Gene 854 (2023) 147111

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Segregation patterns and inheritance rate of copy number variations regions assessed in a Gochu Asturcelta pig pedigree

Katherine D. Arias [a], Juan Pablo Gutiérrez [b], Iván Fernandez [a], Nuria A. Menéndez-Arias [a], Isabel Álvarez [a], Félix Goyache [a,*]

[a] SERIDA-Deva, Camino de Rioseco 1225, 33394-Gijón, Spain
[b] Departamento de Producción Animal, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

ARTICLE INFO

Edited by Chuzhao Lei

Keywords:
Copy Number Variations
Mendelian inheritance
Transmission rate
Inheritance rate
Pedigree
Null alleles

ABSTRACT

Copy Number Variation Regions (CNVR) were subjected to pedigree analysis to contribute to the understanding of their segregation patterns. Up to 492 Gochu Asturcelta pig individuals forming 478 different parents-offspring trios (61 different families) were genotyped using the Axiom_PigHDv1 Array (658,692 SNPs). CNVR calling, performed using two different platforms (PennCNV and QuantiSNP), allowed to identify a total of 344 candidate CNVR on the 18 porcine autosomes covering about 106.8 Mb of the pig genome. Sixty-nine CNVR were identified, to some extent, in both the parents and the offspring and were classified as segregating CNVR. The other candidate CNVR were called in one or more progeny but in neither parent and classified either as singleton or recurrent *de novo* CNVR. Segregating CNVR were, on average, larger and more frequent than the recurrent *de novo* CNVR (444.8 kb vs 287.9 kb long and 34 vs 5 individuals, respectively). In any case, segregating CNVR did not conform to strict Mendelian inheritance patterns: estimates of average paternal and maternal transmission rates ranged from 11.0 % to 13.4 % and mean inheritance rate was below 21 %. This issue should be carefully considered when interpreting the results of CNV studies. Segregating CNVR, present across generations, are unlikely to be artifacts or false positives and can be hypothesized to be important at the population level.

1. Introduction

Genomic segments ranging in size from 50 base pairs (bp) to several megabases (Mb) differing in structure due to the occurrence of insertion, duplication or deletion events are usually referred to as Copy Number Variations (CNV; Feuk et al., 2006; Scherer et al., 2007). CNV have been used as informative markers in association studies for economically important traits in livestock (Chen et al., 2012; Wang et al., 2015; Bergamaschi et al., 2020; Qiu et al., 2021; Ding et al., 2022; Mo et al., 2022) due to their influence on phenotypic variability via the modification of gene structure and gene expression (Conrad et al., 2010; Wang et al., 2021; Wei et al., 2022). However, studies identifying CNV in different populations within species cannot be compared straightforwardly. Some genomic regions seem to be prone to recurrent CNV formation (Carvalho and Lupski, 2016). Accumulation of CNV across these genomic regions is not uniform and different livestock populations show differences in CNV prevalence. Therefore, CNV-based reports tend to show large differences at the breed level, mirroring their particular population histories (Fontanesi et al., 2011; Xie et al., 2016; Rafter et al., 2018; Bovo et al., 2021).

The identification of CNV may be challenged by noisy signals of SNP arrays selected and tested on the basis of their use in SNP genotyping (Winchester et al., 2009). However, CNV are unlikely to be artifacts or false positives if called across generations (Ramayo-Caldas et al., 2010; Fernández et al., 2014; Keel et al., 2019). However, little is known about the inheritance patterns of CNV and studies aiming at the assessment of the accordance of CNV variation with Mendelian inheritance are scant. Although CNV have been assumed as structurally similar to microsatellites, their behavior departs considerably from such kind of markers and, for instance, unlike SNPs and microsatellites, CNV cannot be accurately predicted using imputation algorithms (Rafter et al., 2020). Samarakoon et al. (2011), using a segregating population of cloned progeny lines of *Plasmodium falciparum*, found that only 26.5 % of the CNV identified were present in the parent lines and segregated in the progeny population.

Abbreviations: CNV, Copy Number Variations; CNVR, Copy Number Variations Regions; bp, base pairs; Mb, megabases; HMM, hidden Markov models; sdnCNVR, singleton *de novo* CNVR; rdnCNVR, recurrent *de novo* CNVR; sCNVR, segregating CNVR.

* Corresponding author at: Área de Genética y Reproducción Animal, SERIDA-Deva, 33394-Gijón, Spain.
E-mail address: fgoyache@serida.org (F. Goyache).

https://doi.org/10.1016/j.gene.2022.147111
Received 22 September 2022; Received in revised form 30 November 2022; Accepted 6 December 2022
Available online 9 December 2022

The present research aimed to contribute to the ascertainment of the segregation patterns of CNV regions (CNVR). To deal with this task, a complex pedigree of the endangered Gochu Asturcelta pig breed was available. Gochu Asturcelta is an extremely endangered pig breed that was subjected to a conservation programme in Asturias (Northern Spain) starting from 6 founders only (Menéndez et al., 2016a,b). Gochu Asturcelta is one of the last representatives of the ancient Celtic-Iberian pig strain of the Iberian Peninsula (Menéndez et al., 2016c). The Gochu Asturcelta pedigree was used to follow Mendelian inheritance of CNVR across parents-offspring trios and families to contribute to a better understanding of the nature of those genomic alterations.

2. Materials and methods

2.1. Samples and genotyping

A Gochu Asturcelta pig pedigree including 492 individuals previously analyzed in Arias et al. (2022) was genotyped using the Axiom Porcine Genotyping Array (Axiom_PigHDv1; 658,692 SNPs). The typed individuals were born from 1999 to 2009 (including three founders) and sampled in 14 different farms included in the breeding programme of ACGA. They belonged to 61 different families (descendants of the same parental couple), and formed 478 parents-offspring trios. Offspring genotypes derived from 15 genotyped boars and 28 genotyped sows and obtained from 96 different litters. Family size (number of offspring per parental couple) varied from 1 to 34. The individuals typed are representative of the initial stages of the recovery programme of the Gochu Asturcelta pig breed and included up to four complete generations in their genealogies.

The software Axiom Analysis Suite v4.0.3 (Thermo Fisher Scientific, Waltham, MA) was used to create both genotypic and intensity data useful for CNV calling and standard.ped and.map files useful for complementary analyses. SNPs with ambiguous chromosome locations and SNPs located on either sexual chromosomes or mitochondrial DNA were excluded. SNPs were mapped using the *Sscrofa* genome build 11.1 (Groenen et al., 2012).

2.2. CNV calling and editing

Following a previously published strategy (Goyache et al., 2021, 2022) two different platforms, PennCNV (Wang et al., 2007) and QuantiSNP (Colella et al., 2007), were used to perform CNV calling from the autosomes of each individual in the dataset. PennCNV and QuantiSNP were run using default parameters. Both software suites implement hidden Markov models to detect CNV based on the log of the ratio between the observed and the expected probe hybridization intensity of SNPs and the proportion of B alleles at a SNP. The probability of change in copy number between adjacent SNPs is dependent on their distance. However, whereas PennCNV uses a transition matrix to model realistic copy number transitions, QuantiSNP runs under a Bayesian framework to estimate probabilities for copy number states. In this respect, following the authors' recommendations, CNV identified with QuantiSNP with a Log Bayes Factor lower than 10 were not considered informative and, therefore, filtered out.

CNVR were first constructed within calling platform (either PennCNV or QuantiSNP) through merging, across individuals, overlapping CNV by at least 1 bp using the *merge* function of the software BedTools (Quinlan and Hall, 2010). The upper and lower bounds of these overlaps were considered potential CNVR within software. Finally, candidate CNVR were defined as the upper and lower bounds of the overlaps between potential CNVR identified within each calling platform using the *intersectBed* function of the BedTools software.

The distribution of the candidate CNVR identified across porcine was illustrated using the RIdeogram package of R environment (Hao et al., 2020).

2.3. Pedigree-based analyses

The program COLONY v.2.0.6.8 (Wang, 2019) was used to verify parentage in the pedigree.

Candidate CNVR were arbitrarily coded as dominant markers (presence/absence) according to their identification in a given individual. The program CERVUS 3.0 (Kalinowski et al., 2007) was used to identify the CNVR present in the offspring which were not present in either the father or the mother of the individual.

Following Samarakoon et al. (2011), candidate CNVR were classified into three different categories according to their presence in the parents or the offspring: a) 'singleton *de novo*' CNVR (sdnCNVR), defined as CNVR occurring in one individual (progeny) only; b) 'recurrent *de novo*' CNVR (rdnCNVR), defined as CNVR occurring in multiple progeny but in neither parent; and c) segregating CNVR (sCNVR) were identified in at least one of the parents and in at least one of the progeny.

Consistency of the presence of CNVR with the rules of Mendelian inheritance was assessed using the *–mendel* option of the program PLINK V 1.9. The software iterates through all trios and all variants checking for these errors (Chang et al., 2015). Considering the presence of CNVR in the parents, the origin of the violation of Mendelian inheritance was assigned to the father, the mother, or the offspring.

Following Keel et al. (2019) either the paternal and maternal transmission rates and inheritance rate for each parents-offspring trio and family in the pedigree were computed as follows: (a) in each parent-offspring pair, CNVR in one of the parents (either father or mother) also called in the offspring were counted and then divided by the total number of CNVR calls in the parent; and (b) CNVR calls in the offspring also present in at least one parent were counted and then divided by the total number of CNVR in the offspring. Ideal paternal and maternal transmission rates would be 50 % and inheritance rates would be 100 % (Keel et al., 2019).

2.4. Check for consistency of the results obtained

As an internal control for the consistency of the approach used to the identification of candidate CNVR, sCNVR were also identified on the potential CNVR identified using either the PennCNV or the QuantiSNP calling platforms. Furthermore, paternal and maternal transmission rates and inheritance rate were computed on the sCNVR identified on the potential CNVR identified using either the PennCNV or the QuantiSNP software.

3. Results

PennCNV allowed to identify 5,450 CNV on the 492 individuals typed (Table S1). QuantiSNP mapped 2,558 CNV with a Log Bayes Factor $>$ 10 (781 of them with Log Bayes Factor $>$ 30) on 457 pigs (Table S2).

3.1. Construction and segregation patterns of candidate CNV regions

Overlapping CNV across individuals allowed to construct 2,160 potential CNVR (29 % identified in one individual only) using PennCNV results (Table S3) and 934 potential CNVR (21 % identified in one individual only) within QuantiSNP results (Table S4). These potential CNVR comprised about 82.1 Mb and 161.5 Mb, respectively. Overlapping between the potential CNVR identified using the two calling platforms allowed to identify a total of 344 candidate CNVR on the 18 porcine autosomes (Table S5; Table 1). Candidate CNVR covered about 106.8 Mb, with a mean length of 0.31 Mb. Twenty candidate CNVR were identified in one individual only, whereas 12 candidate CNVR were identified in$>$50 individuals.

The candidate CNVR identified in each individual (and their parents) of the pedigree used are listed in Table S6. According to the frequency of the CNVR in the pedigree (Table S5), 20 candidate CNVR were classified

Table 1
Description of the CNVR identified in Gochu Asturcelta pig per candidate CNVR class.

	CNVR class			
	Singleton	Recurrent	Segregating	Whole
	de novo	*de novo*	CNVR	dataset
Number of CNVR	20	255	69	344
Number of Mendelian mismatches				
assigned to the Father	0	578	2,146	2,724
assigned to the Mother	0	915	2,404	3,319
assigned to the Offspring	20	1,264	1,672	2,956
Mean CNVR length (kb)	132.8 [12.6;409.3][a]	287.9 [3.9; 3,683.6]	444.8 [10.5;3,077.1]	310.4
Mean Number of individuals	1 [1;1]	5 [1;41]	34 [3;187]	11
Mean Number of Families	1 [1;1]	4.5 [1;18]	20 [2;45]	7

[a] Minimum and maximum values are given in brackets.

as singleton *de novo*, 255 (74 % of the total) were classified as recurrent *de novo*, and the remaining 69 candidate CNVR were classified as segregating. Their distribution across porcine autosomes is illustrated in Fig. 1. sCNVR were called in 412 out of 478 parents-offspring trios and 60 out of 61 families. sCNVR covered about 30.7 Mb and were, on average, larger (444.8 kb) than rdnCNVR (287.9 kb; Table 1). Moreover, sCNVR were identified in 418 different parents-offspring trios and, on average, in 34 individuals belonging to 20 families (33 % of the total) whereas the mean number of individuals and families in which rdnCNVR were identified was 5 and 4.5 (7 % of the total number of families), respectively.

Fig. 2 shows two pedigrees illustrating the patterns of familial segregation in two typical examples of rdnCNVR (CNVR256) and sCNVR (CNVR41). Although CNVR256 could be identified in the offspring of a four generations pedigree including different fullsib families, analyses failed in identifying them in the parents. However, although CNVR41 segregated in the offspring of boar 20 and one of its granddaughters (sow 107), it could not be identified in the 107′s father (boar 66). Fig. 2 illustrates the fact that sCNVR could not be identified in all parental generations in a pedigree.

3.2. Inheritance and transmission rates

Both rdnCNVR and sCNVR showed a marked deviation of Mendelian expectations. Most mismatches (69 %) were identified on sCNVR (Table 1). Mendelian inheritance errors assessed for sCNVR were balanced across members of the parents-offspring trio, therefore suggesting that some systematic causes, such as the sex of the parents or pedigree depth, would not affect the failure in identifying the presence of a sCNVR in an individual. Although rdnCNVR were identified in the offspring only, up to 54 % of the mismatches identified for rdnCNVR were assigned to one of the reproductive individuals (Table 1).

A full description of the computations carried out to estimate paternal and maternal transmission rates and inheritance rate are given in Supplementary Table S7 and summarized per both parents-offspring trio and family in Table 2. Both at the trio and the family level, mean values computed were always very low suggesting a strong deviation of Mendelian proportions. A particular case can illustrate this fact (Supplementary Table S7): individual 883 and its father (486) carried 6 different sCNVR having an ideal paternal transmission rate of 50 %; however, pig 883 shared two sCNVR with its mother only (individual 477; maternal transmission rate of 13.3 %). Altogether, the inheritance rate computed for individual 883 was 75 %. Although this was the highest inheritance rate in our data set, it departed from the ideal inheritance rate of 100 %. Average paternal and maternal transmission rates at both the family and the trio levels ranged from 13.4 % to 11.0 %

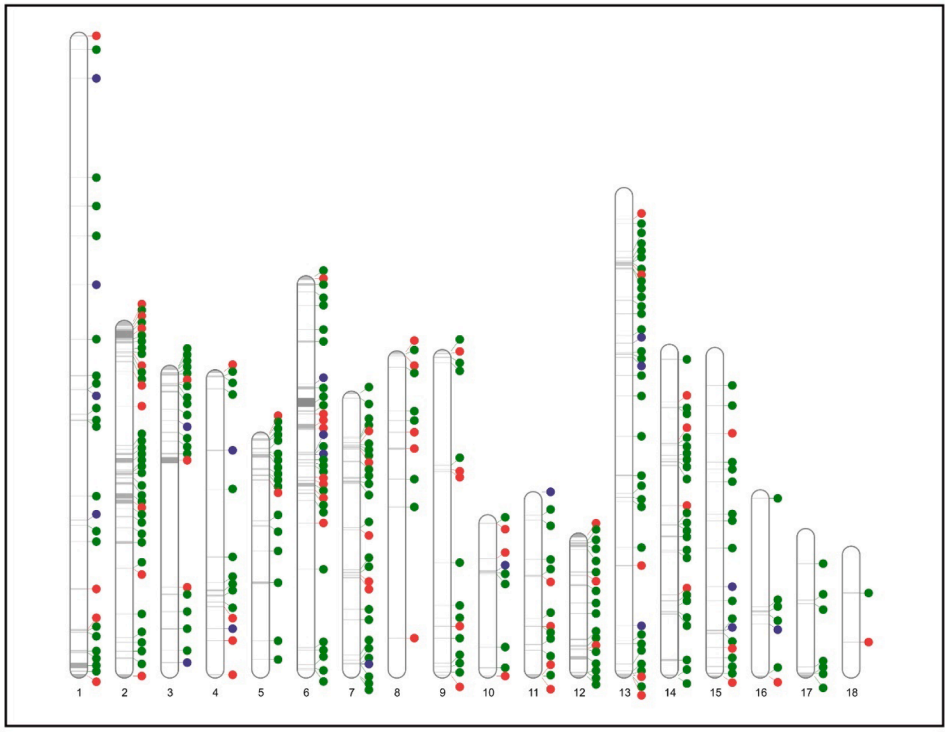

Fig. 1. Ideogram illustrating, per porcine autosome, the distribution of the 20 singleton *de novo* (in blue), the 256 recurrent *de novo* (in green) candidate CNVR, and the 69 candidate segregating (in red) CNVR identified in Gochu Asturcelta pig. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

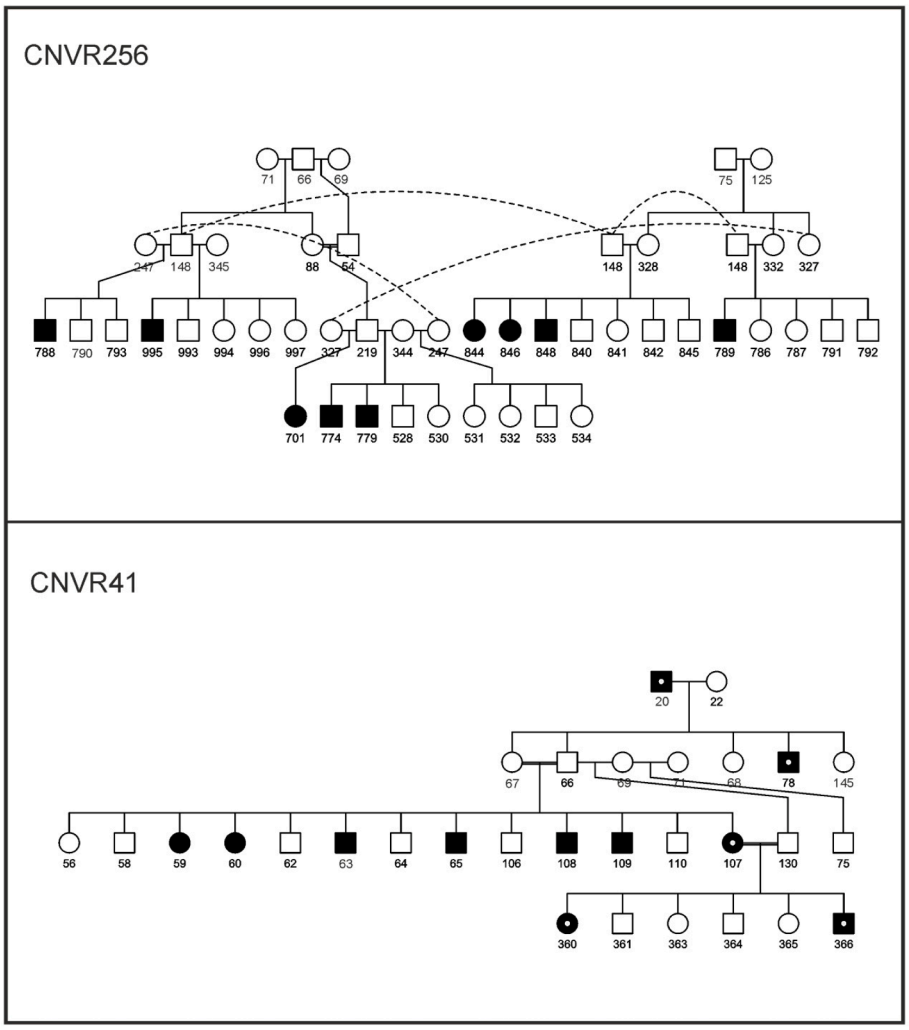

Fig. 2. Two pedigrees illustrating the variation of a recurrent CNVR (CNVR256) and a segregating CNVR (CNVR41). Numbers below squares (males) and circles (females) correspond to the identifications of the individuals. Filled squares and circles indicate the identification of a CNVR in a given individual. Open dots have been included in filled squares or circles to facilitate the identification of segregation events.

and from 12.4 % to 13.4 %, respectively. Paternal and maternal transmission rates were zero for 108 and 76 individuals in the pedigree meaning that they did not share sCNVR with either their fathers or their mothers (Supplementary Table S7). Mean inheritance rate was below 21 % and very similar at both the trio and the family levels.

3.3. sCNVR identification on potential CNVR

Segregation patterns of the potential CNVR identified using either the PennCNV or the QuantiSNP software (Supplementary Tables S8 and S9) allowed to classify 35 (out of 2,160) and 34 (out of 934) potential CNVR as sCNVR for the programs PennCNV and QuantiSNP, respectively (see also Supplementary Tables S3 and S4). Potential sCNVR identified using PennCNV covered about 7.6 Mb of the porcine genome and involved 269 parent-offspring trios. Potential sCNVR identified using QuantiSNP covered roughly 19.7 Mb and involved 272 parents-offspring trios (including the 269 trios identified using PennCNV). Within porcine chromosome, all potential sCNVR identified using PennCNV overlapped with others identified using QuantiSNP (Supplementary Table S10). Furthermore, the 35 and the 34 potential sCNVR identified using the programs PennCNV and QuantiSNP, respectively, could be merged into 33 consensus potential sCNVR covering 20.1 Mb (Supplementary Table S10).

Table 2 gives mean values for the transmission and inheritance rates computed for the potential sCNVR identified within calling platform. Full descriptions of the computations carried out are given in Supplementary Tables S11 and S12. Consistently with the similar sets of parents-offspring trios involved, the mean values obtained were very close no matter the calling platform considered (either PennCNV or QuantiSNP). The paternal and maternal transmission rates were about threefold higher than those computed on the candidate sCNVR identified using the combined results of the two software used. Furthermore, mean inheritance rates were from 164 % to 203 % higher than those computed on the candidate sCNVR identified using the combined results.

4. Discussion

Several technological issues including genotyping platforms, noisy signals, calling algorithms and quality control criteria, affect CNVR identification (Winchester et al., 2009; Zheng et al., 2012; Fernández et al., 2014; Keel et al., 2019). Moreover, CNV are assumed to mirror the particular breeding histories of the populations studied (Chen et al., 2012; Xie et al., 2016; Bovo et al., 2021). Altogether, such factors cause a significant variation between CNVR identified among populations with little CNVR overlap among different reports, which is usually assumed to be lower than 30 % (Xie et al., 2016; Keel et al., 2019; Qiu et al., 2021; Panda et al., 2022).

Table 2
Mean and standard deviation (in brackets) of Paternal transmission rate, Maternal transmission rate and Inheritance rate per trio and family in which segregating CNVR (sCNVR) were called. Results are given for candidate sCNVR and potential sCNVR identified using either the PennCNV or the QuantiSNP software. For descriptive purposes, average values provided in previous studies (Keel et al., 2019; Zheng et al., 2012) are provided.

	Candidate[1] sCNVR	Potential sCNVR		Literature estimates	
	Mean	PennCNV	QuantiSNP	Keel et al. (2019)	Zheng et al. (2012)
Paternal transmission rate				0.377	0.280
per trio	0.110 (0.103)	0.345 (0.352)	0.343 (0.352)		
per family	0.134 (0.112)	0.390 (0.302)	0.392 (0.296)		
Maternal transmission rate				0.414	0.280
per trio	0.134 (0.121)	0.394 (0.359)	0.396 (0.356)		
per family	0.124 (0.084)	0.344 (0.299)	0.349 (0.298)		
Inheritance rate				0.520	0.420
per trio	0.209 (0.209)	0.372 (0.402)	0.343 (0.386)		
per family	0.207 (0.207)	0.420 (0.375)	0.385 (0.351)		

[1] Candidate segregating CNVR constructed by intersecting potential CNVR identified using either the PennCNV or the QuantiSNP software.

Consequently, applying stringent criteria to select informative CNVR is a requisite for the presentation of results of general interest. Although the use of Real-Time PCR is the best way to validate CNV (Ramayo-Caldas et al., 2010; Xie et al., 2016; Ma et al., 2017; Keel et al., 2019), it is not always possible and, frequently, researchers consider CNV only when they are identified in separate analyses using different calling platforms (Rafter et al., 2018; Goyache et al., 2021, 2022). Here we propose the use of pedigree information to select CNVR that are present across generations. CNVR formed by overlapping CNV recalled by at least two different platforms but also showing Mendelian inheritance across generations are unlikely to be artifacts or false positives.

One can argue that the use of one calling platform only may be enough to identify sCNVR properly in the case of the availability of dense SNP array data such as in the current analysis. However, our results suggest that this alternative approach severely underestimates the number of sCNVR identified and, more importantly, limits the identification of target genomic areas putatively spanning CNV alterations of major interest. Merging the results of different CNV calling platforms allows to identify longer genomic areas potentially carrying CNV alterations and, therefore, increases the number of parent-offspring trios in which CNV segregation potentially occur.

In general, sCNVR are relatively large genomic regions identifiable across generations in relatively high frequency that can represent genetic variation of importance at the population level. In contrast, rdnCNVR can be more likely responsible for between-individuals genetic variation. This doesn't mean that rdnCNVR are of negligible genomic importance. Samarakoon et al. (2011) reported that rdnCNVR spanned many genes with a potentially important role in performance. However, if the goal is the ascertainment of the genetic background of a given population the use of rdnCNVR, even if their frequency is relatively high, may not be justified. CNVR studies not using genealogical data are likely to give upward biased estimates of both the number and the length of CNVRs (Zheng et al., 2012; Fernández et al., 2014).

4.1. Inheritance patterns of CNVR

The CNVR identified did not conform to strict Mendelian inheritance patterns. Indeed, both sdnCNVR and rdnCNVR completely departed from Mendelian expectations as they were called in neither parent. However, it is hardly assumable that *de novo* CNVR are due to somatic mutations. The sdnCNVR and the rdnCNVR summed 80 % of the total CNVR identified and somatic mutations could never be so frequent (Samarakoon et al., 2011). Moreover, somatic mutations (i.e. *de novo* CNVR) could affect either paternal and maternal transmission rates but not inheritance rates (Keel et al., 2019). Both sdnCNVR and rdnCNVR probably occur due to the same causes than sCNVR. The segregation patterns of the *de novo* and the segregating CNVR have marked similarities and classical pedigree-based analyses of Mendelian inheritance have problems in separating their behavior. The main difference between sCNVR and rdnCNVR is that the latter could never be identified in a parental generation. It would be expectable that rdnCNVR had Mendelian errors assignable to the offspring only in a similar way than in the case of the sdnCNVR. However, our results informed that Mendelian errors in rdnCNVR were assigned to different members of the parents-offspring trio. The presence of rdnCNVR in individuals acting as fathers or mothers would lead to this deviation from the expectations.

Furthermore, the computation of paternal and maternal transmission rates and inheritance rates clearly illustrates that sCNVR fit Mendelian segregation patterns at a broad scale only. The current transmission and inheritance rates are significantly lower than those previously reported in pig using 12 parents-offspring trios only (Keel et al., 2019) but also lower than those reported in humans (Zheng et al., 2012) using 752 complete family trios. The strategies used to identify sCNVR, the average CNVR length, SNP array density, and the size of the pedigree available may underly the estimates' differences among reports. In this respect, when sCNVR were identified using one calling platform only, paternal and maternal transmission rates were biased upwards when compared with those obtained with the candidate sCNVR approach and similar to the rates computed on the small Keel et al.'s data set (2019). However, they were significantly higher than those computed on the larger Zheng et al.'s (2012) pedigree (Table 2).

In any case, the Gochu Asturcelta pig pedigree used allowed us to check the consistency of Mendelian inheritance using multiple offspring of the same parental couple giving an accurate idea about the 'real' error rates in the transmission of CNVR from one generation to another. Although the deviation from Mendelian inheritance patterns of CNVR is more likely to be due to technical issues associated with SNP arrays calling rather than to the genomic nature of CNVR, the understanding of the error process is necessary if the goal of identifying CNV alterations is their use in association studies (Zheng et al., 2012; Keel et al., 2019).

4.2. General discussion

From a strict genetic marker perspective, CNVR would be affected by Allele-Drop-Out and (ADO) and Allele-Drop-In (ADI) effects. Both sdnCNVR and rdnCNVR can be classified as ADI loci (alleles that are additional to the parental genotypes). In turn, sCNVR could be mainly considered ADO loci (i.e. 'missing' alleles at a locus according to parental genotypes). The deviations of the expected transmission and inheritance rates in our pedigree imply that CNVR should be considered loci displaying either null or partial null alleles (alleles always generating missing data or not, respectively; see Arias et al., 2022, for a recent review on ADI and ADO effects and their relationship with CNV alterations).

Researchers are interested in using CNVR in the ascertainment of the relationship between genomic variation and performance due to the undoubtful importance of Copy Number alterations in phenotypic variation. However, from a genetic marker point of view, pedigree analysis shows that CNVR inheritance does not conform Mendelian rules. This illustrates that major difficulties exist in their satisfactory

identification. CNVR present to some extent in parents and offspring were the minority of the CNVR identified. However, such segregant CNVR are likely to represent 'true' genomic variation. Only an in-depth understanding of their patterns of variation, probably associated with calling errors, can contribute to the appropriate use of CNVR variation in such studies. The development of refined CNVR-calling strategies to identify 'reliable' CNVR will be an issue in genomic research.

Ethical approval.

SERIDA is adhered to the Ethical Committee in Research of the University of Oviedo (Spain) which ensures that all research with biological agents follows Good Laboratory Practices and European and Spanish regulations on biosecurity (Regulation of February 13th, 2014; BOPA no. 47, February 26th, 2014). Tissue and hair root samples used in this project were collected by veterinary practitioners working for the Gochu Asturcelta pig Breeders' Association (ACGA), with the permission and in presence of the owners. For this reason, permission from the Ethical Committee in Research of the University of Oviedo was not required. In all instances, ACGA veterinarians followed standard procedures and relevant national guidelines to ensure appropriate animal care.

CRediT authorship contribution statement

Katherine D. Arias: Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Juan Pablo Gutiérrez:** Investigation, Formal analysis, Writing – review & editing. **Iván Fernandez:** Formal analysis, Writing – review & editing. **Nuria A. Menéndez-Arias:** Data curation, Validation, Project administration. **Isabel Álvarez:** Conceptualization, Project administration, Data curation, Supervision, Validation. **Félix Goyache:** Conceptualization, Supervision, Methodology, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

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.

Data availability

Data will be made available on request.

Acknowledgements

This research was partially funded by AEI-FEDER, grant no. PID2019-103951RB/AEI/10.13039/501100011033. KDA was funded by AEI-ESF, grant no. PRE2020-092905. The genotyping service was carried out at CEGEN-PRB3-ISCIII; it is supported by grant PT17/0019, of the PE I + D + i 2013-2016, funded by ISCIII and ERDF. Authors are indebted to Dr Gonçalo Carvalho (Thermo Fisher Scientific) for his willingness and kind contribution to the execution of this project and to the members of the Gochu Asturcelta Breeders Association (ACGA; https://www.gochuasturcelta.org/) for their support and kind collaboration.

Data Availability Statement.

Pedigree data and all results obtained are provided as Supplementary Data (.xlsx file) containing the following Tables: Table S1: List of Copy Number Variations (CNV) identified using the software PennCNV; Table S2: List of Copy Number Variations (CNV) identified using the software QuantiSNP; Table S3: List of CNV Regions identified using the software PennCNV; Table S4: List of CNV Regions identified using the software QuantiSNP; Table S5: List of candidate Copy Number Variations Regions (CNVR) identified; Table S6: List of CNVR identified in each individual of the pedigree used; Table S7: Details on data used to compute (paternal and maternal) transmission rates and inheritance rate. The SNP array data used and analyzed during the current study, as a part of the K.D. Arias PhD Thesis, are available from the corresponding author on reasonable request.

ORCID.

Katherine D. Arias: https://orcid.org/0000-0001-6710-4208.
Juan Pablo Gutiérrez: https://orcid.org/0000-0003-3142-2356.
Isabel Álvarez: https://orcid.org/0000-0002-6052-8881.
Félix Goyache: https://orcid.org/0000-0002-6872-1045

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.147111.

References

Arias, K.D., Álvarez, I., Gutiérrez, J.P., Fernandez, I., Menéndez, J., Menéndez-Arias, N. A., Goyache, F., 2022. Understanding the occurrence of Mendelian Errors in SNP arrays data using a pedigree of Gochu Asturcelta pig: genomic alterations, family size and calling errors. Scient. Rep., in press. https://doi.org/10.1038/s41598-022-24340-0.

Bergamaschi, M., Maltecca, C., Fix, J., Schwab, C., Tiezzi, F., 2020. Genome-wide association study for carcass quality traits and growth in purebred and crossbred pigs. J. Anim. Sci. 98, skz360. https://doi.org/10.1093/jas/skz360.

Bovo, S., Schiavo, G., Ribani, A., Utzeri, V.J., Taurisano, V., Ballan, M., Muñoz, M., Alves, E., Araujo, J.P., Bozzi, R., Charneca, R., Di Palma, F., Djurkin Kušec, I., Etherington, G., Fernandez, A.I., García, F., García-Casco, J., Karolyi, D., Gallo, M., Martins, J.M., Mercat, M.-J., Núñez, Y., Quintanilla, R., Radović, Č., Razmaite, V., Riquet, J., Savić, R., Škrlep, M., Usai, G., Zimmer, C., Ovilo, C., Fontanesi, L., 2021. Describing variability in pig genes involved in coronavirus infections for a One Health perspective in conservation of animal genetic resources. Sci Rep 11, 3359. https://doi.org/10.1038/s41598-021-82956-0.

Carvalho, C.M.B., Lupski, J.R., 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 17, 224–238. https://doi.org/10.1038/nrg.2015.25.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. https://doi.org/10.1186/s13742-015-0047-8.

Chen, C., Qiao, R., Wei, R., Guo, Y., Ai, H., Ma, J., Ren, J., Huang, L., 2012. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. BMC Genomics 13, 733. https://doi.org/10.1186/1471-2164-13-733.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Research 35, 2013–2025. https://doi.org/10.1093/nar/gkm076.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., MacArthur, D.G., MacDonald, 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., Hurles, M.E., 2010. Origins and functional impact of copy number variation in the human genome. Nature 464 (7289), 704–712.

Ding, R., Zhuang, Z., Qiu, Y., Wang, X., Wu, J., Zhou, S., Ruan, D., Xu, C., Hong, L., Gu, T., Zheng, E., Cai, G., Huang, W., Wu, Z., Yang, J., 2022. A composite strategy of genome-wide association study and copy number variation analysis for carcass traits in a Duroc pig population. BMC Genomics 23, 590. https://doi.org/10.1186/s12864-022-08804-1.

Fernández, A.I., Barragán, C., Fernández, A., Rodríguez, M.C., Villanueva, B., 2014. Copy number variants in a highly inbred Iberian porcine strain. Anim Genet 45, 357–366. https://doi.org/10.1111/age.12137.

Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. Nat Rev Genet 7, 85–97. https://doi.org/10.1038/nrg1767.

Fontanesi, L., Beretti, F., Martelli, P.L., Colombo, M., Dall'Olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D., Russo, V., 2011. A first comparative map of copy number variations in the sheep genome. Genomics 97 (3), 158–165.

Goyache, F., Fernández, I., Tapsoba, A.S.R., Traoré, A., Menéndez-Arias, N.A., Álvarez, I., 2021. Functional characterization of Copy Number Variations regions in Djallonké sheep. J Anim Breed Genet 138, 600–612. https://doi.org/10.1111/jbg.12542.

Goyache, F., Pérez-Pardal, L., Fernández, I., Traoré, A., Menéndez-Arias, N.A., Arias, K. D., Álvarez, I., 2022. Identification and characterization of Copy Number Variations Regions in West African taurine cattle. Animals 12 (16), 2130. https://doi.org/10.3390/ani12162130.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J.,

Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J. K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J.V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., Schook, L.B., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491, 393–398. https://doi.org/10.1038/nature11622.

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., Chen, J., 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. 6, e251.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment: CERVUS LIKELIHOOD MODEL. Molecular Ecology 16, 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x.

Keel, B.N., Nonneman, D.J., Lindholm-Perry, A.K., Oliver, W.T., Rohrer, G.A., 2019. A Survey of Copy Number Variation in the Porcine Genome Detected From Whole-Genome Sequence. Front. Genet. 10, 737. https://doi.org/10.3389/fgene.2019.00737.

Ma, Q., Liu, X., Pan, J., Ma, L., Ma, Y., He, X., Zhao, Q., Pu, Y., Li, Y., Jiang, L., 2017. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. Sci Rep 7, 912. https://doi.org/10.1038/s41598-017-00847-9.

Menéndez, J., Alvarez, I., Fernandez, I., Goyache, F., 2016a. Genealogical analysis of the Gochu Asturcelta pig breed: insights for conservation. Czech J. Anim. Sci. 61, 140–149. https://doi.org/10.17221/8787-CJAS.

Menéndez, J., Álvarez, I., Fernandez, I., Menéndez-Arias, N.A., Goyache, F., 2016b. Assessing performance of single-sample molecular genetic methods to estimate effective population size: empirical evidence from the endangered Gochu Asturcelta pig breed. Ecology and Evolution 6, 4971–4980. https://doi.org/10.1002/ece3.2240.

Menéndez, J., Goyache, F., Beja-Pereira, A., Fernández, I., Menéndez-Arias, N.A., Godinho, R., Álvarez, I., 2016c. Genetic characterisation of the endangered Gochu Asturcelta pig breed using microsatellite and mitochondrial markers: Insights for the composition of the Iberian native pig stock. Livestock Science 187, 162–167. https://doi.org/10.1016/j.livsci.2016.03.013.

Mo, J., Lu, Y., Zhu, S., Feng, L., Qi, W., Chen, X., Xie, B., Chen, B., Lan, G., Liang, J., 2022. Genome-Wide Association Studies, Runs of Homozygosity Analysis, and Copy Number Variation Detection to Identify Reproduction-Related Genes in Bama Xiang Pigs. Frontiers in Veterinary Science 9, 892815. https://doi.org/10.3389/fvets.2022.892815.

Panda, S., Kumar, A., Gaur, G.K., Ahmad, S.F., Chauhan, A., Mehrotra, A., Dutt, T., 2022. Genome wide copy number variations using Porcine 60K SNP Beadchip in Landlly pigs. Animal Biotechnology 1–9. https://doi.org/10.1080/10495398.2022.2056047.

Qiu, Y., Ding, R., Zhuang, Z., Wu, J., Yang, M., Zhou, S., Ye, Y., Geng, Q., Xu, Z., Huang, S., Cai, G., Wu, Z., Yang, J., 2021. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs. BMC Genomics 22, 332. https://doi.org/10.1186/s12864-021-07654-7.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Rafter, P., Purfield, D.C., Berry, D.P., Parnell, A.C., Gormley, I.C., Kearney, J.F., Coffey, M.P., Carthy, T.R., 2018. Characterization of copy number variants in a large multibreed population of beef and dairy cattle using high-density single nucleotide polymorphism genotype data1. Journal of Animal Science 96, 4112–4124. https://doi.org/10.1093/jas/sky302.

Rafter, P., Gormley, I.C., Parnell, A.C., Kearney, J.F., Berry, D.P., 2020. Concordance rate between copy number variants detected using either high- or medium-density single nucleotide polymorphism genotype panels and the potential of imputing copy number variants from flanking high density single nucleotide polymorphism haplotypes in cattle. BMC Genomics 21, 205. https://doi.org/10.1186/s12864-020-6627-8.

Ramayo-Caldas, Y., Castelló, A., Pena, R.N., Alves, E., Mercadé, A., Souza, C.A., Fernández, A.I., Perez-Enciso, M., Folch, J.M., 2010. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics 11, 593. https://doi.org/10.1186/1471-2164-11-593.

Samarakoon, U., Gonzales, J.M., Patel, J.J., Tan, A., Checkley, L., Ferdig, M.T., 2011. The landscape of inherited and de novo copy number variants in a plasmodium falciparum genetic cross. BMC Genomics 12, 457. https://doi.org/10.1186/1471-2164-12-457.

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E., Feuk, L., 2007. Challenges and standards in integrating surveys of structural variation. Nat Genet 39, S7–S15. https://doi.org/10.1038/ng2093.

Wang, J., 2019. Pedigree reconstruction from poor quality genotype data. Heredity 122, 719–728. https://doi.org/10.1038/s41437-018-0178-7.

Wang, Z., Guo, Y., Liu, S., Meng, Q., 2021. Genome-Wide Assessment Characteristics of Genes Overlapping Copy Number Variation Regions in Duroc Purebred Population. Front Genet 12. https://doi.org/10.3389/fgene.2021.753748.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674. https://doi.org/10.1101/gr.6861907.

Wang, L., Xu, L., Liu, X., Zhang, T., Li, N., Hay, E.-H., Zhang, Y., Yan, H., Zhao, K., Liu, G. E., Zhang, L., Wang, L., 2015. Copy number variation-based genome wide association study reveals additional variants contributing to meat quality in Swine. Sci Rep 5, 12535. https://doi.org/10.1038/srep12535.

Wei, X., Shu, Z., Wang, L., Zhang, T., Zhang, L., Hou, X., Yan, H., Wang, L., 2022. Copy Number Variations Contribute to Intramuscular Fat Content Differences by Affecting the Expression of PELP1 Alternative Splices in Pigs. Animals 12, 1382. https://doi.org/10.3390/ani12111382.

Winchester, L., Yau, C., Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. Briefings in Functional Genomics 8, 353–366. https://doi.org/10.1093/bfgp/elp017.

Xie, J., Li, R., Li, S., Ran, X., Wang, J., Jiang, J., Zhao, P., 2016. Identification of Copy Number Variations in Xiang and Kele Pigs. PLoS ONE 11, e0148565. https://doi.org/10.1371/journal.pone.0148565.

Zheng, X., Shaffer, J.R., McHugh, C.P., Laurie, C.C., Feenstra, B., Melbye, M., Murray, J. C., Marazita, M.L., Feingold, E., 2012. Using Family Data as a Verification Standard to Evaluate Copy Number Variation Calling Strategies for Genetic Association Studies. Genet. Epidemiol. 36, 253–262. https://doi.org/10.1002/gepi.21618.