ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 67 128 Number 6, 2019

DETECTION OF AUTOSOMAL HEMIZYGOUS REGIONS IN THE FLECKVIEH POPULATION BASED ON SNP-CHIP DATA AND PARENT OFFSPRING PAIRS

Judith Himmelbauer[1], Gábor Mészáros[1], Johann Sölkner[1]

[1] Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences Vienna (BOKU), Gregor-Mendel-Straße 33, AT-1180 Wien, Austria

To link to this article: https://doi.org/10.11118/actaun201967061447
Received: 9. 8. 2019, Accepted: 11. 11. 2019

To cite this article: HIMMELBAUER JUDITH, MÉSZÁROS GÁBOR, SÖLKNER JOHANN. 2019. Detection of Autosomal Hemizygous Regions in the Fleckvieh Population Based on SNP-chip Data and Parent Offspring Pairs. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 67(6): 1447–1452.

Abstract

A Copy Number Variation (CNV) is a loss or a gain in the DNA sequence, ranging from 50 basepairs to a few megabasepairs. Most studies use whole genome sequencing data to detect deletions. Due to the fact that SNP-chip data is more commonly used in livestock, especially in cattle, the detection of deletions based on SNP-chip data is of interest. In the present study an approach based on SNP chip data and the analysis of Mendelian mismatches in parent-offspring-pairs was developed. Use was made of the fact that deletions appear as homozygous after SNP Chip genotyping. For some SNPs with high number of mismatches, the inheritance of the mismatches could be traced back to one or a few bulls and thereby regions of possible deletions were defined. The study has shown that an approach based on Mendelian mismatches and SNP-chip data is a promising way of detecting deletions.

Keywords: CNV-detection, deletions, Fleckvieh, Mendelian mismatches

INTRODUCTION

Copy number variants (CNVs) are defined as a deletion (loss) or as a duplication (gain) in the DNA sequence in the range of 50 basepairs (bps) up to a few megabasepairs (mbps) (Alkan *et al*., 2011; Conrad *et al*., 2010). Hemizygous deletions (hDEL) show up as homozygous in SNP-chip genotyping, as the machinery does not detect deletions (Amos *et al*., 2003).

The effects of CNVs have not been very well explored so far. For both phenomena, duplications and deletions, beneficial and detrimental effects have been found, whereby the number of examples of negative effects clearly outnumber positive effects. Despite the negative effects of CNVs on the fitness of carriers, these effects are not found in livestock yet, because only very strong and promising animals are genotyped, as only these animals are used for breeding. Some positive or neutral effects of CNVs in livestock have been shown, for example color-sidedness in cattle (Durkin *et al*., 2012), belt, patch and dominant white phenotypes in pig (Rubin *et al*., 2012), white coat-colour in sheep (Fontanesi *et al*., 2011; Norris and Whan, 2008) and goat (Fontanesi *et al*., 2009), late feathering (Elferink *et al*., 2008; Wang *et al*., 2010), pea-comb (Wright *et al*., 2009) and excessive black pigmentation (Shinomiya *et al*., 2012) in chicken and premature hair graying in horse (Pielberg *et al*., 2008). A 600kb deletion with positive effects for milk production and negative effects on fertility in Nordic Red cattle was reported by Kadri *et al*. (2014).

It is important to note that hemizygous deletions (hDEL) show up as homozygous in SNP-chip genotyping, as the machinery does not detect deletions (Amos *et al*., 2003). In this study, we made use of this property to search for deletions by looking for Mendelian mismatches in parent-offspring pairs.

The main objective of this study is to develop an approach to detect inherited regions with possible hemizygosity in the Austrian Fleckvieh population based on Mendelian conflicts in parent-offspring pairs.

MATERIALS AND METHODS

SNP-chip-genotyping Data

The original dataset, which was used for the analysis, consists of 1,799 Fleckvieh bulls, born between 1993 and 2002. The genotyped bulls are mainly Austrian and German animals. The used dataset was used for the project "Entwicklung einer genomischen Zuchtwertschätzung für Fleckvieh" (Development of a genetic breeding value estimation for Fleckvieh), which was conducted between 2008 and 2011 at the University of Natural Resources and Life Sciences in Vienna under the leadership of Prof. Dr. Johann Sölkner. Genotyping was performed by the company Illumina, Inc. San Diego, USA with a 50k SNP-chip.

General Approach

Deletions are not detected by SNP genotyping and the regions with deletions appear incorrectly as homozygous regions in the genome (Amos *et al*., 2003). Due to this wrong indication, a comparison of parent and offspring SNP-chip data based on the Mendelian laws leads to inconsistencies. For example, the sire carries a deletion and the genotype for a specific SNP in this region is B- ("-" stands for the deleted allele). After SNP-chip genotyping the SNP appears as a homozygous SNP, namely BB. For this example, suppose that the son is genotyped as AA for the same specific SNP, having received the allele A from the dam. Based on the Mendelian laws, the sire should be AA or AB for this SNP, but of course not BB. Consequently, this single SNP occurs as a Mendelian mismatch in this case. If several consecutive SNPs show Mendelian mismatches, this indicates a large deletion. Carriers of deletions are expected to transmit this deletion (-) to half of their offspring. Therefore, family patterns of local Mendelian mismatches may be explored

Detection of Mendelian Mismatches and Search for Potential Regions of Deletions Based on Pedigree

The detection of Mendelian mismatches was done with PLINK v 1.9 (Purcell, 2015), with the option "--mendel". This function detects all Mendel errors for all parent-offspring pairs and all SNPs. Due to the fact that the dataset available for this study only contains genotypes of bulls, father-son-pairs were used. Consequently, PLINK detects an error if the father is genotyped as homozygous for one allele and the son as homozygous for the opposing allele.

The file created with PLINK was further processed with R (R Development Core Team, 2018). For the following steps the pedigree was generated, for every individual for 5 generations. More than 5 generations were not possible to compute with the underlying data. Then a different R script checked all positions in the pedigree with sire-son-relationships, whether a Mendelian mismatch occurred at the respective SNPs. The result was the number of Mendelian errors found in the pedigree of each individual for each SNP (= error rate based on the pedigree). Additionally, the number of mismatches at a certain position, which occurred between the individual and some of his sons were counted and used for further analysis. Also, the number of Mendelian mismatches which occurred at a certain position was calculated (= general error rate).

Based on the calculated numbers, SNPs with a high error rate in general and/or in the pedigree of an individual were selected. For these SNPs all sire-son-pairs, in which a Mendelian error was detected at a certain position, were drawn in one figure respectively. If the mismatches do not occur just by chance due to random genotyping errors, the inheritance of the error can be traced back to one or few animals and it is likely that a potential region of deletion was detected.

RESULTS

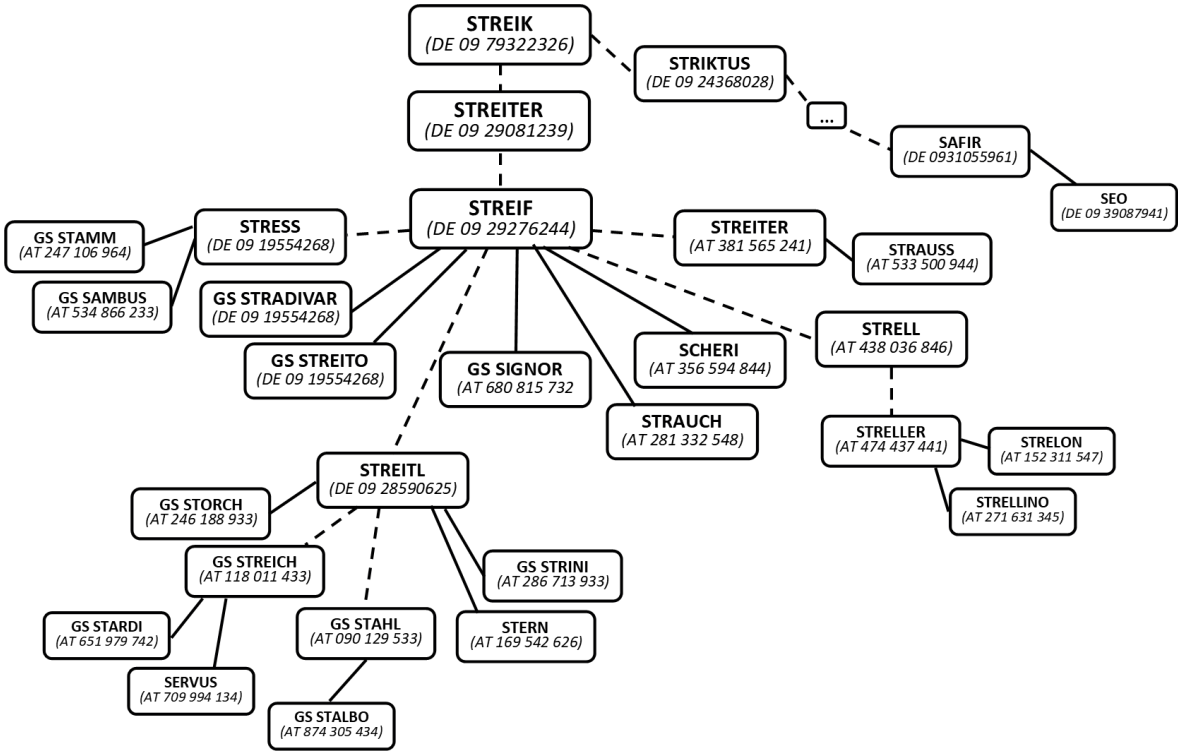
Results of the SNPdata Analysis Based on Mendelian Mismatches

A total of 2,752 errors between sires and sons were found on 336 different SNPs located on all 29 chromosomes. The highest number of errors occurred at the SNP "ARS-BFGL-NGS-113032" located at chromosome 6, position 90,931,190.

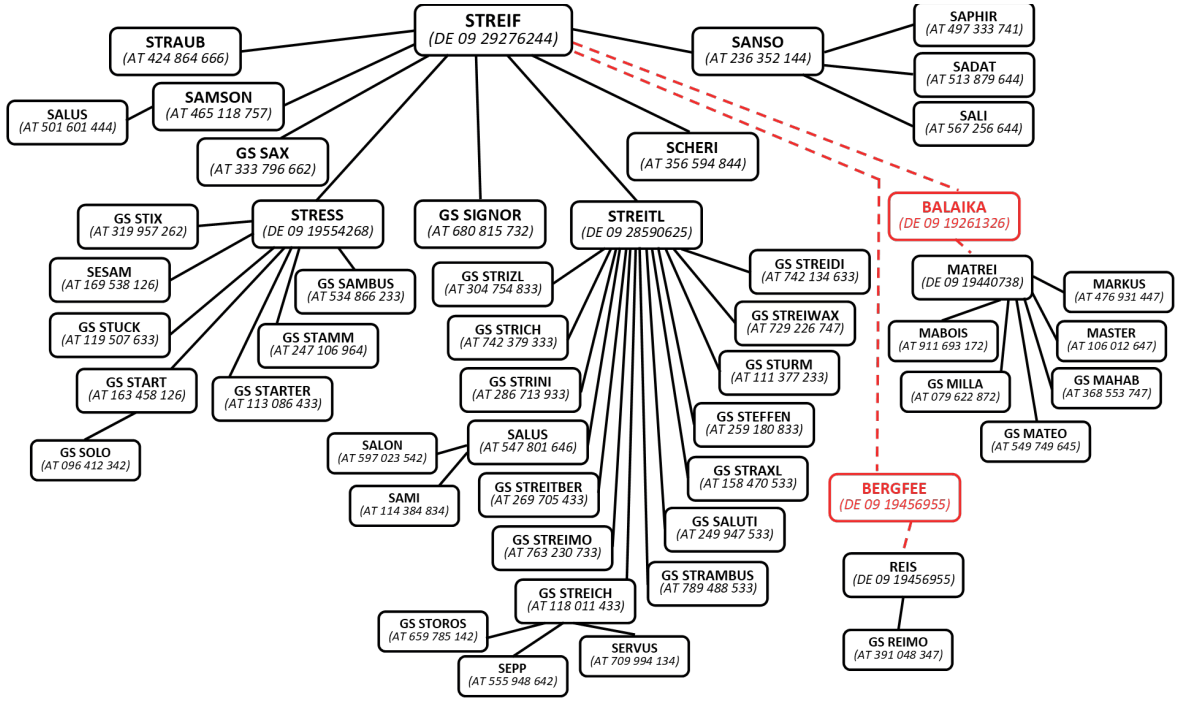
For 75 sire-son combinations a mismatch was detected at this position. Analysis on the pedigree showed that 70 the mismatches could be traced back to 3 important bulls. The majority, 38 sire-son combinations, are somehow connected with STREIF (DE 09 29276244) or maybe even with STREIK (DE 09 79322326), the paternal grandsire of STREIF. In Fig. 1 the mismatches which can be traced back to STREIF are illustrated, only mismatches between sires and their sons are included. Mismatches which were also inherited by STREIF but passed on by the dam are not in the figure. Another 29 mismatches were traced back to HAXL (DE 09 79317838) and 13 mismatches were attributed to POSTNER (DE 09 17355651). Some sire-son combinations were connected to more than one of these three sires and the common ancestor transmitting this deletion is not yet found.

Similar figures were drawn for all following regions described here.

At position 37,262,214 on chromosome 7 in total 23 errors occurred. Out of these 20 can be traced back to an Austrian sire named MARIO (AT 123 994 348). The most important sires in this family



1: *Mendelian Mismatches on chromosome 6, position 90,931,190 between sires and their sons. For each individual the name and the ID number are given. Continuous lines stand for detected mismatches, dashed lines stand for a paternity without a detected mismatch. "... " stands for (maternal and paternal) relationships, which are not shown in detail in the schematic pedigree.*



2: *Mendelian Mismatches on chromosome 28, position 27,762,101 to 27,902,139 between sires and their sons. For each individual the name and the ID number are given. Continuous lines stand for detected mismatches, dashed lines stand for a paternity without a detected mismatch. The individuals in red are female animals.*

are MORELLO (AT 842 871 443), a son of MARIO, and two grand-grand-sons of MARIO: GS MALHAX (AT 153 674 133) and MANDL (AT 410 617 633). The remaining three errors occurred between sons and grandsons of HODACH (DE 09 11331078) but are not related to the bull family described above.

On chromosome 15 several errors occurred at three consecutive SNPs. The three SNPs encompass a region with a length of about 64 kb, from position 80,033,854 to 80,097,824. Several bull families were involved, and HAXL (DE 09 79317838) is the most likely common ancestor having transmitted this likely deletion. the In total there were 41 mismatches detected at the three SNPs.

For the SNP BTA-52615-no-rs on chromosome 21, position 57,479,244, quite high numbers of Mendelian errors were found. In the whole dataset, 40 errors on this position were observed. A closer look reveals that nearly all the detected Mendelian mismatches could be traced back to one sire: STREIF (DE 09 29276244). In 35 out of the 40 sire-son-combinations, where Mendelian mismatches were found on this position, STREIF is somehow involved in the pedigree. Therefore, it is plausible to assume that there is a possible deletion in the region around this SNP.

On chromosome 28, there is a 140 kb long region between position 27,762,101 and 27,902,139 which contains five consecutive SNPs. At all these SNPs, several Mendelian mismatches occurred. The animals involved in the mismatches are mostly the same for all SNPs or at least belong to the same bull-family. In total 88 Mendelian errors were found in the region, many of them in more than one SNP. In Fig. 2 all detected mismatches are shown. In this case STREIF (DE 09 29276244) is the common ancestor. Beside some important sons, there are also two daughters of STREIF, which are also involved in the inheritance of the mismatches. The two daughters are marked in red in Fig. 2. Mendelian errors between STREIF and the daughters, or between the daughters and their sons are not checked, because the used dataset does not include any genotypes of female animals.

DISCUSSION

The method developed in this study to detect potential regions of deletions based on SNP-chip data is considered reliable. For all the selected positions almost all Mendelian mismatches detected in the dataset of 1,799 sires could be traced back to one or two animals in the pedigree. The probability that such a large number of errors accumulate in the progeny of one or two sires just by chance based on genotyping errors is close to zero. Consequently, there are reasonable grounds to suspect deletions in the region at which the Mendelian mismatches occurred. Nandolo *et al.* (2018) also showed that a significant proportion of detected ROH islands in the bovine genome are misidentified due to CNVs or SNP coverage gaps, which underpins the assumption that it is quite possible that the region around the mismatches contain deletions. A similar study on the detection of deletions in the human genome used parent-offspring trios to detect regions with potential deletions in human (Conrad *et al.*, 2006). The authors defined possible regions of deletions as regions with at least two Mendelian mismatches, because a simulation study showed that this is already a sufficiently strict criterion (Conrad *et al.*, 2006). Evaluation of this method revealed that the empirical false positive rate of the approach reached 13% (Conrad *et al.*, 2006), which is quite acceptable. The estimation of the power of the approach is more complex, but in this case a simulation showed that 50% power to detect deletion of at least 25 kb was reached on most chromosomes (Conrad *et al.*, 2006). Summarizing their study, the authors showed that the approach based on Mendelian errors is quite an efficient way to detect deletions. A study to detect deletions associated with diseases in humans worked with a similar approach, also based on mendelian inconsistencies (Kohler and Cutler, 2007). They tested their approach on simulated and real data which showed a very low false-positive rate per SNP (3.12x10-5). The power to find deletions depends on the deletion size, the deletion frequency, the sample size and the SNP density and is for a ≥ 10 kb deletion with a frequency ≥ 5% and a sample size of more than 500 trios at a quite high level (≥ 60%) (Kohler and Cutler, 2007). The specification of the method to detect deletions is, that they used a maximum-likelihood approach to simultaneously estimate deletion frequencies (population frequency and transmission frequency) (Kohler and Cutler, 2007).

In three of the five selected regions the bull STREIF plays an important role in the pedigree and the mismatches can be traced back to this German bull. A quick check shows that there are some more positions with a surprising high rate of detected mismatches where STREIF or sons or grandsons appear in the pedigree of the involved bulls. Years ago, STREIF was a quite important bull, with 123 sons and 339 grandsons with evaluated breeding values. Today STREIF is definitively not one of the top bulls any more with a current total merit index of 57. Compared with all other 105 bulls of the same age class, averaging 60.09, this is a low value. Also, many of the sons and grandsons have very low breeding values. Additionally, none of the current top bulls descends from STREIF. Maybe there is a connection between the high number of possible deletions found and the strong decline of the breeding value and the fact that STREIF and his offspring are not that important anymore.

The next step will be to confirm the results by the use of next generation sequencing (NGS) data of some of the involved bulls. In next generation sequencing methodology, typically short (50–250 bp),

NGS reads are first mapped to a reference genome. Depending on the depth of coverage, the same base pair position is called many times. For positions with duplications, twice the typical number of calls is expected whereas for deletions, it is half the typical number of calls. Differences in call rates may therefore be used to confirm duplications and deletions. Consequently, the coverage rates plotted for a certain region in question for bulls which are identified as carriers of a possible deletion in this region can confirm the results of the method based on Mendelian mismatches. Furthermore, with the coverage rates the actual genome positions of start and end base pairs of deletions can be estimated.

CONCLUSION

The study has shown that the approach based on Mendelian mismatches in parent-offspring pairs is a useful way to define regions of possible deletions in the Fleckvieh population. The great advantage of this method is that it is based on SNP-chip data, which is already available for many cattle populations.

REFERENCES

ALKAN, C., COE, B. P. and EICHLER, E. E. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics,* 12(5): 363–376.

AMOS, C. I., SHETE, S., CHEN, J. and YU, R. K. 2003. Positional identification of microdeletions with genetic markers. *Human Heredity*, 56(1–3): 107–118.

CONRAD, D. F., ANDREWS, T. D., CARTER, N. P., HURLES, M. E. and PRITCHARD, J. K. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genetics*, 38(1): 75–81.

CONRAD, D. F., PINTO D., REDON, R., FEUK, L., GOKCUMEN, O., ZHANG, Y., AERTS, J., ANDREWS, D., BARNES, C., CAMPBELL, P., FITZGERALD, T., HU, M., IHM, C. H., KRISTIANSSON, K., MACARTHUR, D. G., MAC DONALD, J. R., ONYIAH, I., PANG, A. W. C., ROBSON, S., STIRRUPS, K., VALSESIA, A., WALTER, K., WEI, J., TYLER-SMITH, C., CARTER, N. P., LEE, C., SCHERER, S. W. and HURLES, M. E. 2010. Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289): 704–712.

DURKIN, K., COPPIETERS, W., DRÖGEMÜLLER, C., AHARIZ, N., CAMBISANO, N., DRUET, T., FASQUELLE, C., HAILE, A., HORIN, P., HUANG, L., KAMATANI, Y., KARIM, L., LATHROP, M., MOSER, S., OLDENBROEK, K., RIEDER, S., SARTELET, A., SÖLKNER, J., STALHAMMAR, H., ZELENIKA, D., ZHANG, Z., LEEB, T., GEORGES, M. and CHARLIER, C. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*, 482: 81–84.

ELFERINK, M. G., VALLÉE, A. A. A., JUNGERIUS, A. P., CROOIJMANS, R. and GROENEN, M. 2008. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics*, 9: 1–9.

FONTANESI, L., BERETTI, F., MARTELLI, P. L., COLOMBO, M., DALL'OLIO, S., OCCIDENTE, M., PORTOLANO, B., CASADIO, R., MATASSINO, D. and RUSSO, V. 2011. A first comparative map of copy number variations in the sheep genome. *Genomics:* 97(3): 158–165.

FONTANESI, L., BERETTI, F., RIGGIO, V., GÓMEZ GONZÁLEZ, E., DALL'OLIO, S., DAVOLI, R., RUSSO, V. and PORTOLANO, B. 2009. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. *Cytogenetic and Genome Research*, 126(4): 333–347.

KADRI, N. K., SAHANA, G., CHARLIER, C., ISO-TOURU, T., GULDBRANDTSEN, B., KARIM, L., SANDER NIELSEN, U., PANITZ, F., PEDERSEN AAMAND, G., SCHULMAN, N., GEORGES, M., VILKKI, J., LUND, M. S. and DRUET, T. 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLOS Genetics*, 10(1): e1004049.

KOHLER, J. R. and CUTLER, D. J. 2007. Simultaneous Discovery and Testing of Deletions for Disease Association in SNP Genotyping Studies. *The American Journal of Human Genetics*, 81(4): 684–699.

NANDOLO, W., UTSUNOMIYA, Y. T., MÉSZÁROS, G., WURZINGER, M., KHAYADZADEH, N., TORRECILHA, R. B. P., MULINDWA, H. A., GONDWA, T. N., WALDMANN, P., FERENCAKOVIC, M., GARCIA, J., ROSEN, B. D., BICKHART, D., VAN TASSELL, C. P., CURIK, I. and SÖLKNER, J. 2018. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. *Genetics Selection Evolution*, 50(1): 43.

NORRIS, B. J. and WHAN, V. A. 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research,* 18(8): 1282–1293.

PIELBERG, G. R., GOLOVKO, A., SUNDSTRÖM, E., CURIK, I., LENNARTSSON, J., SELTENHAMMER, M. H., DRUML, T., BINNS, M., FITZSIMMONS, C., LINDGREN, G., SANDBERG, K., BAUMUNG, R., VETTERLEIN, M., STRÖMBERG, S., GRABHERR, M., WADE, C., LINDBLAD-TOH, K., PONTÉN, F., HELDIN, C. H., SÖLKNER, J. and ANDERSSON, L. 2008. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics*, 40(8): 1004–1009.

PURCELL, S. 2015. *Plink (version 1.9)*. Available at: http://pngu.mgh.harvard.edu/prucell/plink [Accessed: 2019, October 15].

R DEVELOPMENT CORE TEAM. 2018. R: A language and environment for statistical computing. *The R Project*. [Online]. Vienna, Austria. Available at: http://www.r-project.org [Accessed: 2019, October 15].

RUBIN, C.-J., MEGENS, H.-J., BARRIO, A. M., MAQBOOL, K., SAYYAB, S., SCHWOCHOW, D., WANG, C., CARLBORG, O., JERN, P., JORGENSEN, C. B., ARCHIBALD, A. L., FREDHOLM, M., GROENEN, M. A. M. and ANDERSSON, L. 2012. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences*, 109(48): 19529–19536.

SHINOMIYA, A., KAYASHIMA, Y., KINOSHITA, K., MIZUTANI, M., NAMIKAWA, T., MATSUDA, Y. and AKIYAMA, T. 2012. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in silky chickens. *Genetics*, 190(2): 627–638.

WANG, X., NAHASHON, S., FEASTER, T. K., BOHANNON-STEWART, A. and ADEFOPE, N. 2010. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics*, 11: 351.

WRIGHT, D., BOIJE, H., MEADOWS, J. R. S., BED'HOM, B., GOURICHON, D., VIEAUD, A., TIXIER-BOICHARD, M., RUBIN, C.-J., IMSLAND, F., HALLBÖÖK, F. and ANDERSSON, L. 2009. Copy Number Variation in Intron 1 of SOX5 Causes the Pea-comb Phenotype in Chickens. *PLOS Genetics*, 5(6): e1000512.

Contact information

Judith Himmelbauer: ju.himmelbauer@gmail.com