



LETTER TO THE EDITOR

# Canine transmissible venereal tumor genome reveals ancient introgression from coyotes to pre-contact dogs in North America

Cell Research (2019) 29:592–595; <https://doi.org/10.1038/s41422-019-0183-2>

Dear Editor,

Canine transmissible venereal tumor (CTVT), the oldest known somatic cell line, is a living fossil of the original founder, transmitted from host's cancer cells to other canids during the mating process.<sup>1</sup> Since it was shown ten years ago that living cells from an ancient host could be transmitted among canids, the origin of CTVT has been studied continuously.<sup>2</sup> Recent comparison of the CTVT genetic data with a more comprehensive canine reference panel including pre-contact dogs (PCDs) from North America argued that the CTVT founder (the original canid infected with CTVT) is the closest detectable lineage to PCDs, and that this clade underwent introgression from wild canids in North America.<sup>3</sup>

However, previous studies may not take into account several potential biases in the genotyping methods for CTVT samples and the strategy for collecting loci (Supplementary information, Note), and the genetic ancestry of the ancient founder of CTVT is still unknown. To address these biases and the unknown issue, we collected new CTVT samples and modern canids, and used a newly developed tool and a refined strategy for analyses.

We generated a high-quality reference panel representing canine genetic diversity with the highest spatial and temporal resolution to date, including whole-genome sequencing (WGS) data of 22 newly collected canids and 81 published ancient and modern worldwide dogs and wild canids (Supplementary information, Table S1 and Note). Then, 24.1 M single nucleotide polymorphisms (SNPs) were called from these samples (Supplementary information, Methods), composing a dense reference of makers.

We sequenced two new CTVT samples collected in Kunming, China (Supplementary information, Fig. S1, Methods and Note), and included WGS data of three previously published CTVT samples from Australia, Brazil,<sup>1</sup> and Gambia<sup>3</sup> (Supplementary information, Fig. S2 and Table S2). Together, these five CTVTs from four continents allow us to exclude lineage-specific somatic mutations.

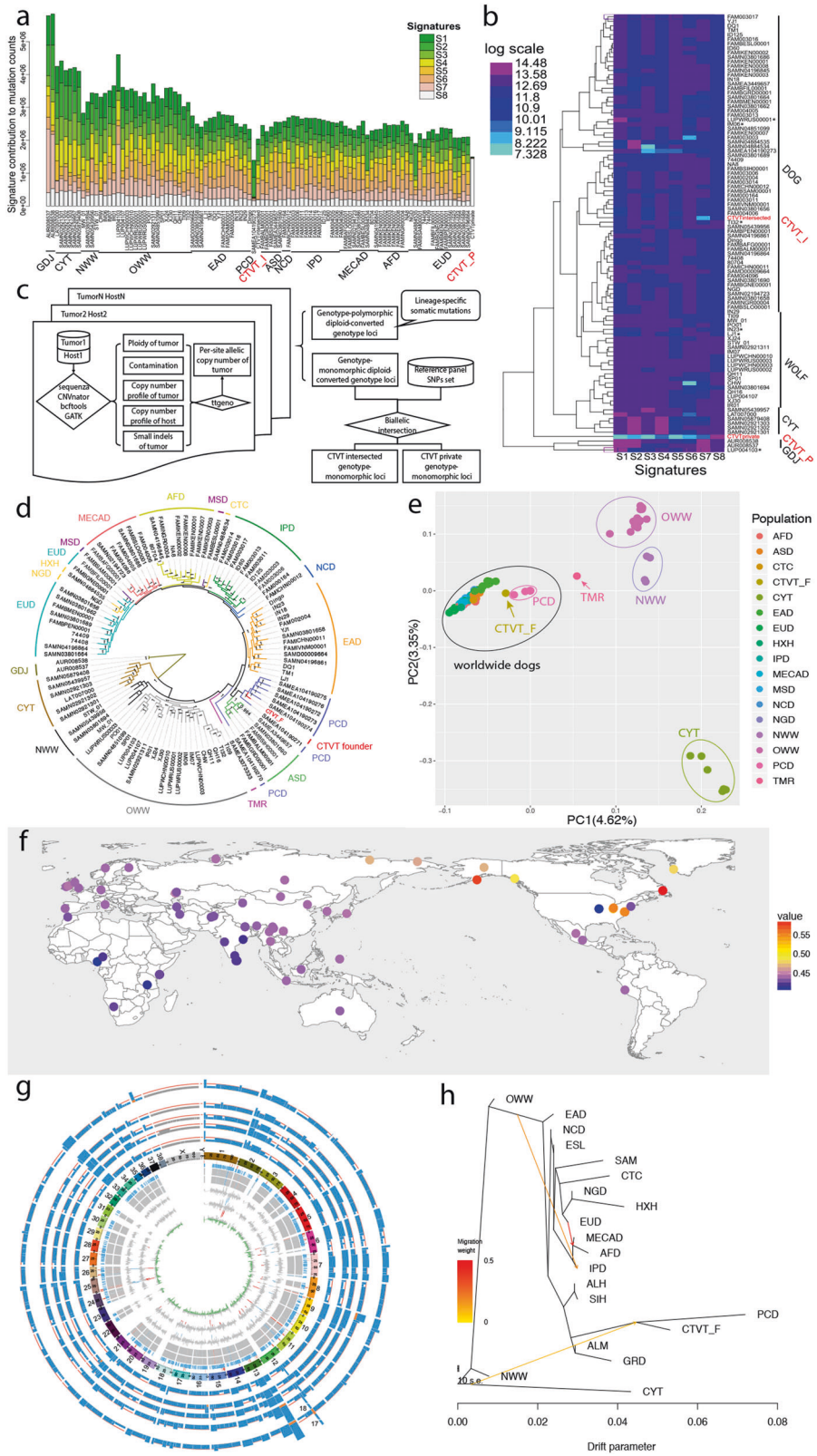
As chromosomal instability is considered the predominant somatic mutational type in the tumorigenesis of CTVT,<sup>4</sup> the copy number variation (CNV) profile is necessary to determine the genotype at local sites. Thus, we developed a *transmissible tumor genotyper* (*ttgeno*), the first genotyping tool designed specifically to analyze WGS data from paired transmissible tumors and their hosts, to obtain per-site allelic copy number of the tumor (Supplementary information, Methods). This tool simultaneously takes into account the ploidy, contamination, local copy number states of both host and tumor, and small indels in the tumor, and removes the subclonal factor, as previous studies have shown that CTVT is almost homogeneous.<sup>1,4</sup> We genotyped each CTVT using this tool, and obtained successful genotyping rates of 95.5%–97.4%.

The genotyped CTVT genome is composed of a mix of different mutations, including systematic errors, alleles inherited by the founder, lineage-specific somatic mutations, and earlier somatic mutations. Assuming a single origin for CTVT,<sup>1</sup> lineage-specific somatic mutations can be distinguished from genotype-polymorphic mutations using multiple worldwide CTVT samples. That is, alleles inherited by the founder and earlier somatic mutations should be genotype-monomorphic among CTVT samples. We found ~1.7 G genotype-monomorphic sites, allowing one missing CTVT sample at each site. Another 2.9 M sites were genotype-polymorphic loci among the five CTVTs, allowing two missing CTVT samples at each site. We used the genotype-polymorphic sites to assess the relationship between these five CTVTs (Supplementary information, Fig. S3) and excluded these from subsequent analyses. We found that of the ~1.7 G sites that were genotype-monomorphic in five CTVT samples, 17.4 M sites (2 M non-ref alleles) are biallelic polymorphic in the reference panel, while 1.5 M sites were private to CTVT samples. Using an assessment strategy based on mutation signatures, we demonstrated that the 17.4 M sites are inherited germline SNPs (Fig. 1a–c; Supplementary information, Figs. S4–7 and Note). Thus, we treated the 17.4 M sites as direct descendants of the suppositional ancient canid “the CTVT founder” and use these sites for subsequent population genetic analyses.

We utilized population phylogeny analysis (Fig. 1d and Supplementary information, Figs. S8–12), principal component analysis (PCA, Fig. 1e and Supplementary information, Fig. S13) and outgroup  $f_3$ (CTVT founder, Pop2; Coyote) statistics<sup>5</sup> (Fig. 1f) to assess the genetic relationship between the CTVT founder and samples in the reference panel (Supplementary information, Note). Our results reveal that the CTVT founder was more closely related to PCDs than to any other populations (Fig. 1d–f), similar to Ní Leathlobhair et al.,<sup>3</sup> but disagreeing on some details (Supplementary information, Note). Meanwhile, the topology of phylogeny (Fig. 1d) and the spacial position of the PCD/CTVT founder cluster in PCA (Fig. 1e) all suggest that introgression from wild canids may exist in this PCD clade. *ADMIXTURE*<sup>6</sup> analysis shows that the CTVT founder also possessed ancestral components found predominantly in wild canids (Supplementary information, Fig. S14 and Note).

To further investigate whether the CTVT founder and PCDs experienced introgression from a population distantly related to dogs, we calculated *D*-statistics<sup>5</sup> to test whether significant asymmetry (positive *D* value,  $Z > 3$ ) exists between Pop1 and Pop2 using the form  $D(\text{Pop1}, \text{Pop2}; \text{Candidate Introgressor}, \text{Andean Fox})$  (Supplementary information, Table S3 and Note). We tested every non-dog group as a candidate introgressor for the CTVT founder using  $D(\text{CTVT founder}, \text{Pop2}; \text{Introgressor}, \text{Andean Fox})$ , where Pop2 was each canid population in turn (Supplementary information, Fig. S15). Only coyotes were found to

Received: 13 May 2018 Accepted: 15 May 2019  
Published online: 3 June 2019



**Fig. 1** **a** Bar plot showing the contributions of the signatures to genome sample mutation counts. **b** Heatmap showing exposures in each genome sample. Samples are grouped according to their levels of exposure to the signatures, as can be seen in the dendrogram on the left. Samples labeled by asterisks are exceptions in the group. **c** A flowchart of the pipeline from genotyping of each tumor to classification of loci. **d** Approximate maximum-likelihood tree. All samples are colored according to the geographical groups. Node labels indicate bootstrap values. **e** Principal component analysis of 102 individuals (excluding golden jackals). Different geographical groups are colored differently. **f** Heatmap of outgroup  $f_3$ -statistics of the form  $f_3(\text{CTVT founder, Pop2; Coyote})$ , where Pop2 represents 69 dogs in the reference panel. Higher  $f_3$  values (warm-toned points) indicate increased genetic drift shared between samples, and therefore higher genetic similarity between the CTVT founder and Pop2 sample. **g** CNV profiles of five CTVTs and coyote introgressed regions in genome of the CTVT founder. Starting from the outer circle and moving to inner circles, each circle shows copy number profiles for T.KM2, T.KM1, T.609, T.79, and T.24, a chromosome ideogram, loci sharing with coyote diagnostic alleles (orange), coyote diagnostic allele positions (blue), two haplotypes of *RFMix* results (gray, Arctic sled dogs' ancestry; blue, New World wolves' ancestry; red, coyotes' ancestry),  $f_{DM}$ -statistics in 500 kb windows by 250 kb step size for sliding window (blue, top 1% negative windows),  $f_{D}$ -statistics in 500 kb windows by 250 kb step size for sliding window (red, top 1% negative windows),  $D$ -statistics in 500 kb windows by 250 kb step size for sliding window (green, negative value). The red reference axis in the 5 outer circles represents the status of diploidy. Data are plotted using *circos*.<sup>15</sup> **h** The maximum-likelihood graph based on *TreeMix* with  $m = 3$ . The scale bar shows ten times the average standard error of the entries in the sample covariance matrix. Weighted colorful arrows indicate relative migration ratio and direction. GDJ, golden jackals; CYT, coyotes; NWW, New World wolves; OWW, Old World wolves; TMR, Taimyr wolf; PCD, pre-contact dogs; ASD, Arctic sled dogs; EAD, East Asia dogs; NCD, Northern China dogs; IPD, India Peninsula dogs; MECAD, Middle East and Central Asia dogs; AFD, African dogs; MSD, mixed sled dogs; EUD, European dogs; NGD, Newgrange dog; HXH, Herxheim dog; CTC, Cherry Tree Cave dog; SIH, Siberian Husky; ALH, Alaskan Husky; ALM, Alaskan Malamute; GRD, Greenland dog; ESL, East Siberian Laika; SAM, Samoyed; CTVT\_I, CTVT intersected with panel's SNPs; CTVT\_P, CTVT private alleles; CTVT\_F, the CTVT founder

be a robust candidate introgressor. Coyotes from Monterey showed significantly positive  $D$ -statistics for most Pop2 populations except the other coyotes, New World wolves, and PCDs ( $Z > 3.7$ ). Similar to previous analysis,<sup>3</sup> two PCDs (i.e., Port au Choix, Weyanoke Old Town) showed significantly positive  $D(\text{CTVT founder, Pop2; PCD, Andean Fox})$  statistics for all Pop2 populations ( $Z > 46$ ), indicating the close relationship between the CTVT founder and PCDs in our panel. Taken together, the CTVT founder is likely an ancient American dog with introgression from populations carrying ancestry related to coyotes from the Monterey area, California, and Alabama. We also tested whether other dogs (Pop1) underwent introgression from coyotes by using  $D(\text{Pop1, Pop2; Coyote, Andean Fox})$ , where Pop2 was tested using all other groups in turn (Supplementary information, Fig. S16). We found no evidence of introgression from coyotes in any dog population except PCDs and the CTVT founder. Due to the CTVT founder's high coverage, we used it as a surrogate for PCDs to test whether any other canids carry ancestry from PCDs (Supplementary information, Fig. S17). Only Arctic sled dogs (ASDs) in North America show more similarity to PCDs, followed by Siberian and Alaskan huskies. However, whether asymmetric  $D$ -statistics indicate introgression from closely related populations or an inheritance relationship cannot be determined without high-density sampling of ancient and modern PCDs and ASDs over a broad geographical region and time frame.

To confirm our result of introgression from coyotes to the CTVT founder shown by  $D$ -statistics analyses, we utilized the coyote-specific diagnostic alleles,<sup>7</sup>  $f_{D}$ -statistics,<sup>8</sup> and  $f_{DM}$ -statistics<sup>9</sup> in sliding windows, as well as *RFMix*<sup>10</sup> to infer the local ancestry in the genome of the CTVT founder (Fig. 1g and Supplementary information, Note). We found that the results were consistent using these methods, with several regions introgressed from coyotes. The introgression rates were estimated to vary in range of 0.9%–2.6% using different methods (Supplementary information, Tables S4–5 and Note).

We used *TreeMix*<sup>11</sup> to investigate the genetic relationship between the CTVT founder, PCDs, other ancient and present-day canids (Supplementary information, Figs. S18–21 and Note). We visualized the matrix of residuals (Supplementary information, Fig. S18b) to determine how the estimated genetic relationship between each pair of canids fits the model. We found three candidate admixture events: (1) between coyotes and the PCD/CTVT founder, (2) between Siberian and Alaskan huskies, and (3) between Indian and African village dogs. In a reticulate maximum-likelihood graph allowing three admixture events, a migration event from the coyote lineage to the PCD/CTVT founder clade is included (Fig. 1h; matrix of residuals in Supplementary informa-

tion, Fig. S21b). Thus, several methods support the presence of gene flow from coyotes to the ancient native dog population represented by the CTVT founder and PCDs. This reticulate graph also demonstrated the concordant result of the *Out of Southern East Asia* hypothesis of living dogs suggested in a previous study<sup>12</sup> (Fig. 1h), which reports that East Asian dogs are the basal clade of all dogs, and two major superclades are found in the dog phylogeny, representing two migration routes to the regions of Far East-America and Indian Peninsula-West Eurasia.<sup>12</sup>

The CTVT founder, inferred from the geographically dispersed CTVT samples, is a useful high-quality proxy for PCDs. The CTVT-private genotype-monomorphic sites will greatly aid cancer evolution studies,<sup>13</sup> and more importantly, the extraction of the CTVT founder genome from genotype-monomorphic sites in CTVT samples is invaluable to canine population studies. Thus, we provide the genotype-monomorphic diploidized sites of the five geographically dispersed CTVTs in the DogDG database of the iDog<sup>14</sup> platform for researchers to conveniently use in future studies.

## ACKNOWLEDGEMENTS

We thank Krishna R. Veeramah and Yong Zhou (BGI, Shenzhen, China) for providing valuable suggestions, Qi-Jun Zhou for collecting CTVT samples and laboratory work, Newton O. Otecko for providing dog samples from Kenya, Ming-Shan Wang for collecting dog samples from Sri Lanka, Shi-Fang Wu for collecting other dog samples, Shao-Jie Zhang for managing re-sequencing data, and the BIG Data Center in Beijing Institute of Genomics, Chinese Academy of Sciences (CAS) for usage of their high-performance computing platform. This work was supported by the Strategic Priority Research Program (B) (XDB13000000) of the CAS, the National Natural Science Foundation of China (31621062, 91731304, 41672021 and 41630102), and the Animal Branch of the Germplasm Bank of Wild Species, CAS (the Large Research Infrastructure Funding). Q.F. is supported by CAS (XDA19050102, QYZDB-SSW-DQC003, and XDPB05), and the Howard Hughes Medical Institute (55008731). G.-D.W. is supported by the Youth Innovation Promotion Association, CAS.

## AUTHOR CONTRIBUTIONS

Y.-P.Z. initiated the project. G.-D.W., Q.F. and Y.-P.Z. managed the project. F.-L.C., S.C. O., A.E., M.M.T., A.D.P. and P.S. collected and provided the samples. T.-T.Y. performed laboratory experiments. T.-T.Y. and X.W. managed the sequencing data. X.W. and B.-W.Z. analyzed and interpreted the data. X.W., M.A.Y., and B.-W.Z. wrote the original manuscript. G.-D.W., Q.F., and Y.-P.Z. reviewed and edited the manuscript.

## ADDITIONAL INFORMATION

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41422-019-0183-2>.

**Competing interests:** The authors declare no competing interests.

Xuan Wang<sup>1</sup>, Bo-Wen Zhou<sup>1,2</sup>, Melinda A. Yang<sup>3</sup>, Ting-Ting Yin<sup>1</sup>, Fang-Liang Chen<sup>4</sup>, Sheila C. Ommeh<sup>5</sup>, Ali Esmailzadeh<sup>6</sup>, Melissa M. Turner<sup>7</sup>, Andrei D. Poyarkov<sup>8</sup>, Peter Savolainen<sup>9</sup>, Guo-Dong Wang<sup>1,2,10</sup>, Qiaomei Fu<sup>3,11</sup> and Ya-Ping Zhang<sup>1,2,10</sup>

<sup>1</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223 Yunnan, China; <sup>2</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650223 Yunnan, China; <sup>3</sup>Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Chinese Academy of Sciences, 100044 Beijing, China; <sup>4</sup>Kunming Police Dog Base of the Ministry of Public Security, Kunming 650204 Yunnan, China; <sup>5</sup>Animal Biotechnology Group, Institute of Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, Nairobi 00200, Kenya; <sup>6</sup>Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman PB 76169-133, Iran; <sup>7</sup>Department of Forestry and Environmental Resources, Fisheries, Wildlife, and Conservation Biology Program, North Carolina State University, Raleigh, NC 27695, USA; <sup>8</sup>Severtsov Institute of Ecology and Evolution, Russian Academy of Science, Leninskiy Prospect, 33, Moscow, Russia 119071; <sup>9</sup>Department of Gene Technology, KTH-Royal Institute of Technology, Science for Life Laboratory, Tomtebodavägen 23A, 17165 Solna, Sweden; <sup>10</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223 Yunnan, China and <sup>11</sup>Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, 100044 Beijing, China

These authors contributed equally: Xuan Wang, Bo-Wen Zhou, Melinda A. Yang.

Correspondence: Guo-Dong Wang (wanggd@mail.kiz.ac.cn) or Qiaomei Fu (fuqiaomei@ivpp.ac.cn) or Ya-Ping Zhang (zhangyp@mail.kiz.ac.cn)

## REFERENCES

1. Murchison, E. P. et al. *Science* **343**, 437–440 (2014).
2. Ostrander, E. A., Davis, B. W. & Ostrander, G. K. *Trends Genet.* **32**, 1–15 (2016).
3. Ni Leathlobhair, M. et al. *Science* **361**, 81–85 (2018).
4. Ujvari, B., Papenfuss, A. T. & Belov, K. *Bioessays* **38**, S14–S23 (2016).
5. Patterson, N. et al. *Genetics* **192**, 1065–1093 (2012).
6. Alexander, D. H., Novembre, J. & Lange, K. *Genome Res.* **19**, 1655–1664 (2009).
7. Monzon, J., Kays, R. & Dykhuizen, D. E. *Mol. Ecol.* **23**, 182–197 (2014).
8. Martin, S. H., Davey, J. W. & Jiggins, C. D. *Mol. Biol. Evol.* **32**, 244–257 (2015).
9. Malinsky, M. et al. *Science* **350**, 1493–1498 (2015).
10. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
11. Pickrell, J. K. & Pritchard, J. K. *PLoS Genet.* **8**, e1002967 (2012).
12. Wang, G.-D. et al. *Cell Res.* **26**, 21–33 (2015).
13. Ostrander, E. A., Dreger, D. L. & Evans, J. M. *Annu. Rev. Anim. Biosci.* **7**, 449–472 (2019).
14. Tang, B. et al. *Nucleic Acids Res.* **47**, D793–D800 (2019).
15. Krzywinski, M. et al. *Genome Res.* **19**, 1639–1645 (2009).