Contribution of copy number variants to phenotypic diversity of domestic dog breeds.

Inna S Povolotskaya, Aitor Serres-Armero

Institute for Evolutionary Biology, Pompeu Fabra University,

C/Doctor Aiguader, 88, Barcelona, Spain, inna.povolotskaya@upf.edu

Tomas Marques-Bonet

tomas.marques@upf.edu

The dog was domesticated from the gray wolf more than 10,000 years ago, although when and where did domestication happen as well as the role of human implication have been focus of intense debate. Approximately 200 years ago, the majority of modern dog breeds became isolated from each other, in parallel with strong artificial selection for specific physical and behavioral phenotypes favored by humans. A large number of dog breeds have been maintained since then, which has resulted in a broad variety of traits and exceptional phenotypic variation [1-2].

Detecting and understanding the footprint that domestication left in the canine genome is an area of active research. Whole genome re-sequencing data from dogs and wolves has now been commonly used for surveys of genome wide patterns of genetic diversity shaped by both natural and artificial selection. Single nucleotide polymorphisms, microsatellites and variants in mitochondrial DNA have been closely examined for links to the specific dog traits and phenotypes or signals of domestication. These studies have uncovered, that nucleotide diversity in dogs and wolves is 1.5 to 2-fold reduced in dogs due to the 9 to 16-fold reduction in effective population size associated with domestication [2]. Selective breeding further led to reduction in variation, longer linkage disequilibrium blocks and a lower number of haplotypes among purebred dogs compared to wolves and "village dogs", which have not gone through any orchestrated breeding process and artificial selection [3]. This reduction in diversity is striking in the light of great phenotypic variation of the modern dog breeds. Remarkably, several studies have focused on the identification of functional variants responsible for phenotypic changes associated with domestication or contributing to phenotypic variation of the modern dog breeds [4-5].

However, copy number variation (CNV), despite its increasingly recognized

importance as a major contributor to phenotypic diversity, has not been extensively explored in canids. In the present study, we aimed to investigate the extent of copy number variable regions in dogs and wolves. However, the analysis of the genome-wide patterns of CNV diversity across a set of various individuals is a non-resolved challenging task and requires precise estimates of the absolute copy number of each CNV locus for each of the individual genomes. Accuracy of all of the existing methods for absolute copy number inference decreases rapidly as copy number increases, and thus, nearly all of the studies of CNV diversity up to date are limited to biallelic loci with segregating alleles CN=1/CN=2 per haplotype [6-7]. In addition, read depth based methods do not provide confidence of the predicted copy number values, which is extremely important for higher copy numbers, to distinguish between true copy number variability and increased technical noise. It was reported, however, that high copy number loci with differentiated copy number between human populations might be an important contributor to phenotypic differences [7]. We designed a new probabilistic framework of the read depth based approach for accurate absolute copy number inference and CNV detection, which enabled us to perform a comprehensive genome-wide analysis of patterns, diversity and dynamics of CNV loci across the whole range of copy numbers in a set of 34 canid genomes. We further investigate the patterns of copy number differentiation between 481 individual dog genomes from a wide range of dog breeds and look for CNV loci which are associated with a comprehensive list of phenotypic features.

Our analyses show that duplicated regions are enriched in genes and hence likely to be of functional importance. We identify 86 loci with large CNV differences between dogs and wolves, enriched in genes responsible for sensory perception, immune response and metabolic processes, among others. In striking contrast to the observed loss of nucleotide diversity in domestic dogs following the two bottlenecks, we find similar proportion of loci with variable copy number in dogs and wolves, suggesting that other dynamics are acting to particularly select for copy number variants with potentially functional impacts. Remarkably, duplications with relatively low mean copy number are consistently more diverse in dogs than in wolves. These low copy number duplications are significantly enriched in genes and partly species specific, suggesting to a certain extent they might be novel and contribute to functional changes that have occurred after the lineages split.

We recover a number of genic CNVs previously reported to be associated with the dog specific phenotypes. Among these genes is the paralogue to the canine alpha-2B-amylase gene (*AMY2B*), which has been recently connected to adaptation to starch digestion in dogs, showed the most remarkable CNV differentiation between the two species [8]. Another interesting case of copy number expansion in dogs is a 150-Kbps duplication in chromosome 24. This duplication spans three members of the signal-regulatory protein (*SIRP*) gene family, which mediate immune-cell regulation. Similarly, the *CBR1* gene, coding for a carbonyl reductase enzyme involved in the degradation of both environmental and biologically synthesized quinones, lies within a region duplicated in most samples with some dog samples having a higher number of copies. One of the examples of natural and artificial selection acting in opposite directions might be the widespread duplication upstream of the *KITLG* gene, which is linked to the increased risk for squamous cell carcinoma. *KITLG* also plays a role in coat color and patterning and high frequency of this duplication might be explained by artificial selection favoring coat color traits preferred by humans. Overall, our findings indicate, that structural variation plays an important role in shaping of the phenotypic diversity of the modern dog breeds.

1. Lindblad-Toh K, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438.

2. Freedman AH, et al. Genome Sequencing Highlights the Dynamic Early History of Dogs. PLoS Genet. 2014;10.

3. Gray MM, et al. Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. Genetics. 2009;181:1493–505.

4. Boyko AR, et al. A Simple Genetic Architecture Underlies Morphological Variation in Dogs. PLoS Biol. 2010;8.

5. Vaysse A, et al. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. PLoS Genet. 2011;7

6. Alkan C, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat. Genet. 2009;41:1061–7.

7. Sudmant PH, et al. Global diversity, population stratification, and selection of human

copy-number variation. Science. 2015;349:aab3761-aab3761.

8. Axelsson E, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013;495:360–4.

9. Karyadi DM, et al. A Copy Number Variant at the KITLG Locus Likely Confers Risk for Canine Squamous Cell Carcinoma of the Digit. PLoS Genet. 2013;9:e1003409.