ. J. Genet.* 2016;52(8):821-827. DOI 10.1134/S1022795416080044

DeNise S., Johnston E., Halverson J., Marshall K., Rosenfeld D., McKenn S., Sharp T., Edwards J. Power of exclusion for parentage verification and probability of match for identity in American kennel club breeds using 17 canine microsatellite markers. *Anim. Genet*. 2004;35(1):14-17. DOI 10.1046/j.1365-2052.2003.01074.x

Eichmann C., Berger B., Reinhold M., Lutz M., Parson W. Canine-specific STR typing of saliva traces on dog bite wounds. *Int. J. Leg. Med*. 2004;118(6):337-342. DOI 10.1007/s00414-004-0479-7

Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol*. 2005;14(8):2611-2620. DOI 10.1111/j.1365-294X.2005.02553.x

Excoffier L., Laval G., Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*. 2005;1:47-50

Fan Z., Silva P., Gronau I., Wang S., Armero A.S., Schweizer R.M., Ramirez O., Pollinger J., Galaverni M., Ortega Del-Vecchyo D., Du L., Zhang W., Zhang Z., Xing J., Vilà C., Marques-Bonet T., Godinho R., Yue B., Wayne R.K. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res*. 2016;26(2):163-173. DOI 10.1101/gr.197517.115

Freedman A.H., Wayne R.K. Deciphering the origin of dogs: from fossils to genomes. *Annu. Rev. Anim. Biosci*. 2017;5:281-307. DOI 10.1146/annurev-animal-022114-110937

Galinskayaa T.V., Schepetov D.M., Lysenkov S.N. Prejudices against microsatellite studies and how to resist them. *Russ. J. Genet.* 2019; 55(6):617-632. DOI 10.1134/S0016675819060043

Godinho R., Llaneza L., Blanco J.C., Lopes S., Alvares F., Garcia E.J., Palacios V., Cortés Y., Talegón J., Ferrand N. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol*. 2011;20(24):5154-5166. DOI 10.1111/j.1365-294x.2011.05345.x

Hall T., Biosciences I., Carlsbad C.J.G.B.B. BioEdit: an important software for molecular biology. *GERF Bull. Biosci*. 2011;2(1):60-61

Halverson J., Basten C. A PCR multiplex and database for forensic DNA identification of dogs. *J. Forensic Sci*. 2005;50(2):352-363

Hedrick P.W. Genetics of Populations. Boston: Jones and Bartlett, 2000

Kalinowski S.T., Taper M.L., Marshall T.C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol*. 2007;16(5):1099-1106. DOI 10.1111/j.1365-294x.2007.03089.x

Korablev M.P., Korablev N.P., Korablev P.N. Genetic diversity and population structure of the grey wolf (*Canis lupus* Linnaeus, 1758) and evidence of wolf × dog hybridisation in the centre of European Russia. *Mamm. Biol*. 2021;101(1):91-104. DOI 10.1007/s42991-020-00074-2

Magory Cohen T., King R., Dolev A., Boldo A., Lichter-Peled A., Kahila Bar-Gal G. Genetic characterization of populations of the golden jackal and the red fox in Israel. *Conserv. Genet*. 2013;14(1):55-63. DOI 10.1007/s10592-012-0423-1

Ministry of Forestry of the Republic of Belarus. Annual Forest Management Report. 2021 (in Russian)

Parker H.G., Dreger D.L., Rimbault M., Davis B.W., Mullen A.B., Carpintero-Ramirez G., Ostrander E.A. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep*. 2017;19(4):697-708. DOI 10.1016/j.celrep.2017.03.079

Peakall R.O.D., Smouse P.E. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*. 2006;6(1):288-295. DOI 10.1111/j.1471-8286.2005.01155.x

Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Exel. Population genetic software for teaching and research – an update. *Bioinformatics*. 2012;28:2537-2539. DOI 10.1093/bioinformatics/bts460

Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-959. DOI 10.1093/genetics/155.2.945

Radko A., Podbielska A. Microsatellite DNA analysis of genetic diversity and parentage testing in the popular dog breeds in Poland. *Genes*. 2021;12(4):485. DOI 10.3390/genes12040485

Revised Validation Guideline – Scientific Working Group on DNA Analysis Methods (SWGDAM). *Forensic Sci. Commun.* 2004;6(3). Available from: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm

Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 1977;74(12):5463-5467. DOI 10.1073/pnas.74.12.5463

Sastre N., Vila C., Salinas M., Bologov V.V., Urios V., Sánchez A., Francino O., Ramírez O. Signatures of demographic bottlenecks in European wolf populations. *Conserv. Genet*. 2011;12(3):701-712. DOI 10.1007/s10592-010-0177-6

Seddon J.M. Canid-specific primers for molecular sexing using tissue or non-invasive samples. *Conserv. Genet*. 2005;6(1):147-149. DOI 10.1007/s10592-004-7734-9

Stronen A.V., Aspi J., Caniglia R., Fabbri E., Galaverni M., Godinho R., Kvist L., Mattucci F., Nowak C., von Thaden A., Harmoinen J. Wolf-dog admixture highlights the need for methodological standards and multidisciplinary cooperation for effective governance of wild × domestic hybrids. *Biol. Conserv.* 2022;266:109467. DOI 10.1016/j.biocon.2022.109467

The Evaluation of Forensic DNA Evidence / Committee on DNA Forensic Science: an Update. National Research Council. Washington D.C.: National Academy Press, 1996

van Asch B., Alves C., Santos L., Pinheiro R., Pereira F., Gusmao L., Amorim A. Genetic profiles and sex identification of found-dead wolves determined by the use of an 11-loci PCR multiplex. *Forensic Sci. Int. Genet*. 2010;4(2):68-72. DOI 10.1016/j.fsigen.2009.05.003

van Oosterhout C., Hutchinson W.F., Wills D.P., Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*. 2004;4(3):535-538. DOI 10.1111/j.1471-8286.2004.00684.x

Verardi A., Lucchini V., Randi E. Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Mol. Ecol*. 2006; 15(10):2845-2855. DOI 10.1111/j.1365-294X.2006.02995.x

Wright S. Variability Within and Among Natural Populations. Evolution and the Genetics of Populations. Chicago: Univ. of Chicago Press, 1978

ORCID

A.E. Hrebianchuk orcid.org/0000-0002-1224-3275
I.S. Tsybovsky orcid.org/0000-0002-8611-8215

Acknowledgements. The authors express their gratitude to the Republican State Public Association "Belarusian Society of Hunters and Fishermen", veterinary center "Drug" and to the State Institution "Minsk city veterinary station" for their help in procuring a collection of biological samples.

Funding information. This research was carried out at Scientific and Practical Center of the State Forensic Examination Committee of the Republic of Belarus and was funded by the State Committee on Science and Technology of the Republic of Belarus (Project No. 20190195).

Conflict of interest. The authors declare no conflict of interest.

Received January 24, 2023. Revised July 4, 2023. Accepted July 11, 2023.