Journal of Equine Veterinary Science 35 (2015) 361–366

Contents lists available at ScienceDirect

Journal of Equine Veterinary Science

journal homepage: www.j-evs.com

Review Article

Genetics After Twilight

Ernest Bailey*

MH Gluck Equine Research Center, University of Kentucky, Lexington, Kentucky, USA

ARTICLE INFO

Article history:
Received 9 February 2015
Accepted 19 March 2015
Available online 8 April 2015

Keywords:
Genomics
Chromosome
Gene mapping
Functional genomics
Gene expression
Ancient DNA

ABSTRACT

In 2006, the whole genome sequence was completed for the Thoroughbred mare, Twilight. This sequence and the tools derived from this work led to the discovery of genetic variants responsible for coat color traits, hereditary diseases, and even some performance traits. Technological advances have made whole genome sequence easier and less expensive. Whole genome sequencing derived from ancient horse DNA has provided insights into the events associated with domestication of the horse and also suggests that one of the costs of domestication has been an increase in the genetic load, specifically, autosomal recessive deleterious genes. The next step for genomic research is the development of approaches for readily investigating functional genomics. Functional genomics is a tool relevant to all fields of equine study, including immunology, nutrition, reproduction, exercise physiology, and veterinary medicine.

1. Introduction

Horse breeders are renowned for the practice of genetics. From the earliest times of domestication, farmers recognized that offspring resembled parents and sought superior parents as breeding stock. The practice is ancient. For thousands of years, Arabian-horse breeders characterized breeding stock in terms of descent from five legendary foundation mares, “Al Khamsa” (extant 570-630 CE) ([1,2] as cited by [3]). In modern times, pedigree analyses were systematically applied to the creation of breed registries and studbooks, such as “The General Stud Book,” published by Weatherby’s, the British registry for the Thoroughbred. Constructing pedigrees enabled a scientific approach to livestock breeding and resulted in the creation of many distinctive populations.

Although the practice of artificial selection is ancient, the science of genetics was rooted in the works of Gregor Mendel and Charles Darwin. In 1859, Darwin reported the existence of extensive, heritable variation among all species and speculated on the effects of natural and artificial selection. In 1865, Mendel demonstrated the particulate nature of heredity, characterizing dominant and recessive nature of elements later characterized as genes. Genetics research during the 1900s showed that genes were located on chromosomes [4], demonstrated that DNA was the hereditary material [5], determined the chemical structure of DNA [6], and ended the century with a project to sequence the entire human genome [7,8].

What were the contributions of science to horse genetics during the 1900s? Early work demonstrated Mendelian rules applied to coat color in horses (e.g., Castle et al, 1947). Technological developments made it possible to investigate blood groups, serum proteins, and even DNA variants all of which demonstrated Mendelian characteristics (reviewed in [9]). Scientists pondered the meaning behind these variants and investigated their possible relationships to performance or health traits. But it is arguable that during the last century, the two largest genetic contributions to the horse industry were:

a) Discovery of the cause of hemolytic disease of the newborn foal [10] and
b) Development of parentage testing in horses using blood groups, biochemical markers, and DNA markers [11].

* Corresponding author at: Dr. Ernest Bailey.
E-mail address: ebailey@uky.edu.

0737-0806/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jevs.2015.03.198

One resulted in routine procedures associated with foaling mares and the other a routine practice for registration of horses.

Scientists have used quantitative genetics to characterize performance and complex disease traits in horses (reviewed in [12]). These techniques have been used to improve warmblood and harness racing horses. However, the impact has been less dramatic for horses than when used in cattle, sheep, pigs, and chickens to improve growth, production, or feed efficiency. In general, quantitative genetic studies of horses largely corroborated the prior observations of breeders, namely, heredity was significant and selection could be applied with great benefit. Most efforts directed at horse genetics during the 1900s confirmed the observations of astute breeders (reviewed in [13]). When traits had a single genetic cause, the traits could be shown to have a dominant, codominant, or recessive mode of inheritance (coat colors, blood groups, biochemical markers, and some diseases). When traits were complex, mathematical expressions created in the field of quantitative genetics were found effective to characterize interactions involving multiple genes and management. Regardless, the practice of horse breeders changed little as a consequence of advances in genetics during the 1900s.

2. Genomics

The past century ended with a grand project to make a gene map and ultimately led to creation of a whole genome sequence for the horse [14]. The horse genome project must be viewed in context of the human genome project. In 1990, scientists established a world-wide collaboration to sequence the entire human genome. The information was projected as useful to better understand human health problems and lead to the creation of new therapies. The project was planned for 15 years and a cost of $3 billion. The project was successfully completed in 2003, 2 years early [7,8]. Major discoveries from the project included: (1) the human genome coded for only 20,000 to 25,000 genes, (2) only 3% of the DNA coded for genes, and (3) the organization of genes among diverse species was remarkably similar. Scientists found the information useful in guiding research in every field of human biology and medicine. The major questions arising from the work were, (1) what is the function of the remaining 97% of the DNA? and (2) how is the expression of genes regulated with respect to timing, amount, and tissue specificity? Presumably, if a section of DNA was important for regulation of a particular gene, then we would expect to see it conserved among mammals. Conserved elements would inform us about function. Scientists at the National Human Genome Research Institute began sequencing the genomes of diverse mammals to identify patterns among the 97% of the genome that did not code for genes.

The horse was among mammals chosen for sequencing. The particular horse chosen for sequencing was Twilight, an inbred Thoroughbred mare from a research herd at Cornell University. Scientists at the Broad Institute in Boston, MA, sequenced Twilight in 2006 as part of the National Human Genome Research Institute program. The DNA sequences were made available to researchers immediately, online, with the formal report published later [14].

The formal report included identification of 20,322 putative horse genes and the chromosome location of the sequences with reference to the horse gene map available at the time [15,16]. In addition to sequencing Twilight, the Broad Institute sequenced small genomic segments for seven other horses and compared them with the sequence of Twilight. This led to identification of more than 1 million genetic variants and their genome location. This was a remarkable advance considering that in 1990 less than 100 polymorphic genes had been identified and only seven of them were localized to chromosomes.

2.1. Tools Derivative of Genome Sequence: Genome Wide Association Studies (GWAS)

The significance of the horse genome sequence cannot be underestimated for equine research. The approximate $20M cost for sequencing is probably the largest single contribution to equine research. The tools created new approaches for investigating horse genetics. Before the genome sequence, the only way to study horse genetics was to identify families segregating for traits of interest. The earliest studies simply determined whether the trait was dominant or recessive [17]. After horse gene maps were created in the 1990s, the chromosome location for a gene might be identified using cosegregation of microsatellite DNA markers. Once a trait was mapped, candidate genes might be identified and sequenced. But family studies were not possible for many hereditary diseases because breeders avoided constructing families for these traits. For complex traits, such as racing performance, it would be difficult to assemble a sufficiently large family for a Mendelian analysis.

In 2011, a new tool, called the Equine SNP50 Chip (Illumina, San Diego, CA), was created to test horses for 54,602 genetic variants called single-nucleotide polymorphisms (SNPs) and distributed across the 31 autosomes and the X chromosome. Because the genome includes 2.7 billion bases, this is a very small representation of the total DNA. Indeed, it is unlikely that important genetic variants are actually present on the chip. Fortunately, the relationship among SNPs is not entirely random. The presence of a particular SNP at one location can be highly predictive for the presence of a particular SNP at a nearby location. This phenomena is a result of a character called linkage disequilibrium and gives the appearance of DNA sequences organized into large blocks. The presence of a particular set of SNPs within a region is predictive of the DNA sequence and particular genes in the region. This led to an approach called genome wide association studies (GWAS). The point was to identify SNP signatures in the DNA and investigate their relationship to traits of interest.

Genome wide association studies are a simple statistical approach providing a valuable alternative to family studies. For example, horses can be divided into two groups based on their phenotypes: those with and those without a trait. For example, we can use the SNP Chip to identify the location of the Extension locus, coding for red