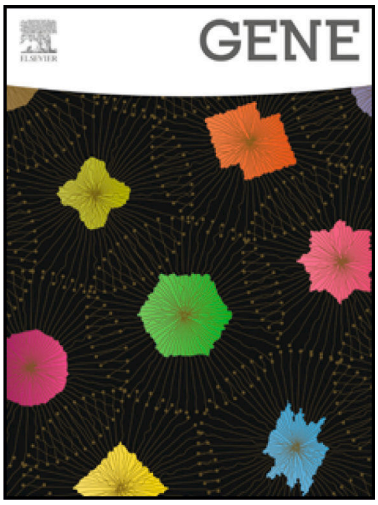


Analysis of Doberman Pinscher and Toy Poodle samples with targeted next-generation sequencing

A. Arizmendi, G. Rudd Garces, J.A. Crespi, L.H. Olivera, L.S. Barrientos, P. Peral García, G. Giovambattista

PII: S0378-1119(22)00889-7
DOI: https://doi.org/10.1016/j.gene.2022.147069
Reference: GENE 147069

To appear in: *Gene Gene*

Received Date: 27 April 2022
Revised Date: 4 November 2022
Accepted Date: 18 November 2022

Please cite this article as: A. Arizmendi, G. Rudd Garces, J.A. Crespi, L.H. Olivera, L.S. Barrientos, P. Peral García, G. Giovambattista, Analysis of Doberman Pinscher and Toy Poodle samples with targeted next-generation sequencing, *Gene Gene* (2022), doi: https://doi.org/10.1016/j.gene.2022.147069

Analysis of Doberman Pinscher and Toy Poodle samples with targeted next-generation sequencing

A. Arizmendi[1,2], G. Rudd Garces[1], J.A. Crespi[1], L.H. Olivera[1], L.S. Barrientos[1], P. Peral García[1], and G. Giovambattista[1]

[1] IGEVET – Instituto de Genética Veterinaria "Ing. Fernando N. Dulout" (UNLP-CONICET LA PLATA), Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, Calle 60 y 118 S/N, 1900, La Plata, Buenos Aires, Argentina. [2]Servicio de Cardiología, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, Calle 60 y 118 S/N, 1900, La Plata, Buenos Aires, Argentina

Corresponding author: Guillermo Giovambattista, IGEVET – Instituto de Genética Veterinaria "Ing. Fernando N. Dulout" (UNLP-CONICET LA PLATA), Facultad de Ciencias Veterinarias UNLP, 60 y 118 S/N, 1900, La Plata, Buenos Aires, Argentina.
E-mail: guillermogiovambattista@gmail.com

HIGHLIGHTS

- GBS is a powerful tool to study simultaneously traits and diseases in large samples
- GBS shows a high degree of accuracy.
- Higher numbers of SNPs are needed to exclude a false progeny in inbred dogs.
- Most genetic conditions were monomorphic in a specific breed or breed group.
- Validation of GBS results for genetic disorders showed 94-100% concordance.

Abstract

Next-generation sequencing (NGS) technologies have enabled the identification of many causal variants of genetic disorders, the development of parentage tests and the analysis of multiple traits in domestic animals. In this study, we evaluated the performance of a Canine Targeted Genotyping-by-Sequencing (GBS) custom panel (Thermo Fisher Scientific, Waltham, Ma, USA) in a cohort of 95 dog DNA samples, comprising 76 Doberman Pinschers and 19 Toy Poodles from Argentina. The used panel included 383 targets (228 parentage SNVs, 137 genetic disorder markers and 18 trait markers). While paternity analysis showed correct duo (97.4%; LOD > $2.98E^{+13}$) and trio (100%; LOD > $2.20E^{+15}$) parentage assignment, the panel resulted still insufficient for excluding close relatives in inbred populations. In this sense, close relatives were wrongly assigned as parents in 12.6% of duos and 0.3% of trios. We detected 17 polymorphic markers (genetic disorders, n=4; hair type, n=3; coat color, n=10) and estimated their allele frequencies in the studied breeds. The accuracy of targeted GBS results were evaluated for three markers that were associated with Progressive rod-cone degeneration, von Willebrand disease type 1 and dilated cardiomyopathy by pyrosequencing and Sanger sequencing genotyping, showing 94-100% concordance among assays. The targeted GBS custom panel resulted cost-effective strategy to study the prevalence of genetic disorders and traits in a large number of samples and to analyze genetic interactions between previously reported variants. Once assays based on AgriSeq technology were standardized, their uses are a good strategy for large-scale routine genetic evaluation of animal populations.

Keywords: *Canis lupus familiaris*, targeted genotyping-by-sequencing, SNVs, genetic disorders, paternity test

Abbreviations list

NGS: Next-generation sequencing

GBS: Genotyping-by-Sequencing

SNP: Single Nucleotide Polymorphism

WGS: whole-genome sequencing

DNA: Deoxyribonucleic Acid

RADSeq: Restriction site-associated DNA sequencing

DP: Doberman pinscher

TP: Toy Poodle

GATK: Genome Analysis Toolkit

STR: short tandem repeat

ISAG: International Society for Animal Genetics

n_a: average number of alleles per locus

$h_{e:}$ expected heterozygosity

Non-EP: Non-expected exclusion power

1-MP: non-match probability

PRCD: Progressive Rod-Cone Degeneration

vWD1: von Willebrand disease type 1

DCM: Dilated Cardiomyopathy

MAF: Minor allele frequency

EP: exclusion power

PDK4: Pyruvate dehydrogenase kinase isozyme 4

FBN2: Fibrillin 2

KRT71: Keratin 71

TBXT: T-Box Transcription Factor T

CBD103: beta-defensin 103

MC1R: melanocortin 1 receptor

CFA: Chromosome Canis Familiaris

1. Introduction

Genetic markers are polymorphic, inheritable and detectable DNA sequences widely used to evaluate parentage relationships and to detect causal mutations for genetic disorders and traits (e.g., coat color) in dogs (Kaelin et al., 2012; Kanthaswamy et al., 2019; Leeb et al. 2017). According to the Online Mendelian Inheritance in Animals database (https://www.omia.org/; accessed on April 20th, 2022), more than 450 likely causal variants affecting 324 genes have been identified in 386 known Mendelian traits/disorders in dogs. The identification of breed-specific genetic variants and its epidemiology has enabled genetic testing and improved breeding programs and, thus, animal welfare (Donner et al., 2018). Compared with whole-genome sequencing (WGS), targeted next-generation sequencing (NGS) is aimed at achieving 'targeted enrichment' of genome subregions to significantly reduce the sequencing of genomic loci of interest (sequencing coverage), increase the number of reads by genomic position (sequencing depth) and reduce costs and efforts (Mertes et al., 2011). This method simultaneously sequences millions of DNA fragments, so that it allows to analyze from the whole genome to several genes or gene regions with a single test (Yohe and Thyagarajan, 2017). The need for highly consistent detection of informative genetic markers is critical for genetic trait/disorder detection. The advantage of targeted genotyping-by-sequencing (GBS) methods over non-targeted GBS approaches (*e.g.*, Restriction site-associated DNA sequencing -RADSeq-) is that they are less susceptible to allele dropouts and missing data. AgriSeq is the trademark of AmpliSeq technology of Thermo Fisher based on targeted GBS assay that was development for agri-genomic application, allowing to genotype hundreds to thousands of markers simultaneously in a highly reproducible manner across diverse sample sets (Siddavatam et al., 2017; Gujjula et al., 2019; Diepenbroek et al., 2020; Elder et al., 2021).

The goal of this study was to evaluate and validate the GBS custom assay as a potential NGS system for genotyping parentage SNV panels and causal variant for genetic disorders and traits in 95 canine samples from Argentina.

2. Material and Methods

DNA samples from 76 related and unrelated Doberman Pinschers (DP) and 19 unrelated Toy Poodles (TP) were genotyped using the Canine Targeted GBS custom panel (Thermo Fisher Scientific, Waltham, Ma, USA) (Tables S1-S4). The data obtained from this beta version allowed the design of two commercial panels (AgriSeqTM Canine SNP Parentage and ID and AgriSeqTM Canine Traits and Disorders, Thermo Fisher Scientific) that were recently released. Genotyping information was obtained from 383 markers (228 parentage SNVs, 137 genetic disorder markers and 18 trait markers). The sequences of the PCR primers were kept proprietary by Thermo Fisher Scientific. The extraction and purification of DNA from blood samples was performed using the Wizard® Genomic DNA Purification Kit (Promega, WI, USA) according to the manufacturer's recommendations. The obtained DNAs were used to build the NGS libraries, which were sequenced using Ion GeneStudio™ S5 system (Thermo Fisher Scientific) at the Thermo Fisher development laboratory (Austin, TX, USA). The FASTq file analysis was performed following the Genome Analysis Toolkit (GATK) recommended workflow (https://gatk.broadinstitute.org/) using the CanFam 3.1 reference genome. Sample call rate, known as the percent of markers generating a genotype call for a specific sample, and marker call rate, the percent of samples generating a genotype call for a specific marker, were calculated. The workflow included a processing step of base (Quality Score) recalibration to detect and correct for patterns of systematic errors that can originate from biochemical processes during library preparation and sequencing, from manufacturing defects in the chips,

or instrumentation defects in the sequencer. The SnpEff version 5.1 software (Cingolani et al., 2012) was used to annotate and predict the functional effect of genetic variants. In addition, this software included in its report the percentage of known variants (SNVs and indels) in the targeted regions present in the dbsnp Canis_Familiaris_v85.vcf SnpEff database, according to the total called ones. Genetic variants were visualized through IGV software (Robinson et al., 2011), and heterozygous animals for hotspot variants were confirmed when both alleles have similar numbers of reads. In addition, 40 DP from the original cohort were selected for short tandem repeat (STR) genotyping using the Canine Genotypes Panel 1.1 (Thermo Fisher Scientific) that encompasses 19 STR loci (18 autosomal and 1 sex-determining, Table S5) of the core panel of loci recommended by the Applied Genetics Committee of Companying Animals of the International Society for Animal Genetics (ISAG; https://www.isag.us/). All samples used in this study belong to the Biobank of the Institute of Veterinary Genetics (IGEVET, for its Spanish acronym, National Scientific and Technical Research Council-National University of La Plata School of Veterinary Sciences). The Institutional Commission for the Care and Use of Laboratory Animals (CICUAL, for its Spanish acronym) from the National University of La Plata School of Veterinary Sciences, Argentina, approved the experimental protocols (CICUAL number 56-1-16T).

Gene frequency, average number of alleles per locus (Na) and expected heterozygosity for each marker (He) of the DP population were estimated using the Genepop 4.7 software (Rousset, 2008). Non-expected exclusion power (Non-EP), non-match probability (1-MP) and paternity index of 228 parentage SNV and 18 STR markers were estimated for different scenarios (two known parents, one known parent, missing parents and individual identification) using Cervus 3.0 software (Kalinowski et al., 2007). Allele frequencies for disease/trait variants were estimated in each breed. The accuracy of targeted GBS results for markers associated with progressive rod-cone degeneration (PRCD), von Willebrand disease

type 1 (vWD1) and a deletion in *PDK4* gene associated with dilated cardiomyopathy (*PDK4*-DCM), was evaluated comparing with results obtained using pyrosequencing and Sanger sequencing, as described by Meurs et al. (2012) and Crespi et al. (2018, 2019). It should be noted, however, that the association of the *PDK4* variant with DCM reported by Meurs et al. (2012) was later refuted by Owczarek-Lipska et al. (2013) and Arizmendi et al. (2020).

3. Results and Discussion

3.1 Kit Performance

The Canine Targeted GBS custom panel showed an average 697x coverage, ranging from 143x to 1688x for individual target regions excluding the failed markers (call rate <98.9%), and 91% sample call rate (Table S6). These values were higher than the recommended minimum read depth for detecting SNVs for genetic disorders (20X) using targeted NGS (Yohe et al., 2015; Rehm., 2013). We identified low performing genetic markers, most of which were discarded from the commercial version of this kit. Analysis of the vcf files with SnpEff software showed 392 polymorphic SNVs and 417 polymorphic indels. Of these, 277 SNVs (70.66%) and 4 indels (0.96%) were present in the dbsnp Canis_Familiaris_v85.vcf database used by this software. Otherwise, 14 of the polymorphic SNVs were multiallelic. Despite the raw data was weighted and refined through the recalibration process, part of this novel indels could be artifacts (sequencing errors) and further validation studies using Sanger sequencing are needed. However, the present results show the power of targeted GBS to detect additional variants and haplotypes when nearby variants are present, making it a useful tool to determine genetic relationship between individuals (Figure 1).

3.2 Genetic Identification and Paternity Testing

The allele frequencies of the SNV and STR parentage panels used are presented in Tables S7 and S8. All parentage markers were polymorphic (average n_a and H_e values, 1.99 and 0.369 for SNVs and 4.22 and 0.455 for STRs). Minor allele frequency (MAF) was > 0.1 in 95% of SNVs. Non-EP varied from 0.0901 (one known parent) to $5.55E^{-0009}$ (individual identification) for STRs, and $2E^{-8}$ (one known parent) to $1.93E^{-0074}$ (individual identification) for SNVs, showing that the SNV panel exhibited higher exclusion power (EP) than the STR panel in all scenarios (Table 1). Similar results were reported by Fernández et al., (2013), who compared the effectiveness of STR and SNV panels in an inbred Angus herd. The SNV panel reached EP values > 0.999, although the second parent was unknown. In contrast, the STRs EP with only one known parent was lower than the recommended discrimination power value threshold (EP = 0.936).

The parentage analysis of DP was carried out using the 228 SNVs of the paternity panel, considering all the possible comparisons between genotyped dogs (945 comparisons). Of those, 27 dogs had a declared paternity and/or maternity (38 duos and 11 trios). These parentage analyses showed 100% correct assignment of the biological parents (Table S9). Subsequently, the analysis of relatives considered as parents resulted in 12.6% duos and 0.3% trios incorrectly assigned as biological parents, while 43.8% duos and 7.5% trios presented doubtful results. When full or half siblings were considered as putative parents, 23.9% of the cases were excluded, 51.1% showed doubtful paternity and the remaining 25% of the duo comparisons were incorrectly assigned. However, when one biological parent was included in the analysis, all full or half siblings were excluded (Table S8). Despite the age of siblings from the same litter would allow to dismiss them as possible parents this becomes difficult in full siblings from different litters or half-siblings, considering that dogs begin their reproductive activity very young and could be therefore misclassified as parents. These results

evidence the importance of increasing the number of markers and testing both parents to avoid false parentage assignment in dogs.

In addition, 217 paternity tests were analyzed using the 18 STR parentage panel. These comparisons included 17 DPs with a declared paternity and/or maternity (13 duos and 4 trios). Only one duo showed a doubtful result (one mismatch) while the remaining declared caseworks were assigned. Otherwise, relatives were wrongly assigned in 54% of duos and 3.2% of trios (Table S9). The STR and SNV panels displayed high performance to assign true parents due to low genotyping errors. However, STRs had a low-resolution power to exclude close relatives from paternity/maternity, in agreement with the non-EP estimated values. The false assignments may be due to the small founder group, the small effective population size and the high consanguinity of the Argentine DP cohort evaluated (Crespi et al., 2018). Furthermore, most of the routine genotyping of dogs done in genetic laboratories consists of the analysis of highly related pedigree animals. In this framework, a marker set should have enough EP to resolve any possible situation, including cases of paternity with multi-putative consanguineous relatives.

3.3 Genetic Disorders

The targeted GBS custom panel included 137 genetic disorder markers. Fourteen of these were not analyzed because they had call rate values below the established quality filters and were excluded from the commercial version of the kit. Wild type alleles were monomorphic in 119 markers in the studied cohort, despite four disorder causal variants were reported in DP and/or TP dogs (Brunson et al., 2004; Chen et al., 2008; Coates et al., 2010; Matthews et al., 1985; Rahman et al., 2013; Zeng et al., 2014). The absence of genetic variants could be explained by the fact that they correspond to diseases with early phenotypic and severe effect, they are distributed within certain pedigrees, and because of their low incidence (Tables S2).

Moreover, we found allelic variation in four markers, corresponding to vWD1, *PDK4*-DCM, PRCD and *FBN2* (Tables S2). The causal disorder allele for vWD1 and the variant in *PDK4* that could be associated with DCM were previously reported in DP, while *PRCD* genetic variant was detected in TP (Brewer et al., 1998; Meurs et al., 2012; Zangerl et al., 2006). The *FBN2* variant associated with canine hip dysplasia (Friedenberg et al., 2011) was polymorphic in both breeds. To the best of our knowledge, this is the first report of polymorphism of this *FBN2* variant in DP and TP breeds, with an allele frequency of 0.92 and 0.50, respectively.

The mutant allele frequency reported in Toy Poodles for the c.5G>A (p.Cys2Tyr) mutation in the *PRCD* gene (Zangerl et al., 2006) varies between 0.45 in Czech Republic and 0.09 in Japan (Dostal et al., 2011; Kohyama et al., 2015). In this study, allele frequency was 0.33 (Table S10), which agrees with previously reported allelic frequencies in Argentina (0.2 to 0.6) (Bernades et al., 2014; Crespi et al., 2018). Pyrosequencing validation of this marker showed 94% concordance (Table S11, Figure S1), with the only exception of an animal with A/A genotype detected by targeted GBS and G/A by pyrosequencing.

In DP, the allelic frequency of c.7437G>A *vWD* type 1 was 0.48, similar with the reported values of 0.41 in another study from Argentina (Crespi et al., 2018) and 0.51 in the United States (https://goo.gl/P1ePkm). This average value can vary significantly within families. In Poodles, the prevalence of the disease (1.64%) and its allelic frequency (0.055) are low (Mattoso et al., 2010, https://goo.gl/P1ePkm), probably explaining the absence of the G>A mutation in the genotyped TP. Pyrosequencing validation of this marker showed 94% concordance, excepting 3 out of 45 genotyped animals presenting an A/A genotype with targeted NGS and G/A with pyrosequencing (Table S11, Figure S3). The wrong genotype call for *PRCD* and *vWD* type 1 variants did not have low depth (individual depth > 180x), so these differences could be explained by an unequal allele amplification during the PCR step due to SNVs in the primer regions. Analyses of BAN files of these animals showed extremely low

reads of the allele drop-out. This could be one of the reasons why the *PRCD* marker was removed in the commercial version.

The development of DCM in Dobermans from the US was associated with a 16-base pair deletion in the *PDK4* gene (0.36 allele frequency) (Meurs et al., 2012), however this *PDK4* variant was not associated with DCM in a European Doberman cohort, in which the allele frequency was 0.16 (Owczarek-Lipska et al., 2013) neither in an Argentine Doberman cohort (0.15 allele frequency) (Arizmendi et al., 2020). The overall mutated allele frequency was 0.22 (Table S10) and had complete concordance with Sanger sequencing results (Table S11, Figure S2). Homozygosity for the del mutation was not detected in the analyzed population. This marker had a low performance with low coverage values in the currently evaluated samples, showing results in only 45 of the 76 DP analyzed. In concordance with this result, INDELs markers for genetic disorders in general exhibited lower call rates (95.94%) and higher number of failed markers (20.4%) than SNVs (97.45% and 3.6%). Noteworthy, the *PDK4* marker was excluded from the commercial version of the kit.

3.4 Trait markers

The custom panel includes 18 trait markers corresponding to 11 genes for tail length, hair type and coat color. All trait markers are described in Table S3. Seven markers were monomorphic in both breeds. Eleven and five markers were polymorphic in TP and DP, respectively. Genotype-phenotype segregations were seen in both breeds, in agreement with previous reports. For instance, the autosomal dominant trait curly coat (Cadieu et al., 2009), was C/C genotype for the *KRT71* gene (CanFam3.1, CFA27 g.2539211C>T) in DP and C/T and T/T genotype in TP, according to their phenotypes. Some conditions such as bobtail (Haworth et al., 2001), which presents the wild type allele C/C (CanFam3.1 CFA1 g.54192143G>C) for the *TBXT* gene, have not been reported in DP and TP. Regarding coat

color genes, genotype-phenotype correlation was observed. For instance, the production of *eumelanin* with the consequent entire black coat phenotype is given by the at least one dominant mutated *CBD103* variant (CanFam3.1 CFA16 g.58965449_58965451del) combined with at least one copy of the wild type *MC1R* receptor (Oguro-Okano et al., 2011; CanFam3.1 CFA5 g.63694334G>A). In this study, despite many of the TPs carried the *CBD103* mutation, they were also homozygous e/e for the *MC1R* gene and, consequently, dominant black could not be expressed.

4. Conclusions

- The GBS assay represents a cost-effective, rapid and high-throughput genotyping strategy to study parentage and the prevalence of genetic disorders and traits in many samples, as well as to analyze genetic interactions between previously reported variants.
- Regarding parentage analysis, the number of genetic markers included in the custom panel resulted insufficient to exclude close relatives in a highly inbred cohort of DP dogs.
- Although only a few genetic disorder markers were polymorphic in the studied cohorts, we confirmed the high call rate (>XX%) in most of the markers (112 out of the 137) included in the custom panel.
- 14 genetic disorder markers with low call rate, including *PDK4* variant, were excluded from the commercial version of the kit. Otherwise the *PRCD* marker was also removed.
- The large number of monomorphic genetic disorder markers could be explained because most of these variants were reported in specific breeds or breed groups.

- The results of trait genetic markers agreed with the phenotype of the genotyped animals.
- The present results contributed to the development of commercial panels, the AgriSeq™ Canine SNP Parentage and ID and the AgriSeq™ Canine Traits and Disorders (Thermo Fisher Scientific).

Availability of data

The genotype database used in the present study is available in Open Science Framework (OSF) Home (https://osf.io/qn8m4/?view_only=fb44e1edd6194c1397627e5421932eb0).

Acknowledgements

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Grant PUE-2016 N° 22920160100004CO and PIP-2021 N° 11220200102074CO) and Universidad Nacional de La Plata (UNLP, Grant V247). The authors thank Christopher C. Adams, Ravindra Ramadhar, Srinivas Chadaram and the Thermo Fisher Scientific BFX & NGS team for their technical support. They also thank breeders and dog owners for contributing with samples, and Adriana Di Maggio for manuscript correction and edition.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Although DNA samples were run by the Thermo Fisher Scientific team, they were not involved in the processing of data and the drafting of the manuscript.

References

Arizmendi, A., Batista, P.R., Crespi, J., Tórtora, M., Vercellini, M., Czernigow, M., Arias, D.O., Giovambattista, G. 2020. Analysis of *PDK4* gene deletion in a population of Doberman Pinschers from Argentina. In Research communications of the 30th ECVIM-CA online congress. *J Vet Intern Med*, 34: 3058-3166. https://doi.org/10.1111/jvim.15924.

Bernades, M., Gómez, N., del Prado, A., Elguero, B., Bernades, L., Martínez, M., Ivanic, J., Castillo, V., Salas, C., Frank, E. 2014. Progressive retinal atrophy (PRA): Prevalence of mutation in the PRCD gene in Poodle and Cocker breeds and its correlation with ophthalmological signs. Proceedings XIV Congreso Nacional de la Asociación de Veterinarios Especializados en Animales de Compañía de Argentina (AVEACA) 2014, 106.

Brewer, G.J., Venta, P.J., Schall, W.J., Yuzbasiyan-Gurkan, V., Li, J. 1998. DNA tests for von Willebrands disease in dobermans, scotties, shelties, and Manchester terriers. *Canine Pract.* 1998, 23:45.

Brunson DB, Hogan KJ. 2004. Malignant hyperthermia: a syndrome not a disease. *Vet Clin North Am Small Anim Pract.* 34(6):1419-1433.

Cadieu, E., Neff, M.W., Quignon, P., Walsh, K., Chase, K., Parker, H.G., VonHoldt, B.M., Rhue, A., Boyko, A., Byers, A., Wong, A., Mosher, D.S., Elkahloun, A.G., Spady, T.C., André, C., Lark, C.G., Cargill, M., Bustamante, C.D., Wayne, R.K., Ostrander, E.A. 2009. Coat variation in the domestic dog is governed by variants in three genes. *Science*. 326(5949), 150–153.

Chen X, Johnson GS, Schnabel RD, Taylor JF, Johnson GC, Parker HG, Patterson EE, Katz ML, Awano T, Khan S, O'Brien DP. 2008. A neonatal encephalopathy with seizures in standard poodle dogs with a missense mutation in the canine ortholog of ATF2. *Neurogenetics*. 9(1), 41-9.

Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* 6, 80–92.

Crespi, J.A., Barrientos, L.S., Giovambattista, G. 2018. von Willebrand disease type 1 in DP dogs: genotyping and prevalence of the mutation in the Buenos Aires region, Argentina. *J Vet Diagn Invest.* Mar;30(2), 310-314.

Crespi, J.A. 2019. Estimación de la prevalencia en Argentina de mutaciones causales de enfermedades caninas: gen de resistencia a multidrogas 1, von Willebrand tipo I y degeneración progresiva de conos y bastones (Tesis doctoral). Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, Argentina.

Coates JR, Wininger FA. 2010. Canine degenerative myelopathy. *Vet Clin North Am Small Anim Pract.* 40, 929–950.

Diepenbroek M, Bayer B, Schwender K, Schiller R, Lim J, Lagacé R, Anslinger K. Evaluation of the Ion AmpliSeq™ PhenoTrivium Panel: MPS-Based Assay for Ancestry and Phenotype Predictions Challenged by Casework Samples. *Genes (Basel)*. 2020 Nov 25;11(12):1398. doi: 10.3390/genes11121398

Donner, J., Anderson, H., Davison, S., Hughes, A.M., Bouirmane, J., Lindqvist, J., Lytle, K.M., Ganesan, B., Ottka, C., Ruotanen, P., Kaukonen, M., Forman, O.P., Fretwell, N., Cole, C.A., Lohi, H. 2018. Frequency and distribution of 152 genetic disease variants in over 100,000 mixed breed and purebred dogs. *PLoS Genet.* 14(4), e1007361.

Dostal, J., Hrdlicova, A., Horak, P. 2011. Progressive rod-cone degeneration (PRCD) in selected dog breeds and variability in its phenotypic expression. *Veterinarni Medicina*. Jun; 56(5), 243-47.

Elder, J.R., Fratamico, P.M., Liu, Y., Needleman, D.S., Bagi, L., Tebbs, R., Allred, A., Siddavatam, P., Suren, H., Gujjula, K.R., DebRoy, C., Dudley, E.G., Yan, X. 2021. A Targeted Sequencing Assay for Serotyping Escherichia coli Using AgriSeq Technology. *Front Microbiol.* Jan 15;11:627997. doi: 10.3389/fmicb.2020.627997

Haworth, K., Putt, W., Cattanach, B., Breen, M., Binns, M., Lingaas, F., Edwards, Y.H. 2001. Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mammalian Genome.* 12, 212–218.

Fernández, M.E., Goszczynski, D.E., Lirón, J.P., Villegas-Castagnasso, E.E., Carino, M.H., Ripoli, M.V., Rogberg-Muñoz, A., Posik, D.M., Peral-García, P., Giovambattista, G. 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genet Mol Biol*. 36(2), 185-191.

Friedenberg, S.G., Zhu, L., Zhang, Z., Foels Wv, Schweitzer, P.A., Wang, W., Fisher, P.J., Dykes, N.L., Corey, E., Vernier-Singer, M., Jung, S.W., Sheng, X., Hunter, L.S., McDonough, S.P., Lust, G., Bliss, S.P., Krotscheck, U., Gunn, T.M., Todhunter, R.J. 2011. Evaluation of a fibrillin 2 gene haplotype associated with hip dysplasia and incipient osteoarthritis in dogs. *Am J Vet Res.* 72(4), 530-40.

Gujjula, K.R., Baselgia, L., Wall, J., Suren, H., Siddavatam, P., Schmidt, J., Conrad, R. 2019. PO0927: The Hallmark of AgriSeqTM Technology: Highly Reproducible Genotype Calls and Identification of Novel Genotypes. *Plant and Animal Genome Conference XVII*. January 12-16 2019, San Diego, CA, USA.

Kaelin, C.B., Barsh, G.S. 2012. Molecular Genetics of Coat Color, Texture and Length in the Dog, in: Ostrander, E.A., Ruvinsky,A. (Eds.), The Genetics of the dog, 2nd Edition. Elaine Ostrander & Anatoly Ruvinsky, USA, pp. 57.

Kalinowski, S.T., Taper, M.L., Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*. 16, 1099-1106.

Kanthaswamy, S., Oldt, R.F., Montes, M., Falak, A. 2019. Comparing two commercial domestic dog (*Canis familiaris*) STR genotyping kits for forensic identity calculations in a mixed-breed dog population sample. *Anim Genet.* 50(1), 105-111.

Kohyama, M.1., Tada, N., Mitsui, H., Tomioka, H., Tsutsui, T., Yabuki, A., Rahman, M.M., Kushida, K., Mizukami, K., Yamato, O. 2016. Real-time PCR genotyping assay for canine progressive rod-cone degeneration and mutant allele frequency in Toy Poodles, Chihuahuas and Miniature Dachshunds in Japan. *J Vet Med Sci.* 78(3), 481–484.

Leeb, T., Müller, E.J., Roosje, P., Welle, M. 2017. Genetic testing in veterinary dermatology. *Vet Dermatol.* 28, 4-e1.

Matthews NS, de Lahunta A. 1985. Degenerative myelopathy in an adult miniature poodle. *J Am Vet Med Assoc.* 186(11), 1213–5.

Mattoso, C.R., Takahira, R.K., Beier, S.L., Araújo, J.P.Jr., Corrente, J.E. 2010. Prevalence of von Willebrand disease in dogs from Sao Paulo State, Brazil. *J Vet Diagn Invest*. Jan;22(1), 55-60.

Mertes, F., Elsharawy, A., Sauer, S., van Helvoort, J. M., van der Zaag, P. J., Franke, A., Nilsson, M., Lehrach, H., Brookes, A. J. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics.* 10(6), 374–386.

Meurs, K.M., Lahmers, S., Keene, B.W., White, S.N., Oyama, M.A., Mauceli, E., Lindblad-Toh, K. 2012. A splice site mutation in a gene encoding for *PDK4*, a mitochondrial protein, is associated with the development of dilated cardiomyopathy in the DP. *Hum Genet.* 131(8), 1319-25.

Oguro-Okano, M., Honda, M., Yamakazi, K., Okano, K. 2011. Mutations in the Melanocortin 1 Receptor, β-Defensin103 and Agouti Signaling Protein Genes, and Their Association with Coat Color Phenotypes in Akita-Inu Dogs. *J. Vet. Med. Sci.* 73(7), 853–858.

Owczarek-Lipska, M., Mausberg, T.B., Stephenson, H., Dukes-McEwan, J., Wess, G., Leeb, T. 2013. A 16-bp deletion in the canine *PDK4* gene is not associated with dilated cardiomyopathy in a European cohort of Doberman Pinschers. *Anim Genet.* Apr;44, 239.

Rahman MM, Yabuki A, Kohyama M, Mitani S, Mizukami K, Uddin MM, Chang HS, Kushida K, Kishimoto M, Yamabe R, Yamato O. 2013. Real-Time PCR Genotyping Assay for GM2 Gangliosidosis Variant 0 in Toy Poodles and the Mutant Allele Frequency in Japan. *J Vet Med Sci.* 76(2), 295-9.

Rehm, H.L., Bale S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., et al. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med.* 15:733–47.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. 2011. Integrative Genomics Viewer. *Nature Biotechnology.* 29, 24–26.

Siddavatam, P., Burrell, A., Ferretti, R., Allred, A., Ridley, K. 2017. Development of a custom ion AgriSeqTM genotyping-by-sequencing panel based on the ISAG bovine core parentage markers. Proceedings of the 22nd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Townsville, Queensland, Australia, 2-5 July 2017, pp.517-520.

Yohe S, Hauge A, Bunjer K, Kemmer T, Bower M, Schomaker M, Onsongo G, Wilson J, Erdmann J, Zhou Y, Deshpande A, Spears MD, Beckman K, Silverstein KA, Thyagarajan B. 2015. Clinical validation of targeted next-generation sequencing for

inherited disorders. *Arch Pathol Lab Med.* Feb;139(2):204-10. doi:10.5858/arpa.2013-0625-OA.

Yohe, S., Thyagarajan, B. 2017. Review of Clinical Next-Generation Sequencing. *Archives of Pathology & Laboratory Medicine*. Vol. 141, No. 11, pp. 1544-1557.

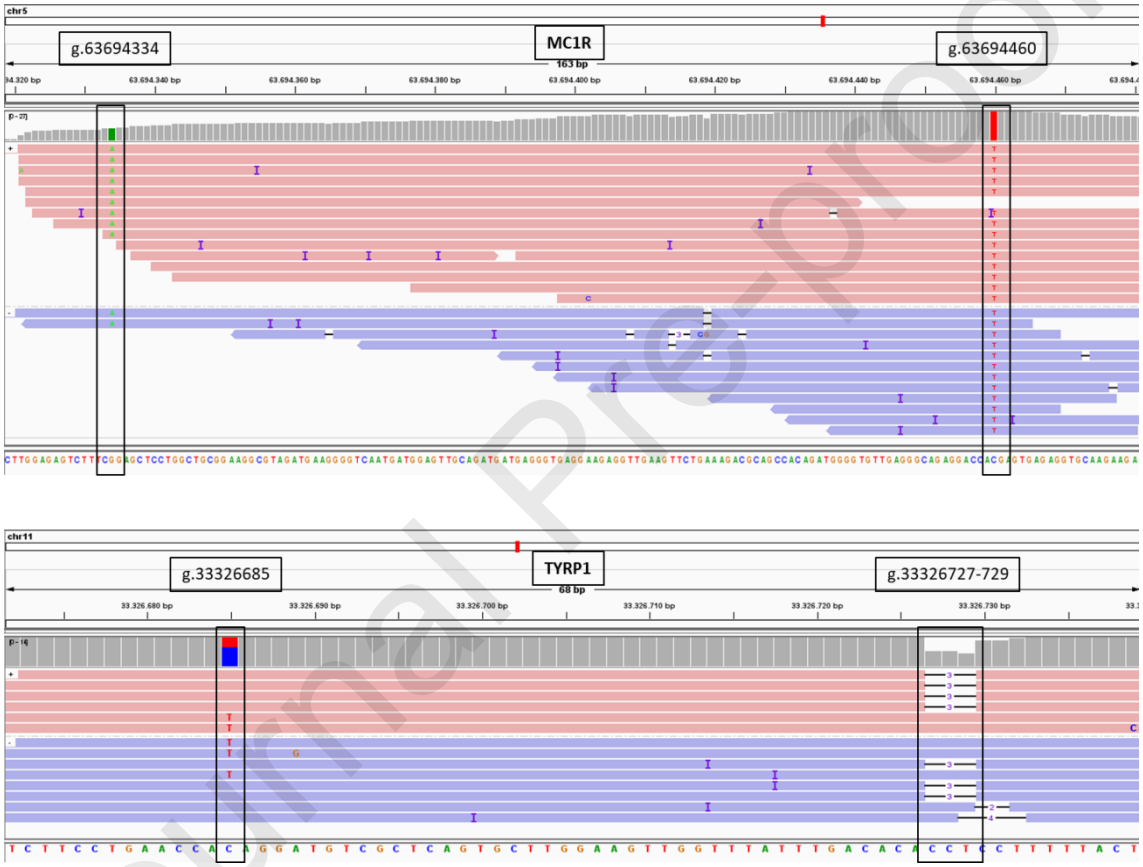
Zangerl, B., Goldstein, O., Philp, A.R., Lindauer, S.J., Pearce-Kelling, S.E., Mullins, R.F., Graphodatsky, A.S., Ripoll, D., Felix, J.S., Stone, E.M., Acland, G.M., Aguirre, G.D. 2006. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics.* 88(5), 551-63.

Zeng R, Coates JR, Johnson GC, Hansen L, Awano T, Kolicheski A, Ivansson E, Perloski M, Lindblad-Toh K, O'Brien DP, Guo J, Katz ML, Johnson GS. 2014. Breed Distribution of SOD1 Alleles Previously Associated with Canine Degenerative Myelopathy. *J Vet Intern Med.* 28(2), 515-21.

Table 1. Different scenarios of the SNV and STR sets of markers in the dog populations studied.

	SNVs	STRs
Number of genotyped individuals	94	40
Number of genotyped markers	228	18
Average number of alleles per locus	1.99	4.22
Average expected heterozygosity	0.369	0.455
Average non-exclusion probability		
one known parent	2.00E^{-8}	9.01E-02
two known parents	1.56E-16	6.49E-03
missing parents	8.89E-27	1.72E-04
individual identification	1.93E-74	5.55E-09
sibling identification	1.83E-38	1.40E-04

Figure 1. Display windows of the BAM files using the IGV 2.8.9 software (Robinson et al., 2011). a) *MC1R* gene region including two SNVs (g.63694460 C>T, allele E^M and g.63694334 G>A, allele *e*), and b) *TYRP1* gene region that includes SNV (g.33326685C>T, allele b^s;) and INDEL (g.33326727-33326729delCCT, allele b^d). The SNVs and the INDEL are indicated with a black box. Forward (pink) and reverse (blue) reads.



Supporting Information

Figure S1. Electropherograms, pyrograms and genotypes of the *PRCD* gene fragment obtained by DNA sequencing, pyrosequencing and targeted NGS, respectively. a) Homozygous normal genotype (G/G); b) heterozygous genotype (G/A); c) homozygous mutated genotype (A/A).

Figure S2. Electropherograms and genotypes of the *PDK4* gene fragment obtained by DNA sequencing and targeted NGS, respectively. a) Homozygous normal genotype; b) heterozygous genotype.

Figure S3. Electropherograms, pyrograms and genotypes of the c.7437G>A mutation of the *vWF* gene obtained by DNA sequencing, pyrosequencing and targeted NGS, respectively. a) Homozygous normal genotype (G/G); b) heterozygous genotype (G/A); c) homozygous mutated genotype (A/A).

Table S1: Parentage markers included in the Canine Targeted GBS custom panel (Thermo Fisher Scientific, USA).

Table S2: Genetic disorder markers included in the Canine Targeted GBS custom panel (Thermo Fisher Scientific, USA). a: low performing markers; b: markers excluded from the current version of the kit; c= polymorphic genetic conditions. ./. (non-functional marker).

Table S3: Trait markers included in the Canine Targeted GBS custom panel (Thermo Fisher Scientific, USA).

Table S4. Number of Doberman Pinscher and Toy Poodle samples obtained from the Institute of Veterinary Genetics Biobank.

Table S5: Detail information for the used microsatellites (STR).

Table S6: Estimated performance of the Targeted NGS Assay.

Table S7: Estimated allele frequency of the SNVs corresponding to the paternity panel.

Table S8: Estimated allele frequency of the microsatellites corresponding to the ISAG paternity panel.

Table S9. Parentage analysis in duos and trios using SNV and STR panels, considering alleged parents and different scenarios of false parents.

Table S10: Polymorphic genetic conditions in the samples analyzed.

Table S11: Concordance between the results of the targeted GBS custom panel with pyrosequencing and capillary sequencing of three selected diseases (n = number of dogs validated).

Authors' contributions

Conceptualization, G.G, A.A. and G.R.G.; data curation, A.A., G.R.G. and L.H.O.; investigation, A.A., G.R.G., J.A.C., L.H.O. and L.S.B.; supervision, G.G. and P.P.G.; writing—original draft, G.G, A.A. and G.R.G.; writing—review and editing, all authors.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Pilar Peral Garcia reports financial support was provided by National University of La Plata. Pilar Peral Garcia reports financial support was provided by Consejo Nacional de Investigaciones Científicas y Técnicas. Pilar Peral Garcia reports a relationship with National University of La Plata that includes: employment and funding grants. Pilar Peral Garcia reports a relationship with Consejo Nacional de Investigaciones Científicas y Técnicas that includes: employment.