VETERINARSKI ARHIV 86 (4), 573-580, 2016

Haplotype analysis of polymorphisms in the American mink (*Neovison vison*) growth hormone gene helps to explain the reasons for their linkage disequilibrium and ancestral haplotypes differentiation

Jakub J. Skorupski[1,2]*

[1]*Department of Ecology and Environmental Protection, Faculty of Biology, University of Szczecin, Szczecin, Poland*

[2]*Department of Genetics and Animal Breeding, Faculty of Biotechnology and Animal Husbandry, West Pomeranian University of Technology, Szczecin, Poland*

SKORUPSKI, J. J.: Haplotype analysis of polymorphisms in the American mink (*Neovison vison*) growth hormone gene helps to explain the reasons for their linkage disequilibrium and ancestral haplotypes differentiation. Vet. arhiv 86, 573-580, 2016.

ABSTRACT

Haplotypes of the American mink (*Neovison vison*) growth hormone gene, phased from diploid multilocus genotypes of 389 individuals, were analysed. Forty-four different haplotypes, for 14 detected polymorphic loci, were identified. Significant gametic disequilibrium occurs in the case of four variable sites, with a correlation coefficient ranging from 50% to 94%. Four haplotype blocks were defined. The results of the D-statistics, conducted for the studied population, suggest that the observed linkage disequilibrium was caused by limited migration and genetic drift. Moreover, the described LD indicates a relatively low frequency of recombination events between variable loci of the American mink growth hormone gene. The developed cladogram reveals the existence of two distinctly different and internally consistent clades, which may indicate the existence of two main patterns of ancestral haplotypes, from which all the others derived. Moreover, the internal structure of the cladogram reveals their clear hierarchical arrangement, which allowed to draw conclusions regarding the origin-relationships of the identified haplotypes.

Key words: American mink, genetic polymorphism, growth hormone gene, haplotype analysis, linkage disequilibrium, Ohta's D-statistics

Introduction

One of the greatest advantages of large-scale, multilocus SNP (single nucleotide polymorphism) analysis is its high power and efficiency in revealing fine-scale population

*Corresponding author:

Jakub Jan Skorupski, PhD., Eng., Wąska 13 St., 71-415 Szczecin, Poland, Phone: +48 512 014 658, E-mail: jakub.skorupski@usz.edu.pl

ISSN 0372-5480
Printed in Croatia

genetic diversity and phylogenetic history (SHRIVER et al., 2005; BRITO and SCOTT, 2009). This is due to their simple patterns of variation, relatively low mutation rates and low levels of homoplasy, as well as the high potential for easy, cost-effective and automated detection (BRITO and SCOTT, 2009). Therefore, the most applicable are analyses of multiple SNPs simultaneously, based on haplotypes, defined as a set of SNP variants on a single chromosome copy (HALLDÓRSSON et al., 2003).

The aim of this study was to analyse haplotypes derived from the American mink (*Neovison vison* Schreb., 1777) growth hormone gene (*mGH*) with multiple polymorphisms.

Materials and methods

The analysis involved genomic DNA isolated from muscle tissue of 389 individual American mink. Samples were collected from the slaughter-waste of ranch mink, carcasses of animals killed on roads and ready-to-use DNA samples (Nova Scotia Agricultural College, West Iceland Centre of Natural History).

PCR amplification of the *mGH* gene (GenBank: JX489617.2) was conducted with primers, following the protocol described by SKORUPSKI (2014). All resulting amplicons were subjected to Sanger sequencing, aimed at detecting genetic polymorphisms within the gene and direct genotyping. For this purpose, the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) was used. The temperature-time profile of the sequential PCR amplification included initial denaturation at 96 °C/1 min, 25 cycles of denaturation at 96 °C/10 s, annealing at 50-52 °C/5 s, and polynucleotide chain synthesis at 60 °C/4 min. The sequential PCR products were cleaned up using the ExTerminator kit (A&A Biotechnology, Poland). The purified products were separated and read in a capillary sequencer, 3730xl DNA Analyzer (Applied Biosystems, USA). Polymorphic sites were identified by direct analysis of obtained sequences.

The PHASE v2.1.1 software was used for reconstructing (phasing) of haplotypes from established diploid multilocus genotypes (STEPHENS and SCHEET, 2005). Analysis of the haplotype diversity was conducted with the Haplotype Analysis extension for Excel (MS Office), and included calculation of the number of observed and possible haplotypes (ELIADES and ELIADES, 2009).

The Haploview v.4.2 software was used to conduct a linkage disequilibrium (LD) analysis between the identified variable loci, as well as to identify the haplotype blocks, using the solid spine of the LD algorithm (BARRETT et al., 2005). The same software was used to calculate the correlation coefficient for each pair of loci (r^2), in relation to the identified haplotypes (BARRETT et al., 2005; SLATKIN, 2008). Multilocus analysis, described by SMOUSE et al. (1983), allowed determination of the average correlation between alleles at all variable loci.

To determine the cause of the discovered gametic disequilibrium, the POPGENE software (YEH and BOYLE, 1997) was used to perform Ohta's two-locus population subdivision analysis (OHTA, 1982a; OHTA, 1982b). Components of the two-locus linkage disequilibrium were determined using D-statistics: D_{IS}^2 (variance of within-subpopulation disequilibrium), D_{ST}^2 (variance of allele correlation on two loci of different gametes in one subpopulation relative to that in the total population), $D'_{IS}{}^2$ (variance of allele correlation on two loci of one gamete in a subpopulation relative to that in the total population), $D'_{ST}{}^2$ (variance of the disequilibrium in the total population) (OHTA, 1982a; OHTA, 1982b).

To illustrate the phylogenetic relationships between the identified haplotypes, a cladogram was generated using the SplitsTree4 v.4.6 software, allowing grouping of closely related haplotypes (HUSON and BRYANT, 2006). The programme uses the minimum spanning tree method, applying the Prim algorithm (PRIM, 1957; WU et al., 2008), based on genetic distances determined by the UPGMA method (the unweighted pair-group method with arithmetic mean) (SNEATH and SOKAL, 1973).

Results

A direct result of the haplotype analysis was the identification of 44 different haplotypes (listed in Table 1) in the study population per 24576 possible haplotypes for 14 detected variable/polymorphic loci (g.616G>C, g.703G>A, g.742G>A, g.748T>C, g.775G>A, g.778G>A, g.837G>C, g.846A>G, g.885delC, g.931C>T, g.1156A>G, g.1219C>G, g.1219_1236del18bp, g.1329T>C) in the *mGH* gene. Thus, in the study population only approx. 1.79‰ of the possible haplotypes were observed. 16 haplotypes occurred only once in the studied population. The most common combination was 5'-G-G-G-T-G-G-G-A-C-C-A-(C-Ins.)-T-3', found in 53.86% of individuals.

The results of linkage disequilibrium analysis proved that the strongest correlation (greatest linkage disequilibrium) could be observed between the polymorphism of g.885delC and g.1219_1236del18bp ($r^2 \times 100\% = 94\%$), and between g.616G>C and g.846A>G ($r^2 \times 100\% = 82\%$). Significant gametic disequilibrium also occurred between polymorphism g.846A>G and g.931C>T, and between g.616G>C and g.931C>T ($r^2 \times 100\%$ values equal 54% and 50%, respectively). Four haplotype blocks were defined: (g.616G>C) - (g.703G>A) - (g.742G>A) - (g.748T>C), (g.775G>A) - (g.778G>A) - (g.837G>C), (g.846A>G) - (g.885delC) - (g.931C>T), (g.1156A>G) - (g.1219C>G) - (g.1219_1236del18bp). Polymorphism g.1329T>C did not belong to any of the identified blocks. The results of the linkage disequilibrium analysis are shown in Fig. 1.

At the same time, it should be noted that the LD map presented does not illustrate the actual relationship between the presence of the polymorphisms g.1219_1236del18bp and g.1219C>G - the occurrence of segregating site in position g.1219 always depends on the absence of deletion g.1219_1236del18bp. This discrepancy is due to the limitations

Table 1. Identified haplotypes of the *mGH* gene

No.	Haplotype	Name	No.	Haplotype	Name
1	G-G-G-T-G-G-G-G-C-C-A-(C-I)-T	HPTP-1	23	G-G-G-C-G-G-G-G-C-C-A-(D)-C	HPTP-23
2	G-G-G-T-G-G-G-A-C-T-A-(C-I)-T	HPTP-2	24	G-G-G-C-G-G-G-A-C-C-G-(C-I)-T	HPTP-24
3	G-G-G-T-G-G-G-A-C-C-G-(G-I)-C	HPTP-3	25	G-G-G-C-G-G-G-A-C-C-A-(C-I)-T	HPTP-25
4	G-G-G-T-G-G-G-A-C-C-G-(C-I)-T	HPTP-4	26	G-G-G-C-G-G-G-A-D-C-G-(D)-T	HPTP-26
5	G-G-G-T-G-G-G-A-C-C-G-(C-I)-C	HPTP-5	27	G-G-G-C-G-G-G-A-D-C-G-(D)-C	HPTP-27
6	G-G-G-T-G-G-G-A-C-C-A-(G-I)-T	HPTP-6	28	G-G-G-C-G-G-G-A-D-C-A-(D)-C	HPTP-28
7	G-G-G-T-G-G-G-A-C-C-A-(G-I)-C	HPTP-7	29	G-G-G-C-G-G-C-G-C-T-A-(C-I)-T	HPTP-29
8	G-G-G-T-G-G-G-A-C-C-A-(C-I)-T	HPTP-8	30	G-A-G-T-G-G-G-G-C-C-A-(G-I)-T	HPTP-30
9	G-G-G-T-G-G-G-A-C-C-A-(C-I)-C	HPTP-9	31	G-A-G-T-A-G-G-G-C-C-A-(G-I)-T	HPTP-31
10	G-G-G-T-G-G-G-A-D-C-G-(D)-C	HPTP-10	32	C-G-G-T-G-G-G-G-C-T-A-(G-I)-T	HPTP-32
11	G-G-G-T-G-G-G-A-D-C-A-(D)-T	HPTP-11	33	C-G-G-T-G-G-G-G-C-T-A-(C-I)-T	HPTP-33
12	G-G-G-T-G-G-C-G-C-T-A-(C-I)-T	HPTP-12	34	C-G-G-T-G-G-G-G-C-C-G-(G-I)-T	HPTP-34
13	G-G-G-T-G-G-C-G-D-C-G-(D)-T	HPTP-13	35	C-G-G-T-G-G-G-G-C-C-A-(G-I)-T	HPTP-35
14	G-G-G-T-G-G-C-A-C-T-G-(C-I)-T	HPTP-14	36	C-G-G-T-G-G-G-G-D-C-G-(D)-T	HPTP-36
15	G-G-G-T-G-G-C-A-C-T-A-(C-I)-T	HPTP-15	37	C-G-G-T-G-G-G-A-C-C-A-(C-I)-T	HPTP-37
16	G-G-G-T-G-G-C-A-C-C-A-(C-I)-T	HPTP-16	38	C-G-G-T-G-A-G-A-C-C-A-(G-I)-T	HPTP-38
17	G-G-G-T-G-G-C-A-D-C-G-(D)-T	HPTP-17	39	C-G-G-T-G-A-G-A-C-C-A-(D)-T	HPTP-39
18	G-G-G-T-G-A-G-A-C-C-A-(G-I)-T	HPTP-18	40	C-G-A-T-G-G-G-G-C-T-A-(C-I)-T	HPTP-40
19	G-G-G-T-G-A-G-A-C-C-A-(C-I)-T	HPTP-19	41	C-A-G-T-G-G-G-G-C-T-A-(G-I)-T	HPTP-41
20	G-G-G-T-G-A-G-A-C-C-A-(C-I)-C	HPTP-20	42	C-A-G-T-G-G-G-G-C-T-A-(C-I)-T	HPTP-42
21	G-G-G-T-A-A-G-A-C-C-A-(C-I)-T	HPTP-21	43	C-A-G-T-G-G-G-G-C-C-A-(G-I)-T	HPTP-43
22	G-G-G-C-G-G-G-G-C-T-A-(C-I)-T	HPTP-22	44	C-A-G-T-G-G-G-G-D-C-G-(D)-T	HPTP-44

Record of alleles for a given locus reflects the actual order of variable loci within the mGH sequence (5'-3'); D - deletion, I - insertion or lack of deletion

of the algorithms in the software used for the analysis of gametic disequilibrium. For the entire population, the average correlation coefficient for alleles at all variable loci is equal to 0.8306 ($\alpha<0.001$), indicating a clear multilocus linkage disequilibrium in the study population.

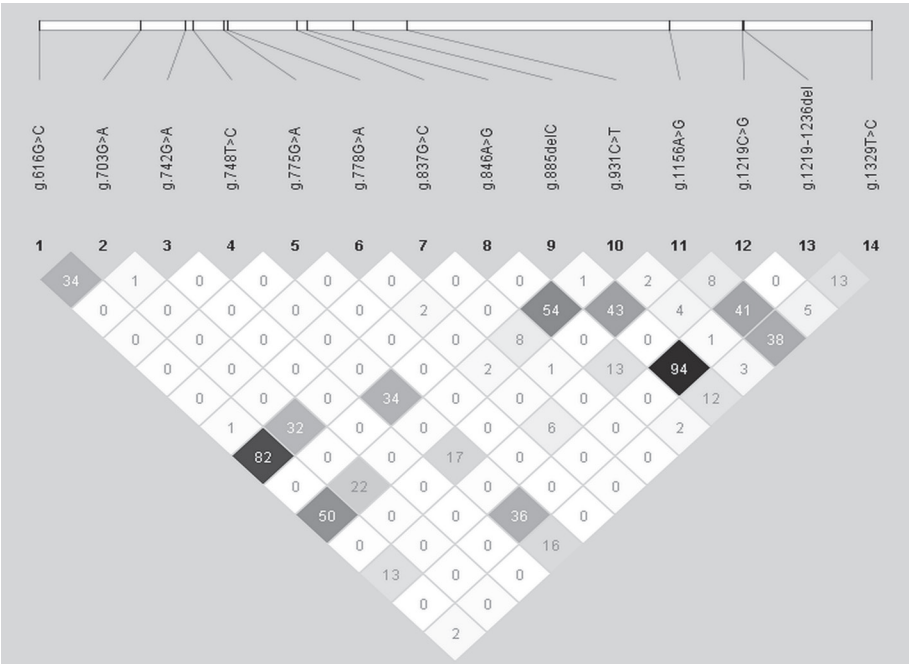

Fig. 1. Linkage disequilibrium ($r^2 \times 100\%$) plot of the *mGH* gene

The following results of Ohta's analysis were obtained: $D_{IS}^2 = 0.0039$, $D_{ST}^2 = 0.0491$, $D'_{IS}{}^2 = 0.0490$, $D'_{ST}{}^2 = 0.0022$.

The cladogram developed clearly separates the two distinct groups (clades), as shown in Fig. 2. The first is heterogeneous, and within it there are four subclades, grouping 32 haplotypes, while the second shows greater homogeneity and includes 12 haplotypes.

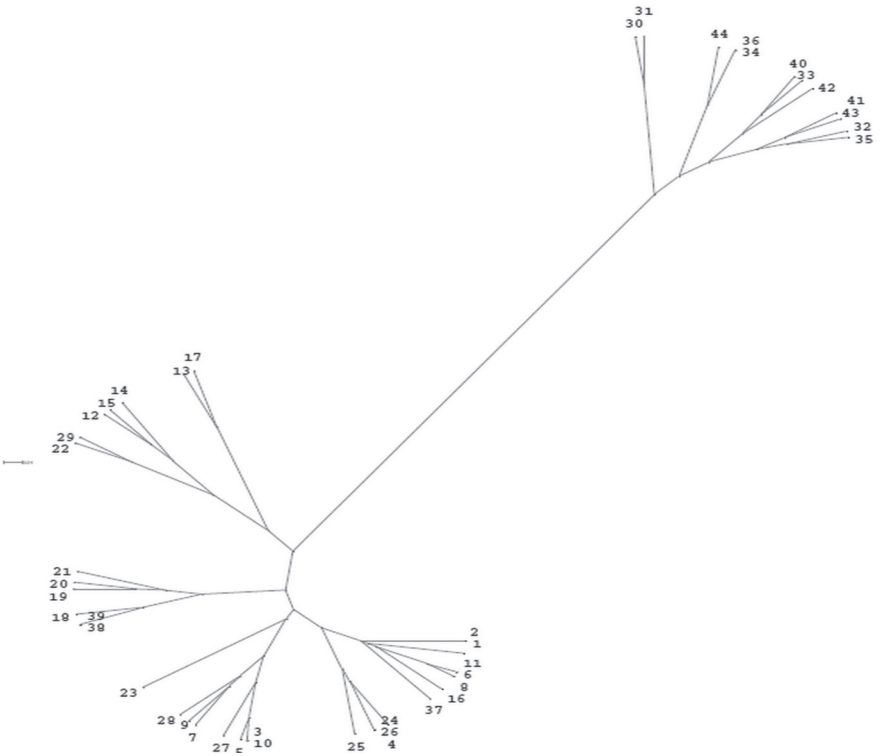

Fig. 2. Cladogram for haplotypes of the *mGH* gene, drawn up using the minimum spanning tree method

Discussion

The described LD indicates a relatively low frequency of recombination events between variable loci of the *mGH* gene, which may be explained by the close distance that divides each variable loci identified for this gene - the shortest distance is 2 bp, while the greatest is 714 bp (OSTRER, 1998).

The relationships $D_{ST}^2 > D_{IS}^2$ and $D'_{IS}{}^2 > D'_{ST}{}^2$ are interpreted as a linkage disequilibrium caused by limited migration and genetic drift, while the relationships $D_{ST}^2 < D_{IS}^2$ and $D'_{IS}{}^2 < D'_{ST}{}^2$ are linkage disequilibrium induced by epistatic natural selection (OHTA, 1982a; OHTA, 1982b). The results obtained for the study population suggest that the

observed linkage disequilibrium, as well as the internal structure of the population were caused by limited migration and genetic drift.

Analysis of a frequency of particular haplotypes does not allow to unambiguously link a particular individual to a given haplotype clade (subclade). This demonstrates the possibility of the existence of an internal structure in the studied population. On the other hand, the existence of two distinctly different and relatively internally consistent clades may indicate the existence of two main patterns of ancestral haplotypes, from which all the others derived. Moreover, the internal structure of the cladograms reveals their clear hierarchical arrangement, which allowed to draw a conclusion regarding the origin-relationships of the identified haplotypes.

Acknowledgements

This research project would not have been possible without the kind support of Prof. Hossain Farid from Nova Scotia Agricultural College, Canada, and Mr Róbert Arnar Stefánsson and Ms Menja von Schmalensee from the West Iceland Centre of Natural History, Iceland. The project was funded by the Polish National Science Centre.

References

BARRETT, J. C., B. FRY, J. MALLER, M. J. DALY (2005): Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265.

BRITO, P. H., V. E. SCOTT (2009): Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135, 439-455.

ELIADES, N. G., D. G. ELIADES (2009): Haplotype analysis: software for analysis of haplotypes data. Distributed by the authors. Forest Genetics and Forest Tree Breeding, Georg-Augst University Goettingen, Germany.

HALLDÓRSSON, B. V., V. BAFNA, N. EDWARDS, R. LIPPERT, S. YOOSEPH, S. ISTRAIL (2003): Combinatorial problems arising in SNP and haplotype analysis. Lect. Notes. Comput. Sc. 2731, 26-47.

HUSON, D. H., D. BRYANT (2006): Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254-267.

OHTA, T. (1982a): Linkage disequlibrium due to random genetic drift in finite subdivided populations. P. Natl. Acad. Sci. USA 79, 1940-1944.

OHTA, T. (1982b): Linkage Disequilibrium With the Island Model. Genetics 101, 139-155.

OSTRER, H. (1998): Non-mendelian Genetics in Humans. Oxford Monographs on Medical Genetics, No 35. Oxford University Press. Oxford.

PRIM, R. C. (1957): Shortest connection networks and some generalizations. Bell. Syst. Tech. J. 36, 1389-1401.

SHRIVER, M. D., R. MEI, E. J. PARRA, V. SONPAR, I. HALDER, S. A. TISHKOFF, T. G. SCHURR, S. I. ZHADANOV, L. P. OSIPOVA, T. D. BRUTSAERT, et al. (2005): Large-

scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum. Genomics 2, 81-89.

SKORUPSKI, J. (2014): Analysis of the nucleotide sequence of the growth hormone gene in American mink (*Neovison vison* Schreb., 1777). PhD, Jagiellonian University in Kraków, Kraków, Poland (in Polish with English abstract).

SLATKIN, M. (2008): Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9, 477-485.

SMOUSE, P. W., J. V. NEEL, W. LIU (1983): Multiple-locus departures from panmitic equilibrium within and between village gene pools of Amerindian tribes at different stages of agglomeration. Genetics 104, 133-153.

SNEATH, P. H. A., R. R. SOKAL (1973): Numerical Taxonomy. W. H. Freeman. San Francisco.

STEPHENS, M., P. SCHEET (2005): Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. Am. J. Hum. Genet. 76, 449-462.

WU, Y., P. R. BHAT, T. J. CLOSE, S. LONARDI (2008): Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 10, e1000212.

YEH, F. C., T. J. B. BOYLE (1997): Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg. J. Bot. 129, 157-163.

Received: 23 June 2015
Accepted: 22 March 2016

SKORUPSKI, J. J.: Analiza haplotipova genski polimorfnog hormona rasta američke vidrice (*Neovison vison*) pomaže objasniti razloge za njihovu vezanu neravnotežu i razlikovanje po predcima. Vet. arhiv 86, 573-580, 2016.

SAŽETAK

Kod 389 američkih vidrica (*Neovison vison*) analizirani su haplotipovi gena za hormone rasta. Od diploidnih multilokusnih genotipova, među kojima je utvrđeno 14 polimorfnih lokusa, identificirana su 44 različita haplotipa. Signifikantna neravnoteža gameta javlja se u slučaju 4 varijabilna mjesta, s koeficijentom korelacije u rasponu od 50% do 94%. Definirana su 4 haplotip bloka. Rezultati D-statistike za istraženu populaciju upućuju da su opažene vezane neravnoteže uzrokovane ograničenom migracijom i genskim pomakom. Osim toga, opisana vezana neravnoteža pokazuje relativno nisku učestalost rekombinacija između varijabilnih genskih lokusa za hormon rasta američke vidrice. Konstruirani kladogram otkriva postojanje dviju potpuno različitih i unutar sebe konzistentnih kladi, što može ukazivati na dva glavna obrasca za haplotipove predaka, od kojih su izvedeni svi drugi haplotipovi. Također, unutarnja građa kladograma otkriva njihov jasan hijerarhijski raspored, što omogućuje zaključke o podrijetlu odnosa između utvrđenih haplotipova.

Ključne riječi: američka vidrica, genetski polimorfizam, gen za hormone rasta, analiza haplotipa, vezana neravnoteža, Ohta D-statistika