Estimation of genomic variation in cervids using cross-species application of SNP arrays

Moravčíková, N., Kirchner, R., Šidlová, V., Kasarda, R., Trakovická, A.

Poljoprivreda/Agriculture

ISSN: 1848-8080 (Online)
ISSN: 1330-7142 (Print)

http://dx.doi.org/10.18047/poljo.21.1.sup.6

Poljoprivredni fakultet u Osijeku, Poljoprivredni institut Osijek

Faculty of Agriculture in Osijek, Agricultural Institute Osijek

ISSN 1330-7142
UDK: 639.111.11:636.082.12+004
DOI: 10.18047/poljo.21.1.sup.6

ESTIMATION OF GENOMIC VARIATION IN CERVIDS USING CROSS-SPECIES APPLICATION OF SNP ARRAYS

Moravčíková, N., Kirchner, R., Šidlová, V., Kasarda, R., Trakovická, A.

Original scientific paper

SUMMARY

The aim of this study was to assess the utility of commercially developed genotyping array for cross-species genotyping in order to estimate the genetic variation across two species from family Cervidae. The genotyping of individuals was carried out using Illumina BovineSNP50 BeadChip. The cross-species application of bovine array was tested overall in 3 farmed and 5 free range Red deer (Cervus elaphus) and 2 free range Fallow deer (Dama dama). After applying data quality control 97.2% of SNPs localized on the chip were removed and only 1,530 autosomal markers showed polymorphism across all analysed individuals. Across all polymorphic SNPs the minor allele frequency reached the average value 0.23±0.16. The analysis based on Bayesian clustering approach clearly showed a partition of deer in two separate clusters in relation to their phylogenetical relationship. Moreover, the PCA analysis indicated that the genetic differences between farmed and free range Red deer caused the division of analysed individuals into the two subpopulations. But the results of cross-species genotyping should be present with caution, because the bovine chip developed primarily for taurine cattle breeds is not fully representative to the evolutionary changes in genome of cervids. Nevertheless, our results suggested that the utility of bovine array alongside microsatellite markers and mtDNA can be very perspective for genetic diversity estimation in deer populations.

Key-words: bovineSNP50 chip, fallow deer, genetic differentiation, polymorphism, red deer

INTRODUCTION

The patterns of genetic variation within wild animal populations may be influenced mainly by natural and anthropogenic factors. Human activity such as agriculture, deforestation, hunting, introduction of alien species or translocation of populations may greatly affect the level of genetic diversity and structure of natural population as well as and reduce their fitness and future adaptive potential (Rosvold et al., 2012). The microevolutionary consequences of human practices might have profound effects not only on threatened species living in small and isolated populations, but also on common and widespread species subject to strong management practices (Olano-Marin et al., 2014). In both present and ancient time deer (*Cervidae*) belong to the most important species representing usable models to assess the consequences of introduction events and breeding practises on genetic diversity. In addition, *Cervidae* is one of the few mammal family for which farmed and wild/feral populations may be found in sympatry (de Garine-Wichatitsky et al., 2009). However, due to heavy hunting and habitat alterations, many populations were severely reduced in numbers in previous centuries (Rosvold et al., 2012). Understanding of population genetic structure is important for management of species as genetically isolated populations with limited diversity are often associated with inbreeding and reduced reproductive fitness. The population bottleneck is one of the factors that can result in the genetic diversity loss. It may cause a reduction in number of alleles and heterozygosity, the fixation of deleterious alleles, and potentially the occurrence of inbreeding depression in populations (Webley et al., 2007; Ernst et al., 2012). The reduction of genetic diversity and the impact of bottleneck effect have been

Ph.D. Nina Moravčíková (nina.moravcikova1@gmail.com), Róbert Kirchner, M. Eng. Veronika Šidlová, M. Eng., Assist. Prof. Radovan Kasarda, Prof. Dr. Anna Trakovická - Slovak University of Agriculture, Faculty of Agrobiology and Food Resources, Tr. A. Hlinku 2, 94976 Nitra, the Slovak Republic

demonstrated with historical records often in accordance with molecular data in several deer populations (Broders et al., 1999; Webley et al., 2007).

Mainly two methods utilizing microsatellite markers and mitochondrial DNA (mtDNA) control region are preferred for the analysis of genetic diversity in deer species. The aim of this study was to assess the utility of cross-species genotyping in the genetic variation estimation within two species from the family *Cervidae*, red deer and fallow deer.

MATERIAL AND METHODS

Two species from family *Cervidae* were included in this study. The cross-species application of genotyping array was tested overall in three farmed and five free range Red deer (*Cervus elaphus*) and two free range Fallow deer (*Dama dama*). All the analysed individuals were originated from Slovakia. The farmed deer samples represented male progeny of sires from New Zealand and dams from Hungary, whereas the free range deer were trophy animals. The single nucleotide polymorphisms (SNPs) genotyping were carried out using Illumina BovineSNP50 BeadChip in commercial lab (Illumina, Inc. San Diego, USA). Firstly any markers with unknown chromosomal position and SNPs located on sex chromosome were removed. Secondly the quality control of genotyping data was applied to eliminate any SNPs with genotyping errors (loci with >10% missing genotypes), less informative markers (MAF<0.01) and markers deviating from Hardy-Weinberg equilibrium limit of 0.001. This resulted in total of 1,530 informative autosomal SNPs which passed the above criteria and were usable for genetic variability analysis within analysed cervids.

The genetic differentiation among the analysed deer was estimated using two approaches, the Bayesian clustering method and principal component analysis (PCA). The basic sample and marker statistic of genotyping data was calculated using SNP & Variation Suite (v7.6.8 Win 64, Golden Helix, Bozeman, MT, USA, www.goldenhelix.com). The Bayesian clustering implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000) was used for estimation of groups number represented by all analysed individuals in relation to the individual admixture proportion. The STRUCTURE analysis was carried out using the default parameters of an admixture model and correlated allele frequencies across all the individuals based on burn-in period of 10,000 followed by 100,000 Markov chain Monte Carlo (MCMC) replications. Ten runs were performed from K=1 to K=10 a4nd the K with the highest likelihood was selected using the STRUCTURE HARVESTER (Earl and von Holdt, 2012) that evaluate the log probability of data (ΔK) according to Evanno et al. (2005). Subsequently, the genetic differences among the analysed deer in relation to the population structure pertaining to the individuals and species were estimated based on principal component analysis (PCA) performed using the SNP & Variation Suite (v7.6.8 Win 64, Golden Helix, Bozeman, MT, USA, www.goldenhelix.com).

RESULTS AND DISCUSSION

After applying the quality control of genotyping data from the total 54,609 SNPs localized on the chip 552 loci with unknown position and 1,171 SNPs related to the sex chromosomes were removed. Subsequently, in the control processes of remaining autosomal SNPs it was found that the 53.89% of markers were genotyped successfully at least 90% of individuals. However most of them were monomorphic and only 5.37% of markers showed polymorphism (2.8% from the total SNPs amount). The proportion of polymorphic SNPs varied across autosomes. The highest part of polymorphic loci was found on autosome 2 and the lowest on autosome 23 (Figure 1). The observed proportion of polymorphic markers was comparable with results published by Haynes and Latch (2012) that similarly tested the application of bovine genotyping array on species from family *Cervidae*.

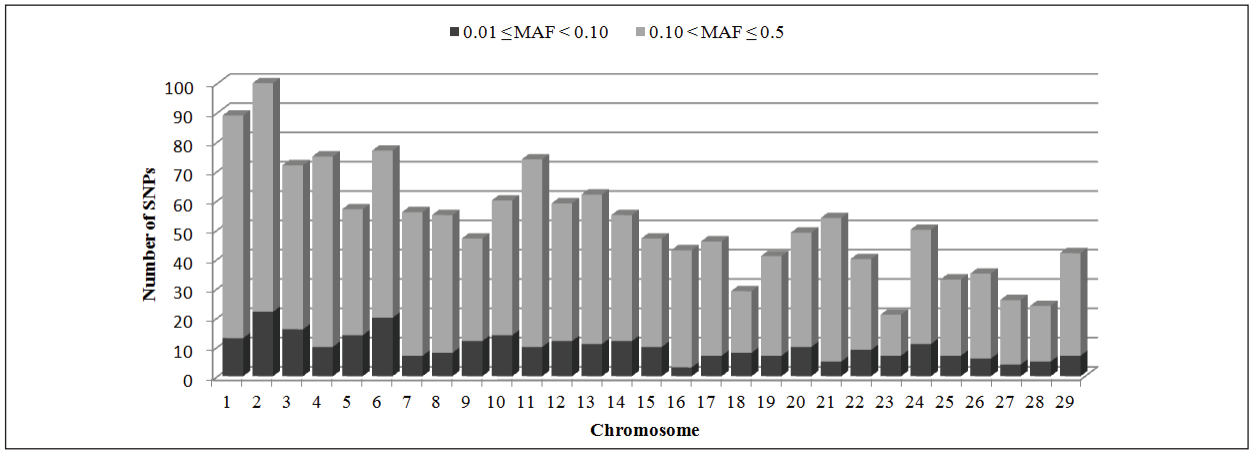

Figure 1. Distribution of polymorphic SNPs across autosomes within two MAF value levels (intermediate and common variants)

The power to detect genetic effects is dependent on minor allele frequency (MAF). In the previous studies it has been demonstrated that rare genotypes are more likely to result in spurious findings, and therefore SNPs with MAF<10% are mostly removed (Tabangin et al., 2009). This study tested only SNPs with MAF limit of 0.01. Across all the polymorphic SNPs the minor allele frequency reached the average value 0.23±0.16. The highest proportion of SNPs with common MAFs variants (0.10<MAF≤0.5) was found on chromosome 16. On the contrary the highest proportion of MAFs<0.10 was observed on autosome 18. Across the all analysed individuals and informative loci the heterozygosity was found at the level 0.28±0.007.

The Bayesian clustering approach revealed a partition of deer in to two separate clusters. However, based on the method assessing log probability of ΔK (Evanno et al., 2005) the local maxima at K=3 (Figure 2) was detected. The observed difference can be attributed to the genetic diversity in Red deer group also demonstrating the principal component analysis. However the distribution observed using STRUCTURE analysis clearly showed division of analysed individuals in relation to the species origin.

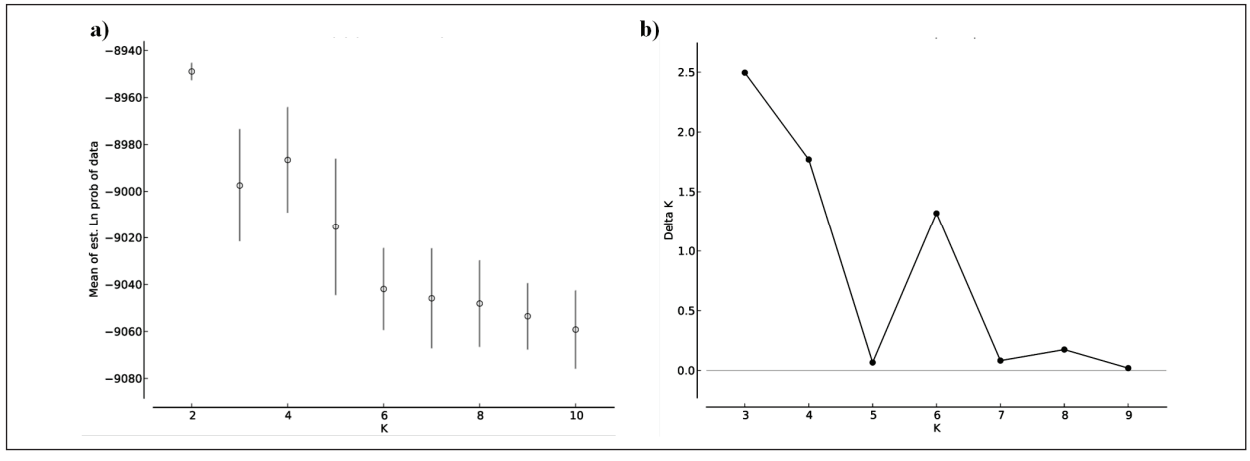

Figure 2. The results of STRUCTURE analysis showing the most likely number of subpopulations (a) mean ± SD of the log likelihood in relation to the different values of K and (b) values of ΔK evaluated according to Evanno et al. (2005)

The principal component analysis performed from the genetic covariance matrix using genotyping data and the correlation coefficient between sample and genotypes for each locus across analysed individuals showed similarly the clear differentiation of animals into the separate clusters related to each species (Figure 3.). Moreover, the PCA analysis indicated that the genetic differences between farmed and free range Red deer caused the division of the analysed individuals into the two subpopulations. The bovineSNP50 array was successfully used in cross-species genotyping in order to estimate genetic diversity in several phylogenetically related species (Michelizzi et al., 2011; Miller et al., 2012). However, the results of these studies should be interpreted with caution because the problem associated with the use of commercially developed array for discovery of novel SNPs in genetically related species is the occurrence of ascertainment bias, the systematic deviation from the expected allele frequency distribution (Haynes and Latch, 2012). This may occur if the SNPs are identified in a small sample of population from part of the species range and also by application to the evolutionary more distant species, because the loci cannot be representative to the evolutionary changes in both species.

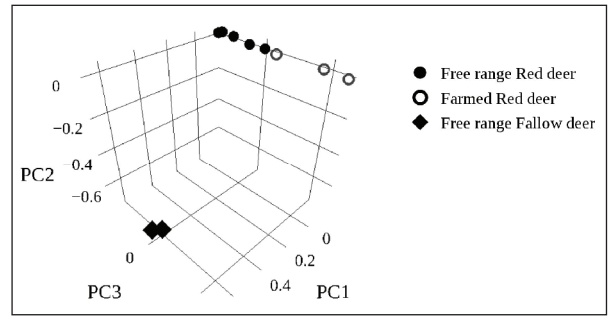

Figure 3. The PCA analysis of the genetic relationship among analysed cervids

CONCLUSION

One of the problems that mainly biased all analyses in relation to the cross-species genotyping is the fact that the applied array is not fully representative to the genome of genotyped species. Moreover, the cervid karyotype is very different as in bovids. However the results of our study indicated that the utility of commercially developed array for model animals can be very perspective to analysis of genetic differentiation within phylogenetically related non-model species. In the situation when the whole genome scan of cervids is not available use of bovine array is one of the ways that can provide usable tools alongside microsatellite

markers and mtDNA for genetic diversity estimation and phylogeny studies in deer populations.

ACKNOWLEDGEMENT

This project was co-financed by the European Union under project no. 26240220080, which supports research activities in Slovakia and Slovak Research and Development Agency under the contract no. APVV-0636-11 and APVV-14-0054.

REFERENCES

1. Broders, H.G., Mahoney, S.P., Montevecchi, W.A., Davidson, W.S. (1999): Population genetic structure and the effect of founder events on the genetic variability of moose, *Alces alces,* in Canada. Molecular Ecology, 8: 1309-1315. doi: http://dx.doi.org/10.1046/j.1365-294x.1999.00695.x
2. De Garine-Wichatitsky, M., de Meeûs, T., Chevillon, C., Berthier, D., Barré, N., Thévenon, S., Maillard, J.C. (2009): Population genetic structure of wild and farmed rusa deer (Cervus timorensis russa) in New-Caledonia inferred from polymorphic microsatellite loci. Genetica, 137(3): 313-323. doi: http://dx.doi.org/10.1007/s10709-009-9395-6
3. Dent, E.A., von Holdt, B.M. (2012): STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4: 359-361. doi: http://dx.doi.org/10.1007/s12686-011-9548-7
4. Emst, M., Putnova, L., Pokoradi, J., Matouskova, J., Zidek, R., Marsalkova, L., Lamka, J. (2012): Microsatellite Analysis DNA for Controlled Breeding of Cervidae and Genetic Breeding of White Colored Cervus elaphus in Czech Republic. Iranian Journal of Applied Animal Science, 2(3): 233-237.
5. Evanno, G., Regnaut, S., Goudet, J. (2005): Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14: 2611-2620. doi: http://dx.doi.org/10.1111/j.1365-294x.2005.02553.x
6. Haynes, G.D., Latch, E.K. (2012): Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus spp.*) using the BovineSNP50 BeadChip. PLoSOne, 7: e36536. doi: http://dx.doi.org/10.1371/journal.pone.0036536
7. Michelizzi, V.N., Wu, X., Dodson, M.V., Michal, J.J., Zambrano-Varon, J., McLean, D.J., Jiang, Z. (2011): A Global View of 54,001 Single Nucleotide Polymorphisms (SNPs) on the Illumina BovineSNP50 BeadChip and Their Transferability to Water Buffalo. International Journal of Biological Sciences, 7: 18-27. doi: http://dx.doi.org/10.7150/ijbs.7.18
8. Miller, J.M., Kijas, J.W., Heaton, M.P., McEwan, J.C., Coltman, D.W. (2012): Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. Molecular Ecology Resources, 12: 1145-1150. doi: http://dx.doi.org/10.1111/1755-0998.12017
9. Olano-Marin, J., Plis, K., Sönnichsen, L., Borowik, T., Niedziałkowska, M., Jędrzejewska, B. (2014): Weak population structure in European roe deer (Capreolus capreolus) and evidence of introgressive hybridization with Siberian roe deer (C. pygargus) in northeastern Poland. PLoS One, 9 (10): e109147. doi: http://dx.doi.org/10.1371/journal.pone.0109147
10. Pritchard, J.K., Stephens, M., Donnelly, P. (2000): Inference of population structure using multilocus genotype data. Genetics, 155(2): 945-959.
11. Rosvold, J., Røed, K.H., Hufthammer, A.K., Andersen, R., Stenøien, H.K. (2012): Reconstructing the history of a fragmented and heavily exploited red deer population using ancient and contemporary DNA. BMC Evolutionary Biology, 12: 191. doi: http://dx.doi.org/10.1186/1471-2148-12-191
12. Tabangin, M.E., Woo, G.J., Martin, L.J. (2009): The effect of minor allele frequency on the likelihood of obtaining false positives. BMC Proceedings, 3(Suppl 7): S41. doi: http://dx.doi.org/10.1186/1753-6561-3-s7-s41
13. Webley, L.S., Zenger, K.R., Hall, G.P., Cooper, D.W. (2007). Genetic structure of introduced European fallow deer (Dama dama dama) in Tasmania, Australia. European Journal of Wildlife Research, 53: 40-46. doi: http://dx.doi.org/10.1007/s10344-006-0069-8

(Received on 30 April 2015; accepted on 6 July 2015)