Canine Genomics: Diversity, Structural variants, and Aging with Health Implications.

AIT MAHFOUD Rahma, M2 DLAD, Marseiile

Abstract

Studies in human and veterinary medicine can be combined to understand evolution, species adaptation and the origins of disease, in order to develop innovative solutions for the benefit of humans and animals. Interdisciplinary genomics offers a powerful opportunity to combine these fields. The aim of this review is to shed light on recent advances in canine genomics, focusing on three major aspects: genomic diversity, the evolution of aging and the impact of structural variants on health.

Introduction

As an interdisciplinary field of research, canine genomics offers a fascinating insight into evolution, species adaptation and the origins of disease, while contributing to significant advances in human and veterinary medicine. Structural variants are changes in the structure of the genome, such as deletions, duplications, inversions and translocations. These variants can affect gene function and contribute to the development of disease. In this review, I will focus on recent advances in the field of canine genomics, shedding light on: genomic diversity, the evolution of aging, and the impact of structural variants on health.

Genomic Diversity and Dog10k Consortium4

Dog10K Consortium is an international project to sequence and analyze the genomes of over 1879 canids, including 1611 dogs (321 breeds), 309 village dogs, 63 wolves and four coyotes (Meadows et al. 2023), using next-generation sequencing technologies, including Illumina sequencing. The sequencing of canid genomes enables us to reconstruct the evolutionary history of dogs and determine their relationship with wolves. Extensive bioinformatics analysis was carried out to identify single nucleotide variants, indels and structural variants across canid genomes, covering autosomes, the X chromosome and mitochondria. Over 75% of the variation was discovered for 239 breeds sampled (Ostrander et al. 2019). Variants were annotated using tools such as SnpEff and Zoonomia phyloP to prioritize functional variants and assess phylogenetic constraint scores, identifying variants likely to have an impact on genomic function and canine health. In-depth study reveals significant diversity in structural variations in the canine genome. Using tools such as Manta and GraphTyper2, with a focus on retrogenes and elements specific to carnivores (SINEC) , are highlighted as significant contributors to genomic diversity in dogs. Their impact is particularly visible in insertion and deletion variants around 200 bp in size(Meadows et al. 2023).

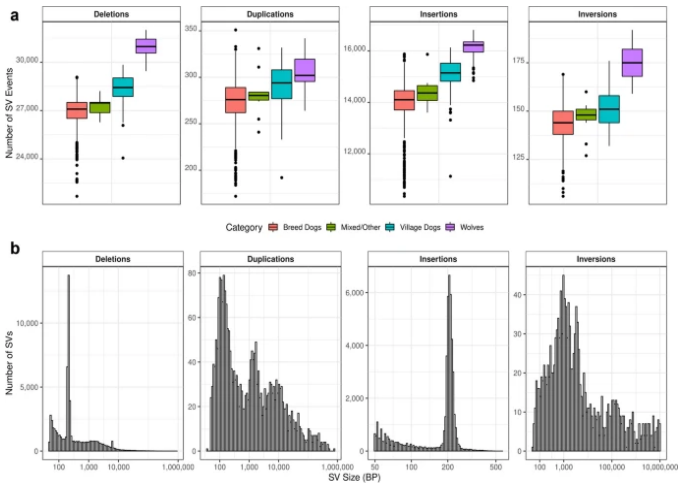

Figure1.*Structural variation detected on 1879 samples.* ***a*** *Boxplots of the number of deletion, duplication, insertion and inversion variants are shown broken down by sample category.* ***b*** *Histograms of the size distribution of each class of structural variants detected are shown. An increase in the number of variants in the ~ 200 bp size group is apparent for both deletions and insertions. This corresponds to the size of the SINEC elements.(Meadows et al. 2023)*

Great Dane Genome Assembly

A genome assembly of a female Great Dane named Zoey was carried out using both long- and short-read data. This assembly improved the gaps present in the

existing reference genome (CanFam) derived from a female Boxer named Tasha(Halo et al. 2021).

The Zoey assembly and 6,857 secondary contigs were aligned on CanFam3.1 to identify large insertion and deletion variants. The Great Dane assembly covers the majority of sequence gaps in the reference genome. The resolved gaps, which have a median GC content of 80.95%, localize more often than expected by chance to transcription start sites and recombination hotspots, suggesting that the stable canine recombinational landscape has shaped the genome architecture. LINE-1 and SINEC have been identified and include retrotransposons. These mobile elements represent the main differences between Great Dane and Boxer assembly, and their continued activity is a major contribution to the diversity and evolution of canine genomes. Retrotransposons may play a role in the regulation of gene expression and be associated with genetic diseases in dogs (Halo et al. 2021), and also with differences in response to medical treatments(Michot, s. d.). Studies have shown that active mobile elements may be associated with an increased risk of tumor and cancer development in dogs. For example, retrotransposons can disrupt the regulation of genes involved in tumor suppression, which may contribute to the formation of soft tissue tumors in large dogs, such as the Great Dane(Halo et al. 2021).

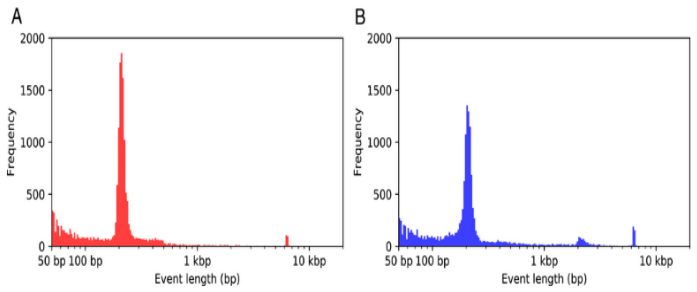

Figure2. *Size of structural variants identified between CanFam3.1 and zoey sets.* ***A:*** *Deletion size distribution.* ***B:*** *Insertion size distribution. Deletions and insertions are types of genetic variation that can affect genome size. The majority of deletions and insertions are small, but there is also a significant population of large variants. These large variants are caused by mobile elements of the genome, such as the LINE1 and SINEC sequences(Halo et al. 2021).*

Structural Variants in the Chinese Population

The article "Structural variants in the Chinese population and their impact on phenotypes, diseases and population adaptation" presents an analysis of structural variants in the Chinese population and their impact on phenotypes, diseases and population adaptation. Long-read sequencing was used to analyze structural variants in 405 unrelated Chinese individuals, with 68 phenotypic and clinical measurements (Wu et al. 2021). The results revealed a landscape of 132,312 non-redundant structural variants, of which 45.2% were novel. Northern and southern Chinese populations were differentiated by annotating 1,929 loss-of-function structural variants affecting the coding sequence of 1,681 genes. The researchers found rare deletions in the HBA1/HBA2/HBB genes associated with anemia. In addition, structural variants linked to immunity were identified. Offering insights into the roles of structural variants in population adaptation (Wu et al. 2021). Specialized bioinformatics tools were used to analyze long-read sequencing data and detect structural variants. Including structural variant calling software such as Sniffles, NanoVar, NanoSV, and other bioinformatics analysis tools for structural variant detection and characterization. A set of 132,312 non-redundant SVs was obtained by merging SVs in all samples, comprising 67,405 DELs, 60,182 INSs, 3,956 DUPs and 769 INVs.

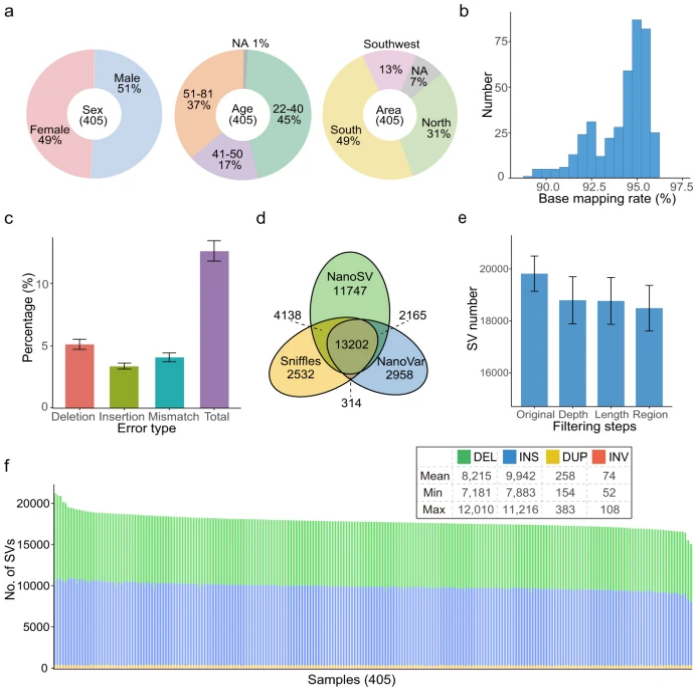

Figure3. ***a****: Overview of information on the sample studied, "N/A" indicates unavailable.* ***b:*** *Distribution of base mapping rate of clean data aligned to the GRCh38 reference genome.*

c: *Error rate represented as a percentage for each type (n = 405 per group).* ***d:*** *Average number of SVs identified by Sniffles, NanoVar and NanoSV per individual and the overlap between them.* ***e:*** *Average number of SVs after each filtering step (n = 405 per group), "Original" represents unfiltered SVs detected by at least two callers, "Depth", "Length" and "Region" represent filtered SVs based on the number of reads supported, extra-long intervals and regions of very low complexity, respectively.* ***f*** *Number of SVs of each type in each individual. Data are represented by the mean ± standard deviation in* ***c*** *and* ***e****(Wu et al. 2021).*

The number of SV of each type varies from one individual to another. In the healthy group, the most frequent SV type is deletion, followed by insertion and translocation(Wu et al. 2021). In the patient group, the most frequent SV type is deletion, followed by duplication.The results of this study suggest that SVs are frequent in the general population, with an average of 5.7 SVs per individual. The most frequent type of SV is deletion, followed by insertion and translocation. The difference in SV frequency between the healthy and diseased groups suggest that SVs may be associated with disease.

Epigenetic clocks and aging

The article "DNA methylation clocks for dogs and humans" presents epigenetic clocks that apply to both dogs and humans. These clocks measure methylation levels in highly conserved stretches of DNA, are capable of predicting chronological age with remarkable accuracy, as well as health status, as DNA methylation levels have been associated with disease and other health problems. Epigenetic clocks can be used to predict the average time to death for dogs and humans, and are likely to target genes involved in longevity, age at menopause, dementia, macular degeneration and mortality risk(Thompson et al. 2017). They are used to estimate chronological age in years (DNAmAge) and relative age, which is the ratio of age and maximum lifespan of the respective species, between 0 and 1 (DNAmRelativeAge). DNA methylation is linked to the lifespan of dogs and humans through chemical marks on the DNA(Sturm et al. 2023). These represent small CH3 groups added to the cytosines of CpG dinucleotides, tipping it into a certain position and pushing it into the next phase of life(Heard et Martienssen 2014). DNA methylation stably but potentially reversibly switches off genes, and is a widely studied epigenetic marker for its role in embryogenesis and carcinogenesis (Heard et Martienssen 2014). DNA methylation is measured in dogs and humans using high-throughput sequencing techniques. These techniques make possible to characterize DNA methylation along the genome using third-generation sequencing methods, such as PacBio third-generation sequencing and Oxford Nanopore, which map genomic regions that are difficult to sequence accurately using short-read sequencing technologies.

Analysis of recent advances and prospects

Recent advances in canine genomics have led to important discoveries about genetic diversity, the evolution of aging and the impact of structural variants on health. The Dog10K project has sequenced the genomes of over 2,000 canines. This revealed great genetic diversity within the canine population, with significant differences between breeds, geographical groups and species. Analyses of the Dog10K data have identified over 48 million unique indel and structural nucleotide variants. These variants are unevenly distributed across the genome, with a higher concentration in genomic regions related to development, immunity and metabolism. Mobile elements, such as retrotransposons, are also very present in the canine genome. These mobile elements can contribute to canine genetic diversity and genome evolution. Studies on DNA methylation have led to the development of epigenetic clocks that can predict the chronological age and biological age of dogs. These clocks are based on the observation that DNA methylation levels change with age, and can be used to identify dogs at increased risk of premature aging or age-related diseases. Structural variant studies have also identified genes and genomic regions associated with aging. Studies of structural variants in dogs have identified variants associated with a variety of diseases, including hereditary diseases, cancers and inflammatory diseases. For example, deletions in the MDR1 gene are associated with sensitivity to antiparasitic drugs(Geyer et Janko 2012), while duplications in the COL1A2 gene are associated with a form of hip dysplasia. The study of the Chinese

population revealed great genetic diversity within this population, with significant differences between geographical groups and ethnicities. This diversity is due to a combination of factors, including migration history, geographical isolation and natural selection. Analyses of the data from this study identified over 132,000 non-redundant structural variants, 45.2% of which are novel. These variants were of high quality, with an estimated false-positive discovery rate of 3.2%. The authors annotated 1,929 loss-of-function structural variants affecting the coding sequence of 1,681 genes. They found rare deletions in the HBA1/HBA2/HBB genes associated with anemia. In addition, structural variants linked to immunity were identified, differentiating northern and southern Chinese populations.

Recent advances in canine genomics offer new perspectives for understanding the health of dogs, as well as application in human medicine. The genetic diversity of dogs play a crucial role in their adaptation to different environments and in understanding how they have evolved in relation to humans. Their genomic structure is essential for understanding the function of genes, regulating their expression and developing approches for diagnosing and treating various diseases. The study of aging in dogs has significant implications for human medicine, paving the way for new approaches to slowing down aging and treating age-related diseases. These advances could improve both dog care and human medicine.

The comparison of tools for identifying and annotating genetic variants is crucial. Criteria such as sensitivity, specificity, functional and clinical accuracy, complexity and cost of the tools need to be taken into account. Combining the strengths of different tools could lead to the development of robust tools for the identification and annotation of all types of variants.

Conclusion

These advances make possible to study genetic diversity between species in depth and to understand the aging process by analyzing genetic variations. They open up prospects in the fields of canine health, human medicine and the prevention of certain diseases. The comparison of different identification and annotation tools is very important for obtaining effective, robust results. This offers a better understanding of genomics, with direct implications for the health and well-being of both animals and humans.

References

1- Geyer, Joachim, et Christina Janko. 2012. « Treatment of MDR1 Mutant Dogs with Macrocyclic Lactones ». *Current Pharmaceutical Biotechnology* 13 (6): 969-86. https://doi.org/10.2174/138920112800399301.

2- Halo, Julia V., Amanda L. Pendleton, Feichen Shen, Aurélien J. Doucet, Thomas Derrien, Christophe Hitte, Laura E. Kirby, et al. 2021. « Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes ». *Proceedings of the National Academy of Sciences* 118 (11): e2016274118. https://doi.org/10.1073/pnas.2016274118.

3- Heard, Edith, et Robert A. Martienssen. 2014. « Transgenerational Epigenetic Inheritance: Myths and Mechanisms ». *Cell* 157 (1): 95-109. https://doi.org/10.1016/j.cell.2014.02.045.

4- Meadows, Jennifer R. S., Jeffrey M. Kidd, Guo-Dong Wang, Heidi G. Parker, Peter Z. Schall, Matteo Bianchi, Matthew J. Christmas, et al. 2023. « Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture ». *Genome Biology* 24 (1): 187. https://doi.org/10.1186/s13059-023-03023-7.

5- Michot, Pauline. s. d. « Utilisation des séquences de génome complet pour l'identification de mutations délétères responsables d'anomalies génétiques récessives chez le bovin. »

6- Ostrander, Elaine A, Guo-Dong Wang, Greger Larson, Bridgett M vonHoldt, Brian W Davis, Vidhya Jagannathan, Christophe Hitte, Robert K Wayne, Ya-Ping Zhang, et Dog10K Consortium. 2019. « Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health ». *National Science Review* 6 (4): 810-24. https://doi.org/10.1093/nsr/nwz049.

7- Sturm, Ádám, Himani Sharma, Ferenc Bodnár, Maryam Aslam, Tibor Kovács, Ákos Németh, Bernadette Hotzi, et al. 2023. « N6-Methyladenine Progressively Accumulates in Mitochondrial DNA during Aging ». *International Journal of Molecular Sciences* 24 (19): 14858. https://doi.org/10.3390/ijms241914858.

8- Thompson, Michael J., Bridgett vonHoldt, Steve Horvath, et Matteo Pellegrini. 2017. « An epigenetic aging clock for dogs and wolves ». *Aging (Albany NY)* 9 (3): 1055-68. https://doi.org/10.18632/aging.101211.

9- Wu, Zhikun, Zehang Jiang, Tong Li, Chuanbo Xie, Liansheng Zhao, Jiaqi Yang, Shuai Ouyang, Yizhi Liu, Tao Li, et Zhi Xie. 2021. « Structural Variants in the Chinese Population and Their Impact on Phenotypes, Diseases and Population Adaptation ». *Nature Communications* 12 (1): 6501. https://doi.org/10.1038/s41467-021-26856-x.