NEWS & VIEWS

C. COLLINS/CORBIS

GENOMICS

The dog has its day

Hans Ellegren

Domestication and selective breeding have transformed wolves into the diversity of dogs we see today. The sequence of the genome of one breed adds to our understanding of mammalian biology and genome evolution.

Dogs have a special place in our society. Man's best friend is not just a valuable hunting partner, guard and herd manager — most of the world's estimated 400 million dogs[1] are pets. Dogs were the first animals to be domesticated (at least 15,000 years ago)[2–4]. They all originate from a single and relatively homogeneous species — the wolf — but modern breeds display an extraordinary diversity of traits (or phenotypes). The hundreds of years of careful inbreeding to produce the many kinds of dog have delivered a geneticist's dream model of human genetic disease (Box 1, overleaf). But to unlock the full potential of this model, we need to understand the genetic basis for the unprecedented diversity and how it has evolved[5]. The high-quality draft sequence of the dog genome described on page 803 of this issue[6] is a good starting-point for that research.

Lindblad-Toh and colleagues[6] invited breed clubs and veterinary schools to suggest an individual dog suitable for genome sequencing. The idea was to identify a highly inbred dog; this was based on the thinking that the animal's genetic homogeneity would simplify the gigantic jigsaw puzzle of assembling millions of sequence reads into a genome sequence. After testing certain genetic markers in a host of dogs, the sequencers settled on a female boxer called Tasha (so there is no Y chromosome in the current sequence).

The assembled sequence from Tasha's DNA spans 2.4×10^9 base pairs (Gbp), which corresponds to an estimated 99% coverage of the canine genome (excluding highly repetitive regions). So, although dogs have 39 pairs of chromosomes (compared with 23 pairs in humans), their genome contains almost 0.5 Gbp less DNA than ours. The difference can be explained mainly by the existence of fewer repetitive elements in the dog lineage, and to some extent by deletion of sequences that were present in an early common mammalian ancestor. Dogs seem to have fewer genes than humans, but the actual numbers might be a bit out for both genomes because identifying genes across whole genomes continues to be a difficult task[7].

The current work is not the first canine genome project. Sequencing of a male poodle (at a lower sequence coverage) recently characterized about 75% of its genome, although with much of the assembled sequence interleaved with gaps of undefined length[8]. However, by comparing it with the boxer genome, the poodle sequence is a useful tool for identifying genetic variants — single nucleotide polymorphisms (SNPs) — in dog populations. Augmented with SNPs identified in the boxer and by limited sequencing of many other dog breeds, 2.5 million variable sites have now been discovered[6]. Comparisons of the different breeds show that there is an average of around 1 SNP per 1,000 base pairs — a similar value to that in human populations.

The SNP data give several evolutionary insights. For instance, analysis of DNA from mitochondria (cellular organelles that have their own genome) has suggested that domestication is associated with a narrow genetic bottleneck where only a few wild ancestors contributed to the domestic gene pool[9]. However, the large genetic diversity seen among dogs is at odds with this hypothesis, and work on other domestic animals shows that they, too, have high levels of variability in their nuclear genes. This implies that, in many cases, back-crosses with wild relatives introduced additional genetic diversity into domesticated animals well after domestication began[10]. The genetic traces of such interbreeding may not be picked up by studies of mitochondrial DNA if the back-crossing occurred mainly between wild males and domestic females, because mitochondrial DNA is inherited only from mothers[11].

The physical positions of the genetic variations within and among breeds create patterns in the genome that give a more detailed perspective on domestication and breed formation. Within breeds, most chromosomes are mosaics of alternating regions of homogeneous sequences — reflecting the recent

Box 1 | From wolf to dog to disease model

The wild ancestor of dogs, the grey wolf, belongs to a large group of mammals called the Carnivora, which includes cats, bears and seals. Roughly 40 million years ago, a family of dog-like carnivores (Canidae) evolved, and about 15 million years ago they diverged into foxes, wolves, jackals and others. A phylogenetic analysis[6] shows that the coyote is the closest living relative of the grey wolf (the two species had a common ancestor one million to two million years ago), followed by, in order of genetic distance, the golden jackal, Ethiopian wolf, dhole and African wild dog. Ancient dog remains from Alaska and Latin America indicate that native American dogs originated from dogs domesticated in the Old World[4]. These dogs must have accompanied late Pleistocene humans across the Bering Straits, which means they were domesticated at least 15,000 years ago, probably in southeast Asia[3].

Modern dog breeds have subsequently been generated by selecting for existing traits among the wild ancestors — a prime example of evolution by selection. The extraordinary variation in shape, size, behaviour and physiology of the breeds makes the dog a unique genetic model; each pure breed is an inbred, isolated genetic population, with simplified genetic structures that can be linked to their physical traits.

Several hundred genetic disorders shared between dogs and humans have been reported, many of which are found in just one or a few breeds. For instance, narcolepsy is seen largely in doberman pinschers, and a hereditary kidney cancer occurs only in German shepherd dogs; the genes underlying both diseases have been identified in dogs. Examples of genetic diseases common to several breeds include blindness, allergy and epilepsy. Using dogs as a model for human genetic disease can not only identify causative genes and aid the development of treatments, but can also provide information on the character of disease-causing mutations. For example, the expansion or insertion of repetitive elements in genes has recently been shown to cause disease in both humans and dogs[16,17]. **H.E.**

common ancestry shared by individual dogs of the same breed — and heterogeneous sequences[6,12,13]. Mathematical simulations can be used to model the way in which population history might be expected to affect genetic diversity and its structural patterns. The model that best fits the observed pattern of SNPs is one that assumes an ancient bottleneck some 9,000 generations ago (domestication), followed by breed-specific bottlenecks 30–90 generations ago (breed formation). However, if repeated back-crossing has occurred, this model would have to be revised.

The dog adds to a growing list of vertebrate species that have had their genome sequenced[14]. A comparative analysis of the human, mouse and dog by Lindblad-Toh *et al.*[6] showed that about 5% of the human genome is being maintained by natural selection — suggesting that it has some essential function. Almost all of this sequence is also present in the dog genome. Only 1–2% of the genomes encodes proteins, so there would seem to be an additional common set (about 3%) of functional elements in mammalian non-coding DNA. These common sequences may constitute, for example, regulatory elements, structural elements or RNA genes. Notably, such regions are found mostly within the 0.8 Gbp of ancestral sequence common to human, mouse and dog.

With the dog genome sequence available, it will be exciting to follow the forthcoming search for associations between certain phenotypes in different breeds and the genes responsible for them (Box 1). It will now be possible, using various genomic approaches, to map breed-specific traits related to morphology, physiology and behaviour[15], which will provide insight into key features of mammalian biology and disease. ■

Hans Ellegren is in the Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.
e-mail: Hans.Ellegren@ebc.uu.se

1. Coppinger, R. & Coppinger, L. *Dogs: A Startling New Understanding of Canine Origin, Behaviour & Evolution* (Scribner, New York, 2001).
2. Vilà, C. *et al. Science* **276,** 1687–1689 (1997).
3. Savolainen, P. *et al. Science* **298,** 1610–1613 (2002).
4. Leonard, J. A. *et al. Science* **298,** 1613–1616 (2002).
5. Sutter, N. B. & Ostrander, E. *Nature Rev. Genet.* **5,** 900–910 (2004).
6. Lindblad-Toh, K. *et al. Nature* **438,** 803–819 (2005).
7. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
8. Kirkness, E. F. *et al. Science* **301,** 1898–1903 (2003).
9. Bruford, M. *et al. Nature Rev. Genet.* **4,** 900–910 (2003).
10. Vilà, C. *et al. Trends Genet.* **21,** 214–218 (2005).
11. Götherström, A. *et al. Proc. R. Soc. Lond. B* **272,** 2337–2344 (2005).
12. Sutter, N. B. *et al. Genome Res.* **14,** 2388–2396 (2004).
13. Parker, H. G. *et al. Science* **304,** 1160–1164 (2004).
14. www.ncbi.nih.gov/Genomes
15. Pollinger, J. P. *et al. Genome Res.* (in the press).
16. Lin, L. *et al. Cell* **98,** 365–376 (1999).
17. Lohi, H. *et al. Science* **307,** 81 (2005).

WATER

Ins and outs of ice nucleation

Srikanth Sastry

Laboratory experiments point to a mechanism by which ice forms from supercooled water with surprising alacrity. Such a mechanism may help to explain ice formation in the atmosphere under certain conditions.

In Kurt Vonnegut's novel *Cat's Cradle*, Earth's waters freeze over on contact with ice-9, a fictional form of ice that is more stable than water. Vonnegut depicts an imaginary and extreme scenario of how the thermodynamics of water might dictate the fate of the planet. But the transformation of water to ice is no less fascinating in reality. Water can remain a liquid even under conditions in which a more stable phase exists. This occurs at temperatures below 0 °C when ice-I (the normal variety) is the stable phase. In these circumstances, water is termed metastable or supercooled. It can be prompted to turn into ice by seeding it with a speck of the stable ice (as in the story), or with a small particle of, for instance, dust or ash.

At Earth's surface, water exists in solid, liquid and gaseous phases at or close to ambient conditions. Transformations between these phases can therefore occur readily (through small changes in temperature, for example), and can have a great influence on the dynamics of Earth's atmosphere. This is the context in which two new publications by Raymond A. Shaw and colleagues[1,2] are set. The authors describe lab studies of a particular pathway, called contact nucleation, by which

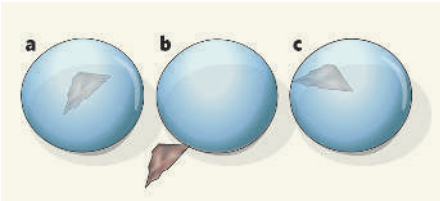

Figure 1 | Three ways in which an ice nucleus may cause crystallization of a water drop. **a**, A nucleus immersed in the bulk drop. **b**, Contact from a nucleus outside the drop. **c**, Contact from within the drop ('contact nucleation inside-out'). Crystallization occurs at higher temperatures in the two surface-contact situations[1,2].

supercooled water can be transformed into ice. They also discuss how their results might help to understand aspects of ice formation in Earth's atmosphere — which in turn affects patterns of rainfall and snowfall, and the influence of clouds on the amount of solar radiation reaching Earth's surface.

When cooled below 0 °C at ambient pressure, water will eventually become ice, a solid with a regular molecular structure and strong attractive interactions between molecules that