Thesis of the (PhD) Dissertation

ESTIMATION OF GENETIC VARIABILITY AMONG TRADITIONAL HUNGARIAN HORSES BY MITOCHONDRIAL DNA ANALYSIS

Nikolett Sziszkosz

Supervisor: Dr. András Jávor CSc.

UNIVERSITY OF DEBRECEN
Doctoral School of Animal Husbandry

TABLE OF CONTENTS

1. BACKGROUND OF THE Ph.D. THESIS ................................................................................. 3
2. AIMS OF THE RESEARCH .............................................................................................. 4
3. METHODS ....................................................................................................................... 5
   3.1 SAMPLE COLLECTION .......................................................................................... 5
   3.2 MITOCONDRIAL DNA TEST ............................................................................... 5
4. MAIN OBSERVATIONS OF THE THESIS ........................................................................ 9
   4.1 MOLECULAR PHYLOGENETIC ANALYSIS OF GIDRAN HORSES ......................... 9
   4.2 VERIFICATION OF MARE FAMILIES USING MOLECULAR MARKERS .................... 12
   4.3 MOLECULAR PHYLOGENETIC ANALYSIS OF HUCUL HORSES ............................. 15
   4.4 COMPARISON OF HUCUL CYTB AND D-LOOP HAPLOTYPES WITH ADDITIONAL PONY SEQUENCES ........................................................................................................ 16
   4.5 VERIFICATION OF HORSE MARE FAMILIES USING TWO MOLECULAR MARKERS ......................................................................................................................... 17
5. NEW AND NOVEL OBSERVATIONS OF THE THESIS .................................................... 20
6. APPLICABILITY OF THE RESULTS ................................................................................. 21
7. PUBLICATIONS IN THE TOPIC OF THE THESIS .......................................................... 25
1. BACKGROUND OF THE PhD THESIS

The traditional Hungarian breeds have been developed by professional breeding work and they became a part of our cultural heritage. Consequently, the preservation of their characteristic attributes and their genes as well as the analysis of their genetic structure is important (Balog, 1997; Mihók, 2016).

Gidrán has one of the smallest populations among the traditional Hungarian breeds, representing a significant cultural and genetic value in Hungarian livestock breeding (Mihók, 2006). Moreover, Hucul horses also come from an endangered group, if we consider the few numbers of mares (Jakabová et al., 2009). Comprehensive mapping of both breeds’ genetic structure is essential for the preservation of their valuable genetic resources. This aspect can be useful for the breed protection plans as well as for understanding their genetic status.

The mitochondrial DNA (mtDNA) is characterized by maternal inheritance and a high degree of polymorphisms. Moreover, it is also suitable for studying the genetic structure of maternal lines (Taberlet et al., 2011; Jianxing et al., 2012). Over the past decade, several studies were conducted using the mitochondrial DNA sequences of cytochrome b (CYTB) gene or the control region - also known as D-loop - for mapping the origins of domesticated and ancestral horses. The non-coding D-loop region is the most polymorphic part of the mitochondrial genome (Lopes et al., 2005). Along with the control region, the examination of the protein coding CYTB sequence polymorphisms plays an important role in origin verification, as well as it is useful for phylogenetic studies.

The main aim of my dissertation was to determine the diversity of Gidrán and Hucul mare lines using the aforementioned mtDNA markers. Our further goal was to reveal possible errors or errors that may occur during the management of stud books. Finally, I also investigated the usefulness of an optimized DNA-based method for the discrimination of male families by mtDNA haplotypes.
2. AIMS OF THE RESEARCH

- Mapping and characterizing the genetic structure of Hungarian horses using mitochondrial DNA markers.

- Identification of Gidrán and Hucul mtDNA haplotypes and analysis of the genetic diversity of species by mitochondrial markers.

- Bioinformatical comparison of the genetic features of Hungarian hucul mare populations with selected pony populations from GenBank.

- Comparison of the Gidrán and Hucul mares’ genealogical data with the molecular phylogenetic results in order to determine possible discrepancies in the management of traditional studbooks.
3. METHODS

3.1 SAMPLE COLLECTION

Altogether, I analysed 250 hair samples of Gidrán mares originated from different Hungarian breeds: Bakonycsernye, Balatonalmádi, Biharkeresztes, Budapest, Csákvár, Debrecen, Győr, Gyűrűs, Lulla, Marócpuszta, Nagycsepely, Réđics, Sárkeresztes, Siklós, Tata, Tiszafüred, Zalaszentgrót.

Furthermore, 267 hucul mares were also involved from different Hungarian breeds: Csemő, Gyűrűs, Hárskút, Izsák, Jósvafő (Aggtelek National Park), Kisoroszi, Ólmod, Pécsely, Pénzesgyőr, Pilismaró, Rakamaz, Solt, Szigetscép, Varbó in our study.

All horses’ hair samples included in this study were teared and posted in well-sealed envelopes by the breeder. To avoid contamination, all samples were kept separated from each other at reduced humidity and room temperature until the laboratory utility.

D-loop sequencing was not performed in case of seventeen hucul samples, the available DNA sequence data of the horses were downloaded in FASTA formats from GenBank database instead. The accession number of haplotype sequences determined by Priskin et al. (2010) were as follows: KC143336.1 - KC143355.1

In the bioinformatics analysis, 22 selected hucul haplotypes were compared with the 35 ponies’ CYTB and D-loop haplotype sequences which were previously obtained from the NCBI (National Center for Biotechnology Information) database. Since the sequences of two markers were investigated, the analysis was limited to samples with full mtDNA sequences.

3.2 MITOCONDRIAL DNA TEST

Genomic DNA extraction was made according to the FAO (FAO/IAEA, 2004) instructions. The freely available Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/) was used for PCR primer design. Primer pairs were designed to target two polymorphic sites which were previously described as suitable markers for phylogenetic studies: 1192 bp D-loops and 1140 bp long CYTB genes for different. Four of the seven primer pairs were specific to the D-loop region, and three primary pairs were specific to the CYTB. The specifications of the primers are shown in Table 1, while the sequences covered by them are shown in Figure 1.
Table 1.: Specifications of primers

<table>
<thead>
<tr>
<th>Primers</th>
<th>Primer sequences</th>
<th>The length of the products</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Dloop1_F</td>
<td>ACGACAACAATTTACCCTCA</td>
</tr>
<tr>
<td></td>
<td>Dloop1_R</td>
<td>GGGGAAAGAGGTGACAGA</td>
</tr>
<tr>
<td>B</td>
<td>Dloop2_F</td>
<td>CCCCCATACCCACCATACC</td>
</tr>
<tr>
<td></td>
<td>Dloop2_R</td>
<td>ATCTAGGGATGCTGTCCT</td>
</tr>
<tr>
<td>C</td>
<td>Dloop3_F</td>
<td>TCAGCAACCTCCCACTAC</td>
</tr>
<tr>
<td></td>
<td>Dloop3_R</td>
<td>ATCTAGGGGATGCTGTCCT</td>
</tr>
<tr>
<td>D</td>
<td>Dloop4_F</td>
<td>ACCTATTCCCGCATACC</td>
</tr>
<tr>
<td></td>
<td>Dloop4_R</td>
<td>ATCTAGGGGATGCTGTCCT</td>
</tr>
<tr>
<td>E</td>
<td>Cyt1_F</td>
<td>TCACACCTGGGAATCTAAC</td>
</tr>
<tr>
<td></td>
<td>Cyt1_R</td>
<td>GTCCGCGAATGCTAATTG</td>
</tr>
<tr>
<td>F</td>
<td>Cyt2_F</td>
<td>GCAATTCTGGCTATGTCCT</td>
</tr>
<tr>
<td></td>
<td>Cyt2_R</td>
<td>GTCCGCGAATGCTAATTG</td>
</tr>
<tr>
<td>G</td>
<td>Cyt3_F</td>
<td>ATCTACATCAAGCTGGTAT</td>
</tr>
<tr>
<td></td>
<td>Cyt3_R</td>
<td>ATCTACATCAAGCTGGTAT</td>
</tr>
</tbody>
</table>

Figure 1: The Equine mtDNA D-loop and CYTB sequences in FASTA format. Identical primer pairs are emphasized with the same colours.
After primer design, the most appropriate primer pair was selected from the seven primers by gradient (with multiple temperatures) PCR and the optimal primer adhesion temperature (62.4 °C) was also determined. During the gel electrophoresis, the research was continued with the "E" primer pair that gave the most distinct bands, which covered a longer fragment within the CYTB gene:

**CYTB gene:**

- 14115F 5’-TTCCACGTGGAATCTAACC-3’
- 15206R 5’-GTCCGCCGATTCATGTTAGT-3’

Within the D-loop region I worked with a primer pair covering 298 bp long mtDNA fragment and was previously used by Priskin et al. (2010):

**D-loop region:**

- 15444F 5’-ACCATCAACACCCAAAGCTG-3’
- 15742R 5’-GCTGATTTCCCGGCTTGGTG-3’ (Priskin et al., 2010)

The primary adhesion temperature was graded at 62.4 °C by the gradient PCR. The amplification was done by PCR using the MJ Research PTC-200 Thermal Cycler (MJ Research, Watertown, MA, USA).

Separation of nucleic acids by gel electrophoresis: 2% Seakem agarose (Lonza, USA), 1X TAE buffer (Thermo Fisher Scientific, Waltham, USA), 0.5 mg/ml Gelred (Biotium, Hayward, CA, USA) for the purpose. The results were checked under UV light.

Purification of the PCR product was carried out using the Viogene DNA / RNA Extraction PCR-M Clean Up System (Viogene-BioTek, Taipei, Taiwan) kit according to the manufacturer's instructions. The quality of the PCR products and DNA concentrations were determined using a NanoDrop ND-1000 spectrophotometer.

Sequencing was performed by Eurofins MWG-Operon (Ebersberg, Germany) and Macrogen Europe (Amsterdam, The Netherlands) with primers also used for PCR.

In the bioinformatic evaluation of the results, the validity of the read nucleotides was evaluated by CodonCode Aligner v. 4.2.7 (CodonCode Corporation, 2014). Sequences obtained from sequencing were compared with sequences downloaded from the NCBI database using the
ClustalW algorithm (Larkin et al., 2007). The nucleotide positions were determined based on the reference sequences available in the GenBank database (acc. nr.: X79547) and (acc. nr.: JN398377). Standard diversity values such as polymorphic sites, haplotype and nucleotide diversity were determined using DnaSP5 (Rozas et al., 2009) software. An Equus asinus (donkey) (acc. nr.: NC001788) sequence was used as an outgroup in case of both markers.

Previously unpublished haplogroups in the NCBI database were identified using the BLAST (Basic Local Alignment Search Tool) online match search engine (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

The DomeTree (http://www.dometree.org/) haplotype analysis was performed in case of both breeds. Phylogenetic trees were visualized by the DARwin for Windows software (http://darwin.cirad.fr/). The nucleotide positions of the DomeTree haplotype are based on the reference sequence (GenBank acc. nr.: JN398377).

The visualization of the mtDNA haplotypes of mare families and the genetic conditions of different pony sequences was visualised by the Network 4.6.0.0. (Bandelt et al., 1999) and PopArt (http://popart.otago.ac.nz) softwares using median-joining network method.

Phylogenetic analyses were performed using the MEGA6 (Tamura et al., 2013) software packages. The maximum likelihood algorithm was used for drawing phylogenetic trees (Hasegawa et al., 1985), moreover bootstrap analysis with 1000 values was also performed (Podani, 2003).

The most suitable nucleotide substitution model for the sequences was selected by the jModelTest (Posada, 2008) software. Accordingly, I used the Hasegawa-Kishino-Yano (HKY) plus gamma model for CYTB (Hasegawa et al., 1985) and the Tamura 3-parameter (T92) plus gamma models for the D-loop marker in case of both breeds.
4. MAIN OBSERVATIONS OF THE THESIS

4.1 MOLECULAR PHYLOGENETIC ANALYSIS OF GIDRÁN HORSES

In this study, 250 and 246 Gidrán samples evaluated in case of CYTB and D-loop markers, respectively. Following the DNA sequencing and bioinformatics normalization of the raw data, altogether 686 nucleotides within the CYTB gene and 202 nucleotides within the D-loop region were considered for further studies. Diversity indicators are summarized in Table 2.

Table 2: The diversity values of the mtDNA markers in Gidrán

<table>
<thead>
<tr>
<th>mtDNA markers</th>
<th>Nr. of nucleotides</th>
<th>n</th>
<th>Nr. of haplotypes</th>
<th>Polymorphic sites</th>
<th>Haplotype diversity ±SD</th>
<th>Nucleotide diversity ±SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYTB</td>
<td>686</td>
<td>250</td>
<td>24</td>
<td>23</td>
<td>0.874 ± 0.011</td>
<td>0.005 ± 0.001</td>
</tr>
<tr>
<td>D-loop</td>
<td>202</td>
<td>246</td>
<td>32</td>
<td>26</td>
<td>0.914 ± 0.008</td>
<td>0.021 ± 0.001</td>
</tr>
<tr>
<td>Merged*</td>
<td>893</td>
<td>242</td>
<td>49</td>
<td>-</td>
<td>0.940 ± 0.006</td>
<td>0.008 ± 0.001</td>
</tr>
</tbody>
</table>

*CYTB and D-loop markers were combined for the analysis

The analysis of the sequences both mtDNA marker analyses confirmed the significant genetic variability of Gidrán horses. Twenty-three polymorphic sites were found in CYTB (twenty-one SNPs, and two insertion/deletion), representing 3.35% polymorphism considering the full length (686 bp) sequence. The D-loop marker exhibited somewhat more 26 variable sites, representing 12.9% of the total 202 bp sequence. Among polymorphic nucleotides, besides the 25 SNPs, only one insertion/deletion was identified. Both marker sequences were rich in adenine (A) / thymine (T), representing 55% for CYTB markers and 64.1% for D-loop.

Both markers showed high haplotype diversity values: 0.8735 ± 0.011 for CYTB and 0.9136 ± 0.008 for D-loop. The nucleotide diversity values were also high: 0.00472 ± 0.00017 (CYTB) and 0.02091 ± 0.00068 (D-loop). The paired genetic distances between haplotypes were 0.001-0.013 (CYTB) and 0.005-0.063 (D-loops). According to the phylogenetic literature, it can be stated that in contrast to other species (Ali et al., 2015) studying the CYTB marker is relatively rare in horses (Li, et al., 2006, Jianxing et al. The observed CYTB nucleotide diversity parameters are similar to the data of Chinese domestic horses. Yues et al. (2012) described 114 different haplotypes by studying the 323 Chinese domestic horses and 84 selected sequences from the GenBank database.
The nucleotide diversity value is between 0.005 and 0.002. In addition, haplotype diversity ranges from 0.706 to 0.975, which also overlaps with my results (Gidrán: 0.874; Hucul: 0.835) (Yue et al., 2012). Qin et al. (2009) identified twenty-two Lichuan horses using 1140 bp length CYTB sequences. Similarly, to the results described above, high haplotype and nucleotide diversity values (0.840 and 0.048) were observed. The observed variability of CYTB in my study also highlights the genetic diversity of Gidrán, however the observed haplotype numbers and nucleotide diversity were lower compared to the D-loop marker.

Taking into account the number of D-loop haplotypes, similar results have been already described in in Lusitan (27 haplotypes / 145 horses) (Lopes et al., 2005), Lipizzan (37 haplotypes / 212 horses) (Kavar et al., 2002), Arabian (27 haplotypes / 200 horses) (Bowling et al., 2000) horses, and higher numbers in case of Kiso horses (7 haplotypes / 136 horses) (Takasu et al., 2014). The observed haplotype and nucleotide diversity values were also consistent with previous research results. Recent results are roughly similar to the Iranian horse population’s data (0.02) (Moridi et al., 2013). Considering the relatively high values, the involved Hungarian horse populations are genetically more variable than the Kerry bog (0.0155 ± 0.0040), the Sulphur (mustard) mustang (0.001 ± 0.002). On the other hand, the genetic structure of Marwari (0.03973 ± 0.01262) or Sorraia (0.104 ± 0.012) is even more diverse than the Gidrán or Hucul (Luís et al., 2006; Prystupa et al., 2012; Devi & Ghosh, 2013). Sequences of the haplotypes were uploaded to the GenBank database with the following identifiers: KT792934 - KT792957 and KT818891 - KT818922.

All phylogenetic analyses regarding both markers were performed separately and in a combination of the two sequences. Of the 31 Gidrán mare families, I successfully isolated 24 different haplotypes using CYTB marker, and 32 different haplotypes using the D-loop marker. Similarly, diverse data were described in a previous Hucul study (Kusza et al., 2013) as well as in Zemaitukan horses, even though both varieties had a heavy bottleneck effect during their history (Cothran et al., 2005; Kusza et al., 2013). The observed high D-loop and CYTB haplotype diversity confirms the fact that Gidrán originates from several maternal lines, which is supported by the fact that the recently known male families are derived from 16 founding mares (Mihók, 2006).

Two different reference sequences were used in the phylogenetic analysis of the Gidrán horses: X79547 and JN398377. Most phylogenetic studies use only the X79547 sequence published in 2005, which makes it possible to classify D-loop haplotypes into haplotypes as defined by Jansen et al. (2002) and Achilli et al. (2012). On the other hand, novel studies, such as the
DomeTree database, use the JN398377 GenBank reference sequence for phylogenetic studies. This latter sequence was realized in 2012 and it is more accurate than the previous X79547 sequence.

Two haplotypes (Ht1\textsubscript{CYTB} and Ht1\textsubscript{D-loop}) of the Gidrán haplotypes was identical to the reference sequence, and haplotypes differed in maximum of 6 (CYTB) and 9 (D-loop) nucleotides.

Of the 24 CYTB haplotypes Ht1\textsubscript{CYTB} (n=54), Ht2\textsubscript{CYTB} (n=49) and Ht6\textsubscript{CYTB} (n=44) were the most common haplotypes. Seven haplotypes, however, represented only one to one individual Ht11\textsubscript{CYTB}, Ht17\textsubscript{CYTB}, Ht18\textsubscript{CYTB}, Ht20\textsubscript{CYTB}, Ht21\textsubscript{CYTB}, Ht22\textsubscript{CYTB}, and Ht24\textsubscript{CYTB}. The phylogenetic "maximum likelihood" tree was represented by the haplotype type 250 of the 250 Gidrán mare. The most common haplotypes for the D-loop marker were Ht6\textsubscript{D-loop} (n=47), Ht16\textsubscript{D-loop} (n=35) and Ht1\textsubscript{D-loop} (n=25). Ten haplotypes, however, were found in only one animal: Ht14\textsubscript{D-loop}, Ht17\textsubscript{D-loop}, Ht18\textsubscript{D-loop}, Ht21\textsubscript{D-loop}, Ht24\textsubscript{D-loop}, Ht26\textsubscript{D-loop}, Ht28\textsubscript{D-loop}, Ht29\textsubscript{D-loop}, Ht30\textsubscript{D-loop}, and Ht31\textsubscript{D-loop}.

The performed Basic Local Alignment Search Tool (BLAST) analysis among NCBI sequences revealed six new CYTB (Ht5\textsubscript{CYTB}, Ht8\textsubscript{CYTB}, Ht11\textsubscript{CYTB}, Ht14\textsubscript{CYTB}, Ht20\textsubscript{CYTB}, and Ht21\textsubscript{CYTB}) and four D-loop (Ht12\textsubscript{D-loop}, Ht28\textsubscript{D-loop}, Ht29\textsubscript{D-loop}, and Ht32\textsubscript{D-loop}) haplotypes.

All haplotypes were also classified into haplogroups. The 32 D-loop haplotypes, defined by Jansen et al. (2002), can be classified into seven main D-loop haplotypes with a distribution of: A: 31\%, B: 3\%, C: 28\%, D: 19\%, E: 3\%, F: 13\%, and G: 3\%. Considering the Achilli’s D-loop classification which contains 18 haploid groups, only the haplogroup E did not represent itself in the examined Gidrán mare population.

According to the DomeTree data (Peng et al., 2015), the CYTB haplotypes were classified into ten (A1b, AQ, B1a2, D, G1a, G1b, G3, H, HQ and MN), while D-loop haplotypes were classified into eighteen (A1d, A3, AB, B1, B1a, D, G, G1, G1a, H1, I, J1, J2, K, L, O, OP, R) haplogroups. Unlike other haplogroup classifications, the DomeTree database is based on a full mtDNA sequence and not only on the D-loop sequences. From this point of view, I also classified the samples based on the two markers. In this case nine haplogroups have been obtained (A1, AB, H, H-I, J-K, M-N, M-Q, B1, B1a).
4.2 VERIFICATION OF MARE FAMILIES USING MOLECULAR MARKERS

The practical usage of mtDNA markers was also analysed for the identification of possible errors / misspellings in the studbooks as well as the investigation of the efficacy of mtDNA analysis in the classification of mare families by haplotypes.

In the first step, phylogenetic trees were prepared separately for each marker. Of the thirty-one mare families, 15 (48.4%) haplotypes were separated by CYTB and 17 (54.8%) haplotypes were separated by D-loop marker. The borodi 14 and 18, as well as mezőhegyesi 1 families could be separated by only the usage of CYTB, while the family of 2, 3, 19 and borodi 2 families could be separated by only the usage of the D-loop markers.

It is important to mention, that some families were not necessarily separated into a distinct haplotype, or multiple mare families share a haplotype. Consequently, the combined use of CYTB and D-loop markers (Pedrosa et al., 2005) could be a possible solution for the more accurate association of mare family and mtDNA haplotypes. In the next phase of my work, I also performed all analysis with the combining markers. In the case of the combined analysis, I found the network representation as the most suitable visualization for the haplotypes of mare families due to the large number of individuals (Figure 2). Based on this observation circular trees were not created for hucul horses.

In line with our expectations, using CYTB and D-loop, I gained more accurate results, because seven additional Gidrán families (mezőhegyesi 7, 8 and 21, borodi 5, 17, and 19, népies 23) were also separated. With the combination of the two markers, borodi 1 and borodi 7 formed a common haplotype. It is important to mention that none of the markers could separate five mare families (mezőhegyesi 5, 11, 13, 14 and borodi 18).
Figure 2: A median-joining phylogenetic network (Bandelt et al., 1999) comprising a total of 242 Gidrán sample sequences using merged CYTB and D-loop markers. The mare families were unknown not known for yellow-labelled 247G and 202G samples. The abbreviation of mare family names in the figure are the following: m = mezőhgyesi; b = borodi; n = népies.

Altogether the mtDNA polymorphisms based studbook classification were performed in 247 Gidrán mare samples. As a result of the molecular study, the matching of haplotypes to the mare family indicated in the studbook were succeeded in altogether 227 (92%) individuals.
Considering the total number of horses in recent study, nine individuals (3.6%) felt into a haplotype that is typical of other mare families indicating misspelling in the studbook management. This rate is excellent in comparison to the stud books of Lipica and Polish horses, where the discrepancy was around 11-11% (Kavar et al., 2002). The differences in the case of two animals affected the mezőhegyesi 2, 14 and 19, and in one case by mezőhegyesi 7, 13 and the népies 23 family.

Almost 1.6% of the samples (of different mare families) belonged to the same haplogroup which can be explained by several reasons: 1.) misspellings in the studbook, 2.) common origin of the horses, or 3.) the low-rate polymorphisms in the examined markers. Consequently, two-two mare families formed a common haplotype (borodi 1 and 7 and mezőhegyesi 5 and 11). Horses from the borodi 7 family did not form a separate haplotype, as did the members of the mezőhegyesi 5 family. The mares of borodi 7 family shared a haplotype with borodi 1, while mares of mezőhegyesi 5 shared a haplotype with the members of mezőhegyesi 11 family. These differences are in line with the data on the Gidrán breeds’ history (Mihók, 2006).

A total of 32 haplotypes were isolated by the D-loop marker, so we could map mare families that could not be isolated by the single CYTB. These were mezőhegyesi 2, 3, and 19, as well as borodi 2. On the other hand, the CYTB sequences were more informative for the examination of borodi 14, borodi 18, the mezőhegyesi 1 families (Figure 2). The significance of the combined use of the two mtDNA markers is emphasized by the fact that the separation of eight Gidrán mare families were possible only by the merged markers, moreover the members of the borodi 19 family were only distinguished by this approach as well.

Interestingly, the members of mezőhegyesi 4 mare family form two distinct haplotypes using both markers. Furthermore, horses of this family formed two separated groups with merged markers. This observation indicates mezőhegyesi 4 mare family can be divided into two molecular sub-families. This phenomenon can be explained by the fact that this mare family had been divided into two families during the time (Mihók, 2006). The mare families were unknown of three horses, but they also formed separated haplotypes.
4.3 MOLECULAR PHYLOGENETIC ANALYSIS OF HUCUL HORSES

Among the collected hucul samples, 265 samples were assessed for the CYTB marker and 267 samples for the D-loop region. The diversity values of the Hucul are summarized in Table 3. Along with the values of the two mtDNA markers – similar to the Gidrán horses – results of combined marker sequences were also presented here.

Table 3: The diversity values of the mtDNA markers in Hucul

<table>
<thead>
<tr>
<th>mtDNA markers</th>
<th>Number of nucleotides</th>
<th>n</th>
<th>Nr. of haplotypes</th>
<th>Polymorphic sites</th>
<th>Haplotype diversity ±SD</th>
<th>Nucleotide diversity ±SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYTB</td>
<td>686</td>
<td>265</td>
<td>13</td>
<td>13</td>
<td>0.835 ± 0.009</td>
<td>0.005 ± 0.001</td>
</tr>
<tr>
<td>D-loop</td>
<td>202</td>
<td>267</td>
<td>22</td>
<td>21</td>
<td>0.878 ± 0.009</td>
<td>0.024 ± 0.001</td>
</tr>
<tr>
<td>Merged*</td>
<td>893</td>
<td>240</td>
<td>30</td>
<td>-</td>
<td>0.889 ± 0.01</td>
<td>0.009 ± 0.001</td>
</tr>
</tbody>
</table>

*CYTB and D-loop markers were combined for the analysis

Compared to the Gidrán breed, the genetic structure of hucul horses has been analysed in earlier publications. These studies are mainly based on polymorphisms of D-loop sequences or microsatellite examinations (Kusza et al., 2013; Priskin et al., 2010; Georgescu et al., 2011). The significance of the CYTB marker analysis has not yet been investigated in Hucul horses. Using the CYTB marker, altogether 13 variable positions were identified (all nucleotide changes were SNPs). Considering the 686 bp length of the sequence, this number indicates 1.9% polymorphism. This value is considerably smaller than Gidrán’s CYTB polymorphism percentages, which implies that CYTB may be less informative about the Hucul diversity. On the other hand, the D-loop was more polymorphic than CYTB. The 202 bp long D-loop marker contained 22 polymorphic sites, covering 10.9% of the total sequence, which is close to the ratio found in Gidrán, and only one insertion / deletion (indel) was identified among polymorphic nucleotides.

Similarly to Gidrán, high haplotype diversity was also observed in hucul horses: 0.835 ± 0.009 for CYTB and 0.878 ± 0.009 for D-loop. In addition, the nucleotide diversity values were also high: 0.00484 ± 0.00013 (CYTB) and 0.02356 ± 0.00055 (D-loop). The genetic distance between the haplotypes in case of CYTB was between 0.001 and 0.013, while in case of the D-loop the distance was between 0.005 and 0.063. Previous studies also report a similar variability in the genetic structure of Hucul, however the Hucul has been affected by two known bottleneck effects (Kusza et al., 2013; Priskin et al., 2010)
Of the 31 Hucul mare families, 13 formed distinct haplotypes by the CYTB marker, while 22 formed distinct haplotypes by D-loop. Of the thirteen CYTB haplotypes Ht1\textsubscript{CYTB} (n = 58), Ht10\textsubscript{CYTB} (n = 57) and H2\textsubscript{CYTB} (n = 56) were the most frequent, while Ht4\textsubscript{CYTB}, Ht8\textsubscript{CYTB}, Ht12\textsubscript{CYTB} haplotypes were identified for only one-one individual. The relationship between the 13 haplotypes of CYTB was visualized on a "maximum likelihood" phylogenetic tree.

The most frequent D-loop haplotypes were Ht5\textsubscript{D-loop} (n=62), Ht4\textsubscript{D-loop} (n=38), Ht3\textsubscript{D-loop} (n=37), Ht6\textsubscript{D-loop} (n=36), and nine haplotypes Ht12\textsubscript{D-loop}, Ht14\textsubscript{D-loop}, H15\textsubscript{D-loop}, Ht16\textsubscript{D-loop}, Ht17\textsubscript{D-loop}, Ht18\textsubscript{D-loop}, Ht19\textsubscript{D-loop}, Ht20\textsubscript{D-loop}, Ht21\textsubscript{D-loop}, Ht22\textsubscript{D-loop} were unique to one and one animal.

The D-loop haplotype 22 can be classified into five main D-loop haplotypes: A: 45%, C: 23%, D: 14%, F: 14%, and G: 5%. Haplogroup A was dominant in Hucul horses, similar to the results of the study by Kusza et al. (2013). According to Jansen et al. (2002) the majority of the Przewalski horses belong to the "A" haplogroup. Interestingly, Przewalski horses are one of the closest relatives of Huculs, so the obtained results further strengthen this relationship between them.

Regarding to the D-loop haplogroups described by Achilli et al. (2012), Hucul mares could be classified into nine haplogroups (A1; A'B'C'D'; E'F'G'; I; L; M; N; O; R). Based on DomeTree classification, I observed altogether 15 (A, A1b, A2, A-I, B1b, E-G, F-G, G, L, L2, L3a2, M-Q, Q2a, R) different haplogroups.

### 4.4 COMPARISON OF HUCUL CYTB AND D-LOOP HAPLOTYPES WITH ADDITIONAL PONY SEQUENCES

The hucul haplotypes obtained during our work were compared with additional ponies’ sequences of different geographic areas by bioinformatics analysis. The main purpose of these investigations was to analyse the genetic diversity among ponies (including Hucul) from different geographic areas. For more reliable results, a total of 35 samples were selected where the CYTB and the D-loop sequences were also available.

Based on our results, the genetic structure of the Hungarian Hucul horses is more diverse, compared to the data of other ponies. Six Hucul CYTB haplotypes formed common haplotypes with Yakut, Shetland, Przewlalskii, Mongolian, Caspian, Chincoteague, Exmoor, Konik, Noriker, Fjord, Icelandic and German riding ponies (except Debao and Welsh ponies). Regarding D-loop, eight haplotypes of the Hucul haplotypes formed a common haplotype with Yakut, Mongolian, Caspian,
Welsh, Chincoteague, Exmoor, Conic, German Riding Pony, and in 14 cases formed separated group. Based on the results obtained, mtDNA markers because of their polymorphic features are useful for mapping the genetic structure within the given species, as well as for isolation mare families. On the other hand, they are not suitable for separating animals according to the geographical habitat.

4.5 VERIFICATION OF HORSE MARE FAMILIES USING TWO MOLECULAR MARKERS

A comparative analysis of mtDNA haplotypes and mare families was performed in a total of 240 mares involved the D-loop sequences of 19 Hucul mares from a study by Priskin et al. (2010).

The mtDNA analysis showed clearly identical results with studbook registry in case of 187 (78%) horses. Aspiráns and Árvácska families from Aggtelek represent the two most frequent mare families among the Hungarian Hucul mare families. Moreover, mares of the 12 Sarata and 4 Kitca originated from Lucina are also common. Based on the molecular results, these individuals formed well-separated, mare-family-specific haplotypes. It is important to point out that the members of Apiráns family formed same haplotypes.

Incomplete studbook classification can happen in case of 30 mares (12.5%). In these cases, the individuals belonging to the above-mentioned families did not create a common haplotype using mtDNA markers. Due to the large number of cases, it is not surprising that most misspellings happened in the following families: Árvácska (6), 4 Kitca (5), 12 Sarata (5). During the comparison, I found three errors in the 882 Gelnica family, which is originated from the formally Slovakian breeds. In addition, the possibility of errors was also found in the studbook classification of Wrona, 5 Plosca, 3 Tatarca, 2 Lucina, 17 Aglaia, 11 Rotunda and 1 Panca families. Mare families with such a large number of individuals should be subjected to greater selection pressure to maintain the genetic balance.

In twenty-four cases (10%), multiple mare families belonged to a single haplotype. In eight cases, the individuals of the 5 Plosca family had a common haplotype with Aspiráns horses. These results may suggest common origins of the two families.

Lower selection pressure should be subjected to families with smaller sample size. Seven mares formed a common haplotype with the 3 Tatarca, 70 Sekacka and Wydra families. I case of
nine horses, 4 Kitca and 17 Aglaia shared a common haplotype. Due to the diminished number of individuals, the position of 3 Tatarca family is endangered among Hucul mare families. Based on our results, members of the 3 Tatarca family did not create a specific haplotype, but shared a common haplotype with the Wydra and the Sekacka horses. It is important to note, that studbook classification of four horses may be registry errors. In case of two members of Árvácska family formed a common haplotype with 4 Kitca and 17 Aglaia horses based on the molecular study. Similarly, horses of 2 Lucina and the 882 Gelnica families also shared a common haplotype with 4 Kitca and 17 Aglaia horses. Increase of the number of these mare families in Hungary is also important, but it is not impossible because there is an adequate number of individuals in Slovakia and Poland (Mihók, 2016). The mtDNS results also confirmed that the members of Wydra and Sekacka families besides 3 Tatarca also deserve a special attention.

In the study, mare families were unknown in the case of eight mares (3%). Unique haplotypes were observed in the case of two mares (5 Plosca and Wrona). Even though there were only one sample available from four (825 Agla, 84 Polonia, 90 Macocha, Bajkalka) families, each of the four families formed a separate haplotype.

The comparative analysis of mtDNA haplotypes and the studbook data was also performed in case 19 Hucul samples previously published by Priskin et al. (2010), which also represented the infrequent mare families like 5 Plosca, 825 Agla, 84 Polonia, Bajkalka, Bukovina, Wolga. The aim of this study was to test the suitability and effectiveness of combined (CYTB - D - loop) markers.

Thirteen individuals (68%) with our combined marker haplotypes were similar to the result of Priskin et al. (2010). The only individual in 84 Polonia family formed a unique haplotype using combined markers.

For four samples, the use of combined markers proved to be more effective in associating mare families with mtDNA haplotypes. Two horses belonging to the Bucovina family formed a separated haplotype. In addition, each horse of the 825 Agla and Bajkalka families also formed a separated haplotype. These horses in Priskin's study formed a common "825 Agla/Lucina/Bukovina" haplotypes using the single D-loop marker. The accuracy of the combined analysis emphasizes that 2 Lucina horses did not overlap with these samples.

While in Priskin's study the "Hroby Bolyhos" belonged to the 5 Plosca family formed a distinct haplotype, based on my results, it shared a haplotype with Aspiráns horses.
Figure 3: A median-joining phylogenetic network (Bandelt et al., 1999) comprising 240 Hucul sequences of the merged CYTB and D-loop markers.
5. NEW AND NOVEL OBSERVATIONS OF THE THESIS

1. In my dissertation, the genetic structure of Gidrán’s and Hucul’s maternal lines was analysed. Based on the best of our knowledge, my study was the first mtDNA based analysis in Gidrán horses. Furthermore, I was the first to describe the sequence polymorphisms of CYTB gene, as well as the first to identify its haplotypes in Hucul horses. I separated 24 Gidrán and 13 Hucul haplotypes according to the polymorphisms of CYTB, and 32 Gidrán and 22 Hucul haplotypes according to the D-loop polymorphisms. In this study, I identified six novel CYTB (Ht5\textsubscript{CYTB}, Ht8\textsubscript{CYTB}, Ht11\textsubscript{CYTB}, Ht14\textsubscript{CYTB}, Ht20\textsubscript{CYTB}, Ht21\textsubscript{CYTB}) and for D-loop (Ht12\textsubscript{D-loop}, Ht28\textsubscript{D-loop}, Ht29\textsubscript{D-loop}, Ht32\textsubscript{D-loop}) haplotypes of Gidrán horses, and two new D-loop (Ht16\textsubscript{D-loop}, Ht18\textsubscript{D-loop}) haplotypes in Hucul horses. Sequences of both CYTB and D-loop haplotypes were uploaded to the freely available GenBank database.

2. I described, that the Hungarian Gidrán and Hucul traditional Hungarian horses assess a high grade of genetic diversity. It is also supported by the observed high diversity values. The haplotype diversity value was 0,874 (Gidrán) and 0,835 (Hucul), as well as 0,914 (Gidrán) and 0,878 (Hucul) regarding the CYTB and D-loop markers, restrictively.

3. A comparative analysis between Hucul and other ponies from different geographic localisation revealed, that both CYTB and D-loop are appropriate markers for mapping genetic heterogeneity of a given breed, but they are not suitable for the separation of haplotypes among different geographical areas.

4. I confirmed, that the management of studbooks is rather adequate. I found that based on the mtDNA haplotypes 91,9% of Gidrán mares and 78% of Hucul horses showed associations with the studbook data.

5. I established, that the combined CYTB and D-loop mtDNA markers allow a more adequate comparison of the molecular data with studbook registry. These markers can be used for mapping the genetic structure of other traditional horse breeds, as well as for the genetic based verification of studbooks.
6. APPLICABILITY OF THE RESULTS

1. In my dissertation, I analysed the diversity of mare families of Gidrán and Hucul horses using mtDNA markers. I made the sequences of the haplotypes available in the GenBank database. I demonstrated, that both breeds assess mare family specific mtDNA haplotypes.

2. One novelty and the result of my research, that the combination of CYTB and D-loop markers efficiently applied to the comparison of studbook data with mtDNA haplotypes.

3. The results of mtDNA analysis, as well as the observed misspellings in the studbooks could be useful in breeding programs of the National Kisbéri-félvér and Gidrán Breeders Association and the National Pony and Small Horse Breeders Association.

4. In my dissertation, I highlighted the importance of the molecular marker based classification of horses. However, studbook’s data are mostly adequate, mtDNA markers could be used for the classification of horses in question into mare families.

5. My results could be also useful for the preservation plans of other breeds, allowing the identification of the genetic status of the species and maintain the rare haplotypes.
7. REFERENCES


8. PUBLICATIONS IN THE TOPIC OF THE THESIS

Candidate: Nikolett Sziszkszó
Neptun ID: KL9K79
Doctoral School: Doctoral School of Animal Husbandry
MTMT ID: 10050429

List of publications related to the dissertation

Hungarian book chapters (1)

Foreign language Hungarian book chapters (1)

Hungarian scientific articles in Hungarian journals (2)

Agrártud. közl. 57, 75-79, 2014. ISSN: 1587-1262.

Foreign language scientific articles in Hungarian journals (1)
5. Sziszkszó, N., Jávor, A., Kuszsa, S.: Development of horse molecular genetics and the diagnostics of most important monogenic hereditary diseases.
Állatteny. takarm. 65 (3), 71-81, 2016. ISSN: 0230-1614.
Foreign language scientific articles in international journals (1)
   DOI: http://dx.doi.org/10.7717/peerj.1894
   IF: 2.177

List of other publications

Hungarian books (1)

Foreign language scientific articles in international journals (2)
   DOI: http://dx.doi.org/10.18388/abp.2014_846
   IF: 1.187


Total IF of journals (all publications): 3,364
Total IF of journals (publications related to the dissertation): 2,177

The Candidate's publication data submitted to the IDEa Tudósitét have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

28 August, 2017