

Title: Canine transmissible venereal tumor genome reveals ancient introgression from coyotes to arctic sled dogs

Xuan Wang^{1,2§}, Bo-Wen Zhou^{1,2§}, Ting-Ting Yin^{1,2}, Fang-Liang Chen³, Ali Esmailizadeh⁴, Melissa M. Turner⁵, Andrei D. Poyarkov⁶, Peter Savolainen⁷, Guo-Dong Wang^{1,2,8*}, Qiaomei Fu^{9,10*}, Ya-Ping Zhang^{1,2,8*}

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China

²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, 650223, China

³Kunming Police Dog Base of the Ministry of Public Security, Kunming, 650204, China

⁴Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, PB 76169-133, Iran

⁵Department of Forestry and Environmental Resources, Fisheries, Wildlife, and Conservation Biology Program, North Carolina State University, Raleigh, NC 27695, USA

⁶Severtsov Institute of Ecology and Evolution, Russian Academy of Science, Leninskiy prospect, 33, Moscow, 119071, Russia

⁷Department of Gene Technology, KTH-Royal Institute of Technology, Science for Life Laboratory, Tomtebodavägen 23A, Solna, 17165, Sweden

⁸Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, 650223, China

⁹Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing, 100044, China

¹⁰Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, 100044, China

[§]Contributed equally.

*Correspondence: Guo-Dong Wang^a, Qiaomei Fu^b, or Ya-Ping Zhang^c

^aE-mail: wanggd@mail.kiz.ac.cn;

^bE-mail: fuqiaomei@ivpp.ac.cn;

^cE-mail: zhangyp@mail.kiz.ac.cn

38 Dear Editor,

39 Ancient genome-sequencing studies have extensively refined our understanding of the
 40 genetic histories and adaptive evolution of humans¹, as well as the domestic process
 41 of livestock and crops². Recent studies about ancient and modern canids all indicate
 42 that canids have intricate histories, and lots of canine populations disappeared during
 43 evolution³⁻⁶. Thus, ancient genomes from canine fossils are in high demand to clarify
 44 canine evolution and the processes leading to dog domestication. Canine transmissible
 45 venereal tumor (CTVT), the oldest known somatic cell line, is a living fossil,
 46 originating from cancer cells transmitted from a host to other canids during the mating
 47 process⁷. Clonal origin analyses hints that the original dog infected with CTVT
 48 (CTVT founder) came from an ancient sled dog or wolf population⁷⁻⁹. However, the
 49 genetic composition of the CTVT founder is still not clear.

50

51 In order to explore this issue, we applied whole genome sequencing (WGS) to two
 52 CTVT samples, their corresponding hosts, and 24 additional canids (Supplementary
 53 Note). Combined with published WGS data of two CTVT samples and high quality
 54 canine WGS data, we constructed a data set containing WGS data of four CTVT
 55 samples a 169-individual reference panel composed of worldwide gray wolves (*Canis*
 56 *lupus*), dogs (*Canis lupus familiaris*), coyotes (*Canis latrans*) and golden jackals
 57 (*Canis aureus*) (Supplementary Note, Table S1).

58

59 We firstly developed a pipeline *transmissible tumor genotyper* to obtain the per-site

60 genotype for each tumor sample (Supplementary Note). All sites with the same
61 genotype among four CTVT samples were used for ascertainment. Next, we
62 successfully ascertained 18.9M biallelic single nucleotide polymorphisms (SNPs) for
63 the CTVT common genotype using 24.5M SNPs from the reference panel
64 (Supplementary Note). We denote these data Set A. SNPs from the CTVT data in set
65 A could be treated as the founder's germline polymorphisms because they showed a
66 typical germline mutation signature (Supplementary Note, Figure S5).

67
68 We utilized population phylogeny analysis, principal component analysis (PCA) and
69 outgroup f_3 (X, CTVT founder; Outgroup) statistical analysis to assess the genetic
70 relationship between the CTVT founder and modern samples in set A (Supplementary
71 Note). Our results reveal that the CTVT founder was more closely related to
72 present-day arctic sled dogs than to any other populations (Figure S6-8), in
73 accordance with very recent results^{7,8}. However, ADMIXTURE analysis showed that
74 the CTVT founder also possessed an ancestral component found predominantly in
75 non-dog populations, a result that we do not observe for any arctic sled dog
76 (Supplementary Note, Figure 1A). Moreover, the CTVT founder did not cluster
77 tightly with arctic sled dogs in the PCA analysis (Figure S7). These results imply that
78 the CTVT founder belonged to a previously unknown arctic dog population that is not
79 represented in the reference panel.

80

81 To further investigate whether the CTVT founder had introgression from a non-dog

82 population, we ascertained set A using the Andean fox (*Lycalopex culpaeus*) genome
83 (Supplementary Note), forming set B. Next, we tested every non-dog group in set B
84 as candidate introgressor using D statistics of the form $D(\text{CTVT founder, Pop2;}$
85 Introgressor, Andean Fox), where Pop2 was tested using all other present-day groups
86 (Supplementary Note, Table S2). Only coyotes were found to be possible candidate
87 introgressors, showing significant ($Z > 4.2$) positive D statistics for all Pop2
88 populations except the New World wolves (Figure 1B). We then tested whether any
89 domesticated dog populations could act as the candidate introgressor for the CTVT
90 founder. Consistent with above results, we find that present-day arctic sled dogs
91 showed significant ($Z > 9.6$) positive $D(\text{CTVT founder, Pop2; Sled dog1, Andean Fox})$
92 statistics for all Pop2 populations (Figure 1C). D statistics suggest that the CTVT
93 founder was an arctic sled dog bearing gene flow from a population most closely
94 related to coyotes. We did not find strong evidence of coyote introgression into
95 present-day arctic sled dogs ($D(\text{Sled dog1, Pop2; Coyote, Andean Fox}) \sim 0$). Finally,
96 we used high frequency coyote private alleles to genotype the CTVT founder, in order
97 to manually check whether there was any evidence of continuous coyote-related
98 segments introgressed into the CTVT founder (Supplementary Note) and found one
99 segment longer than 3Mbps on chromosome 6, and several shorter segments on
100 chromosomes 6, 25 and 26. (Figure 1D). This finding helps us to reject the hypothesis
101 that our results were caused by somatic reverse mutations, rather than introgression.
102 29 protein coding genes are located within these segments in total. Among them, we
103 found four tumor-related genes^{10, 11} and a domestication gene¹² *AMY2B* (Figure 1D).

104 The presence of coyote genotypes at tumor-related genes may contribute to the
105 genetic susceptibility to CTVT tumorigenesis. The coyote genotypes at *AMY2B* may
106 lead to low adaptation on a diet rich in starch and the disappearance of the population
107 represented by the CTVT founder in agrarian age.

108

109 In conclusion, our detailed analyses reveal that the CTVT founder came from an
110 arctic sled dog population that possessed introgression from a population related to
111 coyotes, a result that was not known in previous studies. Considering the habitat of
112 coyotes in North America¹³, we propose two hypotheses: 1) The CTVT founder lived
113 in the arctic region of North America. 2) The CTVT founder lived in the arctic region
114 of the Far East, where arctic dogs possessing the introgressed segments migrated
115 through the Bering Strait in an unknown period. Hence, an ancient story of canine
116 admixture is hidden in the genome of a living fossil, the CTVT. To further test our
117 hypotheses of ancient admixture and to better understand the detailed evolutionary
118 history of dogs from the arctic region and Americas, it is crucial to acquire ancient
119 samples in these regions in future work.

120

121 **Acknowledgments**

122 We thank Melinda A. Yang for her valuable comments, Newton O. Otecko for
123 providing dog samples from Kenya, Ming-Shan Wang for collecting dog samples
124 from Sri Lanka, Shi-Fang Wu for collecting other dog samples, Qi-Jun Zhou for
125 collecting CTVT samples and laboratory work, and the BIG Data Center in Beijing

126 Institute of Genomics, Chinese Academy of Sciences (CAS) for usage of their
 127 high-performance computing platform. This work was supported by the Strategic
 128 Priority Research Program (B) (XDB13000000) of the CAS, National Natural Science
 129 Foundation of China (91531303, 91731303, 41672021, and 41630102). Q.F. is
 130 supported by CAS (XDA19050102, QYZDB-SSW-DQC003, and XDPB05), and the
 131 Howard Hughes Medical Institute (grant number 55008731). G.D.W. is supported by
 132 the Youth Innovation Promotion Association, CAS.

133

134

135 **References**

- 136 1 Yang MA, Fu Q. Insights into Modern Human Prehistory Using Ancient Genomes.
 137 *Trends in Genetics* 2018; **34**:184-196.
- 138 2 Marciniak S, Perry GH. Harnessing ancient genomes to study the history of human
 139 adaptation. *Nature Reviews Genetics* 2017; **18**:659.
- 140 3 Ostrander EA, Wayne RK, Freedman AH, Davis BW. Demographic history,
 141 selection and functional diversity of the canine genome. *Nature Reviews Genetics*
 142 2017; **18**:705.
- 143 4 Freedman AH, Wayne RK. Deciphering the Origin of Dogs: From Fossils to
 144 Genomes. *Annual Review of Animal Biosciences* 2017; **5**:281-307.
- 145 5 Wang G-D, Zhai W, Yang H-C *et al.* Out of southern East Asia: the natural history
 146 of domestic dogs across the world. *Cell Research* 2015; **26**:21.
- 147 6 Botigué LR, Song S, Scheu A *et al.* Ancient European dog genomes reveal

- continuity since the Early Neolithic. *Nature Communications* 2017; **8**:16082.
- 7 Murchison EP, Wedge DC, Alexandrov LB *et al.* Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* 2014; **343**:437-440.
- 8 Decker B, Davis BW, Rimbault M *et al.* Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome research* 2015; **25**:1646-1655.
- 9 Murgia C, Pritchard JK, Kim SY, Fassati A, Weiss RA. Clonal origin and evolution of a transmissible cancer. *Cell* 2006; **126**:477-487.
- 10 An O, Dall'Olio GM, Mourikis TP, Ciccarelli FD. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic Acids Research* 2016; **44**:D992-D999.
- 11 Zhao M, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Research* 2013; **41**:D970-D976.
- 12 Arendt M, Cairns KM, Ballard JWO, Savolainen P, Axelsson E. Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity* 2016; **117**:301.
- 13 vonHoldt BM, Cahill JA, Fan Z *et al.* Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances* 2016; **2**.

168 **Figure Legends**

169 **Figure 1.** Analyses of introgression into the CTVT founder. **(A)** ADMIXTURE
 170 analysis result ($K=3$) for jackal, coyote, gray wolf and dog populations with 163K
 171 SNPs pruned by linkage disequilibrium. Colors indicate different ancestral
 172 components. **(B)** $D(\text{Pop1}, \text{Pop2}; \text{Coyote}, \text{Andean Fox})$, with the Z-score given on the
 173 x axis. When the CTVT founder is Pop2, $Z>3$ for all Pop1 groups except the
 174 New-World wolf which is a descendant of coyotes. **(C)** $D(\text{CTVT founder}, \text{Pop2}; \text{Pop3},$
 175 $\text{Andean Fox})$, with the Z-score given on the x axis. When present-day arctic sled dogs
 176 and coyotes are Pop3, $Z>3$ scores for all Pop2 groups. **(D)** Introgressed segments (red
 177 regions) traced by high frequency coyote private allele positions (blue regions), see
 178 Supplementary Note. Genes within introgressed segments were labeled alongside
 179 ideogram of chromosomes. Copy number loss regions on chromosome 6 were
 180 truncated for clarity.

