The Horse Genome Project and the Caspian Horse

Douglas F. Antczak VMD, PhD
Baker Institute for Animal Health
College of Veterinary Medicine
Cornell University, Ithaca New York 14853 USA

Abstract

Over the past 20 years the Horse Genome Project has produced tremendous advances in our understanding of the genetic makeup of the domestic horse. Important investigations of genes influencing health and performance have been complemented by studies of the evolutionary biology of the horse family. On the 50th anniversary of the re-discovery of the Caspian Horse, it is now possible to apply modern genetic technology for the benefit of this rare, beautiful, and useful breed. Such research has the potential to discover the genetic heritage of the Caspian Horse and its contribution to other breeds in the Middle East and beyond. With the participation of Caspian Horse breeders from Iran and around the world it will be possible to reach these goals.
پروژه زنوم اسب واسب کاسپیان
پروفسور داگلاس آنتزاک
انستیتوی سلامت حیوانات بیکر
کالج دامپزشکی دانشگاه کرنل، ایتاکا، نیویورک، آمریکا

در بیستم سال کشف شده پروژه زنوم (Genome) اسب، پیشرفت هایی به سرایی در آگاهی ما از ساختار زنتیکی اسب داشته است. تحقیقات مهم در مورد زنیایی که بر سلامت و توانایی اسب ها تاثیر گذار هستند، با مطالعات بیولوژی تکاملی خانواده اسب تکمیل شده اند. در پنجاهمین سالگرد کشف مجدد اسب کاسپیان (اسبچه خزر)، این امکان وجود دارد که از تکنولوژی مدرن زنیایی برای بهبود وضعیت این نژاد نادر، زیبا ورزشمند و مفید بهره مند شویم. چنین تحقیقاتی، این امکان را فراهم می‌سازد که از میراث زنتیکی این اسب اطلاع حاصل کنیم و از دخیل بودن آن در تولید ترکه‌های دیگر اسب خارجی بهره مند شویم. آگاهی از آن، از بین بردن بدهی که با شرکت پرورش دهنده‌گان اسب کاسپیان (اسبچه خزر)، در ایران و دیگر کشورها تحقیق این اهداف ممکن می‌شود.
Introduction

Horsemen have selected horses for strength, size, speed, disposition, gait, color and conformation since their domestication approximately 5,500 years ago. At the same time, however, undesirable genes adversely affecting health, performance, or fertility have been inadvertently selected and maintained (Schubert et al., 2014). Thus, knowledge of the equine genome presents two obvious areas for application: identification of disease causing genes, and determination of genes producing the characteristics that are valued by horsemen. A third opportunity is the study of evolution of the closely related members of the horse family in the genus Equus, that is, horses, asses, and zebra (Orlando et al., 2009).

The genome is the collection of all of the genes and related genetic material of an organism. The genome of each species contains the information for building the organism and maintaining its functions. The genome is thus similar to a blueprint for a building or a complex computer software code. The information in the genome is encoded in four simple chemical molecules, called bases, which make up DNA, or Deoxyribo Nucleic Acid. Like letters of an alphabet, those four bases, abbreviated A, T, C, and G, for Adenine, Thymine, Cytosine, and Guanine, can be arranged in an infinite variety of ways to produce the tremendous variation in plants, animals, and microorganisms found on Earth. The genomes of mammals, including humans and horses, each contain approximately 3 billion base pairs of DNA.

Over the past 25 years the technology for determining the DNA sequence of a genome have advanced extremely rapidly. What was once a very challenging task involving years of work by hundreds of scientists and millions of dollars of expense can now be accomplished in a few days by a single individual using a relatively inexpensive DNA sequencing instrument, at the cost of only a few thousand dollars in laboratory supplies. These technological advances have led to rapid progress in determining the complete DNA genome sequences of humans, most domesticated animals, and many free-living species. Large computer databases make the information from these studies available freely to scientists around the world (Table 1).

The State of Knowledge of the Horse Genome
Twenty years ago, in 1995, equine veterinary scientists and geneticists from around the world formed a working group for collaboration on the task of creating genetic maps of the horse. Investigators from over 25 laboratories in 15 countries have met on a regular basis as part of the Horse Genome Project. The group was very successful in developing linkage, physical, and comparative gene maps of the horse (Chowdhary et al., 2003; Swinburne et al., 2000).

In 2006 the US National Human Genome Research Institute partnered with the workshop members of the Horse Genome Project to determine the complete DNA sequence of a single horse, the Thoroughbred mare Twilight (Wade et al., 2009). [Figure 1] Based on the Twilight sequence, the equine genome contains an estimated 2.7 billion base pairs of DNA. This genetic information is distributed in 64 chromosomes (32 pairs) containing an estimated 22,000 genes. The Twilight DNA sequence revealed a remarkable conservation of gene order on the chromosomes of the horse and the human. This has facilitated the elucidation of the equine DNA sequence, by comparison with the better-studied human genome sequence.

Although the Twilight equine genome sequence is very high quality, it is not complete. Some regions of the genome are very difficult to sequence, and the assembly of the DNA in correct order from the small fragments used for sequencing is also challenging. Therefore efforts are continuing to refine the Twilight DNA sequence (Rebolledo-Mendez et al., 2015). In addition, the tremendous decreases in the cost of whole genome DNA sequencing means that it is now routine to obtain the complete genome sequences from horses under study for a variety of scientific purposes.

The Twilight sequence has enabled equine geneticists to develop two types of assays for practical application to genetic studies of the horse. First, partial DNA sequences from horses of a variety of breeds were obtained and used to construct an equine Single Nucleotide Polymorphism (SNP) chip (McCue et al., 2012). SNP chip assays can assess variation between groups of horses at thousands of locations all across the genome. They have been employed to determine the location of mutated genes causing inherited diseases of the horse (Brooks et al., 2010) and genetic regions influencing susceptibility to virus infections (Go et al., 2011).
The second type of assay determines which genes are expressed in individual horse cells or tissues. These so-called expression assays are used to determine patterns of gene expression between horses or between different cell types (Brosnahan et al., 2012; Wang et al., 2013). Expression assays have utility in studies of normal horse physiology and disease states and are likely to be widely used in the future.

The advances in equine genetics and genomics have been so dramatic that two complete books on this subject have been published in the past two years (Bailey et al., 2013; Chowdhary, 2013).

Genes affecting equine health

Many benefits to the equine industry have already accrued from genetic studies of the horse. The genes responsible for over 15 simple inherited equine diseases have been discovered (Table 2) and for many of these diseases diagnostic tests for the defective genes have been developed and commercialized. Such tests enable horse owners to determine if their breeding stock are carriers of autosomal recessive or sex-linked mutant genes. This information can be used when making breeding decisions, for example, to avoid carrier-to-carrier matings. Also prospective buyers of horses can request that genetic tests be performed as a condition for purchase. New and previously recognized equine diseases continue to be investigated by geneticists and veterinary clinicians, and it is likely that additional genes causing diseases with simple Mendelian inheritance patterns will be identified in the coming years.

Complex genetic diseases are caused by the action of two or more genes acting together to produce the disease condition, or by gene-environment interactions, such as exposure to a chemical or infectious agent. Genes that influence such complex traits are usually responsible for only a portion of the disease susceptibility. In the horse several conditions of this type have been described, including allergic respiratory disease (Gerber et al., 2015), insect bite hypersensitivity (Andersson et al., 2012b), and equine sarcoid tumors (Lazary et al., 1985). Several common equine muscle diseases have been described that have a simple genetic
basis. However, for some of these conditions the manifestations can be influenced by a common environmental factor – diet (Mickelson and Valberg, 2015).

Chromosome disorders appear to be relatively rare in horses compared to humans, but new technologies derived from the Horse Genome Project have facilitated the identification of such conditions. The most frequently described disorders are those that result in altered sexual development (Lear and McGee, 2012). These conditions are usually compatible with life, and are only discovered when the animals are used for breeding and found to be infertile. Disorders of autosomal genes, such as trisomies, have also been described (Holl et al., 2013), but less frequently. Recent review papers have summarized the advances in equine medical genetics that have been outlined here (Brosnahan et al., 2010; Finno and Bannasch, 2014; Finno et al., 2009).

Genes influencing equine performance

Compared to the list of known mutations that cause equine inherited diseases, the number of genes that have been shown to influence equine performance is very small. Two examples stand out. First, Emmelline Hill and colleagues in Ireland discovered that the Myostatin (MSTN) gene has SNP variants that are associated with optimal racing distance. One SNP is found commonly in horses that excel at sprinting, while the other form is found more often in horses that race well at longer distances (Hill et al., 2010). As might be expected, American Quarter horses that race one-quarter mile are usually homozygous for the sprint-associated SNP, while Arabian horses used for endurance races are usually homozygous for the distance SNP.

The second example is related to gait. Leif Andersson and colleagues in Sweden identified a nonsense mutation in the DMRT3 gene that favors the ability of a horse to pace or tol. Not surprisingly, this mutation is also associated with performance of harness horses (Andersson et al., 2012a). The DMRT3 gene affects neurological circuits related to limb movement and motion. It is likely that a single mutation event was responsible for this trait, and that horses of all breeds worldwide that show the ability to perform unusual gaits carry the same mutation.
Ancient DNA and the evolution of the horse family

Perhaps the most exciting recent development in equine genetics has been the application of genome sequencing technology to studies of ancient DNA, specifically DNA isolated from horses that lived in the pre-historic era. This research has included the publication of a DNA sequence of obtained from a small horse bone from the Middle Pleistocene era, estimated to be 735,000 years old (Orlando et al., 2013). To date this is the oldest specimen of a large organism that has been successfully sequenced. This study of ancient horse DNA is a wonderful example of how modern molecular biology techniques based on gene sequencing is merging with conventional approaches of archeology and paleontology based on physical examination of bones and other fossils to develop new understanding about evolution. This study pushed back the estimated date of the most recent common ancestor of the genus Equus to 4.0 – 4.5 million years before present. The research also provided evidence that the Przewalski’s horse is a distinct species that has been separated from the domestic horse for about 50,000 years.

In another study Pruvost and colleagues have provided convincing evidence that pre-historic horses from approximately 25,000 years before present carried the LP (leopard) mutation that underlies the well-known Appaloosa coat color patterns (Pruvost et al., 2011). This is consistent with the recognizable coat patterns depicted in Paleolithic cave paintings in Pech-Merle, France.

Full sequencing and analysis of the complete genomes of all of the living members of the genus Equus, plus the recently extinct Quagga, has complemented the results obtained with ancient DNA (Jonsson et al., 2014). This work provided new details about the evolutionary relationships among the horse family members, and also revealed evidence for gene flow between species of the genus Equus that is similar to that described for early humans, Neanderthals, and Denisovan peoples (Paabo, 2015).

Studies of the equine Y chromosome and mitochondrial DNA, which is inherited only from the mother, have provided an interesting hypothesis about horse domestication. Very little variation between horses has been detected in horse Y chromosomes, suggesting that male horses (stallions) were domesticated only rarely (Lindgren et al., 2004). In contrast, the large number of
mitochondrial DNA types found in domestic horses indicated multiple domestication events which may have been limited largely to mares (Cieslak et al., 2010).

Genetic studies of the Caspian Horse

The earliest published record of the Caspian Horse is from 1969 by Louise Firouz, who described the re-discovery of the breed and its unique characteristics (Firouz, 1969). Because of her passionate efforts to promote and preserve this breed, many subsequent research studies have followed, and several of these have involved genetics. An early report claimed that some Caspian horses carried 65 chromosomes, rather than the 64 of domestic horses, and was interpreted that Caspian Horses had interbred with the Przewalski’s Horse that has 66 chromosomes (Hatamimonazah and Pandit, 1979). This report has not been confirmed by any of the more recent studies, and so should be left as an open question. A study of red blood cell antigens from the mid-1990s identified a new genetic variant of the D blood group in a family of Caspian Horses (Cothran and Long, 1994).

Two more recent studies using microsatellite markers demonstrated considerable genetic diversity within Caspian horses in Iran (Amirinia et al., 2007; Shasavaran and Rahimi-Mianji, 2010). An investigation of Y chromosome diversity in Iranian indigenous breeds found no evidence of variation (Rafeie et al., 2011), consistent with earlier findings in horses from other regions (Lindgren et al., 2004). A large study of 24 European breeds typed using a panel of 12 microsatellite loci identified the Caspian region as a site of very high genetic diversity in the Oriental horse populations (i.e. Arabian, Akhal Teke, and Caspian Horse) that lived in this region during the Holocene period. The Caspian Horse samples tested specifically showed high levels of heterozygosity and low levels of inbreeding (Warmuth et al., 2011).

Several studies have shed light on the genetic relationship between the Caspian Horse and other Oriental breeds. First, a study of variants of maternally inherited mitochondrial DNA in a large sample of over 1900 horses from several breeds showed that Oriental horses, including the Barb, Arabian, Akhal Teke, and the Caspian Horse, clustered together, suggesting a similar ancestry (Bower et al., 2011). Second, the equine SNP chip was applied in a multi-breed study
of over 800 horses from 36 breeds, including 18 Caspian Horses (Petersen et al., 2013). This research used several bioinformatic analysis methods to demonstrate the genetic relationships between breeds. Using the distance based neighbor joining tree, the Arabian, Akhal Teke, and Caspian Horses were grouped together, with no other breeds close to them. However, when the parsimony method was applied, the Caspian Horses were separated from the Arabians and Akhal Tekes. Although more research is needed to obtain a more definite answer to this question, the evidence for a close relationship between the Caspian Horse and other Oriental horse breeds seems quite strong. Consistent with these genome-wide studies is a case report of Severe Combined Immunodeficiency Disease (SCID) in a Caspian Horse in the United States (Larson et al., 2011). Because the SCID mutation has been found only in Arabian or part-Arabian horses, it may be that this mutation is an old one that pre-dates the separation of the Arabian and Caspian Horse breeds.

The Case for the Caspian Horse

The information currently available from genetic studies of the Caspian Horse supports the likelihood of an ancient origin for this breed and a close relationship to other Oriental breeds. However, the number of genetic studies has been limited, and only a small sample of Caspian Horses has been tested. It is generally agreed that the global population of Caspian Horses is very small, and therefore likely to suffer from inbreeding. Fortunately, most of the population studies have estimated a high level of genetic diversity among the Caspian Horses.

The continuing decreases in the cost of genome sequencing makes it feasible to consider the prospects for a Caspian Horse Genome Project. Such an undertaking could include full genome sequencing of selected Caspions, plus genome-wide scanning of the entire Caspian population using the second-generation SNP chip assay, and a comprehensive determination of mitochondrial DNA types of the Caspian maternal lines. The information from such a genetic study would provide important information on the evolutionary history of the Caspian breed, and medical genetic data that could be used by Caspian Horse breeders and owners. Finally, the genetic information could be used to generate universal pedigrees for the Caspian Horse that
would facilitate breeding decisions and efforts to preserve and expand this magnificent breed of horse.

Acknowledgements
This manuscript was produced with support from the Dorothy Russell Havemeyer Foundation, Inc.
## Tables and Figures

### Table 1. Selected Internet sites for genomic sequence databases of the horse

<table>
<thead>
<tr>
<th>Site</th>
<th>WWW Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horse Genome Project (Univ. Kentucky)</td>
<td><a href="http://www.uky.edu/Ag/Horsemap/">http://www.uky.edu/Ag/Horsemap/</a></td>
</tr>
<tr>
<td>NIH NCBI</td>
<td><a href="http://www.ncbi.nlm.nih.gov/genome?term=txid9796%5Borgn%5D">http://www.ncbi.nlm.nih.gov/genome?term=txid9796%5Borgn%5D</a></td>
</tr>
<tr>
<td>UCSC Genome Browser</td>
<td><a href="http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=438659539_WHpBrJ7Z9n3B7amrZ98Ajczmkvxn&amp;clade=mammal&amp;org=Horse&amp;db=0">http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=438659539_WHpBrJ7Z9n3B7amrZ98Ajczmkvxn&amp;clade=mammal&amp;org=Horse&amp;db=0</a></td>
</tr>
<tr>
<td>Ensembl</td>
<td><a href="http://www.ensembl.org/Equus_caballus/Info/Index">http://www.ensembl.org/Equus_caballus/Info/Index</a></td>
</tr>
<tr>
<td>Broad Institute</td>
<td><a href="http://www.broad.mit.edu/mammals/horse">http://www.broad.mit.edu/mammals/horse</a></td>
</tr>
</tbody>
</table>
### Table 2. Selected inherited diseases of the horse

<table>
<thead>
<tr>
<th>Condition</th>
<th>Breeds</th>
<th>Gene / Chromosome</th>
<th>Inheritance</th>
<th>Key Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital Stationary Night Blindness (CSNB)</td>
<td>Appaloosa</td>
<td>TRPM1 ECA1</td>
<td>Incomplete Dominance</td>
<td>(Bellone et al., 2013)</td>
</tr>
<tr>
<td>Foal Immunodeficiency Syndrome (FIS)</td>
<td>Fell and Dale Ponies</td>
<td>SLC5A3 ECA26</td>
<td>Autosomal Recessive</td>
<td>(Fox-Clipsham et al., 2011)</td>
</tr>
<tr>
<td>Glycogen Branching Enzyme Deficiency (GBED)</td>
<td>Quarter Horse</td>
<td>GBE1 ECA26</td>
<td>Autosomal Recessive</td>
<td>(Ward et al., 2004)</td>
</tr>
<tr>
<td>Gray Horse Melanoma</td>
<td>Many</td>
<td>STX17 ECA25</td>
<td>Autosomal dominant</td>
<td>(Rosengren Pielberg et al., 2008)</td>
</tr>
<tr>
<td>Hereditary Equine Regional Dermal Asthenia (HERDA)</td>
<td>Quarter Horse</td>
<td>PPIB ECA1</td>
<td>Autosomal Recessive</td>
<td>(Tryon et al., 2007)</td>
</tr>
<tr>
<td>Hyperkalemic Periodic Paralysis (HYPP)</td>
<td>Quarter Horse Paint</td>
<td>SCN4A ECA11</td>
<td>Autosomal dominant</td>
<td>(Rudolph et al., 1992)</td>
</tr>
<tr>
<td>Junctional Epidermolysis Bullosa (JEB)</td>
<td>American Saddlebred</td>
<td>LAMA3 ECA8</td>
<td>Autosomal Recessive</td>
<td>(Graves et al., 2009)</td>
</tr>
<tr>
<td>Lavender Foal Syndrome (LFS)</td>
<td>Arabian</td>
<td>MYO5A ECA1</td>
<td>Autosomal Recessive</td>
<td>(Brooks et al., 2010)</td>
</tr>
<tr>
<td>Malignant Hyperthermia</td>
<td>Quarter Horse</td>
<td>RYR1 ECA10</td>
<td>Autosomal dominant</td>
<td>(Aleman et al., 2009)</td>
</tr>
<tr>
<td>Overo Lethal White Syndrome (OLWS)</td>
<td>Paint</td>
<td>EDNRB ECA17</td>
<td>Autosomal Recessive</td>
<td>(Metallinos et al., 1998)</td>
</tr>
<tr>
<td>Polysaccharide Storage Myopathy (PSSM)</td>
<td>Quarter Horse Draft Horses</td>
<td>GSY1 ECA10</td>
<td>Autosomal dominant</td>
<td>(McCue et al., 2008)</td>
</tr>
<tr>
<td>Severe Combined Immunodeficiency (SCID)</td>
<td>Arabian</td>
<td>DNA-PK&lt;sub&gt;cs&lt;/sub&gt; ECA9</td>
<td>Autosomal Recessive</td>
<td>(Shin et al., 1997)</td>
</tr>
<tr>
<td>Hoof wall separation disease (HWSD)</td>
<td>Connemara pony</td>
<td>SERPINB11 ECA8</td>
<td>Autosomal Recessive</td>
<td>(Finno et al., 2015)</td>
</tr>
</tbody>
</table>
Figure 1. Twilight, the DNA donor of the Horse Genome Project equine sequence. Twilight is stabled at the Equine Genetics Center of the Baker Institute for Animal Health at Cornell University. The mare was bred in a small herd of horses selected for homozygosity at the Major Histocompatibility Complex.
References:


