ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

Screening and detection of chromosomal copy number alterations (CNA) in the domestic horse by SNP-array genotyping data

Journal:	*Animal Genetics*
Manuscript ID	AnGen-20-09-0278.R2
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Pirosanto, Yamila; Universidad Nacional de la Plata, Producción Animal; Consejo Nacional de Investigaciones Cientificas y Tecnicas, IGEVET Laseca, Nora; Universidad de Córdoba, Departamento de Genética Valera, Mercedes; Universidad de Sevilla, Departamento de Agronomía Molina, Antonio; Universidad de Córdoba, Departamento de Genética Moreno Millan, Miguel; University of Córdoba, Department of Genetics Bugno Poniewierska, Monika; University of Agricultire, Department of Animal Reproduction, Anatomy and Genomics Ross, Pablo; University of California-Davis, Animal Science Azor, Pedro; Asociación Nacional de Criadores de Caballos de Pura Raza Española (ANCCE). Sevilla. Spain, Libro genealógico Demyda-Peyrás, Sebastián; Universidad Nacional de la Plata, Producción Animal; Consejo Nacional de Investigaciones Cientificas y Tecnicas, IGEVET
Keywords:	SNP, Copy number alteration (CNA), equines, chromosomal abnormalities, infertility

Screening and detection of chromosomal copy number alterations (CNA) in the domestic horse using SNP-array genotyping data

Yamila Pirosanto[ab] Nora Laseca[c], Mercedes Valera [d], Antonio Molina[c], Miguel Moreno-Millán[c], Monika Bugno-Poniewierzka[e], Pablo Ross[f], Pedro Azor[g], and Sebastián Demyda-Peyrás[ab*]

[a]*Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata. Calle 60 y 118 s/n, 1900, La Plata, Argentina.*

[b] *IGEVET (UNLP-CONICET LA PLATA), Facultad de Ciencias Veterinarias UNLP, Calle 60 y 118 s/n, 1900, La Plata, Argentina.*

[c] *Laboratorio de Diagnóstico Genético Veterinario. Departamento de Genética. Universidad de Córdoba, CN IV KM 396, Edificio Gregor Mendel, Campus Rabanales, Córdoba 14071, España.*

[d] *Departamento de Agronomía. Escuela Técnica Superior de Ingeniería Agronómica. Universidad de Sevilla. Ctra. de Utrera km 1. 41013-Sevilla, España*

[e] *Department of Animal Reproduction, Anatomy and Genomics, University of Agricultire, Krakow, Poland*

[f] *Department of animal science, University of California, Davis. One Shields av, Davis, CA, USA*

[g] *Asociación Nacional de Criadores de Caballos de Pura Raza Española (ANCCE). Sevilla. Spain.*

* Corresponding author. Tel.: +54 2214211799

E-mail address: sdemyda@fcv.unlp.edu.ar (S. Demyda Peyrás).

Abstract

Chromosomal abnormalities are a common cause of infertility in horses. However, they are difficult to detect using automatized methods. Here, we propose a simple methodology based on single nucleotide polymorphism (SNP)-array data which allows us to detect the main chromosomal abnormalities in horses in a single procedure. As proof of concept, we were able to detect chromosomal abnormalities in 33 out of 268 individuals, including monosomies, chimerisms and male and female sex-reversions, by analyzing the raw signal intensity produced by an SNP array-based genotyping platform. We also demonstrated that the procedure is not affected by the SNP density of the array employed or by the inbreeding level of the individuals. Finally, the methodology proposed in this study could be performed in an open bioinformatic environment, thus permitting its integration as a flexible screening tool in diagnostic laboratories and genomic breeding programs.

Keywords: SNP, copy number alteration, equines, chromosomal abnormalities, infertility

Introduction

The use of single nucleotide polymorphism (SNP) array-based genomic information is becoming a key part of breeding programs in domestic animals (Pryce & Daetwyler 2012; Wiggans *et al.* 2017). Nowadays, the number of livestock animals genotyped is increasing exponentially, even in equines, where the development of genomic programs is becoming increasingly common worldwide (McCoy *et al.* 2019; McGivney *et al.* 2020). Therefore, it is expected that the number of horses being genotyped will also increase considerably in the coming years.

The detection of copy number alterations (CNA) using SNP data is a technique commonly used in human genetic counseling (Harton *et al.* 2013). However, it has rarely been employed in domestic animals and even less as a screening methodology (Raudsepp & Chowdhary 2016). To date, only one large-scale systematic screening was performed, in which more than 100,000 heifers were analyzed (Berry *et al.* 2017). However, in horses, only two small studies including 2 foals (Holl *et al.* 2013) and 55 fetuses (Shilton *et al.* 2020) have been reported.

The incidence of sex-related chromosomal aberrations in the domestic horse is high (around 2% in some populations according to Bugno *et al.* (2007a)) in comparison with other domestic species (Villagómez & Pinton 2008). Among these, true 63,X (Gamo *et al.* 2019) or mosaic 64,XX/63,X (Kjöllerström *et al.* 2011) ECAX monosomy, and sex reversal mares (64,XYDSD) are the most common (Villagómez *et al.* 2011). Both syndromes are also difficult to detect at an early age, since phenotypic abnormalities appear after puberty (Anaya *et al.* 2014). In contrast, sex reversal males (64,XXDSD) and true chimerism (64,XX/64,XY) were also detected in horses, but to a lesser extent (Power & Leadon 1990). All of them are usually associated with ambiguous genitalia, a fused vulva, and an enlarged clitoris (Bannasch *et al.* 2007; Lear & McGee 2012; Albarella *et*

al. 2018), which simplifies its detection. However, chimerism was only detected, albeit with a lower prevalence, in the blood tissue (Anaya *et al.* 2018), showing a normal phenotype and fertility even during adulthood (Demyda-Peyras *et al.* 2013)

CNA in horses has been diagnosed using different approaches, including karyotyping (Neuhauser et al. 2019), *in situ* hybridization (Bugno et al. 2007b), short tandem repeat (STR) allele counting (Anaya *et al.* 2017), and more recently, droplet digital PCR (ddPCR) (Szczerbal *et al.* 2020). These techniques are expensive and/or highly specific to a particular syndrome, making their integration into routine screening problematic. This is even more difficult in individuals carrying chromosomal mosaicisms, where detection is not possible using these DNA-based methods.

In mammals, there is a specific region which shows a high degree of sequence homology (98-100%) between sex chromosomes, known as the pseudoautosomal region (PAR) (Raudsepp *et al.* 2012). This region is located in the proximal part of the *p* arm of the ECAX in the horse, and includes 1.8 Mb and 18 genes only (Raudsepp & Chowdhary 2008). Since the PAR is the only region with common markers in ECAY and ECAX, we hypothesize that it could be an interesting option to analyse their heterozygosity to screen abnormalities in the sex chromosome pair.

Nowadays, only two SNP genotyping arrays are available in horses: the Axiom™ Equine Genotyping Array (Thermofisher), which includes 670,796 SNP markers, and the Equine GGP array (Illumina) which includes 65,175 SNP markers (Schaefer & McCue 2020). Both arrays are highly reliable, but in some cases, can be problematic when trying to generate a consensus dataset, not only since they include different SNP markers but also because the results are presented in different formats. Despite that, both platforms determine the allelic variants in each locus and individual by comparing two parameters

obtained from the hybridization intensity of a locus-specific probeset (two probes, one per allele), as follows:

$$\text{LRR} = \log_2\left(\frac{(iA + iB)}{(irA + irB)}\right) \qquad \text{BAF} = \log_2\left(\frac{(iB)}{(irA + irB)}\right)$$

where ***iA*** and ***iB*** are the hybridization intensities detected in alleles A and B, respectively, in a sample, and ***irA*** and ***irB*** are the median total intensity for A and B alleles in a set of selected reference samples (normal and diploid) (ThermoFisher 2019). At each particular locus, LRR indicates the presence of CN losses or gains and BAF is a measure of the heterozygosity. Therefore, any CNA detection based on the analysis of BAF and LRR could easily be implemented regardless of the genotyping platform employed.

With this in mind, we report here a simple, robust, and semiautomated technique for the detection of chromosomal abnormalities in horses based on analyzing the raw data from the two most common SNP-genotyping platforms.

Materials and methods

Animals

We analyzed 19 individuals belonging to the Pura Raza Español (PRE) breed, which had been previously diagnosed with different chromosomal abnormalities at the Cytogenetic and Molecular Laboratory of the University of Córdoba, as well as 14 individuals showing phenotypic abnormalities in the reproductive tract (ambiguous genitalia or hypoplastic ovaries and/or uterus) or abnormal sexual behavior (mares trying to mount other mares). A short description of the individuals, as well as the origin of the sample, is provided in Supplementary Table 1. In addition, we analyzed 22 PRE stallions and 209 PRE mares with proven fertility as controls. The selection of these individuals

was based on two premises: first, that they showed normal morphology and reproductive function in the complete breeding soundness examination performed by the ANCCE prior to being accepted as breeders; and second, that all these animals had produced at least one normal offspring at the time of the study.

Karyotyping and molecular screening

Chromosomal complements were determined in 19 individuals by Giemsa-stained karyotyping (n=7, complete karyotype), dual-labeled whole chromosome fluorescence in situ hybridization (FISH, n=6, sex pair complements only), or both techniques (n=6). In all cases, at least 100 cell metaphases obtained from lymphocyte cultures were assessed following our protocols (Bugno et al. 2007a; Demyda-Peyras et al. 2013).

In addition, DNA was retrieved from blood samples of 14 additional cases and 231 controls (22 stallions and 209 mares, obtained directly from the ANCCE) using the Canvax blood DNA extraction kit (Canvax Biotech, Cordoba Spain). Finally, sex-pair chromosomal complements were determined in all the samples (n=264) using the STR-based methodology previously validated for the PRE breed (Anaya et al. 2017).

SNP array genotyping

All the individuals (n=264) were genotyped using the SNP HD Axiom™ Equine Genotyping Array (Thermofisher, Madrid, Spain), consisting of ~670 thousand SNPs (Schaefer et al. 2017). Raw files (.CEL) were processed using the Axiom Analysis Suite 5.0 software (Thermofisher, Spain) following the *"best genotyping practices"* workflow with the default parameters (DQC ≥ 0.82 and call rate ≥ 97). Only SNP markers showing

a high-quality genotyping rate (SNP CR > 95% and FLD > 3.6; (Thermofisher 2013)) were kept. The final dataset (HD) included all the SNP located on the ECA10 (n=17,965; used as autosomal control), ECAX (n= 24,854) and ECAY (n=1). Thereafter, ECAX chromosome markers were grouped into two different regions for analysis purposes as follows: the pseudoautosomal region (PAR) included SNPs located from positions 1 to 1.86 Mb of EACX (Raudsepp and Chowdhary (2008)), and the NON-PAR region ECAX:1.861-128.21 Mb, included 675 and 24,179 SNPs, respectively. Finally, *log R ratio* (LRR) and *b allele frequency* (BAF) values were obtained per individual and marker following the standard workflow of the AXIOM CNV tool software (Thermofisher 2013). EquCab 3.0 was used as the reference genome (Beeson *et al.* 2019).

To test if the array density could affect the results, we repeated all the analyses using the information from medium-density (MD) and low-density (LD) reduced datasets. The MD was created including only the SNPs available in both horse arrays (Axiom ™ 670K and GGP 65K), while the LD was created by randomly selecting 10,000 SNPs from the MD dataset. After additional filtering and processing, LRR and BAF values of 4956 markers in MD (2107 in ECA10, 26 in PAR and 2823 in NON-PAR) and 844 markers in LD (358 in ECA10, 7 in PAR and 479 in NON-PAR) per individual were kept for further analysis.

Finally, we tested if individual inbreeding value could affect the determinations, since some of them are based on the analysis of heterozygous calls. This fact could be particularly important in PRE horses where inbreeding values could be as high as 50% (Perdomo-González *et al.* 2020). To this end, we first determined the molecular inbreeding value of the ECAX (F_{ROHX}) in the control mares using the DetectRUNS package (Biscarini *et al.* 2018). Minimum ROH length was set at 1Mb, and missing and heterozygous calls were set per chromosome based on our methodology (Goszczynski *et*

al. 2018). Thereafter, the individuals were clustered into two different groups according to F_{ROHX} as follows: inbred mares ($F_{ROHX} > 0.50$; n=16, HOM-MARES) and outbred mares ($F_{ROHX} < 0.5$, n = 193, MARES).

CNA analysis

CNA detection was performed based on the analysis of LRR and BAF. To achieve this, we first determined the percentage of heterozygosity (HET), as the relation between true heterozygous calls (those with BAF between 0.25 and 0.75, according to Popova *et al.* (2009)) and total markers, in each individual and chromosomal region (ECA10, PAR and NON-PAR).

Thereafter, CNA calls were made by comparing the LRR and HET patterns (low, high or intermediate) in each region among control groups (inbred and outbred mares, and males) and each case type with CNA. A detailed explanation of the methodological approach is included in the Results section.

Statistical analysis

Differences among CNAs (each type) and controls (inbred and non-inbred) were determined by a generalized linear model (GLM) and a Bonferroni post-hoc test ($p<0.05$). The model included the group (7 levels: 2 outbred and 1 inbred controls and 4 CNA types) as a fixed factor. In addition, we tested the effect of the array density using a GLM model, including group (the same 7 levels) and array type (3 levels) as fixed factors and using a merged dataset including the LD, MD and HD data. Finally, the similarity coefficient (RV; (Robert & Escoufier 1976)) among each pair of matrix parameters (HET, LRR and BAF) was determined using a Monte Carlo resampling approach to estimate the p-values. All the analyses were performed in the *R statistical environment* (R-Core-Team 2020),

using the *data.table* (Dowle & Srinivasan 2019), *dplyr* (Hadley *et al.* 2020), and *MatrixCorrelations* (Indahl *et al.* 2018) packages. The figures were generated using *karyoploteR* (Gel & Serra 2017).

Results

Molecular screening and karyotyping

Results of karyotyping (n=13) and/or in *situ hybridization* (n=12) revealed the existence of two 63,X, nine 64,XYDSD, four 64,XXDSD and four 64,XX/64,XY individuals (Suppl. Table 1). Only chimeric animals showed a complex karyotype, with percentages of 64,XY/64,XX cells ranging from 42/58% to 73/27%. No mosaicism or chimerism was detected in the other individuals analyzed.

STR genotyping agreed with the cytogenetic results in all the individuals analyzed (n=19; Supl table 1). In addition, the 14 individuals selected by the phenotypic abnormalities were diagnosed as sex reversal mares (64,XYDSD; n=9) and males (64XX DSD, n=5). Finally, the mares (n=209) and stallions (n=22) used as the control were diagnosed as normal, with sex-chromosomal complements according to their phenotypic sex.

SNP-based CNA detection

The results of the CNA calls obtained using the HD dataset are shown in Table 1. In ECA10, no CNAs were detected, since the LRR and HET values in the control groups and those with confirmed sex-chromosomal abnormalities (X0, XXDSD, XYDSD and XX/XY) were similar ($p>0.05$) in all cases.

On the contrary, all the groups showed differential combinations of HET and LRR in the sex-chromosomes. Mares (inbred and non-inbred) showed an increased HET and an LRR near 0 ($P < 0.05$) in PAR and NON-PAR. In contrast, males showed an increased HET and LRR in PAR (since the markers located in that region have a complementary allele on ECAY) and a hemizygous pattern (HET close to 0 and LRR close to -0.5) in the NON-PAR. Mares with ECAX monosomy (X0) showed a hemizygous pattern in both ECAX regions, due to the lack of a complementary region due to the presence of a single sex-chromosome. Finally, sex reversal individuals (64,XYDSD mares and 64,XXDSD horses) showed results compatible with their chromosomal sex (males and males respectively) regardless of the phenotype observed. In all cases (n=27), CNA analysis showed the same results as those detected by karyotyping or STR genotyping. These results could be also easily detected by analyzing the graphical patterns of HET and LLR scatter plots of each group (Figure 1).

The results for chimeric individuals (64,XX/64,XY) are also shown in Table 1. In ECA10, and PAR there were no differences in the mean HET, LRR and BAF, since the chimerism was balanced. In this case, the genotype of each marker is formed by a combination of two possible genotypes from the paternal (AA) and maternal lines (BB). However, homozygous combinations (AAAA; BAF>0.75 and BBBB, BAF<0.25)) were discarded during the estimation of HET. Chimerism can therefore be detected by the analysis of BAF dispersion, which shows a curve with three peaks associated with the AAB, AABB, and ABB genotypes (Fig. 2a). In contrast, NON-PAR, BAF showed a curve with only two peaks (Fig 2b) associated with the AAB or ABB genotypes, since the maternal line has two possible alleles (A or B) while the paternal line is hemizygous, and therefore, only has a single possible allele (A). For that reason, an AABB peak (located near 0.5 in PAR and ECA10) cannot be present in mosaic individuals, which allows us to

discriminate between individuals carrying balanced (chimeric) and unbalanced (mosaic) CNAs. In addition, male/female chimeric individuals were characterized by an intermediate LRR value (near -0.22) only in NON-PAR ($P < 0.05$; Table 1) in comparison with individuals carrying one (LRR near -0.5) or two (LRR near 0) copies of NON-PAR. This pattern occurs since LRR is an indirect estimation of the DNA content of the sample, and therefore, intermediate values also indicate the presence of mosaicism/chimerism in which one of the cell lines is aneuploid.

In addition, our results showed that while HET was lower in HOM-MARES than in MARES ($P < 0.05$), it was higher in comparison with individuals carrying a single ECAX copy (MALES, 64,XYDSD and 63,X), demonstrating that this method is also valid for analyzing highly inbred individuals.

Finally, all the results obtained using HD data were replicated using the reduced the MD and LD datasets (Supp. Tables 2 and 3 respectively). No statistical analysis interaction between the group and array type was found (p-values of 0.35, 0.99, and 0.96 for HET, LRR, and BAF respectively). In addition, the RV coefficients between each parameter's dataset pair were 0.861 (HD-LD), 0.901 (MD-LD) and 0.958 (HD-MD), showing great consistency (with a p-value of <0.0001 in all cases), regardless of the array density.

Discussion

In this study, we have proposed a simple, robust methodology based on SNP-array data to detect individuals carrying four of the most common types of CNA in horses. Despite a similar methodology being routinely employed in preimplantation aneuploidy diagnosis (PGD-A) in humans, its use in domestic animals is still uncommon.

Our technique was able to detect four of the most common sex chromosomal alterations reported in horses (Power 1990; Villagómez *et al.* 2011) in 36 individuals. In addition, the results obtained in male-female chimeras demonstrate that SNP-CNA analysis can also detect individuals carrying ECAX mosaicisms. This is an important point considering that 63,X/64,XX, the third most important sex-pair chromosomal alteration reported in this species (Bugno *et al.* 2001), cannot be detected by using the molecular methods currently available (Anaya *et al.* 2017; Szczerbal *et al.* 2020). However, all the chimeric samples analyzed in this study showed similar percentages of each cell line (64,XX or 64,XY in this case). This was shown by the wide gap observed between modal BAF peaks and by the intermediate LRR value observed in PAR of XX/XY individuals, which have previously been associated with the degree of chimerism/mosaicism in human samples (Markello *et al.* 2012). In contrast, there are reports in horses in which the grade of chimerism or mosaicism detected is low (Albarella *et al.* 2018). Since the accuracy in detecting these complex karyotypes in humans by CNA is determined by the existence of a minimum grade of chimerism (Goodrich *et al.* 2017), further studies are needed to determine the minimum detection threshold of this technique in horses.

Our results also showed a robust detection and discrimination of 63,X, and 64,XYDSD individuals based on the analysis of the PAR and NON-PAR. Both cases revealed a hemizygous pattern in NON-PAR, which was also observed in the PAR of Turner´s subjects (63,X). In comparison, 64,XYDSD depicted a heterozygous PAR, in agreement with a normal male, while the lack of compatibility between the genetic and the phenotypic sex allows its detection. It is worth mentioning that these genomic patterns can be detected from birth, and even by PGD analysis of embryo biopsies, thus allowing

the breeders to implement the early and appropriate management of any individuals affected.

One point to note in our study is that CNA detection was possible despite the density of markers of the array employed (low, medium or high). This was demonstrated by the RV coefficient obtained between each pair of arrays (close to 0.9 on average, $p<0.0001$). However, the fact that the LD results were reliable suggests that CNA detection can be integrated into genomic breeding programs based on low-density customized arrays (Bolormaa *et al.* 2015; Lopes *et al.* 2018), especially when the minor differences observed using the LD dataset in the PAR can be easily overcome by increasing the number of markers included. In addition, the entire operation was performed on R, an open, flexible, and collaborative bioinformatics platform, and by analyzing raw data (LRR and BAF), which can be obtained from any hybridization-based genotyping platform. Both facts make this methodology flexible, customizable, easy to replicate and independent of the genotyping array employed.

Another advantage of SNP-based CNA detection is the flexibility to screen the whole genome with a single analysis (Tang & Amon 2013). For instance, we easily detected the chimerism existing in the ECA10 in XY/XX individuals. It is also theoretically possible to detect CNAs in any other chromosome by using the same bioinformatics approach proposed in the study but varying the data analyzed. This was recently suggested by Shilton *et al.* (2020), who detected several autosomal CNAs in aborted fetuses using a similar NGS-SNP combined approach. However, it is important to consider the selection of the tissue used to obtain the DNA before the analysis, since non-blood samples will not allow us to determine the presence of blood chimerism. In fact, the molecular methodologies currently available for detecting CNAs in horses (e.g., FISH, PCR, ddPCR, or STR analysis) are extremely inflexible, since all of them are based on

chromosome-specific probes or markers, and therefore must be adapted to each case on an empirical basis. Consequently, SNP-based CNA screening detection could be an interesting tool to increase our existing knowledge regarding autosomal aberrations in horses.

One unexpected finding was the existence of a small percentage of heterozygosis (HET >0) in the NON-PAR of individuals carrying a single copy of ECAX (X0, 64,XYDSDS, and males). In these, HET was almost 2% on average, which is 20 times higher than that reported in humans using the same technology (0.1%, according to Saunders *et al.* (2007)). These false calls were previously ascribed to diverse causes, but mostly to the quality of the samples (Gunnarsson *et al.* 2008; Singh *et al.* 2016). However, our samples were collected and processed in ideal conditions, and the QC values of the genotyping process were optimal. Therefore, our best hypothesis is that those "impossible" NON-PAR heterozygous calls in hemizygous individuals could be produced by the remapping of SNPs to different chromosomes (up to 5% according to Beeson *et al.* (2019) occurring after the release of the updated version of the equine reference genome in 2018 (Kalbfleisch *et al.* 2018), especially when both SNP-arrays available in horses were developed using the previous version of the reference genome released in 2007 (Schaefer & McCue 2020). In any case, such abnormal calls did not alter the robustness of the methodology employed in the present study.

Finally, we concluded that inbreeding did not alter our CNA detection ability. This is an important finding in horses, since inbreeding levels are usually high (Petersen *et al.* 2013), but especially so in the PRE breed, where it is not difficult to find individuals with F >25% (Perdomo-González *et al.* 2020). We tested the hypothesis in a small dataset of mares with extreme F_{ROHX} values (higher than 50%), where most of the SNPs employed in the analysis were monomorphic and non-informative. However, the large number of

SNP employed in each chromosome analyzed (even in the MD and LD datasets) has reduced the probability of obtaining an erroneous CNA by chance due to the occurrence of a homozygous pattern in an entire chromosome region.

To sum up, we have developed a simple and robust analytical technique to detect some of the most important chromosomal abnormalities reported in horses by analyzing the raw intensity data produced by SNP-based genotyping arrays. We have also demonstrated that this methodology could be performed using high and medium density genotyping chips, and in individuals with increased inbreeding values, without affecting its precision. Since the methodology is utilized in an open, flexible bioinformatic environment, their integration into routine laboratory workflows and breeding programs is perfectly feasible.

Acknowledgments

We thank the Asociación Nacional de Criadores de Caballos de Pura Raza Española for providing the samples. Funds for this project were obtained from PICT2016-0359 and PICT-2018-0227 grants (ANPCyT, Argentina, PI: Sebastián Demyda Peyrás) and AGL2017-84217-P grant (MINECO, Spain, PI: Antonio Molina). Yamila Pirosanto is an ANPCyT Ph.D. fellow. We would also like to thank Dr. Ruth Barral-Arca (Thermofisher, Spain) for her expertise and technical support in the analysis of genotyping data and A. Di Maggio for the manuscript correction.

Data Availability

The raw dataset employed in this study is available on the Mendeley repository at the following access link: http://dx.doi.org/10.17632/bnt6vvwcr7.1

Conflict of interest

The authors have no conflicts of interest to declare.

References

Albarella S., De Lorenzi L., Catone G., Magi G.E., Petrucci L., Vullo C., D'Anza E., Parma P., Raudsepp T., Ciotola F. & Peretti V. (2018) Diagnosis of XX/XY Blood Cell Chimerism at a Low Percentage in Horses. *Journal of Equine Veterinary Science* **69**, 129-35.

Anaya G., Fernández M.E., Valera M., Molina A., Azcona F., Azor P., Solé M., Moreno-Millán M. & Demyda-Peyrás S. (2018) Prevalence of twin foaling and blood chimaerism in purebred Spanish horses. *The Veterinary Journal* **234**, 142-4.

Anaya G., Molina A., Valera M., Moreno-Millan M., Azor P., Peral-Garcia P. & Demyda-Peyras S. (2017) Sex chromosomal abnormalities associated with equine infertility: validation of a simple molecular screening tool in the Purebred Spanish Horse. *Animal Genetics* **48**, 412-9.

Anaya G., Moreno-Millán M., Bugno-Poniewierska M., Pawlina K., Membrillo A., Molina A. & Demyda-Peyrás S. (2014) Sex reversal syndrome in the horse: Four new cases of feminization in individuals carrying a 64,XY SRY negative chromosomal complement. *Animal Reproduction Science* **151**, 22-7.

Bannasch D., Rinaldo C., Millon L., Latson K., Spangler T., Hubberty S., Galuppo L. & Lowenstine L. (2007) SRY negative 64,XX intersex phenotype in an American saddlebred horse. *The Veterinary Journal* **173**, 437-9.

Beeson S.K., Schaefer R.J., Mason V.C. & McCue M.E. (2019) Robust remapping of equine SNP array coordinates to EquCab3. *Animal Genetics* **50**, 114-5.

Berry D.P., Wolfe A., O'Donovan J., Byrne N., Sayers R.G., Dodds K.G., McEwan J.C., O'Connor R.E., McClure M. & Purfield D.C. (2017) Characterization of an X-chromosomal non-mosaic monosomy (59, X0) dairy heifer detected using routinely available single nucleotide polymorphism genotype data. *Journal of Animal Science* **95**, 1042-9.

Biscarini F., Cozzi P., Gaspa G. & Marras G. (2018) DetectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes in R.

Bolormaa S., Gore K., Van Der Werf J.H.J., Hayes B.J. & Daetwyler H.D. (2015) Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Animal Genetics* **46**, 544-56.

Bugno M., Słota E. & Kościelny M. (2007a) Karyotype evaluation among young horse populations in Poland. *Schweizer Archiv fur Tierheilkunde* **149**, 227-32.

Bugno M., Słota E., Pieńkowska-Schelling A. & Schelling C. (2007b) Detection of equine X cromosome mosaicism in a mare using an equine X whole chromosome painting probe (WCPP) - A case report. *Acta Veterinaria Hungarica* **55**, 207-12.

Bugno M., Słota E. & Zabek T. (2001) Two cases of subfertile mares with 64,XX/63,X mosaic karyotype. *Annals of Animal Science* **1**, 7-11.

Demyda-Peyras S., Bugno-Poniewierska M., Pawlina K., Anaya G. & Moreno-Millán M. (2013) The use of molecular and cytogenetic methods as a valuable tool in the detection of chromosomal abnormalities in horses: A Case of sex chromosome chimerism in a Spanish Purebred colt. *Cytogenetic and Genome Research* **141**, 277-83.

Dowle M. & Srinivasan A. (2019) data.table: Extension of data.frame.

Gamo S., Tozaki T., Kakoi H., Hirota K.I., Nakamura K., Nishii N., Alumunia J. & Takasu M. (2019) X monosomy in the endangered Kiso horse breed detected by a parentage test using sex chromosome linked genes and microsatellites. *J Vet Med Sci* **81**, 91-4.

Gel B. & Serra E. (2017) karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-90.

Goodrich D., Xing T., Tao X., Lonczak A., Zhan Y., Landis J., Zimmerman R., Scott R.T., Jr. & Treff N.R. (2017) Evaluation of comprehensive chromosome screening platforms for the detection of mosaic segmental aneuploidy. *Journal of Assisted Reproduction and Genetics* **34**, 975-81.

Goszczynski D., Molina A., Terán E., Morales-Durand H., Ross P., Cheng H., Giovambattista G. & Demyda-Peyrás S. (2018) Runs of homozygosity in a selected cattle population with extremely inbred bulls: Descriptive and functional analyses revealed highly variable patterns. *PLOS ONE* **13**.

Gunnarsson R., Staaf J., Jansson M., Ottesen A.M., Göransson H., Liljedahl U., Ralfkiær U., Mansouri M., Buhl A.M., Smedby K.E., Hjalgrim H., Syvänen A.C., Borg Å., Isaksson A., Jurlander J., Juliusson G. & Rosenquist R. (2008) Screening for copy-number alterations and loss of heterozygosity in chronic lymphocytic leukemia - A comparative study of four differently designed, high resolution microarray platforms. *Genes Chromosomes and Cancer* **47**, 697-711.

Hadley W., Romain F., Lionel H. & Müller K. (2020) dplyr: A Grammar of Data Manipulation.

Harton G.L., Munné S., Surrey M., Grifo J., Kaplan B., McCulloh D.H., Griffin D.K. & Wells D. (2013) Diminished effect of maternal age on implantation after preimplantation genetic diagnosis with array comparative genomic hybridization. *Fertility and Sterility* **100**, 1695-703.

Holl H.M., Lear T.L., Nolen-Walston R.D., Slack J. & Brooks S.A. (2013) Detection of two equine trisomies using SNP-CGH. *Mammalian Genome* **24**, 252-6.

Indahl U.G., Næs T. & Liland K.H. (2018) A similarity index for comparing coupled matrices. *Journal of Chemometrics* **32**, e3049.

Kalbfleisch T.S., Rice E.S., DePriest M.S., Walenz B.P., Hestand M.S., Vermeesch J.R., O'Connell B.L., Fiddes I.T., Vershinina A.O., Saremi N.F., Petersen J.L., Finno C.J., Bellone R.R., McCue M.E., Brooks S.A., Bailey E., Orlando L., Green R.E., Miller D.C., Antczak D.F. & MacLeod J.N. (2018) Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology* **1**, 197.

Kjöllerström H.J., Collares-Pereira M.J. & Oom M.M. (2011) First evidence of sex chromosome mosaicism in the endangered Sorraia Horse breed. *Livestock Science* **136**, 273-6.

Lear T.L. & McGee R.B. (2012) Disorders of sexual development in the domestic horse, Equus caballus. *Sexual Development* **6**, 61-71.

Lopes F.B., Wu X.L., Li H., Xu J., Perkins T., Genho J., Ferretti R., Tait R.G., Jr., Bauck S. & Rosa G.J.M. (2018) Improving accuracy of genomic prediction in Brangus cattle by adding animals with imputed low-density SNP genotypes. *Journal of Animal Breeding and Genetics* **135**, 14-27.

Markello T.C., Carlson-Donohoe H., Sincan M., Adams D., Bodine D.M., Farrar J.E., Vlachos A., Lipton J.M., Auerbach A.D., Ostrander E.A., Chandrasekharappa S.C., Boerkoel C.F. & Gahl W.A. (2012) Sensitive quantification of mosaicism using high density SNP arrays and the cumulative distribution function. *Molecular genetics and metabolism* **105**, 665-71.

McCoy A.M., Beeson S.K., Rubin C.-J., Andersson L., Caputo P., Lykkjen S., Moore A., Piercy R.J., Mickelson J.R. & McCue M.E. (2019) Identification and validation of genetic variants predictive of gait in standardbred horses. *PLOS Genetics* **15**, e1008146.

McGivney B.A., Han H., Corduff L.R., Katz L.M., Tozaki T., MacHugh D.E. & Hill E.W. (2020) Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Scientific Reports* **10**.

Neuhauser S., Handler J., Schelling C. & Pieńkowska-Schelling A. (2019) Fertility and 63,X Mosaicism in a Haflinger Sibship. *Journal of Equine Veterinary Science* **78**, 127-33.

Perdomo-González D.I., Sánchez-Guerrero M.J., Molina A. & Valera M. (2020) Genetic Structure Analysis of the Pura Raza Español Horse Population through Partial Inbreeding Coefficient Estimation. *Animals* **10**, 1360.

Petersen J.L., Mickelson J.R., Cothran E.G., Andersson L.S., Axelsson J., Bailey E., Bannasch D., Binns M.M., Borges A.S., Brama P., da Câmara Machado A., Distl O., Felicetti M., Fox-Clipsham L., Graves K.T., Guérin G., Haase B., Hasegawa T., Hemmann K., Hill E.W., Leeb T., Lindgren G., Lohi H., Lopes M.S., McGivney B.A., Mikko S., Orr N., Penedo M.C.T., Piercy R.J., Raekallio M., Rieder S., Røed K.H., Silvestrelli M., Swinburne J., Tozaki T., Vaudin M., M. Wade C. & McCue M.E. (2013) Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLOS ONE* **8**.

Popova T., Manié E., Stoppa-Lyonnet D., Rigaill G., Barillot E. & Stern M.H. (2009) Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biology* **10**, R128.

Power M.M. (1990) Chromosomes of the horse. In: *Domestic animal cytogenetics* (pp. 131-67. Academic Press, Inc., San Diego, Ca.

Power M.M. & Leadon D.P. (1990) Diploid-triploid chimaerism (64, XX/96,XXY) in an intersex foal. *Equine Veterinary Journal* **22**, 211-4.

Pryce J.E. & Daetwyler H.D. (2012) Designing dairy cattle breeding schemes under genomic selection: A review of international research. *Animal Production Science* **52**, 107-14.

R-Core-Team (2020) R: A language and environment for statistical computing v4.0.3 "Bunny-Wunnies Freak Out". (ed. by R Foundation for Statistical Computing V, Austria).

Raudsepp T. & Chowdhary B.P. (2008) The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. *Cytogenetic and Genome Research* **121**, 102-9.

Raudsepp T. & Chowdhary B.P. (2016) Chromosome aberrations and fertility disorders in domestic animals. *Annual Review of Animal Biosciences* **4**, 15-43.

Raudsepp T., Das P.J., Avila F. & Chowdhary B.P. (2012) The Pseudoautosomal Region and Sex Chromosome Aneuploidies in Domestic Species. *Sexual Development* **6**, 72-83.

Robert P. & Escoufier Y. (1976) A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **25**, 257-65.

Saunders I.W., Brohede J. & Hannan G.N. (2007) Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. *Genomics* **90**, 291-6.

Schaefer R.J. & McCue M.E. (2020) Equine Genotyping Arrays. *Veterinary Clinics of North America - Equine Practice* **36**, 183-93.

Schaefer R.J., Schubert M., Bailey E., Bannasch D.L., Barrey E., Bar-Gal G.K., Brem G., Brooks S.A., Distl O., Fries R., Finno C.J., Gerber V., Haase B., Jagannathan V., Kalbfleisch T., Leeb T., Lindgren G., Lopes M.S., Mach N., da Câmara Machado A., MacLeod J.N., McCoy A., Metzger J., Penedo C., Polani S., Rieder S., Tammen I., Tetens J., Thaller G., Verini-Supplizi A., Wade C.M., Wallner B., Orlando L., Mickelson J.R. & McCue M.E. (2017) Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* **18**, 565.

Shilton C.A., Kahler A., Davis B.W., Crabtree J.R., Crowhurst J., McGladdery A.J., Wathes D.C., Raudsepp T. & de Mestre A.M. (2020) Whole genome analysis reveals aneuploidies in early pregnancy loss in the horse. *Scientific Reports* **10**.

Singh R.R., Mehrotra M., Chen H., Almohammedsalim A.A., Sahin A., Bosamra A., Patel K.P., Routbort M.J., Lu X., Ronald A., Mishra B.M., Virani S., Medeiros L.J. & Luthra R. (2016) Comprehensive Screening of Gene Copy Number Aberrations in Formalin-Fixed,

Paraffin-Embedded Solid Tumors Using Molecular Inversion Probe–Based Single-Nucleotide Polymorphism Array. *Journal of Molecular Diagnostics* **18**, 676-87.

Szczerbal I., Nowacka-Woszuk J., Kopp-Kuhlman C., Mackowski M. & Switonski M. (2020) Application of droplet digital PCR in diagnosing of X monosomy in mares. *Equine Veterinary Journal* **52**, 627-31.

Tang Y.-C. & Amon A. (2013) Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**, 394-405.

Thermofisher (2013) Axiom CNV Summary Tool user manual. URL https://tools.thermofisher.com/content/sfs/manuals/axiom_cnv_summary_tool_usermanual.pdf.

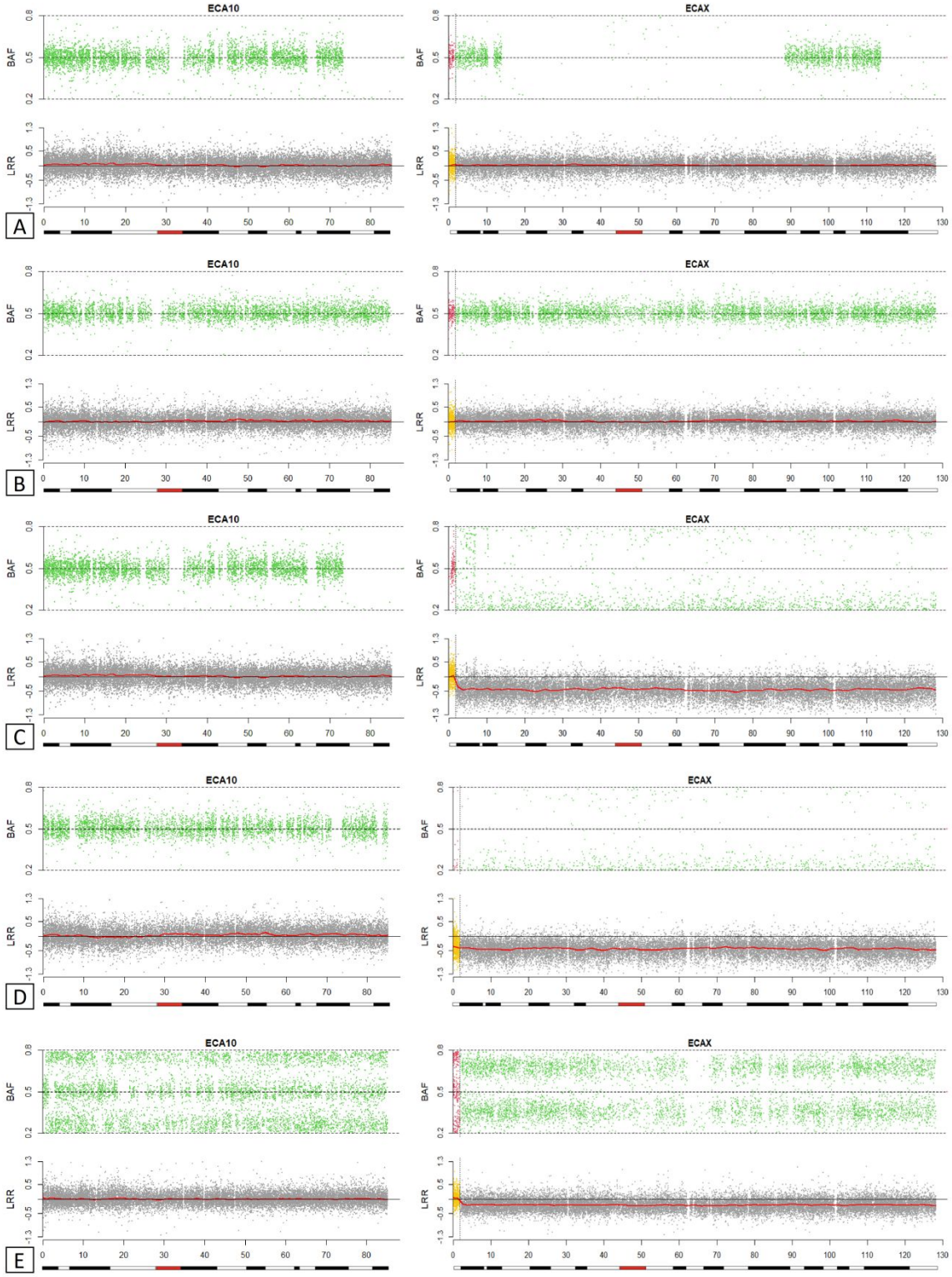
ThermoFisher (2019) Axiom copy number analysis. URL https://assets.thermofisher.com/TFS-Assets/GSD/Technical-Notes/tech-note-axiom-cnv.pdf.

Villagómez D.A.F., Lear T.L., Chenier T., Lee S., McGee R.B., Cahill J., Foster R.A., Reyes E., St John E. & King W.A. (2011) Equine disorders of sexual development in 17 mares including XX, SRY-negative, XY, SRY-negative and XY, SRY-positive genotypes. *Sexual Development* **5**, 16-25.

Villagómez D.A.F. & Pinton A. (2008) Chromosomal abnormalities, meiotic behavior and fertility in domestic animals. *Cytogenetic and Genome Research* **120**, 69-80.

Wiggans G.R., Cole J.B., Hubbard S.M. & Sonstegard T.S. (2017) Genomic Selection in Dairy Cattle: The USDA Experience. *Annual Review of Animal Biosciences* **5**, 309-27.

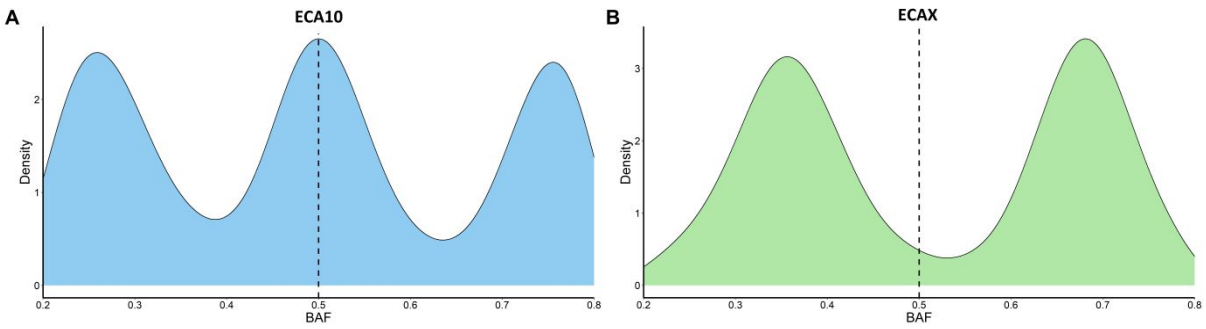
Figure 1



Footnote

b allele frequency (BAF) (green) and *log R ratio* (LRR) (grey) values of HOM-mares (A), mares(B), males (C), 63,X (D) and 64,XX/64,XY (E) groups. In each figure, ECA10 is plotted on the left, and ECAX is plotted on the right. In ECAX, BAF and LRR values of the PAR region are plotted in red and yellow respectively. X-axis values are expressed in Mb. HOM-MARES (Fig. 1a), MARES (Fig. 1b) and 64,XXDSD (Fig. 1e) showed a similar pattern compatible with females on PAR and NON-PAR, whereas MALES (Fig. 1c) and 64,XYDSD (Fig. 1e) showed a similar pattern compatible with males. In contrast, 63,X mares (Fig. 1d) showed a differential pattern in PAR ECAX (low LRR), allowing a clear differentiation from the other groups.

Figure 2



Density plot of BAF values in the 64,XX/64,XY group. ECA10 (A) shows a trimodal peak. ECAX (B) shows a bimodal peak produced by the chromosomal imbalance between cell lines from the same individual.

552 **Table 1: Analysis of LRR and HET means in 7 groups of PRE horses with different sex-karyotypes using a HD SNP dataset**

553

	ECA10		PAR ECAX		NON-PAR ECAX		ECAY
	HET	LRR	HET	LRR	HET	LRR	
HOM-Mare	19.99±5.42	0.01±0.27	20.62±8.85^{A}	-0.01±0.32^{A}	9.00±2.74^{b}	0.01±0.26^{a}	NO
Mare	22.78±3.82	0.00±0.26	24.57±6.55^{A}	-0.01±0.28^{A}	20.35±3.20^{a}	0.00±0.24^{a}	NO
Male	23.88±3.61	0.01±0.27	27.19±4.14^{A}	0.00±0.30^{A}	1.67±0.92^{c}	-0.42±0.28^{c}	YES
X0	25.37±1.75	0.02±0.33	0.44±0.21^{B}	-0.32±0.35^{B}	1.05±0.65^{c}	-0.42±0.30^{c}	NO
XXDSD	23.53±2.36	0.00±0.27	29.73±1.56^{A}	0.01±0.29^{A}	21.77±2.52^{a}	-0.01±0.26^{c}	NO
XYDSD	25.09±2.25	0.01±0.29	26.49±5.50^{A}	0.04±0.31^{A}	2.25±2.20^{c}	-0.46±0.30^{a}	YES
XX/XY	22.09±3.50	0.00±0.23	23.30±3.35^{A}	0.00±0.26^{A}	15.18±6.18^{a}	-0.22±0.23^{b}	YES

Values are stated as mean ± SD. Within columns, different superscripts show statistical differences at $p<0.05$ (lower case) and $p<0.01$ (upper case)

554