

Molecular Approaches to Understanding Animal Domestication: What Have We Learned So Far?

Michael W. Bruford*, Biodiversity and Ecological Processes Group, School of Biosciences,
Cardiff University, PO Box 915, Cathays Park, Cardiff CF10 3TL, Wales, UK.

Summary

Molecular genetic analysis has been applied to the study of animal domestication since the advent of allozyme markers in the 1960's and 70's. However, it is only recently that DNA sequence-based studies have had a major impact on understanding domestication in animal species. A complex picture of domestication in both the New World and Old World has emerged. Through analysis of multiple marker types and especially mitochondrial, Y-chromosomal and nuclear DNA in modern breeds and their potential wild and domestic ancestors, and by using both 'modern' and 'ancient' DNA sources, we continue to gain new insights into the events that produced the domestic animals of today. Of note has been the surprisingly high number of domestication events inferred from these data and the diverse locations in which they appear to have taken place. These results do not only alter our perceptions of the domestication process, but will also influence our approach to domestic animal diversity conservation, especially in the area of prioritisation. Here I update a recent review some of these studies and highlight some recent advances in domestic animal evolutionary genetics, and discuss their implications for future research directions.

Introduction

Understanding the history of human populations and the animals we domesticated was until recently squarely in the domain of evolutionary anthropology and archaeozoology. Recently, molecular genetics has changed this situation and has allowed the identification of the wild ancestors of many modern domestic animals and the nature of the expansion of domestic animal populations in recent millennia (Bruford et al 2003; Goodrich and Wiener, 2005).

Many (but not all) animal domestication events are now thought to have occurred at three locations: southwest Asia, eastern Asia and the Americas, 8–10,000 years ago. More specifically, these areas comprised the 'fertile crescent' and its eastern margin, towards the Indus valley region, China and mainland south-east Asia) and the Andean chain of South America. Cattle (*Bos taurus/indicus*), sheep (*Ovis aries*), goats (*Capra hircus*), pigs (*Sus scrofa*), camels (*Camelus* spp) and buffalo (*Bubalus* spp) are now thought to have been domesticated in the two Asian regions, whereas llamas (*Lama guanicoe*), alpacas (*Vicugna pacos*) and guinea pigs (*Cavia porcellus*) were domesticated in South America (Diamond, 2002). A few notable exceptions exist. Genetic evidence for the horse (*Equus caballus*) points to more recent domestication at a number of locations in more northerly latitudes (Lister et al, 1998; Vila et al, 2001; Jansen et al, 2002). The geographical origin of the dog (*Canis familiaris*) is still debated, with multiple maternal lineages evident and the region responsible primarily inferred as East Asia, (Vila et al 1997; Savolainen et al 2002; Leonard et al 2002). To further complicate the picture, recent analysis of largely ancient DNA samples from wild boar (*Sus scrofa*; Larson et al 2005) has thrown new light on pig domestication and shows that pigs were domesticated from a larger number of genetic lineages, found widely dispersed across Eurasia. Finally, a new domestication arena (Africa) has recently been proposed for the donkey (Beja-Pereira et al, 2004).

Molecular markers utilised in domestication studies

Primarily molecular domestication studies have focused on a phylogenetic approach to date, and as such DNA sequence analysis and especially the mitochondrial (mt) DNA control region (CR) has been used. This is the marker of choice for many studies of vertebrate phylogeography (the study of DNA lineages in a geographic context; Avise, 2000). Since mtDNA is predominantly maternally inherited in a single copy (haplotype) and because the evolutionary rate of mtDNA and particularly the CR can be orders of magnitude faster than nuclear DNA, studies using the CR have found that it is both highly polymorphic and structured enough across the range of domestic species and their potential wild relatives so that phylogenetic affinities and the approximate geographic region of domestication can be inferred. For example, Luikart et al (2001) could examine the phylogenetic affinities of over 300 distinct haplotypes in domestic goats (*Capra hircus*) which were identified from a study of just over 400 individuals. MtDNA is not only a powerful tool for establishing the levels of genetic diversity and phylogenetic structure within a species, but it can also tell us about the recent demographic processes affecting a population, for example whether a population has undergone a demographic expansion. However, although mtDNA can be extremely informative in domestication studies (both within and among species), its uses are limited. It can be a poor predictor of overall genomic diversity, because it behaves like a single locus and is an extra-nuclear genetic marker with specific evolutionary dynamics. Since mtDNA is maternally inherited, it does not detect male gene flow, which has a powerful influence on the evolution of domestic species in modern times (MacHugh et al 1997; Kadwell et al 2001). Further, the segregation of an mtDNA lineage within a livestock population can only have occurred through the domestication of a wild female or through the incorporation of a female into an existing domestic stock, which might not always be the case. The capture of male genetic lineages during domestication can potentially be tracked using the Y chromosome, as has recent been demonstrated in the wisent (Verkaar et al 2004). However, this chromosome is much less variable within species than most other genomic sequences, which prevents its routine use for phylogenetic analyses. Although they lack the phylogenetic precision of mtDNA sequences, microsatellites (Bruford and Wayne, 2003) are being used more in domestic animal studies – particularly at the breed level, where they have been used to great effect to describe structure within domestic species at a variety of levels, but primarily at the level of the breed. This approach has been applied in most domestic animals, including chickens (Rosenberg et al 2001), dogs (Parker et al 2004) and many livestock populations (e.g. MacHugh 1994, Hanotte et al 2002). The advantage of codominant markers is obvious in a situation where domestication and introgression may be bidirectional and the real power of these markers is when they are used in combination, as in the case for example, of the llama and alpaca (Kadwell et al 2001) and African cattle (MacHugh et al 1997; Hanotte et al 2000).

Domestication questions

Ancestors: In cases where material (modern or ancient) is available from the putative wild ancestors of modern domestic species, it is possible to compare these species or populations in a phylogenetic context to establish the most likely progenitor. This has been a major focus for the last ten years and a number of surprising results have emerged. Perhaps the best example to date (due to its ‘completeness’) is in cattle, where studies of both zebu and taurine types initially posed the question of whether both types were merely differentiated forms that originated from a single domestication event,

or whether they represented the products of separate domestications (Loftus et al 1994a,b; Bradley et al 1996). Taurine cattle from Europe and Africa and zebu cattle from India and Africa were compared at the molecular level. Indian zebu cattle have been found to have profoundly different mtDNA control-region sequences when compared with both European and African taurines and African zebu. The sequence divergence is consistent with a shared common ancestor of the order hundreds of thousands of years ago. Since cattle domestication has occurred during the last 10,000 years, the most probable explanation is that different subspecies of the ancestral wild cattle *Bos primigenius* has been domesticated in different regions of Europe and Asia.

MtDNA has also revealed a similarly complex pattern in goats (Luikart et al 2001), with at least three divergent lineages resulting from at least two domestication events in the Near East and in Asia. Further, at three control-region lineages have also been described in sheep (*Ovis aries*; Guo et al 2005; Bruford and Townsend, 2005) which correlate well, but not completely, with modern fat- and thin-tailed phenotypic varieties, and indicate three possible domestication events for this species, just as in goats. However to date a paucity of modern wild sheep and goat samples has hampered more accurate pinpointing of where and when these events took place. Two mtDNA lineages representing swamp and water buffalo (*Bubalus bubalis*) have also been described, which indicates that there were independent domestication events (Takanaka et al 1996; Lau et al 1998). While in South America studies of the endemic wild and domestic camelids (Kadwell et al 2001) asked the question of whether the modern-day alpaca (*Vicugna pacos*) and llama (*Lama glama*) descend from either the wild guanaco (*Lama guanicoe*) or vicuña (*Vicugna vicugna*). MtDNA analysis showed that the llama and alpaca have haplotypes that originate from both wild species, but that most of the individuals have guanaco-type mtDNA. However, microsatellites showed that this inference was primarily the result of hybridization between male alpaca and female llama. Finally, in Africa a recent study of the domestic donkey (*Equus asinus*; Beja Pereira et al 2004) showed that donkeys were domesticated twice, from two wild ass subspecies in Nubia and Somalia.

Ancient DNA: It is becoming clear that studies attempting to pinpoint wild ancestors of domestic species will need to rely, where possible, on DNA from ancient specimens taken from potential sites of domestication. For example, the origins of European taurine cattle were enhanced by a large-scale study of mtDNA CR variation in 400 cattle from Europe, Africa and the Near East (Troy et al 2001). Skeletal remains found in the United Kingdom from four wild aurochs were found to be 3,720–7,320 years old and their phylogenetic relationships were determined. Interestingly, they were found to be unrelated to modern-day UK cattle, which were instead more similar to modern cattle in the Near East, indicating that at least the mtDNA origins of European cattle are consistent with the spread of cattle by cattle farmers from the Fertile Crescent, and not local, independent domestication. Clearly such an analysis needs to be supplemented with nuclear DNA if possible, since local introgression from aurochs bulls would not be picked up in such a study. Most recently, a study by Larson et al (2005) used ancient DNA methods to disprove an earlier hypothesis (Guiffra et al, 2000) that pigs were domesticated at two locations in eastern and western Asia. MtDNA CR sequences from 686 wild and domestic pigs were instead revealed multiple centers of domestication across Eurasia and that European, rather than Near Eastern, wild boar are the principal source of modern European domestic pigs. This last inference is extremely interesting in that it is the first concrete indication that Europe might have played a role in the domestication of modern human-associated species.

The use of ancient DNA has also proved crucial in understanding two much more complex species – the dog and the horse. For example, analysis of both ancient and modern DNA in the horse (*Equus caballus*) implies a highly complex process (Lister et al, 1998; Vila et al, 2001; Jansen et al, 2002). Jansen et al (2002) found mtDNA CR sequences grouped into at least 17 lineages. Although, some groups clustered with modern breeds or regions, but many did not. The authors estimated that minimally 77 mares must have been domesticated to produce modern levels of CR diversity and it therefore seems likely that horses were domesticated numerous times. A summary of these findings is presented in Table 1.

Table 1 – inferred domestication events from molecular data for some modern domestic animal species (updated from Bruford et al 2003)

Species	Possible ancestors	Extant ancestors	# inferred domestications	Average CR divergence among haplogroups
<i>Bos indicus</i> <i>Bos taurus</i>	<i>B. primigenius primigenius</i> <i>B. primigenius namadicus</i>	0	2 (Asia and Africa) 1	0.057
<i>Ovis aries</i>	<i>O. orientalis/musimon</i> <i>O. ammon</i> <i>O. vignei</i>	3	3 0 0	0.045
<i>Capra hircus</i>	<i>C. aegagrus</i> <i>C. falconeri</i>	2 (<i>C. aegagrus</i> has 3 subspecies)	3 0	0.041
<i>Equus caballus</i>	<i>E. caballus</i>	0	Multiple (Eurasia)	0.040
<i>Equus asinus</i>	<i>E. africanus</i> <i>E. hemionus</i> <i>E. kiang</i>	3 (<i>E. africanus</i> has 2 subspecies)	2 (Nubia / Somalia) 0 0	0.033
<i>Sus scrofa</i>	<i>S. scrofa</i> <i>S. indicus</i>	2	Multiple (Eurasia): up to six	Not described
<i>Bubalus bubalis</i>	<i>B. bubalis bubalis</i> <i>B. bubalis carabanensis</i>	2	2	0.074
<i>Canis familiaris</i>	<i>C. lupus</i>	1	Multiple	0.047
<i>Lama glama</i>	<i>L. guanicoe</i> <i>V. vicugna</i>	2 (both species have two subspp.)	Not described	Not described
<i>Vigugna pacos</i>	<i>L. guanicoe</i> <i>V. vicugna</i>	2 (both species have two subspp.)	Not described	Not described

Contemporary patterns

Modern day patterns of diversity in domestic animals have been influenced by ancient and recent human movements, selection, introgression, admixture, genetic drift and inbreeding. A combination of markers has been used to allow inference of these events, and have particularly involved examining diversity in populations with reference to populations found in putative historic centres of domestication as identified above. For example, recently Cymbron et al (2005), following on from work originally done by Loftus et al (1999) examined hypotheses on the expected genetic legacy of a putative Neolithic spread of agriculture across Europe following two main paths: a ‘Danubian route’ whereby Neolithic farmers expanded north across the central European plains and a ‘Mediterranean’ route, where migration is supposed to have occurred along the Mediterranean coast. By examining 20 European cattle breeds chosen to test these hypotheses the authors supported the two

distinct cattle migrations. However, this study also reinforced previous evidence that Mediterranean cattle breeds may have had more recent input (introgression) from both the Near East and Africa.

Disentangling historical patterns of movement from modern breeds has therefore proved to be a challenge. For example, studies using microsatellites and Y-chromosome DNA showed that the nuclear DNA of African zebu cattle was most similar to that found in other zebu cattle populations, in stark contrast to the results from mtDNA. The reason for this difference was that mtDNA studies were not able to detect zebu genes (passed into the local taurine population by zebu bulls), because male mitochondrial DNA alleles are not transmitted to descendent generations (MacHugh et al 1997). Geographical patterns of genetic diversity in livestock change as rapidly as new studies are published and as new lineages are discovered – recent examples in sheep and pigs (Guo et al, 2005; Bruford and Townsend 2005; Larson et al 2005) suggest that the picture is only going to become more complex, especially with species which are easily moved (e.g. goats where a major centre of current genetic diversity is in the Alps; Luikart et al, 2001) or where multiple human migrations and local agricultural practices combine to produce complex patterns of admixture and multiple origins of livestock (e.g. Wood and Phua 1996; Hanotte et al 2002).

In other species, major studies of geographical patterns of genetic diversity have either yet to be published or the results seem to be very different. In horses there is some association between mitochondrial lineages and geography. Northern European ponies share a distinct mitochondrial lineage, as do Iberian and northwest African breeds (Jansen et al 2002) although it is hard to find other geographic patterns, which is likely to be the result of the extensive movement of horses. Horses have accordingly, low intercontinental population differentiation compared to other livestock species, which is consistent with their use mainly for transportation. The combined effects of portability and/or mobility (e.g. in goats and horses) and introgression (e.g. in cattle and sheep) has shaped the distribution of genetic diversity that we see in livestock globally today.

Limitations of current data and future directions

It seems clear from the above that to effectively document and conserve the vanishing diversity which occurs within and between modern domestic animal breeds, it is imperative that different marker types are used to assay variation. The utility of mtDNA and microsatellites are hopefully clear from the material presented here, but given the huge influence that male-mediated gene flow can have on diversity, sufficiently variable Y-chromosome markers will be important for integrated genetic studies in the future. Further, the molecular markers used to characterize diversity in livestock have, until recently, had little to do with the genes under selection for economically important traits. One recent example showed the power of including such genes. Beja-Pereira et al (2003) studied geographic patterns of variation in genes encoding for six milk proteins in European cattle. They found a geographic correlation between genetic diversity in milk proteins, the putative locations of Neolithic cattle farming sites and lactose tolerance in humans. The identification of mutations for phenotypic variation will add a new dimension to the characterization of animal domestication, as it will allow researchers to trace selection and the spread of economically important alleles, as above. Such analysis is potentially very powerful. The recent publication of the chicken genome and associated SNP markers (Wong et al 2004) open up the possibility of high resolution analysis of genetic diversity in ancient and modern chicken populations. The exciting prospect is in front of us of being able to simultaneously

detect male and female demographic history and the signatures of selection within the genomes of our domestic animals.

Most importantly, recent advances in ancient-DNA analysis and collaboration between geneticists and archaeologists are set to advance domestic animal genetic studies and enhance our ability to infer their evolutionary and demographic history (e.g. Larson et al 2005). Ancient DNA is already helping us to pinpoint the geographic locations of origin for the different domestic animals, and to determine the timing and routes of their spread across the globe. The addition of new species (such as the domestic chicken) and the addition of more samples and methods of analysis, mean that the story for many species is likely to increase in complexity and to completely overturn previous hypotheses and suppositions about man's influence on his most important non-human companions.

References

1. Avise JC (2000) *Phylogeography : The History and Formation of Species*. Harvard University Press, Harvard.
2. Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Chamberlain AT, Nunes TP, Metodiev S, Ferrand N, Erhardt G (2003) Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genet* 35: 311-313.
3. Beja-Pereira A, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, Mashkour M, Jordana J, Taberlet P, Luikart G (2004) African origins of the domestic donkey. *Science* 304: 1781-1781.
4. Bradley DG, MacHugh DE, Cunningham P and Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl Acad. Sci. USA* 93: 5131–5135.
5. Bruford MW and Wayne RK (1993) Microsatellites and their application to population genetics. *Curr. Opin. Genet. Dev.* 3: 939–943.
6. Bruford MW, Bradley DG and Luikart G (2003) Genetic analysis reveals complexity of livestock domestication. *Nature Rev. Genet.* 4: 900-910.
7. Bruford MW and Townsend SJ (2005) Case studies in the genetics of animal domestication: sheep. In “*Documenting Domestication: New Genetic and Archaeological Paradigms*” Eds. M. Zeder, D. Decker-Walters, D. Bradley, and B. D. Smith. California University Press.
8. Cymbron T, Freeman AR, Malheiro MI, Vigne JD, Bradley DG (2005) Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proc R. Soc Lond. B.* 272: 1837-1843.
9. Diamond J (2002) The evolution, consequences and future of plant and animal domestication. *Nature* 418: 700–707.
10. Goodrich J, Wiener P (2005) A walk from the wild side: the genetics of domestication of livestock and crops. *Bioessays* 27: 574-576.
11. Guiffra E, Kijas JMH, Armager V, Carlborg Ö, Jeon J-T and Andersson L (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154: 1785-1791.
12. Guo J, Du L-X, Ma Y-H, Guan W-J, Li H-B, Zhao Q-J, Li X and Rao S-Q (2005) A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim Genet* 36: 331.
13. Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J, Rege JEO (2000) Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol. Ecol.* 9 (4): 387-396.
14. Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW and Rege EO (2002) African pastoralism: genetic imprints of origins and migrations. *Science* 296: 336-339.
15. Jansen T, Forster P, Levine MA, Oelke H, Hurler M, Renfrew C, Weber J and Olek K (2002) Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl Acad. Sci. USA* 99: 10905–10910.
16. Kadwell M, Fernandez M, Stanley HF, Wheeler JC, Rosadio R and Bruford MW. (2001) Genetic analysis reveals the wild ancestors of the llama and alpaca. *Proc. R. Soc. Lond B.* 268, 2575-2584.
17. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson, H, Brand, T, Willerslev, E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307: 1618 – 1621.
18. Lau CH, Drinkwater RD, Yusoff K, Tan SG, Hetzel DJ and Barker JSF (1998) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial D-loop and cytochrome b sequence variation. *Anim Genet* 29: 253-264.

19. Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S, Vila C (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* 298: 1613-1616.
20. Lister AM, Kadwell M, Kaagan LM, Jordan WC, Richards MB and Stanley HF (1998) Ancient and modern DNA from a variety of sources in a study of horse domestication. *Anc. Biomol.* 2: 267-280.
21. Loftus RT, MacHugh DE, Bradley, DG, Sharp, P. M. and Cunningham, P. Evidence for two independent domestications of cattle. (1994a) *Proc. Natl Acad. Sci. USA* **91**, 2757-2761.
22. Loftus RT, MacHugh DE, Ngere LO, Balain DS, Badi AM, Bradley DG, Cunningham EP. (1994b) Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim.Genet.* 25: 265-271.
23. Loftus RT, Ertugrul O, Harba AH, El-Barody MAA, Machugh DE, Park SDE, Bradley DG (1999) A microsatellite survey of cattle from a centre of origin: the Near East. *Mol. Ecol.* 8: 2015-2022.
24. Luikart G, Gielly L, Excoffier L, Vigne JD, Bouvet J, Taberlet P (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. USA* 98: 5927-5932.
25. MacHugh DE, Loftus RT, Bradley DG, Sharp PM and Cunningham P (1994) Microsatellite DNA variation within and among European cattle breeds. *Proc. R. Soc. Lond. B* 256: 25-31.
26. MacHugh, DE, Shriver, MD, Loftus, RT, Cunningham, P and Bradley, DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146: 1071-1086.
27. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L (2004) Genetic structure of the purebred domestic dog. *Science* 304: 1160-1164.
28. Rosenberg NA, Burke T, Elo K, Feldmann MW, Freidlin PJ, Groenen MAM, Hillel J, Maki-Tanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* 159: 699-713.
29. Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610-1613.
30. Tanaka K, Solis CD, Masangkay JS, Maeda K, Kawamoto Y, and Namikawa T. (1996) Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of the cytochrome *b* gene. *Biochem Genet* 34: 443-452.
31. Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain A, Sykes BC and Bradley DG (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410: 1088-1091.
32. Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, Lenstra JA (2004) Maternal and paternal lineages in cross-breeding bovine species: has wisent a hybrid origin? *Mol. Biol. Evol* 21: 1165-1170.
33. Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. *Science* 276: 1687-1689.
34. Vila C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, Liden K, Wayne RK and Ellegren H (2001) Widespread origins of domestic horse lineages. *Science* 291: 474-477.
35. Wong GKS et al. (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432: 717-722.
36. Wood, NJ and Phua SH (1996). Variation in the control region sequence of the sheep mitochondrial genome. *Anim Genet.* 27: 25-33.