Genes of domestic mammals augmented by backcrossing with wild ancestors

Carles Vilà, Jennifer Seddon* and Hans Ellegren

Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

Both archaeological data and the presence of few mitochondrial DNA lineages suggest that most wide-spread domestic mammals (cattle, sheep, goats, pigs and dogs) derive from only a handful of domestication events. However, each of these species shows a high level of diversity at the nuclear genes of the major histocompatibility complex (MHC). Through simulations incorporating various degrees of population subdivision, growth rate and selection, we demonstrate that the numerous MHC *DRB* alleles that are present in modern domestic mammals implies that substantial backcrossing with wild ancestors, either accidental or intentional, has been important in shaping the genetic diversity of our domesticates. These results support the view that, contrary to common assumption, domestic and wild lineages might not have been clearly separated throughout their history.

Introduction

The domestication of plants and animals dramatically changed the way humans live. It enabled a greater degree of independence from environmental fluctuations, made it possible to concentrate food resources, transformed global demography and was a prerequisite to the rise of civilization [1]. Domestication is commonly seen as a rare event that has occurred only in a limited number of regions [1–3]. Genetic analyses of domestic mammals have provided concordant evidence because mitochondrial DNA (mtDNA) haplotypes typically cluster into a few lineages. This pattern is consistent with a few independent domestication events [1–3]. However, because mtDNA only reveals information about the female contribution to domestication and because it is highly sensitive to random genetic drift, mtDNA might be a poor estimator of the actual number of individuals involved in domestication.

The major histocompatibility complex (MHC) encodes proteins that are essential to the normal functioning of the immune system. In contrast to neutral markers such as mtDNA, genes of the MHC typically show a high level of genetic diversity that is maintained by balancing selection

[4]. One consequence of this form of selection is that MHC alleles are maintained for longer periods of time than expected under neutrality [5]. Hence, these alleles can provide evolutionary information over a long time frame and have been used to provide estimates of effective population sizes in humans [6,7] and to calculate founding population sizes for other species [8]. Similarly, the long persistence times of MHC alleles can be exploited to investigate the number of animals involved in the origins of domestic mammals and to gain insight into the husbandry practices at the time of the domestication.

Genetic diversity at MHC loci

Several MHC DR *DRB* alleles have been collated in public databases for cattle, sheep, goats, pigs and dogs. We extracted exon 2 MHC *DRB* alleles in sheep, goats and pigs from GenBank, exon 2 *DRB1* sequences for dogs [9] and *DRB3* alleles for cattle (http://www.projects.roslin.ac.uk/bola/bolahome.html). Between 39 and 66 alleles were identified per species (Table 1). The majority of these alleles differ at one or more non-synonymous sites (37–63 alleles). New alleles in domestic mammals can only arise by mutation or gene flow from the ancestor, with selection and demography altering their frequencies. Considering the estimated substitution rate for MHC class II alleles [10] and the mean number of non-synonymous sites, the average time expected for a single non-synonymous substitution to occur was calculated to be > 800 000 years for each species (Table 1). This value greatly exceeds the time in which the domestication process began for most of these species, which is estimated to be 8000–14 000 years ago [2]. Furthermore, the probability of non-synonymous substitutions occurring within a sequence in a given time period can be estimated by the expression $1-(1-\mu)^{tN}$, or approximated as μtN, where μ is the non-synonymous mutation rate, N is the number of non-synonymous sites and t is time. For the *DRB* sequences in Table 1, the probability of a non-synonymous substitution occurring in the past 20 000 years is < 0.025 for all species. Even considering an effective population size in the order of 10^6, most of the alleles would be expected to have originated before domestication. Hence, all or almost all of the 37–63 *DRB* alleles that differ by at least one non-synonymous substitution are likely to have been present in the wild progenitors and inherited by the derived domestic species.

Corresponding author: Vilà, C. (carles.vila@ebc.uu.se).
* Current address: School of Veterinary Science, University of Queensland, St Lucia QLD 4072, Australia.

Table 1. MHC *DRB* allelic diversity in different species[a,b]

Species	Number of alleles	Mean number of non-synonymous sites $\pm$ SE[c]	Time per non-synonymous substitution	Number of alleles (Nd > 0)
Dog	42	200.39 $\pm$ 3.36	846 000	41
Pig	50	171.94 $\pm$ 7.51	986 000	49
Goat	39	197.15 $\pm$ 4.93	860 000	37
Sheep	41	187.59 $\pm$ 6.24	894 000	39
Cattle	66	189.53 $\pm$ 5.90	894 000	63

[a]Abbreciation: *Nd*, number of non-synonymous substitutions.
[b]The total number of different MHC *DRB* alleles in public databases, mean number of non-synonymous sites per allele, estimated time required per non-synonymous substitution and number of alleles that are differentiated by one or more ($Nd > 0$) non-synonymous substitutions in each species.
[c]Standard error (SE) was calculated using 500 bootstrap iterations.

Estimation of the number of founders

For the 37–63 persistent *DRB* alleles to pass into the domestic mammals, the absolute minimum number of individuals involved in the domestication process is one-half of these values (i.e. 19–32 individuals). However, this makes the unlikely assumption that all founding individuals are heterozygous for different alleles and, importantly, it does not take into account the effects of genetic drift, which will lead to a loss of MHC alleles while the population size is small [11].

To assess the relative influence of growth rate, selection coefficient and the number of alleles present in the founding population, we simulated the effects of genetic drift for a population founded by a fixed number of adults (Box 1). A comparison of the number of MHC alleles retained showed that the number of alleles in the ancestral population and the strength of the balancing selection had minor effects compared with the rate of population growth (Figure 1). Balancing selection is expected to maintain increased numbers of alleles and it has been suggested that alterations to living conditions, and exposure to new environments and their pathogens, could lead to increased selective pressures and the maintenance of more MHC alleles in domestic mammals [12,13]. However, even if the change in living conditions associated with domestication implied an extreme increase of the selective coefficient from 0.01 (the average value for variable MHC class II genes [14]) to 0.50 (implying that half of the homozygote individuals do not

breed; as observed in some extreme cases [15]), this would produce only a maximal twofold increase in the number of alleles remaining (with a population growth of 1% per generation). By contrast, the more important effect of rapid population expansion is shown by the three- to sixfold increase in the number of alleles remaining when the rate of population growth is increased from 1% to 10%. Star-like mtDNA phylogenies suggest that a rapid population growth – on an evolutionary time scale – occurred in many domestic animals [16–18].

A second set of simulations (Figure 2a) estimated that if the domesticated species derived from a single domestication event (from a single population) and the population grew at ~5% per generation, a founding population of >200 equally reproducing individuals was required to retain as much diversity as we currently observe. With a much faster population growth of 10% per generation, ~150 individuals would suffice. However, this rate might represent an unrealistic scenario because it is difficult to imagine that human societies that existed ~10 000 years ago would begin to manage herds for food production that had so many reproducing individuals and that expanded so rapidly. A growth of 10% per generation implies that a population involving 150 founders would reach 1000 breeders in just 20 generations or ~60 years, if we assume a generation time of three years in, for example, goats and sheep. Archaeological research supports the notion that population growth in some of the early domesticates was not fast, because they were subjected to heavy exploitation

---

Box 1. Simulation of the effect of drift on major histocompatibility complex alleles

We used computer simulations to estimate the minimum number of founders that is necessary to retain the current level of major MHC diversity. We simulated the loss of alleles in populations following the model in Ref. [8], assuming that one or several groups of individuals of the ancestral species were isolated at once and became the founders of the domestic species.

MHC diversity was introduced into the domestic species by the founders and random genetic drift removed it. Although mutation (including recombination and gene conversion) could have also generated new alleles, we consider that it is unlikely that many alleles have arisen in this way since the time of domestication (see main text). The effect of the drift was assessed considering:
  (i) the diversity in the ancestral species
  (ii) founding population size
  (iii) number of independent domestication events
  (iv) selection coefficient
  (v) population growth rate
  For each simulation, various numbers of diploid founder individuals were randomly drawn from an unlimited ancestral population, in which

60 or 80 MHC alleles were present at identical frequencies. All founder individuals had the same probability of becoming breeders. The sampled population was mated at random with each new generation being formed by randomly sampling pairs of alleles from the previous generation. The population was allowed to grow geometrically at a constant rate between 1% and 10% per generation with non-overlapping generations, until a population size of 1000 breeding individuals was reached. At this point, the effect of drift is negligible and we conservatively assume that the alleles present in the population at that time have been retained in modern domesticates. We assumed a selection coefficient of 0.01, typical for MHC alleles [13], implying that 1% of the homozygotes do not contribute to the next generation, and we also compared the effect of different selection coefficients from 0.01 to 0.50. The total number of retained alleles was recorded when the population size reached 1000 breeding individuals. When several founding populations were considered, they were allowed to grow simultaneously until each reached 1000 individuals and the total number of different alleles was counted across all populations. Simulations were repeated 1000 times under each set of parameters.
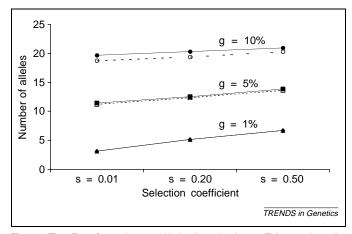
Figure 1. The effect of growth rate and balancing selection coefficient on the major histocompatibility complex (MHC) diversity in an expanding population. The number of alleles retained by a population founded by 60 individuals after growing to a population size of 1000 (Box 1) under different balancing selection coefficients (s), different geometric growth rates (g, percent increase per generation) and assuming that the ancestral population contained 80 (continuous line) or 60 (broken line) different alleles with identical frequencies (lines are almost overlapping at g=1%). Each point represents the average number of alleles retained after 1000 simulations.

and their remains did not start to appear in neighboring areas until centuries after the initial management of herds [19,20].

A more likely scenario is that there were several independent foci of domestication, as suggested by the mtDNA and archaeological data, and that the population growth was <10% per generation. In a third set of simulations (Figure 2b), we divided the founders into a variable number of populations, each growing at a rate of 5% per generation, and used selection coefficients of 0.01 and 0.20. Even in the presence of strong balancing selection, the simulations continued to show that relatively high numbers of breeding founders were required. For example, ~70–100 founders were required from four separate populations and even when six populations were involved (more than implied by most mtDNA studies [3]), 50–60 founders would be required. If the process of the domestication was independently initiated in numerous populations, the number of founders that would be required to lead to the current level of diversity would be smaller. For example, if the process was initiated in 15 populations, only a pair of individuals per population would be required. However, the small number of mtDNA haplogroups detected in domestic mammals and the archaeological evidence suggest that the separate involvement of many populations in the domestication process is unlikely.

Using the 'ancient' MHC alleles, we can infer that many founders were involved in the domestication of several mammalian species. However, the 50–100 founders suggested by these simulations are necessarily an underestimate of the true number of individuals involved. First, it is likely that extended analyses of the domestic species will reveal additional MHC alleles, because not all breeds in all continents have been equally well studied. Second, the number of founders estimated in these simulations refers to the number of unrelated individuals and, third, the simulations assume an equal contribution to
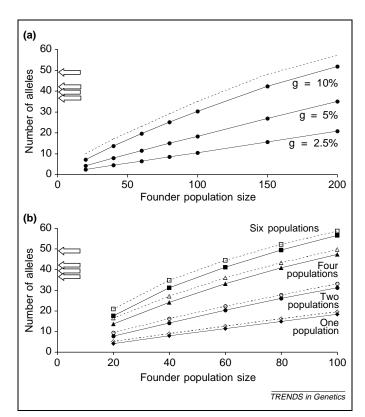

Figure 2. The effect of random genetic drift on the retention of major histocompatibility complex (MHC) alleles depending on the founding population size. Each point represents the average number of alleles retained after 1000 simulations. In each simulation, the population is allowed to grow at a specified constant rate (g) from the size indicated by the total number of founders to a size of 1000 breeding individuals per population. Founders are drawn from a population with 80 equally frequent alleles. The arrows point to the number of alleles that have at least one non-synonymous substitution in the studied domestic mammals (except cattle, for which the number of alleles would be 63; Table 1). **(a).** Different growth rates, assuming only one founding population and selection coefficient s=0.01 (the broken line represents the 95th percentile for g=10%). **(b).** Different numbers of founding populations, assuming a growth rate of 5% per generation and selection coefficient s=0.01 (continuous line) and s=0.20 (broken line).

reproduction from all individuals, which relates to the concept of effective population size (Ne). However, the contribution to reproduction is likely to have had a large variance and members of the same herd that were captured together are likely to have been related and shared alleles. Natural populations of these species, with the exception of dogs, are polygynous [21,22]. The bias in the contribution of each sex that this implies would be further accentuated during the early stages of the domestication of livestock. These stages are characterized in the archaeological record by the remains of juvenile males and some adult females, indicating that males were consumed, whereas females were retained for reproduction [19,20,23]. It is difficult to estimate the total number of individuals, N, from Ne, although the presence of non-reproducing individuals and the sex bias in reproductive contribution should result in a total population size that far exceeds the estimated founding population size. In natural wildlife populations, estimates of the ratio of N to Ne are in the order of ten [24]. Therefore, the results of these simulations suggest a founding population of many hundreds of individuals, perhaps up to 1000, was required to explain the observed level of MHC diversity in these domestic mammals.

On the one hand, our calculations might be over-estimated by ignoring the role of inter- or intra-genic recombination in the creation of MHC alleles [13], which has been predicted in several species, including cattle [25], sheep [26] and dogs [27]. It has been suggested that increased selection in new environments might occasionally lead to an increased evolutionary rate at MHC loci [28,29]. However, the rate of gene conversion mutations in class II genes has been estimated to be 8.6×10^{-6} in mice [30], or an average time of 116 000 generations per mutation. Accordingly, the allelic diversity that could have been originated since the time of domestication is not likely to be different from that generated by mutation as estimated above. On the other hand, a coalescent-based analysis of dog alleles [31] indicates that several of them might have originated through gene conversion or recombination. However, there is no evidence that many alleles could have been generated since domestication, and the presence of multiple alleles shared between dogs, wolves and coyotes (despite the poor sampling of these taxa), suggests that it is likely that most alleles originated before domestication. Making small adjustments in the number of MHC alleles to account for the effects of intra- and inter-genic recombination does not greatly influence our estimates.

The role of backcrosses

The estimated number of individuals from the ancestral species with a genetic contribution to our domestic mammals can be considered too high, given the probable capacity for herd management and food production at the time of domestication. The effect of selection and demography were insufficient in our simulations to justify the observed diversity with a smaller number of founders. The domestication of mammals in the Fertile Crescent (in the Middle East) depended on an increasing use of plant resources to feed the animals [20,32]. The emerging picture resulting from archaeological research is that herds of early domesticates were commonly small and complemented other food sources. Given these management limitations, it is possible that the current genetic diversity does not derive solely from the time of domestication, but could have been increased through extended gene flow from the wild species to the domesticates in the form of backcrosses [16,33]. Our interpretation supports the observation of slow morphological change during centuries after domestication in goats, sheep, cattle and pigs. This observation, together with the patterns of exploitation deduced from the remains of the first domestic herds, has been interpreted as a signature of frequent restocking of the managed herds with wild animals [19,20]. For the species whose ancestors were geographically restricted, such as sheep [34] and goats [16], the augmentation of genetic diversity in the managed herds might have been restricted in time and/or space to the centers of domestication. However, for the species where domestic and wild lineages have co-existed over extensive areas, such as cattle, dogs and pigs [22], occasional backcrosses might have increased the diversity of the domestic lines. Hybridization between domestic dogs and their ancestor, the grey wolf (*Canis lupus*) [35],

has been documented when they are found in sympatry [36–38] and occasionally might have been deliberately induced by some human groups [39]. It is unclear whether backcrosses were most often intentional or accidental. Although open herd management could replenish genetic diversity and increase the stock numbers, it could also result in the loss of traits that had been selected for. Perhaps backcrosses were more often permitted during the early stages of the domestication process. Our results, together with the archaeological evidence, suggest that the separation between wild and domestic might not have always been as well defined in mammals as suggested by previous genetic studies [3].

The high MHC diversity and low mtDNA diversity could be explained by preferential backcrossing of wild males and domesticated females. The hybrid offspring of domestic females could be more easily integrated into the domestic population than the reverse mating. Such backcrossing would permit the incorporation of new MHC alleles into the domestic population without the necessity of numerous founders and would minimize inbreeding depression problems that could be expected in isolated populations founded by a limited number of individuals. Backcrosses involving the incorporation of some wild males into the domestic population would have been particularly efficient in reducing the possible deleterious effects that are associated with inbreeding depression if there was a larger contribution of those males to reproduction compared with females, as occurs in modern breeding practices [40]. Consequently, our simulations suggest that the contribution of the ancestral species to the diversity of our domesticates could have gone far beyond the initial domestication events. Unfortunately, the difficulties in modeling all of the factors that could affect the role of backcrosses (including the degree and time of overlap between the distribution ranges of the domestic and wild taxa or the rate of backcrossing) hinder an estimation of their frequency. The extensive contribution of wild ancestors has left our domestic mammals with the legacy of increased genetic diversity and, hence, we have domestic mammals with the potential to adapt to disease and alterations in the environment.

References

1 Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707
2 Clutton-Brock, J. (1999) *A Natural History of Domesticated Animals*, Cambridge University Press
3 Bruford, M.W. *et al*. (2003) DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4, 900–910
4 Hughes, A.L. and Yeager, M. (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32, 415–435
5 Klein, J. *et al*. (1998) Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* 29, 1–21

6 Klein, J. *et al*. (1990) The major histocompatibility complex and human evolution. *Trends Genet.* 6, 7–11

7 Ayala, F.J. *et al*. (1994) Molecular genetics of speciation and human origins. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6787–6794

8 Vincek, V. *et al*. (1997) How large was the founding population of Darwin's finches? *Proc. R. Soc. Lond. B. Biol. Sci.* 264, 111–118

9 Kennedy, L.J. *et al*. (2001) Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. *Anim. Genet.* 32, 193–199

10 Klein, J. *et al*. (1993) The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol.* 11, 269–295

11 Seddon, J.M. and Ellegren, H. (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. *Proc. R. Soc. Lond. B. Biol. Sci.* 271, 2283–2291

12 Paterson, S. *et al*. (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3714–3719

13 Hickson, R.E. and Cann, R.L. (1997) Mhc allelic diversity and modern human origins. *J. Mol. Evol.* 45, 589–598

14 Satta, Y. *et al*. (1994) Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. U. S. A.* 91, 7184–7188

15 Aguilar, A. *et al*. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3490–3494

16 Luikart, G. *et al*. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5927–5932

17 Troy, C.S. *et al*. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088–1091

18 Savolainen, P. *et al*. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298, 1610–1613

19 Zeder, M. (2003) Hiding in plain sight: The value of museum collections in the study of the origins of animal domestication. In *Decyphering Ancient Bones; The Research Potential of Bioarchaeological Collections* (Grupe, G. and Peters, J., eds), pp. 125–138, Verlag Marie Leidorf

20 Zeder, M. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. In *The First Steps of Animal Domestication: New Archaeozoological Approaches* (Vigne, J-D. *et al.*, eds), Oxbow Press (in press)

21 Shackleton, D.M. ed. (1997) *Wild Sheep and Goats and their Relatives. Status Survey and Conservation Action Plan for Caprine*, IUCN

22 Nowak, R.M. (1999) *Walker's Mammals of the World*, The John Hopkins University Press

23 Zeder, M.A. and Hesse, B. (2000) The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science* 287, 2254–2257

24 Frankham, R. (1995) Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66, 95–107

25 Andersson, L. *et al*. (1991) Concerted evolution in a segment of the first domain exon of polymorphic MHC class II beta loci. *Immunogenetics* 33, 235–242

26 Schwaiger, F.W. *et al*. (1994) Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. *Mol. Biol. Evol.* 11, 239–249

27 Seddon, J.M. and Ellegren, H. (2002) MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics* 54, 490–500

28 Belich, M.P. *et al*. (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* 357, 326–329

29 Watkins, D.I. *et al*. (1992) New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* 357, 329–333

30 Martinsohn, J.T. *et al*. (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50, 168–200

31 McVean, G. *et al*. (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160, 1231–1241

32 Hillman, G. (1996) Late Pleistocene changes in wild plant-foods available to hunter-gatherers of the northern Fertile Crescent: possible preludes to cereal cultivation. In *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (Harris, D.R., ed.), pp. 159–203, Smithsonian Institution Press

33 Hedrick, P.W. *et al*. (2000) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity* 85, 617–624

34 Hiendleder, S. *et al*. (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc. R. Soc. Lond. B. Biol. Sci.* 269, 893–904

35 Vilà, C. *et al*. (1997) Multiple and ancient origins of the domestic dog. *Science* 276, 1687–1689

36 Randi, E. and Lucchini, V. (2002) Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conserv. Genet.* 3, 29–43

37 Andersone, Z. *et al*. (2002) Hybridisation between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. *Mamm. Biol.* 67, 79–90

38 Vilà, C. *et al*. (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. *Heredity* 90, 17–24

39 Schwartz, M. (1997) *A History of Dogs in the Early Americas*, Yale University Press

40 Cunningham, E.P. *et al*. (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet.* 32, 360–364