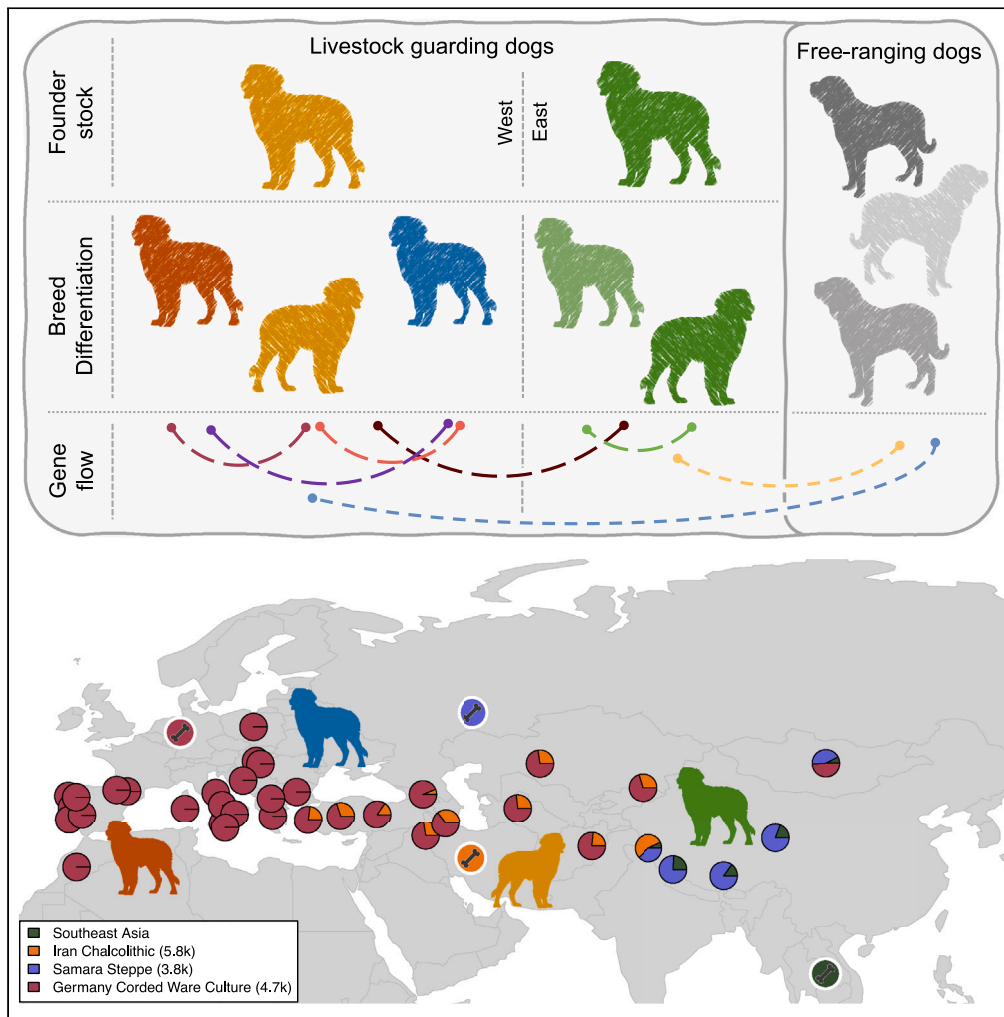## Article

Multiple ancestries and shared gene flow among modern livestock guarding dogs

Diogo Coutinho-Lima, Dayna L. Dreger, Ignacio Doadrio, ..., Greger Larson, Elaine A. Ostrander, Raquel Godinho

diogofclima5@cibio.up.pt (D.C.-L.)
dayna.dreger@nih.gov (D.L.D.)
greger.larson@arch.ox.ac.uk (G.L.)
eostrand@mail.nih.gov (E.A.O.)
rgodinho@cibio.up.pt (R.G.)

Highlights

Livestock guarding dog (LGD) breeds are part of two contemporary lineages

Extensive gene flow among distinct LGD breeds

Reproductive isolation may not be necessary to preserve highly specialized dogs

LGDs mostly used as pets exhibit higher inbreeding, opposed to working dogs

iScience

Article

Multiple ancestries and shared gene flow among modern livestock guarding dogs

Diogo Coutinho-Lima,[1,2,3,10,12,*] Dayna L. Dreger,[4,10,*] Ignacio Doadrio,[5] Heidi G. Parker,[4] Hamid R. Ghanavi,[6] Laurent Frantz,[7,8] Greger Larson,[9,11,*] Elaine A. Ostrander,[4,11,*] and Raquel Godinho[1,2,3,11,*]

SUMMARY

Livestock guarding dogs (LGDs) have been used to protect livestock for millennia. While previous works suggested a single origin of modern LGDs, the degree and source of shared ancestry have not been tested. To address this, we generated genome-wide SNP data from 304 LGDs and combined it with public genomic data from 2,183 modern and 22 ancient dogs. Our findings reveal shared ancestry and extensive gene flow among modern LGD breeds which we attribute to historical livestock migrations. Additionally, admixture between LGDs and free-ranging dogs argues against reproductive isolation as a core mechanism for maintaining the specialized skills of LGDs. Finally, we identify two lineages within modern LGDs and uncover multiple ancestries tracing back to distinct Eurasian ancient dogs, concordant with the absence of a single ancestor. Overall, our work explores the complex evolutionary history of LGDs, offering valuable insights into how human and livestock co-migrations shaped this functional group.

INTRODUCTION

Dogs (*Canis lupus familiaris*) played a pivotal role in shaping hunter-gatherer communities and later agrarian societies. Notably, the domestic dog stands as the sole domesticated animal predating farming, adapting to work alongside humans in many ways.[1] Adaptations include occupational necessities, such as assisting humans in herd management tasks including livestock guarding. Dogs also experienced physiologic adjustments such as the ability to digest increased levels of starch.[2,3]

Livestock guarding dogs (LGDs) were likely indispensable for the expansion of agrarian communities due to their role in protecting livestock from predation.[4] Owing to their close association with pastoral societies, frequently concomitant with nomadic lifestyles, LGDs spread across Eurasia.[5] Thus, understanding the relationships between distinct LGD breeds might yield insights into human migrations associated with the diffusion of livestock practices. Modern LGDs have diversified into distinct breeds, adapting to various livestock species and to local environments.[6] For instance, the Tibetan Mastiff excels in the low-oxygen conditions of the high-altitude Himalayas.[7] LGDs include both breeds recognized by international registering bodies and native landrace populations. The former have strict ranges of morphological and behavioral variation, and selection is additionally guided by appearance, while native landrace populations exhibit less strict aesthetic requirements and selection is mostly driven by their specific roles.[8] Notwithstanding this distinction, here we use the term "breeds" to refer to both locally defined landrace populations and recognized breeds.

Popular lore suggests that LGDs may have originated in the Fertile Crescent, i.e., modern-day Turkey, Iraq, and Syria, where they played a role in early livestock management.[6,9] This region has also been identified as the geographic origin of domestic sheep, goats, pigs, and cattle.[10] Nonetheless, the geographic origin, and whether LGDs emerged from a single or multiple lineages, has yet to be tested. Additionally, information regarding LGD relationships with other dog populations is limited. Dutrow et al.[11] recently described a genetic association between free-ranging and purebred dogs tied to the same geographic region, including LGDs, suggesting historical and contemporary admixture. This may imply that maintaining barriers to gene flow between free-ranging dogs and dogs selected for functional roles may not be crucial for preserving those specialized skills.

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal
[2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
[3]BIOPOLIS - Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal
[4]Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
[5]Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, Spain
[6]Department of Biology, Lund University, Lund, Sweden
[7]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK
[8]Palaeogenomics Group, Faculty of Veterinary Sciences, Ludwig-Maximilians-University of Munich, Munich, Germany
[9]Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
[10]These authors contributed equally
[11]Senior author
[12]Lead contact
*Correspondence: diogofclima5@cibio.up.pt (D.C.-L.), dayna.dreger@nih.gov (D.L.D.), greger.larson@arch.ox.ac.uk (G.L.), eostrand@mail.nih.gov (E.A.O.), rgodinho@cibio.up.pt (R.G.)
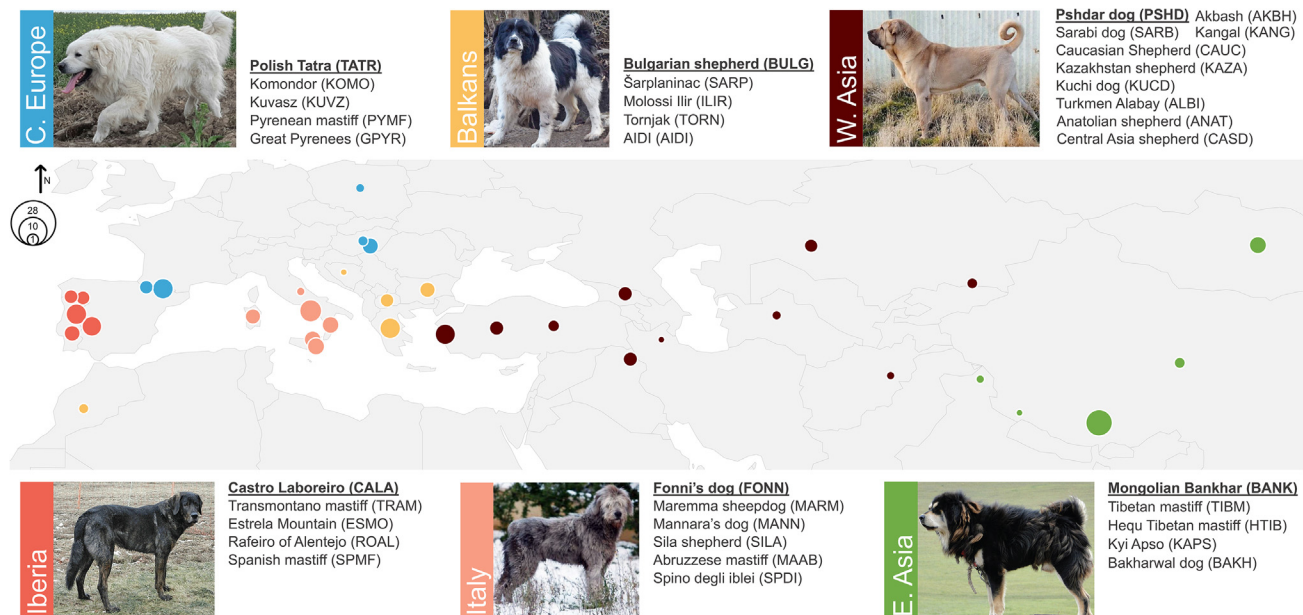https://doi.org/10.1016/j.isci.2024.110396

Figure 1. Livestock guarding dog breeds sampled in our study

Each circle represents the sample size and approximate geographic location of a breed, colored at the group level. Breeds are grouped as follow: Iberia, Italy, Central Europe (C. Europe), Balkans, West Asia (W. Asia), and East Asia (E. Asia). Breeds in bold and underlined are depicted in pictures. Abbreviations are provided in parentheses. Full information is in Table S3. Photograph credits are listed in the acknowledgments.

In this study, we tested the single-origin hypothesis regarding LGDs by generating genome-wide SNP data from 304 LGDs, representing 36 contemporary breeds (Figure 1), and analyzed it along with whole-genome sequencing data from 22 ancient dogs, spanning 10 thousand years (ky) before present (BP, Figure 2A). Additionally, we used SNP data from 165 modern Eurasian free-ranging dogs to explore potential connections between LGDs and free-ranging dogs in overlapping regions, allowing us to ask if reproductive isolation is a primary mechanism for preserving the specialized skills observed in LGDs.

RESULTS AND DISCUSSION

Ancient dogs reveal multiple ancestries within LGDs

Ancestry patterns among domestic dogs are mostly explained by the geographic origin of each breed.[12,13] However, the ancestry background of LGDs, one of the most widely distributed functional dog types, remains unclear. To explore genetic structure within the LGD group, we performed principal-component analysis (PCA, Figures 2B, S1, and S2) and uniform manifold approximation and projection (UMAP, Figure S3). Notably, LGD breeds from the same geographic region clearly overlap. When projected onto the PCA of LGD variability, the ancient dogs are localized with breeds hailing from the same geographic regions (Figure 2B). Neighbor-joining phylogeny (Figure S4) and admixture (Figure 2C) analyses also replicate this geographic association. This suggests that patterns of LGD ancestry are primarily determined by a breed's geographical origin and rooted in ancient dogs from the same region.

To ascertain the proportions of distinct ancestries within LGD breeds, we selected seven ancient dogs from the last 10 ky, along with one modern New Guinea singing dog, to represent potential sources of ancestry. The New Guinea singing dog was selected as a representative of the Southeast Asia lineage for which there is no ancient genomic data available.[12] Samples were selected to specifically test ancestries linked to geographical or temporal patterns. The best-fitting models generated by the qpAdm function of Admixtools[14] support East Asian LGDs, such as the Tibetan Kyi Apso, as a blend of Southeast Asia ancestry (modern New Guinea singing dog) with large contributions from the Samara Steppe ancient dog (3.8 ky BP, Figure 2D). The prominence of Steppe-related ancestry in East Asian LGDs may be explained by a strong genetic turnover of the local genetic ancestry into Steppe-related diversity following the eastward migration of Steppe pastoralists toward East Asia <5 ky ago.[12] A recent study on the establishment of dairy pastoralism in the Tibetan Plateau described the introgression of West Eurasian ancestry into Tibetan dogs.[15] We hypothesize that the ancestry of East Asian LGDs might also be linked to the emergence of dairy pastoralism in the region.

The best-fitting models for European LGDs, like the Estrela Mountain dog or Kuvasz, rely solely on the genetic background of an ancient dog from Germany (4.7 ky BP), in accordance with other modern European breeds.[12] By comparison, the genetic background of LGDs from West Asia finds a better fit in models that include the ancient dogs from Germany and Chalcolithic Iran (Figure 2D). Modern free-ranging dogs and non-LGD breeds from West Asia display a similar blend of ancestries.[12] This is concordant with the expansion of a single dog population in Europe, which completely replaced other early European dogs and later expanded into Asia.[12] The dynamics that triggered or facilitated this ancestry replacement in Europe remain unclear.
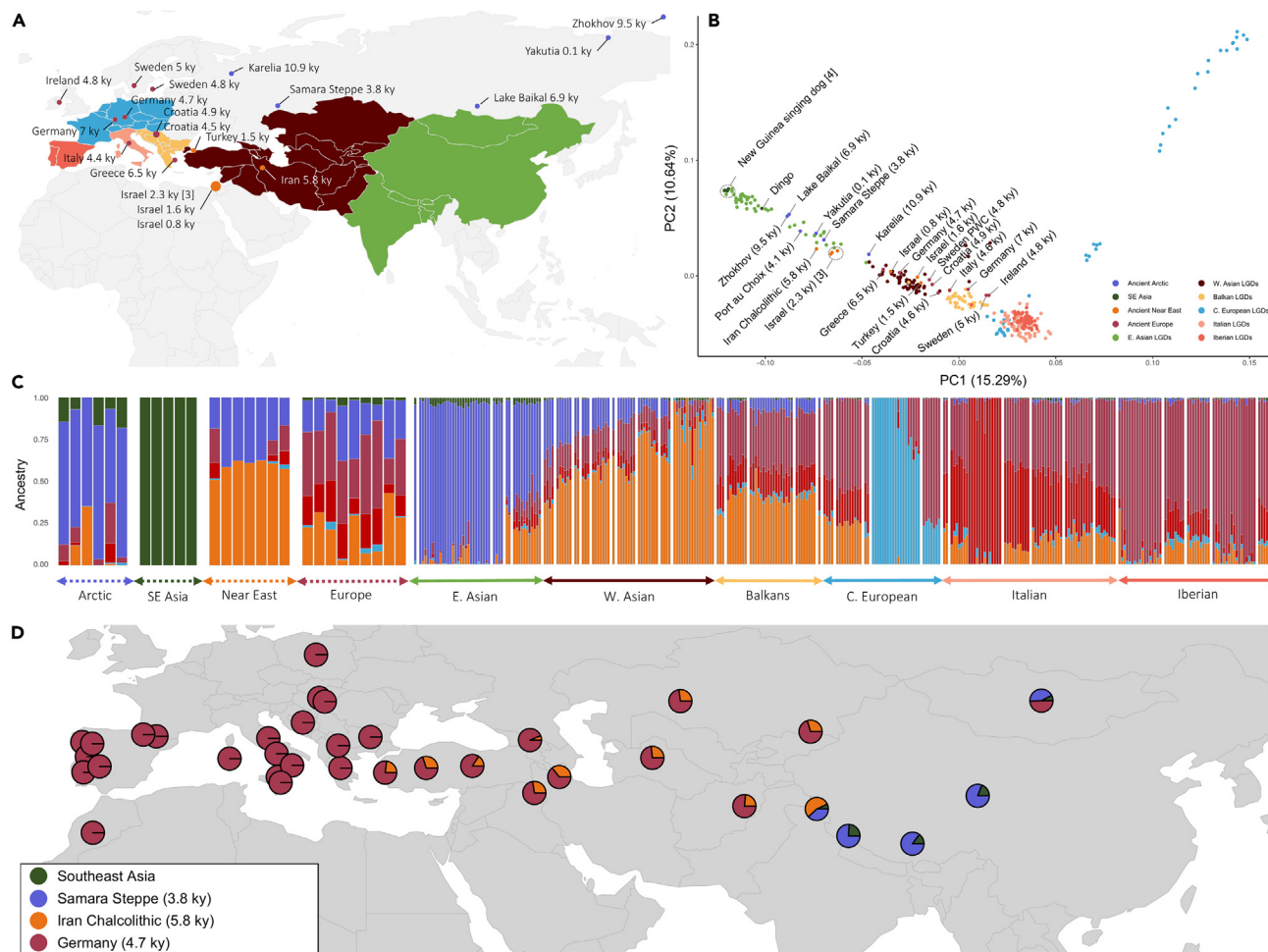
Figure 2. Genomic ancestry of livestock guarding dogs

(A) Geographic distribution of LGD breeds sampled in this study (colored shades) and locations of ancient dogs. ky, 1,000 years. The ancient American dog (AL3197, 4.1 ky – Port au Choix, Canada) does not appear on the map.

(B) Population structure of groups of LGDs and ancient dogs exposed by PCA. Each colorful point represents one individual of a defined group: Iberia, Italy, Central Europe (C. Europe), Balkans, West Asia (W. Asia), and East Asia (E. Asia).

(C) The proportion of membership of 304 individuals from 36 LGD breeds, 22 ancient dogs, 4 New Guinea singing dogs, and 1 Dingo at 6 estimated divisions, as calculated by the ADMIXTURE software. The dotted arrows correspond to the ancient dog samples and the Southeast Asia lineage (SEA) and the solid arrows to LGDs; arrows colored according to the defined group.

(D) qpAdm ancestry proportions of LGD populations using ancient dogs and the New Guinea singing dog (Southeast Asia) as proxies for ancestry sources. See also Figures S1–S4.

While multiple ancestries within LGDs seemingly contradict a single-origin scenario for this functional group, we cannot rule this hypothesis out, as local admixture could have masked the signal of a shared common ancestor. It remains possible that LGDs from East Asia shared a common ancestor with other LGDs but underwent a complete turnover in ancestry, erasing any discernible trace of a single origin in the genome of modern LGDs. Instances of complete ancestry replacements have been described before, such as the substitution of native American dogs with European ancestry following the Age of Discovery.[16] The alternative hypothesis of a multiple origin would, in turn, also be consistent with our findings. In this scenario, LGDs independently emerged in distinct regions in response to the rising demands of livestock management, followed by extensive diffusion. This process could have been accompanied by convergent selection or shared gene flow, ultimately leading to the development of similar phenotypes across all LGD breeds.

Multiple lineages among modern LGDs

Modern dog breeds often form clusters based on their distinct roles and function within human societies, such as hunting, especially those within the same geographic region.[17] To test if LGDs are a monophyletic group, we built neighbor-joining phylogenies of LGDs with other modern breeds (Figures 3A, S5, and S6). LGDs typically group by breed, and deviations often result in clustering with breeds from the same
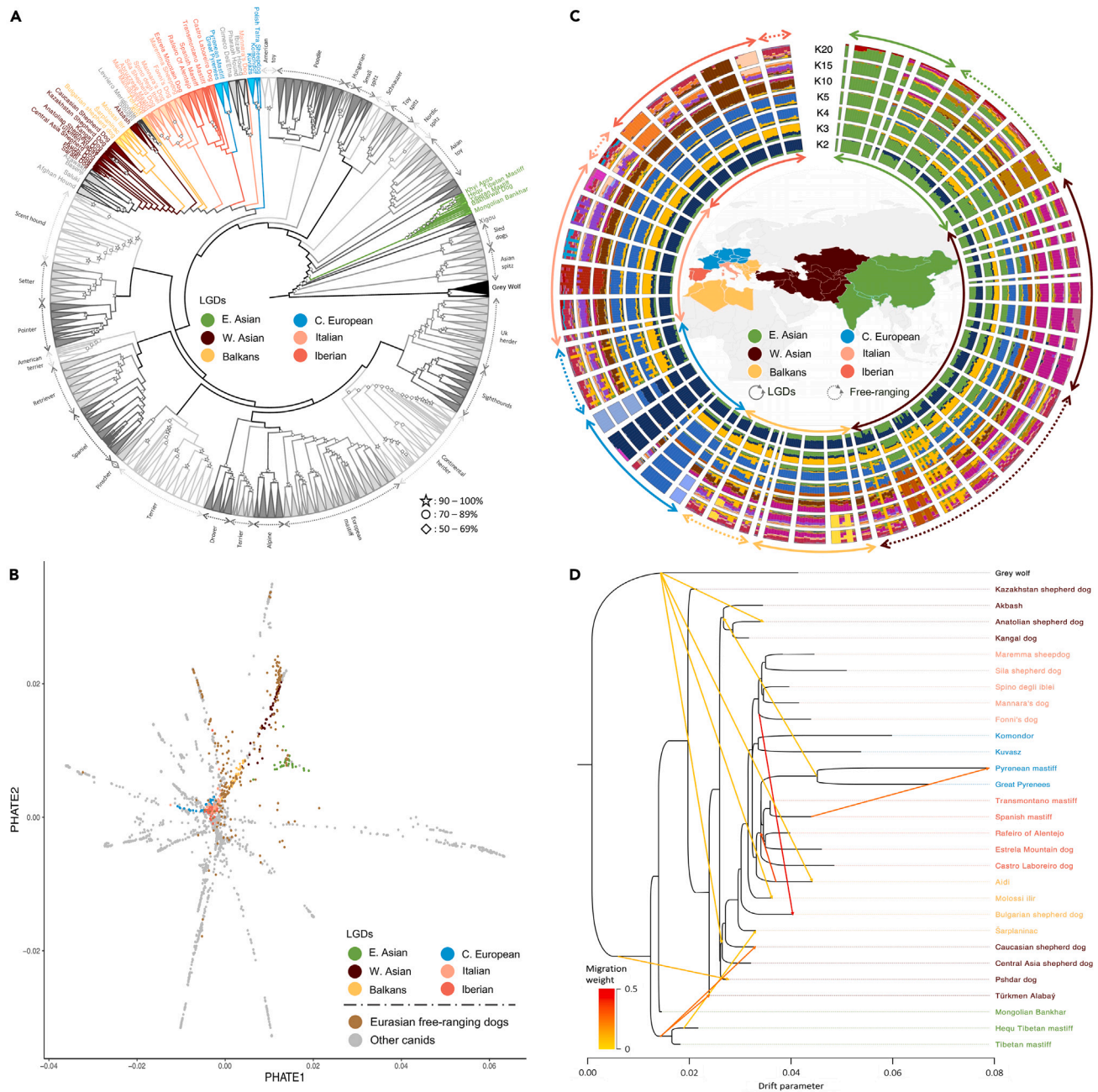
Figure 3. Population structure and gene flow of livestock guarding dogs

(A) Neighbor-joining tree of the domestic dog breeds rooted to the gray wolf (black). Livestock guarding dogs are colored according to the defined geographic group (Iberia, Italy, Central Europe [C. Europe], Balkans, West Asia [W. Asia[, East Asia [E. Asia]), and the remaining canids are colored gray.

(B) PHATE plot. Livestock guarding dogs are colored according to their defined geographic group, Eurasian free-ranging dogs are colored light-brown, and the remaining canids are colored gray.

(C) Proportion of membership of LGD breeds and free-ranging dog populations, as calculated by the ADMIXTURE software. The dotted arrows correspond to free-ranging dogs, and the solid arrows to LGDs. Breed and free-ranging dog population names for each cluster in Figure S8.

(D) Maximum likelihood tree with 13 migration events, using the gray wolf as root. Admixture boundaries are denoted with arrows in the direction from the migrant's origin to the recipient breed and heat colored according to the migration weight. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch. See also Figures S5–S14 and Tables S1 and S2.

region (Figure S6). We observe LGDs mostly clustering in the previously defined "Mediterranean clade"[17] (Figures 3A and S5). However, LGDs from East Asia do not cluster with LGDs from Europe and West Asia. Instead, they form a distinct clade, which is consistent with a possible dual origin for LGDs.

To further test between single and multiple origins for LGDs, we applied PHATE (potential of heat-difusion for affinity-based trajectory embedding), a computational tool for visualizing both local and global structures in high-dimensional data. We aimed to ascertain whether LGDs form a single group within the broader spectrum of canine groups, such as retrievers or scent hounds, which would be suggestive of a single functional origin. Again, we observe two clear groups, one consisting of East Asian LGDs and the other formed by the remaining LGD breeds (Figure 3B). While we cannot definitively distinguish between multiple origins or a single origin succeeded by the replacement of East Asian LGDs, we argue that the latter would entail the sharing of at least a minor proportion of ancestry across all breeds, a pattern not evident in our analyses. Regardless, our findings strongly support the hypothesis that modern LGD breeds from East Asia and the rest of Eurasia are part of two lineages that have evolved independently for millennia.

Similar patterns for the development of abilities have also been identified for other functional groups. A recent study addressing the origins of Eurasian and African sighthounds suggested a multiple-origin scenario,[18] aligning with our finding for LGDs. Together, both studies highlight abilities that were developed multiple times across a functional group. Interestingly, the evolutionary link between LGDs and sighthounds has also been previously described.[11,17] Sighthound breeds form three groups in a phylogenetic analysis: Western/Northern Eurasia, Mediterranean, and Chinese Xigou. PHATE analyses show Chinese Xigou integrating with East Asian LGDs, and Mediterranean sighthounds integrating with West Eurasian LGDs (Figure S7). We contend that this may be explained by ancestry patterns observed in modern dogs being predominantly shaped by the geographic origins of each breed.

Livestock practices mediated gene flow among LGDs

Multiple investigations report that human and livestock migrations sustained gene flow between LGD breeds.[8,19,20] To test this, we used ADMIXTURE[21] and TreeMix.[22] Our findings indicate a consistent admixture pattern among LGDs, with a substantial number of breeds displaying mixed ancestry (Figures 3C and S8). Moreover, most of the migration edges predicted by TreeMix were identified between breeds separated by large geographical distances, such as that observed between Fonni's dog and Bulgarian shepherd dogs (Figures 3D and S9). This likely indicates very high similarity among geographically close breeds.[22] We also computed pairwise breed differentiation through the fixation index (F_{ST}, Table S1) and detected low overall differentiation between breeds. Finally, we employed D and f_3-statistics in the form of $D(P_1,P_2,P_3,O)$ and $f_3(P_1,P_2,O)$, respectively, to detect gene flow between each pair of LGD breeds (Figures S10 and S11). As anticipated, LGDs exhibit higher allele sharing with breeds from the same or neighboring regions, reflecting recent shared ancestry and extensive gene flow.

Widespread gene flow supports the hypothesis that livestock migrations have played a role in shaping the genomic background of LGDs.[8,19,20] Albeit restricted due to modern political boundaries, livestock practices persist across Eurasia, maintaining their historical importance and cultural significance.[23,24] Transhumance migration, a biannual cultural movement of livestock between high- and lowlands that played a pivotal role in shaping the Eurasian landscape, greatly impacted LGDs.[25,26] During these migrations, LGDs moved and followed the livestock, serving as a dynamic conduit for gene flow across distinct regions and breeds. For instance, Spanish Mastiff dogs exhibited no detectable variation in population structure despite distances of up to 700 km between sampling sites, likely attributed to the participation of most sampled individuals in transhumance migrations. Similarly, LGDs within West Asia, a region where transhumance migrations persist, revealed a clear admixture pattern among all breeds and a lack of genetic distinctiveness (Figure 3C).

Other studies of Balkans and Italian LGD breeds have also described this association between gene flow among LGDs and transhumance migrations.[8,19] This highlights the influence of transhumance, which acted as a strong driver of homogenization and gene flow between distinct LGD breeds.[8,19,20] Thus, cultural traditions in livestock management, particularly transhumance, play a paramount role in shaping the genomic background of LGDs, actively supporting not only gene flow between distinct breeds but also the preservation of a single breed across extensive geographical spans. As such, transhumance offers crucial insights into the complex dynamics between human and livestock cultural migrations and LGDs' genetic diversity.

Signatures of shared ancestry between LGDs and free-ranging dogs

A previous study involving 226 breeds and free-ranging dogs from 47 countries underlined shared ancestry and admixture among geographically overlapping dog types.[11] Moreover, admixture between LGDs and free-ranging dogs may have been facilitated, given the frequent unsupervised periods LGDs experience while protecting livestock. To test this, we incorporated free-ranging dogs into our data, together with modern breeds and wolves, and analyzed the data using PHATE[27] (Figure 3B). We observe that Eurasian free-ranging dogs cluster primarily with LGDs. By comparison, free-ranging dogs from other parts of the world, such as the Americas and Africa, are mostly localized to the center of the plot, as previously reported.[11] Phylogenetic, population structure, and admixture analyses (Figures 3C and S12–S14) reveal that dogs from either group but within the same geographic region share similar genomic backgrounds.

To further identify gene flow between free-ranging dogs and LGDs, we conducted an analysis of haplotype sharing (Figure 4A). Consistent with previous analyses, LGDs exhibit more extensive sharing with free-ranging dogs from the same or nearby regions. Breeds of European origin, such as the Maremma sheepdog and the Castro Laboreiro dog, exhibit significant haplotype sharing with all Eurasian free-ranging dog populations. By comparison, LGDs from East Asia primarily exhibit increased haplotype sharing with free-ranging dogs within the same region, although in less proportion and with smaller haplotypes compared to other LGDs, except for the Mongolian Bankhar. While Dutrow
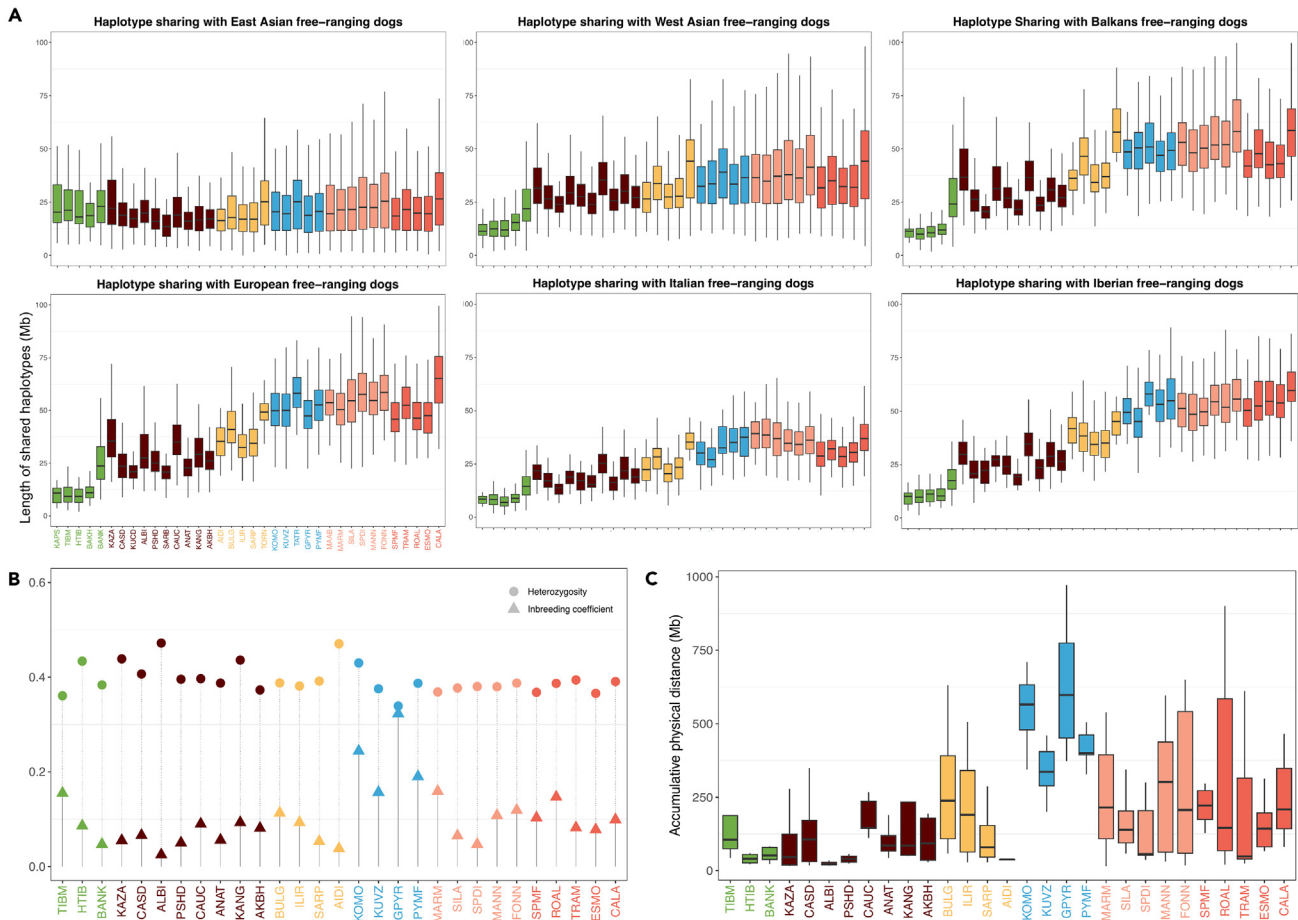
Figure 4. Genomic variability and relationship between LGD and Eurasian free-ranging dog

(A) Haplotype sharing between LGDs and Eurasian free-ranging dogs. Combined haplotype length is displayed on the y axis, and LGD breeds are listed on the x axis, colored according to the defined group.

(B) Graphical representation of expected heterozygosity and inbreeding coefficient for LGD breeds.

(C) The physical length of runs of homozygosity (ROHs) across LGD breeds. Livestock guarding dogs are colored according to their defined geographic group, and the remaining canids are colored gray. The full name of each breed is listed in Figure 1 and Table S3. See also Figures S15 and S16 and Table S3.

et al.[11] first described the association between purebred and regionally linked free-ranging dogs, our study expands this connection, highlighting its prevalence primarily with LGDs and emphasizing its consistency across Eurasia.

Whereas breed clubs and registering bodies forbid dog owners from crossbreeding to dogs from other breeds for the purpose of maintaining traits, such restrictions are not imposed on working landrace populations and, as such, may be challenging to maintain in working dogs frequently left unattended. While the detected signal could stem from free-ranging dogs primarily originating from breeds present in a specific region, we cannot disregard the possibility of ongoing gene flow. Therefore, strict barriers to gene flow between highly skilled and non-specialized dogs might not be essential for preserving specialized skill sets, at least in the case of LGDs.

Genomic diversity and breeding practices in LGDs

Commonly used as working dogs, some LGD breeds are now often or exclusively used as pet dogs, which may have an impact on the genetic diversity of the breed and translate into a loss of guarding skills, at least among breeding lines. To explore this further, we calculated genetic diversity estimates for breeds for which there were a minimum of three individuals (Tables S2 and S3). Our findings align with past studies as we observe no variation in heterozygosity (Table S3 and Figure 4B) when comparing LGDs based on geographical origin.[8,19,20] Furthermore, we used metadata associated with each sample to classify dogs into two categories, working and pet (see Table S4), and averaged inbreeding coefficients for comparison between working and pet dogs. LGDs that are primarily kept as pet dogs display higher inbreeding coefficient (F = 0.18) than those maintained for working purposes (F = 0.08). Lower inbreeding coefficients are likely to be common in working dogs, as pedigree management is less prevalent than in breeds used as pets.[20,28,29]

To further investigate the impact of breeding practices on LGDs, we evaluated genome-wide autozygosity as runs of homozygosity (ROHs, Figures 4C and S15) and genome-wide linkage disequilibrium (LD, Figure S16). In breeds mostly composed of pet dogs, such as the Great

Pyrenees or the Kuvasz, we observe longer ROHs and reduced LD decay which are usually associated with recent inbreeding.[30] Additionally, these breeds exhibit a greater proportion of the genome with ROHs in lower-generation classes (2–64, Figure S15), indicating recent inbreeding events. This highlights clear differences among LGD breeds, probably caused by a recent transition from working landraces to a registered system of pedigree in pet breeds, which often depends on the genetic contributions of a few popular sires.[1] In addition to genomics, behavioral and performance tests are necessary to comprehend the influence of breeding practices on the functional performance ability of LGDs.

Conclusion

LGDs followed humans moving across Eurasia, resulting in distinct regional breeds shaped by human-driven selection to suit specific tasks and environments. Our work unveils two distinct modern lineages and multiple ancestries tracing back to distinct Eurasian ancient dogs within LGDs, akin to other functional groups like sighthounds.[18] We contend that our data are consistent with multiple origins among LGDs, likely driven by a common need in livestock management shared by distinct societies. However, a single origin followed by the complete replacement of East Asian LGDs remains possible. It is essential to understand if similar selection pressures on both lineages resulted in convergent evolution. Shared gene flow among LGD breeds highlights the historical impact of livestock practices, such as transhumance, on shaping these dogs. Notably, as we observe gene flow with free-ranging dogs, our findings inquire about the need for reproductive isolation to maintaining specialized dog abilities. Identifying potential regions under selection is necessary to comprehend the role of genomics, behavior, or a combination of both in the functional ability of LGDs.

Limitations of the study

In this study we explored LGDs' relationships and genomic ancestry across Eurasia, using ancient dogs as reference. However, our study may face limitations due to low sampling in certain breeds and regions, particularly in Central and East Asia. Additionally, the use of a reduced representation of the dog genome (~100 k SNPs) constrained the number of SNPs recovered from ancient genomes, limiting the statistical power of the analyses. Ultimately, obtaining high-coverage whole genomes of LGDs, additional ancient genomes, and including LGD samples from currently underrepresented regions will be crucial to deepen our knowledge on the complex evolutionary history of LGDs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Sample collection and DNA extraction
  - Genome alignment, SNP genotyping and data preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Population structure and ancestry

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110396.

AUTHOR CONTRIBUTIONS

R.G., I.D., L.F., E.A.O., and G.L. conceived and conceptualized the study. R.G., I.D., H.R.G., and E.A.O. provided the DNA samples. D.C.-L., H.G.P., and D.L.D. assembled the modern dataset from new and published sequencing data. D.C.-L. assembled the ancient dataset from published data. D.C.-L. and D.L.D. performed the bioinformatic analysis. All the authors discussed and interpreted the results. D.C.-L. and D.L.D. wrote the original draft with inputs from the other authors. All the authors revised and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Ostrander, E.A., Wayne, R.K., Freedman, A.H., and Davis, B.W. (2017). Demographic history, selection and functional diversity of the canine genome. Nat. Rev. Genet. *18*, 705–720. https://doi.org/10.1038/nrg.2017.67.

2. Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å., and Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature *495*, 360–364. https://doi.org/10.1038/nature11837.

3. Wang, G.-D., Shao, X.-J., Bai, B., Wang, J., Wang, X., Cao, X., Liu, Y.-H., Wang, X., Yin, T.-T., Zhang, S.-J., et al. (2019). Structural variation during dog domestication: insights from gray wolf and dhole genomes. Natl. Sci. Rev. *6*, 110–122. https://doi.org/10.1093/nsr/nwy076.

4. Russell, N. (2015). Neolithic Human-Animal Relations (Groniek), p. 48.

5. Gehring, T.M., VerCauteren, K.C., and Landry, J.M. (2010). Livestock Protection Dogs in the 21st Century: Is an Ancient Tool Relevant to Modern Conservation Challenges? Bioscience *60*, 299–308. https://doi.org/10.1525/bio.2010.60.4.8.

6. Hancock, D. (2014). Dogs of the Shepherds: A Review of the Pastoral Breeds (The Crowood Press).

7. Wang, M.S., Wang, S., Li, Y., Jhala, Y., Thakur, M., Otecko, N.O., Si, J.F., Chen, H.M., Shapiro, B., Nielsen, R., et al. (2020). Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol. Biol. Evol. *37*, 2616–2629. https://doi.org/10.1093/molbev/msaa113.

8. Talenti, A., Dreger, D.L., Frattini, S., Polli, M., Marelli, S., Harris, A.C., Liotta, L., Cocco, R., Hogan, A.N., Bigi, D., et al. (2018). Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. Ecol. Evol. *8*, 2911–2925.

9. Rigg, R. (2001). Livestock Guarding Dogs: Their Current Use Worldwide (The Canid Specialist Group), p. 133.

10. Frantz, L.A.F., Bradley, D.G., Larson, G., and Orlando, L. (2020). Animal domestication in the era of ancient genomics. Nat. Rev. Genet. *21*, 449–460. https://doi.org/10.1038/s41576-020-0225-0.

11. Dutrow, E.V., Serpell, J.A., Ostrander, E.A., Dutrow, E.V., Serpell, J.A., and Ostrander, E.A. (2022). Domestic dog lineages reveal genetic drivers of behavioral diversification. Cell *185*, 4737–4755.e18. https://doi.org/10.1016/j.cell.2022.11.003.

12. Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., Girdland-Flink, L., Lin, A.T., Storå, J., Sjögren, K.-G., Anthony, D., et al. (2020). Origins and genetic legacy of prehistoric dogs. Science *370*, 557–564. https://doi.org/10.1126/science.aba9572.

13. Bergström, A., Stanton, D.W.G., Taron, U.H., Frantz, L., Sinding, M.-H.S., Ersmark, E., Pfrengle, S., Cassatt-Johnstone, M., Lebrasseur, O., Girdland-Flink, L., et al. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. Nature *607*, 313–320. https://doi.org/10.1038/s41586-022-04824-9.

14. Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., and Reich, D. (2023). On the limits of fitting complex models of population history to f-statistics. Elife *12*, e85492.

15. Peng, M.-S., Liu, Y.-H., Shen, Q.-K., Zhang, X.-H., Dong, J., Li, J.-X., Zhao, H., Zhang, H., Zhang, X., He, Y., et al. (2023). Genetic and cultural adaptations underlie the establishment of dairy pastoralism in the Tibetan Plateau. BMC Biol. *21*, 208. https://doi.org/10.1186/s12915-023-01707-x.

16. Leathlobhair, M.N., Perri, A.R., Irving-Pease, E.K., Witt, K.E., Linderholm, A., Haile, J., Lebrasseur, O., Ameen, C., Blick, J., Boyko, A.R., et al. (2018). The evolutionary history of dogs in the Americas. Science *361*, 81–85. https://doi.org/10.1126/science.aao4776.

17. Parker, H.G., Dreger, D.L., Rimbault, M., Davis, B.W., Mullen, A.B., Carpintero-Ramirez, G., and Ostrander, E.A. (2017). Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. Cell Rep. *19*, 697–708. https://doi.org/10.1016/J.CELREP.2017.03.079.

18. Li, W.-L., Liu, Y.-H., Li, J.-X., Ding, M.-T., Adeola, A.C., Isakova, J., Aldashev, A., Peng, M.-S., Huang, X., Xie, G., et al. (2023). Multiple origins and genomic basis of complex traits in sighthounds. Mol. Biol. Evol. *40*, msad158.

19. Janeš, M., Zorc, M., Ferenčaković, M., Curik, I., Dovč, P., and Cubric-Curik, V. (2021). Genomic characterization of the three Balkan livestock guardian dogs. Sustainability *13*, 1–17. https://doi.org/10.3390/su13042289.

20. Bigi, D., Marelli, S.P., Liotta, L., Frattini, S., Talenti, A., Pagnacco, G., Polli, M., and Crepaldi, P. (2018). Investigating the population structure and genetic differentiation of livestock guard dog breeds. Animal *12*, 2009–2016. https://doi.org/10.1017/S1751731117003573.

21. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. Genome Res. *19*, 1655–1664.

22. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967.

23. Patai, R. (1951). Nomadism: Middle Eastern and Central Asian. Southwest. J. Anthropol. *7*, 401–414. https://doi.org/10.1086/soutjanth.7.4.3628514.

24. Spooner, B. (1972). The status of nomadism as a cultural phenomenon in the Middle East. J. Asian Afr. Stud. *7*, 122–131.

25. Silva, R.F. (2000). Transumância no Portugal central: diversidade e organizacão do território [Transhumance in central Portugal: diversity and land organization] (Associacão da defesa do património Arouquense).

26. Coppinger, R., and Coppinger, L. (2001). Dogs: A Startling New Understanding of Canine Origin, Behavior, and Evolution (Scribner).

27. Moon, K.R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D.B., Chen, W.S., Yim, K., Elzen, A.v.d., Hirn, M.J., Coifman, R.R., et al. (2019). Visualizing structure and transitions in high-dimensional biological data. Nat. Biotechnol. *37*, 1482–1492. https://doi.org/10.1038/s41587-019-0336-3.

28. Amiri Ghanatsaman, Z., Adeola, A.C., Asadi Fozi, M., Ma, Y.P., Peng, M.S., Wang, G.D., Esmailizadeh, A., and Zhang, Y.P. (2018). Mitochondrial DNA sequence variation in Iranian native dogs. Mitochondrial DNA A DNA Mapp. Seq. Anal. *29*, 394–402. https://doi.org/10.1080/24701394.2017.1289375.

29. Marinov, M., Teofanova, D., Gadjev, D., Radoslavov, G., and Hristov, P. (2018). Mitochondrial diversity of Bulgarian native dogs suggests dual phylogenetic origin. PeerJ *6*, e5060. https://doi.org/10.7717/peerj.5060.

30. Boccardo, A., Marelli, S.P., Pravettoni, D., Bagnato, A., Busca, G.A., and Strillacci, M.G. (2020). The German Shorthair Pointer Dog Breed (*Canis lupus familiaris*): Genomic Inbreeding and Variability. Animals. *10*, 498. https://doi.org/10.3390/ani10030498.

31. Hayward, J.J., Castelhano, M.G., Oliveira, K.C., Corey, E., Balkman, C., Baxter, T.L., Casal, M.L., Center, S.A., Fang, M., Garrison, S.J., et al. (2016). Complex disease and phenotype mapping in the domestic dog. Nat. Commun. 7, 10460. https://doi.org/10.1038/ncomms10460.

32. Shannon, L.M., Boyko, R.H., Castelhano, M., Corey, E., Hayward, J.J., McLean, C., White, M.E., Abi Said, M., Anita, B.A., Bondjengo, N.I., et al. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. Proc. Natl. Acad. Sci. USA 112, 13639–13644. https://doi.org/10.1073/pnas.1516215112.

33. Dreger, D.L., Rimbault, M., Davis, B.W., Bhatnagar, A., Parker, H.G., and Ostrander, E.A. (2016). Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. Dis. Model. Mech. 9, 1445–1460. https://doi.org/10.1242/dmm.027037.

34. Yang, Q., Chen, H., Ye, J., Liu, C., Wei, R., Chen, C., and Huang, L. (2019). Genetic diversity and signatures of selection in 15 Chinese indigenous dog breeds revealed by genome-wide SNPs. Front. Genet. 10, 1174. https://doi.org/10.3389/fgene.2019.01174.

35. Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Parker, H.G., and Ostrander, E.A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat. Commun. 10, 1489. https://doi.org/10.1038/s41467-019-09373-w.

36. Lobo, D., López-Bao, J.V., and Godinho, R. (2023). The population bottleneck of the Iberian wolf impacted genetic diversity but not admixture with domestic dogs: A temporal genomic approach. Mol. Ecol. 32, 5986–5999. https://doi.org/10.1111/mec.17171.

37. Natoli, E., Bonanni, R., Cafazzo, S., Mills, D.S., Pontier, D., and Pilot, M. (2021). Genetic inference of the mating system of free-ranging domestic dogs. Behav. Ecol. 32, 646–656. https://doi.org/10.1093/beheco/arab011.

38. Pilot, M., Malewski, T., Moura, A.E., Grzybowski, T., Oleński, K., Ruść, A., Kamiński, S., Ruiz Fadel, F., Mills, D.S., Alagaili, A.N., et al. (2015). On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. Proc. Biol. Sci. 282, 20152189. https://doi.org/10.1098/rspb.2015.2189.

39. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.

40. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88.

41. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

42. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. 17, 60. https://doi.org/10.1186/s13059-016-0918-z.

43. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

44. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

45. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

46. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinf. 15, 356. https://doi.org/10.1186/s12859-014-0356-4.

47. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. 3, 861. https://doi.org/10.21105/joss.00861.

48. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak : a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387.

49. Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J.F., Ajmone-Marsan, P., and Colli, L. (2017). BITE: an R package for biodiversity analyses. Preprint at bioRxiv. https://doi.org/10.1101/181610.

50. Wickham, H., Chang, W., and Wickham, M.H. (2016). Package 'ggplot2.' Create elegant data visualisations using the grammar of graphics. Version 2, 1–189.

51. South, A. (2011). rworldmap: A new R package for mapping global data. R J. 3, 35.

52. Felsenstein, J. (1989). PHYLIP: Phylogeny Inference Package. Cladistics 5, 164–166.

53. FigTree. http://tree.bio.ed.ac.uk/software/figtree/.

54. Fitak, R.R. (2021). OptM : estimating the optimal number of migration edges on population trees using Treemix. Biol. Methods Protoc. 6, bpab017. https://doi.org/10.1093/biomethods/bpab017.

55. Neuwirth, E., and Brewer, R. (2014). RColorBrewer: ColorBrewer Palettes. http://ftp.auckland.ac.nz/software/CRAN/doc/packages/RColorBrewer.pdf.

56. Petr, M., Vernot, B., and Kelso, J. (2019). admixr - R package for reproducible analyses using ADMIXTOOLS. Bioinformatics 35, 3194–3195. https://doi.org/10.1093/bioinformatics/btz030.

57. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194, 459–471. https://doi.org/10.1534/genetics.113.150029.

58. Excoffier, L., Lischer, H.E.L., and Schneider, S. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

59. Bertrand, A.R., Kadri, N.K., Flori, L., Gautier, M., and Druet, T. (2019). RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. Methods Ecol. Evol. 10, 860–866. https://doi.org/10.1111/2041-210X.13167.

60. Druet, T., and Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. Mol. Ecol. 26, 5820–5841. https://doi.org/10.1111/mec.14324.

61. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

62. Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat. Rev. Genet. 10, 639–650. https://doi.org/10.1038/nrg2611.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
118 LGD genotypes	This paper	https://osf.io/qp9eg/?view_only=8638cd71a4024e0aba66c40521661e7f
2 wolf genotypes	This paper	https://osf.io/qp9eg/?view_only=8638cd71a4024e0aba66c40521661e7f
Deposited data		
186 LGD genotypes	Wang et al.,[7] Talenti et al.,[8] Dutrow et al.,[11] Parker et al.[17] Hayward et al.,[31] Shannon et al.,[32] Dreger et al.,[33] Yang et al.,[34] Plassais et al.,[35] Lobo et al.[36]	Dryad: https://doi.org/10.5061/dryad.078nc, https://doi.org/10.5061/dryad.0vt4b8gvt, https://doi.org/10.5061/dryad.76hdr7srv, https://doi.org/10.5061/dryad.266k4 | NCBI GEO: GSE121027, GSE213053, GSE96736, GSE83160, GSE123368 | OSF: https://doi.org/10.17605/OSF.IO/N3CDR
1398 non-LGD genotypes	Talenti et al.,[8] Parker et al.,[17] Shannon et al.,[32] Yang et al.[34]	Dryad: https://doi.org/10.5061/dryad.76hdr7srv, https://doi.org/10.5061/dryad.v9t5h | NCBI GEO: GSE96736, GSE121027
780 free-ranging dog genotypes	Shannon et al.,[32] Natoli et al.,[37] Pilot et al.[38]	Dryad: https://doi.org/10.5061/dryad.078nc, https://doi.org/10.5061/dryad.v9t5h, https://doi.org/10.5061/dryad.stqjq2c2q
19 wild canid genotypes	Wang et al.,[7] Talenti et al.,[8] Plassais et al.,[35] Lobo et al.[36]	Dryad: https://doi.org/10.5061/dryad.0vt4b8gvt | NCBI GEO: GSE121027 | OSF: https://doi.org/10.17605/OSF.IO/N3CDR | NCBI SRA: SAMN02921301
Ancient dog sequencing reads	NCBI	NCBI BioProject: PRJEB38079, PRJNA608847, PRJNA319283, PRJEB22026
New guinea singing dog and dingo sequencing reads	NCBI	NCBI BioProject: PRJNA274504, PRJNA263947, PRJNA232497, PRJNA448733
Software and algorithms		
GenomeStudio, Genotyping Module version 2.0	Illumina	https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html
PLINK v1.9	Purcell et al.[39]	https://www.cog-genomics.org/plink/
AdapterRemoval v2.3.2	Schubert et al.[40]	https://github.com/MikkelSchubert/adapterremoval
BWA-ALN algorithm v0.7.17	Li et al.[41]	https://bio-bwa.sourceforge.net/bwa.shtml; RRID:SCR_010910
dedup v0.12.8	Peltzer et al.[42]	
GATK v3.8	DePristo et al.,[43] McKenna et al.[44]	https://software.broadinstitute.org/gatk/; RRID: SCR_001876
mapDamage v2.0.9	Jónsson et al.[45]	http://ginolhac.github.io/mapDamage/; RRID:SCR_001240
ANGSD v0.933	Korneliussen et al.[46]	https://github.com/ANGSD/angsd
umap v0.2.7.0	McInnes et al.[47]	https://umap-learn.readthedocs.io/en/latest/
phateR v1.0.7	Kopelman et al.[48]	https://cran.r-project.org/web/packages/phateR/readme/README.html
ADMIXTURE v1.3.0	Alexander et al.[21]	http://www.genetics.ucla.edu/software/admixture/; RRID:SCR_001263
CLUMPAK	Kopelman et al.[48]	https://tau.evolseq.net/clumpak/
BITE v1.2.0	Milanesi et al.[49]	https://github.com/marcomilanesi/BITE
Admixtools v2.0.0	Maier et al.[14]	https://github.com/DReichLab/AdmixTools; RRID: SCR_018495
ggplot2 v3.4.0	Wickham et al.[50]	https://cran.r-project.org/web/packages/ggplot2/index.html
Rworldmap v1.3.6.	South et al.[51]	https://cran.r-project.org/web/packages/rworldmap/index.html
PHYLIP v3.6	Felsenstein[52]	https://phylipweb.github.io/phylip/
FigTree v1.4.2	FigTree[53]	http://tree.bio.ed.ac.uk/software/figtree/
Treemix v1.13	Pickrell and Pritchard[22]	https://speciationgenomics.github.io/Treemix/
OptM v0.1.5	Fitak[54]	https://cran.r-project.org/web/packages/OptM/readme/README.html
RColorBrewer v1.1.3	Neuwirth and Brewer[55]	https://cran.r-project.org/web/packages/RColorBrewer/index.html

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
admixR v0.9.1	Petr et al.[56]	https://github.com/bodkan/admixr
Beagle v4.1	Browning and Browning[57]	http://faculty.washington.edu/browning/beagle/beagle.html; RRID: SCR_001789
ARLEQUIN v3.5.2.2	Excoffier et al.[58]	http://cmpg.unibe.ch/software/arlequin35/
RZooRoH v0.3.2.1	Bertrand et al.,[59] Druet and Gautier[60]	https://cran.r-project.org/web/packages/RZooRoH/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Diogo Coutinho-Lima (diogofclima5@cibio.up.pt).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw genotype data have been deposited at https://osf.io/qp9eg/?view_only=8638cd71a4024e0aba66c40521661e7f and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

For this study genome-wide SNP data was generated for 118 modern livestock guarding dogs (*Canis lupus familiaris*) and 2 wolf (*Canis lupus*) samples. Additionally, genome-wide SNP or whole-genome resequencing data from ancient dogs and modern dogs and wolves, were retrieved from public databases (see Supplemental Tables; Tables S4, S5, S6, S7, S8, and S9).

METHOD DETAILS

Sample collection and DNA extraction

We genotyped 118 dogs encompassing 23 LGD breeds in order to characterize the diversity of LGDs (Table S4). The Ecogenomics research group (CIBIO-InBio/BIOPOLIS, Vairão, Portugal) obtained blood samples from 82 dogs representing 18 LGD breeds, most of them employed as working dogs in their country of origin. We also genotyped additional samples from two Russian wolves. We extracted genomic DNA from blood samples using the Qiagen DNeasy Blood & Tissue Kit according to the manufacturer's instructions and its concentration was quantified using the Qubit dsDNA HS assay from Invitrogen. The NHGRI Dog Genome Project provided samples from 36 additional dogs representing 9 breeds, located either in the USA or imported from the country of origin. We used standard phenol-chloroform extraction[61] to extract genomic DNA from blood samples and PERFORMAgene swabs to collect saliva samples followed by DNA extraction using the manufacturer's protocols (DNAgenotek, Ontario, Canada). Biological samples were collected under approved animal care and use protocols (National Human Genome Research Institute Animal Care and Use Committee at the National Institutes of Health, protocol GFS-05-01).

Genome alignment, SNP genotyping and data preparation

We used the Illumina CanineHD Whole-Genome Genotyping BeadChip, containing 173,662 variants, to genotype all samples. Genotype calling was performed using GenomeStudio software (Illumina) following Illumina's guidelines (GenomeStudio, Genotyping Module version 2.0 Software Guide, 2016), and exported to Plink format.[39] We combined the newly genotyped samples with previously published Illumina CanineHD SNP array or whole-genome sequence data for 186 dogs representing 21 LGD breeds,[7,8,11,17,31–36] mostly provided by the NHGRI Dog Genome Project and by Lobo et al.[36] This resulted in a combined total of 304 dogs from 36 LGD breeds. We additionally merged the LGD dataset with previously published genotypes from an additional 1398 dogs representing non-LGD modern dog breeds,[8,17,32,34] 780 free-ranging dogs[32,37,38] with a worldwide distribution and 20 gray wolves[7,8,36] (Tables S5, S6, S7, S8, and S9).

    After the initial quality assessment, we obtained an LGD dataset (∼120K SNP markers) and performed additional filtering with PLINK v1.9.[39] We filtered out markers with more than 10% missing data (GENO = 0.1), and minor allele frequency (MAF = 0.01) of <0.01, as well as markers on sex chromosomes. Individuals with a missing rate (MIND = 0.1) > 10% were excluded. We used the "–genome" function of PLINK v1.9[39] to calculate relatedness between each pair of individuals. Closely related individuals (identity-by-descent ≥0.5) were removed. We applied

pruning based on linkage disequilibrium (LD) to analyses that are sensitive to LD. Pruning was performed using an r^2 threshold of 0.8, with a sliding-window size of 50 SNPs. The window was shifted and recalculated every 10 SNPs.

Publicly available whole-genome sequences of ancient DNA (aDNA) of 22 dogs were trimmed and adapters removed with AdapterRemoval v2.3.2.[40] Processed reads were mapped against the dog reference genome (CanFam3.1) using the alignment tool, BWA-ALN algorithm v0.7.17.[41] Duplicate reads and reads that mapped to multiple locations in the reference genome were discarded using dedup v0.12.8.[42] To improve the local mapping of reads that span indels, we used GATK v3.8[43,44] to locally realign reads toward minimising the number of mismatches around the indels for each of the samples. We used mapDamage v2.0.9[45] to assess aDNA damage patterns in the ancient samples and additionally to rescale the quality scores of bases inferred to be affected by DNA damage.

Pseudo-haploid calling was performed on the ancient dog samples using the -doHaploCall utility in ANGSD v0.933.[46] During the pseudo-haploid calling, random bases were sampled with a minimum mapping quality of 30 and a minimum base call of 20. Only biallelic SNPs overlapping with the Illumina CanineHD Whole-Genome Genotyping BeadChip coordinates were retained. The ANGSD output was converted to PLINK format with the haploToplink utility from ANGSD. Because the genotypes of the ancient dogs are in pseudo-haploid form, we "pseudo-haploidized" the modern-day data by randomly sampling a single allele at each site, using a custom-made script. The PLINK dataset was merged with the LGDs, and publicly available data of 4 New Guinea singing dogs and 1 Dingo.[35] The inclusion of the latter served to represent a domestic dog lineage originating from Southeast Asia, for which ancient data is not currently available. The data was filtered for linkage disequilibrium by removing SNPs with an R^2 value greater than 0.5 with any other SNP within a sliding window of 50 SNPs and advanced by 10 SNPs each time. This filtering resulted in the retention of ~50k SNPs.

In order to perform the desired comparative analyses, we generated five distinct datasets that we term Subsets: the "Guardian Subset"; "Free-ranging Subset"; "Breed Subset"; "BigDog Subset"; and "Ancient Subset". The relevant composition and metrics of these subsets are described in Table S2. To prevent sampling bias, the LGD sample sizes in the subsets Free-ranging, Breed, and BigDog were limited to a maximum of 10 individuals per breed. For the same reason, sample sizes for free-ranging dogs and modern breeds were also restricted to a maximum of 10 individuals per geographic region and breed, respectively, in the Free-ranging and Breed subsets. As all Italian free-ranging dogs were sourced from a single population, which included related individuals, we computed pairwise relatedness for this dataset and selected 10 unrelated dogs (identity-by-descent <0.5) for downstream analysis to mitigate possible bias. To enhance comprehension of the analyses and their outcomes, we categorized LGD breeds based on their geographical origins and clustering methods described below. The detailed information regarding specific breed, the corresponding dog numbers, and their respective group classifications can be found in Table S3.

QUANTIFICATION AND STATISTICAL ANALYSIS

Population structure and ancestry

To investigate the genetic structure between LGDs we carried out a Principal Components Analysis (PCA) with PLINK v1.9[39] using the "–pca" command. In the Ancient subset, the ancient dogs and modern New Guinea Singing dogs and Dingo were projected onto the LGD PCs using the "–pca-clusters" of PLINK v1.9.[39] The "umap" v0.2.7.0[47] R package was applied to generate Uniform Manifold Approximation and Projections (UMAP) for the Guardian and Free-ranging subsets. The first 20 principal components (PCs) were used as input for BigDog subset to generate the embedding using the PHATE method ("phateR" v1.0.7).[27]

To access population structure and estimate ancestry proportions among LGD breeds, we employed the ADMIXTURE v1.3.0[21] software. The analysis was performed individually for both the Free-ranging and Ancient subsets. This separation was necessary because including every dog in each analysis would have led to a confusing outcome. Population divisions (K) ranged from two to N (where N is the number of breeds that have at least two individuals sampled), with each K being assessed using 100 bootstraps, in 5 independent iterations. The "–cv" flag was used to find the best partition model to split the populations (lowest cross-validation error), according to Admixture best practices. We used CLUMPAK[48] to truncate the different runs of each K and the R package "BITE" v1.2.0[49] was used to plot the results, using the "membercoeff.circos" and "membercoeff.plot" functions.

To identify the most appropriate models to explain the ancestry proportions of each LGD breed, we employed a rotation approach through Admixtools v2.0.0[14] in the Ancient subset. We tested all possible one, two and three-source models for each breed and ranked based them on their p-values. Models with inferred ancestry proportions <0 or >1 were discarded. To accommodate the pseudo-haploid data, qpAdm was executed with the "allsnps = TRUE" and "inbred = TRUE" parameters. The ancestry sources consisted of a core group of ancient dogs along with a modern New Guinea singing dog, selected as follows: America (AL3194, 4.1 ky), Arctic (CGG6, 9.5 ky), Karelia (OL4061, 10.9 ky), Lake Baikal (OL4223, 6.9 ky), Iran Chalcolithic (AL2571, 4.8 ky), Samara Steppe (C5, 3.8 ky), Germany (CTC, 4.7 ky), and Southeast Asia (NGSD01, modern). We selected these sources to account for ancestry patterns associated with their geographical/temporal origin, which are also commonly featured in top-ranking ancestry models, as described by Bergstrom et al.[12] The coyote (*Canis latrans*)[32] was included as an outgroup (Table S9). The output was plotted on R using packages "ggplot2" v3.4.0[50] and "Rworldmap" v1.3.6.[51]

Phylogenetic relationships and gene flow

To investigate the phylogenetic relationships within LGDs and their connections to free-ranging dogs and modern breeds, we constructed multiple neighbour-joining trees with the gray wolf as the root. Genetic distances between individuals were estimated using the "–distance" function of PLINK v1.9[39] and the "1-ibs," "square," and "flat-missing" modifiers. The dataset was bootstrapped 100 times to ensure statistical power using a custom script (https://github.com/pdroslva84/Plink_IBS_bootstraps) that subsamples the original dataset. We created

neighbour-joining phylogenies for every subset, using the PHYLIP v3.6 software package[52] and later depicted in FigTree v1.4.2.[53] In the Breed and BigDog subset trees, 1000 bootstraps were performed instead of 100 to increase the accuracy of the analysis. Furthermore, when the confidence support reached a threshold of 50% or higher, we collapsed the clades based on the corresponding breeds.

To search for gene flow between LGDs we performed an extended analysis of the relationships among breeds in the Guardian subset using Treemix v1.13.[22] Allele frequencies and missing data for each marker were calculated in PLINK v1.9[39] using the "–freq" and "–missing" options and the individuals were grouped according to their population using the "–within" function. Subsequently, an analysis with Treemix[22] was conducted, in which the number of admixing events ranged from 0 (absence of gene flow) to N, where N is the number of breeds with three or more individuals. To account for the fact that nearby SNPs are not independent, we used the "–k" flag to group SNPs in windows of 300, a length that exceeds the known extent of LD in modern dogs.[34] R package "OptM" v0.1.5[54] was used to estimate the optimal number of migration edges on the population trees. The dataset was bootstrapped 100 times to ensure statistical power using a script from BITE v1.2.0.[49] The gray wolf was used as the root for the maximum likelihood tree. The "RColorBrewer" v1.1.3 package of the R software[55] was used to plot the trees and residual errors for the optimal migration identified.

To test for significant admixture among the LGDs we calculated D-statistics and f3-statistics in populations with three or more individuals using the "admixR" v0.9.1[56] package. The D-statistics and f3-statistics were computed with $D(P_1,P_2,P_3,O)$ and $f_3(P_1,P_2,O)$, respectively, with the gray wolf as outgroup. Haplotype sharing was determined by identity-by-descent (IBD) estimations among individuals and performed on the Free-ranging subset with Beagle v4.1.[57] The dataset was analyzed in windows of 1,000 SNPs with an overlap of 25 SNPs. Haplotype sharing was considered significant when median values were above the 95th percentile of all breed pairs. Boxplots of haplotype sharing distributions and the D-statistics were plotted using R package "ggplot2" v3.4.0.[50]

Genomic diversity

Genomic variability among LGD breeds was assessed in the Guardian subset. We used the ARLEQUIN v3.5.2.2[58] software package to query diversity and population differentiation. The genetic differentiation between breeds was calculated with the aid of pairwise mean F_{ST}.[62] The significance for the F_{ST} was tested using the typical criteria of ARLEQUIN v3.5.2.2[58] (10,000 permutations, with significance set at $\alpha = 0.05$). We used the function "–het" of Plink v1.9[39] to calculate the inbreeding coefficient (F) for each individual dog and breed-level F was calculated by averaging individual F. Additionally, individual F values were averaged for comparison between samples classified as working dogs and pet dogs. Dogs labeled: "Assumed working dog" and "Possible pet dog" were considered as working and pet dogs, respectively. Samples classified as "Unknown" were not considered for this analysis.

To find fragments in homozygosity (ROH) across the genome we used "–homozyg-snp" and "–homozyg-kb" in PLINK v1.9.[39] The software aligned a moving window of 50 SNPs across the genome of each individual to identify long contiguous ROHs with allowance for one heterozygous site, five missing calls per window, and a minimum length of 200 kb for an ROH. Additionally, we characterized the individual ROHs using the R package RZooRoH v0.3.2.1.[59] RZooRoH adopts a model-based approach to classify the length of ROHs into generation classes, providing information on the timing of inbreeding events. Each class is associated with a rate (R_k) equal to the size of the inbreeding loop in generations.[60] Based on the density of the SNP panel, we applied a MixKR model comprising 12 classes (with R_k equal to 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 and 2048).

Finally, we used PLINK v1.9[39] to calculate pairwise LD by computing the genotype correlation coefficient (r^2) using the command "–r2 –ld-window-kb 500 –ld-window-r2 0". This was performed for breeds with five or more individuals. To prevent potential bias due to unequal sample sizes all breeds were limited to a maximum of 10 individuals if more than 10 dogs had been sampled. All analyses were plotted through the R package "ggplot2" v3.4.0.[50]