ALPACA WORLD

# Alpaca DNA: From Wool to Genes (And Back Again!)

*Mike Bruford*

## Background

Over the last twenty years the impact of DNA technology in all sectors of society has been immense, and the science has advanced at a rapid (some would say, frightening) pace. Until the invention of DNA fingerprinting, by Sir Alec Jeffreys at Leicester University in 1985, molecular geneticists (those strange people like me who work with DNA in its rawest form) had beavered away largely unnoticed in laboratories, developing methods now widely known to the public such as DNA sequencing and genetic modification. It was in the mid-eighties, however, that it became clear that it would be soon possible to automate these labour-intensive techniques, create genetic ‘maps’ of chromosomes using DNA ‘markers’ as signposts and that reading the entire sequence of DNA from a living organism would indeed be possible in the foreseeable future. The application of these new methods quickly turned towards livestock, and ten years later, the first outline genetic maps of cattle, sheep and pigs had been produced. But rather like the early maps produced by cartographers of the world’s immense continents such as Africa and the Americas, we only had an ‘outline view’ at that stage and, for example, although we knew roughly where the five major chromosomal regions were that controlled milk production traits in cattle, we did not know precisely where, or what, the genes were.

A massive increase in lab capacity and automation of the 1990’s, including the setting up of the world’s first genome factories, meant that this quickly changed. The full sequence (all 3,000,000,000 units – or bases) of the 23 chromosomes that make up the human genome was published in draft in 2001 and was confirmed last year: the cattle and pig genomes are now well on their way to completion. We now know precisely what some of the genes are which control milk quality, and the use of DNA profiling (the modern equivalent of Jeffreys’ DNA fingerprinting) is routine in forensic science, relying on the enormous National DNA Database containing the genetic profiles of over a million Britons.

These methods and the vast libraries of information they produce are set to revolutionise our understanding of the genetic basis of any trait we care to examine, including for example disease susceptibility and performance characteristics. The possibilities seem endless, but the consequences are equally profound. The genetic diversity we see in the natural world today is the result of millions of years of evolution – even domesticated animals and crop plants have a unique history, thousands of years old. It is, ultimately, up to society to decide how we use this information – the power is potentially immense, and has already resulted in such controversial advances as ‘disease profiling’ in humans, GM food and the cloning of Dolly the sheep.

DNA methods have therefore led a biological revolution and presented society with many possibilities and choices. These same possibilities and choices are beginning to become available to the alpaca breeder and some forms of genetic testing are already being used in different countries. What I intend to do in the remainder of this article is outline the DNA work we have done so far, how that has changed our views about the alpaca, what tests are available and what is likely to be available in the future.

## What IS an alpaca?

This seems like a simple enough question to answer. However, nothing could be further from the truth! And for a while, during studies my colleagues Jane Wheeler, Miranda Kadwell, Raul Rosadio, Mati Fernandez and I carried out in the late 1990's, the origin and identity of the modern alpaca seemed as clear as that of the Push-Me-Pull-You… What we did know is that four South American camelids were recognised, two of which are wild species, the guanaco (*Lama guanicoe*) and the vicuña (*Vicugna vicugna*) with two domestic forms, the alpaca, classified at the time as *Lama pacos* and the llama, *Lama glama*. The origins of the domestic forms had long been debated and remained far from clear when we started our work. We also knew from genetic and other work carried out by Jane, Miranda, Raul and Helen Stanley in the early 1990's that the guanaco and vicuña evolved from a common ancestor 2-3 million years ago and that they are the only representatives of the wild South American camelids to survive the Pleistocene epoch (a period of major climate fluctuations spanning roughly 2 million years to 20,000 years ago).

This meant that we had an unusual and exciting opportunity to solve the origin of the domestic forms, since all potential ancestral forms still existed. This is sadly not the case in some species such as the modern horse and cattle, posing tremendous problems for zoologists interested in their origins. Our motivation for trying to solve this puzzle was a realisation that understanding the origin and current diversity of South American camelids (SACs hereon) was an issue of increasing importance due both to the potential economic benefits of proper management and the huge the threats that face marginal and extensive agriculture today, especially in the Andean region.

The difficulty with alpaca and llama is that, in contrast with other livestock, there is no written history associated with how they were originally reared. The presumably rich oral traditions of transferring herding knowledge appears to have been largely lost during the Spanish conquest, and the archeological record shows that many SAC breeds disappeared when both the human and native livestock populations were reduced by 80-90% during the first 100 years of contact. As will be well known to many readers of this magazine, currently, although llama and alpaca rearing is a central element of the economy in the high Andes, it is often not profitable due to the poor quality of the animals and their fibre. We felt at the time (and still feel) that the reconstruction of fine fibre breeds could be greatly aided by the contribution of genetic analysis, but how?

Traditionally, both the llama and alpaca were thought to have been domesticated from a guanaco ancestor, although tantalising archeological evidence had suggested a link between the alpaca to the vicuña, dating domestication to 6 - 7,000 years ago in the Peruvian Andes. Unfortunately, understanding the domestication of llama and alpaca using methods such as skull, teeth and skeletal measurements had been hampered due to a lack of statistical power, and because archaeological remains are often in poor condition. Fibre analysis had been pursued at sites where fleece has been preserved, largely by Jane Wheeler and colleagues, but the decimation of traditional llama and alpaca breeds at the time of the conquest and  the likelihood of hybridisation among the survivors made understanding the results in the absence of genetic evidence practically impossible.

This was a problem crying out for a DNA-based approach. Such methods had already been used to identify the origins or domestic species such as cattle, sheep, goat and pig, so why not the alpaca? The key to solving the problem was to collect and analyse of a good sample of both the wild and domestic populations throughout their entire geographic range. This was potentially an enormous task, but fortunately one that had been largely carried out already by Jane Wheeler, Helen Stanley and Raul Rosadio during their previous studies. Collaboration with our colleague Ricardo Baldi in Patagonia, who studies the wild guanaco in the southern end of its range, allowed us to approach an analysis with confidence for the first time.

We were able to sample both vicuña subspecies (*V. v. vicugna* and *V. v. mensalis*) from Argentina, Chile and Peru, guanaco (*L. g. guanicoe* and *L. g. cacsiliensis*) from Argentina and Peru, llama (a wide range of morphological types) from Argentina, Bolivia and Peru, alpaca (including 'suri and 'huacaya'), from Argentina, Bolivia, Chile and Peru and both wari and pacovicuña. Samples were taken only from those individuals whose phenotype conformed to accepted criteria for domestic forms. Our sample included 440 vicuña, 122 guanaco, 60 llamas and 141 alpacas.

## DNA Methods

We needed to analyse two types of DNA (and this is where it gets a little technical). The first type was mitochondrial (mt) DNA – a rather unique sequence found in the mitochondria within our cells (small compartments or ‘organelles’ responsible for energy production). MtDNA is used by evolutionary biologists to construct ‘trees’ of species, linking them back to a common ancestor just as in a family genealogy. The key with mtDNA is that it accumulates new mutations up to 10-fold more rapidly than any other sequence in  the genome. These mutations allow us to find species-specific sequence characters (and often many of them) and by comparing between species, work out which species is related to which on the basis of their relative similarity. This approach has been used for thousands of species and is the method used by Jane Wheeler and colleagues in their first study of 1994. What we were hoping to find was an association between the mtDNA of the domestic forms and their potential wild ancestors, solving the problem once and for all. However, this approach didn’t give promising results in the 1994 study because the small sample of llama and alpaca were found to share mtDNA sequences with both vicuña and guanaco. We hoped this inconclusive result would be improved by our much expanded sample. There was one problem: mtDNA is only inherited in the maternal line, which does not matter when comparing different species, but could prove misleading when looking at domestic forms which might have undergone hybridization. The mtDNA of a male llama who produces a cria with a female alpaca will not be detected, even though the offspring is a hybrid.

What we found really puzzled us. We found a lot of diversity (good news!), with 26 unique sequences present. All vicuña and all guanaco grouped together but were mutually separated by 9 mutations, in accordance with their divergence over 2-3 million years. We used Arabian camel as an evolutionary ‘outgroup’ and this was indeed more distinct with 21 mutations separating it from New World species. The domestic forms were more enigmatic: 81% were found grouped with guanaco, including nearly all llamas and 75% of alpaca. A much smaller number grouped with vicuña – these were nearly all alpaca and pacovicuñas.

So what could explain this? The fact that the great majority of domestic samples grouped with guanaco could be seen as support of a guanaco origin for both domestic forms, but the occurrence of alpaca in signficant numbers in both groups was confusing – perhaps they had a dual origin? Alternatively, hybridization and 'movement' of mtDNA between species, via females might be obscuring the picture. It was for this reason that we decided to find genetic markers that would discriminate among the wild species in the chromosomes found in the nucleus of the cell, where 99.9% of a mammal's 45,000-or-so genes reside. This could then potentially tell us whether the alpaca had a guanaco, hybrid or much more unlikely (according to the mtDNA data), vicuña origin.

However, finding genetic markers that could definitively distinguish between vicuña and guanaco would not be straightforward. The only commonly used chromosomal markers in New World camelids were those developed in the mid 90's for paternity testing. These markers, called microsatellites, are highly variable, with up to thirty variants (alleles) per marker. A mammal will possess up to two different alleles, since one allele is inherited from each parent. It is these markers that are routinely used for the National DNA Database in the UK. The use of just six of these markers together allows the genetic identification of individuals with an error rate of less than a million to one – every human, with the exception of identical twins, in the UK has a unique profile using such markers, with so much inherent variation, surely they could not be used to discriminate among species? In what can only be described as one of those rare, 'eureka!' moments, in 1998 my colleague Miranda Kadwell had been testing a number of the 30 available markers for a project we had just started on wild vicuña and guanaco. Her aim was to see if the markers, developed in llamas and alpacas in the USA, were variable in a few individuals from the Peruvian wild populations. We noticed with surprise that one of the markers (named YWLL 46) consistently produced alleles in vicuña that were completely different from the guanaco samples. Could we have stumbled upon the marker we were looking for by accident? Extensive testing on many more wild samples confirmed the pattern and, furthermore, we found that one other of the 30 available markers behaved in the same way. We finally had a way to address the domestication question using markers inherited from both parents.

## Our results and their implications

When we tested these markers in alpaca and llama, we were in for another big surprise. The vast majority of alpaca shared alleles at these markers with the vicuña, not the guanaco, whereas llamas were much more similar to the guanaco. Analysis of further markers, which did not show completely distinct alleles from the wild species further backed up these initial findings: alpacas were more similar to vicuña than to guanaco in their chromosomal DNA. However, a consistent and detectable level of hybridization is present in both domestic forms, confirming the findings of the mtDNA study. There seem to be llama alleles in alpaca and alpaca alleles in the llama – their frequency in the 'wrong' species can be as high as 11%.

To get a fuller picture, we decided to put together both the mtDNA results and the microsatellite markers in 211 individuals for which we had complete data. The results were intriguing and disturbing at the same time. The picture is very different for both markers, with the proportion of guanaco (more likely llama) mtDNA in modern day alpacas being very high at around 70%, but the chromosomal DNA markers show only approximately 15% hybridization. Overall, if you combine the evidence, while 60% of llamas showed no evidence of hybridization, the same can be said for only 20% of alpacas.

What underlies this pattern? The results give a complex picture of how modern day New World camelid genetics been shaped by domestication with the added impact of events relating to the Spanish conquest. A similar complex picture has emerged from studies of African cattle where the genes from humped cattle (zebu) show very similar patterns of occurrence in taurine (non-humped, indigenous) cattle in sub-Saharan countries. Here mtDNA has barely entered the taurine population due to the exclusive use of zebu bulls during the spread of Islamic pastoralists, but the picture with microsatellites and Y-chromosome (male) DNA is one of much greater hybridization. In our study, only 27% of the alpacas we sampled are mitochondrially 'vicuña', although 40% of alpaca possess only vicuña microsatellite alleles, but have guanaco mtDNA haplotypes. This suggests that the introduction of guanaco (or more likely llama) mtDNA some time in the past may have occurred a number of times within alpaca populations but may have been accompanied more recently by a reversion to line or stock breeding within local alpaca populations, resulting in the partial elimination of llama chromosomes, while the mtDNA 'contamination' has remained fixed in the alpaca lines.

The implications of these data are potentially important for Peru and elsewhere. In our sample, only 35% of domestic animals had not undergone any detectable hybridisation based on mtDNA and the two diagnostic mcrosatellites. In particular there are a very large number of hybrids within the alpaca population (80%). The reason for this may be very recent in origin. During the last 20-25 years large-scale hybridisation between llamas and alpacas has been carried out in the Andes. Specifically, male alpacas have been bred to female llamas to increase the population of animals producing higher priced "alpaca" fibre, and male llamas have been bred to female alpacas to obtain greater fleece weights, increasing income. With sale price traditionally determined by weight, and little consideration given to fineness, the quality of alpaca fibre is thought to have decreased over the past 25 years. Indigenous Quechua and Aymara speaking herders subdivide the hybrids into llamawari or waritu (llama-like) and pacowari or wayki (alpaca-like) respectively, depending upon physical appearance. The first generation offspring are fertile, usually intermediate in size and can be back-crossed to either parental type. In addition recent intensive selection for white fleece in modern alpaca may also have involved both types of hybridisation. A combination of these recent breeding practices with our results could explain the confusion surrounding the domestic forms in the recent past, since it is likely that many specimens used in previous studies were hybrids.

Given the extreme hybridisation in present day South American alpacas, DNA analysis has been crucial to resolving the origin of this domestic form. Our results strongly supported the vicuña as the ancestor of the alpaca, and we therefore proposed that the classification of the alpaca should be changed from *Lama pacos* to *Vicugna pacos*. This proposal has now been accepted. The degeneration of quality and value in present day SACs in Peru has therefore probably been due to extensive hybridisation, probably beginning with the Spanish conquest and continuing to the present day. When it was believed that these crosses were between different forms of a single domestic animal descended from the guanaco, there was little concern about the economic impact of this hybridization. However, given that our work has shown that the alpaca is much more likely to be descended from the vicuña, the impact of such crosses is now clear.

The use of DNA analysis to identify hybrid animals in the breeding pool in Peru seems appropriate, since the ongoing hybridisation process makes it impossible to accurately identify all hybrids on the basis of how animals look or measure. However, the knowledge that the alpaca descends from the vicuña opens new routes for the improvement of alpaca fibre production not only through the identification of hybrids and their removal from pure bred elite herds, but via the back crossing of pure-bred alpacas to their vicuña ancestor in order to possibly improve fibre fineness.

While 90% of the alpaca fibre produced in Peru has a diameter greater than 25 µm and fetches low prices on the world market ($3-30/kg 1980-1995), preconquest animals produced fibre of 17-22 µm, similar to cashmere (15-17 µm: $60-120/kg 1980-1995). It is possible therefore that identification of the remaining pure alpacas may aid in recovery of the fine fibre characteristics of preconquest animals. Such a recovery would indeed have a major impact on the economy of the Andes.

**The future**

I am often asked about what the implications of these results are for alpaca producers outside of Peru and how genetic tests may help the UK industry in particular. First, identification and databasing of individual animals and pedigree verification seem to be an entirely appropriate use of genetic markers when carried out in registry management. Indeed a compete and accurate picture of the relationships of animals within a registry seems necessary to maintain a healthy (i.e. non-inbred and non-crossbred), proved and benchmarked population. Such an approach is being used extensively already in the US with microsatellite markers. It seems highly likely that the UK registry will go in the same direction over the next year or so. The genetic diversity in UK alpacas can undoubtedly be better managed as a result of knowing how related animals are to each other, and such information may help guide the industry further in its deliberations about the benefits of operating a closed registry.

Second, the genetic markers we have used to address the origins of the alpaca are a crude and potentially misleading way to measure 'purity' in alpacas or llamas. The relationship between the proportion of vicuña alleles using the markers we have analysed and fibre fineness in the alpaca is not yet clear, although it is currently under study. It is likely to be different from country to country, according to the genetic make-up of the founders and more recently imported individuals. Our guess, however, is that the hybridization we have detected in South American camelids is likely to predate the export of the first founders to countries outside the continent. Further, even if an animal displays only vicuña alleles using these three markers, that does not mean that his/her genome does not contain guanaco/llama genes, because one only examines three markers in a genome with 45,000 genes on 37 chromosomes!

Further research is needed to map more genetic markers that distinguish among the wild species. It is possible that eventually a test will be available which might involve as many as 100 markers, able to give a percentage vicuña value for the nuclear chromosomes. Such markers may be developed as a result of ongoing genetic mapping studies in the US and UK, which are aimed primarily locating those genes which are responsible for fibre characteristics desirable to the industry. If fibre quality alleles can be used in Gene Assisted Selection programs, we may in the future be able to produce alpaca populations yielding wool of the quality of preconquest South American breeds.

*Mike Bruford is Professor of Biodiversity at the School of Biosciences, Cardiff University. He gratefully acknowledges the collaboration of Jane Wheeler, Miranda Kadwell, Helen Stanley, Raul Rosadio, Mati Fernandez, Ricardo Baldi, the Darwin Initiative for the Survival of Species, the Natural Environment Research Council and* *all those people who donated samples to the project, without whom none of this research would have been possible.*