

## RESEARCH ARTICLE SUMMARY

## CANCER EVOLUTION

# Somatic evolution and global expansion of an ancient transmissible cancer lineage

Adrian Baez-Ortega *et al.*

**INTRODUCTION:** The canine transmissible venereal tumor (CTVT) is a sexually transmitted cancer that manifests as genital tumors in dogs. This cancer first arose in an individual “founder dog” several thousand years ago and has since survived by transfer of living cancer cells to new hosts during coitus. Today, CTVT affects dogs around the world and is the oldest and most prolific known cancer lineage. CTVT thus provides an opportunity to explore the evolution of cancer over the long term and to track the unusual biological transition from multicellular organism to obligate conspecific asexual parasite. Furthermore, the CTVT genome, acting as a living biomarker, has recorded the changing mutagenic environments experienced by this cancer throughout millennia and across continents.

**RATIONALE:** To capture the genetic diversity of the CTVT lineage, we analyzed somatic mutations extracted from the protein-coding genomes (exomes) of 546 globally distributed CTVT tumors. We inferred a time-resolved phyloge-

netic tree for the clone and used this to trace the worldwide spread of the disease and to select subsets of mutations acquired at known geographical locations and time periods. Computational methods were applied to extract mutational signatures and to measure their exposures across time and space. In addition, we assessed the activity of selection using ratios of nonsynonymous and synonymous variants.

**RESULTS:** The CTVT phylogeny reveals that the lineage first arose from its founder dog 4000 to 8500 years ago, likely in Asia, with the most recent common ancestor of modern globally distributed tumors occurring ~1900 years ago. CTVT underwent a rapid global expansion within the past 500 years, likely aided by intensification of human maritime travel. We identify a highly specific mutational signature dominated by C>T mutations at GTCCA pentanucleotide contexts, which operated in CTVT up until ~1000 years ago. The number of mutations caused by ultraviolet light exposure is correlated with latitude of tumor collection,

and we identify CTVTs with heritable hyperactivity of an endogenous mutational process. Several “driver” mutation candidates are identified in the basal trunk of the CTVT tree, but there is little evidence for ongoing positive selection. Although negative selection is detectable, its effect is largely confined to genes with known essential functions, thus implying that CTVT predominantly evolves through neutral processes.

**CONCLUSION:** We have traced the evolution of a transmissible cancer over several thousand years, tracking its spread across continents and contrasting the mutational processes and selective forces that molded its genome with those

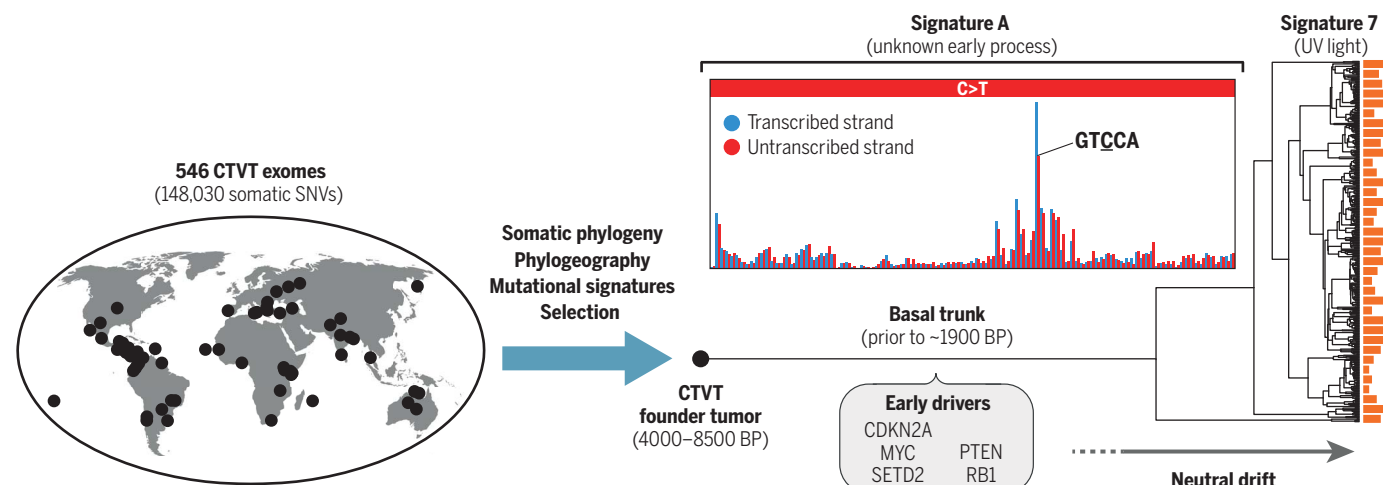
## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aau9923>

described in human cancers. The identification of a highly context-specific mutational process that operated in the past but subsequently vanished, as well as correlation of ultraviolet light-induced DNA damage with latitude, highlight the potential for long-lived, widespread clonal organisms to act as biomarkers for mutagenic exposures. Our results suggest that neutral genetic drift is the dominant evolutionary force operating on cancer over the long term, in contrast to the ongoing positive selection that is often observed in short-lived human cancers. The weakness of negative selection in this asexual lineage may be expected to lead to the progressive accumulation of deleterious mutations, invoking Muller's ratchet and raising the possibility that CTVT may be declining in fitness despite its global success. ■

The list of authors and their affiliations is available in the full article online.

Cite this article as A. Baez-Ortega *et al.*, *Science* 365, eaau9923 (2019). DOI: 10.1126/science.aau9923



**Cancer evolution over thousands of years.** The canine transmissible venereal tumor (CTVT) is an ancient contagious cancer with a global distribution. We sequenced the exomes of 546 CTVT tumors and identified somatic single-nucleotide variants (SNVs). These were used to construct a time-resolved phylogenetic tree, yielding

insights into the cancer's phylogeography, mutational processes, and signatures of selection across thousands of years. Notably, a highly context-specific mutational pattern named signature A was identified, which was active in the past but ceased to operate about 1000 years ago. BP, years before present.

## RESEARCH ARTICLE

## CANCER EVOLUTION

# Somatic evolution and global expansion of an ancient transmissible cancer lineage

Adrian Baez-Ortega<sup>1</sup>, Kevin Gori<sup>1\*</sup>, Andrea Strakova<sup>1\*</sup>, Janice L. Allen<sup>2</sup>, Karen M. Allum<sup>3</sup>, Leontine Bansse-Issa<sup>4</sup>, Thinlay N. Bhutia<sup>5</sup>, Jocelyn L. Bisson<sup>1,6</sup>, Cristóbal Briceño<sup>7</sup>, Artemio Castillo Domracheva<sup>8</sup>, Anne M. Corrigan<sup>9</sup>, Hugh R. Cran<sup>10</sup>, Jane T. Crawford<sup>11</sup>, Eric Davis<sup>12</sup>, Karina F. de Castro<sup>13</sup>, Andriago B. de Nardi<sup>14</sup>, Anna P. de Vos<sup>15</sup>, Laura Delgadillo Keenan<sup>16</sup>, Edward M. Donelan<sup>2</sup>, Adela R. Espinoza Huerta<sup>17</sup>, Ibikunle A. Faramade<sup>18</sup>, Mohammed Fazil<sup>19</sup>, Eleni Fotopoulou<sup>20</sup>, Skye N. Fruean<sup>21</sup>, Fanny Gallardo-Arrieta<sup>22</sup>, Olga Glebova<sup>23</sup>, Pagona G. Gouletson<sup>24</sup>, Rodrigo F. Häfelin Manrique<sup>25</sup>, Joaquim J. G. P. Henriques<sup>26</sup>, Rodrigo S. Horta<sup>27</sup>, Natalia Ignatenko<sup>28</sup>, Yaghouba Kane<sup>29</sup>, Cathy King<sup>3</sup>, Debbie Koenig<sup>3</sup>, Ada Krupa<sup>30</sup>, Steven J. Kruzeniski<sup>17</sup>, Young-Mi Kwon<sup>1</sup>, Marta Lanza-Perea<sup>9</sup>, Mihran Lazyan<sup>31</sup>, Adriana M. Lopez Quintana<sup>32</sup>, Thibault Losfelt<sup>33</sup>, Gabriele Marino<sup>34</sup>, Simón Martínez Castañeda<sup>35</sup>, Mayra F. Martínez-López<sup>36,37</sup>, Michael Meyer<sup>38</sup>, Edward J. Migneco<sup>39</sup>, Berna Nakanwagi<sup>40</sup>, Karter B. Neal<sup>41</sup>, Winifred Neunzig<sup>3</sup>, Máire Ní Leathlobhair<sup>1</sup>, Sally J. Nixon<sup>42</sup>, Antonio Ortega-Pacheco<sup>43</sup>, Francisco Pedraza-Ordoñez<sup>44</sup>, Maria C. Peleteiro<sup>45</sup>, Katherine Polak<sup>46</sup>, Ruth J. Pye<sup>47</sup>, John F. Reece<sup>48</sup>, Jose Rojas Gutierrez<sup>49</sup>, Haleema Sadia<sup>50</sup>, Sheila K. Schmeling<sup>51</sup>, Olga Shamanova<sup>52</sup>, Alan G. Sherlock<sup>47</sup>, Maximilian Stammnitz<sup>1</sup>, Audrey E. Steenland-Smit<sup>4</sup>, Alla Svitch<sup>53</sup>, Lester J. Tapia Martínez<sup>17</sup>, Ismail Thoya Ngoka<sup>54</sup>, Cristian G. Torres<sup>55</sup>, Elizabeth M. Tudor<sup>56</sup>, Mirjam G. van der Wel<sup>57</sup>, Bogdan A. Vițălaru<sup>58</sup>, Sevil A. Vural<sup>59</sup>, Oliver Walkinton<sup>47</sup>, Jinhong Wang<sup>1</sup>, Alvaro S. Wehrle-Martinez<sup>60</sup>, Sophie A. E. Widdowson<sup>61</sup>, Michael R. Stratton<sup>62</sup>, Ludmil B. Alexandrov<sup>63</sup>, Iñigo Martincorena<sup>62</sup>, Elizabeth P. Murchison<sup>1†</sup>

The canine transmissible venereal tumor (CTVT) is a cancer lineage that arose several millennia ago and survives by “metastasizing” between hosts through cell transfer. The somatic mutations in this cancer record its phylogeography and evolutionary history. We constructed a time-resolved phylogeny from 546 CTVT exomes and describe the lineage’s worldwide expansion. Examining variation in mutational exposure, we identify a highly context-specific mutational process that operated early in the cancer’s evolution but subsequently vanished, correlate ultraviolet-light mutagenesis with tumor latitude, and describe tumors with heritable hyperactivity of an endogenous mutational process. CTVT displays little evidence of ongoing positive selection, and negative selection is detectable only in essential genes. We illustrate how long-lived clonal organisms capture changing mutagenic environments, and reveal that neutral genetic drift is the dominant feature of long-term cancer evolution.

**T**ransmissible cancers are malignant somatic cell clones that spread between individuals by direct transfer of living cancer cells. Analogous to the metastasis of cancer to distant tissues within a single body, transmissible cancers “metastasize” as allogeneic grafts between individuals within a population (1). Such clones have been observed only eight times in nature, suggesting that they arise rarely; however, once established, transmissible cancers can spread rapidly and widely and persist through time (1, 2). Such cancers provide an opportunity to explore the evolution of cancer over the long term and to track the unusual biological transition from

multicellular organism to obligate conspecific asexual parasite.

The canine transmissible venereal tumor (CTVT) is the oldest and most prolific known contagious cancer (2, 3). It is a sexually transmitted clone that manifests as genital tumors in dogs. This cancer first arose from the somatic cells of an individual “founder dog” that lived several thousand years ago (2). The cancer survived beyond the death of this original host by transfer of cancer cells to new hosts. Subsequently, this cancer has spread around the world and is a common disease in dog populations globally, although it declined and largely disappeared from many Western

countries during the 20th century owing to the management and removal of free-roaming dogs (4).

Similar to cancers that remain in a single individual, CTVT accumulates somatic mutations. These result from the activities of endogenous and exogenous mutational processes, and genetically imprint a cancer’s history of mutagenic exposures (5). Thus, the CTVT genome can be considered a living biomarker that records the changing mutagenic environments experienced by this cancer throughout millennia and across continents. Although most somatic mutations in cancer have no functional effect and are considered neutral “passenger” mutations, a subset of mutations are positively selected “driver” mutations that confer the proliferation and survival advantages that spur cancer growth (6). Ordinary cancers, which remain in a single host, often acquire additional driver mutations during tumor progression (7); however, it is unknown whether transmissible cancers that survive for hundreds or thousands of years similarly continue to adapt. It seems possible that the evolution of long-lived cancers such as CTVT may instead be dominated by negative selection acting to remove deleterious mutations. Finally, in addition to recording a history of exposures and signatures of selection, somatic mutations provide a tool for tracing CTVT phylogeography, potentially revealing how dogs, together with humans, moved around the world over the past centuries. Here, we use somatic mutations extracted from the protein-coding genomes (exomes) of 546 globally distributed CTVT tumors to trace the history, spread, diversity, mutational exposures, and evolution of the CTVT clone.

## CTVT phylogeny

We sequenced the exomes (43.6 megabases, Mb; mean sequencing depth ~132×) of 546 CTVT tumors collected between 2003 and 2016 from 43 countries across all inhabited continents (datasets S1 and S2). Candidate somatic mutations were defined as single-nucleotide variants (SNVs) or short insertions and deletions (indels) identified in one or more CTVT tumors, but not found in 495 normal dog exomes from the CTVT tumors’ matched hosts. This approach yielded 160,207 variants (148,030 SNVs, 3392 per Mb; 12,177 indels, 279 per Mb; table S1). The features of this set, including its variant allele fraction distribution, phylogenetic structure, comparison with the distribution of private germline variants in the dog population, mutational signature composition, and nonsynonymous-to-synonymous mutation ratio [details in (8)], suggest that it is very highly enriched for somatic mutations. However, some minimal germline variation may remain, possibly including rare germline variants from the founder dog and residual contaminating alleles from matched hosts.

We identified the subset of the candidate somatic mutations belonging to a clocklike mutational process [specifically, cytosine-to-thymine (C>T) substitutions at CpG sites (8, 9)] and used these to construct a time-resolved phylogenetic tree for the CTVT lineage (Fig. 1A). The mutation rate was inferred by applying a Bayesian

Poisson model to previously ascertained empirical observations (10) and was estimated as  $6.87 \times 10^{-7}$  C>T mutations per CpG site per year (8). The topology of the CTVT phylogenetic tree reveals a long basal trunk (Fig. 1A), representing the chain of CTVT transmissions from its origin ~6220 years ago [95% highest posterior density interval (HPDI) 4148 to 8508 years ago] to the earliest detected node ~1938 years ago (95% HPDI 993 to 3055 years ago). This node splits a set of five tumors collected in India from the remaining population (groups labeled 57 and 58; Fig. 1A). The second and third most basal nodes (respectively ~1004 years ago, 95% HPDI 497 to 1570 years ago, and ~829 years ago, 95% HPDI 424 to 1310 years ago) separate 16 tumors from Eastern Europe and the Black Sea region, and three tumors from Northern India, from the remaining set, respectively (groups labeled 54 to 56 and 1; Fig. 1A). Together with evidence that the founder dog shared ancestry with ancient dog remains recovered in northeast Siberia and North America (10), the CTVT phylogeny supports a model whereby CTVT originated ~4000 to 8500 years ago in Central or Northern Asia and remained within the area for the subsequent 2000 to 6000 years. Starting less than ~2000 years ago, CTVT escaped from its founding population, perhaps due to contact between previously isolated dog groups, and spread to several locations in Asia and Europe (Fig. 1B).

The more recent history of CTVT is marked by rapid global expansion (17) (Fig. 1C and fig. S1). CTVT was introduced to the Americas with early colonial contact (~500 years ago, 95% HPDI 284 to 888 years ago), probably initially to Central America, and further into North and South America (red sublineage 1; Fig. 1, A and C). About 300 years ago, this sublineage spread out of the Americas in an almost polytomous global sweep that brought CTVT into Africa at least five times and reintroduced the disease to Europe and Asia (black sublineage 1; Fig. 1, A and C). In parallel, a second

tumor sublineage spread out of Asia or Europe into Australia and the Pacific (sublineage 2; Fig. 1, A and D). This second sublineage is also detected in North America, and its tumors were introduced to Africa on at least two occasions. By ~100 years ago, CTVT was present in dog populations worldwide, establishing local lineages that have since remained largely in situ. The CTVT phylogeny thus suggests that dogs, together with their neoplastic parasites, were extensively transported around the world in the 15th to early 20th centuries, probably by sea travel.

### Mutational processes in CTVT

The CTVT mutational spectrum, a representation of the six substitution types together with their immediate 5' and 3' base contexts, is dominated by C>T mutations, as previously described (12, 13) (Fig. 2A). Applying Markov chain Monte Carlo sampling on a Bayesian model of mutational signatures (8, 14), we extracted signatures of five mutational processes from the CTVT mutation load. These include three signatures that closely resemble COSMIC (15) signatures 1, 5, and 7 (Fig. 2B). These signatures, which have previously been described in CTVT (12), reflect endogenous mutational processes (signatures 1 and 5) and exposure to ultraviolet (UV) light (signature 7) (5). A fourth signature displaying some similarity (cosine similarity 0.81) to COSMIC signature 2, which is associated with activity of APOBEC enzymes (5), was also detected (labeled signature 2\*, Fig. 2B).

The fifth signature extracted from CTVT does not resemble any previously described mutational pattern. This signature, which we designate signature A, is characterized by C>T mutations at NCC contexts (mutated nucleotide underlined) and shows substantial pentanucleotide sequence preference for GTCCA (TGGAC on the complementary strand; Fig. 2, B and C, and fig. S2). This extended sequence preference is markedly more pronounced than previously reported pentanu-

cleotide context biases, such as those associated with UV light or DNA polymerase epsilon deficiency (Fig. 2C) (16–18), and is not explained by the sequence composition of the canine exome (fig. S3). It is possible that signature A's causative mutagen is highly context-specific, or, alternatively, that this signature's associated repair processes are ineffective at certain sequence contexts ("repair shielding") (19). In addition, signature A displays strong transcriptional strand bias, with more mutations of guanine on the untranscribed compared to the transcribed strand of genes, indicating that its causative lesion is likely a guanine adduct subject to transcription-coupled repair (TCR). Notably, the guanine-directed transcriptional strand bias of signature A at TCC contexts counteracts the cytosine-directed transcriptional strand bias of signature 7 at TCC, such that no overall transcriptional strand bias is observed at this context in the CTVT mutational spectrum (Fig. 2A).

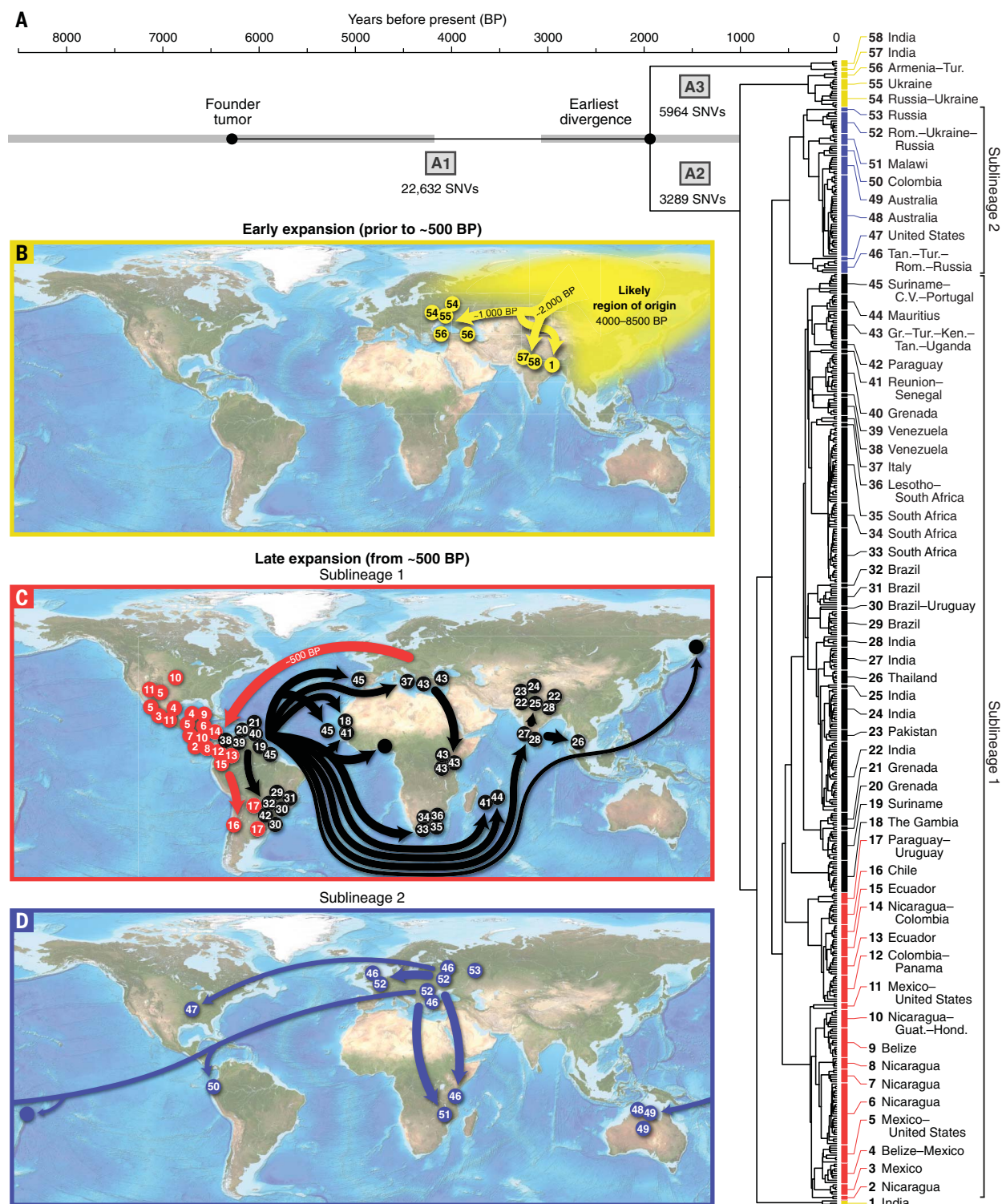
Using the CTVT phylogenetic tree to isolate subsets of mutations, we explored variation in mutational signature exposure across time and space (figs. S4 and S5 and dataset S3). Notably, this revealed that signature A was highly active prior to ~2000 years ago (causing ~35% of mutations in the basal trunk of the tree, branch A1) and persisted in parallel at lower levels in the two basal branches after the first node (~12 and ~9% of mutations in branches A2 and A3, respectively), but then abruptly vanished (Fig. 2C and fig. S5). Moreover, signature A is not detectable within the germ line of a global population of 495 dogs (fig. S6). It is possible that signature A reflects the activity of an exogenous mutagen that was exclusively present in the environment that CTVT inhabited before its escape from its founding population. Alternatively, it is plausible that signature A may result from an endogenous DNA-damaging agent that occurred in CTVT cells early during the lineage's history, but which ceased to accumulate

<sup>1</sup>Transmissible Cancer Group, Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. <sup>2</sup>Animal Management in Rural and Remote Indigenous Communities (AMRRIC), Darwin, Australia. <sup>3</sup>World Vets, Gig Harbor, USA. <sup>4</sup>Animal Shelter, Stichting Dierenbescherming Suriname, Paramaribo, Suriname. <sup>5</sup>Sikkim Anti-Rabies and Animal Health Programme, Department of Animal Husbandry, Livestock, Fisheries and Veterinary Services, Government of Sikkim, India. <sup>6</sup>Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, Roslin EH25 9RG, UK. <sup>7</sup>ConserLab, Animal Preventive Medicine Department, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago, Chile. <sup>8</sup>Corozal Veterinary Hospital, University of Panamá, Panama City, Republic of Panama. <sup>9</sup>St. George's University, True Blue, Grenada. <sup>10</sup>The Nakuru District Veterinary Scheme Ltd, Nakuru, Kenya. <sup>11</sup>Animal Medical Centre, Belize City, Belize. <sup>12</sup>International Animal Welfare Training Institute, UC Davis School of Veterinary Medicine, Davis, CA, USA. <sup>13</sup>Centro Universitário de Rio Preto (UNIRP), São José do Rio Preto, São Paulo, Brazil. <sup>14</sup>Department of Clinical and Veterinary Surgery, São Paulo State University (UNESP), São Paulo, Brazil. <sup>15</sup>Ladybrand Animal Clinic, Ladybrand, South Africa. <sup>16</sup>Veterinary Clinic Sr. Dog's, Guadalajara, Mexico. <sup>17</sup>World Vets Latin America Veterinary Training Center, Granada, Nicaragua. <sup>18</sup>National Veterinary Research Institute, Vom, Nigeria. <sup>19</sup>Animal Clinic, Mombasa, Kenya. <sup>20</sup>Intermunicipal Stray Animals Care Centre (DIKEPAZ), Perama, Greece. <sup>21</sup>Animal Protection Society of Samoa, Apia, Samoa. <sup>22</sup>Faculty of Veterinary Science, University of Zululand, Maracaibo, Venezuela. <sup>23</sup>Veterinary Clinic BIOCONTROL, Moscow, Russia. <sup>24</sup>Faculty of Veterinary Medicine, School of Health Sciences, University of Thessaly, Karditsa, Greece. <sup>25</sup>Veterinary Clinic El Roble, Animal Healthcare Network, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago de Chile, Chile. <sup>26</sup>OnevetGroup, Hospital Veterinário Berna, Lisboa, Portugal. <sup>27</sup>Universidade Vila Velha, Vila Velha, Brazil. <sup>28</sup>Veterinary Clinic Zoovetserwis, Kiev, Ukraine. <sup>29</sup>École Inter-états des Sciences et Médecine Vétérinaires de Dakar, Dakar, Senegal. <sup>30</sup>Department of Small Animal Medicine, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands. <sup>31</sup>Vetexpert Veterinary Group, Yerevan, Armenia. <sup>32</sup>Veterinary Clinic Lopez Quintana, Maldonado, Uruguay. <sup>33</sup>Clinique Veterinaire de Grand Fond, Saint Gilles les Bains, Reunion, France. <sup>34</sup>Department of Veterinary Sciences, University of Messina, Messina, Italy. <sup>35</sup>Facultad de Medicina Veterinaria y Zootecnia, Universidad Autónoma del Estado de México, Toluca, Mexico. <sup>36</sup>School of Veterinary Medicine, Universidad de las Américas, Quito, Ecuador. <sup>37</sup>Cancer Development and Innate Immune Evasion Lab, Champalimaud Center for the Unknown, Lisbon, Portugal. <sup>38</sup>Touray & Meyer Vet Clinic, Serrekunda, The Gambia. <sup>39</sup>Hillside Animal Hospital, St. Louis, MO, USA. <sup>40</sup>The Kampala Veterinary Surgery, Kampala, Uganda. <sup>41</sup>Asavet Veterinary Charities, Tucson, AZ, USA. <sup>42</sup>Vets Beyond Borders, Bylakuppe, India. <sup>43</sup>Faculty of Veterinary Medicine, Autonomous University of Yucatan, Merida, Mexico. <sup>44</sup>Laboratorio de Patología Veterinaria, Universidad de Caldas, Manizales, Colombia. <sup>45</sup>Interdisciplinary Centre of Research in Animal Health (CIISA), Faculty of Veterinary Medicine, University of Lisbon, Lisboa, Portugal. <sup>46</sup>Four Paws International, Vienna, Austria. <sup>47</sup>Vets Beyond Borders, The Rocks, Australia. <sup>48</sup>Help in Suffering, Jaipur, India. <sup>49</sup>Veterinary Clinic Dr José Rojas, Los Andes, Chile. <sup>50</sup>Department of Biotechnology, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan. <sup>51</sup>Corozal Veterinary Clinic, Corozal Town, Belize. <sup>52</sup>Veterinary Clinic Vetmaster, Ramenskoye, Russia. <sup>53</sup>State Hospital of Veterinary Medicine, Dniprodzerzhynsk, Ukraine. <sup>54</sup>Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya. <sup>55</sup>Laboratory of Biomedicine and Regenerative Medicine, Department of Clinical Sciences, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago, Chile. <sup>56</sup>Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Melbourne, Australia. <sup>57</sup>Animal Anti Cruelty League, Port Elizabeth, South Africa. <sup>58</sup>Clinical Sciences Department, Faculty of Veterinary Medicine Bucharest, Bucharest, Romania. <sup>59</sup>Department of Pathology, Faculty of Veterinary Medicine, Ankara University, Ankara, Turkey. <sup>60</sup>Faculty of Veterinary Sciences, National University of Asuncion, San Lorenzo, Paraguay. <sup>61</sup>Lilongwe Society for Protection and Care of Animals (LSPCA), Lilongwe, Malawi. <sup>62</sup>Wellcome Sanger Institute, Hinxton, UK. <sup>63</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.

\*These authors contributed equally to this work.

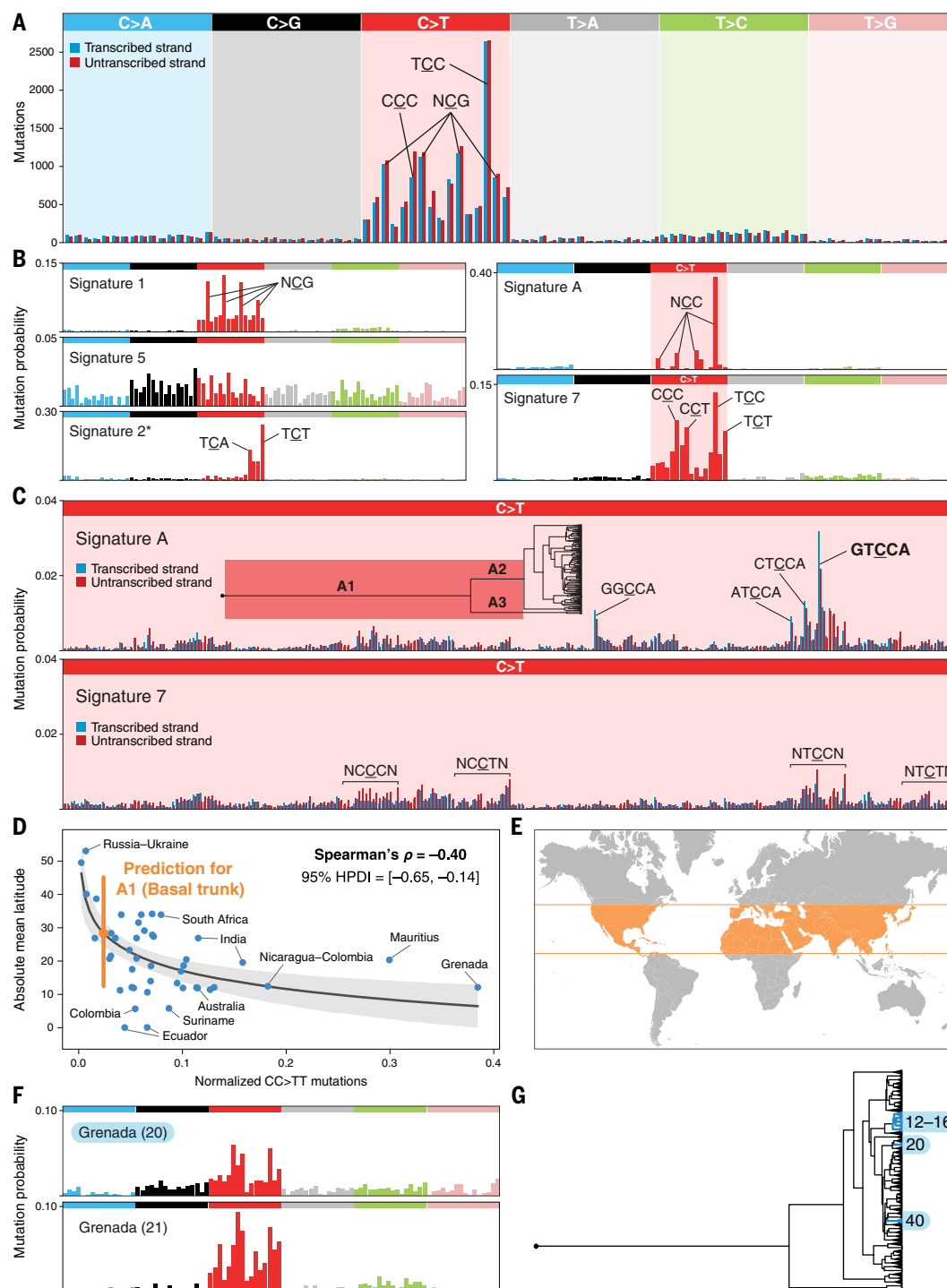
†Corresponding author. Email: epm27@cam.ac.uk





**Fig. 1. Phylogeny and geographical expansion of CTVT.** (A) Time-resolved phylogenetic tree inferred from clocklike exonic somatic variation in CTVT. Each tip indicates a tumor, and sampling locations are labeled. Numbers refer to phylogenetic groups displayed on maps in (B) to (D). Sublineages 1 and 2, referred to in (C) and (D), are marked. Three groups of ancestral somatic variation (A1, A2, A3) and their respective numbers of SNVs are indicated. The age of the CTVT founder tumor and the earliest detected node are indicated in years before present (BP), with gray bars depicting Bayesian 95%

HPDIs. (B to D) Maps presenting likely routes of early and late expansion of CTVT. Numbered circles indicate the locations of phylogenetic groups labeled in (A); arrows represent inferred geographical movements. Circle and arrow colors indicate different sets of movements, as labeled in (A). Thin arrows indicate expansion routes for which phylogenetic evidence is limited; dots without numbers indicate tumors that are not represented in the tree. C.V., Cape Verde; Gr., Greece; Guat., Guatemala; Hond., Honduras; Ken., Kenya; Rom., Romania; Tan., Tanzania; Tur., Turkey.



**Fig. 2. Mutational processes in CTVT.** (A) Trinucleotide-context mutational spectrum of somatic SNVs in a single CTVT tumor. Horizontal axis presents 96 mutation types displayed in pyrimidine context. Relevant mutation contexts are indicated. (B) Mutational spectra of extracted mutational signatures with relevant mutation contexts indicated. (C) Pentanucleotide-context mutational spectra of signature A (top) and signature 7 (bottom). Horizontal axis presents 256 C>T mutation types with relevant mutation contexts indicated. The inset tree shows the phylogenetic branches exposed to signature A. (D) Bayesian logarithmic regression and Spearman's correlation between absolute mean latitude and normalized CC>TT mutations in phylogenetic groups shown in Fig. 1A. Normalized CC>TT mutations

represent the ratio between group-specific CC>TT mutations and group-specific C>T changes at CpG dinucleotides. The black line and shadowed area indicate the regression curve and associated 95% HPDI. The orange dot and bars represent predicted absolute mean latitude and associated 90% prediction interval for the basal trunk ancestral variation (group A1). Posterior median and 95% HPDI of the correlation coefficient are shown. (E) Map showing the latitude range corresponding to the 90% prediction interval for group A1, presented in (D), in the Northern Hemisphere. (F) Mutational spectra of a phylogenetic group showing evidence of signature 5 hyperactivity (top) and a closely related unaffected group (bottom). (G) Diagram indicating the phylogenetic situation of the tumor groups displaying signature 5 hyperactivity.

from ~1000 years ago, perhaps as a result of a cellular metabolic change. Although the nature of such a change is unknown, the replacement of possibly defective mitochondrial DNA by horizontal transfer, which likely occurred in parallel in branches A2 and A3 within the past ~1690 years (11), may have altered the metabolic environment within CTVT cells.

Although CTVT usually occurs within the internal genital tract, it may sometimes protrude from the genital orifice or spread to perineal skin, resulting in sporadic exposure to solar UV radiation (12, 13). The amount of UV radiation reaching Earth, however, varies substantially across global environments (20). We investigated whether latitude influenced the degree of UV exposure in CTVT tumors by estimating the contribution of signature 7 within subsets of mutations acquired at known latitudes. Indeed, qualitative assessment of mutational spectra of location-specific CTVT mutation subsets suggests extensive variation in UV exposure; for example, the mutational spectra of tumors collected in Mauritius show considerably more evidence of signature 7 compared with those of tumors collected in Russia (fig. S4). Using CC>TT dinucleotide mutations (21) as a proxy for signature 7 (fig. S7), we identified a nonlinear association between latitude and UV exposure (Spearman's correlation -0.40, 95% HPDI -0.65 to -0.14; Fig. 2D). By fitting CC>TT mutations observed in the basal trunk of the CTVT tree to this curve, we estimated the latitude of the CTVT founder population (Fig. 2, D and E) (8).

Examining the contribution of signature 5 across the CTVT lineage, we observed three independent phylogenetic groups of tumors that appear to have acquired signature 5 hyperactivity phenotypes (groups labeled 12 to 16, 20, and 40; Fig. 2, F and G, and figs. S4 and S5). In one case, involving tumors collected in several South and Central American countries (groups 12 to 16), the phenotype has been maintained for ~150 years. This phenotype is likely to result from signature 5 and not from the double-strand DNA repair deficiency-mediated COSMIC signature 3, which presents a similar mutational profile (5, 22), as we failed to observe the enrichment for indels that co-occurs with signature 3 (22, 23). It is possible, however, that these tumors were exposed to another, as yet undescribed, mutational process. Signature 5 is widespread in cancer and normal tissues and has unknown etiology, although it may be partly associated with endogenously generated adducts subject to nucleotide excision repair (5, 9, 18). We annotated nonsynonymous mutations occurring in the three groups' respective clonal ancestors, providing a catalog of genes that may play a role in generation or suppression of signature 5 (dataset S4).

### CTVT mutations and gene expression

The prevalence of substitution mutations in CTVT decreases with increasing gene expression, likely reflecting the activity of TCR operating on DNA damage associated with signatures 7 and A, as well as a signature 1 preference for genes with lower expression (16, 24, 25) (fig. S8, A and B). We

observed that exons have a higher substitution prevalence than introns, possibly as a result of sequence context (figs. S8A and S9). The prevalence of indels is positively correlated with increasing gene expression, as has been observed in human cancers, and may reflect transcription-associated damage (26) (fig. S8A).

We assessed the contribution of TCR in two temporally distinct subsets of mutations: those acquired before the earliest detectable node in the phylogenetic tree (~8500 to 2000 years ago; branch A1 in Fig. 1A) and those acquired subsequent to this node (~2000 years ago to the present). Although C>T mutations acquired at TCC contexts in highly expressed genes in branch A1 have little strand bias, likely because of the opposing transcriptional strand preferences of signatures 7 and A at this context, those genes with very low expression predominantly show the transcriptional strand bias associated with signature A (fig. S8C). Assuming that the transcriptional strand bias observed in these low-expressed genes reflects earlier expression and subsequent silencing of genes, this suggests that there may have been an early period in CTVT evolution when the lineage was exposed to signature A more intensely than it was to signature 7. This may reflect variation in the climate or environment to which CTVT was exposed early in its history.

### Selection in CTVT

CTVT has a massive mutation burden, which exceeds that observed in even the most highly mutated human cancer types (Fig. 3A). Each CTVT tumor carries on average 37,800 SNVs across its predominantly diploid (12) exome (~2 million SNVs genome-wide; table S2). Indeed, the tally of somatic mutations that have accumulated in CTVT since it departed its original host is comparable with the number of germline variants that distinguish some pairs of outbred dogs (fig. S10). Within the set of 546 tumors, 14,412 (~73%) protein-coding genes carry at least one nonsynonymous mutation, and 5704 (~29%) have mutations predicted to cause protein truncation (Fig. 3B).

We searched for evidence of positive selection in CTVT. The driver mutations that initially caused CTVT, and promoted its transmissible phenotype, will have occurred in the basal trunk of the CTVT tree. *SETD2*, *CDKN2A*, *MYC* [previously described (12, 27)], *PTEN*, and *RBI*, known cancer genes that frequently harbor driver mutations in human cancers (15), carry biallelic loss-of-function or potential activating mutations in the trunk and may be early drivers of CTVT (Fig. 3C and table S3). To search for late drivers, which may have been acquired in more recent parallel CTVT lineages, we identified independent mutations that occurred repeatedly across the tree, and measured the normalized ratio of nonsynonymous to synonymous mutations (dN/dS) per gene after correcting for mutational biases and context effects (8). This approach only yielded two uncharacterized genes with dN/dS > 1 (*q*-value < 0.05), predicted to encode a neuroligin precursor and a roundabout homolog (dataset S5). The potential for these genes to act as late drivers in CTVT cannot

be assessed, and it is possible that local sequence structures may result in higher-than-expected recurrent mutation rates at these loci (28). Overall, we find little evidence that CTVT is continuing to adapt to its environment.

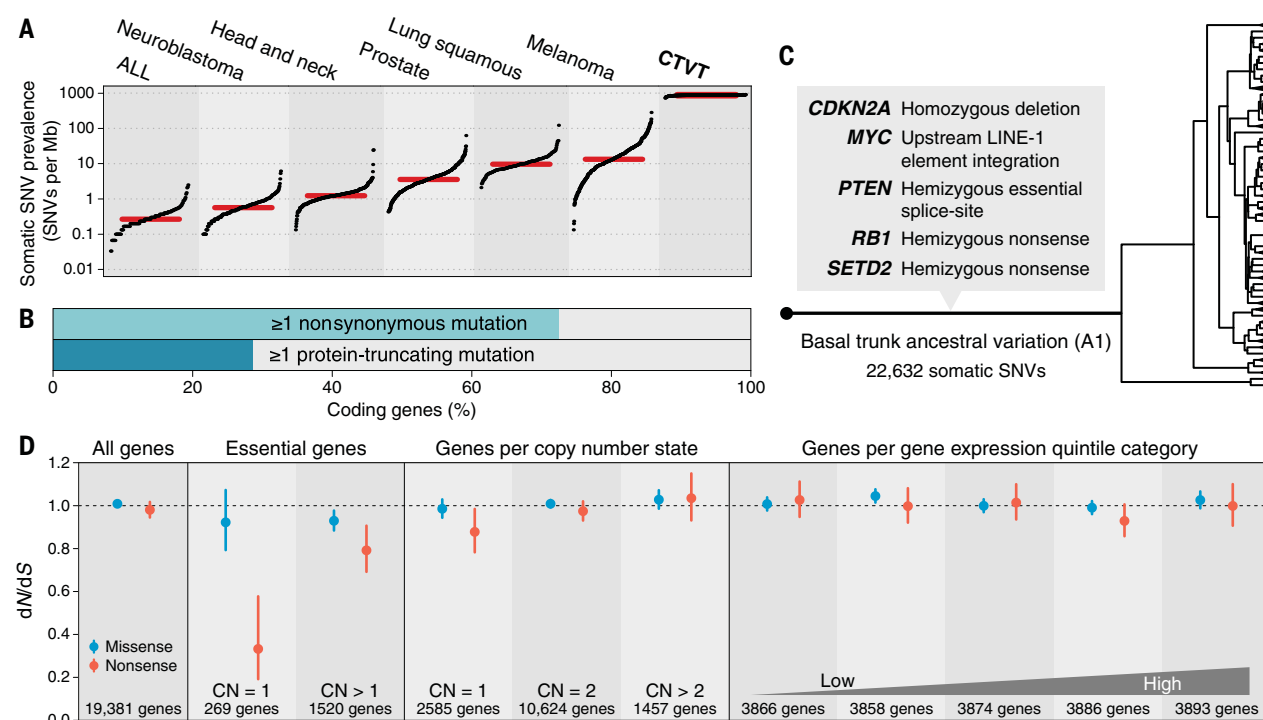
Negative selection, which acts to remove deleterious mutations, is very weak in human cancers (17, 29, 30). Human cancers have short life spans, and their evolution is dominated by sweeps of strong positive selection, thus reducing the potential for negative selection to act (17). Given its long life span, high mutation burden, and lack of ongoing positive selection, it is possible that negative selection may be a more dominant force in CTVT evolution. Further, unlike in ordinary cancers, intertumor competition may offer more opportunities for negative selection to manifest in CTVT, purging lineages less able to infect new hosts and spread through the host population. Indeed, negative selection has been detected operating on CTVT mitochondrial genomes (11). Our analysis of dN/dS in CTVT across all genes, however, yielded dN/dS ≈ 1 for both missense and nonsense mutations, indicating near-neutral evolution (Fig. 3D and dataset S5). Similarly, dN/dS did not differ from neutrality in genes categorized by expression level (Fig. 3D). Negative selection, acting both on missense and nonsense mutations, could be detected, however, in sets of genes with known essential functions (Fig. 3D) and was particularly pronounced for nonsense mutations in essential genes occurring in haploid regions (dN/dS = 0.33, *p* < 10<sup>-4</sup>). A slight signal of negative selection acting on nonsense mutations in haploid regions (dN/dS = 0.88, *p* = 0.027) is explained by 269 essential genes, as negative selection was not detected after removal of these genes (Fig. 3D and dataset S5). These results imply that CTVT largely evolves by neutral genetic drift. This may partly reflect functional obsolescence of many mammalian genes in this relatively simple parasitic cancer, as well as the buffering effect of CTVT's largely diploid genome (12). However, it is also likely that transmission bottlenecks between hosts render weak selection inefficient. This may be expected to lead to the progressive accumulation of deleterious mutations in the population (Muller's ratchet) (31), raising the possibility that CTVT may be declining in fitness despite its global success.

### Discussion

Studies of cancer evolution typically focus on how malignant clones alter during the first years, or perhaps decades, of their existence. We have tracked the evolution of a cancer over several thousand years, and compared the mutational processes and selective forces that molded its genome with those described in short-lived human cancers.

Our results suggest that neutral genetic drift may be the dominant evolutionary force operating on cancer over the long term, in contrast to the ongoing positive selection that is often observed in human cancers (7, 17). Thus, our results suggest that CTVT may have optimized its adaptation to the transmissible cancer niche early in its history. Subsequently acquired advantageous mutations may have offered incremental change





**Fig. 3. Selection in CTVT.** (A) Somatic SNV prevalence across six human cancer types and CTVT. Dots represent individual tumors; red lines indicate median SNV prevalence. ALL, acute lymphoblastic leukemia. (B) Bars showing the percentage of protein-coding genes in the CTVT genome harboring  $\geq 1$  nonsynonymous somatic mutation (SNV or indel; 14,412 genes) and  $\geq 1$  somatic protein-truncating somatic mutation (5704 genes). (C) Diagram presenting the putative driver events found in the set of basal trunk ancestral variants (group A1,

Fig. 1A). A description of each somatic alteration is shown next to the corresponding gene symbol. (D) Exome-wide dN/dS ratios estimated for somatic SNVs in all protein-coding genes (left) and in sets of genes defined according to gene essentiality, copy number state, and expression level. Estimates of dN/dS are presented for missense (blue) and nonsense (orange) mutations in each gene group. The dashed line indicates dN/dS = 1 (neutrality); error bars indicate 95% confidence intervals.

of minimal benefit, such that they were insufficient to overcome the neutral effects of drift. Notably, since the 1980s, CTVT has been routinely treated with vincristine, a cytotoxic microtubule inhibitor (32). Despite the strong selection pressure imposed by vincristine treatment, we find no evidence of convergent evolution of vincristine resistance mechanisms in CTVT at the level of point mutations or indels.

The mechanisms whereby CTVT is tolerated by the host immune system, despite its status as an allogeneic graft, are poorly understood (33, 34). The weakness of negative selection beyond genes essential for cell viability implies that there are negligible selective pressures imposed by immunoeediting of somatic neoepitopes at a genome-wide level. This is perhaps unsurprising, given the massive antigenic burden already presented by allogeneic epitopes. These findings support evidence that CTVT largely circumvents the adaptive immune system, at least during its initial stages of progressive tumor growth, perhaps in part through down-regulation of major histocompatibility complex molecules (13, 34–36).

Our analyses reveal a mutational signature, signature A, that occurred in the past but ceased to be active from about 1000 years ago. A recent study (37) detected evidence for an excess of C>T mutations at TCC contexts, the mutation type most

prevalent in signature A, accumulating in the human germ line between 15,000 and 2000 years ago. If this human mutation pulse is due to signature A, it could indicate a shared environmental exposure that was once widespread but has now disappeared. However, we find no evidence of an excess of C>T mutations at GTCCA pentanucleotides in the dog germ line, suggesting that dogs were not systemically exposed to signature A in their past. Further research will be required to elucidate the biological origin of signature A and the mechanism of its pronounced pentanucleotide sequence bias; however, this study highlights the potential for long-lived, widespread clonal organisms to act as biomarkers for the activity of mutational processes.

Genomic instability and ongoing positive selection are often considered key hallmarks of carcinogenesis (38). CTVT does not have an intrinsically high mutation rate (“genomic instability”), at least at the level of SNVs, and its vast mutation burden simply reflects the lineage’s age. We find no clear evidence for continued positive selection beyond initial truncal events. Thus, CTVT illustrates that, once spawned and sufficiently well-adapted to its niche, neither hallmark is necessary to sustain cancer over the long term.

CTVT is a singular biological entity. It is the oldest, most prolific, and most divergent cancer

lineage known in nature; it has spread throughout the globe and has seeded its tumors in many thousands of dogs. Here, we have traced this cancer’s route through the steppes of Asia and Europe and as an unwelcome stowaway on global voyages. We have observed the patterns in its mutational profiles reflecting the dynamics of its exogenous and endogenous environment. Further, we have shown that CTVT largely evolves by neutral processes, and that the mutations that it continues to acquire may pose a threat, rather than an advantage, to its long-term fitness.

### Materials and methods summary

The protein-coding genomes of 1051 CTVT and matched host samples were captured and sequenced on an Illumina HiSeq2000 instrument. Germline and somatic variants were identified using a bespoke computational pipeline based on Platypus (39) and annotated with the Ensembl Variant Effect Predictor (40). A phylogenetic tree was inferred using RAXML (41), and time of origin was estimated with a Poisson Bayesian model using information about clonal somatic variation from a case of direct CTVT transmission (10). Probabilistic estimates of tree branch lengths and divergence times were obtained using BEAST (42). Mutational signatures and exposures were inferred from subsets of somatic variants using the sigfit R package (14). Selection

was assessed by global and genewise estimates of somatic dN/dS, obtained using the dNdScv R package (17).

## REFERENCES AND NOTES

1. M. J. Metzger, S. P. Goff, A Sixth Modality of Infectious Disease: Contagious Cancer from Devils to Clams and Beyond. *PLoS Pathog.* **12**, (2016). doi: [10.1371/journal.ppat.1005904](https://doi.org/10.1371/journal.ppat.1005904); pmid: [27788268](https://pubmed.ncbi.nlm.nih.gov/27788268/)
2. A. Strakova, E. P. Murchison, The cancer which survived: Insights from the genome of an 11000 year-old cancer. *Curr. Opin. Genet. Dev.* **30**, 49–55 (2015). doi: [10.1016/j.gde.2015.03.005](https://doi.org/10.1016/j.gde.2015.03.005); pmid: [25867244](https://pubmed.ncbi.nlm.nih.gov/25867244/)
3. D. P. Blaine, *A Domestic Treatise on the Diseases of Horses and Dogs* (T. Boosey, London, 1810).
4. A. Strakova, E. P. Murchison, The changing global distribution and prevalence of canine transmissible venereal tumour. *BMC Vet. Res.* **10**, 168 (2014). doi: [10.1186/s12917-014-0168-9](https://doi.org/10.1186/s12917-014-0168-9); pmid: [25186078](https://pubmed.ncbi.nlm.nih.gov/25186078/)
5. L. B. Alexandrov et al., Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013). doi: [10.1038/nature12477](https://doi.org/10.1038/nature12477); pmid: [23945592](https://pubmed.ncbi.nlm.nih.gov/23945592/)
6. M. R. Stratton, P. J. Campbell, P. A. Futreal, The cancer genome. *Nature* **458**, 719–724 (2009). doi: [10.1038/nature07943](https://doi.org/10.1038/nature07943); pmid: [19360079](https://pubmed.ncbi.nlm.nih.gov/19360079/)
7. N. McGranahan, C. Swanton, Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell* **168**, 613–628 (2017). doi: [10.1016/j.cell.2017.01.018](https://doi.org/10.1016/j.cell.2017.01.018); pmid: [28187284](https://pubmed.ncbi.nlm.nih.gov/28187284/)
8. Materials and methods are available as supplementary materials.
9. L. B. Alexandrov et al., Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015). doi: [10.1038/ng.3441](https://doi.org/10.1038/ng.3441); pmid: [26551669](https://pubmed.ncbi.nlm.nih.gov/26551669/)
10. M. Ni Leathlobhair et al., The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018). doi: [10.1126/science.aao4776](https://doi.org/10.1126/science.aao4776); pmid: [29976825](https://pubmed.ncbi.nlm.nih.gov/29976825/)
11. A. Strakova et al., Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. *eLife* **5**, (2016). doi: [10.7554/eLife.14552](https://doi.org/10.7554/eLife.14552); pmid: [27185408](https://pubmed.ncbi.nlm.nih.gov/27185408/)
12. E. P. Murchison et al., Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* **343**, 437–440 (2014). doi: [10.1126/science.1247167](https://doi.org/10.1126/science.1247167); pmid: [24458646](https://pubmed.ncbi.nlm.nih.gov/24458646/)
13. B. Decker et al., Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* **25**, 1646–1655 (2015). doi: [10.1101/gr.190314.115](https://doi.org/10.1101/gr.190314.115); pmid: [26232412](https://pubmed.ncbi.nlm.nih.gov/26232412/)
14. K. C. Gori, A. Baez-Ortega, sigfit: Flexible Bayesian inference of mutational signatures. *bioRxiv* 372896 [Preprint]. 10 July 2018. <https://doi.org/10.1101/372896>
15. S. A. Forbes et al., COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015). doi: [10.1093/nar/gku1075](https://doi.org/10.1093/nar/gku1075); pmid: [25355519](https://pubmed.ncbi.nlm.nih.gov/25355519/)
16. E. D. Pleasance et al., A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010). doi: [10.1038/nature08658](https://doi.org/10.1038/nature08658); pmid: [20016485](https://pubmed.ncbi.nlm.nih.gov/20016485/)
17. I. Martincorena et al., Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–1041.e21 (2017). doi: [10.1016/j.cell.2017.09.042](https://doi.org/10.1016/j.cell.2017.09.042); pmid: [29056346](https://pubmed.ncbi.nlm.nih.gov/29056346/)
18. L. B. Alexandrov et al., The repertoire of mutational signatures in human cancer. *bioRxiv* 322859 [Preprint]. (15 May 2018). <https://doi.org/10.1101/322859>
19. S. S. Lange, K. M. Vasquez, HMGBl: The jack-of-all-trades protein is a master DNA repair mechanic. *Mol. Carcinog.* **48**, 571–580 (2009). doi: [10.1002/mc.20544](https://doi.org/10.1002/mc.20544); pmid: [19360789](https://pubmed.ncbi.nlm.nih.gov/19360789/)
20. R. L. McKenzie, P. J. Aucamp, A. F. Bais, L. O. Björn, M. Ilyas, Changes in biologically-active ultraviolet radiation reaching the Earth's surface. *Photochem. Photobiol. Sci.* **6**, 218–231 (2007). doi: [10.1039/b700017k](https://doi.org/10.1039/b700017k); pmid: [17344959](https://pubmed.ncbi.nlm.nih.gov/17344959/)
21. G. P. Pfeiffer, Y. H. You, A. Besaratina, Mutations induced by ultraviolet light. *Mutat. Res.* **571**, 19–31 (2005). doi: [10.1016/j.mrfmmm.2004.06.057](https://doi.org/10.1016/j.mrfmmm.2004.06.057); pmid: [15748635](https://pubmed.ncbi.nlm.nih.gov/15748635/)
22. S. Nik-Zainal et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016). doi: [10.1038/nature17676](https://doi.org/10.1038/nature17676); pmid: [27135926](https://pubmed.ncbi.nlm.nih.gov/27135926/)
23. S. Morganello et al., The topography of mutational processes in breast cancer genomes. *Nat. Commun.* **7**, 11383 (2016). doi: [10.1038/ncomms11383](https://doi.org/10.1038/ncomms11383); pmid: [27136393](https://pubmed.ncbi.nlm.nih.gov/27136393/)
24. N. J. Haradhvala et al., Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* **164**, 538–549 (2016). doi: [10.1016/j.cell.2015.12.050](https://doi.org/10.1016/j.cell.2015.12.050); pmid: [26806129](https://pubmed.ncbi.nlm.nih.gov/26806129/)
25. E. Letouze et al., Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017). doi: [10.1038/s41467-017-01358-x](https://doi.org/10.1038/s41467-017-01358-x); pmid: [29101368](https://pubmed.ncbi.nlm.nih.gov/29101368/)
26. S. Jinks-Robertson, A. S. Bhagwat, Transcription-associated mutagenesis. *Annu. Rev. Genet.* **48**, 341–359 (2014). doi: [10.1146/annurev-genet-120213-092015](https://doi.org/10.1146/annurev-genet-120213-092015); pmid: [25251854](https://pubmed.ncbi.nlm.nih.gov/25251854/)
27. N. Katzir et al., “Retroposon” insertion into the cellular oncogene c-myc in canine transmissible venereal tumor. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1054–1058 (1985). doi: [10.1073/pnas.82.4.1054](https://doi.org/10.1073/pnas.82.4.1054); pmid: [2983328](https://pubmed.ncbi.nlm.nih.gov/2983328/)
28. X. Zou et al., Short inverted repeats contribute to localized mutability in human somatic cells. *Nucleic Acids Res.* **45**, 11213–11221 (2017). doi: [10.1093/nar/gkx731](https://doi.org/10.1093/nar/gkx731); pmid: [28977645](https://pubmed.ncbi.nlm.nih.gov/28977645/)
29. D. Weghorn, S. Sunyaev, Bayesian inference of negative and positive selection in human cancers. *Nat. Genet.* **49**, 1785–1788 (2017). doi: [10.1038/ng.3987](https://doi.org/10.1038/ng.3987); pmid: [29106416](https://pubmed.ncbi.nlm.nih.gov/29106416/)
30. J. Van den Eynden, S. Basu, E. Larsson, Somatic mutation patterns in hemizygous genomic regions unveil purifying selection during tumor evolution. *PLOS Genet.* **12**, (2016). doi: [10.1371/journal.pgen.1006506](https://doi.org/10.1371/journal.pgen.1006506); pmid: [28027311](https://pubmed.ncbi.nlm.nih.gov/28027311/)
31. H. J. Muller, The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964). doi: [10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8); pmid: [14195748](https://pubmed.ncbi.nlm.nih.gov/14195748/)
32. E. I. Amber, R. A. Henderson, J. B. Adeyanju, E. O. Gyang, Single-drug chemotherapy of canine transmissible venereal tumor with cyclophosphamide, methotrexate, or vincristine. *J. Vet. Intern. Med.* **4**, 144–147 (1990). doi: [10.1111/j.1939-1676.1990.tb00887.x](https://doi.org/10.1111/j.1939-1676.1990.tb00887.x); pmid: [2366223](https://pubmed.ncbi.nlm.nih.gov/2366223/)
33. D. Frampton et al., Molecular Signatures of Regression of the Canine Transmissible Venereal Tumor. *Cancer Cell* **33**, 620–633.e6 (2018). doi: [10.1016/j.ccell.2018.03.003](https://doi.org/10.1016/j.ccell.2018.03.003); pmid: [29634949](https://pubmed.ncbi.nlm.nih.gov/29634949/)
34. H. V. Siddle, J. Kaufman, Immunology of naturally transmissible tumours. *Immunology* **144**, 11–20 (2015). doi: [10.1111/imm.12377](https://doi.org/10.1111/imm.12377); pmid: [25187312](https://pubmed.ncbi.nlm.nih.gov/25187312/)
35. C. Murgia, J. K. Pritchard, S. Y. Kim, A. Fassati, R. A. Weiss, Clonal origin and evolution of a transmissible cancer. *Cell* **126**, 477–487 (2006). doi: [10.1016/j.cell.2006.05.051](https://doi.org/10.1016/j.cell.2006.05.051); pmid: [16901782](https://pubmed.ncbi.nlm.nih.gov/16901782/)
36. A. Fassati, N. A. Mitchison, Testing the theory of immune selection in cancers that break the rules of transplantation. *Cancer Immunol. Immunother.* **59**, 643–651 (2010). doi: [10.1007/s00262-009-0809-1](https://doi.org/10.1007/s00262-009-0809-1); pmid: [20033157](https://pubmed.ncbi.nlm.nih.gov/20033157/)
37. K. Harris, J. K. Pritchard, Rapid evolution of the human mutation spectrum. *eLife* **6**, (2017). doi: [10.7554/eLife.24284](https://doi.org/10.7554/eLife.24284); pmid: [28402220](https://pubmed.ncbi.nlm.nih.gov/28402220/)
38. P. C. Nowell, The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976). doi: [10.1126/science.959840](https://doi.org/10.1126/science.959840); pmid: [959840](https://pubmed.ncbi.nlm.nih.gov/959840/)
39. A. Rimmer et al., Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014). doi: [10.1038/ng.3036](https://doi.org/10.1038/ng.3036); pmid: [25017105](https://pubmed.ncbi.nlm.nih.gov/25017105/)
40. W. McLaren et al., The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016). doi: [10.1186/s13059-016-0974-4](https://doi.org/10.1186/s13059-016-0974-4); pmid: [27268795](https://pubmed.ncbi.nlm.nih.gov/27268795/)
41. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033); pmid: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/)
42. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012). doi: [10.1093/molbev/mss075](https://doi.org/10.1093/molbev/mss075); pmid: [22367748](https://pubmed.ncbi.nlm.nih.gov/22367748/)
43. A. Baez-Ortega, Somatic evolution and global expansion of an ancient transmissible cancer lineage, University of Cambridge Repository (2019); doi: [10.17863/CAM.24962](https://doi.org/10.17863/CAM.24962)

## ACKNOWLEDGMENTS

We acknowledge the Core Sequencing Facility, IT groups, and members of the Cancer Genome Project at the Wellcome

Sanger Institute. We thank the following individuals for useful information and for their help obtaining samples for this project: I. Airikkala-Otter, J. Alzate-Ocampo, D. Argüello, J. I. Arias, C. L. Arnold, S. Barrass, E. Batrakova, R. Bortolotti Viéra, N. Brown, F. C. Casas, J. Cooper, A. C. Cotacachi, S. M. Cutter, J. de Vos, L. Dmytro, P. Farnham, A. Fassati, A. Fernandez-Riomalo, R. Gaitan, D. Hanzlíček, R. R. Hupples, J. M. Igundu, M. Jimenez-Coello, D. Kamstock, P. Kelly, T. Korytina, A. Kuznetsova, G. E. Lavalle, O. A. Lawal, T. Lerotholi, M. Lima-Maigua, J. Loayza-Feijoo, M. López-Bucheli, M. Maina, M. Mancero-Albuja, C. Marchiori Bueno, L. Martínez-López, A. Martínez-Meza, B. M. Masuruli, T. M. Morata Raposo, J. Mulholland, C. Murgia, A. Murison Swartz, F. Nargi, M. M. Onsare, E. Ortiz-Rodríguez, E. Peach, L. Pellegrini, G. Polton, F. Proaño-Pérez, J. C. Ramirez-Ante, C. Raw, C. Smedo, I. Stoikovic, I. Swarisch, M. Tinucci Costa, E. Turitto, M. R. Vural, D. Walker, R. Weiss, K. Xie, M. Zandvliet, staff at Animal Medical Centre Belize City (Belize), veterinary surgeons and staff at Help in Suffering (Jaipur, India), staff at Hopkins Belize Humane Society (Belize), veterinary workers at Pet Centre (UVAS, Lahore, Pakistan), students from St. George's University (True Blue, Grenada, West Indies) who assisted with sample collection, staff at Veterinary Clinic “El Roble” (Chile), staff and volunteers at World Vets (Gig Harbor, USA), and staff at the WVS International Training Centre in Ooty (India). We are grateful to the following organizations for helpful information: American College of Veterinary Internal Medicine (ACVIM), Animal Balance, Animal Care Association (The Gambia), Animal Management in Rural and Remote Indigenous Communities (AMRRIC), Associação Bons Amigos de Cabo Verde, Humane Society of Cozumel, Humane Society Veterinary Medical Association—Rural Area Veterinary Services (HSVMA—RAVS), Israel Veterinary Medical Association, Italian Veterinary Oncology Society, Rural Vets South Africa, Veterinary Cancer Society, Veterinary Society of Surgical Oncology (VSSO), VetPharma, Vets Beyond Borders, VIDAS and Coco's Animal Welfare, The Spanky Project, VWB/VSF Canada, West Arnhem Land Dog Health Program (WALdHeP), World Small Animal Veterinary Association (WSAVA), МИР ВЕТЕРИНАРИИ (World Veterinary Medicine). **Funding:** This work was supported by Wellcome (102942/Z/13/A) and by a Philip Leverhulme Prize from the Leverhulme Trust. A.St. was supported by a Postgraduate Student Award from the Kennel Club Charitable Trust. **Author contributions:** E.P.M. designed and directed the project. A.B.-O. developed methods and led computational data analysis. K.G. developed methods and assisted with computational analysis. A.St. collected samples, performed laboratory work, designed exome probes, oversaw sequencing and provided conceptual advice. J.L.A., K.M.A., L.B.-I., T.N.B., J.L.B., C.B., A.C.D., A.M.C., H.R.C., J.T.C., E.D., K.F.d.C., A.B.d.N., A.P.d.V., L.D.K., E.M.D., A.R.E.H., I.A.F., M.F., E.F., S.N.F., F.G.-A., O.G., P.G.V., R.F.H.M., J.J.G.P.H., R.S.H., N.I., Y.K., C.K., D.K., A.K., S.J.K., M.L.-P., M.L., A.M.L.Q., T.L., G.M., S.M.C., M.F.M.-L., M.M., E.J.M., B.N., K.B.N., W.N., S.J.N., A.O.-P., F.P.-O., M.C.P., K.P., R.J.P., J.F.R., J.R.G., H.S., S.K.S., O.S., A.G.S., A.E.S.-S., A.S.v., L.J.T.M., I.T.N., C.G.T., E.M.T., M.G.v.d.W., B.A.V., S.A.V., O.W., A.S.W.-M. and S.A.E.W. provided clinical samples. Y.-M.K., M.N.L., and M.S. assisted with analysis and contributed to interpretation of results. J.W. contributed to sample management and curation. M.R.S., L.B.A., and I.M. provided technical advice and assisted with interpretation of results. A.B.-O. and E.P.M. wrote the manuscript and designed the figures. All authors commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Whole-exome sequence data have been deposited in the European Nucleotide Archive (ENA; [www.ebi.ac.uk/ena](https://www.ebi.ac.uk/ena)) under overarching accession number ERP109580. Variant calling data and other data supporting analyses have been deposited in the University of Cambridge Repository ([www.repository.cam.ac.uk](https://www.repository.cam.ac.uk)) (43). Custom algorithms employed for data processing and analysis are available in GitHub (<https://github.com/baezortega/TCG2019>).

## SUPPLEMENTARY MATERIALS

[science.sciencemag.org/content/365/6452/eaau9923/suppl/DC1](https://science.sciencemag.org/content/365/6452/eaau9923/suppl/DC1)  
Materials and Methods  
Figs. S1 to S16  
Tables S1 to S4  
Datasets S1 to S6  
References (44–70)

22 August 2018; accepted 20 June 2019  
[10.1126/science.aau9923](https://doi.org/10.1126/science.aau9923)



## Somatic evolution and global expansion of an ancient transmissible cancer lineage

Adrian Baez-Ortega, Kevin Gori, Andrea Strakova, Janice L. Allen, Karen M. Allum, Leontine Bansse-Issa, Thinlay N. Bhutia, Jocelyn L. Bisson, Cristóbal Briceño, Artemio Castillo Domracheva, Anne M. Corrigan, Hugh R. Cran, Jane T. Crawford, Eric Davis, Karina F. de Castro, Andriago B. de Nardi, Anna P. de Vos, Laura Delgadillo Keenan, Edward M. Donelan, Adela R. Espinoza Huerta, Ibikunle A. Faramade, Mohammed Fazil, Eleni Fotopoulou, Skye N. Fruean, Fanny Gallardo-Arrieta, Olga Glebova, Pagona G. Gouletsou, Rodrigo F. Häfelin Manrique, Joaquim J. G. P. Henriques, Rodrigo S. Horta, Natalia Ignatenko, Yaghoub Kane, Cathy King, Debbie Koenig, Ada Krupa, Steven J. Kruzeniski, Young-Mi Kwon, Marta Lanza-Perea, Mihran Lazyan, Adriana M. Lopez Quintana, Thibault Losfelt, Gabriele Marino, Simón Martínez Castañeda, Mayra F. Martínez-López, Michael Meyer, Edward J. Migneco, Berna Nakanwagi, Karter B. Neal, Winifred Neunzig, Máire Ní Leathlobhair, Sally J. Nixon, Antonio Ortega-Pacheco, Francisco Pedraza-Ordoñez, Maria C. Peleteiro, Katherine Polak, Ruth J. Pye, John F. Reece, Jose Rojas Gutierrez, Haleema Sadia, Sheila K. Schmeling, Olga Shamanova, Alan G. Sherlock, Maximilian Stammnitz, Audrey E. Steenland-Smit, Alla Svitich, Lester J. Tapia Martínez, Ismail Thoya Ngoka, Cristian G. Torres, Elizabeth M. Tudor, Mirjam G. van der Wel, Bogdan A. Vitalaru, Sevil A. Vural, Oliver Walkinton, Jinhong Wang, Alvaro S. Wehrle-Martinez, Sophie A. E. Widdowson, Michael R. Stratton, Ludmil B. Alexandrov, Iñigo Martincorena and Elizabeth P. Murchison

*Science* **365** (6452), eaau9923.  
DOI: 10.1126/science.aau9923

### It's a dog's life

Canine transmissible venereal tumor is one of the few cancer lineages that is transferred among individuals through contact. It arose millennia ago and has been evolving independently from its hosts ever since. Baez-Ortega *et al.* looked at the phylogenetic history of the cancer and describe several distinctive mutational patterns (see the Perspective by Maley and Shibata). Most notably, both positive and negative selection show only weak or distant signals. This suggests that the main driver of the lineage's evolution is neutral genetic drift. Understanding the influence of drift may reshape how we think about long-term cancer evolution.

*Science*, this issue p. eaau9923; see also p. 440

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/365/6452/eaau9923>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/07/31/365.6452.eaau9923.DC1>

#### RELATED CONTENT

<http://science.sciencemag.org/content/sci/365/6452/440.full>

#### REFERENCES

This article cites 69 articles, 11 of which you can access for free  
<http://science.sciencemag.org/content/365/6452/eaau9923#BIBL>

Use of this article is subject to the [Terms of Service](#)

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.