Review of the recent studies regarding the equine mitochondrial genome

By David Howard, EMB3

Introduction

Equine (Equus caballus) mitochondria are organelles that are found in the cytoplasmic matrix of the eukaryotic cells (Dolezal et al. : 2006), see Figure 1; contained within each eukaryotic cell there are likely to be several thousand mitochondria present (IDT : 2005). The mitochondria are semi-autonomous organelles possessing their own unique genome (Bereiter-Hahn and Vöth : 2005, Huynen et al. : 2009); separate to that of the equine cell's own genome. The mitochondria, see Figure 2, are thought to have originally existed as free living alphaproteo-bacteria and adapted to a symbiotic relationship within the equine cells, approximately 1.5 billion years ago (IDT : 2005).

The mitochondria are exclusively passed down the maternal line directly from the dam to her offspring (Hartl : 2009). This is due to the structure of the male sperm, which contains the stallion's mitochondria in the mid-section, rather than in the head section. The mid-section does not permeate the oocyte during fertilisation and therefore the male mitochondria are lost and only the dam's oocyte mitochondria are passed on to the offspring (Jeon : 2005). This unique hereditary trait allows the maternal line to be investigated without influence from any paternal factors. The study of the inherited mitochondrial genome and examining the mutations that occur to it over time will provide information on the phylogenetic and phylogeographic properties of the equine species.
The mitochondria are responsible for producing 95% of the eukaryotic cells energy (Ning et al. : 2010), through the process of oxidative phosphorylation using the tricarboxylic acid cycle. The mitochondria contain an adenosine triphosphate (ATP) synthase, which is able to generate ATP from adenosine diphosphate. ATP is able to transport chemical energy thereby allowing the metabolic pathway for cellular respiration. The efficiency of mitochondrial pathways is demonstrated by the ability to generate 38 molecules of ATP from the oxidation of one molecule of glucose’ (Valberg and MacLeay : 2010). The important role of the mitochondria in the production of cell energy, allows the investigation of the mitochondrial genome’s effect on the potential performance of the horse.

The mitochondria have also been shown to regulate the ongoing process of apoptosis of the equine cells (Wang : 2001). Apoptosis is the process of programmed cell death, which occurs when a cell is damaged or becomes infected by a virus as well as maintaining homeostasis, replacing older cells as new ones are formed. Second mitochondria-derived activator of caspases (SMACs) are proteins that are released into the cytosol, binding to inhibitor of apoptosis proteins (IAPs). The SMACs deactivate the IAPs, which normally block the apoptotic process, thereby allowing apoptosis to occur (Fesik and Shi : 2001). Excessive levels of apoptosis can lead to atrophy (Dupont-Versteegden : 2006), whereas restricted apoptosis may result in cancerous cell proliferation (Evan and Vousden : 2001).

**The Mitochondrial Genome**

The mitochondria within the animal’s cells will have an identical genome, consisting of a mitochondrial deoxyribonucleic acid (mtDNA) structure. The ‘mammalian mtDNA is usually 15,000 to 18,000 base pairs’ (Sadava : 1993) in length and the human mitochondrial genome has been identified as having 16,569 base pairs (Anderson et al. : 1981). The mtDNA has a closed circular structure, Figure 3, and provides coding for 37 genes of which 13 are for protein subunits, two code for ribosomal ribonucleic acid and 22 are for transfer ribonucleic acid. The genome contains

![Figure 3. The mitochondrial genome](Staveley : 2009)
very little junk DNA (Wong and Sunshine : 1996), with over half of the coding sequences not ending with a stop codon. Studies have also shown that differing amino acids are produced for the same coding sequences, depending on the species of animal (IDT : 2005).

The rate of mutation of the mitochondrial genome is five to ten times greater than that of regular DNA (Brown et al. : 1979). This is due to a lack of protein protection and a deficient DNA repair mechanism on the mitochondrial genome. There is also an increased generation of mutagenic oxygen radicals, which contributes to the higher mutation rate of the mtDNA (Orrenius et al. : 2007). These mitochondrial genetic mutations have been associated with diseases related to areas of high aerobic demand, such as mitochondrial encephalomyopathy, myoclonic epilepsy and Leber's hereditary optic neuropathy (Merck : 2010).

The section of the mitochondrial genome that is of particular interest to studies regarding hereditary is the displacement loop (D-loop). This is an area on the genome that is approximately 1,100 base pairs in length (Ishida et al. : 1995) and contains no coding sequences. The site also contains the promoters for the transcription of ribonucleic acid (RNA) and is the landing site for the RNA polymerase (Chang and Clayton : 1985). The rate of genetic mutation in the D-loop is 2.8 to five times greater than that of the remainder of the mtDNA (Aquadro and Greenberg : 1982; Cann et al. : 1984), therefore making it of greater importance for hereditary research. The polymerase chain reaction (PCR) technique allows the amplification of the mtDNA of the D-loop to allow comparisons between samples to identify the genetic mutations and variance between subjects.

**Genetic Traits for Performance and Athletic Ability**

**Thoroughbred Racehorse Performance**

The importance of mtDNA in energy metabolism, through oxidative phosphorylation, may provide a link between mitochondrial haplotype and the performance potential of the animal. Studies involving human mtDNA have identified 17 genes effecting metabolism (Rankinen et al. : 2006) of which seven affected fitness and performance. Studies involving the horse have been centred on the performance potential of the Thoroughbred racehorse due to the recording of racing performance. There were 14,669 horses registered as being in training for 2009 (BHA : 2009), which provides a large number of potential subjects and allows meaningful comparisons of performance
between subjects. The Weatherbys’ General Stud Book (GSB) contains the pedigrees of all registered Thoroughbred horses and therefore also allows histological performance assessments, as the maternal lineage and likely mtDNA genome can be established.

Harrison and Turrion-Gomez (2006) collected 1,000 mtDNA samples, consisting of the majority of European female lines representing 33 distinct families, traceable back to the original founder mares of the Thoroughbred population. They identified 17 different haplotypes for the sample population collected, although this was across the entire mtDNA genome. The D-loop area only provided 14 different haplotypes, with the other three haplotypes identifiable by loci outside of the D-loop. The determination of performance was achieved by collating the winners of 21 major races for each year from 1954 to 2003. By establishing the haplotype of each winner and then comparing it to that haplotypes general occurrence in the horse population, provided a ‘race index’ of that haplotype’s success.

When comparing the 17 different haplotypes using race distance as a factor, five of the haplotypes had a statistical significance with regards to performance. Two haplotypes demonstrated a strong correlation with greater performance over longer distances; whilst the remaining three were strongly correlated with improved performance over shorter distances, see Figure 4.

![Figure 4. Regression lines describing the relationship of Race Index vs. race distance for five haplotypes. (Harrison and Turrion-Gomez :2006)](attachment)

The results revealed by this study confirm the belief that the maternal line carries a greater influence over performance over a particular distance; with over fifty percent of
the three year old Thoroughbred population identified as belonging to one of the five mitochondrial haplotypes that have a strong correlation to performance over an extreme of distance. The study however allowed the repetition of a subject where it achieved multiple race wins, thus creating a bias in that haplotypes ‘race index’. The Harrison and Turrion-Gomez (2006) study also found that 19 of the 33 distinct family lines studied had anomalous mtDNA inheritance, which was studied in greater depth by Hill et al. (2002).

**Identifying the Genes for Performance**

The Thoroughbred racehorse has been systemically bred to maximise the performance potential of any progeny. This has created a breed that is exceptionally adapted to high levels of performance and therefore carries a genetic phenotype that is a suited to producing an athletically fit animal. Gu et al. (2009) conducted a genome scan of the Thoroughbred to identify the genes responsible for athletic ability, which will in turn reveal the genes within the complex molecular networks underlying obesity and its consequential pathologies, for the human population. The study mapped the entire Thoroughbred genome, including mtDNA, and identified areas responsible for fatty acid oxidation, increased insulin sensitivity and muscle strength.

The isolation of the mitochondrial fission regulator 1 gene was achieved, which encodes for an inner mitochondrial membrane protein that promotes mitochondrial fission. This gene ‘may impose strong selection pressure in the Thoroughbred by protection of mitochondria-rich tissue against oxidative stress’ (Gu et al. : 2009). This work was aimed at improving the understanding of human metabolic pathologies related to obesity, such as type 2 diabetes. Further studies could investigate the relationship between the mitochondrial fission regulator 1 gene and racehorse performance.

**The Effect of the Environmental on the Mitochondria**

A study by Ning et al. (2010) investigated the potential for the mitochondrial genome to adapt to environmental factors over time. This involved the sequencing of 509 Chinese plateau horses for the NADH dehydrogenase 6 (ND6) gene, which is involved in oxidative phosphorylation and energy metabolism. The horses were segmented into three groups depending on the altitude in which they lived (>2,200m, 1,200 – 1,700m, <900m). Mutations on the ND6 gene within the different altitude groups were then investigated. The results showed that the horses living at a high altitude showed a statistically significant drop in genetic diversity on the ND6 gene compared to the other
groups. This is likely to have occurred as ‘advantageous mutations will increase in frequency to adapt; after they have spread through the population, the level of diversity will decrease’ (Ning et al.: 2010). Therefore the mtDNA of the horses living at the higher altitude underwent an adaptation to cope with the environmental conditions.

From the studies published thus far, there is definitive evidence of the contribution of mtDNA on the potential for performance of the horse. The continued mapping of the mitochondrial genome and the identification of haplotypes, will allow further development and understanding with regards to the level of potential performance. It is unlikely that there will ever be a failsafe method for performance selection solely on the basis of genotype, as there are many uncontrollable hereditary and external environmental factors to be taken into consideration. The potential for the identification of genes responsible for performance related pathologies is an area that is of significant importance. This is due to the increasingly obese human population with ‘thirty-nine per cent of adults had a raised waist circumference in 2008 compared to 23% in 1993’ (NHS: 2010).

**Phylogenetic Studies**

The hereditary nature of the mitochondrial genome, which is passed exclusively down the maternal line, allows the origin of a particular species or breed to be investigated. The predictable nature of the rate of mutation of the D-loop mtDNA is able to provide evidence of the degree of separation between subjects and the likely date for inter- and intra-evolutionary division, thereby revealing the origin of the breeds and the genus.

**The Extant Species of the genus Equus**

The variance between the mtDNA samples taken from six extant species of the genus Equus is presented in a literature review by Forstén (1992). The degree of variation between the mitochondrial genome of the various Equus species allows the relationship between them to be revealed, as well as an approximate date for morphology. Forstén (1992) compared the mtDNA data collected by George and Ryder (1986) to the fossil and dental morphology records, which existed prior to the development of mitochondrial analysis.

The extant species covered were categorised as the Grevyi zebra (E. grevyi), the Mountain zebra (E. zebra), the Plains zebra (E. burchelli), the African ass (E. asinus), the
Asian wild ass (E. hemionus) and the true horse (E. caballus). The results of the mtDNA analysis and the existing paleontological records are presented in Figure 5.

The fossil records and the mtDNA data were comparatively close in agreement in terms of the evolutionary structure of the Equus species, however some of the dates for morphology differed. Most notably was the date for the branching of the Equus Caballus, which according to mtDNA occurred 3.9 Ma ago, however no callalloid fossils have been dated older than 1.8 Ma ago. This may have been caused by a lack of fossil teeth discovered between 1.8-3.9 Ma ago, dental morphology occurring at a different time to the overall morphology of the species or the presence of several mtDNA lineages in the mid-Pleistocene period. Substitutions of mtDNA mutations have been shown to occur at varying rates in related and differing lineages (Galtier et al. : 2010, Nabholz et al. : 2009), which may have caused misleading timelines for the Equus Caballus, in the George and Ryder (1986) study.

The Thoroughbred Breed

The accurate recording of matings and the closed breeding rules regarding the thoroughbred breed, will in theory provide a clear family tree for the racehorse, which can be dated back to the original founding mares for the breed. All matings are recorded in the GSB, which is maintained by Weatherbys and dates back to 1791 and lists the 78
founder mares. A study by Cunningham et al. (2001) using microsatellite loci from 211 Thoroughbreds, found that ten of the founder mares are currently responsible for 72% of the maternal lineages. However, the Cunningham et al. (2001) study made use of microsatellite loci, which can be passed on from the sire or the dam and so is susceptible to stochastic effects (Mullins: 2004).

The study conducted by Hill et al. (2002) also concentrated on the genetic diversity within the Thoroughbred bred with regards to the founder mares, whilst using the 381 base pairs from the mitochondrial D-loop as the site for analysis. The mtDNA is passed only down the maternal line and therefore should provide an accurate picture of the original founder mares. According to Bruce Lowe's (1895) Family Figure System, the original 78 founder mares belonged to 43 distinct families. Hill et al. (2002) analysed 100 Thoroughbreds that belong to 19 of the most common families and found 17 different haplotypes, which led them to conclude that as few as twelve founder mares may have contributed to the 19 common families studied.

The data collected was unable to provide the original genetic origins of the Thoroughbred. However, it did reveal that nearly half of the subjects had genetic anomalies compared to that expected with regards to the pedigree stud book. This is likely to be caused by inaccurate matings occurring on the stud farm or clerical errors in the recordings of pedigrees. A replication of the study should be conducted with an increased number of subjects, ensuring that all 43 distinct family lines are included to provide further evidence of the original founder mares.

The Lipizzaner Breed

The Lipizzaner breed was established in the Austrian Empire in the second half of the sixteenth century (LANA: 2010), by the Habsberg family. The Lipizzaner breed is subject to strict breeding rules similar to those of the racing Thoroughbred, with only maternal lines from the Lipica Stud prior to the Second World War regarded as classical maternal lines. The recording of matings has provided much of the historical and hereditary knowledge with regards to the Lipizzaner breed.

Kaver et al. (2002) conducted a study sequencing the Lipizzaner mitochondrial genome, with the intention of providing empirical evidence of the genetic basis and diversity of the breed. MtDNA was collected from 212 Lipizzaners and revealed 37 distinct haplotypes within the breed, which was regarded as a high level of genetic diversity. This is not unexpected as the origin of the breed is thought to have involved from Karst,
Spanish, Italian, Kladruber and Arabian breeds (Nürnberg : 1993). The study revealed that there were discrepancies between the expected genealogies and the actual mtDNA encountered for at least 11% of the subjects. An assumption was also made that due to the mtDNA mutation rate of between 2-4 x 10^{-8} per site per year, it is likely some of the existing Lipizzaner haplotypes are identical to the original founder mares of the breed.

The Zemaitukai Breed

The Zemaitukai is a rare breed of Equus caballus living in Lithuania, and by 2003 had declined to just 147 animals, consisting of two paternal lines, and five maternal family lines (Macijauskiene : 2002). A study by Cothran et al. (2005) addressed the potential for the ongoing conservation and management of the breed with the aim of maximising the level of genetic diversity. Mitochondrial D-loop samples were taken from five animals covering the five surviving maternal lineages, which revealed the existence of five different mtDNA haplotypes.

The five Zemaitukai samples were compared with two other Lithuanian breeds, the Lithuanian Heavy Draught and the Zemaitukai Heavy Type, as well samples from ten European breeds. In comparison to the additional samples the Zemaitukai was shown to have a high nucleotide and sequence diversity within the breed. The origin of the Zemaitukai breed was also unclear with the five haplotypes failing to demonstrate any clear pattern of relationship with the ten European breeds. The sample size was very small (n=5) and did not provide any new information about the ancestry of the Zemaitukai. However, they concluded the anticipated pedigrees were correct and that efforts should be made to ensure the five maternal families continue to be used for breeding, to maintain the genetic diversity of the breed.

The Argentinian Creole and the Peruvian Paso Breeds

Mirol et al. (2002) conducted a study on two South American breeds, the Argentinian Creole and the Peruvian Paso, to identify and compare their genetic variance and diversity to both the Arab and Thoroughbred breeds. Amplifying the mitochondrial D-loop of the subjects and analysing them by single stranded conformational polymorphism (SSCP) provided the diversity between the breeds. A total of 100 horses were used in the study, which revealed 14 different SSCP variants, with the greatest variability found within the Arab breed. By analysing the inter-breed location of the variants it was possible to produce a ‘phenogram based on a discrete two state
character data matrix using the Wagner parsimony method’ (Mirol et al. : 2002), see Figure 6.

Figure 6. Tree obtained using the Wagner parsimony method. ARB, Arabian; TB, Thoroughbred; PPA, Peruvian Paso; ARC, Argentinian Creole (Mirol et al. : 2002)

The results of the study reveal that the Peruvian Paso has a closer relationship with the Thoroughbred than the other breeds. These are both in turn more closely linked with the Argentinian Creole than the Arabian breed, which solely occupied a basal clade position. The Arabian and Thoroughbred are considered to the foundation stock for many horse breeds (Bowling : 1994) and Mirol et al.’s (2002) study reveals that, due to the maternal hereditary traits associated with the mitochondria, the two South American breeds studied may have come from Thoroughbred mares and Arabian stallions. Of the 100 subjects analysed only seven were Thoroughbreds, a more even distribution of breeds in the sample population should have been sought to ensure equal variance.

Phylogeographic Studies

Due to the association with particular breeds to a geographical territory, it is possible to produce a phylogeographical representation, using the mitochondrial genome, of the evolution and migrationary patterns of the Equine caballus over time. The mapping of the migration of the horse, since its domestication approximately 6,000 years ago (Goodwin : 2009), will aid in providing a clearer understanding of Neolithic anthropology and the trade routes that existed. Combining the migrationary patterns with data regarding breed and time scale will assist with establishing the evolution morphology and genetic clusters of the current domesticated breeds.
The Iberian Peninsula

An area of specific interest for equine migration is the Iberian Peninsula (Figure 7), which incorporates the countries, Spain, Portugal and Gibraltar. Due to the location of the Peninsula between Africa and Europe, it has been peripheral to major human and livestock migration and expansions. The breeds located within this area are likely to be the predecessors of the Garrano from Portugal, the Losino, Andalusian and Carthusian Spanish breeds (Lopes et al. : 2006), the Barb from North Africa (Jansen et al. : 2002) and the Celtic ponies of the British Isles (Luís et al. : 2006).

Royo et al. (2005) collected mtDNA data from 138 Iberian horses, 14 Barbs and 20 Exmoor ponies for analysis. Their study revealed a ‘close genetic relationship, at a mtDNA level, between Northern African and Iberian horse populations’ (Royo et al. : 2005), but not between the Iberian and the Exmoor ponies. This suggests that trade routes were certainly established between North Africa and the Iberian Peninsula, but not from the Peninsula into Northern Europe. However, Royo et al. (2005) suggests that ‘Iberian horses may be the result of a male-mediated introgression of Northern European ponies’, which is consistent with ‘sex-biased ancient breeding practices’ (Vilà et al. : 2001). A similar study conducted on the Y-chromosome may reveal whether this was the case.

The Sites for Domestication

The issue regarding the location or locations for domestication of the horse has yet to be fully resolved, and was addressed in a study by Jansen et al. (2002). They analysed a total of 652 mtDNA samples and found 93 different haplotypes, which could be grouped into 17 clusters, each specific to a set of breeds and/or geographical areas. By calculating the estimated maximum level of mtDNA mutation (1/100,000 yrs) and the earliest time of predicted domestication (9400BC) it is possible to remove 12 of the haplotypes as post-domestication mutations. Due to 43 minor haplotypes being only one mutation distant from their immediate ancestor, it was deduced that a maximum of 4.4 of those haplotypes also postdate domestication, leaving a minimum of 77 original...
founder mares for domestication. Jansen et al. (2002) concluded that this indicates that there were several distinct horse populations involved in the domestication of the horse. However, it was still unclear as to whether domestication occurred at a single site and domesticated horses were supplied along trade routes or the knowledge for domestication spread along trade routes and domestication occurred at separate locations.

McGahern et al. (2006) conducted an investigation of the geographic affiliation of mtDNA haplotypes for the domesticated horse. Prior to this study no statistically significant data had been produced linking a particular haplotype clade to a specific region, due to the complex structure of breeds within each haplotype. Although the Jansen et al. (2002) study did reveal links between clade and location, no statistical testing was conducted, as that was not the purpose of that particular study.

During the McGahern et al. (2006) study they collected 118 mtDNA sequences from seven previously unsampled breeds (Orlov, Mesenskaya, Vyatskaya, Akhal Teke, Yakut, Mongolian and Guan Mountain) from central and north-eastern Eurasia and China combined with previously collected data from another 844 horses. The total data set was categorised into seven haplotypes, A-G, as described by Jansen et al. (2002) and their geographical locations split into three regions, European (EUR), the Middle East & Africa (MEA) and the Far East (FE), Figure 8.

Figure 8. Mitochondrial DNA haplogroup distributions in Eurasian horse populations (EUR, n = 619; MEA/FE, n = 343). Geographical locations for newly sampled populations are denoted by the position of the individual pie charts. The size of each circle is proportional to the number of sequences sampled from that region. Haplogroups are coloured as follows: dark blue = A, light blue = B, red = C, green = D, brown = E, orange = F and pink = G. (McGahern et al. : 2006)
The newly sampled breeds revealed two new sub-haplotypes on the A and F clusters. This new data combined with the existing samples provided a geographical division for the F cluster and the frequency was distributed with EUR = 9.7%, MEA = 16.5% and FE = 16.8%. When comparing the frequency of haplotype F using Fisher's exact test of independence in the MEA and FE regions with the EUR region, reveals $p < 0.000005$ and therefore is highly statistically significant. This reveals the likelihood of separate horse populations used for domestication in Europe compared to the Middle Eastern, African and Far East regions and therefore domestication appears to have been achieved at more than one location.

The ongoing mtDNA sequencing of unsampled breeds and relating the data collected to geographical locations, combined with archaeological and anthropological studies will continue to provide greater clarity regarding the domestication of the horse and the associated timescales and patterns of migration. Improvements in mitochondrial genome sequencing techniques will potentially reveal greater clarity regarding the sites and rates for mutation of the mtDNA, thus providing a clearer indication of the lineage and evolutionary pattern of the domestic horse.

**Conclusion**

The functional aspects of the mitochondria for the control of energy metabolism warrant the mapping of its genome for determining the genes for athletic performance potential. The Thoroughbred racehorse provides the most viable subject for equine studies regarding performance, due to the regular intense competitive environment in which it competes. Although it is unlikely that a 'perfect' gene sequence will be established, due to the many external factors involved in racing performance, racehorse owners will continue to strive to maximise their chance of success by gaining as much information about their animals as possible.

The potential for the performance genes of Thoroughbred's to aid in identifying pathological deficiencies in humans is an area of immense importance. The ongoing mapping of the mitochondrial genome of unsampled species and subjects will increase our understanding of the potential defects that can be caused by mutations of the mtDNA, such as atrophy, cancer, mitochondrial encephalomyopathy and myoclonic epilepsy.

The semi-autonomous and maternal hereditary nature of the mitochondrial genome presents an alternative method of investigating the origins of the equine species and the
history of its domestication. Understanding the genetic relationships between breeds has formed the basis for many recent studies, some of which have been covered in this report. An understanding of the genetic morphology of the genus Equus will help to clarify the historical role of the domesticated horse, as well as the extant species. The data collected thus far has been of interest and further studies centred on the male Y-chromosome may corroborate the evidence regarding mtDNA hereditary patterns. The migrationary patterns of the various equine species and the sites of domestication will also enhance our understanding of the human activity and trade routes prior to recorded history or where accounts are incomplete.
References


IDT (2005) *Mitochondrial dna (mtDNA)*. Integrated DNA Technologies


Lowe, B. (1895) Breeding horses by the figure system. Horace Cox: London


