

# Whole genome resequencing of the Iranian native dogs and wolves to unravel variome during dog domestication

**Zeinab Amiri Ghanatsaman**

Shahid Bahonar University of Kerman

**Guo-Dong Wang**

Kunming Institute of Zoology Chinese Academy of Sciences

**Hojjat Asadollahpour Nanaei**

Shahid Bahonar University of Kerman

**Masood Asadi Fozi**

Shahid Bahonar University of Kerman

**Min-Sheng Peng**

Kunming Institute of Zoology Chinese Academy of Sciences

**Ali Esmailzadeh** (✉ [aliesmaili@uk.ac.ir](mailto:aliesmaili@uk.ac.ir))

Shahid Bahonar University of Kerman <https://orcid.org/0000-0003-0986-6639>

**Ya-Ping Zhang**

Kunming Institute of Zoology Chinese Academy of Sciences

---

## Research article

**Keywords:** Single nucleotide variant, Copy number variant, Structural variant, Fertile crescent

**Posted Date:** January 6th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.15926/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on March 4th, 2020. See the published version at <https://doi.org/10.1186/s12864-020-6619-8>.

# Abstract

Background Advances in genome technology have simplified a new comprehension of the genetic and historical processes crucial to rapid phenotypic evolution under domestication. To get new insight into the genetic basis of the dog domestication process, we conducted whole-genome sequence analysis of three wolves and three dogs from Iran which covers the eastern part of the Fertile Crescent located in Southwest Asia where the independent domestication of most of the plants and animals has been documented and also high haplotype sharing between wolves and dog breeds has been reported. Results Higher diversity was found within the wolf genome compared with the dog genome. A total of 12.45 million SNPs were detected in all individuals (10.45 and 7.82 million SNPs were identified for all the studied wolves and dogs, respectively) and a total of 3.49 million small Indels were detected in all individuals (3.11 and 2.24 million small Indels were identified for all the studied wolves and dogs, respectively). A total of 10,571 copy number variation regions (CNVRs) were detected across the 6 individual genomes, covering 154.65 Mb, or 6.41%, of the reference genome (canFam3.1). Further analysis showed that the distribution of deleterious variants in the dog genome is higher than the wolf genome. Also, annotation of genomic variations showed that the proportion of genomic variations in the intron and intergenic regions in the wolf genome is higher than that in the dog genome, while the proportion of the coding sequences and 3'-UTR in the dog genome is higher than that in the wolf genome. The genes related to the olfactory and immune systems were enriched in the set of the structural variants (SVs) identified in this work. Generally, genes engaged in digestion and metabolism and neurological process had an important role in the process of dog domestication. Conclusions Our results showed more deleterious mutations and coding sequence variants in the domestic dog genome than those in wolf genome. By providing the first Iranian dog and wolf variome map, our findings contribute to understanding the genetic architecture of the dog domestication.

## Background

The dog (*Canis familiaris*) was likely the earliest domesticated animal and the only one humans' friend in the past [20, 67]. Genetic studies and archaeology findings show that dogs have a common ancestor with the gray wolf (*Canis lupus*) [21, 64, 69]. In the Southwest Asia, major-scale farming extended within the so-named Fertile Crescent (FC) where the independent domestication of plants and animals had led to shifting from gathering and hunting to sedentary farming following expansion of the first complex societies [22, 74]. Mostly, agricultural *developments* happened in the eastern horn of FC especially Elam (covering a region of southern Iraq and Iran), joining Mesopotamia and Iranian plateau [4]. Dogs are often drawn in art at *ancient* times in several parts of Southwest Asia [20, 51]. Therefore, one of the most theories about the geographical origin of the domestic dog has been that they originated in Southwest Asia, presumably in the FC [20]. In addition, the Middle East has been proposed as the beginning of domestic dog for great haplotype sharing between the Middle Eastern wolves and dog breeds [65] although this hypothesis has been questioned due to dog-wolf introgression in the Middle East [6, 7, 29] rather than an indication of Middle Eastern origins. The dog is a notable instance of variation under

domestication, however the evolutionary processes underlying the genesis of this diversity are weakly realized.

In recent years, advance in high-capacity genome examining *techniques*, especially whole genome sequencing, SNP genotyping array and comparative genomic hybridization (CGH) arrays have authorized the recognition of genome-wide structural variants. The *array methods* have limited resolution and low sensitivity because their performance is strongly depending on the marker frequency and particularly constructed non polymorphic markers [5, 42, 53], thus they cannot detect small copy number variations (CNVs) (< 10 kb) and cannot precisely identify boundaries of CNVs [73]. Next-generation sequencing methods provide a high-accuracy base-by-base vision of the genome and capture all variants by different size that might otherwise be missed, and all these are important and have significant effects on an extensive range of traits in domesticated animals. For examples: Fear and anxiety will be increased by increasing of expression of GRIK2 gene in domesticated species than their wild species including rabbit, guinea pig, dog and chicken [39], *MC1R* makes coat color variants in pig [27] and mutation in *TSHR* influences seasonal reproduction in chicken [56]. CNVs can also have major phenotypic changes in animals. For example, previous studies have found *CNVs involved in traits* related to pea-comb and late feathering in chicken [70, 26], polledness in goat [49], hair ridge in dog [32], health and production in cattle [12] and adaptability in dog [9, 68]. In this work for the *first time*, we sequenced the whole genomes of 6 canids from the same geographical range (three Iranian wolves and three Iranian dogs) with an average depth of 16X. . One of the sequenced dogs, Qahderijani, is a mastiff ecotype dog originating in Qahderijan, Iran, which is located in FC belt (surrounding areas of FC). Other two sequenced dogs were sampled from the Saluki breed, a hunting dog breed, belonged to the FC region, and also is considered as the long marathon runner of the canine, as its incredible endurance enables it to run for many miles.

In our analysis of the Iranian dog and wolf sequences, we applied assembly version canFam3.1 as a reference sequence [40]. SNPs and small Indels were detected in this research as differences between the recently *gained* genome sequences and reference sequence, and detected 12.45 million SNPs and 3.48 million small Indels. Valid algorithms were applied to analyze 6 genomes to get highly reliable CNVs and SVs. The potentially breed-specific CNVRs were defined and the functional relation of the SV and CNVR-covering genes was further evaluated with GO enrichment. Genome-wide analysis indicates more genetic diversity in the dog genome than that in the wolf genome. The annotation of the results from different types of genomic variations proposed that increasing the percentage of genomic variations in the coding and the regulatory regions of genes than that in intron and intergenic regions during domestication is the substantial contributor to the currently detected difference between dog and wolf. Also, comparison of the genomic variations effects between dog and wolf genomes showed that generally genes engaged in neurological and digestion and metabolism *processes* had a considerable effect in the progress of dog domestication. The CNVs reported in this research are enriched for olfactory and immune system genes.

## Results

### Sequencing output

Illumina Paired-end sequencing was performed for all 6 individuals (Additional file 1: Table S1 and Figure S1). After filtering, the range of total high-quality sequence data was from 42.1 Gb (Sample ID: #GW1) to 51 Gb (#DogQI), and the coverage varied from 14.51 (#GW1) to 17.15 (#GW2) (Additional file 1: Table S2). Relatively high mean sequencing depth can increase the accuracy of CNV calling through read depth method [1], and using the paired-end DNA sequencing reads together with the relatively long read length (by a uniform length of 125) will be useful to identify Indels [39, 54]. Here, to increase confidence of base calls and *accuracy of detecting genomic variations*, sequencing was done with relatively high mean depth for all samples (Additional file 1: Table S1). We also used uniform depth of coverage across individual genomes for increasing reliability of CNV calling (Additional file 1: Table S2).

## SNP detection and annotation

SNPs were detected through *aligning* sequences to the reference genome. A total of 12.45 million SNPs were detected in all individuals (10.45 and 7.82 million SNPs were identified for all studied wolves and dogs, respectively) (Additional file 1: Table S3 and Figure S2).

We obtained the ratio of *transitions to transversions* ( $Ti/Tv$ ) for the SNPs and the number of heterozygous and homozygous in the SNPs across the 6 individual genomes. The number of heterozygous SNPs was higher than the number of homozygous SNPs in 6 individuals. The  $Ti/Tv$  ratio in SNPs varied from 1.99 (#DogQI) to 2.07 (#GW3) (Additional file 1: Table S4). Figure 1 illustrates the proportion of SNPs present in each genomic regions, including intergenic, introns, exon, transcript, upstream, downstream, 3' untranslated regions (3'-UTR) and 5' untranslated regions (5'-UTR). Our results indicate that most of the SNPs are located in the intergenic (53.57%) and intron (31.99%) regions (Additional file 1: Table S5). The total number of synonymous SNPs (silent SNPs, 68899) were more than the total number of non-synonymous SNPs (nonsense and missense SNPs, 46789) (Additional file 1: Table S6). Also, results of SNPs annotation showed that the proportion of SNPs in intron (31.85 vs 31.81) and intergenic (53.92 vs 53.52) regions in wolf genome was higher while the percentage of the SNPs in exon regions (0.81 vs 0.84) and 3'-UTR (0.43 vs 0.46) in wolf genome was lower than that in dog genome.

## Small Indels detection, annotation and gene ontology

Indels were detected using *aligning* sequences to the reference genome. The number of Indels was calculated for all individuals (Additional file 1: Table S3). A total of 3.48 million Indels were detected across the 6 individual genomes, 2.24 million and 3.11 million of which were for 3 dogs and 3 wolves, respectively. We calculated the number of heterozygous and homozygous Indels across the 6 individual genomes (Additional file 1: Table S4). The proportion of heterozygous Indels (52.12) was higher than the proportion of homozygous Indels (47.59) in the 6 individuals. The total number of small insertions across the 6 individual genomes was 1.58 million and the total number of small deletions across the 6 individual genomes was 1.9 million (Additional file 1: Table S7). We drew the *Indel length* histogram for 3 dogs (Additional file 1: Figure S3), 3 wolves (Additional file 1: Figure S4) and *across six individual genomes* (Additional file 1: Figure S5). The results showed that the *Indels* of 1 bp in *length* across the 6 individual genomes had the highest frequency and *the deletions of the same size* were more frequent than the

insertions. Annotation of the results from small Indels showed that most of the Indels are located in intergenic (53.79%) and intron regions (34.45%) (Additional file 1: Table S8). Of the total number of small Indels, 53.79, 34.778, 0.25, 0.002, 5.54, 4.95, 0.46, and 0.14% were located within the intergenic, introns, exon, transcript, upstream, downstream, 3'-UTR and 5'-UTR regions, respectively. The percentage of small Indels that are located in upstream, 5'-UTR, 3'-UTR, exon and transcript regions across 3 dog genomes was higher than that across 3 wolf genomes, but the percentage of Indels that are located in downstream, introns and intergenic regions across 3 wolf genomes was higher than that across 3 dog genomes. We obtained 21,104 genes from ensemble through the annotation of a total of 3.48 million small Indels. After, we carried out gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for these genes (Additional file 1: Table S9). Gene Ontology (GO) analysis categorized genes related to small Indels in the *three main* classes (molecular function, biological process and cellular component) (Additional file 1: Table S9). The KEGG pathway analysis showed that two pathways related to cancer and Melanoma (usually but not always, a cancer of the skin) were enriched among the small Indels in both dog and wolf.

### **SVs detection, annotation and gene ontology**

We obtained genomic SVs including insertions, deletions, tandem duplication, translocations (inter and *intra* chromosomal) and inversions from three dogs and three wolves (Additional file 1: Table S10; Additional file 2: Table S16, Additional file 3: Table S17 and Additional file 4: Table S18). To obtain potential functional roles related to the different types of SVs, all genes that were completely or partially overlapped with genomic regions including, Indels (insertion and deletion), inventions and complex SVs (inter and *intra* chromosomal translocations) were retrieved from Ensemble (Additional file 1: Table S11). Annotation of results from SVs showed that in general the percentage of coding sequences variants in dog genome is higher than that in wolf genome (Additional file 1: Figures S6-S13). Results of gene set enrichment analysis showed three enriched categories related to “covering molecular function”, “biological process” and “cellular component” (Additional file 1: Table S12). The most conspicuous cluster terms related to dog and wolf individuals were “cellular carbohydrate metabolic process ( $P$ -value, 0.04)” and “nervous system development ( $P$ -value, 0.03)”, *respectively*. We also identified some candidate genes associated with olfactory and immune systems (Additional file 1: Table S12).

### **CNV detection**

We obtained putative CNVs for 6 individuals using CNVnator program and the mean number of CNVs per individual was 4143.83 ranging from 2871 to 5437 (Additional file 1: Table S13). For all of the autosomal CNVs categorized as gain, the mean copy number value of six individuals was 3.57 and the maximum copy number assessment was 174.472 on chromosome 7 (chr7) of wolf. The results showed that the number of gains in the three dog genomes was higher than those in the three wolf genomes (Additional file 1: Table S13). A total of 10571 CNVRs were obtained from overlapping of all CNVs across the 6 individuals (Additional file 5: Table S19), including 1-38 and X chromosomes, ranging in size from 1.05 kb to 3433.35 kb with an average of 14.63 kb and a median of 7.05 kb, covering 154.65 Mb, or 6.41%, of the

assayed canFam3.1 genome (Table 1). CNVRs were divided into three groups, including 6400 loss, 3916 gain and 255 both (gain and loss) events (Additional file 5: Table S19). Deletion:duplication ratio in the total CNVRs was 1.96. Among all CNVRs, 6,105 (57.75%) were found in a single individuals (singleton), 1,522 (14.39%) shared in two individuals, and 2,944 (27.84%) shared in at least three individuals (Figure 2B). A number of 6702 (63.4%) CNVR events were less than 10 Kb while 494 (4.7%) of the CNVRs were longer than 50 kb in size (Table 1 and Figure 2A). The highest and lowest numbers of CNVRs belonged to chromosomes 18 and 35, respectively (Additional file 1: Figure S14 and Additional file 6: Table S20).

## **CNV annotation and gene ontology analysis**

*The annotation of results from CNVs* showed that the percentage of CNVs in coding sequences (14% vs. 6%) and 3'-UTR (6% vs 0) in the dog genome was greatly higher than that in the wolf genome, but the percentage of CNVs in the intergenic regions (22% vs. 14%) in wolf genome was greatly higher than that in the dog genome (Additional file1: Figures S15 and S16). To achieve potential functional roles related to the putative CNVs, all genes that completely or partially overlapped with these CNVs were detected from Ensemble. A total of 8595 genes were retrieved, including 6703 of the CNVs. Results of gene ontology (Go) analysis showed that some general genes associated with olfactory and immune systems are enriched among the CNV gains in dog and wolf (Additional file 1: Table S14). All the terms related to olfactory system are over-represented ( $P$ -value < 0.01) in the wolf compared with those in the dog (Additional file 1: Table S14). The term "Starch and sucrose metabolism" is enriched in the dog CNV gains. Also, our result showed some categories including "cardiac conduction ( $P$ -value <0.03)", "actin filament ( $P$ -value, 0.037)", "muscle filament sliding ( $P$ -value, 0.02)", "ATP binding ( $P$ -value, 3.46E-04)" and "calcium ion binding ( $P$ -value, 0.001)" are enriched among the CNV gains in the Saluki breed (Additional file 1: Table S14).

## **Comparison with previous dog CNV studies**

To compare the identified CNVRs in this work with those of the published studies, all previous CNVR coordinates from canFam2 were migrated to canFam3 using the UCSC leftover program. In our results, 4454 CNVRs (42.1%) were overlapped by four previous studies, and the remaining 6117 (57.865%) were considered as novel CNVRs (Additional file 1: Table S15 and Additional file 7: Table S21).

## **Visualization of Structural Genomic Variation**

*For visualizing similarities and differences of positional relationships and genome structure between dog and wolf genomes, we drew maps of circular genomes for dog and wolf (Figure 3).*

## **Identification of Deleterious Mutations**

Population genetic processes due to reduced population size, such as inbreeding depression and bottlenecks, have a profound impact on the genetic makeup of a species including levels of deleterious variation [15, 34, 36]. Our results indicated that the proportion of deleterious mutations varied between

wolf and doge chromosomes (Figure 4), and more deleterious mutations are in dog genome, compared with their wild ancestor.

## Discussion

Analysis of high-quality next-generation sequencing data clearly showed the difference of the distribution and impact of the genomic variations between dog and wolf. The ratio of *transition to transversion* (ti/tv) is an indicator of false positive ratio for SNP calling [10, 31], the ratios calculated for all individuals (1.99 to 2.07) (Supplementary Table S4) indicate the precision of the identification of single-nucleotide mutations in our research. In addition, the results of this research similar to previous studies [58] showed that most of the SNPs belong to within introns or between genes and the number of synonymous SNPs was higher than non-synonymous SNPs. The majority of small Indels (95.89 % in dog and 95.64.% in wolf) were less than 10 bp in *length*, similar results were reported in a study of Indels in chicken [72].

We detected 10571 CNVRs with a mean of 4143.83 CNVs per sample in the canine genome. Similar to those reported in dog and wolf [18, 44, 47, 48], human [23, 55] and mouse [30], loss events were more prevalent than gain events in our results (1.63 fold). This may mirror the greater relative hardness of identifying gains because of the smaller relative alteration in copy number (3:2 versus 2:1). Loss events included shorter genomic sequences than gains on median (4.499 kb vs. 11.699 kb), mean (7.387625 kb vs. 21.38724 kb) and total (47.280800 Mb vs. 83.752434 Mb) (Table 1). This could show that duplications are less likely to be cleaned by purifying selection [5]. A total of 4466 (42.25%) CNVRs are seen in at least two individuals and 6105 (57.75%) CNVRs present in only one individual. Percentage of singletons obtained in this work is in agreement with that reported in previous studies related to identification of CNV in human [55], dog [47] and chicken [73]. We realized that the CNVRs were non-randomly distributed across the canid genome (Table S20). Chromosome 32, for example, has 2.03% of sequences displaying copy number variable, whereas chromosome 18 has 42.79% of sequences with copy number variation (Supplementary Table S20). In general, the chromosomes 9 (13.03%), 26 (14.97%) and 18 (42.78%) showed a high percentage of the CNVRs.

The terms “sensory perception of smell”, “detection of chemical stimulus” and “Olfactory transduction” are involved in sensory perception and were enriched among the CNV gains in dog and wolf and all of them were over-represented among the CNV gains in wolf ( $P < 0.01$ ). Both wolf and dog develop olfaction, audition and vision by 2 weeks, 4 weeks and 6 weeks of age on average, respectively [41]. Wolf pups start to investigate their environment at 2 weeks of age while they are blind and deaf, and must depend mainly on sense of smell, while dog pups start to investigate their environment at 4 weeks of age [41]. In a previous study, the fraction of olfactory receptor pseudogenes in dog and wolf was 17.78 and 12.08%, respectively, however, difference between these values in dog and wolf was not significant [76]. In one another study, no difference in the olfactory capacity of the dog breeds, which have been chosen for their smelling ability and the hand-bred grey wolves, was reported [50]. However, our results suggest an importance rule for olfaction during dog domestication. Six of the GO terms belonged to CNV gains in this study are also similar to those that were presented using aCGH method in dog [11].

Gene ontology term enrichment analysis showed that gene families involved in sense of smell and immune system commonly rapid growing for their importance in the organism terms answering to fast changes in the environment and fitness, also they have been frequently identified in CNV regions of multiple mammalian *genomes* [2, 71, 78]. Go terms related to *heart and muscle functions* such as “cardiac conduction” and “actin filament” were only enriched in the CNV gains in Saluki dog. These results can be expected because Saluki is a hunting dog breed which is considered as the long marathon runner of the canine world and its incredible endurance enables the dog to run for many miles. It has been presented that endurance exercise training makes a number of cardiac adaptations to marathon running [59]. Recently, Di Gerlando et al (2019) reported specific CNVs related to hunting in the BRA dog.

A fundamental number of the CNVs (~ 42%) from this work are compatible with those identified in previous studies in dogs and wolves. In addition, a substantial number of the Go terms that are enriched among the CNV in this study are concordant with the Go terms (~ 31%) related to previous study of copy number variations in dogs and wolves [11]. This compatibility with the previous studies, in conjunction with the identification of the CNVs specific to the Saluki breed, lends more support to the CNVs identified in this work. The difference between the CNVs detected in the study herein and those described previously can be related to the particular breeds studied and also the difference between the methods used. Generally, the CNVs that are identified by read-depth analysis are on average much smaller than those detected by aCGH.

The total numbers of (10.45 million vs 7.82 million) SNPs, (3.11 million vs 2.24 million) Indels, deletions (18628 vs 13059), inversions (401 vs 334), inter (520 vs 706) and intra (421 vs 359) chromosomal translocations regions were higher in the wolf genome than those in the dog genome, while the total number of CNVs located at gain (2277 vs 521) and insertions (352 vs 311) regions in the dog genome were higher than those in the wolf genome. It has been accepted that gene duplication through yielding material for selection, mutation and drift can be a chief source of recentness in evolution [77].

Our results from the genome analysis of dog and wolf revealed reduction of *genomic diversity* during dog domestication. A population bottleneck occurred in the wolves thousand years ago after a population expansion occurred by human through artificial selection on specific traits leading to different breeds of dogs [3, 29]. The effective population size in wolves is higher than that in dogs so higher genome diversity in wolves is expected compared to dogs [3, 29]. Our results from two components of genetic variation sources including SVs and CNVs confirmed that the novel adaptations permitted the primal ancestors of recent dogs to live on a diet with high starch compared to the carnivorous diet of wolves, which is an essential step in the primal domestication of dogs [8, 9, 29, 60, 69]. The term “nervous system development” was enriched among SVs in wolf and is indicative of reducing aggression in the first steps of animal domestication [66]. “Nervous system development” is defined as a process that particular result is the development of nervous tissue over time from its production to its developed shape.

Annotation of results from different types of genomic variations showed that in general the percentage of genomic variations in intron and intergenic regions in wolf genome is higher than that in dog genome

while in *coding sequences* and 3'-UTR in dog genome is higher than that in wolf genome.

It seems that domestication and its related processes such as relaxed selection have an important role in increasing the percentage of genomic variation in the coding and the regulatory *sequences* of dog genome. The relaxation of selection likely increases the functional genetic diversity throughout the genome of the dog and this diversity includes both the genes and the elements involved in gene expression [13, 24]. Previous studies have shown that the extensive selection for phenotypic and behavioral traits, have resulted in morphological diversity within the domestic dog [63]. Also, it has been suggested that after domestication some subtle sources of genomic diversification such as changes in the interactions among genes products and in the timing of gene expression may have influenced the diversity of the forms observed in the domestic dog [16, 44]. More deleterious mutations were detected in dog genome, compared with their wild ancestor. Our results similar to previous studies [14, 16] confirmed that domestication has increased deleterious mutations in domesticated animals than those in their wild ancestors.

It should be noted that mammalian genomes possess a complex structure with a **diverseness** of repetitive elements that complicates extensive genome-wide analyses [62]. To better acknowledge this result, there is still the need for using mate pair sequences or merging long-insert mate pair and short-insert paired-end sequences to analyze the dog and wolf genomes and elucidate the difference of the distribution and impact of the genomic variations between dog and wolf *during dog domestication*.

## Conclusions

*We resequenced* the whole genomes of 6 canids from the Middle East for the first time and we compared the effect and *distribution of the* genomic variations between dog and wolf genomes. Whole genome resequencing of three dogs and three wolves detected 7.82 *million* and 10.45 *million* SNPs, respectively. Numerous putatively CNVs were identified through an analysis of read depth difference. Furthermore, we have identified SVs which could be useful for marker based population genetic investigation. Downstream analysis of the identified SVs and CNVs revealed the changes between dog and wolf genome during dog domestication.

## Methods

### Sampling and sequencing

The sources of the animals used in this study were as follow: one wolf was sampled from Kerman zoo, South of Iran, two wolves were used from Eram Park Zoo, Tehran, Iran; two Saluki dogs were sampled from private farms in Kurdistan province, west of Iran and one Qahderijani dog was used from a private farm in Isfahan, Iran. We collected blood samples from three captive Iranian wolves (Additional file 1: Figure S17) and three Iranian dogs including a Qahderijani (Additional file1: Figure S18) and two Saluki dogs (Additional file1: Figure S19) with the consent of the owners. Sampling locations are reported in

Table 2. DNA was prepared with phenol/chloroform technique. Pair-end sequence data for all 6 individuals were generated using Hiseq 2500 Illumina.

### **Quality control and mapping**

The quality of the reads was evaluated with FastQC program and outputs of quality control showed that all reads had high-quality and were without adaptor contamination. *Aligning data against* the genome assembly canfam3.1 was done with Burrows-Wheeler Aligner program (BWA) [37]. The SAMtools [38] was applied to change the Sequence Alignment MAP (SAM) files to the Binary Alignment MAP (BAM) files and sort and index them. All of the .bam files were cleaned from PCR duplicates with Picard program. The accuracy of mapping was evaluated using of two criteria including percentage of *aligning against* the reference genome and mean depth with SAMtools

### **Short Indel and SNP detection**

*Genome Analysis Toolkit* (GATK) program [43] was applied to detect SNPs and Indels. All .bam files were preprocessed in two steps; i) local realignment around Indels was done using known Indels, ii) recalibrating base quality scores was done to increase quality score for each base. The purified data belonged to the same individual were jointly used to create genome variant call format (gVCF) files by GATK HaplotypeCaller, followed by merging the gVCF files belonged to all individuals employing the GATK GenotypeGVCFs. Finally, SNPs and Indels were separated from the resulted raw variant file and filtered using GATK Select Variants and GATK Variant Filtration, respectively.

### **SVs detection**

SVs including deletions, inversions, translocations (inter and intra chromosomal) and insertions were detected by using both of BreakDancer-1.1 [17] and DELLY [54] software. SVs were filtered using BreakDancer with read coverage  $\geq 10$ , the score  $\geq 80$  and size  $\geq 50$  bp.

### **SNP and Indel annotation**

Functional consequence analysis of SNPs and short INDELS were predicted using SnpEff 4.0e [19]. The transition to transversion and homozygous to heterozygous ratios for single nucleotide variants were calculated with SnpSift [57].

### **Prediction of deleterious mutations**

To predict the deleterious mutations within all individual canid genomes, we used the SIFT (Sorting Intolerant from Tolerant) algorithm [46]. If this normalized value is less than 0.05, the substitution is predicted to be deleterious, and those greater than or equal to 0.05 are predicted to be tolerated.

### **CNV Calling**

Putative CNVs on the 38 Canine autosomes and X chromosome were detected based on read depth method using CNVnator [1]. We run CNVnator with a bin size of 150 bp and GC correction (default) for our data. Filtering putative CNVs was done using different criteria including size > 1 kb, P-value < 0.01 and q0 (zero mapping quality) < 0.5. We removed all un-localized chromosome CNVs (chrUn). Putative CNVRs were obtained using Bedtools software [52] from overlapping of 1bp or greater CNVs on chromosomes 1-38 and X chromosome in 6 individuals as suggested before [55]. All CNVRs were categorized into three classes, e.g., “Loss” (including deletion), “Gain” (including duplication) and “Both” (including both deletion and duplication). To compare the putative CNVRs from this study with the CNVRs reported in the previous studies, all coordinates related to CNVRs of the previous studies were converted from canFam2.0 to canFam3.1 using the lift over tools (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).

### **Gene contents and gene ontology analysis**

Dog gene IDs that covered small Indels, SVs and CNVRs were retrieved from Ensemble annotation [28]. All dog gene IDs were changed to human gene IDs. Gene orthologous connection between dog and human was obtained from Ensembl. Gene ontology (GO) was done using DAVID program [33].

### **Visualization of structural genomic variation**

We drew the physical distribution of CNVRs on chromosomes 1-38 and X chromosomes using vcstools [35]. RCircos package [75] was used to draw *circular genetic maps* for visualizing similarities and differences of positional relationships and genome structure between dog and wolf.

## **List Of Abbreviations**

BAM: Binary Alignment MAP; BWA: Burrows-Wheeler Aligner program; chrUn: Un-localized chromosome; CGH: Comparative genomic hybridization; CNVRs: *Copy number variation* regions; CNVs: Copy number variations; FC: Fertile crescent; GATK: *Genome Analysis Toolkit*; GO: Gene ontology; gVCF: Genomic variant call format; GW: Gray wolf; Indels: Insertion and deletion; KEGG: Kyoto Encyclopedia of Genes and Genomes; QI: Qahderijani; SAM: Sequence Alignment MAP; SI: Saluki; SVs: Structural variants; 3'-UTR: 3' untranslated region; 5'-UTR: *5'untranslated region*

## **Declarations**

### **Ethics approval and consent to participate**

This study had Institutional Animal Care and Use Committee (Kunming Institute of Zoology, approval ID: SYDW-2013021) approval. We collected peripheral blood samples from 3 Iranian dogs with the consent of owners and 3 gray wolves after obtaining authorization for research from the Department of Environmental Protection in Iran (No. 93/34089, dated 14 October 2014).

### **Consent for publication**

Not applicable.

## Availability of data and materials

**Data deposition:** Raw sequence reads data have been deposited in the Genome Sequence Archive (<http://gsa.big.ac.cn/>) under accession CRA0001324 for raw data of genomes.

## Competing interests

The authors express no competing interests.

## Funding

This research was funded by the National Natural Science Foundation of China (No. 91531303), the international cooperation program of bureau of international cooperation of Chinese Academy of Sciences (No.GJHZ1559), and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding). A.E. was supported by the Chinese Academy of Sciences President's International Fellowship Initiative (No. 2016VBA050). MSP and GDW appreciate the assistances from the Youth Innovation Promotion Association, Chinese Academy of Sciences.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Authors' contributions

AE, Y-PZ and M-SP conceived and *planed the study*. ZAG and MAF provided samples. ZAG prepared the *genomic DNAs of the six samples*. ZAG and G-DW analyzed and interpreted the data. ZAG drafted the *manuscript*. HAN, M-SP and AE revised the manuscript. G-DW *prepared* whole genome resequencing data. All authors have read and approved the final version of the manuscript.

## Author details

<sup>1</sup>Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, PB 76169-133, Kerman, Iran

<sup>2</sup>Yong Researchers Society, Shahid Bahonar University of Kerman, PB 76169-133, Kerman, Iran

<sup>3</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences No. 32 Jiaochang Donglu, Kunming, Yunnan, 650223, China.

<sup>4</sup>State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming 650091, China.

<sup>a</sup>Co-first authors.

## Acknowledgements

This research was carried out as part of PhD thesis at Shahid Bahonar University of Kerman, Iran. We appreciate sampling assistance from the dog owners and staff from department of natural resources in Tehran and Kerman, Kerman Zoo, Tehran Eram Zoo and Shiraz Zoo in Iran. Also, we thank Dr. Hosein Rashidi and Dr. Iman Memarian for their assistance in sampling wolf in Kerman Zoo and Tehran Eram Zoo, respectively.

## References

1. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011; 21:974-984.
2. Agam A, Yalcin B, Bhomra A, Cubin M, Webber C, Holmes C, et al. Elusive copy number variation in the mouse genome. *PloS One.* 2010; 5:e12839.
3. Alam M, Han KI, Lee DH, Ha JH, Kim JJ. Estimation of effective population size in the Sapsaree: a Korean native dog (*Canis familiaris*). *Asian Austral J Anim.* 2012; 25:1063-1072.
4. Alizadeh A. The rise of the highland Elamite state in southwestern Iran. *Curr Anthropol.* 2010; 51:353–383.
5. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* 2011; 12:363-376.
6. Amiri Ghanatsaman Z, Adeola AC, Asadi Fozzi M, Ma YP, Peng MS, Wang GD, et al. Mitochondrial DNA sequence variation in Iranian native dogs. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2017; 17:1-9.
7. Ardalan A, Kluetsch CF, Zhang AB, Erdogan M, Uhlén M, Houshmand M, et al. Comprehensive study of mtDNA among Southwest Asian dogs contradicts independent domestication of wolf, but implies dog–wolf hybridization. *Ecol Evol.* 2011; 1:373-385.
8. Arendt M, Fall T, Lindblad-Toh K, Axelsson E. Amylase activity is associated with *AMY2B* copy numbers in dog: implications for dog domestication, diet and diabetes. *Anim Genet.* 2014; 45:716-722.
9. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 2013; 495:360-364.
10. Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol.* 2011; 12:R68.
11. Berglund J, Nevalainen EM, Molin AM, Perloski M, Andre C, Zody MC, et al. Novel origins of copy number variation in the dog genome. *Genome Biol.* 2012; 13:R73.

12. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 2012; 22:778-790.
13. Björnerfeldt S, Webster MT, Vilà C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* 2006; 16:990-994.
14. Bosse M, Megens HJ, Derks MF, de Cara ÁM, Groenen MA. Deleterious alleles in the context of domestication, inbreeding, and selection. *Evol Appl.* 2019; 12: 6-17.
15. Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 2009; 10:195–205.
16. Chen J, Ni P, Li X, Han J, Jakovlić I, Zhang C, et al. Population size may shape the accumulation of functional mutations following domestication. *BMC Evol Biol.* 2018; 18:4.
17. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat methods.* 2009; 6: 677-681.
18. Chen WK, Swartz JD, Rush LJ, Alvarez CE. Mapping DNA structural variation in dogs. *Genome Res.* 2009; 19:500-509.
19. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 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.* 2012; 6:80-92.
20. Clutton-Brock J. Domesticated animals: from early times, Heinemann in assoc. with British Museum (natural history), London; 1981.
21. Clutton-Brock J. Origins of the dog: domestication and early history. In: Serpell J, editor. *The domestic dog: its evolution, behaviour, and interactions with people.* New York: Cambridge University Press. p7–20; 1995.
22. Colledge S, Conolly J, Shennan S, Bellwood P, Bouby L, Hansen J, et al. Archaeobotanical evidence for the spread of farming in the Eastern Mediterranean 1. *Curr Anthropol.* 2004; 45: S35–S58.
23. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010; 464:704-712.
24. Cruz F, Vilà C, Webster MT. The legacy of domestication: accumulation of deleterious mutations in the dog genome. [Mol Biol Evo.](#) 2008; 25:2331-2336.
25. Di Gerlando R., Mastrangelo S, Sardina MT., Ragatzu M, Spaterna A, Portolano B, et al. A Genome-Wide Detection of Copy Number Variations Using SNP Genotyping Arrays in Braque Français Type Pyrénées Dogs. *Animals.* 2019; 9:77.
26. Elferink MG, Vallée AA, Jungerius AP, Crooijmans RP, Groenen MA. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics.* 2008; 9:1-9.
27. Fang M, Larson G, Ribeiro HS. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* 2009; 5. e1000341.

28. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. *Nucleic Acids Res.* 2011; 40:D84-D90.
29. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, et al. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* 2014; 10: e1004016.
30. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, et al. A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* 2007; 3:e3.
31. Guo Y, Li J, Li CI, Long J, Samuels DC, Shyr Y, et al. The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics.* 2012; 13:666.
32. Hillbertz NH, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat Genet.* 2007; 39:1318-1320.
33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44-57.
34. Keller L, Waller D. Inbreeding effects in wild populations. *Trends Ecol Evol.* 2003; 17:230–
35. Kim H, Sung S, Cho S, Kim TH, Seo K, Kim H. VCS: Tool for Visualizing Copy Number Variation and Single Nucleotide Polymorphism. *Asian Austral J Anim.* 2014; 27:1691-1694.
36. Kimura M. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb. Symp. Quant. Biol.* 1955; 20:33–53.
37. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics.* 2009; 25:1754-1760.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009; 25:2078-2079.
39. Li Y, Wang GD, Wang MS, Irwin DM, Wu DD, Zhang YP. Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome Biol Evol.* 2014; 6:3115-3121.
40. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 2005;438:803–819.
41. Lord K. A comparison of the sensory development of wolves (*Canis lupus lupus*) and dogs (*Canis lupus familiaris*). *Ethology.* 2013;119:110-120.
42. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet.* 2008;40:1166-1174.
43. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce frame work for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20:1297-1303.
44. Molin AM, Berglund J, Webster MT, Lindblad-Toh K. Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics.* 2014; 15:1-10.

45. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One*. 2013; 8:e77910.
46. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res*. 2002; 12: 436–446.
47. Nicholas TJ, Baker C, Eichler EE, Akey JM. A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. *BMC Genomics*. 2011; 12:414.
48. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res*. 2009; 19:491-499.
49. Pailhoux E, Vigier B, Chaffaux S, Serval N, Taourit S, Furet JP, et al. A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nat Genet*. 2001; 29:453-458.
50. Polgár Z, Kinnunen M, Újváry D, Miklósi Á, Gácsi M. A Test of Canine Olfactory Capacity: Comparing Various Dog Breeds and Wolves in a Natural Detection Task. *PLoS One*. 2016; 11:e0154087.
51. Przewdziecki XJ B, Paris G. Our levriers: the past, present and future of all sighthounds. Les Amis de Xavier Przewdziecki, La Colle-sur-Loup. France; 2001.
52. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841-842.
53. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*. 2009; 4: e6524.
54. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012; 28:i333-i339.
55. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006; 444:444-454.
56. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010; 464:587-591.
57. Ruden DM, Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, et al. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet*. 2012; 3:35
58. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, et al. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*. 2011; 12:1-14.
59. Thompson PD. Cardiovascular Adaptations to Marathon Running. *Sports Med*. 2007; 37:444-447.
60. Tonoike A, Hori Y, Inoue-Murayama M, Konno A, Fujita K, Miyado M, et al. Copy number variations in the amylase gene (AMY2B) in Japanese native dog breeds. *Anim Genet*. 2015; 46:580-583.

61. Turner FS. Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. *Front Genet.* 2014; 5:5.
62. van Heesch S, Kloosterman WP, Lansu N, Ruzius FP, Levandowsky E, Lee CC, et al. Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genomics.* 2013; 14:257.
63. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* 2011; 7:e1002316.
64. Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, et al. Multiple and ancient origins of the domestic dog. *Science.* 1997; 276:1687-1689.
65. VonHoldt BM et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature.* 2010; 464:898–902.
66. Wang GD, Zhai W, Yang HC, Fan RX, Cao X, Zhong L, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun.* 2013; 4:1860.
67. Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 2016; 26:21–33.
68. Wang GD, Shao XJ, Bai B, Wang J, Wang X, Cao X, et al. Structural variation during dog domestication: insights from gray wolf and dhole genomes. *Natl Sci Rev.* 2018; 6:110-122.
69. Wayne RK. Molecular evolution of the dog family. *Trends Genet.* 1993; 9:218-224.
70. Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, Vieaud A, et al. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. *PLoS Genet.* 2009; 5:e1000512.
71. Xu L, Hou Y, Bickhart DM, Zhou Y, Song J, Sonstegard TS, et al. Population-genetic properties of differentiated copy number variations in cattle. *Sci Rep.* 2016; 6:23161
72. Yan Y, Yi G, Sun C, Qu L, Yang N. Genome-wide characterization of insertion and deletion variation in chicken using next generation sequencing. *PLoS One.* 2014; 8: e104652.
73. Yi G, Qu L, Liu J, Yan Y, Xu G, Yang N. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics.* 2014; 15:962.
74. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A.* 2008; 105:11597-11604.
75. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics.* 2013; 14: 244.
76. Zhang HH, Wei QG, Zhang HX, Chen L. Comparison of the fraction of olfactory receptor pseudogenes in wolf (*Canis lupus*) with domestic dog (*Canis familiaris*). *J For Res.* 2011; 22:275-280.
77. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol.* 2003; 18: 292-298.
78. Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, et al. Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics.* 2016; 17:379.

## Tables

Table 1-Size distribution of CNVRs detected by CNVnator

Summary statistic of CNVRs	Gain	Loss	Both (loss and gain)	Total
Number of CNVRs	3916	6400	255	10571
Total length(Mb)	83.75	47.28	23.62	154.65
Mean length(Kb)	21.39	7.39	92.62	14.63
Median length(Kb)	11.70	4.49	38.99	7.05
1 ≥ Kb to <5 Kb	555 (14.17%)	60 (0.94%)	-	3996 (37.80%)
5 ≥ Kb to <10 Kb	1119 (28.57%)	3441(53.76%)	14 (5.49%)	2706 (25.59%)
10 ≥ Kb to <20 Kb	1160 (29.62%)	1573 (24.57%)	45 (17.64%)	2252 (21.30%)
20 ≥ Kb to <50 Kb	750 (19.15%)	1047 (16.35%)	189 (74.11%)	1123 (10.62%)
50 ≥ Kb	332 (8.47%)	279 (4.35%)	7 (2.74%)	494 (4.67%)

Table 2- Sampling location and ecotypes

Sample	IDSample	Location	Ecotype	The latitude and longitude of each location
Dog	DogSI1	Sanandaj, Iran	Saluki (Tazi)	35 18' 52" N, 46 59' 32" E
Dog	DogSI2	Bijar, Iran	Saluki (Tazi)	35 52' 22" N, 47 36' 10" E
Dog	DogQI	Esfahan, Iran	Qahderijani	32 38' 0" N, 51° 39' 0" E
Wolf	GW1	Hamadan, Iran	-	34 48' 0" N, 48° 31' 0" E
Wolf	GW2	Tehran, Iran	-	35 41' 46" N, 51 25' 23" E
Wolf	GW3	Kerman, Iran	-	30 17' 0" N, 57 5' 0" E

## Additional Files

**Additional file 1:** Tables S1-S15 and Figures S1- S19

**Additional file 2:** Table S16. Genomic structural variants including insertions, tandem duplication, deletions, translocations (inter and *intra* chromosomal) and inversions for three dogs.

**Additional file 3:** Table S17. Genomic structural variants including insertions, tandem duplication, deletions, translocations (inter and *intra* chromosomal) and inversions for three wolves.

**Additional file 4:** Table S18. The total number of deletions and inversions in dog and wolf genomes.

**Additional file 5.** Table S19. The total number of CNVRs

**Additional file 6:** Table S20. *Statistics* of the detected CNVs for Canine autosomes and X chromosome.

**Additional file 7:** Table S21. Comparison with previous dog CNV studies

## Figures

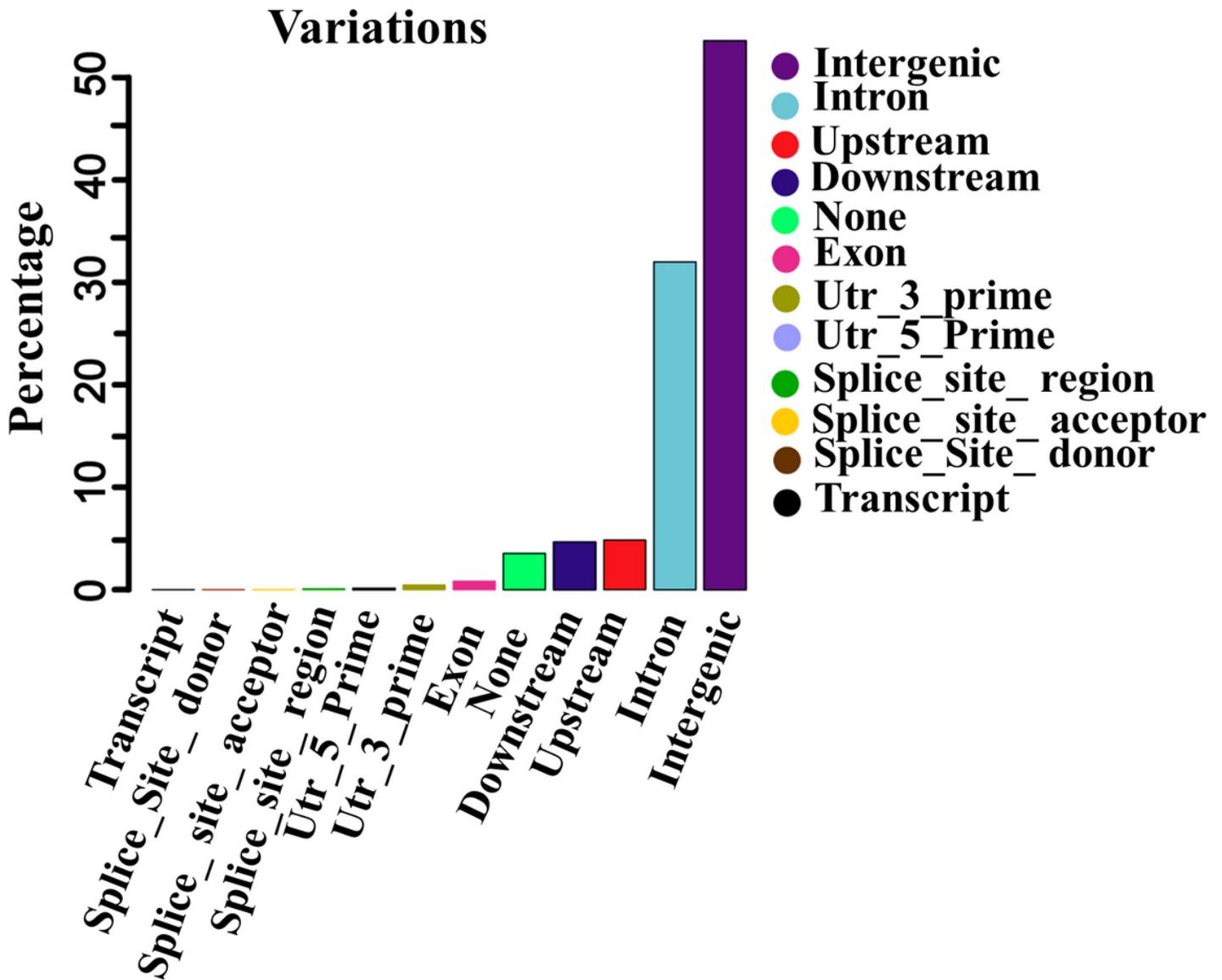


Figure 1

The proportion of SNPs present in each genomic regions, including intergenic, introns, exon, transcript, upstream, downstream, three prime untranslated regions (3'-UTR) and five prime untranslated.

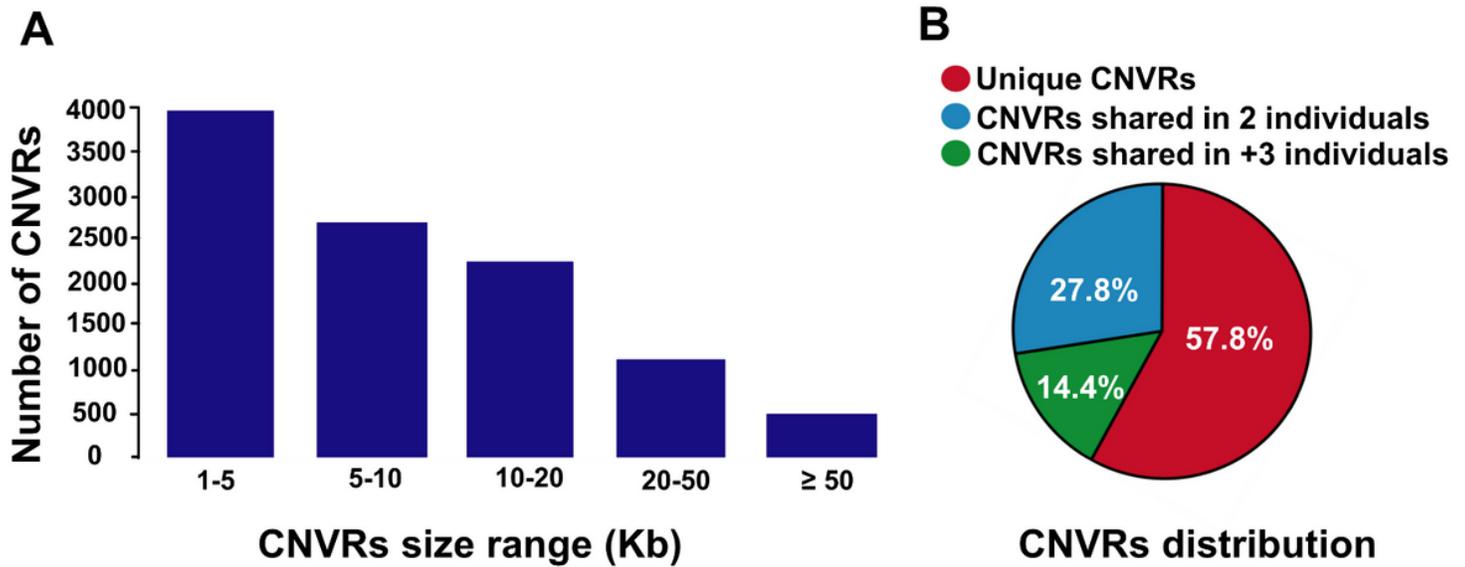


Figure 2

The length and distribution of CNVRs. (a) a total of 6702 (63.39%) and 494 (4.67%) out of all CNVRs had sizes ranging from 1.049 to 10kb and longer than 50 kb in size, respectively. (b) 4466 (42.25%) CNVRs are shared in at least two individuals and 6105 (57.75%) CNVRs present in only one individual.

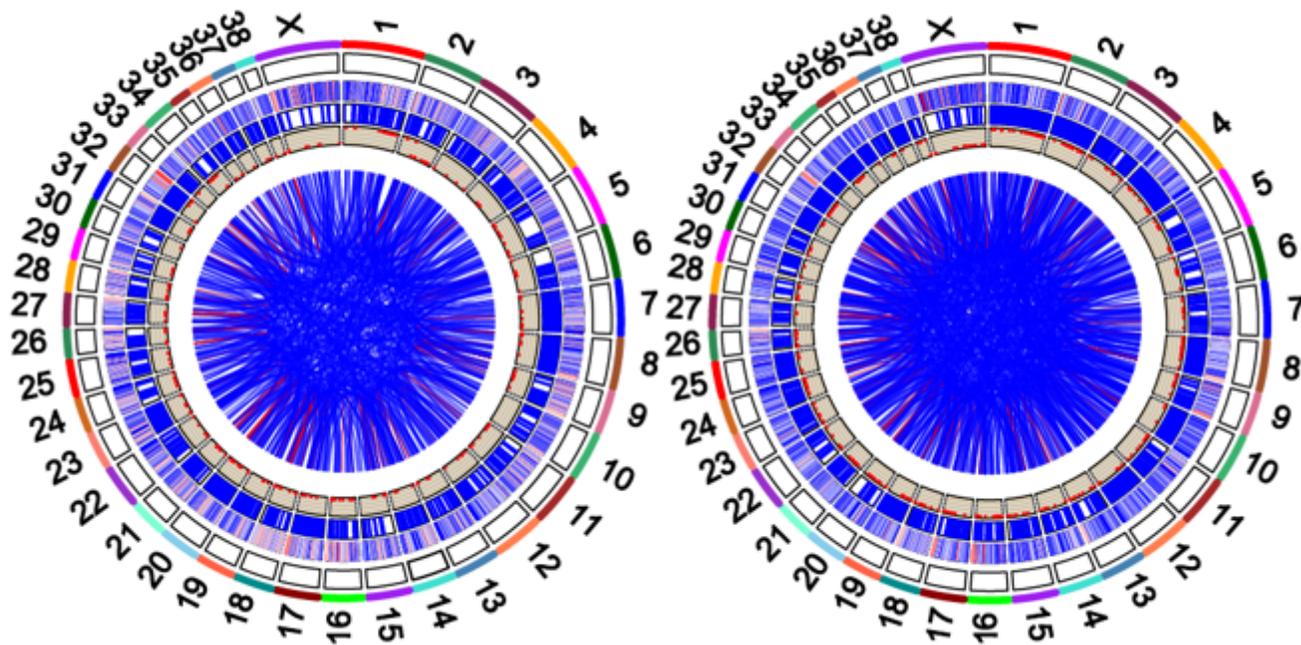


Figure 3

Graphical visualization of predicted SVs for dog (left) and wolf (right). Starting from outside of the circle, the following features are shown: chromosome ideograms, heatmap plot of copy number variation with

color according to the CNV value computed by CNVnator, genomic locations of tandem duplications, genomic locations of inversions and genomic locations of intra and inter-chromosomal links

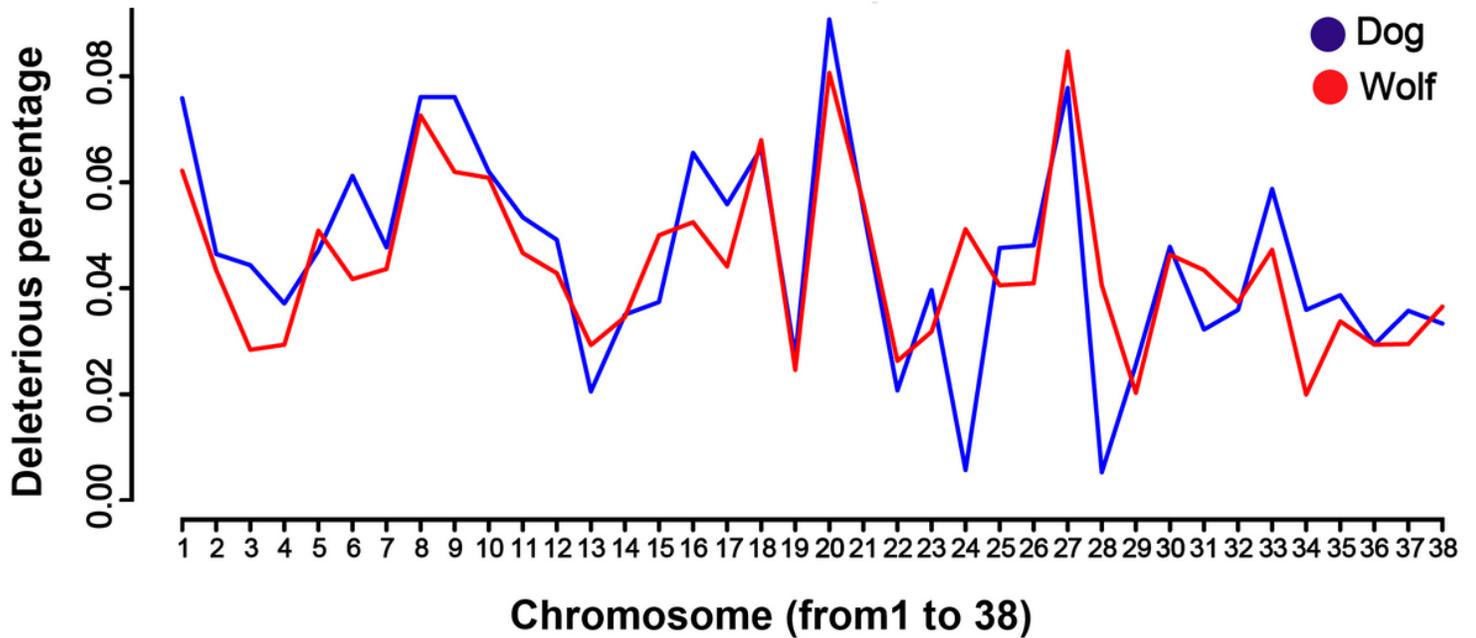


Figure 4

The proportion of deleterious mutations in wolf and doge chromosomes.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S19.xls](#)
- [S17.xlsx](#)
- [Supplementary.doc](#)
- [S20.xlsx](#)
- [S18.xlsx](#)
- [S21.xls](#)
- [S16.xlsx](#)