MAPPING ATHLETIC PERFORMANCE RELATED GENES IN THE EQUINE GENOME AND A GENOME SCAN FOR SUPERIOR ATHLETIC PERFORMANCE IN THE THOROUGHBRED

A Dissertation

by

KEITH WILLIAM DURKIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2009

Major Subject: Genetics

MAPPING ATHLETIC PERFORMANCE RELATED GENES IN THE EQUINE GENOME AND A GENOME SCAN FOR SUPERIOR ATHLETIC PERFORMANCE IN THE THOROUGHBRED

A Dissertation

by

KEITH WILLIAM DURKIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Bhanu P. Chowdhary
Committee Members,	Terje Raudsepp
	Loren C. Skow
	James E. Womack
	David Adelson
Intercollegiate Faculty Chair	Craig J. Coates

May 2009

Major Subject: Genetics

ABSTRACT

Mapping Athletic Performance Related Genes in the Equine Genome and a Genome Scan for Superior Athletic Performance in the Thoroughbred. (May 2009)

Keith William Durkin, B.S., University of Limerick

Chair of Advisory Committee: Dr. Bhanu Chowdhary

The primary goal of the Thoroughbred industry is to breed and train superior equine athletes capable of excelling on the racetrack. To date, research into the genetic underpinnings of athletic ability has been limited in the horse. Advances in equine genomics and the genetics of athletic performance in humans have opened up the possibility of investigating this important trait in the Thoroughbred.

Initially, 46 candidate genes associated with human athletic performance were mapped in the equine genome by radiation hybrid (RH) and fluorescent *in situ* hybridization (FISH) mapping. RH data and later the draft equine genomic sequence allowed us to identify microsatellites adjacent to these and other candidate genes (95 in total). Additional microsatellites were added to increase genome coverage, producing a final panel of 186 markers. All the potential markers were initially screened on a pool of DNA for 16 Thoroughbreds to ensure they were polymorphic. The panel was genotyped on 162 Thoroughbreds in total; Centimorgans (cM) between microsatellites were determined with CRI-MAP. The animal's athletic ability was estimated using career winnings $\log_e$ transformed to create a linear trait; unraced animals were treated as

missing data. Linkage analysis was carried out using the MERLIN program, and association analysis was carried out using the QTDT program. Appropriate thresholds for statistical significance were determined by carrying out 1000 simulated genome scans based on the structure of the original data. LOD scores above 1.54 met the criteria of statistical significance (with a 5% chance of type I error). In the actual genome scan, the marker *L12.2* had the highest observed LOD score of 1.16 and p-value of 0.01 and consequently was not significant; the association analysis also did not detect significant association with performance on the track.

Given the complexity of the phenotype under investigation and the modest sample size, the lack of linkage/association was not unexpected. Nevertheless, this study has contributed to the RH and FISH maps of the equine genome. Additionally, the development of the genome scanning panel for this study has provided useful information on the most informative microsatellites for linkage or association studies in the Thoroughbred.

DEDICATION

To my family, for their love and support over the years

ACKNOWLEDGEMENTS

I am grateful to a large number of people without whom I could not have carried out this work. Firstly my committee chair, Dr. Chowdhary, for his help, guidance and encouragement. My committee members: Dr. Raudsepp for her patience in teaching me many of the techniques I have used in this work and her always helpful advice, Dr. Skow for his open office door, helpful discussion and suggestions and Dr. Womack and Dr. Adelson for their helpful comments and discussion.

I would especially like to thank all the graduate students, student workers, technicians and post docs that have worked in Dr. Chowdary's, Dr. Skow's and Dr. Womack's labs during my time at Texas A&M, especially Glenda Goh, Avni Santani, Candice Brinkmeyer Langford, Sandra Afonso, Ashley Gustafson Seabury and Christopher Seabury for their help in my work.

I would also like to thank Dr. Varner for his assistance in obtaining many of the stallion samples and Stonerside stables for providing the majority of the sample population. Thanks also to Dr. Huber and Dr. Taylor for their advice on analysis of the data and all those who I have left out.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES x

CHAPTER

I INTRODUCTION 1

Rationale of the study 1
Origins of the Thoroughbred 6
Athletic ability in the Thoroughbred 7
Energy demands of racing and genetic contribution 9
Measurement of athletic performance 11
Heritability (^2h) of athletic ability in the Thoroughbred 13
Breeding for superior athletic ability and effect of selection 18
Genetics of athletic ability in humans and candidate genes 22
Athletic performance in model organisms 39
Mapping QTLs in livestock species and potential difficulties 42
Equine QTLs and genomics 46
Study objectives 52

II RADIATION HYBRID AND FLUORESCENT IN SITU HYBRID MAPPING OF 46 HUMAN GENES ASSOCIATED WITH ATHLETIC PERFORMANCE IN THE EQUINE GENOME 55

Introduction 55
Materials and methods 58
Results 63
Discussion 70

CHAPTER Page

III CONSTRUCTING A GENOME SCANNING PANEL FOR USE IN THE THOROUGHBRED 76

Introduction 76
Materials and methods 83
Results 90
Discussion 99

IV A GENOME SCAN FOR ATHLETIC PERFORMANCE IN THE THOROUGHBRED 109

Introduction 109
Materials and methods 113
Results 118
Discussion 123

V FUTURE PROSPECTS AND CONCLUSIONS 132

Future prospects 132
Insights from other species 135
Conclusion 137

REFERENCES 139

APPENDIX A 165

APPENDIX B 217

VITA 220

LIST OF FIGURES

FIGURE Page

2-1 Representative images of double color FISH hybridizations on partial equine metaphase using BAC clones containing the following loci 68

2-2 Location of the 95 human candidate genes in the equine genome 69

3-1 Genotypes for pooled DNA and the individuals making up the pool for microsatellites *COR033* (six alleles) and *COR025* (one allele) 91

3-2 Representative figures showing location of candidate genes and microsatellites.. 96

3-3 Peak morphology before and after primer redesign 99

4-1 Distribution of lifetime winnings after log_e transformation for all animals, including those that were not genotyped.. 120

4-2 Plots for the six chromosomes with the highest LOD scores calculated by the MERLIN program using the “--qtl” option 122

LIST OF TABLES

TABLE Page

2-1 Human candidate genes mapped in the equine genome in this study, published in Raudsepp *et al*. (2008a) 65

3-1 Summary of marker information, cM distances and information content for markers on each equine chromosome.. 95

4-1 Summary of family information showing number of animals in each family, number of animals genotyped in each family and the average $\log_e$ transformed winnings for each family 119

CHAPTER I

INTRODUCTION

Rationale of the study

Thoroughbred racing is an important industry in both the United States and throughout the world. In 2004 alone, it was estimated that Thoroughbred racing had a 20,271 million dollar impact (includes direct, indirect and induced effects) on the gross domestic product of the United States economy (http://www.jockeyclub.com/-factbook.asp?section=18). The primary goal of this global industry is to produce, train and race animals capable of excelling on the racetrack. Consequently athletic ability/performance is of paramount importance to those involved in the Thoroughbred industry. Despite the central role athletic ability/performance plays in the Thoroughbred, investigations into the genetic underpinnings of this trait have been largely confined to estimating the heritability of race winnings or handicap ratings using race and pedigree records (Ricard *et al.* 2000). Unlike other domestic species such as cattle, pig and chicken where numerous studies have been conducted to identify regions of the genome associated with complex and economically important phenotypes, no systematic studies have as yet been undertaken to discover regions of the equine genome contributing to athletic performance.

Within the Thoroughbred industry there is a firm belief that performance on the track is heritable, a fact reflected in the multimillion dollar sales prices routinely paid for

---

This dissertation follows the style of *Animal Genetics*.

untested yearlings from successful bloodlines. Research into the heritability of racing performance and performance over different distances and track surfaces has shown a moderate heritability for these traits (Tolley *et al.* 1985; Williamson & Beilharz 1998; Ricard *et al.* 2000). As a consequence it is reasonable to assume that there are genes influencing athletic performance segregating within the Thoroughbred population.

Athletic performance is obviously a complex trait influenced by both environmental (nutrition, training, management etc) and genetic factors. This genetic contribution is most likely controlled by a number of genes, each with differing levels of influence on the phenotype. Identifying the underlying polymorphisms that affect such a trait is more challenging than locating polymorphisms affecting simple Mendelian traits as the correlation between polymorphism and phenotype is often substantially weaker (Andersson & Georges 2004). However despite this caveat, research in humans (Rankinen *et al.* 2006) and model organisms (Ways *et al.* 2007; Lightfoot *et al.* 2008), has shown that it is possible to identify regions of the genome harboring genes influencing athletic performance, using whole genome scanning panels of microsatellite markers and the candidate gene approach. Work in these species has also produced an extensive list of candidate genes that may also influence athletic performance in the Thoroughbred.

At the present time breeding and the prediction of racing potential in the Thoroughbred remains a largely subjective process, mainly based on pedigree records. The identification of genomic polymorphisms influencing athletic performance has the

potential to place the selection for, and the identification of superior athletic performance on a more objective footing. Such polymorphism when identified will be valuable in recognizing animals with a genotype conducive to superior performance on the track, determining better crosses between stallions and mares and identifying which distance/track surface are best suited to the animal. While such goals remain aspirational at the moment, this study is a preliminary attempt to identify regions of the equine genome containing polymorphisms that influence athletic ability in the Thoroughbred.

Compared to the other economically important domestic species, genomics in the horse had a late start (Chowdhary & Bailey 2003). The early development of gene and genetic maps in cattle and pig combined with the availability of experiential populations allowed researchers in these species to identify a large numbers of Quantitative Trait Loci (QTLs) over the past decade (http://www.animalgenome.org/cgi-bin/QTLdb/SS/summary, http://www.animalgenome.org/QTLdb/cattle.html). However in the horse progress on this front has been limited. While it remains problematic to produce experimental populations in the horses, especially for a trait such as athletic performance, it is possible (although often difficult) to collect large half-sib families for analysis. Additionally, in the Thoroughbred, pedigree and phenotypic data is often available going back many generations, a fact that makes the Thoroughbred an attractive model for studying the genetics of athletic performance, provided the necessary genomic tools are available.

The last number of years has seen remarkable progress in equine genomics. Initial work using ZOO-FISH revealed the remarkable conservation between the horse

and human genomes (Raudsepp *et al.* 1996). While linkage (Guerin *et al.* 1999; Swinburne *et al.* 2000a; Guérin *et al.* 2003; Penedo *et al.* 2005; Swinburne *et al.* 2006) and radiation hybrid (RH) maps (Chowdhary *et al.* 2003; Perrocheau *et al.* 2006; Raudsepp *et al.* 2008a) have provided a wealth of microsatellite markers suitable for genome scanning and the mapping information to identify which candidate genes are located beside these microsatellites. The CHORI-241 equine BAC library has also been an invaluable tool for constructing BAC based contigs (Gustafson *et al.* 2003; Brinkmeyer-Langford *et al.* 2008), providing FISH probes to anchor RH and linkage groups to specific chromosomes and producing BAC based physical maps of the equine genome (Leeb *et al.* 2006; Woehlke *et al.* 2008). Finally, the recent sequencing of the equine genome at the Broad Institute of MIT and Harvard, and the release of a draft assembly of the sequence is a major milestone in the rapid development of equine genomics (http://www.broad.mit.edu/mammals/horse/). These developments have for the first time made investigations into the genetic underpinning of complex traits in the horse practical. The application of these resources is already beginning to bear fruit as can be seen from the increasing number of monogenic and even complex traits mapped/studied in the horse (Dierks *et al.* 2007; Diesterbeck *et al.* 2007; Chowdhary & Raudsepp 2008).

As has been pointed out already athletic performance is the primary trait of importance in the Thoroughbred. Research has demonstrated its heritability and investigations in other species have shown the feasibility of identifying genes and other polymorphisms associated with this phenotype. As a consequence it appears prudent to

begin investigations into the genetic underpinnings of this important phenotype in the horse. This initial chapter will provide a review of the available literature relevant to this project. It is hoped this will give an overview of the progress in the field and provide an insight into the rationale behind the approaches taken in this study. The chapter covers a number of areas, beginning with the origins of the Thoroughbred, its remarkable athletic abilities and the significant literature devoted to the measurement and heritability of athletic performance. This information is important for deciding how to measure the phenotype and gives an indication of the likelihood of identifying genes associated with this phenotype.

In addition to developments in the field of equine genomics an equally important impetus that spurred us to begin this investigation was the rapid growth of studies examining athletic performance and exercise related phenotypes in humans and model organisms. These studies have provided a wealth of candidate genes to investigate for association with performance in the horse. The most promising candidates are discussed along with some of the whole genome studies carried out in humans and rodents.

While it is possible to point to a number of success stories, attempting to map complex traits such as athletic performance remains a difficult task. In complex traits the correlation between phenotype and genotype is often weak making the identification of regions of the genome contributing to the phenotype difficult. Additional complications such as epistasis and imprinting can conspire to further obscure the picture. These potential pitfalls are outlined. Finally some of the recent examples of QTL mapping in the horse are discussed and the status of equine genomics outlined. It is hoped that this

review will show that the approach taken in this project was not arrived at arbitrarily and was reasonable given the resources and tools at our disposal.

The second chapter will deal with the initial mapping of candidate genes previously associated with athletic performance in humans, a brief discussion of the approach taken is given along with some of the salient mapping results. Chapter III covers the development of the genome scanning panel of microsatellites, outlining the process of marker identification and genotyping of the sample population. Chapter IV details the analysis of the genotypes using linkage and association analysis, in an attempt to identify regions of the genome containing polymorphisms contributing to superior athletic performance. Chapter V briefly discusses the prospects for mapping complex traits in the horse given some of the recent developments in other species and outlines some brief conclusions.

Origins of the Thoroughbred

The Thoroughbred is one of the oldest and best documented domestic breeds with pedigree records dating back to 1791 and the establishment of the studbook by James Weatherby. The breed was originally developed in England from a small number of stallions of Middle Eastern and North African origin that were crossed with native English mares to produce a breed of horse specifically for racing (Willett 1981; Cunningham 1991). Pedigree records indicate that only 31 animals (12 mares and 19 stallions) are responsible for 80% of the genetic make up of the Thoroughbred, with four of these stallions responsible for one third of the alleles in the current population

(Cunningham 1991; Cunningham *et al.* 2001). A small amount of additional genetic variability may have been introduced in the 1800s when the Thoroughbred spread to Australia and the United States and was crossed with local Australian and American mares (Bailey 1998). Recent molecular data shows a similar pattern, using 13 microsatellite markers genotyped in 211 Thoroughbred horses Cunningham *et al.* (2001) demonstrated that 78% of alleles in the population can be traced to just 30 founding animals. From this limited number of founders, the Thoroughbred population has expanded dramatically over the subsequent centuries to the point where the current annual foal crop exceeds 100,000 registered animals worldwide (http://www.jockeyclub.com/factbook.asp?-section=17).

Athletic ability in the Thoroughbred

The horse is a born athlete. As a prey animal natural selection has shaped its physiology to facilitate rapid escape from predators. In the case of the Thoroughbred, humans have further enhanced these traits through selective breeding for improved performance on the racetrack. Athletic ability is a complex trait, arising from the interplay of both environmental and genetic factors. In more specific terms, athletic ability is the result of interactions between a number of physiological and psychological factors, including:

1) Aerobic capacity, influenced by:
   a) Heart rate
   b) Stroke volume

c) Gas exchange in lungs/muscle

d) Hemoglobin concentration

e) Percent Type I muscle fibers

2) Anaerobic capacity, influenced by:

a) Percent Type II muscle fibers

b) Substrate availability

3) Biomechanics (efficiency of the animals stride).

4) Psychological make up or "the will to win."

5) Training state or response to training (Rose & Hodgson 1994).

Research in the field of equine sports medicine has highlighted the remarkable athletic abilities of the different breeds. At one extreme the Quarter horse is capable of reaching almost 70 km/h in races over 400m, while at the other extreme Arabians can complete endurance races that cover between 80 and 160 km in a day at speeds of 16 km/h (Ridgway 1994). Much of the horses, and especially the Thoroughbreds, remarkable athletic ability can be attributed to its prodigious aerobic capacity. In order to highlight this it is useful to get a point of reference by comparing some physiological parameters between the Thoroughbred and human athletes.

- VO_2max (maximal amount of oxygen an athlete is capable of transporting and utilizing, expressed per Kg of bodyweight): Thoroughbreds have a VO_2max (160ml O_2/Kg/min) twice that of the élite human athlete (69-85ml O_2/Kg/min). In other words when the disparity in body weight is factored in, the Thoroughbred is capable of delivering and utilizing double the volume of oxygen

an élite human athlete is capable of (untrained human males have a VO_2max of 45ml O_2/Kg/mim) (Derman & Noakes 1994).

- The Thoroughbred can increase its heart rate almost tenfold from a resting rate in the low 20s to almost 250 beats per min; while the human counterpart can only manage a four-fold increase, from a resting rate of between 40 and 60 beats per min to a rate in the range of 180 to 200 beats per min (Derman & Noakes 1994).
- The resting stroke index (stroke volume/body weight) for the Thoroughbred is 1.3 to 2.3 ml/kg; this can increase to between 2.5 and 2.7 ml/kg during maximal exercise. In humans the index is 1.1 to 1.4 ml/kg, and it increases to 1.5 ml/kg during maximal exercise (Derman & Noakes 1994).
- In humans resting hematocrit is 40% to 50%, and it rises on average by 4% during exercise, mainly due to fluid loss from the circulation (Laub *et al.* 1993). Compared to this resting hematocrit in the Thoroughbred is 32% to 46%; during maximal exercise this can rise to between 60% and 70%. (Derman & Noakes 1994).

From the figures above it can be seen that the Thoroughbred has a remarkable ability to take in and transport oxygen and this capacity obviously plays a large role in the Thoroughbreds remarkable athletic abilities.

Energy demands of racing and genetic contribution

Thoroughbred races are generally run between five furlongs (1006 m) and two miles (3219 m). Consequently the metabolic demands are likely to be different over a

five furlong sprint in comparison to a two mile race (Evans 1994). Over shorter distances it was initially expected that Thoroughbreds would show a greater reliance on anaerobic pathways for energy production, with aerobic respiration coming to dominate as the length of race increases. Surprisingly, even over short five-furlong races, 70% of the required energy is supplied aerobically, while in longer races the vast majority of energy requirements are met by aerobic pathways (Eaton *et al.* 1995). Only in Quarter Horse races over 400m, that last for ~22s, do we see anaerobic pathways provide more than 50% of the required energy (Eaton *et al.* 1995). The surprisingly large role of aerobic respiration over short distances compared to humans competing over equivalent distances is mainly attributable to the remarkable oxygen carrying capacity of the horse (Derman & Noakes 1994). However while anaerobic capacity may not play as large a role in Thoroughbred racing as initially assumed, it is still likely to play a significant role in shorter races where small differences in performance can make the difference between first and second.

Muscular contraction during exercise is fueled primarily by glucose metabolism (stored in the muscles as glycogen) producing Adenosine-5'-triphosphate (ATP, the "currency" of molecular energy in the cell) in two ways. During the initial stages of a race when O_2 is limited, anaerobic pathways rapidly produce energy from glucose and glycogen, by producing two and three molecules of ATP per molecule of glucose/glycogen respectively along with lactic acid as a byproduct (Eaton 1994; Snow & Valberg 1994). As the cardiovascular system begins delivering increased supplies of O_2 the more efficient aerobic metabolic pathways predominate. If exercise intensity

exceeds the aerobic capacity, anaerobic metabolism attempts to make up the difference. During a race the Thoroughbred exercises at an intensity that requires a contribution from both aerobic and anaerobic metabolism (Eaton 1994; Snow & Valberg 1994).

As has been pointed out previously a number of physiological and psychological factors contribute to athletic performance, as a consequence a considerable number of genes have the potential to play a role in this trait. Most of these genes are thought to exert only a small effect on the phenotype and as a consequence it is not possible to detect their influence without very large sample populations (Falconer & Mackay 1996b). Additionally such minor effects would provide little additional information to the breeder or trainer of the Thoroughbred, only becoming meaningful at the population level. However in complex traits in addition to the many genes exerting a minor effect, there are often a small number of genes that play a major role in the phenotype (Andersson & Georges 2004). Given the central role physiological factors such as aerobic/anaerobic capacity play in performance it is the genes controlling these traits that represent some of the best potential candidates for explaining performance variability seen in the Thoroughbred.

Measurement of athletic performance

When mapping any genetic trait, accurately measuring the phenotype under examination is vital to identifying linkage or association between the phenotype and the polymorphic markers employed in the study (Carlson *et al.* 2004). In the Thoroughbred the ultimate test of athletic ability has always been performance on the racetrack.

However deciding on how to measure race performance to facilitate comparison of athletic ability between horses can be a difficult task. Over the years attempts to quantify athletic performance in the Thoroughbred have fallen into three main categories based on:

1. Race time
2. Handicap rating
3. Earnings/race rank

1. Race time While speed against the clock may at first glance seem the most logical measurement of performance it must be remembered that races are not run against the clock but against other horses. Race times for classic races such as the Belmont, English Oaks and Derby have shown little improvement in winning time over the last number of decades despite continued selection for athletic ability (Gaffney & Cunningham 1988; Cunningham 1991). Additionally, as will be outlined in grater detail below, heritability for race time in some studies is close to zero (Chico 1994; Langlois 1996), arguing that speed alone does not equate with success on the track.

2. Handicap rating In Thoroughbred racing, horses that perform well on the track are assigned heavier weights to carry in subsequent races. Therefore in a handicapped race, such as the Melbourne cup, all animals should, in theory, have an equal chance of winning. These handicap weights can be expressed as a rating to reflect the past performance of the horse and can be used to compare the athletic ability of different

animals (Tolley *et al.* 1985). While the handicap rating of an animal has a large subjective component, it probably more accurately reflects athletic ability than race times, especially over longer racing distances (Langlois 1980).

3. Earnings/race rank As mentioned earlier the goal of Thoroughbred racing is to be first past the post. As a consequence the most appropriate measurement of performance in the Thoroughbred is race ranking or winnings. Langlois (1996) favored using race ranking as a measure of performance but also points to the usefulness of earnings once they have been log transformed to create a linear rather than exponential trait. In racing, as the prize money increases better quality horses tend to enter the race, thus increasing the difficulty of winning. As a result lifetime earnings can be seen as a reflection of the horses ability to perform well on the track against its peers. While far from a perfect measurement of performance, information on earnings are readily accessible for the majority of Thoroughbreds and is probably the most commonly used indicator of athletic performance at the present time (Langlois & Blouin 2004).

Heritability (^{2}h) of athletic ability in the Thoroughbred

"The most effective way to become a champion athlete is to be selective when choosing ones parents" (Derman & Noakes 1994). While this approach is not practical (or ethical) in humans, the large stud fees and exorbitant prices paid for well bred horses demonstrates that the Thoroughbred industry places a very real price on "good genes". Despite the extensive pedigree and performance records available for the Thoroughbred

there has been a paucity of research into the genetics of athletic performance. This research is strikingly limited, particularly when compared to research devoted to the identification of complex and economically important traits in other livestock species (Ricard *et al.* 2000). To date the majority of investigations into the genetics of athletic performance in the Thoroughbred have concentrated on examining the heritability of racing performance, using measurements such as race time, handicap rating and earnings. Examining the heritability of a trait is an important step in determining if it is feasible to identify associated genes. A trait with little or no heritability is a poor candidate for genetic analysis as the alleles passed down the generations would have little or no effect on the variation observed in the population under examination (Falconer & Mackay 1996a).

Heritability of race time

Tolley *et al.* (1985) in an extensive review of the literature dealing with the heritability of racing performance in horses identified ten studies that examined the heritability of racing time, with estimates ranging from 0.09 to 0.78. Richard *et al.* (2000) pointed out that many of these early studies were based on relatively small numbers of animals, in the studies with reasonable sample sizes most estimates fell between 0.1 and 0.2.

Chico (1994) analyzed Spanish Thoroughbreds competing in Grand-Prix races (élite races, where competitors carry equal weight) and found heritability to be zero for racing time. Looking at over 25,000 Thoroughbreds in Japan, Oki *et al.* (1995) examined

the heritability of racing time over a variety of distances on turf and dirt tracks. Heritability was broadly similar on both racing surfaces, however a clear decrease in racing time heritability was seen as the length of race increased, heritability at 1000 m was 0.25 and this dropped to 0.08 at 2000 m.

In order to provide some context for the above figures it is worth mentioning the heritability for some other economically important traits in livestock species. In cattle heritability for milk yield is 0.35, and for bodyweight is 0.65, in pigs back-fat thickness is 0.7, weight gain per day 0.4 and litter size 0.05 (Falconer & Mackay 1996a). The low heritability seen for race time in the more recent studies indicates that measuring race time is a poor way to quantify performance on the track, especially in longer races (Langlois 1996). However it is interesting to note that over shorter races a higher heritability is observed indicating that raw speed is more important over these distances. As anaerobic metabolism plays a greater role over shorter distances genes affecting this system may be good candidates for influencing performance in sprinters.

Heritability of handicap rating

Tolley *et al.* (1985) provides an excellent review of the earlier studies examining heritability of handicap rating. In total eight studies were discussed. Heritability for this indicator of performance showed a good deal of variability with estimates based on regression to the sire and paternal half sib showing very large values. These values were generally discounted due to a correlation between phenotype of the sire and the environment of the offspring as breeders that can afford stallions with high stud fees are

likely to provide higher quality early care and training (Ricard *et al.* 2000). For regression based on the dam, the majority of these studies showed heritability between 0.3 and 0.4 (Tolley *et al.* 1985). Gaffney and Cunningham (1988) using the Timeform handicap ratings over the period 1961 to 1985 for 31,263 three year olds, the offspring of 2,087 sires, found heritability to be 0.39 ± 0.013 when regressed on the mare. As seen in previous studies, regression on the stallion showed very high heritability (0.76 ± 0.023) indicating that an environmental bias was inflating the heritability estimate.

Heritability of earnings/race rank

A good overview of the literature prior to 1985 is again provided by Tolley *et al.* (1985), with five studies examining the heritability of race earnings reviewed. The earnings were log transformed to normalize the distribution and heritability estimates fell between 0.23 and 0.56, with the majority of the estimates towards the lower end of the series. Recent studies have produced more modest estimates. Chico (1994) looked at Thoroughbreds in Spain and found heritability for earnings to be 0.1. In Poland Sobczynska and Lukaszewicz (2004) found heritability for earnings to be 0.12, while Svobodova *et al.* (2005) looking at Thoroughbreds in the Czech Republic found heritability for career earnings to be 0.32. Finally, in a study likely to upset the owners of stallions with high stud fees, Wilson and Rambaut (2008) looked at stallions in the U.K. and U.S.A. and estimated that while lifetime earnings were 0.095 heritable, there was no genetic correlation between a stallions stud fee and the lifetime earnings of the animals offspring.

From these studies on handicap rating and earnings it can be concluded that performance on the track is moderately heritable in the Thoroughbred population. The modest heritability of the trait does make the process of mapping genes associated with the phenotype more difficult. Any major QTL segregating within the population is likely to explain a relatively modest portion of the phenotypic variation, making it difficult to observe the signal of linkage among the noise generated by other variables. Nevertheless QTLs for complex traits with a similar heritability, such as milk yield in the cow (Boichard *et al.* 2003; Ashwell *et al.* 2004) and osteochondrosis in the horse (Dierks *et al.* 2007; Wittwer *et al.* 2007) have been successfully mapped to specific genomic regions using panels of microsatellite markers.

Specific genes associated with athletic performance in the horse

While a number of studies have shown that athletic performance has a modest heritable component in the horse, to date only one study has shown an association between specific genomic polymorphisms and athletic performance. Harrison and Turrion-Gomez (2006) showed that different mitochondrial haplotypes were correlated with performance at different racing distances in English Thoroughbreds competing in classic races. Suggesting that the mitochondrial genome also plays a role in athletic performance in the Thoroughbred.

Breeding for superior athletic ability and effect of selection

Since the foundation of the breed, the Thoroughbred has been bred with the primary goal of producing an animal capable of excelling on the track. This has traditionally been achieved by selectively breeding animals that have succeeded on the racetrack. Also considered in the selection process is the pedigree of the animal, taking into account the performance of ancestors and relatives. In addition to performance on the racetrack other traits such as conformation and temperament are evaluated to ensure that the offspring are capable of standing up to the physical and psychological rigors of training and racing.

With such extensive pedigree/performance records and selection for a single (although complex) trait it might be assumed that the Thoroughbred industry would be well positioned to exploit modern breeding methods. However unlike other domestic species such as cattle, pig and chicken, to date the Thoroughbred industry has been largely recalcitrant to selection using specific breeding values based on the past athletic performance of the individual animal and relatives (Cunningham 1991; Langlois 1996). Despite the apathy of the Thoroughbred industry, the Best Linear Unbiased Predictor (BLUP) animal model (Tavernier 1988) has been adapted to estimate breeding values in the Thoroughbred and other sport horses (Tavernier 1988; Langlois *et al.* 1996; Langlois & Blouin 2004). The application of such a model has the potential to improve the efficiency of selection for superior athletic performance in the Thoroughbred over traditional methods. However, the conservative and fragmented nature of the industry

make widespread application of such a model unlikely in the near future (Langlois & Blouin 2004).

Artificial selection in agricultural important species based on the observable quantitative phenotypes has been extremely successful over the last 50 years in increasing productivity in both plants and animals (Dekkers & Hospital 2002). The past twenty years have also seen the increasing application of molecular genetics to dissect the molecular underpinnings of economically important traits in agriculturally important species. Efforts have also been made to utilize this information in marker assisted selection for superior productivity and disease resistance (Andersson 2001; Dekkers 2004). If specific alleles affecting athletic ability can be located in the horse the information could be utilized in selecting for superior athletic performance and in identifying promising animals before they have been tested on the track. Additionally, unlike other livestock species where the individual animals monetary value is low and genotyping costs can quickly erode a tight profit margin. In the Thoroughbred identifying important alleles via genotyping would only represent a tiny fraction of the animal's value.

Cunningham's Paradox- Static race times and genetic progress

Studies examining heritability of racing performance while far from perfect, do at least show a consistent if often modest heritability for performance on the racetrack. Selection intensities in the Thoroughbred are at around 50% in the mare and 5% in the stallion and generation intervals average 10 years (Langlois 1980; Gaffney &

Cunningham 1988). As a consequence one would expect genetic gain in the Thoroughbred population for performance on the track. Using Timeform handicap rating gathered for 11,328 animals that raced as three year olds between 1961 and 1985 Gaffney and Cunningham (1988) used a sire model best linear unbiased prediction (BLUP) analysis to estimate a genetic gain in Timeform rating of 0.94 ± 0.13 a year. More recently Wilson and Rambaut (2008) looked at stallion's in the U.K. and U.S.A. and estimated that while lifetime earnings were 0.095 heritable, there was no genetic correlation between a stallions stud fee and the lifetime earnings of the animals offspring. Similar results have also been observed in the Quarter Horse that races over much shorter distances (320m) than the Thoroughbred. Wilson *et al.* (1988) after examining over 1 million race records between 1960 and 1983 found that that finish times were decreasing by 0.0088, 0.0090 and 0.0037 seconds per year for the 320, 366 and 402 m races, respectively, due to effect of selective breeding.

Despite the apparent heritability of performance on the track and selection for the phenotype, over the last number of decades it has been observed that Thoroughbreds do not appear to be getting appreciably. Winning time for the Classic English races the St Ledger, the Derby and the Oaks, and the longer leg of the American triple crown the Belmont have improved little in modern times. It is worth mentioning that some improvement has been observed in the two shorter legs of the American triple crown, the Kentucky Derby and the Preakness (Gaffney & Cunningham 1988; Cunningham 1991). This apparent discrepancy between heritability and response to selection was christened by Prof Alan Robertson in 1975 as "Cunningham's paradox" (Langlois 1980).

A number of suggestions have been put forward to explain this discrepancy. Firstly, many of the early investigations into heritability produced large heritability estimates that have since been discounted as they used regression on the sire methods that tend to inflate the estimates, due to the environmental correlations between offspring and the phenotype of the sire (Ricard *et al.* 2000). As a result in many cases, calculated rates of genetic improvement based on these inflated heritabilities were overestimated. However more carefully controlled estimations of genetic gain still point towards an expected improvement in performance on the track.

Operating under the assumption that the reported heritabilities for performance are based on some genetic contribution to superior performance, Dr Cunningham himself speculated in 1975 that the observed heritability is the result of a depletion of the additive genetic variation in the Thoroughbred population (Tolley *et al.* 1985). A similar point was again made by Gaffney and Cunningham (1988) when they pointed out that while race times in some élite races have remained static, these times are posted by the top performers in the population. They believe that the observed genetic gain is occurring in the overall population, not in élite animals competing in the top races. As a result the average horse is getting better while the élite remain static.

Langlois (1980) highlighted the point that speed and racing ability are not the same thing. In the Thoroughbred what counts most is which horse wins, not which animal is capable of the fastest race time. He speculated that during the foundation stages of the breed raw speed was of greater importance. However, as the level of athletic ability in the population improved due to the influence of selection, a number of

different and subtler factors (such as temperament) increased in importance. This may be an especially important point in herd animals such as the horse. Many Thoroughbreds may have similar athletic abilities, but it is the animal that is most willing to break away from the field that wins. Therefore in the modern Thoroughbred it may be variation in less obvious genes affecting traits such as temperament, which are mainly contributing to performance on the track. As a consequence while it is important to first examine more conventional candidate genes it is prudent to also consider a genome wide approach to identify less obvious genes affecting performance on the racetrack.

Genetics of athletic ability in humans and candidate genes

It has long been realized that athletic ability in humans has a large genetic component. Even with the best training and the will to succeed, relatively few individuals are capable of achieving élite athlete status. Initial studies in humans investigated the extent to which genetics affects the physiological systems (such as: VO_2max, anaerobic capacity, stroke volume, etc) considered important in athletic performance. In the case of VO_2max, heritability estimates have ranged from 0.4 to 0.6 (Bouchard *et al.* 1992; Bouchard *et al.* 2000; Feitosa *et al.* 2002; Rupert 2003). While studies into the heritability of anaerobic capacity have also shown high estimates, ranging from 0.31 to 0.86 (Bouchard *et al.* 1992). As regards identifying specific genes associated with athletic performance in humans, a good deal of work has been carried out over the last ten years, with a number of groups investigating athletic performance using both whole genome scans and candidate gene approaches. The interest in this field

is highlighted by the rapid increase in the number of articles being published on the subject. The initial human gene map of performance and health related phenotypes published in 2001 contained 42 autosomal genes and quantitative trait loci (QTL) (Rankinen *et al.* 2002b) while the 2005 update contained 165 autosomal genes and QTLs (Rankinen *et al.* 2006). The weight of evidence supporting the association between athletic performance and many of these genes is often confined to a single small-scale study (Rankinen *et al.* 2006). In other complex phenotypes such as schizophrenia, reports of linkage between the phenotype and region of the genome have often been difficult to replicate (Glazier *et al.* 2002). As a consequence it is likely that many of these candidate genes will fail to show statistically significant linkage/association with athletic performance in different populations. However, a few genes such as the *ACE* I/D polymorphism have repeatedly shown association with superior athletic performance in different populations (Rankinen *et al.* 2006). This and other promising candidate genes are discussed in greater detail below.

The Angiotensin-Converting Enzyme (ACE)

To date the most extensively studied gene associated with athletic ability/performance phenotype in humans is the angiotensin I converting enzyme (*ACE*) gene. The *ACE* gene is involved in the conversion of angiotensin I to angiotensin II and the degradation of vasodilator kinins. As a result the ACE enzyme is an important component of the endocrine rennin-angiotensin system that helps regulate blood pressure (Myerson *et al.* 1999; Bray 2000). In humans there are two major variants of the *ACE*

gene, the first variant the insertion (I) allele, contains an extra 287 base pair fragment, while the second known as the deletion (D) allele lacks this insertion. The D allele is associated with relatively higher ACE activity (Myerson *et al.* 1999).

There have been conflicting reports in the literature as regards this I/D polymorphism. A number of studies have associated the I allele with increased endurance performance in different groups, including but not limited to: British high-altitude mountaineers (Montgomery *et al.* 1998), British distance runners (Myerson *et al.* 1999), Australian rowers (Gayagay *et al.* 1998) and Russian middle distance athletes (Nazarov *et al.* 2001). Additionally, in patients with congestive stable heart failure, individuals that are homozygous for the D allele have decreased exercise capacity (Abraham *et al.* 2002). However other studies did not find such an association. For example, Taylor *et al.* (1999) looking at 120 Australian athletes competing in highly aerobic sports, Rankinen *et al.* (2000b) looking at 192 endurance athletes from Canada, Germany, Finland and the United States and Scott *et al.* (2005) looking at 291 Kenyan endurance athletes, failed to find an association between the *ACE* I/D polymorphism and performance (Taylor *et al.* 1999). Finally in the HERITAGE family study (Gagnon *et al.* 1996) with a sample population of 724 individuals, undergoing 20 weeks of endurance training, no association between *ACE* I/D polymorphism and cardiorespiratory performance was found (Rankinen *et al.* 2000a).

As regards the D allele, there is some evidence that this allele may be associated with performance in events that require a large contribution from anaerobic respiration. An excess of the D allele has been seen among Russian athletes competing in power

orientated events (Nazarov *et al.* 2001). In addition, an excess of the D allele was seen in British swimmers and runners competing over short distances (Myerson *et al.* 1999).

The *ACE* I/D polymorphism has also been investigated in relation to its effect on response to training. In a study using British army recruits an association was found between the I allele and increased duration of performance of elbow flexions when holding a 15 kg barbell following a 10 week training program (Montgomery *et al.* 1998). Another study also using Army recruits found an association between the I allele and improvements in delta efficiency (% ratio of change in work performed per min to the change in energy expanded per min) when exercising on a cycle ergometer following an 11 week physical training program (Williams *et al.* 2000). As regards the D allele, carriers showed a greater response to quadriceps strength training than I homozygous individuals (Rankinen *et al.* 2001). However in a study using 147 U.S. army recruits, no association was found between *ACE* genotype and improvements in aerobic performance and performance in sit-ups and push-ups following 8 weeks of basic training (Sonna *et al.* 2001). Finally Myerson *et al.* (2001) using 141 British Army recruits found individuals homozygous for the D allele had a greater increase in left ventricular mass when compared to subjects with the II genotype after completing a 10 week exercise training program.

The literature cites a number of conflicting opinions in relation to the *ACE* gene. The problem has also been exacerbated by the lack of clarity in how exactly differing levels of the angiotensin converting enzyme could affect performance. Some authors have cast doubt on the relationship between the *ACE* I/D polymorphism and endurance

performance. Rankinen *et al.* (2000b) pointed out that studies showing a positive correlation with endurance tend to have a small sample size and less rigorous phenotypic measurement, while larger studies with controlled phenotypic measurement show no correlation with superior endurance performance. On the other hand Nazarov (2001) points out that examining a large number of athletes from an array of disciplines without proper stratification according to standard of athlete and duration of event results in a masking of the association between the *ACE* I/D polymorphism and superior endurance performance. It has also been suggested that the *ACE* genotype exerts its effect through a local muscle effect and as a consequence is independent of VO_2max and this may account for the findings in some of the negative studies (Woods *et al.* 2000; Perusse *et al.* 2003). Another factor that may account for the conflicting results observed is the strong linkage disequilibrium between the *ACE* I/D polymorphism and many potentially important functional polymorphisms in the *ACE* gene in European populations (Scott *et al.* 2005). In other populations this strong linkage disequilibrium is absent (Scott *et al.* 2005). Finally it has also been suggested that the effect of the *ACE* I/D polymorphism could be the result of linkage disequilibrium with other genes, such as the adjacent human growth hormone gene (*hGH*) (Rankinen *et al.* 2000a; Rupert 2003). On balance, it appears that the *ACE* I/D polymorphism is pointing to some underlying genomic polymorphism contributing to athletic performance in humans. However, better controlled and larger scale studies that account for ethnic background are required to make definitive conclusions about the role of this gene.

Skeletal Muscle-Specific Creatine Kinase (CKMM)

CKMM is an enzyme found in vertebrates and is responsible for catalyzing the transfer of a high-energy phosphate from phosphocreatine to ADP, with the resultant formation of ATP and creatine. In horses, the CKMM enzyme activity in fast twitch type II muscle fibers is twice that seen in slow twitch type I muscle fibers (Snow & Valberg 1994). Genetically modified mice deficient for the CKMM enzyme show improved endurance but lose the ability to perform short bursts of high intensity activity (van Deursen *et al.* 1993). While in humans Rivera *et al.* (1997) in the HERITAGE family study found an association between a polymorphism in the 3' untranslated region of *CKMM* recognized by the Ncol restriction enzyme and VO_2max response to a 20 week regime of endurance training. In this study *CKMM* genotype accounted for at least 9% of the variation seen in VO_2max.

α-Actinin-3 (ACTN3)

The sarcomere is the basic unit of the myofibril (a cylindrical organelle within muscle cells) and consequently of contraction in muscles. α-actinin filaments are an important component of the sarcomere, making α-actinin an important structural component of muscle (Snow & Valberg 1994). In humans two genes encode the skeletal α-actinins; *ACTN2* is expressed in all muscle fiber types while *ACTN3* is confined to type II fast twitch fibers. In the human population up to 18% of individuals are homozygous for a premature stop codon (577XX) in *ACTN3.* As a consequence, these

individuals lack the ACTN3 protein (Yang *et al.* 2003), however a lack of the protein has no apparent detrimental effect.

In a study examining the frequency of the allele among a population of Australian athletes it was observed that sprinters showed a significantly higher frequency of the normal gene, while endurance athletes showed a higher frequency for the 577XX genotype (Yang *et al.* 2003). Similarly, Niemi and Majamaa (2005) looked at élite endurance and sprint athletes and found the 577XX genotype underrepresented in sprinters and absent in élite sprinters. However there was no apparent correlation between *ACTN3* genotype and endurance performance. Moran *et al.* (2007) found a negative association between the 577XX genotype and performance in a 40m sprint in 992 Greek adolescents. Finally, Druzhevskaya *et al.* (2008) also found the 577XX genotype underrepresented in power orientated Russian athletes. In addition to human studies, MacArthur *et al.* (2007) showed *ACTN3* knockout mice displayed a shift in muscle metabolism towards greater reliance on aerobic pathways. Skeletal muscle in the *ACTN3 -/-* mice showed increased activity for NADH-tetrazolium reductase (NADH-TR) and succinate dehydrogenase (SDH), both markers of aerobic metabolism. Additionally mitochondria density increased in the skeletal muscle of *ACTN3 -/-* mice. In the same study analysis of the genomic region surrounding the 577X allele in humans included in the HapMap project showed evidence of recent positive selection in individuals of European and Asian descent. As a consequence it is plausible that the 577XX genotype may have been selected for due to a positive effect on endurance capacity in carriers.

Ciliary Neurotrophic Factor (CNTF)

CNTF is a polypeptide protein hormone originally named for its role in promoting survival of chick ciliary neurons (DeChiara *et al.* 1995). Subsequent research showed the CNTF protein has a role in the survival of a number of neuronal cell types, with most research in the gene focusing on its role in the maintenance of motor neurons (DeChiara *et al.* 1995). The CNTF receptor is composed of three subunits, one of these subunits, the CNTF receptor-α is abundantly expressed in skeletal muscles suggesting a role for the CNTF protein in muscle function (Roth *et al.* 2001). Research in rats has shown that levels of the protein are associated with muscular fiber area and strength (Guillet *et al.* 1999). Within the human population there is a *CNTF* gene variant (known as the A allele) that produces a nonfunctional protein due to the alteration of a splice site. Surprisingly given, its influence on a number of neuronal cell types, individuals homozygous for the A/A allele do not demonstrate increased rates of neuromuscular disease (Takahashi *et al.* 1994). As a consequence it appears that the CNTF protein is not vital for development, but may play a role in responding to injury (DeChiara *et al.* 1995). Roth *et al.* (2001) examined the effect of the *CNTF* G/A allele and muscular strength in 494 healthy men and women. Subjects exhibiting the G/A genotype possessed significantly greater muscle quality and muscular strength at fast contraction speeds when compared to individuals homozygous for either allele. The A/A individuals showed the lowest muscular strength and contraction speed (Roth *et al.* 2001).

Uncoupling Protein 2 and 3 (UCP2/3)

Uncoupling proteins (UCP) are a closely related group of proteins that are thought to be involved in thermogenesis, fatty acid regulation and are also implicated in obesity (Buemann *et al.* 2001). *UCP2* is expressed in a variety of tissues including adipose tissue and skeletal muscle, while *UCP3* is primarily expressed in skeletal muscle (Lanouette *et al.* 2002). Astrup *et al.* (1999) investigated 24 hour energy expenditure and substrate oxidation in 60 healthy subjects. The authors reported subjects with the val/val-55 *UCP2* genotype had greater metabolic activity when compared to subjects with the ala/ala-55 *UCP2* genotype. Buemann *et al.* (2001) carried out a sub-maximal exercise test comparing 8 subjects of the val/val genotype and 8 subjects of the ala/ala genotype and found that exercise efficiency at a range of work levels was significantly higher in the val/val subjects. The authors speculated that different exercise efficiencies observed between the two genotypes are the result of linkage disequilibrium with unidentified polymorphisms in a neighboring region. The val/ala-55 domain of the protein does not have a known function and the adjacent *UCP3* gene is considered as a likely location for the polymorphism. Interestingly, polymorphisms in *UCP3* have been implicated in body composition changes (body mass index, % fat, etc) in response to regular exercise (Lanouette *et al.* 2002).

α2A-adrenic receptor (ADRA2A)

α2A-adrenic receptors are found throughout the peripheral and central nervous system. These receptors play a role in the regulation of adipose tissue lipolysis, an

important contributor to the energy demands of endurance exercise. Wolfarth *et al.* (2000) reported a weak association between a restriction site in the *ADRA2A* gene recognized by the DraI enzyme and élite endurance athletes when compared to sedentary controls.

β2-adrenergic receptor (ADRB2)

*β*2-adrenergic receptors are involved in the regulation of glycogenlysis and gluconeogenesis in the liver, relaxation of smooth muscle and cell metabolism in skeletal muscle. Moore *et al.* (2001) looked at a polymorphism in codon 27 (Glu27Gln) of the gene in 63 postmenopausal women. Individuals homozygous for a Glu27 variant had a significantly lower VO_2max when compared to heterozygotes and Gln27 homozygotes.

β1-adrenergic receptor (ADRB1)

The *β1*-adrenergic receptor is a G-protein coupled receptor involved in the mediation of the hormones epinephrine and the neurotransmitter norepinephrine. Polymorphisms in the *ADRB1* gene have been implicated in regulating heart rate and obesity (Dionne *et al.* 2002). Wagoner *et al.* (2002) examined the effect of two common polymorphisms (codon 49 Ser/Gly and codon 389 Gly/Arg) in the β1-adrenergic receptor gene in 263 patients with idiopathic or ischemic cardiomyopathy. Individuals homozygous for the Gly389 allele showed significantly lower peak VO_2max compared with Arg389 homozygotes (14.5 ± 0.6 vs. 17.7 ± 0.4 ml/kg/min), heterozygotes showed a VO_2max between the two homozygotes. When the Ser/Gly49 polymorphism

was also considered it was found that individuals homozygous for Ser49 and Gly 389 had the lowest peak VO_2max (14.4 ± 0.5 ml/kg/min), while individuals homozygous for Gly49 and Arg389 had the highest peak VO_2max (18.2 ± 0.8 ml/kg/min).

G protein ß3 (GNB3)

G proteins are heterotrimers composed of α, β and γ subunits, and are involved in signal transduction. Polymorphisms in the *GNB3* subunit have been associated with hypertension and obesity (Downes & Gautam 1999). Rankinen *et al.* (2002c) investigated the impact of a splice variant created by a C to T transition in exon 10 of the *GNB3* gene on hemodynamic and body composition phenotypes before and after 20 weeks of training. They found that individuals homozygous for the C/C allele showed a significant reduction in heart rate during submaximal exercise following training (an indicator of improved aerobic capacity). Interestingly the greatest effect was observed in blacks highlighting the importance of taking the ethnic background of the subjects into account.

Peroxisome proliferators-activated receptor α (PPARα)

PPARα is a transcription factor that regulates genes involved in inflammation, fatty acid uptake and oxidation. Jamshidi *et al.* (2002) looked at 144 British Army recruits participating in a 10-week physical training program, it was observed that subjects homozygous for the C allele of a G/C polymorphism in intron 7 had a three fold greater increase in left ventricular mass when compared to individual with the GG

genotype. The mechanism of action for this allele remains unclear as the change in nucleotide does not alter the amino acid inserted. The authors speculate that this allele is in linkage disequilibrium with some unidentified polymorphism in the regulatory regions of the *PPARα* gene that is affecting transcriptional activity.

Vitamin D receptor (VDR)

VDR is involved in regulating bone homeostasis through the vitamin D endocrine system (Rabon-Stith *et al.* 2005). Grundberg *et al.* (2004) investigated the relationship between the size of a poly A repeat in the 3' UTR and a BsmI restriction site with muscle strength and body mass index (BMI). The shorter allele of the repeat was associated with higher hamstring strength and BMI. Lorentzon *et al.* (2001) examined several polymorphisms in *VDR* and their effect on bone mineral density in 99 Caucasian girls. They found that individuals heterozygous for an ApaI restriction site within the gene had increased bone mineral density. The authors also postulated that levels of physical activity might interact with these polymorphisms leading to alterations in bone mineral density. More recently Rabon-Stith *et al.* (2005) found an association between a FokI restriction site in exon 2 and bone mineral density response to strength training.

Myostatin (MSTN) and the myostatin pathway

Myostatin (also known as growth differentiation factor 8, *GDF8*) is a member of the transforming growth factor-β super family and a potent negative regulator of skeletal muscle growth. Myostatin is highly conserved in mammals and mutations in the gene

cause dramatic phenotypes in a number of species. Mice lacking the myostatin gene show muscle weight twice that of wild-type animals (McPherron *et al.* 1997). In cattle some breeds such as the Belgian Blue show a "double muscling" phenotype that has been attributed to a 11bp deletion in the third exon of the gene (Grobet *et al.* 1997; McPherron & Lee 1997). In the dog, the whippet is a breed used for racing, with a slight build resembling a small greyhound. In contrast to the normal phenotype, "bully" whippets are heavily muscled with well-developed leg and neck muscles. The animals do not experience any apparent health problems apart from muscle cramping but are often euthanized as they do not conform to the breed standard (Mosher *et al.* 2007). It was recently reported that "bully" whippets are homozygous for a 2bp deletion in the third exon of the myostatin gene, creating a premature stop codon. Additionally animals heterozygous for this deletion have increased muscle mass and superior racing performance (Mosher *et al.* 2007).

In humans there has been one report of a child homozygous for a loss of function mutation in the myostatin gene. As observed in other mammals the child displayed muscle hypertrophy, but was otherwise normal (Schuelke *et al.* 2004). Interestingly the mother was a former professional athlete and a number of family members were unusually strong. Huygens *et al.* (2004), using 329 young Caucasian male sibs carried out linkage analysis examining estimated muscle cross-sectional area and 11 polymorphic markers adjacent to 10 candidate genes in the myostatin pathway. They found suggestive linkage with the markers adjacent to the *GDF8* (HSA2q32), *CDKN1A* (HSA6p21) and *MYOD1* (HSA11p15) genes. A follow up study was carried out by the

same group using a different population of 367 male siblings from Belgium (Huygens *et al.* 2005). In total 9 candidate genes from the myostatin pathway were examined using 29 microsatellites; in this instance linkage was observed between markers flanking *CDK2* (12q13), *RB1* (13q14) and *IGF1* (12q23) and knee muscle strength. The lack of significant association between markers adjacent to the myostatin gene indicates that it plays a minor role in explaining the variation in muscle strength in humans (Huygens *et al.* 2005). A final point worth mentioning is the lack of overlap in genes showing significant linkage for the two studies. This highlights the difficulty of reproducing signals of linkage for complex phenotypes, particularly when different methods of measurement are used (muscle cross-sectional area vs. knee muscle strength) and different populations examined.

Applicability to the horse

There is no guarantee that genes affecting athletic performance in humans will play a similar role in the horse. As has been mentioned previously, given the apparent lack of heritability for raw speed and static race times in longer races, genes affecting aerobic capacity (the major supplier of energy over these distances) may have a minor role in explaining race performance variation in the Thoroughbred. However the continued improvement in race times and modest heritability for speed over shorter distances suggests that genes affecting sprinting ability (e.g. genes affecting anaerobic capacity) are under selection in the Thoroughbred population. Rational arguments can be made as to why a particular candidate gene has the potential to affect performance in the

Thoroughbred. However, given the availability of polymorphic markers adjacent to the majority of these genes it is prudent to examine as many as possible for evidence of linkage or association with performance. While many of these genes are unlikely to show similar effects in horse and human the example of the myostatin gene vividly demonstrates how different species can produce remarkably similar phenotypes when the same gene is mutated.

Current commercial applications

One of the goals of studying the genetic component of athletic performance in humans and horses is that by identifying the genes associated with superior athletic performance it might be possible to identify individuals endowed with a genotype conducive to superior athletic performance. Given the complexity of the genetics of athletic performance, such a goal is not yet feasible. However, it is likely that information on an individual's genotype for a number of genes could be used in the near future to determine whether they are more suited to events that rely heavily on aerobic or anaerobic respiration. As has been outlined above, alleles located near or within a number of genes have been associated with superior performance in endurance or strength orientated events. Some companies are already offering genetic tests based on some of these polymorphisms. In humans, *ACTN3* testing is being offered (http://www.gtg.-com.au/humandnatesting/index.asp?menuid=070.110.020) while in the horse tests are being offered to identify haplotypes associated with superior performance in longer races (http://Thoroughbredgenetics.com/standard_services.htm). Offering such

tests to the general public appears premature because these single gene tests only explain a small fraction of the individuals athletic potential. Given the publics often misguided belief in genetic determinism results from such tests may lead individuals or trainers to discount events or training regimes that when all factors are considered are actually suitable for the athlete (human or equine). Therefore, while such genetic information could potentially be useful for athletes and trainers, a greater number of alleles need to be considered to reach a level where such tests would be meaningful for the individual athlete.

Genome wide linkage scans in humans

In addition to the numerous candidate gene studies, a small number of genome scans have been carried out in humans to identify regions of the genome linked to performance related phenotypes. Bouchard *et al.* (2000) carried out a genome scan for VO_2max before and after a 20 week endurance training program in 481 individuals from the HERITAGE family study using 289 microsatellites. A number of suggestive linkages were identified for both VO_2max pre and post training. Interestingly, the region of the genome harboring the *ACE* I/D polymorphism failed to show any linkage to VO_2max. However the region containing the *CKMM* gene did show suggestive linkage with VO_2max response to 20 weeks of training. In a follow up to Bouchard *et al.* (2000), Rico-Sanz *et al.* (2003) again looked at VO_2max and maximal power output but added an additional 220 markers to the genome scanning panel for a total of 509 microsatellite markers. The sample size was 351 and 102 sibling pairs for whites and blacks

respectively for the sedentary state measurements and 329 and 90 respectively, for the training response. The most significant linkage for baseline VO_2max was seen at HSA11p15 ($P < 0.023$) in whites, in blacks suggestive linkage ($0.01 > P > 0.0023$) was seen at HSA1p31, HSA7q32 and HSA7q36. As regards post training VO_2max response to training, blacks showed promising linkage at HSA1p31, while whites showed suggestive linkage at HSA4q27, 7q34 and 13q12. Finally a region at HSA5p23 showed promising linkage to maximal power output post training in whites, while blacks showed promising linkage at HSA1q22 and 13q11.

Rankinen *et al.* (2002a) carried out a genome scan for exercise stroke volume and cardiac output before and after a 20 week endurance training program in the HERITAGE family study. It was interesting to note that some regions of linkage were not the same in blacks and whites. Whites showed promising linkage at HSA14q31.1 and 10p11.2 for stroke volume at baseline and post training, while blacks showed linkage for baseline stroke volume at HSA1p21.3, 3q13.2 and 12p13.2.

The findings from this and the previous studies highlight the importance of taking into account the genetic background of the sample population given the different linkage signals seen in blacks and whites. Additionally *CKMM* was highlighted as a candidate gene in the Bouchard *et al.(2000)* paper because of suggestive linkage to the gene and a previous study (Rivera *et al.* 1999) linking it to VO_2max. The lack of linkage between VO_2max and the region of the genome harboring *CKMM* in the Rico-Sanz *et al.* (2003) study again highlights the difficulty of replicating results in complex phenotypes.

Athletic performance in model organisms

The rat and athletic performance related phenotypes

Epidemiology studies in humans have made it increasingly evident that many exercise related phenotypes such as maximal aerobic capacity and exercise endurance are linked with positive health outcomes in a number of chronic diseases, such as coronary heart disease and diabetes (Booth *et al.* 2000; Lightfoot 2006). However given, the complexity of the phenotypes and inherent limitations of human subjects (inability to selectively breed for a phenotype, control diet and environment closely, etc) efforts have been made to produce rodent models to assist in the study of exercise related phenotypes. Initial investigations to determine the genetic component of endurance phenotypes in rats used 11 inbred strains (6 males and 6 females from each strain) that were run on a treadmill to estimate aerobic performance. The strains showed a range of aerobic capacities, as regards the extremes, it was found that COP rats covered 298 ± 30 m when run to exhaustion, while DA rats covered 840 ± 64 m (Barbato *et al.* 1998).

In an attempt to create a strain of rat with both high and low aerobic capacity, Koch & Britton (2001) began with a genetically heterogeneous rat population of 168 individuals (N:NIH stock). Overall the founding population covered 355 ± 11 m on a treadmill before exhaustion. Following six generations of divergent artificial selection the high capacity line could cover 839 ± 21 m while the low line could cover 310 ± 8 m. Comparison between the lines revealed the differences in aerobic capacity were mainly attributable to changes in peripheral skeletal muscle, showing increased capillary density and increased citrate synthase and β-hydroxyacyl-CoA dehydrogenase

concentration in the high aerobic capacity rats (Henderson *et al.* 2002; Howlett *et al.* 2003). After 15 generations of selection continued divergence was observed between the strains, with improvement in VO_2max in the high capacity strain and reduction in VO_2max in the low capacity strain. The divergence was attributable to an increased ability to deliver O_2 to the exercising muscles in the high capacity strain and decline in O_2 delivery in the low capacity strain (Gonzalez *et al.* 2006).

Ways *et al.* (2002) performed a genome scan for loci associated with aerobic running capacity in rats. In the study three significant quantitative trait loci were identified, with the strongest linkage observed on chromosome 16 (RNO16). The authors point to a number of candidate genes within the region that could contribute to aerobic performance, including lipoprotein lipase (*LPL*) enzyme, the β-3 adrenergic receptor (*ADRB3*) and the neuropeptide Y5 receptor (*Npy5r*). A tentative linkage was also located on the p-terminus of chromosome 3 (RNO3). A subsequent study created congenic strains by introducing RNO16 and a portion of RNO3 from a high performance strain of rat into a low performance stain. Recipients of the RNO16 chromosome showed increased aerobic running capacity while recipients of the portion of RNO3 showed increased but not statistically significant aerobic running capacity (Ways *et al.* 2007).

The mouse and athletic performance related phenotypes

Not to be outdone by its larger rodent cousin the mouse has also been recruited in the study of performance related phenotypes. Lightfoot *et al.* (2001) examined ten inbred strains of mouse to estimate aerobic capacity. Strains were not selected based on

exercise related phenotypes but based on their availability and the widespread use of the strain in the genetics literature. A significant difference in maximal duration of exercise accomplished was observed between strains, with 58% to 73% of the variability observed in aerobic capacity attributable to the genetic background of the animal. Lightfoot *et al.* (2007) carried out a genome scan for QTLs associated with maximal endurance in intercrosses between two strains of mice previously selectively bred to have either high or low endurance capacity. Significant linkage was observed on the X chromosome and suggestive linkage observed on chromosome 8.

Lightfoot *et al.* (2008), using an interval mapping approach, identified a number of QTLs affecting distance, duration and speed voluntarily run by the mice. The observed QTLs explained 11% to 34% of the variation seen for these traits. These QTLs were further investigated to identify evidence of epistatic interactions. The authors found evidence for a number of interactions between the QTLs and calculated that these interactions accounted for 26% of the total variation among the 3 traits examined (Leamy *et al.* 2008).

In addition to attempting to identify genes associated with exercise related phenotypes in the mouse a number of groups have also sought to investigate the function of candidate genes by creating genetically modified strains, either by removing or altering the expression of the gene of interest. Wang *et al.* (2004) engendered a mouse that over-expressed an activated form of the peroxisome proliferator-activated receptor δ (*PPARGδ*) in skeletal muscle. The resultant mouse was capable of continuous running for up to twice the distance of its wild type littermates and showed an increased number

of type I muscle fibers in skeletal muscle. Hanson and Hakimi (2008) created a transgenic mouse with a cDNA for the PEPCK-C enzyme attached to a human α-skeletal actin gene promoter. The animals showed increased physical activity and longevity (living almost 2 years longer than controls), had elevated numbers of mitochondria in skeletal muscle, produced lower concentrations of lactate during exercise and showed a remarkable ability to exercise at high intensity for extended periods of time.

Mapping QTLs in livestock species and potential difficulties

Athletic performance, like many traits of economic importance in livestock species, is a complex quantitative trait influenced by a number of environmental and genetic factors. Originally it was assumed that quantitative traits were controlled by a large number (perhaps hundreds) of genes each exerting a small effect on the observed phenotype and as a consequence the prospects of identifying the molecular underpinnings of QTLs appeared poor (Flint & Mott 2001). Research in inbred model organisms provided some hope, showing that in many cases quantitative traits were influenced by a relatively small number of QTLs with large effect. However, attempts to replicate these findings in more outbred species have often been disappointing (Flint & Mott 2001).

Over the last 25 years literally thousands of Mendelian traits have been dissected at the molecular level in humans, however in comparison, relatively few genes contributing to complex traits have been identified (Glazier *et al.* 2002). In animals the number of genes identified that underlie Mendelian traits is more modest. The Online

Mendelian Inheritance in Animals (OMIA) database includes 135 animal species and lists 241 phenotypes characterized at the molecular level (http://omia.angis.org.au/). As regards more complex quantitative traits there has been no shortage of mapped QTLs in domestic species. In the pig over 1,800 QTL have been identified (http://www.animalgenome.org/cgi-bin/QTLdb/SS/summary) but only a very small portion (<10) of these QTLs have been investigated to the point where the underlying molecular cause (the quantitative trait nucleotide, QTN) has been identified (Rothschild *et al.* 2007). In the cow a similar pattern is observed with over 1100 QTLs identified (http://www.animalgenome.org/QTLdb/cattle.html). However, again only a handful of causative mutations (<5) have been identified at the molecular level (Womack 2005; Ron & Weller 2007).

Difficulties associated with mapping QTLs and QTNs

There are a number of factors that complicate the task of mapping QTLs and the especially enigmatic QTN. Firstly the ability to detect a QTL is primarily based on the size of the population the QTL is segregating in and the proportion of the phenotype that is explained by the QTL (Andersson 2001). As many QTLs exert a modest influence on the phenotype, the population size necessary to ensure they are identified can be substantial (generally over 1000); (Darvasi & Pisanté-Shalom 2002). However, provided a sufficient sample size is available, sufficient marker density achieved across the genome and the phenotype correctly defined, identifying QTLs for a trait of interest has become relatively routine as is highlighted by the large number of QTLs identified in the

pig and cattle. Nevertheless, going from a multi centimorgan linkage peak to the QTN remains a major stumbling block.

One of the features of QTLs contributing to this problem is the lack of a clear relationship between genotype and phenotype. Each QTL is generally only responsible for a fraction of the observed phenotypic variability. As a consequence it is difficult to fine map and identify the actual QTN among the numerous polymorphisms under a linkage peak (Andersson & Georges 2004). A further complicating factor is that many of the mutations underlying QTLs may be in regulatory regions that have not been fully annotated. Due to this it can be difficult to differentiate between potentially important polymorphisms and neutral variation in the regions between the more fully annotated coding regions (Andersson & Georges 2004). Finally some QTLs initially identified with a major effect on the phenotype can actually be a number of QTLs of small effect in loose linkage (Andersson & Georges 2004).

Epistatic interactions and imprinting

In addition to the difficulties mentioned above, the parent from which a QTL is inherited and the genotypes of other QTLs in the individual can influence the effect a particular QTL has on the phenotype. In general, QTL mapping experiments make no attempt to detect epistatic interactions and as a consequence may be ignoring a substantial component of the observed phenotypic variation (Carlborg & Haley 2004). Epistatic interactions have been identified in the chicken (Carlborg *et al.* 2003), mouse (Leamy *et al.* 2008) and other model organisms (Carlborg & Haley 2004). However,

while undoubtedly an important factor in many complex traits, attempting to take epistatic interactions into account increases the complexity of data analysis significantly and requires larger sample sizes, thus making it difficult to test for epitasis at present (Andersson & Georges 2004; Carlborg & Haley 2004; Georges 2007). An additional complication not always considered is the role imprinting can play in QTL expression. Genomic imprinting is parent of origin specific silencing of a gene (Morison *et al.* 2005). Therefore imprinted QTLs show different effects depending on the contributing parent. The number of imprinted genes in the genome is believed to be relatively modest, 56 human genes are currently listed as imprinted in the Geneimprint database (www.geneimprint.com) and this figure increases to just over 180 when predicted and genes with conflicting/provisional data are included. Imprinted genes are often involved in growth and development and as a result may play a role in many economically important traits in livestock species. Work in pigs has already identified a number of imprinted QTLs (de Koning *et al.* 2000; Rothschild *et al.* 2007) highlighting the need to take this phenomenon into consideration.

Copy number variation

One of the surprising findings to emerge over the last number of years is the remarkable amount of submicroscopic structural variation observed in the human genome. Recent work has shown that these copy number variations (CNV) are an important source of inter-individual genetic variation (Feuk *et al.* 2006). While much of this variation is benign it is becoming increasingly apparent that structural variations

also play a role in complex disorders such as autism and schizophrenia (Abrahams & Geschwind 2008; Stefansson *et al.* 2008). Preliminary results from cattle (Liu *et al.* 2007) and pigs (Fadista *et al.* 2008) show that a significant number of CNVs are present in these species and as a consequence are expected to be present in the horse. It is safe to assume that this variability at the genome level is likely to contribute to the phenotypic variability in domestic species.

Unfortunately at the present time the genomic sequences of many of the domestic species are not of the same quality as the human assembly. Due to this, identification and cataloging of CNV in domestic species may be more difficult than in humans. However at the same time work in humans is producing new single nucleotide polymorphism (SNP) chip designs and computational approaches that are capable of identifying CNV while performing genotyping experiments (McCarroll *et al.* 2008). Additionally, the increasing availability of next generation sequencing technologies (Mardis 2008) will hopefully improve the quality of genomic sequences of domestic species and assist in the identification and investigation of CNVs. Hence, future QTL mapping experiments should also be capable of considering the potential contribution of CNV to phenotypic variability.

Equine QTLs and genomics

In the horse a modest number of Mendelian traits have been characterized at the molecular level, with 15 phenotypes listed in the OMIA, the majority of which are coat color variants (http://omia.angis.org.au/). As regards QTLs, relatively few studies have

been reported in the horse when compared to other domestic species. Some of this can be attributed to an initial lack of genomic resources such as linkage and physical maps of the equine genome. However the major contributing factor is probably inadequate population sizes and the impracticality of setting up experimental populations in order to examine the segregation of QTLs of interest, as has been done in many of the other domestic species.

The few QTL studies that have been carried out in the horse have concentrated on diseases of the locomotory system. Wittwer *et al.* (2007) searched for QTLs associated with osteochondrosis (OC) and palmar/plantar osseous fragments (POF) in 219 South German Coldblood horses using 250 microsatellites and found 17 suggestive regions of linkage with one or both of the phenotypes. In a follow up study by the same group, one of the regions on chromosome 4 that showed linkage with osteochondrosis disease (OCD) was investigated in greater detail. Using a total of 22 SNPs and one microsatellite in the region of linkage it was found that three SNPs in the *AQAH* gene were significantly associated with OCD in the fetlock (Wittwer *et al.* 2008). Dierks *et al.* (2007) also carried out a genome scan for osteochondrosis using 260 microsatellites in 211 Hanoverian Warmblood horses. The authors reported genome wide significant QTLs on equine chromosomes 2, 4, 5 and 16 while chromosome wide significant QTLs were observed on equine chromosomes 2, 3, 4, 5, 15, 16, 19 and 21. There was very little overlap between the QTLs identified in the genome scans for osteochondrosis in Hanoverian Warmbloods and South German Coldblood horses. It may be that many of these putative QTLs were cases of spurious linkage or examples of population specific

QTLs. Diesterbeck *et al.* (2007) carried out a genome scan looking at pathological changes in the navicular bone in 192 animals using 214 microsatellites. Chromosome wide significant QTLs were located on equine chromosome 2, 3, 4, 10 and 26, while genome wide significant QTLs were seen on equine chromosomes 2 and 10.

The lack of concordance between QTLs for osteochondrosis in Hanoverian Warmbloods and South German Coldblood reflects a pattern often seen in linkage studies investigating complex phenotypes in humans. Altmüller *et al.* (2001) created a database of 101 studies that examined 31 complex disease. They noted that the majority (66.3%) of the studies failed to show significant linkage using the criteria of Lander & Kruglyak (1995) (LOD score >3.6). Additionally, different studies examining the same disease generally failed to show similar patterns of linkage. The authors pointed to many of the problems associated with QTL mapping outlined previously in this chapter and noted that successful studies had large population sizes, well defined phenotype and were confined to a single ethnic group.

Status of equine genomics

Despite a late start, over the last number of years great strides have been made in our understanding of the structure of the equine genome (Chowdhary & Bailey 2003; Chowdhary & Raudsepp 2008). A number of complementary approaches have been employed producing low to medium resolution synteny (Shiue *et al.* 1999), linkage (Guerin *et al.* 1999; Swinburne *et al.* 2000a; Guérin *et al.* 2003; Penedo *et al.* 2005; Swinburne *et al.* 2006) cytogenetic (Raudsepp *et al.* 1996; Lear *et al.* 2001; Mariat *et al.*

2001; Milenkovic *et al.* 2002; Raudsepp *et al.* 2002) and radiation hybrid (Chowdhary *et al.* 2003; Perrocheau *et al.* 2006; Raudsepp *et al.* 2008a) maps of the equine genome. High-resolution radiation hybrid maps of individual equine chromosomes have also been produced (Lee *et al.* 2004; Raudsepp *et al.* 2004a; Brinkmeyer-Langford *et al.* 2005; Gustafson-Seabury *et al.* 2005; Wagner *et al.* 2006; Goh *et al.* 2007). These radiation hybrid maps also integrated available cytogenetic, linkage and comparative mapping data from other vertebrates to improve the accuracy of the map and identify evolutionally conserved regions of synteny and chromosome breakpoints. In addition, over 1800 microsatellite markers have been identified and mapped in the equine genome (http://dga.jouy.inra.fr/cgi-bin/lgbc/loci_micro.operl?BASE=horse).

The combined availability of linkage and radiation hybrid maps, with the wealth of comparative data from species with better characterized genomes has been a powerful tool for investigating traits of interest in the horse. Microsatellites from the linkage maps opened up the possibility of using linkage analysis to identify regions of the genome contributing to phenotypes of interest. RH maps showed what genes were in the vicinity of linkage peaks, while comparative data expanded the list of candidate genes in the region and provided functional information about the genes. The development of these tools has made it possible to map and identify the underlying molecular causes for a number of Mendelian traits in the horse (Chowdhary & Raudsepp 2008). The recent sequencing of the equine genome has supplanted these maps to a certain extent. However these maps will still play an important role in helping to correctly assemble the

equine genome sequence and provide information about regions not fully assembled in the current horse sequence map.

The three equine BAC libraries (INRA, TAMU, and CHORI-241) have played a vital role in the development of the aforementioned maps by providing the many FISH probes required to anchor linkage and RH groups to specific chromosomes. BAC based contigs have also been developed to cover important regions of the genome (Gustafson *et al.* 2003; Raudsepp *et al.* 2004b; Brinkmeyer-Langford *et al.* 2008; Raudsepp & Chowdhary 2008; Raudsepp *et al.* 2008b). Whole genome physical maps based on the CHORI-241 library have also been developed. Leeb *et al.* (2006) published a human-horse comparative map constructed using BAC end sequences from the CHORI-241 library. While a BAC map based on end sequences and fluorescent fingerprints of the CHORI-241 library is also under development and will be an important tool in determining the accuracy of the equine genomic sequence assembly (Woehlke *et al.* 2008).

Despite its diminutive size and small number of genes, the equine Y chromosome has been the subject of a concerted and long running effort to map and characterize the euchromatic part of the chromosome. Raudsepp *et al.* (2004b) produced an initial physical map covering about 20-25% of the euchromatic region. Continued work has greatly expanded the map with only about 2 Mb of the estimated 13 Mb of euchromatic DNA missing in the new map (Raudsepp *et al.* 2008b). Efforts are currently underway to characterize the gene content of the equine Y chromosome and apply these finding to stallion fertility (Chowdhary & Raudsepp 2008).

One of the most exciting developments in the field of horse genomics in recent years has been the sequencing of the equine genome. A female Thoroughbred (Twilight) provided the DNA and was sequenced to 6.8X coverage at the Broad Institute of MIT and Harvard. The initial assembly of the sequence was released in January 2007, with a second assembly made available in September 2007 (http://www.broad.mit.edu/ftp/pub/assemblies/mammals/horse/Equus2/). The sequencing of the equine genome will greatly simplify identifying regions of the genome harboring polymorphisms with an impact on health, reproduction and performance. Candidate genes and adjacent genomic markers suitable for association analysis can now be easily identified. While regions showing association or linkage to a trait of interest in whole genome approaches can be rapidly investigated for genes or regulatory regions likely to harbor important polymorphisms.

In addition to Twilight's DNA, 100,000 paired-end shotgun sequence reads from randomly distributed regions of the genome were generated from seven modern and ancient horse breeds (Arabian, Andalusian, Akhal-teke, Icelandic, Standardbred, Quarter Horse and Thoroughbred) for the purpose of SNP detection. The SNPs generated from this comparison in combination with those identified by comparing Twilight's reads resulted in the identification of 948,609 SNPs (http://www.broad.mit.edu/mammals/horse/). The identification of these SNPs has paved the way for the construction of a genotyping beadchip with over 54,000 SNPs spread throughout the equine genome (http://www.illumina.com/pages.ilmn?ID=285). This technology opens up the possibility of carrying out whole genome association mapping

experiments to identify regions of the genome associated with both Mendelian and complex traits. In addition to the SNP chip, new technologies are becoming available to investigate the expression of genes from the equine genome. Further equine cDNA and oligonucleotide arrays have also been developed, with the latest being a 22,000 element 70-mer oligoarray developed at Texas A&M (Chowdhary & Raudsepp 2008). This development makes it possible to track the changes in gene expression in tissues at different time points during growth and development, over the progression of a disease or in cases and controls. Consequently it will be possible to identify which genes play a central role in the phenotype under investigation. These new technologies open up a large number of new opportunities to investigate traits that affect equine health, reproduction and performance, creating exciting opportunities for the equine genetics community.

Study objectives

Given the important role athletic performance plays in the Thoroughbred, information on the genetic underpinnings of this trait would find application in breeding, identifying and training equine athletes. As has been pointed out previously, it has been shown that athletic ability/performance has a heritable component and research in other species has demonstrated the feasibility of identifying genes associated with the trait. In order to build on and expand our understanding of the athletic ability/performance in the horse it is necessary to begin in-depth investigations into the genetic component of this phenotype. The recent work in the field of equine genomics has provided the necessary

tools to identify genetic markers associated with athletic performance in the horse. The availability of a large number of published microsatellites and the draft sequence assembly of the equine genome allowed for the identification of polymorphic markers adjacent to candidate genes. These resources facilitated the mapping of candidate genes and allowed for the construction of a genome scanning panel of microsatellites to search for regions of the genome linked or associated with athletic performance in the Thoroughbred.

Research in humans and other livestock species has demonstrated the many difficulties associated with attempting to map complex traits like athletic performance. As a consequence the likelihood of identifying regions of the genome linked or associated with athletic performance in the Thoroughbred was modest. However by setting up clear and achievable objectives it was possible to maximize our chances of success given the recourses available, additionally the fulfillment of these objectives would have a utility beyond the primary goal of this project.

In summary the project had three main objectives:

a) Map genes associated with human exercise and athletic performance in the equine genome.

b) Develop a genome scanning panel of microsatellite markers covering the euchromatic portion of the equine genome, with an emphasis on markers adjacent to candidate performance genes.

c) Conduct a genome wide scan on a pedigreed population of Thoroughbreds to search for genomic regions linked or associated with superior athletic performance.

It was hoped that by carrying out this study new insights into the genetics of athletic performance would be gained. Additionally, mapping the candidate genes would add to the list of genes mapped in the equine genome, while the construction of the genome scanning panel would provide an important resource for mapping traits of economic importance in the Thoroughbred.

CHAPTER II

RADIATION HYBRID AND FLUORESCENT IN SITU HYBRID MAPPING OF 46 HUMAN GENES ASSOCIATED WITH ATHLETIC PERFORMANCE IN THE EQUINE GENOME

Introduction

Research in the field of human sports medicine has highlighted the significant role genetic factors play in athletic performance (MacArthur & North 2005; Brutsaert & Parra 2006; Rankinen *et al.* 2006). Initial studies highlighted the substantial influence genetics plays in traits such as VO_2max (heritability estimates range from 0.4 to 0.6) and anaerobic capacity (estimates range from 0.31 to 0.86) (Bouchard *et al.* 1992; Bouchard *et al.* 2000; Feitosa *et al.* 2002; Rupert 2003). Encouraged by these findings researchers began to search for the specific genes influencing athletic performance in humans. Two main approaches have been taken in this endeavor. Some have taken the whole genome approach and carried out genome scans looking for regions of the genome linked to specific phenotypes affecting athletic performance, such as maximal oxygen uptake (Bouchard *et al.* 2000) and exercise stroke volume (Rankinen *et al.* 2002a). However most investigators have taken the candidate gene approach. Candidate genes have been investigated because of their role in biological processes likely to influence athletic performance (e.g. *CKMM,* involved in metabolism in the muscle, *ACTN3* structural protein found in skeletal muscle). These genes are tested for association with athletic performance using polymorphic marker (SNPs, microsatellites etc) in or around the gene

of interest using a suitable sample population. These two approaches have generated a large list of genes associated with athletic performance over the last 10 years, a fact highlighted in the 2005 update of the human performance gene map that lists 165 genes and QTLs (Rankinen *et al.* 2006). Many of these genes represent excellent candidates for influencing athletic performance in the horse. The genes *ACE* and *ACTN3* have the potential to affect aerobic capacity, *CKMM* and *MSTN* may play a role in sprinting ability, while *VDR* could play a role in keeping the horse sound during the rigors of training and racing.

Research in the horse has also demonstrated that success on the track has a genetic component. Investigations into the heritability of race winnings and handicap ratings have shown consistent (although modest) heritabilities for both measurements of performance (Ricard *et al.* 2000). The last number of years have seen remarkable progress in the filed of equine genomics with the development of linkage and radiation hybrid maps and the recent sequencing of the equine genome (Chowdhary & Raudsepp 2008). These developments finally provide the tools necessary to utilize the wealth of information gathered in humans and rodent models to investigate athletic performance in the horse. In order to test these candidate genes for association/linkage with athletic performance polymorphic markers must be identified close to these genes.

Prior to the availability of the draft genome sequence of the horse, two approaches were practical for identifying such markers. Firstly, single nucleotide polymorphisms (SNPs) could be identified within individual genes (by sequencing gene specific PCR products in a number of individuals and searching for SNPs) and tested for

association with athletic performance. Secondly, previously published microsatellite markers could be identified adjacent to these genes and tested for linkage or association with athletic performance. The second option was chosen as the recent development of a number of equine genetic maps had produced a large number of potentially useful microsatellite markers (Penedo *et al.* 2005; Swinburne *et al.* 2006). However, at the initiation of this project many of the candidate genes had not been mapped in the equine genome and as a consequence it was not possible to identify adjacent microsatellites. In order to rectify this situation it was decided to localize all unmapped genes from the most recent update (when the project was initiated this was the 2004 update) of the "human gene map for performance and health related phenotypes" (Wolfarth *et al.* 2005). By doing this it would be possible to identify the nearest previously published microsatellite marker for each candidate gene. These microsatellite markers could then be genotyped in a suitable population and tested for linkage/association with superior athletic performance.

At the time this work was undertaken, our lab was working towards producing the second-generation radiation hybrid (RH) map of the equine genome. As a result, the candidate genes mapped by RH analysis and fluorescent in situ hybridization (FISH) were included in this map. The inclusion of these genes added to the RH portion of the map and made a particularly significant contribution to the cytogenetic portion of the integrated map eventually produced (Raudsepp *et al.* 2008a).

Materials and methods

Primer design

The human exonic sequence of each candidate gene was obtained from the Ensembl genome browser (http://www.ensembl.org/index.html) and entered into the BLASTN, NR and EST_OTHERS search engines (http://www.ncbi.nlm.nih.gov/BLAST) to obtain sequence for the orthologs in other mammals. When an expressed sequence tag (EST) was available for the equine ortholog of the gene, it was used for primer design utilizing the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). For the remainder of the genes, a multiple alignment method of primer design was used. Briefly, the sequence of the gene was obtained from as many mammals as possible (generally included human, rat, mouse, cow, pig, sheep, dog) and aligned using the Clustal program (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The aligned sequence was searched for regions that showed high conservation and were therefore likely to have the same sequence in the equine ortholog. Primers were designed within a single exon and with a PCR product of less than 1 kb. The primers were designed with two to three mismatches between the mouse/rat consensus sequences and the remainder of the species to produce equine specific amplification in the rodent DNA background of the radiation hybrid (RH) panel.

Primer optimization

Primer pairs were optimized on horse and hamster genomic DNA to ensure horse specific amplification. Each PCR reaction contained 50 ng of DNA as template, 0.2 mM dNTPs, 1X buffer (Sigma Aldrich, MO), 0.3 pmol of each primer 1.5, 2, 3, or 4 mM $MgCl_2$ and 0.25 U JumpStart RED*Taq* DNA polymerase (Sigma Aldrich, MO). Reactions were carried out with an initial 30 second denaturation at 94°C, 1 cycle at 94°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec. This was followed by 30 cycles at 94°C for 30 sec, annealing temperature ranging from 50°C to 66°C for 30 sec, and 72°C for 30 sec, with a final extension of 72°C for 5 min. All reactions were carried out using either a 96-well Touchgene Gradient thermal cycler (Techne Inc., NJ) or a DNA Engine Dyad (MJ Research Inc., MA). PCR amplification products were resolved on 2.0% agarose gels containing 0.25 ug/ml ethidium bromide. Primers were initially tested with an annealing temperature of 58°C and 1.5 mM $MgCl_2$. If a single amplified band was observed with horse DNA and no amplification was seen with the hamster DNA, no further optimization was carried out. However, if no amplification was observed, annealing temperature was decreased and $MgCl_2$ concentration was increased to reduce the stringency of the PCR reaction. If amplification was observed with hamster template DNA, annealing temperature was increased in order to increase the specificity of the reaction. Only primers that showed equine specific amplification were used for further analysis.

PCR product sequencing

To ensure that the correct gene was being amplified, all primers designed using the multiple alignment method were sequenced to confirm the origin of the PCR product. PCR reactions with a volume of 50ul were performed for each primer and excess nucleotides and primer removed using the QIAquick PCR Purification Kit (Qiagen, CA) following the manufactures instructions. Size and quality of the amplified DNA was tested by electrophoresis on a 2% agarose gel and further quantified with a SmartSpec Plus spectrophotometer (BioRad, CA) at 260nm wavelength.

Final concentration of DNA required in the 10ul sequencing reaction was determined using the following formula: (ng/ul) = 9 x fragment size (kb). (For example if a PCR product was 450bp the final concentration of DNA in the sequencing reaction is 9 x 0.450 = 4ng/ul, therefore the reaction required a total of 40ng template DNA). In addition to the appropriate volume of template DNA, the reaction contained 1.5ul of the forward or reverse primer (concentration 5uM), 0.5ul MasterAmp™ (Epicentre, WI) and 2ul of BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, CA), the volume of the reaction was brought up to 10 ul using ddH_2O The reactions were carried out with an initial denaturation at 94°C of 3 min and then cycled at 95°C 30 sec, 50°C 20 sec, 60°C 4 min for 35 cycles with a final extension of 60°C 10 min.

Unincorporated nucleotides were removed from the sequencing reactions using Sephadex G-50 columns (Biomax Inc., MD) following the manufacturer's instructions. Purified products were dried down and stored at -20 ^{o}C until loaded into an ABI3730 automated capillary DNA sequencer (Applied Biosystems, CA). The resultant sequence

of the PCR products were confirmed using BLAST (http://www.ncbi.nlm.nih.gov/BLAST) and BLAT (http://www.genome.ucsc.edu/cgi-bin/hgBlat) as described earlier (Chowdhary *et al.* 2003).

Radiation hybrid mapping

The 5000-rad horse x hamster radiation hybrid (RH) panel was genotyped in duplicate for each primer pair by PCR as described in Chowdhary *et al.* (2003). Genotypes were scored manually by first resolving the PCR products on a 2% agarose gel and inspecting each clone in the panel for the presence or absence of PCR amplification. Scoring data was initially analyzed by two-point analysis using RHMAPPER-1.22 at http://equine.cvm.tamu.edu/cgi-bin/ecarhmapper.cgi to identify the closest framework marker from the equine first generation whole genome radiation hybrid map as described in Chowdhary *et al.* (2003). This data was also included in the second generation radiation hybrid map of the equine genome (Raudsepp *et al.* 2008a). In the second generation map the analysis was carried out using rh_tsp_map (Agarwala *et al.* 2000; Schaffer *et al.* 2007) CONCORDE (Applegate 2006) and Qsopt (http://www.isye.gatech.edu/~wcook/qsopt) as described in Brinkmeyer-Langford *et al.* (2005) and Murphy *et al.* (2007). Briefly, markers were initially assigned to linkage groups using two-point analysis with a threshold LOD score of 7.6 to create an initial robust "MLE-consensus map" (Agarwala *et al.* 2000). This cutoff point ensured that few markers would be dropped from analysis with inter-chromosomal scores above threshold and allowed more linkage groups to be combined. When a marker did not have a LOD

score above 7.6 with another marker it was dropped from analysis. Markers that were dropped from the initial consensus map were placed in between consecutive markers when the best placement was at least 0.1 LOD units better than the next best placement. Centi-Ray (cR) distances between markers were assigned by solving instances of a restricted traveling salesman problem. The remaining markers were binned where they spanned at most three consensus intervals in the consensus map. Orientation and order of linkage groups were conformed using FISH, with previously published linkage maps providing additional conformation (Penedo *et al.* 2005; Swinburne *et al.* 2006).

Fluorescent in Situ Hybridization (FISH)

PCR primers for all the genes mapped by RH analysis were used to isolate bacterial artificial chromosome (BAC) clones containing the gene of interest from the CHORI-241 equine BAC library using available pools and super pools. Individual clones were initially cultured (cultures were grown overnight at 37 ^{o}C, with constant agitation) in 1ml of 2YT media containing 30ul/ml chloramphenicol. The culture was plated on LB agar plates containing 30ul/ml chloramphenicol to obtain single colonies. Presence of the gene of interest was conformed by PCR and the positive single colonies used to inoculate 100ml of 2YT media containing 30ul/ml chloramphenicol. DNA was extracted using the Qiagen midiprep kit (Qiagen, Chatsworth, CA) according to the manufactures instructions. For each BAC 1ug of the resultant DNA was labeled with either biotin or digoxigenin following the manufacturers instructions using the BioNick. Labeling System (Invitrogen, Carlsbad, CA) and DIG-Nick Translation Mix (Roche,

Indianapolis, IN). Labeled clones were hybridized in pairs (one labeled with biotin, the other with digoxigenin) to horse metaphase spreads to identify their chromosomal location. Clone isolation, labeling of DNA, in situ hybridization, signal detection and image analysis were performed as previously reported (Raudsepp *et al.* 1999; Chowdhary *et al.* 2003; Raudsepp *et al.* 2008).

Results

In total 55 genes from the "human gene map for performance and health related phenotypes" (Wolfarth *et al.* 2005) had not been mapped in the equine genome when this project was initiated (Chowdhary *et al.* 2003). Equine specific PCR primers were developed for 46 of these genes. These PCR primers were then used to genotype the 5000-rad RH panel for the purposes of RH mapping and to isolate BACs from the CHORI-241 BAC library for use in FISH. The gene symbol, chromosomal location as determined by FISH and closest mapped marker in the radiation hybrid map are detailed in Table 2-1, primer information is detailed in Table A-1. (Tables with the A prefix can be found in the appendix). The genes *ADRB1, ADRA2A*, *ANG*, *CETP, LIPG*, *APOE*, *ACADVL*, *IL15RA,* and *KCNQ1* could not be mapped despite at least four attempts to design equine specific primers for these genes.

Representative images of FISH hybridizations on equine metaphase spreads using some of the isolated BAC clones are shown in Figure 2-1. The location of the mapped genes in the equine genome and the candidate genes mapped by others is depicted in Figure 2-2. All cytogenetic localizations are in agreement with their location

determined by RH mapping. The BAC clone containing the gene *CASQ2* (CHORI clone 178D1) showed strong hybridization to all centromeres. Despite 40X excess of genomic blocking DNA, this nonspecific hybridization to the centromeres could not be prevented. RH mapping puts *CASQ2* close to the centromere of ECA1 suggesting that the BAC clone containing this gene also contains centromeric repeats, resulting in the hybridization to the centromeres.

When these genes were mapped all except one were localized to a region previously identified as homologous between the horse and the human genomes (Chowdhary *et al.* 2003; Perrocheau *et al.* 2006). The sole exception was *RYR2.* Initial two-point analysis with the genotyping data for *RYR2* showed no linkage to the framework markers in the first generation equine RH map. However, FISH mapping placed the gene at ECA1q13. When the genotyping data for *RYR2* was analyzed in conjunction with additional RH mapping data from Wagner *et al.* (2004c) using RHMAPPER-1.22 (Slonim *et al.* 1997), *RYR2* was placed 12.8 cR from *UMNe196*. The microsatellite *UMNe196* had been previously linked by RH mapping to the marker *RET,* this marker is located at ECA1q13 according to the first generation RH map of the horse (Chowdhary *et al.* 2003; Wagner *et al.* 2004c). The second generation RH map of the horse has since conformed this finding with additional genes from the same region of HAS1 mapping to ECA1 (Raudsepp *et al.* 2008a).

Table 2-1 Human candidate genes mapped in the equine genome in this study, published in Raudsepp *et al*. (2008a).

Gene Symbol	Chr No.	FISH	EquCab2	Map Position (cR)	Placement LOD	Assignment	Nearest marker by RH	Human	Human Mar. 2006
GPR10 (aka PRLHR)	1	1p15-16	13.7	40.79	1.51	frame	1CA30 (0 cR)	10q26.13	*120.3*
AGT	1	1q12-13	67.1	97.2	3.13	placed	ASB12ms (29.66 cR)	1q42-q43	228.9
RYR2	1	1q13	74.1	12.87	3.29	frame	UMNe196 (12.87 cR)	1q43	235.2
LIPC	1	1q21.3-22	132.9	363.33	1.9	placed	RORA (25 cR)	15q21-q23	56.5
CYP19A1	1	1q22	*139.1*	472.92	1.89	placed	SPPL2A (7.71 cR)	15q21.1	*49.2*
CPT2	2	2p17-18	6.1	---	[PPAP2B - OSBPL9]	binned	PPAP2B - OSBPL9	1p32.3	53.3
FGA	2	2q14.3-21.2	79.3	419.62	4.07	frame	FGG (0 cR)	4q31.3	155.7
FGB	2	2q14.3-21.2	79.3	---	[TKY645 - TKY850]	binned	TKY645 - TKY850	4q31.3	155.8
UCP1	2	2q22	90.9	629.65	2.9	frame	TKY335 (3.94 cR)	4q31.21	141.7
FABP2	2	2q22-24	108.2	68.38	4.14	frame	UMNe336 (5.92 cR)	4q28-q31	120.4
HP	3	3p13	21.8	392.54	1.83	frame	AHT022 (0 cR)	16q22.1	70.6
PPARGC1A	3	3q23	100.7	714.02	3.98	frame	DHX15 (14.67 cR)	4p15.2	23.4
PGAM2	4	4p12-13	14.8	413.34	0.95	frame	GCK (8.34 cR)	7p13	43.8
IGFBP1	4	4p12	16.2	435.96	6.03	frame	UMNe404 (5.27)	7p13-p12	45.8
PON1	4	4q13-14	38.6	261.44	1.45	frame	TKY210 (0 cR)	7q21.3	94.7
PON2	4	4q13-14	38.6	261.44	1.45	frame	TKY210 (0 cR)	7q21.3	94.8
NPY	4	4q21.1-21.3	55.7	551.86	1.84	frame	CBX3 (3.31 cR)	7p15.3	24.2
ATP1B1	5	5p17-18	5.5	106.53	3.51	frame	F5 (14.72 cR)	1q24.2	167.3
APOA2	5	5p13-14	35.8	487.98	2.01	placed	DEDD (9.5 cR)	1q23.3	159.45

Table 2-1 (continued)

Gene Symbol	Chr No.	FISH	EquCab2	Map Position (cR)	Placement LOD	Assignment	Nearest marker by RH	Human	Human Mar. 2006
ATP1A2	5	5p12-13	37.4	0	0.57	frame	COPA (15.41 cR)	1q23.2	156.8
S100A1	5	5p12	43.9	134.89	2.4	placed	S100A9 (3.57 cR)	1q21.3	151.86
CASQ2	5	cen	53.1	129.62	4.25	frame	COR023 (17 cR)	1p13.1	116
AMPD1	5	5q12-13	54	93.57	2.27	frame	TSHB (3 cR)	1p13.1	115
LEPR	5	5q17	95	350.75	5.89	frame	5STS03 (4.16 cR)	1p31	*65.65*
GNB3	6	6q13	34.3	211.62	3.33	frame	TKY377 (0 cR)	12p13.31	6.8
VDR	6	6q21.3-22	65.5	557.47	2.38	placed	COL2A1 (4.13 cR)	12q13.11	46.5
UCP2	7	7q14-15	69.8	101.3	6.19	frame	TKY690 (13.32 cR)	11q13.4	73.3
UCP3	7	7q14-15	69.8	101.3	6.19	frame	TKY690 (13.32 cR)	11q13.4	73.3
LDHA	7	7q16-18	86.6	4.79	0.8	placed	SOX6 (4.79 cR)	11p15.1	18.3
PNMT	11	11p12	22.7	1134.32	2.04	frame	CACNB1 (7.04 cR)	17q12	35
SGCA	11	11q13	25.8	---	[TKY304 - TKY988]	binned	TKY304 - TKY988	17q21.33	45.5
CNTF	12	12q13	18.9	0	5.55	frame	UMNe331 (22.99 cR)	11p12.1	58.1
SCGB1A1	12	12p13-14	22.2	113.66	2.7	frame	EEF1G (7.91 cR)	11q12.3	61.9
PYGM	12	12q14	24.9	160.65	6.74	frame	COR030 (2.6 cR)	11q13.1	64.2
ACTN3	12	12p13-14	26.5	204.32	2.13	frame	LRFN4 (18.67 cR)	11q13-q14	66.07
SERPINE1	13	13p14-15	*9*	155.62	4.78	frame	UMNe217 (0 cR)	7q21.3-q22	*100.55*
PPARG	16	16q12-13	4.8	176.53	1.6	frame	RAF1 (19.09 cR)	3p25.2	12.3
SGCG	17	17q13	*3.8*	48.46	0.24	placed	TKY373 (0 cR)	13q12.12	*22.65*
CASR	19	19q21-22	*37.1*	791.69	0.35	placed	TKY686 (0 cR)	3q21-q24	*123.3*

Table 2-1 (continued)

Gene Symbol	Chr No.	FISH	EquCab2	Map Position (cR)	Placement LOD	Assignment	Nearest marker by RH	Human	Human Mar. 2006
TNF	20	20q16-21.1	31.3	605.17	2.1	frame	UMNe214 (0 cR)	6p21.33	31.6
PLCG1	22	22q16-17	*30.9*	---	[SRC - SFRS6]	binned	[SRC - SFRS6]	20q12-q13.1	*39.19*
CNTFR	23	23q18-19	50.4	---	[NDUFB6 - UMNe560]	binned	NDUFB6 - UMNe560	9p13.3	34.5
BDKRB2	24	24q16.2-16.3	38.6	487.15	4.92	frame	SERPINA1 (15.63 cR)	14q32.2	95.7
ADRB3	27	27q12-13	7.6	0	0	frame	TKY764 (0 cR)	8p12	37.9
PPARA	28	28q18	42	420.03	3.51	frame	UMNe345 (0 cR)	22q13.31	44.9
GK	X	Xp21	*23.9*	--	[UMNe058 - DMD]	binned	UMNe058 - DMD	Xp21.3	*30.58*

EquCab2 Refers to megabase position in the Sep. 2007 *Equus caballus* draft assembly produced by The Broad Institute (http://www.broad.mit.edu/mammals/horse)

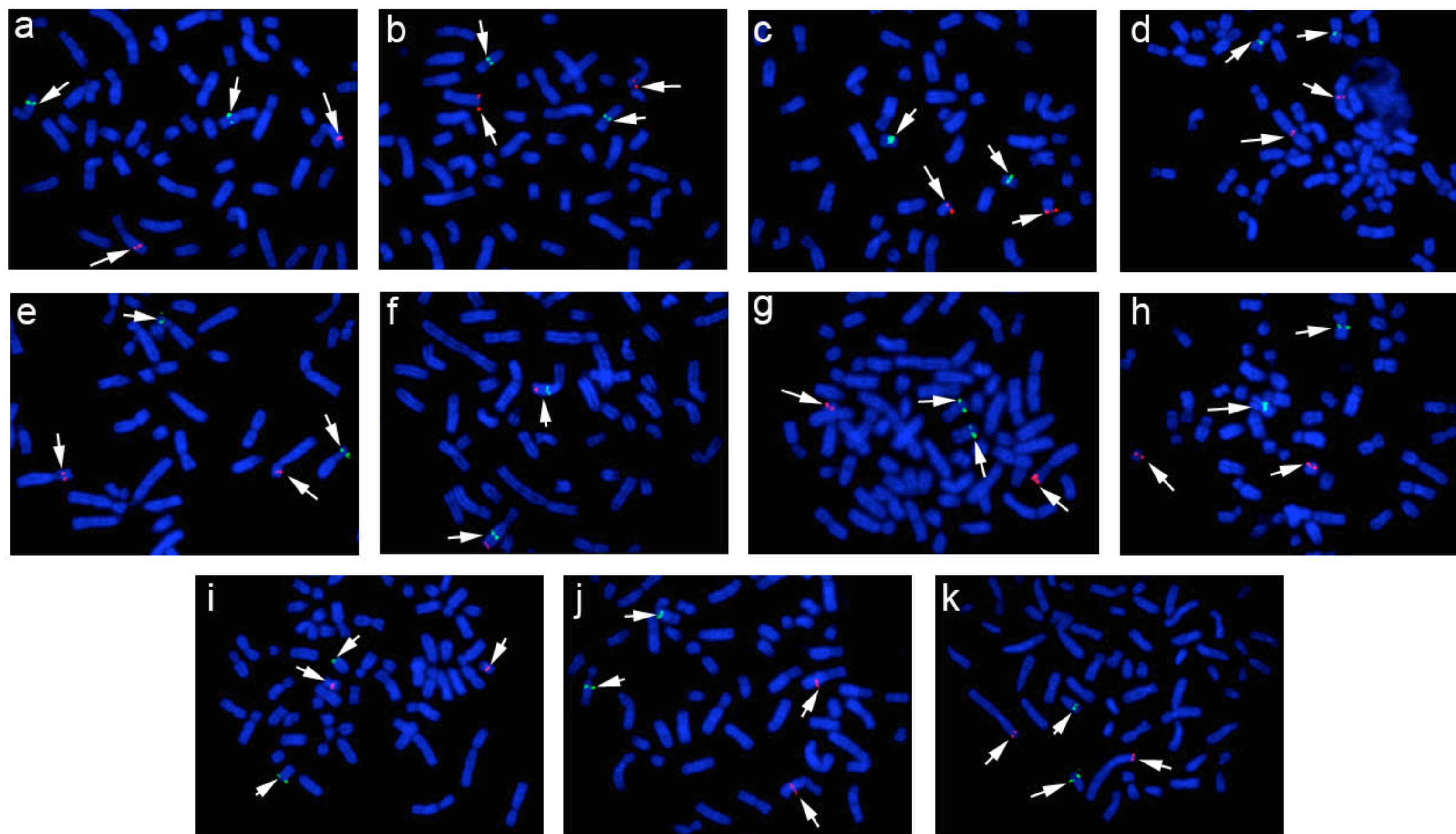

Figure 2-1 Representative images of double color FISH hybridizations on partial equine metaphase using BAC clones containing the following loci: a. UCP1 (red) + UCP3 (green), b. SUR (red) + TNF (green), c. SGCB1A1 (red) + SGCA (green), d. RYR2 (red) + S100A1 (green), e. HP (red) + TGFB1 (green), f. FABP2 (red) + FGA (green), g. CPT2 (red) + ENO3 (green), h. ADRB3 (red) + AMPD1 (green), i. ADRB3 (red), j. CNTF (red) + CNTFR (green), k. FGB (red) + GNB3 (green) and l. GPR10 (red) + PLCG1 (green)

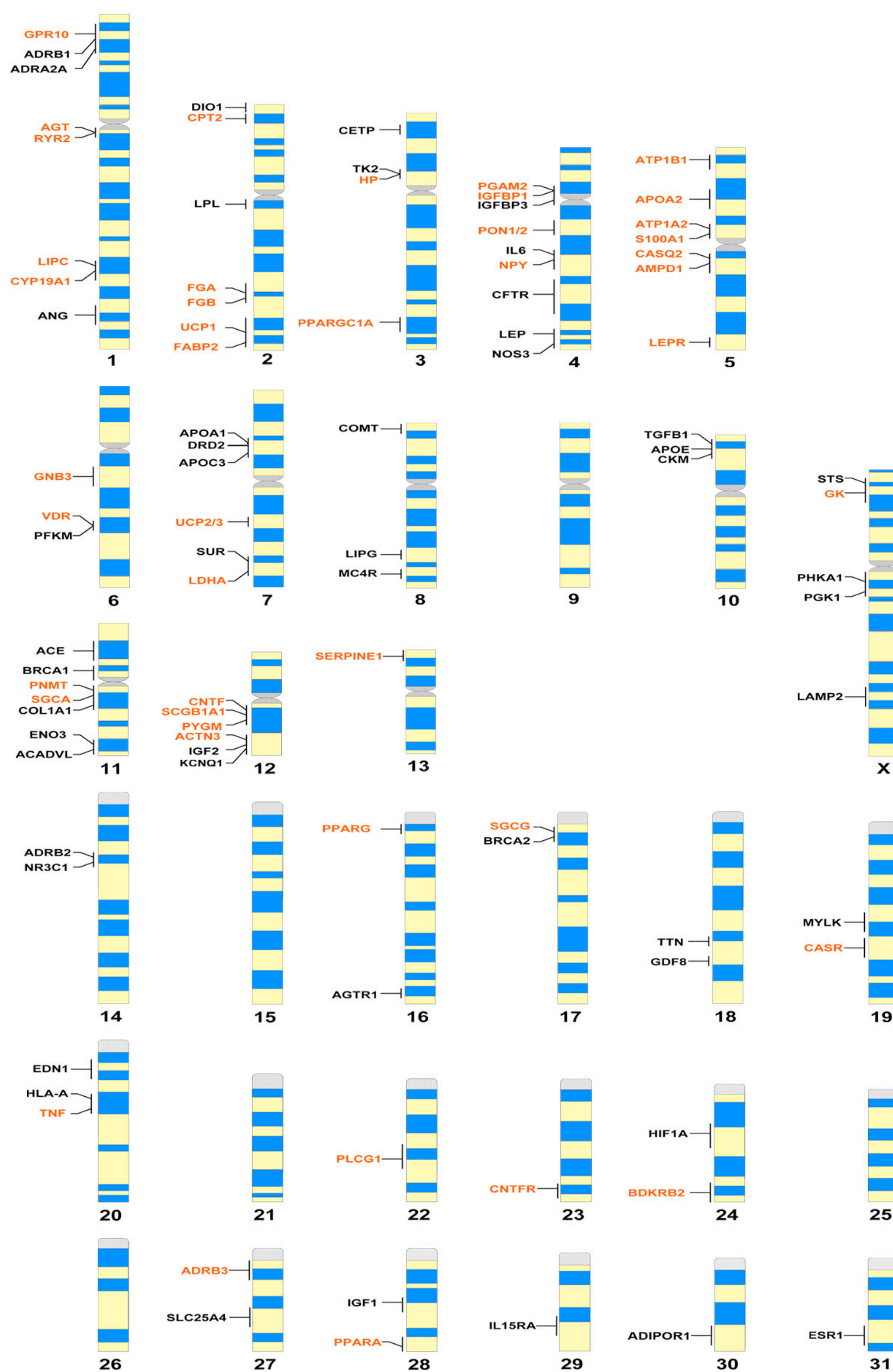

Figure 2-2 Location of the 95 human candidate genes (Rankinen *et al.* 2006) in the equine genome. Those in red (46) were mapped in this study (see Tables 2-1 & A-1), in black (49) mapped by others (see Table A-2).

Discussion

Candidate performance genes

The aim of this portion of the overall project was to map human nuclear genes associated with athletic performance in the equine genome. These candidate genes were identified in the 2004 update of "The human gene map for performance and health-related fitness phenotypes" (Wolfarth *et al.* 2005). Of the 85 nuclear genes identified in this paper 30 had been previously mapped in the equine genome (Chowdhary *et al.* 2003). These 85 candidate genes are likely to have similar biological roles in the horse and human, making them obvious candidates for influencing athletic performance in the horse. In order to test for linkage/association with performance we had to first identify polymorphic markers in or near to these candidate genes (Evans & Cardon 2004). As a consequence the primary rationale behind carrying out this work was to assist in identifying which of the over 1800 published equine microsatellites (http://dga.jouy.inra.fr/cgi-bin/lgbc/loci_micro.operl?BASE=horse) are located near these candidate genes. Once identified these markers could then be genotyped on a population of Thoroughbreds and tested for association/linkage with superior athletic performance.

A recent example of the utility of the candidate gene approach in dissecting the underlying genetics of a phenotype is provided by the Myostatin (*MSTN)* gene that plays a key role in muscle mass regulation in mice (McPherron *et al.* 1997), cattle (Grobet *et al.* 1997; McPherron & Lee 1997), humans (Schuelke *et al.* 2004) and sheep (Clop *et al.* 2006). Information from other species made *MSTN* an obvious candidate gene in whippets displaying the "bully" phenotype. As was pointed out in the initial chapter

whippets heterozygous for a mutation in this gene show increased muscle mass and superior racing performance (Mosher *et al.* 2007). Myostatin also has the potential to play a role in equine athletic performance. It is unlikely that a loss of function mutation similar to the one seen in the whippet is segregating in the Thoroughbred, given the absence of the double muscle phenotype generally associated with mutations in this gene. It is possible that polymorphisms in and around the gene are affecting expression or function of the gene in more subtle ways. These changes in expression may not be sufficient to cause the double muscle phenotype but they could still affect muscle mass and race performance.

Of the 95 genes depicted in Figure 2-2, the likes of *CKMM, ACTN3* and *MSTN* are obvious candidate genes for athletic performance, given their intimate role in metabolism and muscle growth. Other genes included in the map, at first sight do not appear to be especially strong candidates. One such gene is *VDR,* a gene involved in regulating bone homeostasis. Some authors have speculated since the Thoroughbred has been subject to selective breeding for performance over a large number of generations the relative importance of genes affecting speed has declined. Conversely the genes affecting other factors such as temperament and the ability to stay sound in the face of the rigors of racing and training became more important (Langlois 1996; Bailey 1998). Polymorphisms in *VDR* have been linked to variability in bone mineral density and it has been speculated that polymorphisms in this gene may interact with levels of physical activity (Lorentzon *et al.* 2001; Rabon-Stith *et al.* 2005). As a consequence *VDR* could

potentially influence bone homeostasis during growth and training and as a consequence may assist in keeping the horse sound and capable of performing on the track.

At the moment, discussion of the potential role different candidate genes could play in athletic performance in the Thoroughbred is mainly speculation. While the candidate gene approach has been successful in the past in humans and other species, there is no guarantee that these genes are contributing to athletic performance in the horse. However with this caveat in mind, these genes represent the best candidates for athletic performance in the Thoroughbred at the moment. This is especially true given the lack of any published investigations linking polymorphisms in nuclear genes to athletic performance in the horse. As a consequence, mapping these genes was an important first step in investigating athletic performance in the Thoroughbred.

Contribution to the equine RH and cytogenetic map

The RH and FISH mapping of these 46 genes has also contributed to the increasing list of genes mapped in the equine genome. The FISH mapping is particularly useful as this data provides valuable anchors for RH linkage groups on the chromosomes. Additionally, where marker order is ambiguous and where discrepancies exist between linkage and RH maps, FISH data in combination with comparative data can be helpful in determining the most likely position of the markers in question.

When the genotyping data for the 46 mapped candidate genes was initially analyzed in combination with genotyping data from first generation RH map using RHMAPPER-1.22 it was found that nine of the genes showed no evidence of linkage.

Following the addition of genotyping data for all the markers included in the second generation RH map, all 46 genes showed linkage to a marker. Of the nine genes that initially failed to show linkage, FISH localizations with BAC clones containing the genes all hybridized to previously identified regions of homology between the equine and human genomes, giving us confidence about the accuracy of the RH mapping data. The one exception was the marker *RYR2* located on human chromosome 1. This marker showed linkage by RH to markers on ECA1 and FISH localized it at 1q13, a region not previously identified as homologous between the equine and human genomes in the first generation RH map (Chowdhary *et al.* 2003). *RYR2* showed linkage to the microsatellite *UMNe196* previously RH mapped to a region overlapping the FISH localization of *RYR2* (Wagner *et al.* 2004c). This finding mirrors the observations of Faber & Wedrano (2004) and Tozaki *et al.* (2007) that flanking sequence of *1CA01*, *TKY3232*, and *TKY1924* (microsatellites mapped to ECA1 by linkage analysis) share homology to HSA1. Cross-species chromosome painting also identified this region of homology between HSA1 and ECA1 (Yang *et al.* 2004) and this homology was further confirmed by the second generation map (Raudsepp *et al.* 2008a) and the draft assembly of the equine genome sequence (UCSC EquCab1, http://genome.ucsc.edu/).

Genes that were not mapped

The majority of this mapping work was carried out prior to the availability of the draft assembly of the equine genome sequence (http://genome.ucsc.edu/cgi-bin/hgGateway). Additionally at this time there was limited equine sequence data

available in the various online DNA databases such as NCBI (http://www.ncbi.nlm.nih.gov/). As a consequence primer design had to depend on inferring the sequence of the equine ortholog by comparing the gene in species with more abundant sequence information such as human, cattle, rat and mouse. The fact that it was possible to accurately map 46 of the 55 genes highlights the usefulness of this approach for designing gene specific primers for RH mapping. The availability of a draft sequence of the equine genome has largely negated the necessity of this approach in the horse. However, this technique will still be useful in species lacking the type of genomic tools now available for the horse.

While this approach to designing primers is robust in certain cases it does not always prove successful. Designing primers for RH genotyping without sequence for the target species can be difficult. Primers must be designed in regions of the gene that are sufficiently conserved to allow amplification in the species of interest. However there must also be some differences between the target species and the rodent to prevent amplification of the rodent ortholog.

The release of the draft equine sequence allowed us to identify the location of the unmapped genes in the equine genome along with the position of 10 additional candidate genes reported in the latest update of the human performance gene map (Rankinen *et al.* 2006). As regards the genes that were not mapped three main problems were identified that prevented successful primers design. First, when few comparative sequences were available it was not possible to design primers to amplify either the horse or hamster ortholog. This was the case with *CETP.* Secondly despite a number of attempts we failed

to design primers that produced horse specific amplification (without associated hamster amplification). Most likely this is the result of high conservation between horse and rodent in the regions the primers were designed. This problem was encountered with *LIPG.* Finally in some cases equine specific amplification was achieved, however the sequence of the PCR product was found to be from a closely related member of the same gene family. In the case of *ADRA2A* all the primers designed amplified the closely related *ADRA2B*.

Conclusion

This portion of the study provided a starting point for identifying specific genomic polymorphisms that are associated with athletic performance in the horse. Mapping these human candidate genes makes it possible to identify microsatellite markers adjacent to the genes. This in turn allowed us to test for association and linkage between these microsatellite markers and athletic performance in our sample population. Finally the mapping of these genes has contributed significantly to the second-generation radiation hybrid map of the equine genome and especially to the equine cytogenetic map.

CHAPTER III

CONSTRUCTING A GENOME SCANNING PANEL FOR USE IN THE THOROUGHBRED

Introduction

The primary goal of this project was to identify regions of the genome influencing athletic performance in the Thoroughbred. Given the complexity of athletic performance this phenotype is likely to be influenced by a number of genes (referred to as a quantitative trait loci or QTLs). Most of these loci will probably have a minor impact on the phenotype, however as was pointed out in the initial chapter, many complex quantitative phenotypes are influenced to a large extent by a relatively small number of QTLs (Andersson 2001; Andersson & Georges 2004). The basic strategy for identifying a QTL is essentially the same as that used in mapping polymorphisms responsible for a simple Mendelian trait. Fundamentally both are based on a search for linkage between an allele at a marker locus and an allele at a linked QTL/gene. As a consequence the major requirements for identifying the location of a QTL is, i) a population that shows variation for the trait of interest, and ii) a set of mapped polymorphic markers uniformly distributed across the genome (Falconer & Mackay 1996b). It is the goal of this chapter to outline the rationale and approach taken to develop a whole genome scanning panel of microsatellites for use in the Thoroughbred.

Markers suitable for mapping a gene/QTL must meet a number of requirements. They should be: i) polymorphic, ii) plentiful and located throughout the genome, iii) have

no effect on the phenotype and iv) be co-dominant, so that all genotypes can be identified (Falconer & Mackay 1996b). Prior to the development of DNA based markers, mapping genes/QTLs was very difficult, with neutral protein variations such as blood antigens being the only available polymorphic markers. Interestingly one of the earliest attempts to identify genes associated with athletic performance involved examining blood antigens in athletes competing in the 1968 Mexico Olympics (Rankinen *et al.* 2001). Subsequent advances in molecular biology have provided a wealth of DNA based markers such as: restriction fragment length polymorphisms, microsatellites and single nucleotide polymorphisms, all of which fulfill the requirements outlined above. Genome scanning panels constructed from such markers can be used to map both simple Mendelian traits (e.g. grey coat color in horses) and more complex phenotypes (e.g. % milk fat in dairy cattle) in domestic species.

In theory genome scanning is deceptively simple. A typical genome scan involves genotyping a pedigreed population with a set of genomic markers distributed at regular intervals throughout the genome. The genotypes are then analyzed for each marker individually or more commonly now in combination with adjacent markers. Genotyped individuals are assigned to different groups based on the alleles they carry at each marker. The average phenotype for the animals in each group is then examined to see if it differs significantly from the population mean. This process is repeated across the genome and as a result it is possible to identify regions of the genome contributing to the phenotype under examination (Mackay 2001; Georges 2007).

Genome scanning panels of microsatellites have been used extensively in humans to search for genes influencing complex traits such as asthma, autism, osteoarthritis and schizophrenia (Altmüller *et al.* 2001; Stefánsson *et al.* 2003; Yonan *et al.* 2003; Maziade *et al.* 2005; Pillai *et al.* 2006). They have also been used extensively in livestock species such as cattle, pig and chicken to identify a large number of QTLs affecting a range of economically important traits such as milk production, back fat thickness and egg production (Abasht *et al.* 2006; Ron & Weller 2007; Rothschild *et al.* 2007). In the horse, genome scanning panels have been utilized in mapping both simple Mendelian traits, such as the appaloosa and grey coat colors (Locke *et al.* 2002; Terry *et al.* 2004), and in dissecting the genetic causes of more complex phenotypes affecting the locomotory system such as osteochondrosis (Dierks *et al.* 2007; Diesterbeck *et al.* 2007; Wittwer *et al.* 2007).

In human studies, genome scanning panels are generally composed of 300-400 microsatellite markers with an average of 10 cM between each marker (Altmüller *et al.* 2001). In livestock species, panels of microsatellites generally range between 100-300 markers, with an average distance of 5-20 cM between markers (Georges 2007). In the horse a genome scanning panel containing 101 markers has been available for a number of years (http://www.uky.edu/Ag/Horsemap/Resources/HorseScan.htm). This resource has been successfully employed to identify regions of the genome harboring alleles influencing a number of simple Mendelian phenotypes such as the grey coat color (Locke *et al.* 2002). However, the large cM distances between markers increases the chances that a causative mutation may not be strongly linked (and as a consequence fail to show

evidence of linkage) with one of the markers in the genome scanning panel. In some instances the large distance between markers also reduces the information available from adjacent markers to assist in tracking the segregation of markers through a pedigree. Furthermore where linkage is observed, linkage peaks will cover very large regions of the genome, making identification of the causative mutation relatively difficult. As a consequence more recent investigations have drawn from the increased number of microsatellites available in the horse, to produce panels of microsatellites containing in the region of 200 markers (Dierks *et al.* 2007; Diesterbeck *et al.* 2007; Wittwer *et al.* 2007). The limited number of markers included in the previously developed horse genome scanning panel, combined with the lack of information on the level of polymorphism for these markers in the Thoroughbred, spurred us to develop a new panel of microsatellites, specifically for genome scanning in the Thoroughbred.

Information content of genome scanning panel

The ability of a genome scanning panel of mapped genetic markers to detect linkage is closely related to its information content. Information content is a measure of the fraction of inheritance information extracted by a panel of markers relative to that extracted by an infinitely dense panel of mapped markers. Information content of 1 reflects complete information, information content of 0 represents no information (Kruglyak 1997). Information content is affected by the degree of polymorphism at each locus and by the density of markers in the panel (Kruglyak 1997; Evans & Cardon 2004). Markers with a higher numbers of alleles are more likely to be heterozygous in each

individual genotyped. In heterozygous individuals each homologue will obviously have a different allele, simplifying the task of tracking the segregation of each homologue through the pedigree. Additionally as the density of markers increases, information from closely linked markers can also be utilized. Simulation studies have shown that a genome scanning panel of microsatellite markers (with 5 equally abundant alleles) placed every 2 cM is required to extract 100% of the inheritance information from a pedigree (Evans & Cardon 2004). Such a dense panel was not possible for practical and cost based reasons in the current study. However the availability of mapping information that provided the location of human candidate genes in the equine genome (outlined in the previous chapter) afforded us the opportunity to strategically select and develop our microsatellites. By placing microsatellites adjacent to these genes, if one or more shows a similar influence on athletic performance in Thoroughbreds, the adjacent microsatellite would be in a better position to track the segregation of this gene through the pedigree. At the present time the genes described in Chapter II are the best candidates, however there is no guarantee that they are contributing to athletic performance in the Thoroughbred. As a consequence it was decided to expand beyond the microsatellites adjacent to the candidate genes to create a panel of microsatellites that provide more complete coverage of the genome. Additionally, as has been mentioned earlier, by increasing the density of markers information from adjacent markers can be utilized to increase the power to detect putative linkage with a QTL. The original goal of this work was to develop a panel of microsatellites with one marker per 10 cM. While this marker density is not capable of extracting all the inheritance information from a pedigree, when both parents are

genotyped and the markers sufficiently polymorphic (5 alleles), at the marker itself the percent information content (PIC) is 80%. In other words, at this location 80% of the time it will be possible to tell if the same allele in two sibs is identical by state (IBS) or identical by decent (IBD). Alleles IBD are inherited from the same ancestor, alleles IBS are not. Distinguishing between the two states in turn allows us to identify regions of the genome that tend to segregate with the trait of interest. In the case of the midpoint between the two markers the PIC falls to 70% (Evans & Cardon 2004). As a consequence such a panel is still more than capable of identifying a QTL segregating within a pedigreed population.

Developments in equine genomics

In order to make our panel as informative as possible and therefore retrieve as much information as possible from the genotyped family material it is important that we increase marker density and utilize markers that are polymorphic as possible in the Thoroughbred. Earlier studies investigating the origins of the Thoroughbred and examining microsatellites polymorphism have observed that many microsatellites have fewer alleles in the Thoroughbred when compared to other breeds (Cunningham 1991; Cunningham *et al.* 2001; Tozaki *et al.* 2003). This reduced polymorphism in the Thoroughbred is a legacy of the limited founding population of the breed (Cunningham *et al.* 2001). Therefore many of the published equine microsatellites will provide little information if included in the genome scanning panel because of a lack of polymorphism.

As a consequence markers must be first examined for polymorphism in the Thoroughbred prior to inclusion in the panel, hence ensuring their usefulness.

The rapid pace of development in the field of equine genomics has greatly simplified the search for new polymorphic markers and thus the creation of a new genome scanning panel. The availability of over 1800 published equine microsatellites and their utilization in two recent linkage maps (Penedo *et al.* 2005; Swinburne *et al.* 2006) has proven useful in creating a larger panel of markers for the Thoroughbred and other breeds. Additionally the development of new physical maps of the equine genome based on radiation hybrid analysis (Chowdhary *et al.* 2003; Raudsepp *et al.* 2008a), BAC based comparative mapping (Leeb *et al.* 2006) and the draft assembly of the equine genome sequence (UCSC EquCab2, http://genome.ucsc.edu/), greatly simplifying locating previously published markers and candidate genes as well as establishing their relative distribution and proximity to one another. Finally the availability of the whole genome sequence in conjunction with bioinformatic tools such as sputnik (http://cbi.labri.fr/outils/Pise/sputnik.html) makes the identification of new microsatellite markers a straightforward process.

As this project was primarily concerned with athletic performance the selection of markers for the genome scanning panel prioritized the inclusion of microsatellites adjacent to candidate genes considered to be associated with athletic performance in humans. Subsequent efforts then focused on identifying markers from the rest of the genome, with the goal of producing a panel of microsatellites that has markers distributed on average at 10 cM intervals. The development of this resource will also provide

invaluable information to any equine research group wishing to undertake genome scanning in the Thoroughbred.

Materials and methods

Marker selection and primer design

The initial set of microsatellite markers were selected based on their proximity to candidate genes that influence performance and health-related phenotypes in humans (Wolfarth *et al.* 2005). Scoring data for each candidate gene was initially analyzed by two-point analysis using RHMAPPER-1.22 at http://equine.cvm.tamu.edu/cgi-bin/ecarhmapper.cgi. This analysis identified the closest framework marker from the equine first generation whole genome radiation hybrid map (Chowdhary *et al.* 2003) and consequently the chromosome the candidate genes were located on. Radiation hybrid scoring data for microsatellites on the same chromosome, generated during the construction of the second generation equine RH map, was then obtained. The microsatellite and candidate gene radiation hybrid scoring data was then analyzed together using two-point analysis in the RHMAPPER software, as described in Chowdhary *et al.* (2003). As a result it was possible to identify which microsatellite markers were closest to the candidate gene. The subsequent release of the draft assembly of the equine genome sequence (UCSC EquCab1, http://genome.ucsc.edu/) allowed us to identify the location of previously unmapped genes and select adjacent previously published microsatellites.

The release of the draft assembly of the equine genome sequence (UCSC EquCab1, http://genome.ucsc.edu/) also allowed the determination of megabase distance between the published microsatellites and the candidate genes. Additionally, previously published microsatellite primers could be checked in the draft assembly to ensure they had not been designed within a repetitive element. This was achieved by using the in-silico PCR utility in the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start) to identify the location of the microsatellite. The surrounding DNA sequence (generally 2000bp) was then exported to the RepeatMasker web server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). Around 2000 bp of sequence surrounding each previously published microsatellite was also imported into the primer3 program (http://frodo.wi.mit.edu/) and the published primers checked for mismatches with the reference sequence. The primers were also examined to ensure they matched the default primer design guidelines implemented in the Primer3 software (http://fokker.wi.mit.edu/primer3/input-help-040.htm#PRIMER_SELF_ANY). When the original primers produced genotypes that were difficult to interpret, e.g. due to excessive stutter bands or poor peak morphology, the primers were redesigned using the draft genome sequence adjacent to the original microsatellite repeat.

Where previously published microsatellites were not adjacent to a candidate gene the relevant gene was located in EquCab1 using the UCSC genome browser (http://genome.ucsc.edu/). Generally 125 kb of DNA sequence from either side of the candidate gene was exported and analyzed for repeats using Sputnik (http://cbi.labri.fr/outils/Pise/sputnik.html) to eventually identify new microsatellites.

Sputnik locates microsatellites by identifying repeated patters of nucleotides between 2 and 5 bp in length. A score is assigned to each repeat based on the number of times the pattern of nucleotides is repeated, higher number of repeats give a higher score. The number of insertions or deletions interrupting the repeat reduces the score. The repeat with the highest score determined by sputnik were then located in EquCab1 (http://genome.ucsc.edu/). The microsatellite and the 1000 bp of sequence surrounding the microsatellite (500 bp either side) were initially exported to RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) to mask common repetitive elements in the sequence. The resultant sequence was then used to design primers using the Primer3 software (http://frodo.wi.mit.edu/).

Using the microsatellites adjacent to the candidate genes as a starting point, we sought to identify additional microsatellites in the regions between these markers to provide more complete genome coverage. Existing equine linkage maps (Swinburne *et al.* 2006; Penedo *et al.* 2005) and the integrated map of the equine genome (Raudsepp *et al.* 2008a) were used to identify microsatellite that could potentially fill these gaps in the panel.

The previously published markers considered for inclusion in the panel and the newly identified microsatellites, were initially used to genotype a pool of DNA from 16 Thoroughbreds to ensure they were polymorphic (these animals were all sourced from a single farm in Kentucky). The horses included in the pool were randomly selected from our sample population, as the identities and sex of the animals was not divulged at first the sex and relationships between these animals was not initially known. After the release

of the names of the animals it was found that there were 11 females and 5 males in the pool. It was also observed that 2 animals were half sibs by one stallion, while an additional 3 animals were half sibs by a different stallion. Two animals were half sibs with the same mother; finally there was one mother and son pair. DNA concentration for each horse was diluted to 25 ng/ul and then mixed in equal proportions for use in genotyping. This ensured that each animal contributed an equal amount of template DNA to the pool and allowed us to get a rough approximation of the copy number for each allele based on the relative peak height (see the figure on p.91). Markers with more that three apparent alleles were subsequently genotyped by PCR with DNA from the individual animals to ensure that peaks observed after genotyping the pooled DNA were actually alleles and not stutter bands.

Genotyping

Genotyping was carried out using the three primer method outlined by Oetting *et al.* (1995). In this approach, the forward primer has a 19-bp extension identical to the M13 sequencing primer attached to its 5’end (5’-TTTCCCAGTCACGACGTTG-3’). The reverse primer remains unmodified. This modified forward primer is added to the reaction in much smaller quantities (see below) than the reverse primer and is exhausted in the initial rounds of PCR replication. However, an additional fluorescently labeled primer with a sequence identical to the M13 tail on the forward primer is also included in the reaction at the same concentration as the reverse primer. In subsequent rounds of PCR amplification this fluorescently labeled primer replaces the exhausted M13 tailed forward

primer. Consequently the newly amplified DNA fragments will include the attached fluorochrome (either 6-FAM, VIC, NED, or PET). This approach allowed us to evaluate a large number of microsatellites for their level of polymorphism in the horse, without incurring the expense of obtaining directly labeled fluorescent primers. An important consideration, as many of the markers would not be used in the eventual genome scanning panel due to a lack of polymorphism in the Thoroughbred or poor amplification.

Each PCR reaction contained 50 ng of DNA as template, 0.1 mM dNTPs, 1X buffer (Sigma Aldrich, MO), 1.0 ul of MasterAmp™ (Epicentre, WI), 1 pmol of universal fluorescently labeled M13 primer, 1 pmol of the reverse primer and 0.06 pmol of the M13 tailed forward primer. Additionally 1.5 or 2.0 mM $MgCl_2$ and 0.25 U JumpStart RED*Taq* (without dye) DNA polymerase (Sigma Aldrich, MO) were added to the mix. Reactions were carried out with an initial 30 sec denaturation at 94°C, 1 cycle at 94°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec, followed by 35 cycles at 94°C for 30 sec, annealing temperature ranged from 50°C to 66°C for 30 sec, and extension at 72°C for 30 sec, with a final extension of 72°C for 5 min. All reactions were carried out on either a 96-well Touchgene Gradient thermal cycler (Techne Inc., NJ) or a DNA Engine Dyad (MJ Research Inc., MA). PCR amplification products were resolved on 2.0% agarose gels containing 0.25 ug/ml ethidium bromide. The fluorescently labeled PCR products were stored at -20 °C prior to analysis. After PCR the resultant samples were pooled where possible to reduce the number of genotyping injections necessary on the ABI 3730 (Applied Biosystems, Foster City, CA, USA). Care was taken when pooling samples to ensure that size ranges did not overlap and markers tagged with the

same fluorochrome had product sizes that differed by at least 100 bp. Typically eight samples were pooled, taking 0.25 ul of the PCR product for each sample. The pooled samples were then added to 7.5 ul Hi-Di™ Formamide (Applied Biosystems, Foster City, CA, USA) and 0.5ul GeneScan™ –500 LIZ® Size Standard. The volume of PCR product added was adjusted empirically based on the intensity of the band observed when resolved on the 2% agarose gel and based on peak height after genotyping. The pooled samples were all genotyped on a ABI 3730 (Applied Biosystems, Foster City, CA, USA) following the manufactures instructions.

Sample population

Our sample population mainly consisted of a 137 Thoroughbreds from a single farm in Kentucky. An additional 25 Thoroughbreds, (mainly stallions standing in Kentucky) obtained as part of previous work in our lab were also included in the sample population. This gave a total of 162 animals. The genotypes for the 16 animals included in the DNA pool were obtained during the initial screening of the markers for polymorphism. The markers eventually included in the final genome scanning panel were then genotyped on the remaining 146 Thoroughbreds. The 162 animals (16+146) genotyped could be divided into 59 families based on their pedigree going back three generations.

Data analysis

The resultant genotyping data was analyzed using the GeneMapper® software version 3.5 (Applied Biosystems). Pedigree relationships were confirmed using the PEDSTATS program (Wigginton & Abecasis 2005). The markers were also examined for deviations from Hardy-Weinberg equilibrium using this program. As the markers included in this panel were sourced from more than one linkage map and also included a number of newly identified microsatellites it was necessary to estimate cM distances between the markers. This was achieved using CRI-MAP v2.4 (Green *et al.* 1990) with the BUILD option. Location of the markers in relation to one another was identified by consulting both the integrated map of the horse genome (Raudsepp *et al.* 2008a) and the draft assembly of the equine genome (UCSC EquCab2, http://genome.ucsc.edu/). Accuracy of the approximate cM distances obtained from CRI-MAP was evaluated, where possible, by examining the cM distances between the same microsatellites in existing linkage and physical maps.

Percent information content and multipoint information content for the final set of markers was calculated as descried by Rijsdijk & Sham (2002), using the online genetic power calculator (http://ibgwww.colorado.edu/~pshaun/gpc/mpic.html). The information content of the panel, when genotyped on the available sample population (162 animals) was estimated with the MERLIN program using the "--information" option (Abecasis *et al.* 2001).

Results

Marker selection

In total 454 microsatellites were considered for inclusion in the genome scanning panel. During the initial genotyping of the DNA pool of 16 randomly selected Thoroughbreds it was observed that 51 microsatellites showed no amplification when the M13 tail was added to the forward primer (listed in Table A-3). An additional 28 had poor genotype peak morphology and excessive stutter bands (listed in Table A-4). Hence these 79 (28+51) microsatellites were excluded from the panel. The remaining 375 markers showed satisfactory amplification. An example of the typical genotypes observed during the initial phase of testing is shown in Figure 3-1. The figure shows the genotypes for the markers *COR033* and *COR025* with the pooled DNA at the top and the genotypes of the individual animals used in the pool below. As is evident from the figure, the approach used in this study successfully identified all the alleles possessed by the 16 animals in the pool.

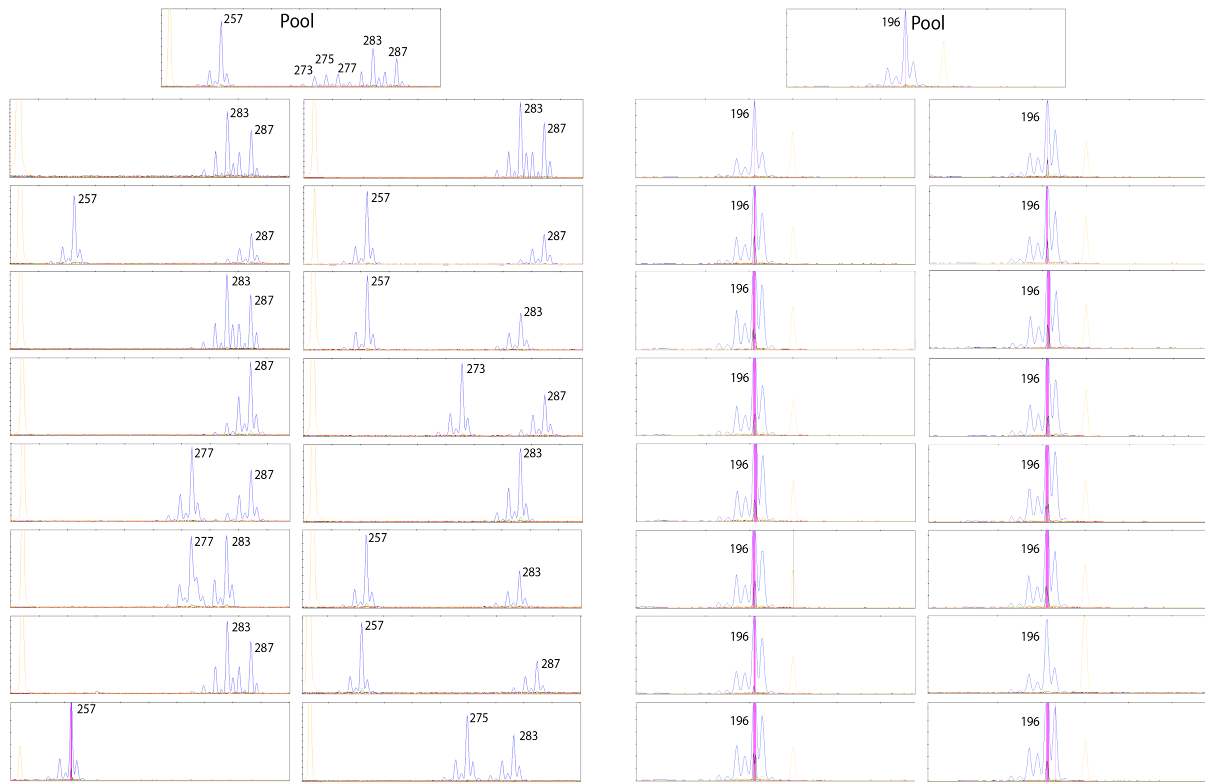

Figure 3-1 Genotypes for pooled DNA and the individuals making up the pool for the microsatellites *COR033* (six alleles) and *COR025* (one allele).

Of the 375 markers, 94 microsatellites were not included in the panel, because they had only one or two alleles. These markers are listed in Table A-5. An additional 62 microsatellites were not used, because they were adjacent to an already available and more polymorphic marker. The majority of these 62 markers possessed three alleles in our 16 Thoroughbreds (listed in Table A-6), however they are likely to be more polymorphic in other breeds. This left a panel of 219 markers that were eventually used to genotype the remaining 146 Thoroughbreds in our sample population. In this panel 192 of the microsatellites had been selected from the pool of previously published equine

microsatellites (http://dga.jouy.inra.fr/cgi-bin/lgbc/main.pl?BASE=horse). The remaining 27 microsatellites were identified from the draft equine genome sequence and chosen because of their proximity to a candidate gene. For the 219 markers selected the mean number of alleles identified after genotyping the 16 animals in the DNA pool was 3.6 (range 2 to 7). When the remaining animals were genotyped this increased to 4.2 (range 2 to 10). The pool of 16 animals used for initial polymorphism analysis tended to underestimate the number of segregating alleles in the population. The latest linkage map of the equine genome produced by Swinburne *et al.* (2006) has a length of 2772 cM. Based on this figure the markers in the panel are distributed on average, at 12.7 cM intervals.

After genotyping the 162 animals in the sample population the markers were checked for deviations from Hardy-Weinberg equilibrium. It was observed that 15 markers showed a significant (<0.05) deviation and were therefore excluded from further analysis (listed in Table A-7). Next, during the estimation of cM distances between the markers it was observed that 18 markers showed unrealistic cM distances based on the cM/Mb distances observed for these markers in other linkage/physical/sequence maps. These markers were also eliminated from further analysis (listed in Table A-8).

Quality of final marker set

After the removal of the markers with questionable integrity a total 186 markers remained in the final panel. Based on the cM distances estimated using CRI-MAP v2.4 the markers in this panel span 2875.5 cM. Consequently the average distance between the

markers is 16.4 cM. For the final panel of 186 markers, the mean number of segregating alleles per marker was 4.2, with a range between 2 to 10 alleles. Further, mean heterozygosity among the markers was 56% with a low of 3.7% for *COR032* and a high of 87% for *TKY936*. The potential single point information content (PIC) for each marker (based on having genotypes for offspring and parents) closely mirrored the figures for heterozygosity. *COR032* showed the lowest PIC at 3.7% while *COR070* showed the highest PIC at 79.8%. As a result if a QTL was located adjacent to *COR032* it is unlikely that this marker would display evidence of linkage to the phenotype. On the other hand the high PIC for *COR070* indicates that it should be capable of showing evidence of linkage to an adjacent QTL. The average single point information content for the panel was 51.7%. (This figure only considers the individual markers; it does not take into account information that can be extracted from adjacent markers, and the ability of the panel to track regions of the genome between the markers.)

The mean multi point information content (MPIC), measured at 1cM intervals, for the genome as a whole was 43.8%. This figure takes into account the ability of the markers to track regions of the genome between markers and also takes into account information that can be extracted from adjacent markers. The mean MPIC for the individual chromosomes ranged between 16.8% for ECA30 and 63.2% for ECA12 and ECA17. Consequently a QTL on ECA30 has a poor chance of being identified, while a QTL on ECA12 and ECA17 had a reasonably good chance of being detected when using this panel.

The PIC calculated by MERLIN for each marker takes into account information from adjacent markers and is based on the actual sample population used in this study. The mean PIC for the genome as whole was 28.2%. Given this low figure, it suggests that we are unlikely to identify a QTL segregating in this population. However regions adjacent to the more polymorphic markers have the potential to show evidence of linkage if a QTL of large effect is present. Mirroring the results for the MPIC, the lowest mean PIC of 8.3% was observed for ECA30, while the highest of the range, 39.9% was seen for ECA12. As regards the individual markers, the lowest PIC observed (calculated by MERLIN) was for *HMS18* (6.7%) and highest observed for *GKms* (51.33%).

A total of 22 of the 27 newly identified markers were included in the final panel (3 were dropped during estimation of cM distances, 2 were dropped as they showed a significant deviation from Hardy-Weinberg equilibrium). These new markers showed a similar range for a number of alleles (4.2), level of heterozygosity (51.8%), potential PIC (44.2%) and actual PIC (28.5%). A summary of the total cM distances, marker information and information content for each chromosome is provided in Table 3-1. Information on allele number, heterozygosity, location in the genome and information content for each marker is detailed in Table A-9. Figure B-1 (Figures with the B prefix can be found in appendix) shows the location of the candidate genes and the markers used in the final panel. A representative selection of chromosomes is shown in Figure 3-2. The blue line represents the MPIC for the respective chromosome, while the actual PIC for each marker is represented by a red square. The location of candidate genes and microsatellite markers on the chromosomes are also depicted.

Table 3-1 Summary of marker information, cM distances and information content for markers on each equine chromosome

Chr No.	Map distance in cM	No. of markers	Average cM distance	Average No. of alleles	HET (%)	Potential Individual PIC	MPIC	Actual PIC
1	235.9	17	13.9	3.4	46	39.3	42.4	23.3
2	147.8	9	16.4	4.1	54	45.6	43.7	23.8
3	101	9	11.2	4.7	60	52.7	59.6	30
4	120.8	11	11	4	53	44.3	52.7	22.1
5	158.2	10	15.8	3.9	55	46.6	43.8	26.6
6	101	7	14.4	5.4	58	53.2	50.2	31.4
7	69.6	6	11.6	4	57	48	50.4	28.1
8	93.5	5	18.7	3.8	54	50.2	38.8	27.5
9	66.8	5	13.4	5.2	62	56	50.1	34
10	77.4	8	9.7	3.6	56	47.7	57.5	30.4
11	98.4	7	14.1	4.1	53	49	47.1	29.3
12	71.9	7	10.3	5.3	65	59.5	63.2	39.9
13	95.1	4	23.8	3.8	58	52.5	35.0	25.6
14	136.7	5	27.3	3.8	55	49.2	29.4	24
15	109	6	18.2	4.2	54	47.8	39.0	23.3
16	85.2	4	21.3	4.8	67	59.4	41.4	33.9
17	62.1	7	8.9	4.3	55	52.8	63.2	32.5
18	111.4	7	15.9	3.6	54	54.4	43.2	22.7
19	38.8	3	12.9	4.3	64	57.4	48.7	32.8
20	122.7	5	24.5	4.6	63	55.8	29.9	36.3
21	29.6	3	9.9	5	71	64.5	62.4	32.2
22	93.1	4	23.3	4.25	58	51	30.4	28.1
23	66.5	4	16.6	2.5	49	39.5	36.4	17.2
24	33.7	3	11.2	5.7	51	62.2	52.3	33.2
25	46.1	2	23.1	4	50	43.7	22.9	17.2
26	74.7	3	24.9	3.7	52	47.8	30.2	25.5
27	65.8	6	11	4.3	59	52.6	52.6	35.8
28	17.9	3	6	4.7	47	48.8	57.4	24.9
29	52.6	3	17.5	4	67	61	41.6	37
30	51	2	25.5	4	42	56.6	16.8	8.3
31	55.4	3	18.5	5	59	54.4	32.9	31.6
X	185.8	8	23.2	8	55	49.7	37.5	34
	2875.5	**186**	**16.4**	**4.4**	**56**	**51.7**	**43.8**	**28.2**

HET % = percent heterozygosity; PIC = percent information content
MPIC = Multipoint information content

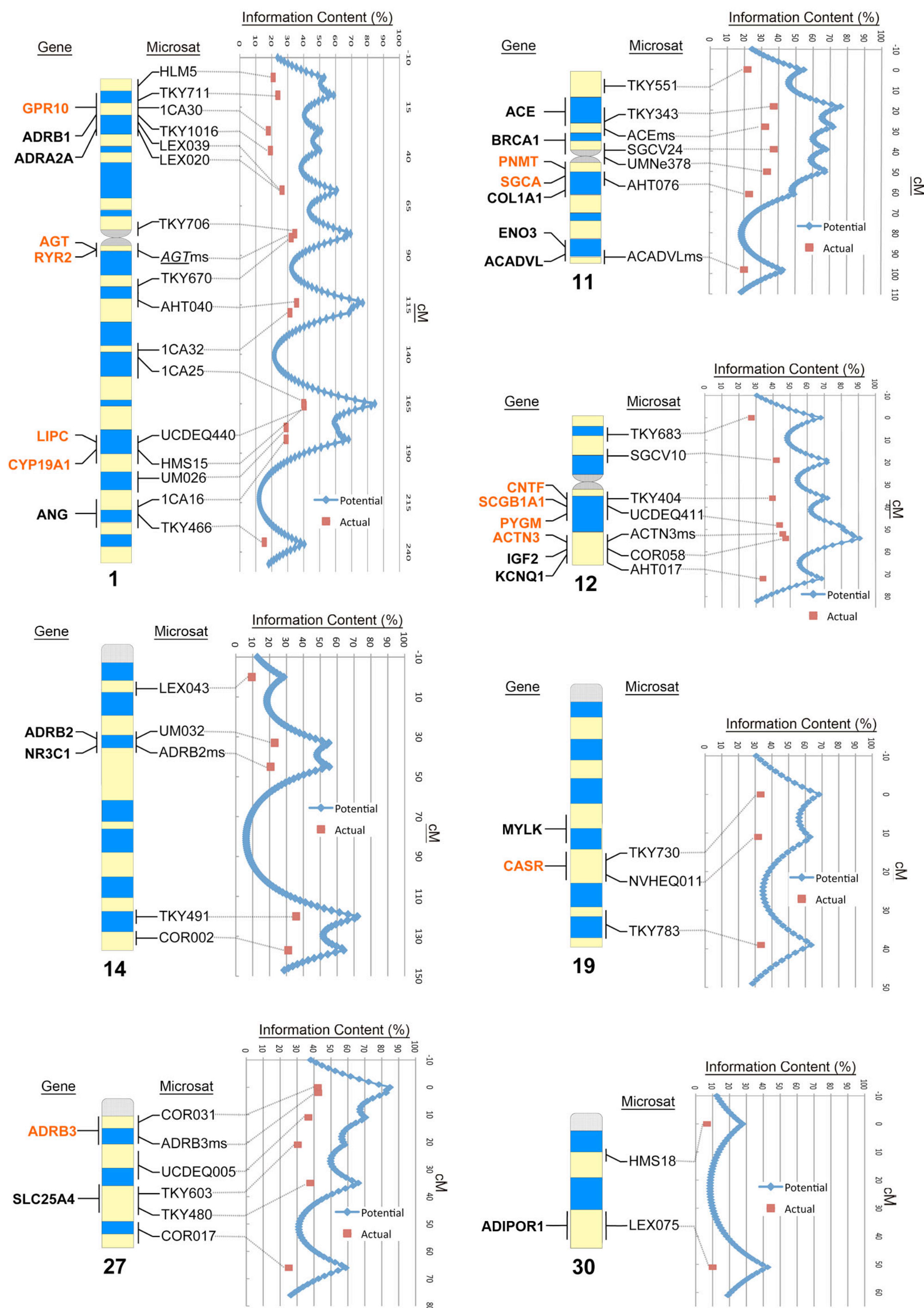

Figure 3-2 Representative figures showing location of candidate genes (genes in red mapped in this study) and microsatellites. Blue line represents the multipoint information content (MPIC) calculated based on marker density and level of polymorphism for each marker. Red boxes represents the actual percent information content calculated in MERLIN and based on the sample population used in the study.

Distance of markers from candidate genes

The draft equine genome sequence (UCSC EquCab2, http://genome.ucsc.edu/) allowed us to determine the distance (in Mb) between candidate genes and the nearest microsatellite included in the genome scanning panel (see Table A-1 & A-2). It was observed that the mean distance between a candidate gene and the nearest microsatellite was 1.18 Mb. In the case of the newly identified markers (derived from the draft genome sequence) the mean distance reduced to 77 kb. Location of the candidate genes and markers used in the panel are shown in Figure B-1. A total of seven candidate genes did not have any adjacent microsatellite. Four of these genes did have an adjacent marker in the original panel of 219 markers, however these markers were subsequently eliminated from the panel during data analysis. For the remaining three genes attempts had been made to identify new microsatellites adjacent to these genes. However no suitable microsatellite could be identified within 250 kb of the respective gene (125 kb each side).

Published primer quality

The majority of the markers used in the panel were evaluated and selected for inclusion in the panel prior to the release of the draft assembly of the equine genome. However, following the release of the assembly, we examined the location of the primer pair used to amplify each microsatellite. It was observed that around one third of the primers had been designed within a repetitive element such as SINE or LINE.

In most cases this had no apparent detrimental effect as determined by the morphology of the peaks produced during genotyping or based on segregation of the alleles. However, it was invariably found that when a microsatellite showed very poor peak morphology and a large number of stutter bands the primers were designed in a repetitive element (see Table A-4). When genotyped some of the published primers had poor peak morphology and excessive stutter making the identification of the individual alleles difficult. This was exacerbated by the fact that the identities of the animals were kept anonymous when the samples were submitted and were divulged only after genotyping had been completed. As a consequence we could not use information from relatives when trying to identify alleles in individuals with ambiguous genotype peaks. In such cases, attempts were made to redesign the primers around the microsatellite. An example of the peak pre- and post-primer redesign is shown in Figure 3-3. In total 54 markers were redesigned, 40 of which we included in the final panel. Markers where primers were redesigned are denoted with an asterisk in Tables A-7, A-8 and A-9.

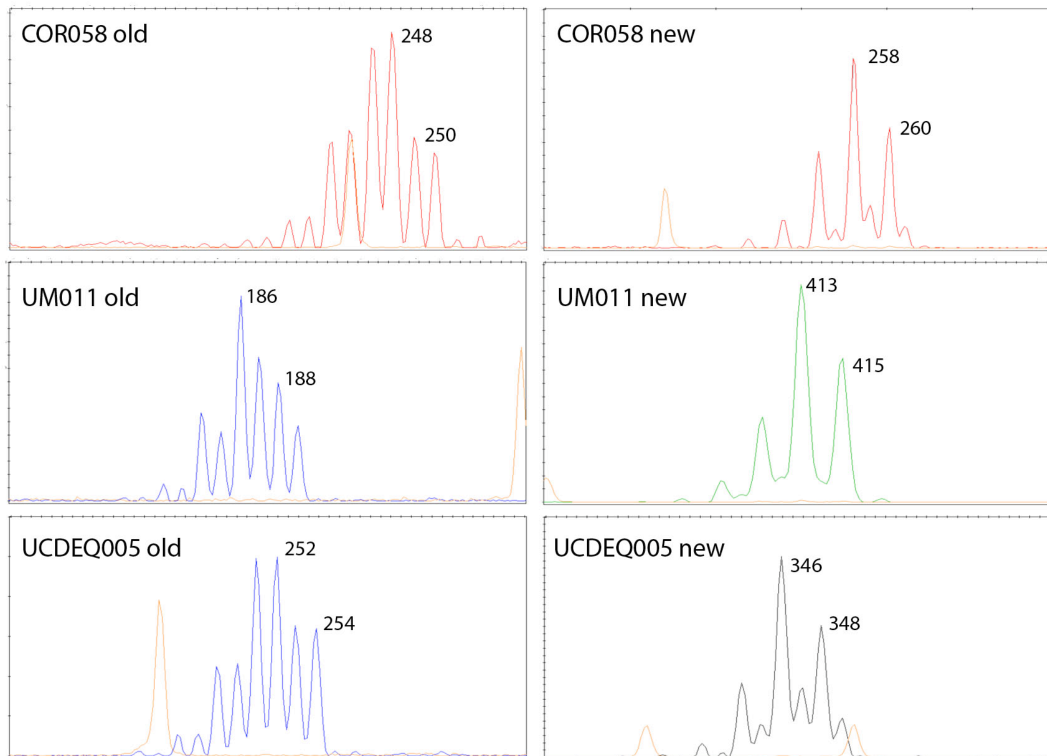

Figure 3-3 Peak morphology before and after primer redesign.

Discussion

Rationale for developing a new genome scanning panel

One of the basic requirements for identifying the underlying genomic variation contributing to a phenotype of interest using linkage analysis is a panel of mapped polymorphic markers. The utility of a panel is influenced to a large extent by two variables. i) The level of polymorphism and consequently the heterozygosity of the markers used. High levels of heterozygosity improve the ability of the panel to identify regions of the genome overrepresented in individuals that deviate significantly from the mean for the trait studied. ii) The density of markers used in the panel. As the density or

markers increases information from adjacent markers can be exploited to track the segregation of a region of the genome. For example despite the biallelic nature and often low level of heterozygosity for SNPs, a panel spaced every 1 cM can extract the same amount of inheritance information as a panel of microsatellites spaced every 2 cM (Kruglyak 1997; Evans & Cardon 2004).

As has been pointed out previously, at the initiation of this project there was a genome scanning panel of just over 100 microsatellite markers available for use in the horse. While this panel had been successfully utilized in the past to identify regions of the equine genome involved in a number of simple Mendelian phenotypes, it had a number of weaknesses. The first was the paucity of markers used in the panel. Based on the length of latest linkage map (Swinburne *et al.* 2006), there is on average over 27 cM between each marker. As a consequence, QTLs between markers are likely to be missed and the extra information that can be extracted by considering adjacent markers together is reduced.

Simulation studies have demonstrated that in order to extract 100% of the inheritance information from a pedigree, microsatellites are required at intervals of 2 cM (Kruglyak 1997; Evans & Cardon 2004). Such a marker density was not feasible because of the cost and effort required to genotype such a large number of markers. Typical genome scanning panels with markers placed every 10 cM are not capable of extracting all of the inheritance information from a pedigree (Kruglyak 1997; Evans & Cardon 2004). At the mid point between two markers the PIC can fall below 60% (Evans & Cardon 2004). However provided the marker has sufficient, equally abundant alleles (>3)

the PIC in the markers immediate vicinity (within a few cM) remains close to 80% (Evans & Cardon 2004). In this study we have tried to find a balance by strategically placed markers adjacent to candidate genes. By doing this we have sought to extract the most inheritance information possible in the regions of the genome most likely to harbor a QTL.

Intermarker distances and cM total

In our genome scanning panel the average distance between a candidate gene and the nearest marker was 1.18 Mb. Thus, if the candidate gene was contributing to athletic performance the close proximity of a marker would increase the likelihood that linkage would be observed. While we were reasonably successful in placing markers adjacent to candidate genes our second goal of producing a panel with one marker per 10 cM was not realized. The actual final intermarker distance in the panel was 16.4 cM. While this was below our target, genome scanning panels with similar numbers of markers have been successfully employed in the past to identify genomic regions harboring QTLs in the horse (Dierks *et al.* 2007; Diesterbeck *et al.* 2007) and other domestic species (Georges 2007). The total cM distance for the panel (2,877.5 cM) is slightly above that seen in the most recent equine linkage map, which spans 2,772 cM (Swinburne *et al.* 2006), but below that seen in the Penedo *et al.* (2005) linkage map, which spans 3,740 cM. The cM distances calculated between the markers included in the panel generally corresponded well with the physical distances between the markers in existing physical maps and the cM distances when the two markers were found in one of the existing linkage maps.

However, these cM distances should still be treated with some caution. The sample population utilized in estimating the cM distances was not optimized for this task. Many of the families used had incomplete genotyping information and as a result cM distances are generally based on a limited number of meiosis.

Our initial panel of markers, genotyped on the 162 animals making up the sample population contained 219 markers. However 33 of these markers had to be excluded from the analysis due to deviation from Hardy-Weinberg equilibrium and unlikely cM distance between markers. The problems observed for these markers were most likely caused by genotyping error rather than any inherent problem with the markers. Many of these markers have been previously used for linkage mapping in the horse (Penedo *et al.* 2005; Swinburne *et al.* 2006). Additionally as these markers are polymorphic in the Thoroughbred, any future attempts to carry out genome scanning in the Thoroughbred should consider including these markers.

Microsatellite polymorphism

The majority of modern Thoroughbreds can trace their ancestry back to a handful of animals (<80 animals) and this narrow genetic base is reflected in the overall low levels of polymorphism observed for microsatellites in the Thoroughbred (Cunningham *et al.* 2001). As has been pointed out previously, fewer alleles generally equates to lower heterozygosity, this in turn reduces the PIC and makes it more difficult to identify QTLs segregating in the population. For example, in the existing 101 microsatellite genome scanning panel (http://www.uky.edu/Ag/Horsemap/Resources/HorseScan.htm), the mean

number of alleles per marker is 7.5 (generally based on genotyping a small number of individuals from a variety of breeds). When these markers were used to genotype our pool of 16 Thoroughbreds it was observed that the mean number of alleles fell to 3.3. Some reduction in the number of alleles is expected as we are using a small number of animals, from a single breed. However the difference in the number of alleles observed was quite striking in some cases. For example the markers *LEX071* and *VHL209* are both listed as possessing seven alleles, however, in our pool of sixteen Thoroughbreds only a single allele was observed. The microsatellite *COR100* is reported to have thirteen alleles and *LEX073* is reported to have fourteen alleles; however in our pool of 16 Thoroughbreds only two alleles were observed for both markers. In general our pool identified only about half of the reported alleles for each marker. However some markers did defy this trend. For example *COR070* is listed as having nine alleles; seven of these alleles were identified in the sixteen animals. Subsequent genotyping for the remaining animals in the sample population identified an additional two alleles.

The final genome scanning panel of markers has an average of 4.2 alleles per microsatellite after genotyping with the sample population of 162 animals. This number of alleles per microsatellite was comparable or slightly lower than that seen in previous studies examining the level of microsatellite polymorphism in the Thoroughbred. Cunningham *et al.* (2001) genotyped 13 microsatellites on 211 Thoroughbred and saw an average of 4.7 alleles per marker. Tozaki *et al.* (2003) genotyped 20 microsatellites on 25 Thoroughbreds and observed on average 4.2 alleles per marker. Finally Glowatzki-Mullis *et al.* (2006) genotyped 50 microsatellites on 31 thoroughbreds and observed an average

of 4.7 alleles per marker. The slightly lower level of polymorphism observed, is most likely the result of including a number of microsatellites in our panel that were not as polymorphic as desired. These markers were included as they were located adjacent to a candidate gene or a region of the genome that lacked another suitable marker.

Overall information content of the panel

The two variables discussed above (marker density and level of polymorphism/heterozygosity) are the major factors affecting the utility of a genome scanning panel. A vivid demonstration of the ability of the panel to extract inheritance information is provided in Figure 3-2 that shows a selection of chromosome with the multipoint information content for each chromosome (for the rest of the chromosomes see Figure B-1). It is noteworthy that the line representing multi point information content (blue) generally fluctuates around the 50% point showing that this panel of markers is constrained in its ability to extract inheritance information from the genome. This is partly attributable to the relatively low level of polymorphism observed in the markers and partly to the low marker density of the panel. In the blue line the location of microsatellites can be identified by the peaks, with more polymorphic markers eliciting higher peaks. As these peaks represent higher MPIC values, these are the regions of the genome the panel is best equipped to track through the pedigree. Consequently QTLs in such regions have the best chance of being identified. This figure also clearly demonstrates the effect of large cM distances between markers. In the case of ECA14 the

75.4 cM gap between *ADRB2ms* and *TKY491* creates a large region of the genome where our ability to identify a QTL is severely constrained.

The figure also demonstrates the impact of marker density and heterozygosity on information content. In the case of the short arm of ECA1 there are six microsatellites with an average of just under 10 cM between each marker. Despite the relatively close proximity of the markers the blue line generally remains below 50%. This low MPIC can be partially explained by the low average number of alleles (2.8) and heterozygosity (46.6%) seen for these markers. This observation highlights the need for highly polymorphic markers in genome scanning panels with relatively large intermarker cM distances. However, the low number of alleles generally observed for microsatellites in the Thoroughbred poses a major challenge when assembling a genome scanning panel of microsatellite markers in this breed.

Conversely on ECA27, the markers *COR031* and *ADRB3ms* are located 2.2 cM apart. *COR031* has 4 alleles, heterozygosity of 75.5% and a PIC of 66.8% while *ADRB3ms* had 5 alleles, 54.9 % heterozygosity and a PIC of 54.7%. In Figure 3-2 the MPIC in the region where these two markers are located actually exceeds 80%, demonstrating that when adjacent markers are analyzed together they are better able to extract inheritance information from that region of the genome. This region also demonstrates that the number of alleles possessed by a microsatellite is not always the best indicator of its utility. Despite having 5 alleles, *ADRB3ms* only has a heterozygosity of 54.9% as many of the alleles were not very common in the population.

Figure B-1 provides an overview of the panel and helps identify what regions have a dearth of markers. ECA19 appears to have a reasonable MPIC at 48.7% (see Table 3-1). However this number is based on only three markers that are all concentrated towards the distal part of the chromosome. MPIC is only calculated for 10 cM beyond the last marker. As a consequence the MPIC observed for this chromosome is somewhat misleading because it does not take into account the location of the markers in the context of the overall size of the chromosome.

The MPIC diagramed in Figures 3-2 and A-1 is calculated assuming that genotyping information will be available for both parents and offspring. However, the majority of the families included herein lacked genotyping information for one of the parents. The effect of these missing genotypes is clearly apparent when the PIC was calculated in MERLIN. The small red boxes in Figure 3-2 and A-1 represent the actual PIC for each marker. From these figures and Table 3-1 (which gives a summary for MPIC, and PIC calculated by MERLIN) it is obvious that the actual amount of information extracted by the markers in the panel is consistently below their actual potential. This observation highlights the importance of obtaining genotyping data from all family members.

Association analysis

By selecting markers in close proximity to the candidate genes we also opened up the possibility of carrying out association analysis using a portion of the microsatellites in our panel. Association analysis is based on linkage disequilibrium present in the population, linkage analysis is based on linkage present within a family (Georges 2007). In association analysis the marker being tested must be in linkage disequilibrium with the underlying causative allele if it is to show evidence of association. As a consequence this method is only useful if the polymorphic marker has been placed adjacent to a candidate gene (within 30-500 kb depending on the species and breed) or if a dense map of markers is used (one marker every 30-500 kb, again depending on the species and breed). The extent of linkage disequilibrium in the horse falls between that seen in the dog (0.5-1 Mb in some breeds) and humans (10-30 kb) (Ardlie *et al.* 2002; Lindblad-Toh *et al.* 2005; Wade *et al.* 2008). This would suggest that only markers within ~250kb of a gene are useful for this approach. However as we are examining athletic performance, a trait that is under positive selection in the Thoroughbred, a candidate gene under selection should show an increased region of linkage disequilibrium in the area surrounding the allele due to the “hitch-hiking” effect (Maynard Smith & Haigh 1974; Andersson & Georges 2004). Consequently linkage disequilibrium may slightly extend beyond its usual limits in areas of the genome surrounding an allele under positive selection.

New technologies vs. microsatellite panel

When this project to study athletic performance in Thoroughbreds was initiated, the only logical way to take a whole genome approach was the use of a panel of microsatellites distributed throughout the genome. However, the last couple of years have seen remarkable developments within the field of equine genomics, salient among which are the sequencing of the equine genome and the development of whole genome expression and genotyping arrays (see Chowdhary & Raudsepp 2008). Providing new platforms and methodologies to in investigate phenotypes of interest.

While the new SNP array based technologies do possess a great deal of potential, they have only recently become available through commercial vendors and their cost may be prohibitive for many laboratories. Moreover, because labs will already have primers for a number of equine microsatellites their use in genome scanning, especially in phenotypes controlled by single genes, is likely to continue for some time. The information on polymorphism/heterozygosity for the markers evaluated during the construction of this panel will be useful for those choosing markers for a specific region of the genome. Finally, despite the uneven distribution of markers and the presence of some large gaps, this new panel of markers will be valuable for linkage analysis in the Thoroughbred. Especially as additional markers can be readily developed in these gaps using the draft equine genome sequence.

CHAPTER IV

A GENOME SCAN FOR ATHLETIC PERFORMANCE IN THE THOROUGHBRED

Introduction

The main goal of the Thoroughbred industry is to breed and train superior equine athletes, capable of excelling on the racetrack. Despite the obvious importance of athletic ability in the Thoroughbred, there has been little progress in identifying the specific differences on the DNA level that distinguish a classic winner from an "also ran" (Ricard *et al.* 2000). Most studies investigating the genetic component of athletic ability in the Thoroughbred have concentrated on determining the heritability of different measurements of racing performance, such as race winnings, handicap rating and race time. These estimates provide an indication of the level of genetic variation available in the population for this composite trait and indirectly give a glimpse of what is required to study this trait at the molecular or sequence level. Such studies have consistently demonstrated a modest heritability for race winnings and handicap ratings, with heritability estimates for these traits generally ranging between a low of 0.1 and a high of 0.5. As regards race times, lower heritability estimates are observed, generally ranging between a low of 0.1 and a high of 0.2. (Tolley *et al.* 1985; Ricard *et al.* 2000).

Despite the detection of heritability for athletic performance in Thoroughbreds by various researchers, to date only one study has demonstrated an association between specific genomic polymorphisms and athletic performance. Using mitochondrial

haplotypes Harrison and Turrion-Gomez (2006) showed that certain mitochondrial haplotypes were correlated with performance at different racing distances in English Thoroughbreds. Given the importance of athletic performance to the Thoroughbred industry the lack of published research investigating how nuclear genes affect athletic performance appears surprising. However, it must be remembered that many of genomic tools (linkage and physical maps) required for investigating a complex phenotype like athletic performance have only recently become available in the horse. This study is one of the first attempts to investigate the contribution of nuclear genes to athletic performance in the Thoroughbred.

Athletic performance in humans

The field of human sports science has become increasingly interested in the genetic component of athletic performance. This has been reflected in a rapidly growing literature reporting a large number of genes and QTLs associated with athletic performance (Rankinen *et al.* 2006). Two main approaches have been employed to investigate athletic performance in the humans: association studies and linkage analysis.

Association studies attempt to identify significant differences in the distribution of alleles for a polymorphism (SNPs, insertions, deletions, etc) between two sample populations. The polymorphisms selected are typically in or adjacent to a gene that has the potential to influence athletic performance (muscle enzyme, growth hormone, etc). While the sample populations used usually consist of elite athletes and non-elite controls, or a population of individuals that display a range of values for the phenotype under

examination. (For example in the case of the *ACE* I/D polymorphism described in Chapter I, it has been reported that the insertion (I) allele tends to occur at higher frequency in a populations of athletes competing in events requiring a high aerobic capacity). The majority of the studies listed in the human performance gene map have employed this type of approach (Rankinen *et al.* 2006).

In comparison to association analysis, linkage studies typically employ panels of highly polymorphic microsatellites markers located at regular intervals (generally 10 cM apart) throughout the genome (Altmüller *et al.* 2001). These markers are used to genotype a population of related individuals in which the phenotype of interest has been measured. The resultant data is then examined for evidence of linkage between the markers and the phenotype of interest. As the markers must be tracked through the generations, pedigreed families are essential in this approach.

One prominent example of the use of this approach in humans was a number of genome scans carried out as part of the HERITAGE family study (Gagnon *et al.* 1996). Over the course of this study exercise related phenotypes such as: maximal exercise capacity, exercise stroke volume, cardiac output and maximal oxygen uptake were investigated. (The panels of microsatellites used contained between 289 and 509 markers and their distribution over the genome ranged from 11 cM per marker to 6.2 cM per marker.) As a result a large number of QTLs influencing these phenotypes were identified throughout the genome (Bouchard *et al.* 2000; Rankinen *et al.* 2002a; Rico-Sanz *et al.* 2003).

In addition to whole genome scans, some researchers have employed a targeted approach by carrying out linkage analysis using microsatellites adjacent to candidate genes. In humans a number of genes involved in the myostatin pathway were investigated using this approach, successfully identifying suggestive evidence of linkage to muscle strength for several genes in the pathway including *RB1* (13q14) and *IGF1* (12q23) (Huygens *et al.* 2004; Huygens *et al.* 2005). Investigations into the genetics of athletic performance in the mouse and rat have also taken advantage of genome scanning panels to identify QTLs affecting endurance capacity (Ways *et al.* 2002; Lightfoot *et al.* 2007).

Sample population

Linkage or association studies require a suitable sample population. In a best-case scenario different breeds are crossed to facilitate mapping the trait of interest in the resultant offspring. Alternately, the large commercial herds and the widespread use of artificial insemination in cattle, pigs and chicken make it possible to collect large half-sib families (Andersson 2001). As a consequence sample populations in some livestock species often exceed 1000 animals for QTL mapping experiments. In the Thoroughbred, breeding operations are more fragmented, with a large number of small farms breeding mares to a variety of different stallions by natural coverings, thus making it much more difficult to obtain large sample populations in this breed. Moreover, these farms tend to be quite insular making the possibility of enrolling several farms in a combined study difficult. Consequently, the smaller sample sizes generally obtainable in the horse will hamper the chances of identifying a QTL. Nevertheless, as was discussed in Chapter I

previous studies in the horse examining multifactorial phenotypes such as osteochondrosis disease and radiological changes in the navicular bone, have been successful in identifying QTLs affecting these phenotypes when using modest sample sizes (<200 animals) (Diesterbeck *et al.* 2007; Wittwer *et al.* 2007). As a consequence it was not unprecedented to attempt to investigate a complex phenotype such as athletic performance using a sample population of modest size.

Materials and methods

Sample population

The majority of the sample population (137) was from a single Thoroughbred farm in Kentucky. Samples were primarily collected in 2004 and mainly consisted of mares and their offspring from the two previous years. Names and relationship between the animals were kept anonymous at the time of sample collection and genotyping by giving each animal a unique identifier. Names were only divulged after all the foals had reached four years of age and had the opportunity to race. An additional 25 Thoroughbreds were also included in the analysis, these were mainly stallions standing in Kentucky that had been obtained as part of previous work in our lab, giving a total of 162 animals. Once the names of the animals were divulged, pedigree information was obtained from the Thoroughbred Database (http://www.pedigreequery.com/) and a pedigree going back two generations was constructed for each genotyped animal.

Phenotype

Track winnings were treated as a quantitative measurement of athletic performance. Winnings have been previously shown to be a reliable indicator of performance on the track and are easily obtained (Tolley *et al.* 1985; Langlois 1996; Langlois & Blouin 2004). Lifetime race winnings for each animal were obtained from the Thoroughbred Database (http://www.pedigreequery.com/). When reported in currencies other than U.S. dollars (USD) winnings were converted into USD based on the exchange rate at the time the winnings were earned. Earnings were $\log_e$ transformed to convert winnings from an exponential to linear trait (Wilson & Rambaut 2008) and the resultant data was treated as a quantitative trait.

Marker selection and genome scanning panel

Isolation, polymorphism content analysis and mapping by genotyping, of the 186 markers are discussed in chapter III and details regarding the markers are provided in Table A-9. For each of the 95 candidate genes associated with athletic performance in humans (Rankinen *et al.* 2006) the closest microsatellite marker included in the panel was identified using the draft equine genome sequence (EquCab2, http://genome.ucsc.edu/). These markers are listed Table A-1 and A-2 and also presented as underlined markers in Table A-9. A total of seven candidate genes have no adjacent microsatellite (within 5 Mb) in the panel. Four of these candidate genes did have an adjacent marker in the original panel of 219 markers. However, during the process of data analysis some markers was eliminated due to a deviation from Hardy-Weinberg equilibrium or because

their inclusion produced inconsistent cM distances between adjacent markers. The nearest microsatellite to each candidate gene (66 markers in total, some markers were in close proximity to more than one candidate gene) was also utilized for family based association analysis using the QTDT program (Abecasis *et al.* 2000).

Pedigree checking and linkage analysis

Pedigree relationships and genotyping results were checked using the PEDSTATS program (Wigginton & Abecasis 2005). Linkage analysis for the autosomes was carried out with the MERLIN software package, version 1.1.2 (Abecasis *et al.* 2001) using the "--qtl" analysis option. This option carries out a non-parametric test for linkage and therefore does not require a specified model for the trait (dominant, recessive, etc). Markers on the X chromosome were analyzed using the separate executable MINX (MERLIN in X) (http://www.sph.umich.edu/csg/abecasis/merlin/reference.html). MERLIN implements a linear model described by Kong and Cox (1997) that was created to identify small increases in allele sharing across a large number of families, as would be expected in the case of a complex phenotype such as athletic performance. At a most basic level the "--qtl" option calculates a score for each allele by determining how much the phenotype of the individuals carrying the allele differs from the population mean. The segregation of the alleles is then tracked through the pedigree (establishes if alleles are identical by descent (IBD) or identical by state (IBS) in related individuals) to determine if a particular region of the genome is linked to the phenotype of interest (http://www.sph.umich.edu/csg/abecasis/Merlin/reference/qtl.html). The procedure for

converting the individual inheritance scores into a Z-score is described in Whittemore and Halpern (1994). Merlin uses the Z-score to calculate a likelihood ratio test for linkage and produces a LOD score statistic utilizing the method described by Kong and Cox (1997).

To determine the significance of the LOD scores obtained with the current data set the "--simulate" option in MERLIN was used. This option replaces the input data with simulated data that matches the original data set for marker spacing, allele frequency and family structure but is unlinked with the trait of interest. As a result it is possible to get an estimate of the LOD scores likely to arise by chance alone. In total 1000 replicates were carried out, the 50 top LOD scores were identified. LOD scores below those observed for the 50 highest were not considered as statistically significant therefore providing the cutoff point for statistical significance. LOD scores observed in the top 50 are usually encountered only once per 20 genome scans and this corresponds to genome wide p-value of 0.05 (Georges 2007).

Association analysis

The markers were tested for evidence of association with track winnings using the Quantitative Transmission Disequilibrium Tests (QTDT) program (Abecasis *et al.* 2000). This program carries out tests for linkage disequilibrium within families (commonly referred to as transmission disequilibrium tests) and is capable of handling both quantitative and discrete traits. This method is limited by the fact that linkage disequilibrium does not extend beyond a few hundred kb (varies between species and

breeds), but it has greater power to identify modest phenotypic effects when compared to linkage analysis (Abecasis *et al.* 2000; Sham *et al.* 2000). For this analysis only the 66 microsatellite markers adjacent to the candidate genes were included in the analysis to reduce the problem of determining statistical significance with multiple testing. Data from the X chromosome was not considered because the QTDT program cannot handle genotypes from this chromosome.

The QTDT program tests for association using the phenotypes of related individuals. As a consequence the program requires information on whether the alleles found in two individuals in a pedigree are identical by state or identical by descent. This information is required to distinguish between linkage and association. The IBD matrices were calculated using MERLIN with the "--ibd" command (http://www.sph.umich.edu/csg/abecasis/Merlin/tour/ibd.html). In order to help describe the similarities between individuals in the pedigree, QTDT allows variance components to be specified. The following variance components suggested by the QTDT tutorial were utilized: **e** (Non-shared Environment, environmental effects unique to each family member), **g** (Polygenic, effects related to degree of relatedness of individuals), and **a** (Additive Major Gene Effect) (http://www.sph.umich.edu/csg/abecasis/QTDT/tour/). QTDT examines each allele separately by default, for microsatellites this would result in a large number of tests. By using the "--multi-allelic" option a global p-value is calculated for each microsatellite, rare alleles (frequency <= 5%) are lumped together and individual effects for the other alleles are estimated to produce a final p-value for the marker (http://www.sph.umich.edu/csg/abecasis/QTDT/docs/multi.html). Only the

genotypes for the 124 animals with available race winnings information were utilized in the association analysis.

Results

Sample population

In total 162 animals were genotyped, while an additional 222 animals was included in the pedigrees to define the relationships between the individuals. The latter group of animals was not genotyped and was not available for analysis. A summary of the families is shown in Table 4-1. Information on race performance was available for 324 of the animals, with 124 of the genotyped animals having race data available. The overall and sex specific distribution of the $\log_e$ transformed winnings is outlined in Figure 4-1.There is a clear tendency for males to have won more prize money, due mainly to the much stronger selection practiced by breeders for superior performance in males compared to females.

For the majority of the chromosomes analyzed only 21 of the families provided sufficient genotyping data to allow MERLIN to track the alleles through the pedigrees. Many of the families had only one or two genotyped individuals and as a result there was insufficient information to track the segregation of the alleles and calculate IBD. In the case of ECA19, ECA23 and ECA31, 20 of the families were informative; for ECA25 and ECA30, only 19 families were informative and for ECAX 18 families provided data that was useful for analysis. These chromosomes all possess a relatively small number of markers compared to the other chromosomes. Consequently if a family displayed a high

level of homozygosity for the markers on one of these chromosomes it was not possible to calculate IBD, thus causing the family to be eliminated from the analysis for that chromosome.

Table 4-1 Summary of family information showing number of animals in each family, number of animals genotyped in each family and the average $\log_e$ transformed winnings for each family.

Family ID	No. animals in ped	No. genotyped	Mean $\log_e$ winnings
101	12	6	12.63
102	7	3	12.92
103	7	3	12.01
104	7	3	11.14
105	7	3	12.39
106	9	4	11.84
107	9	4	12.49
108	17	8	12.21
109	8	4	11.69
110	11	5	12.99
111	10	4	11.15
112	11	5	12.97
113	9	4	12.56
114	15	7	10.66
115	7	3	12.34
116	8	4	11.42
117	14	8	12.79
118	11	5	11.9
119	12	6	11.91
120	5	2	11.21
121	5	2	12.47
122	7	3	11.22
123	11	4	12.19
124	6	2	13.53
125	9	3	12.44
126	7	3	12.69
127	7	3	12.89
128	5	2	12.41
129	5	2	12.83
130	5	2	12.72
131	5	2	12.05
132	5	2	12.33
133	5	2	13.25
134	5	2	13.35
135	5	2	12.89
136	15	7	12.26
137	5	2	12.79
138	5	2	11.42
139	5	2	12.03
140	7	3	11.8
141	3	1	11.23
142	3	1	13.4
143	3	1	13.46
144	3	1	13.77
145	3	1	11.57
146	3	1	11.52
147	3	1	14.26
148	5	1	12.61
149	3	1	11.51
150	3	1	11.78
151	3	1	12.12
152	3	1	12.53
153	3	1	13.21
154	3	1	11.88
155	3	1	11.36
156	3	1	11.66
157	3	1	13.02
158	3	1	14
159	3	1	12.43
Total	**384**	**162**	**12.27**

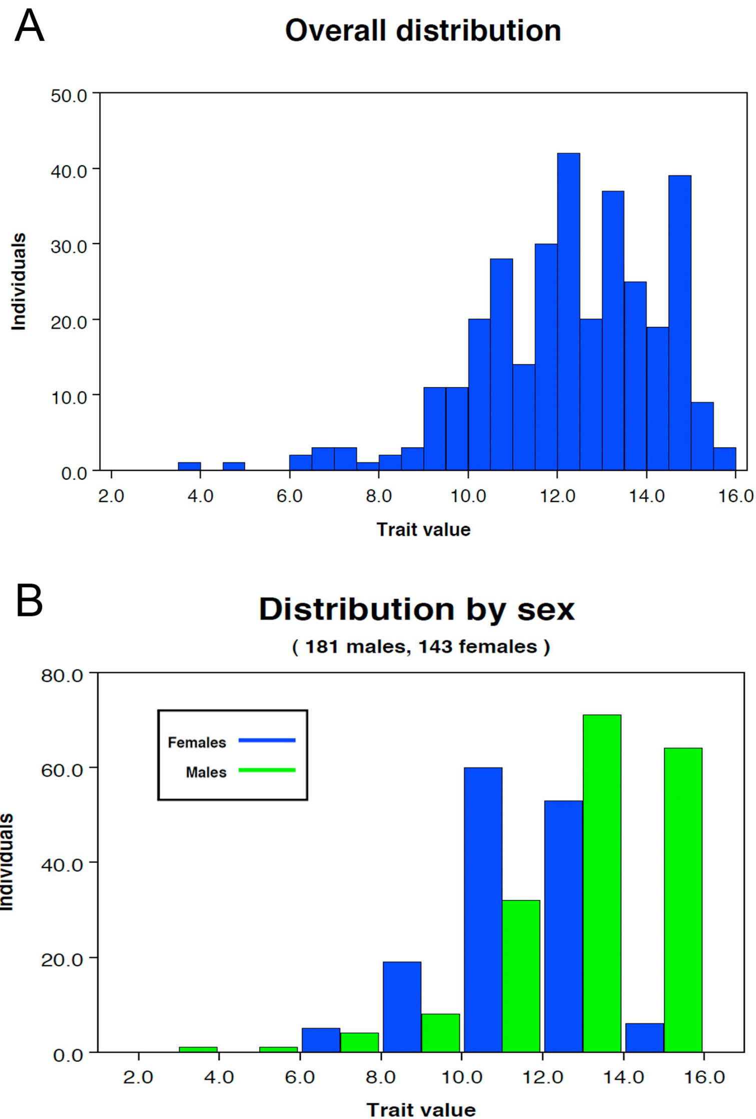

Figure 4-1 Distribution of lifetime winnings after $\log_e$ transformation for all animals, including those that were not genotyped. **A.** Shows overall distribution of winnings for the population as a whole. **B.** Shows distribution of winnings by sex, the pronounced tendency for males to have higher values is the result of the much stronger selection exerted on males.

Whole-genome scan

Results of the linkage analysis for the autosomes are detailed in Table A-10 with the results for the X chromosome in Table A-11. The maximal achievable LOD score was 5.46 for the majority of the chromosomes. For ECAX the LOD score was 5.38, for ECA19, ECA23 and ECA31 it was 5.27, for ECA25 it was 5.1 and finally for ECA30 it was 5.06. The lower maximum LOD score for these chromosomes reflects the fact that fewer families provided useful genotyping data in these cases.

The microsatellite *L12.2* located on ECA29 at 52.6 cM, showed the highest LOD score in the data set, with a LOD score of 1.16 and p-value of 0.01. Plots diagramming the highest LOD scores observed in the genome scan are shown in Figure 4-2 and the LOD scores and cM distance correspond to the data outlined in Table A-10.
The empirically determined threshold for statistical significance was a LOD score of 1.54 (following 1000 simulations). This corresponds to a genome wide p-value of 0.05. The 75 highest LOD scores observed in the simulation are listed in Table A-12. An example of the highest LOD scores observed from one simulation is shown in Figure 4-2 B. The highest LOD score observed in our data set (1.16) is well below the threshold for whole genome statistical significance (1.54). As a consequence it was apparent that our set of markers did not show any evidence of significant linkage to prize money won in the current sample population.

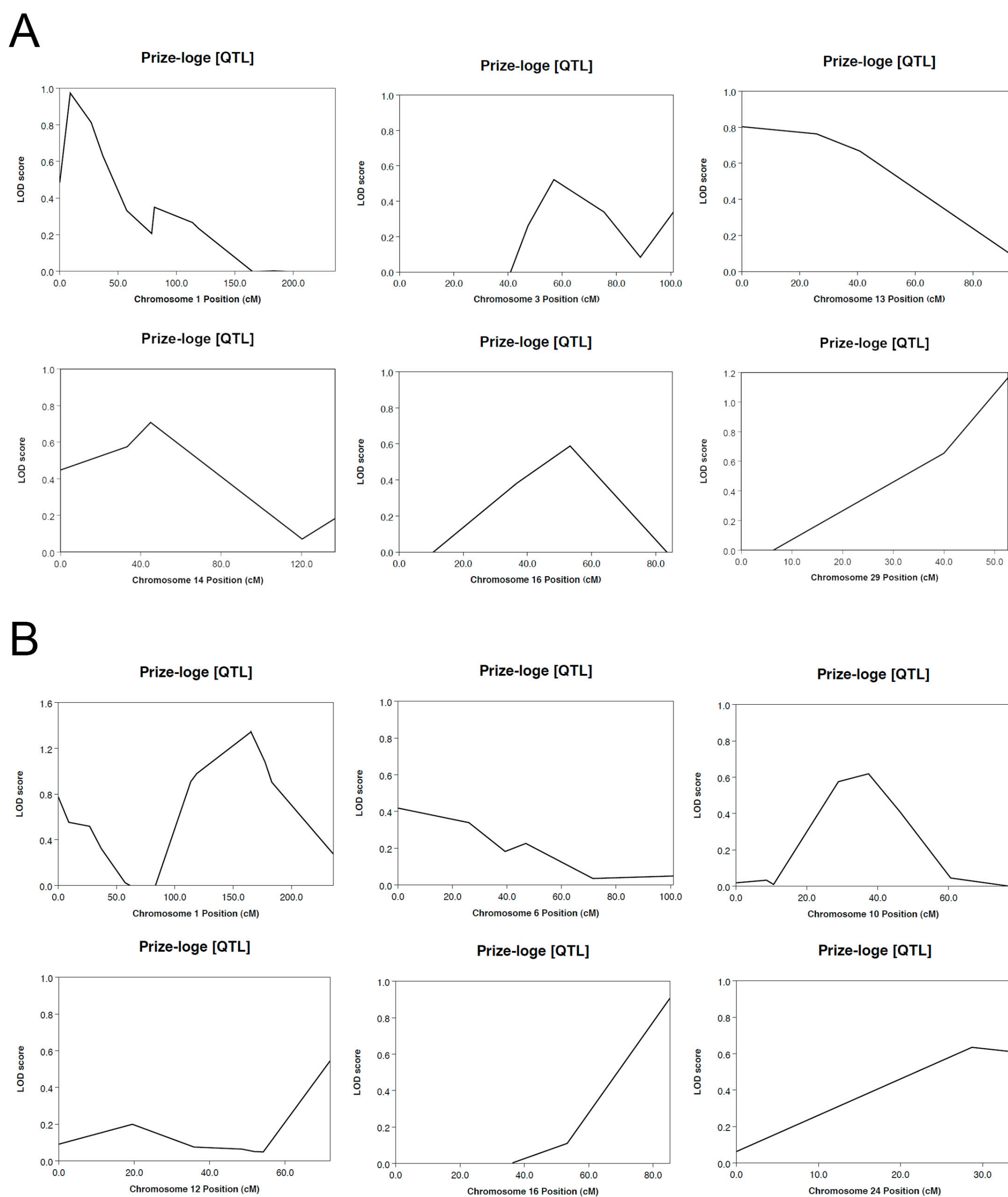

Figure 4-2 Plots for the six chromosomes with the highest LOD scores calculated by the MERLIN program using the "--qtl" option **A.** The actual data set, corresponds to the results presented in Table A-10. **B.** Example from a single simulated data set. The simulated data set has the same maker spacing and level of polymorphism as observed in the original data set. However the simulated genotypes are not linked to the phenotype of interest.

Association analysis

In the case of the family based association analysis implemented by QTDT, no significant evidence of association was observed. The marker displaying the strongest association was the microsatellite *VDRms* with a p-value of 0.4047. However this value is not statistically significant. The output from QTDT is shown in Table A-13. These results were obtained using the default association model and the typical variance components suggested in the QTDT tutorial (http://www.sph.umich.edu/csg/abecasis/QTDT/). Different combinations of variance component and other association models implemented in QTDT were also tested using the genotype data for the 66 adjacent markers. However, no evidence of association with the phenotype was observed.

Discussion

Measuring the phenotype

A number of measurements have been suggested as indicators of performance on the track, including rank at finish, handicap weight, time to finish, best time, earnings, average earnings and log of earnings (Hintz 1980). All are essentially attempts to quantify the ranking of the horse in the races they have participated in and to a certain extent can be seen as interchangeable (Langlois 1996). Information on race winnings is one of the most readily available pieces of information for any Thoroughbred. As a consequence it is an attractive indicator to use in evaluating performance on the track and is one of the most commonly used measurements when determining the heritability of athletic performance in the Thoroughbred (Tolley *et al.* 1985; Langlois 1996; Langlois &

Blouin 2004). A review of the available literature by Tolley *et al.* (1985) found that estimates of the heritability for race winnings ranged between 0.23 and a rather high estimate of 0.56, with the majority of the estimates falling towards the lower end of the range. However when Chico (1994) examined this parameter in Spanish Thoroughbreds and estimated the heritability of race earnings to be a much more modest 0.1. In Poland, Sobczynska and Lukaszewicz (2004) found heritability for earnings to be 0.12, while Svobodova *et al.* (2005) looking at Thoroughbreds in the Czech Republic found heritability of race earnings to be 0.32. Overall, the ready availability of information on prize money and the relatively consistent heritability observed made it the most practical and accurate measurement of performance presently available for the sample population under examination.

One of the main weaknesses of measuring performance with an indicator such as prize money or handicap rating is that in certain cases the indicator may not actually be measuring the exact same phenotype in different animals. Animals that compete over five furlongs (1006 m) will rely to a larger extent on anaerobic pathways than animals competing over two miles (3219 m) (Eaton *et al.* 1995). As a consequence the genes influencing anaerobic pathways would be more important in sprinters than in stayers. Williamson and Beilharz (1998) looking at Australian Thoroughbreds found a high heritability for the distance at which animals preformed best. Ricard *et al.* (2000) reported a similar trend, suggesting that different genetic factors are involved in performance, depending on the race length and racing surface the animal performs best at. An observation that lends some credence to the old saying “horses for courses”.

Unfortunately in the present study it was not possible to separate the sample population based on the animal's optimal racing distance as this information was not available. In future studies it would be prudent to attempt to obtain information on the race distance and track surface favored by each animal in order to analyze animals separately based on their favored race distance and track conditions.

In addition to the difficulty outlined in measuring and quantifying performance of the animals analyzed in this study a second issue faced was how to deal with animals that had not raced. It could be argued that unraced animals should be classified as poor performers because they were most likely judged incapable of racing due to conformational problems or were not considered suitable for racing after evaluation at a training yard. Therefore despite never being tested on a track they should be assigned a lifetime winnings of zero by default. However there are a large number of potential factors (with injury being the most prominent) that conspire to prevent an animal from reaching the track. As a consequence, irrespective of the underlying reason the phenotype of the animals that had not raced was treated as missing data.

Genome scanning panel

One of the assumptions underlying the present investigation is that in addition to QTLs with small effect there are QTLs with a significant effect on the phenotype segregating within the Thoroughbred population. Given the modest sample size and the number of markers available, it was very unlikely that QTLs of small effect would be identifiable in the present study (Andersson & Georges 2004). We had also hoped to

increase our chances of identifying significant linkage by strategically selecting microsatellite markers adjacent to candidate genes. However despite this targeted approach we did not identify significant linkage with the phenotype.

The recent publication of an additional 2,400 microsatellites (Tozaki *et al.* 2007) and the availability of the draft equine genome sequence (UCSC EquCab2, http://genome.ucsc.edu/) increase the prospects of improving marker density and filling some of the gaps in the current panel. However as has been pointed out previously, in order to extract 100% of the inheritance information from a pedigree over 1000 evenly spaced microsatellites are required. Additionally as was outlined in the previous chapter microsatellites tend to have a modest number of alleles in the Thoroughbred, necessitating an even higher marker density to compensate for the lack of heterozygosity. The prospect of attempting to genotype thousands of microsatellites in a large number of animals is daunting. However the development of new SNP based genotyping platforms such as the EquineSNP50 BeadChip (http://www.illumina.com/pages.ilmn?ID=285) appears as an attractive alternative. Preliminary investigations using this SNP chip have shown the majority of the SNPs on the chip are polymorphic in the Thoroughbred. Additionally the chip has been successfully used to remap the chestnut (*MC1R*) and black (*ASIP*) coat color loci in a number of breeds (Mickelson *et al.* 2009). Consequently while it may be prudent to attempt to fill some of the gaps in the panel described in this study, attempting to improve the panel to a point where it is capable of extracting 100% of the information content is not prudent given the availability of alternate approaches.

Sample population

The largest impediment to identifying QTLs affecting athletic performance in the current study was the sample population utilized. The total number of animals genotyped was 162, as has been pointed out a number of times previously when attempting to investigate complex phenotypes large sample population are generally required. The utility of the sample population in the current study was further hampered by two factors. Firstly, 38 of the genotyped animals had never raced and as a result their phenotype was treated as missing data. Secondly, when pedigrees were constructed it was observed that a number of the resultant families had a large number of missing individuals, especially sires. Missing genotypes reduces the ability of MERLIN to track the alleles through the pedigree and determine if an allele is identical by descent or identical by state in two related individuals. This was reflected by the fact that MERLIN was only capable of utilizing the genotype information from 21 of the families in the majority of the chromosomes (see Table A-10 & A-11). MERLIN does not indicate which families are uninformative. Taking a best-case scenario and assuming that families with the most genotyped individual were the families utilized, this would mean that only 100 individuals were actually tested for linkage with the phenotype of interest. Additionally, when we consider the fact that a portion of these animals would also lack race performance data, the actual number falls below 100. In the light of these observations the lack of significant linkage was unsurprising.

In the future, in order to have a realistic change of identifying regions of the genome affecting athletic performance larger sample populations will be required.

Additionally, if methodologies requiring pedigreed populations are utilized, attempts should be made to either obtain genotypes from both parents or increase the number of siblings genotyped to more accurately track the segregation of alleles in the pedigree.

Other confounding factors

In the current study the marker set and sample population are the most cogent factors in relation to our inability to identify significant linkage or association to the trait of interest. However other factors such as epistatic interactions, can also influence our ability to identify QTLs (Carlborg & Haley 2004). Searching for such interactions requires large sample sizes and increases the complexity of data analysis (Andersson & Georges 2004; Georges 2007). Given the modest sample size of the population available in the current study the probability of detecting such interactions was very low. An additional factor that has the potential to complicate the search for QTLs affecting performance is imprinting. While it was not possible to take its potential influence into account in the non-parametric linkage analysis carried out by the MERLIN program (Abecasis *et al.* 2001) it was possible to asses its impact in the QTDT program (Abecasis *et al.* 2000) used for association analysis. However, as with the other variables/models tested using this program no significant association was observed between any marker and the phenotype while assuming imprinting.

Determining significance

A feature of many linkage studies investigating complex phenotypes is the lack of reproducibility of QTLs across populations (Altmüller *et al.* 2001; Glazier *et al.* 2002). Some of this can be attributed to the fact that different QTLs are segregating in different populations for the same phenotype. However, some of this discrepancy may be due to spurious linkage. In the current genome scan almost 200 microsatellites were used, producing a large number of tests for linkage with performance. Therefore by chance alone, some markers will show apparent linkage with the trait of interest. To compensate for this problem of multiple testing, permutation analysis is often employed (Georges 2007). This approach simulates a large number of genome scans based on versions of the study data unlinked to the trait of interest to arrive at a threshold for significance (Es 2003). In the current study the marker *L12.2* showed a LOD score of 1.16 and a p-value of 0.01. At first glance one might be tempted to describe this as significant. However as can be seen from the 1000 simulations carried out in MERLIN (Abecasis *et al.* 2001) high LOD scores (the highest LOD score observed in the simulations was 2.09, with a p-value of 0.001, see Table A-12) can occur by chance and LOD scores above 1.16 are routinely observed in simulated data sets. The simulated data in this study determined that only LOD scores above 1.54 met the criteria for genome wide statistically significance (accepting a 5% chance of a false positive). While it may be worthwhile to examine additional microsatellites in the region around *L12.2* for linkage to performance, the current LOD score for this marker is at best suggestive and well below the threshold for genome wide significance.

Distance to candidate genes

Association analysis is better suited to analyzing complex phenotypes as this approach is capable of detecting QTLs of small effect (as low as 0.01%) in a suitable sample population (Sham *et al.* 2000). However in this approach the genotyped marker must be in strong linkage disequilibrium with the underlying QTN if this association is to be detected. In the present study over half of the microsatellite markers were over 1 Mb from their respective candidate gene and as a consequence were unlikely to be in strong linkage disequilibrium. As a result even if a candidate gene was contributing to athletic performance in the study population, the association may have been missed due to weak linkage disequilibrium between the candidate gene and its respective microsatellite marker, especially because of the small sample size. In contrast having a microsatellite immediately adjacent to the candidate gene is not as vital in linkage analysis, as linkage extends over a number of cM in families. As a consequence microsatellites within 1 or 2 Mb of their respective candidate gene will generally remain linked to the marker as they pass through the pedigree.

Conclusion

The lack of significant evidence of linkage or association with the markers utilized and prize money won on the track, while disappointing, was not surprising. The inherit limitations in the panel of markers utilized, the small sample population and the missing genotypes in the pedigrees all contributed to our inability to detect significant association/linkage. However given the pool of previously published microsatellites

available in the horse at the initiation of this work, it would have been difficult to produce a more informative panel for use in the Thoroughbred. The last number of years has seen rapid progress in the field of equine genomics, with the production a number of new technologies for genotyping and expression analysis. As a result, it seems that future studies investigating athletic performance in the horse may be more constrained by their ability to collect a robust sample population, rather than a lack of the necessary genomic tools.

CHAPTER V

FUTURE PROSPECTS AND CONCLUSIONS

Future prospects

Given the important role athletic performance plays in the Thoroughbred it is certain that future studies will also attempt to identify the molecular underpinnings of this phenotype in the Thoroughbred. Given what has been outlined in this manuscript it is evident that such investigations should include a larger and more comprehensive sample population than utilized in the present study. Despite the development of novel genomic tools and technologies, the fundamental need for a large well documented population will persist. However, given the dispersed and conservative nature of the industry this is likely to remain a major stumbling block.

Provided a suitable sample population can be obtained, the next question is what approach should be adopted for analysis of the phenotype of interest. As mentioned previously, even QTLs of large effect exert a relatively small influence on the overall phenotypic variability in complex phenotypes. Simulation studies carried out by others have shown that in the case of complex phenotypes where parental data is missing, the sample sizes required to be confident of identifying linkage or association are generally much smaller (Sham *et al.* 2000). Consequently future efforts are more likely to benefit from association based approaches given the smaller sample sizes required (hundreds rater than thousands) and the ability to identify more modest phenotypic effects when compared to linkage analysis (Sham *et al.* 2000).

One of the major drawbacks with association analysis is that markers must be in strong linkage disequilibrium with the causative mutation. As a consequence two approaches can be taken, either polymorphic markers can be identified adjacent to candidate genes or a whole genome panel of high-density markers can be employed. In the present study the first approach was utilized by selecting microsatellites adjacent to candidate genes. Future studies attempting to carry out association analysis using the candidate genes approach could also use microsatellites as markers. However, in the horse almost one million SNPs have already been identified as part of the horse genome sequencing effort (http://www.broad.mit.edu/mammals/horse/snp/) and the possibility always remains for individual investigators to re-sequence regions of the genome to identify additional SNPs to tag a region of interest. As a result carrying out association analysis using a panel of SNPs in and around a candidate gene has become a relatively straightforward endeavor.

The great limitation of this candidate genes approach is that selected candidate genes may not be contributing to the variability observed in the population under examination. Even within species, different QTLs can affect the same phenotype in sub populations. This phenomenon has been observed in the literature dealing with athletic performance in humans where ethnic groups can show linkage to different regions of the genome when examining the same phenotype (Rankinen *et al.* 2006). As a consequence if the selected candidate genes are not influencing the variability observed in the sample population, obviously no association with the markers will be observed.

The majority of the discussion so far in relation to candidate genes has been based on genes identified in other species. An attractive alternate approach is to use the recently developed equine oligonucleotide array (Chowdhary & Raudsepp 2008) to identify equine candidate gene for athletic performance. This can be achieved by looking for genes that show altered levels of expression during and after exercise. Investigations into the transcriptional response to exercise in equine muscle have already been carried out on cDNA arrays (McGivney *et al.* 2008) and hopefully these and future investigations will identify candidate genes that can be investigated for association with athletic performance in the horse. While this approach holds promise there are also drawbacks, only a limited number of tissues can be readily accessed to identify transcription changes during exercise. As a consequence only a portion of the potentially important changes in transcription during exercise will be identifiable.

The second option is to go the whole genome route. As was mentioned earlier the equine SNP chip has been successfully used to remap the regions of the genome harboring the recessive loci responsible for two different coat colors (Mickelson *et al.* 2009). This platform could also be utilized to investigate more complex phenotypes in the future, although the large sample population required may be cost prohibitive in the near term. However, the rapid pace of development in genotyping technologies and next generation sequencing will eventually make whole genome approaches more affordable in large sample populations. A whole genome approach may be especially important in examining athletic performance in the Thoroughbred. As was discussed as some length in the chapter I, race times in a number of élite Thoroughbred races have improved little

over the last number of decades, causing some to speculate that Thoroughbreds may have reached a physiological limit that prevents further improvements in race time (Gaffney & Cunningham 1988). Despite this, studies examining the heritability of performance on the track show that genetic variability for racing performance is still present in the Thoroughbred population (Ricard *et al.* 2000). This observation has lead some to speculate that it may be genetic variability in the genes influencing traits such as the will to win or the ability to stay sound during the rigor of training and competing that are important in the modern Thoroughbred rather then genes affecting anaerobic and aerobic respiration (Langlois 1980; Bailey 1998). As a consequence it is difficult to anticipate what genes are influencing performance in Thoroughbred and select plausible candidates, making it prudent to take a whole genome approach in order to cast as wide a net as possible when searching for the regions of the genome affecting performance.

Insights from other species

The sequencing of the equine genome and the development of resources such as the equine SNP chip provides important new tools to assist in the identification of complex traits in the horse. However, as has been seen in humans where high-resolution SNP chips and microarrays have been available for a number of years, even with these technologies in place dissecting the underlying molecular causes of complex traits remains a difficult task. Unlike the case in humans, the underlying population structure and history of the horse provides a number of advantages. As in other domestic species it is possible (but difficult) to collect samples from large numbers of half-sibs, additionally

pedigree records and phenotypic data that extend back for many generations is often available. More importunately domestication and selective breeding will have left marks on the equine genome that should simplify the identification of regions undergoing selection and harboring disease genes (Karlsson & Lindblad-Toh 2008).

In the dog, domestication and subsequent breed creation produced population bottlenecks that reduced genetic variability in the population and created long breed specific haplotype blocks that are then slowly eroded by recombination (Lindblad-Toh *et al.* 2005). These long haplotype blocks can be exploited to identify regions of the genome harboring traits of interest using a relatively small genome wide panel of SNPs. Within breeds a panel containing only 15,000 SNPs is sufficient for association mapping (Karlsson & Lindblad-Toh 2008). However within breeds relatively large regions of the genome will be associated with the phenotype of interest. To narrow down the candidate region identified in one breed it is possible to compare the same large segment of the genome in different breeds exhibiting the same phenotype to identify which segment they have in common. This approach can reduce the candidate region to an extent that it is possible to begin re-sequencing to identify the underlying molecular polymorphisms and a similar strategy is possible in the horse (Karlsson *et al.* 2007). In the horse selective pressure is likely to have been less severe and the incidence of cross breeding higher than that seen in the dog. As a consequence haplotype blocks are shorter (but still considerably longer than seen in humans) and there is likely to be greater haplotype block sharing between breeds (Karlsson & Lindblad-Toh 2008; Wade *et al.* 2008). Nevertheless despite these caveats the history of domestication and breed creation in the horse are major

advantages for mapping single gene traits and hopefully more complex phenotypes in the horse.

A second factor that may help in mapping traits of economic importance in the horse is the effect selection has on the region of the genome harboring the advantageous allele. Selection for a favorable trait will eventually drive the allele responsible to fixation in the population. In addition flanking loci will also become homozygous due to 'hitch-hiking" and as a consequence a region of homozygosity surrounding the allele under selection will be observed. Once the allele is fixed in the population this region of homozygosity will be eroded by recombination (Andersson & Georges 2004). The longer regions of linkage disequilibrium surrounding a polymorphism under positive selection may make it easier to detect these alleles when carrying out whole genome association analysis

Conclusion

Equine genomics has witnessed a remarkable transformation over the last number of years. From an era where the equine genome was essentially a black box to the point where we have a sequenced equine genome. In turn, new genomic tools such as the SNP chip and oligonucleotide array provide powerful new ways to investigate both economically import and health related traits in the equine genome. As a consequence we can hope that the equine genetics community will utilize these tools to improve the efficiency of the equine industry by identifying markers associated with economically important traits. But most importantly it is hoped that the genetic underpinnings of

heritable equine diseases will be elucidated, a development that may also provide insights into human health.

The failure of this study to identify any QTLs associated with athletic performance in the Thoroughbred was disappointing, but not entirely surprising given the inherent limitations in the present study. However this work has not been a futile exercise, mapping candidate genes in the equine genome has contributed to the RH and especially to the FISH map of the equine genome. Additionally the development of a genome scanning panel specifically for the Thoroughbred will be useful for future microsatellite based linkage studies. The rapid developments in the field of equine genomics means that if the study were initiated today a rater different approach would be utilized, taking advantage of new tools such as the equine SNP chip and oligonucleotide array. While the development of such tools is welcome, they are no guarantee of success. The lack of progress in identifying the underlying causes of many complex diseases and phenotypes in human genetics attests to the fact that even with such tools at ones disposable it remains a challenge to dissect the underlying molecular causes of complex phenotypes. However given some of the potential advantages associated with mapping complex traits in a domestic species outlined above we can be hopeful that more concrete insights into the genetic underpinnings of athletic performance in the Thoroughbred are not far off.

REFERENCES

Abasht B., Dekkers J.C.M. & Lamont S.J. (2006) Review of quantitative trait loci identified in the chicken. *Poultry Science* **85**, 2079.

Abecasis G.R., Cardon L.R. & Cookson W.O.C. (2000) A general test of association for quantitative traits in nuclear families. *The American Journal of Human Genetics* **66**, 279-92.

Abecasis G.R., Cherny S.S., Cookson W.O. & Cardon L.R. (2001) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* **30**, 97-101.

Abraham M.R., Olson L.J., Joyner M.J., Turner S.T., Beck K.C. & Johnson B.D. (2002) Angiotensin-converting enzyme genotype modulates pulmonary function and exercise capacity in treated patients with congestive stable heart failure. *Circulation* **106**, 1794-9.

Abrahams B.S. & Geschwind D.H. (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Genetics* **9**, 341.

Agarwala R., Applegate D.L., Maglott D., Schuler G.D. & Schaffer A.A. (2000) A fast and scalable radiation hybrid map construction and integration strategy. *Genome Research* **10**, 350.

Altmüller J., Palmer L.J., Fischer G., Scherb H. & Wjst M. (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *The American Journal of Human Genetics* **69**, 936-50.

Andersson L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2**, 130-8.

Andersson L. & Georges M. (2004) Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics* **5**, 202-12.

Applegate D.L. (2006) *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, Princeton, New Jersey.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299-309.

Ashwell M.S., Heyen D.W., Sonstegard T.S., Van Tassell C.P., Da Y., VanRaden P.M., Ron M., Weller J.I. & Lewin H.A. (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science* **87**, 468-75.

Astrup A., Toubro S., Dalgaard L.T., Urhammer S.A., Soerensen T.I.A. & Pedersen O. (1999) Impact of the v/v 55 polymorphism of the uncoupling protein 2 gene on 24-h energy expenditure and substrate oxidation. *International Journal of Obesity* **23**, 1030-4.

Bailey E. (1998) Odds on the FAST gene. *Genome Research* **8**, 569-71.

Bailey E., Skow L., Bernoco D., DelValle A., Scavone M.D.P., Bowling A.T. & Murray J.D. (2000) Equine dinucleotide repeat loci *LEX071* through *LEX078*. *Animal Genetics* **31**, 286-7.

Barbato J.C., Koch L.G., Darvish A., Cicila G.T., Metting P.J. & Britton S.L. (1998) Spectrum of aerobic endurance running performance in eleven inbred strains of rats. *Journal of Applied Physiology* **85**, 530-6.

Bjornstad G., Midthjell L. & Roed K.H. (2000) Characterization of ten equine dinucleotide microsatellite loci: *NVHEQ21*, *NVHEQ54, NVHEQ67, NVHEQ70, NVHEQ75, NVHEQ77, NVHEQ79, NVHEQ81, NVHEQ82* and *NVHEQ83*. *Animal Genetics* **31**, 78-9.

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y. & Leveziel H. (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35**, 77-102.

Booth F.W., Gordon S.E., Carlson C.J. & Hamilton M.T. (2000) Waging war on modern chronic diseases: primary prevention through exercise biology. *Journal of Applied Physiology* **88**, 774-87.

Bouchard C., Dionne F.T., Simoneau J.A. & Boulay M.R. (1992) Genetics of aerobic and anaerobic performances. *Exercise and Sport Sciences Reviews* **20**, 27-58.

Bouchard C., Rankinen T., Chagnon Y.C., Rice T., Perusse L., Gagnon J., Borecki I., An P., Leon A.S. & Skinner J.S. (2000) Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE Family Study*. *Journal of Applied Physiology* **88**, 551-9.

Bray M.S. (2000) Genomics, genes, and environmental interaction: the role of exercise. *Journal of Applied Physiology* **88**, 788-92.

Breen M., Lindgren G., Binns M.M., Norman J., Irvin Z., Bell K., Sandberg K. & Ellegren H. (1997) Genetical and physical assignments of equine microsatellites--first integration of anchored markers in horse genome mapping. *Mammalian Genome* **8**, 267-73.

Brinkmeyer-Langford C., Raudsepp T., Gustafson-Seabury A. & Chowdhary B.P. (2008) A BAC contig map over the proximal approximately 3.3 Mb region of horse chromosome 21. *Cytogenetic and Genome Research* **120**, 164.

Brinkmeyer-Langford C., Raudsepp T., Lee E.J., Goh G., Schäffer A.A., Agarwala R., Wagner M.L., Tozaki T., Skow L.C. & Womack J.E. (2005) A high-resolution physical map of equine homologs of HSA19 shows divergent evolution compared with other mammals. *Mammalian Genome* **16**, 631-49.

Brutsaert T.D. & Parra E.J. (2006) What makes a champion? Explaining variation in human athletic performance. *Respiratory Physiology & Neurobiology* **151**, 109-23.

Buemann B., Schierning B., Toubro S., Bibby B.M., Sørensen T., Dalgaard L., Pedersen O. & Astrup A. (2001) The association between the val/ala-55 polymorphism of the uncoupling protein 2 gene and exercise efficiency. *International Journal of Obesity* **25**, 467-71.

Caetano A.R., Pomp D., Murray J.D. & Bowling A.T. (1999) Comparative mapping of 18 equine type I genes assigned by somatic cell hybrid analysis. *Mammalian Genome* **10**, 271-6.

Carlborg O. & Haley C.S. (2004) Epistasis: too often neglected in complex trait studies. *Nature Reviews Genetics* **5**, 618-25.

Carlborg O., Kerje S., Schutz K., Jacobsson L., Jensen P. & Andersson L. (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Research* **13**, 413-21.

Carlson C.S., Eberle M.A., Kruglyak L. & Nickerson D.A. (2004) Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446-52.

Chico M.D. (1994) Genetic analysis of thoroughbred racing performance in Spain. *Annales de Zootechnie* **43**, 393–7.

Chowdhary B.P. & Bailey E. (2003) Equine genomics: galloping to new frontiers. *Cytogenetic and Genome Research* **102**, 184-8.

Chowdhary B.P. & Raudsepp T. (2008) The Horse Genome Derby: racing from map to whole genome sequence. *Chromosome Research* **16**, 109-27.

Chowdhary B.P., Raudsepp T., Kata S.R., Goh G., Millon L.V., Allan V., Piumi F., Guerin G., Swinburne J., Binns M., Lear T.L., Mickelson J., Murray J., Antczak D.F., Womack J.E. & Skow L.C. (2003) The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. *Genome Research* **13**, 742-51.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibe B., Bouix J., Caiment F., Elsen J.M. & Eychenne F. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* **38**, 813-8.

Coogle L. & Bailey E. (1997) Equine dinucleotide repeat loci *LEX049-LEX063*. *Animal Genetics* **28**, 378.

Coogle L. & Bailey E. (1999) Equine dinucleotide repeat loci *LEX064* through *LEX070*. *Animal Genetics* **30**, 71-2.

Coogle L., Reid R. & Bailey E. (1996a) Equine dinucleotide repeat loci from *LEX025* to *LEX033*. *Animal Genetics* **27**, 289.

Coogle L., Reid R. & Bailey E. (1996b) Equine dinucleotide repeat loci *LEX015-LEX024*. *Animal Genetics* **27**, 217.

Coogle L., Reid R. & Bailey E. (1997) Equine dinucleotide repeat loci *LEX034-LEX048*. *Animal Genetics* **28**, 309.

Cunningham E.P. (1991) The genetics of thoroughbred horses *Scientific American* **264**, 92-8.

Cunningham E.P., Dooley J.J., Splan R.K. & Bradley D.G. (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics* **32**, 360-4.

Darvasi A. & Pisanté-Shalom A. (2002) Complexities in the genetic dissection of quantitative trait loci. *Trends in Genetics* **18**, 489-91.

de Koning D.J., Rattink A.P., Harlizius B., van Arendonk J.A.M., Brascamp E.W. & Groenen M.A.M. (2000) Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proceedings of the National Academy of Sciences* **97**, 7947-50.

DeChiara T.M., Vejsada R., Poueymirou W.T., Acheson A., Suri C., Conover J.C., Friedman B., McClain J., Pan L. & Stahl N. (1995) Mice lacking the CNTF receptor, unlike mice lacking CNTF, exhibit profound motor neuron deficits at birth. *Cell* **83**, 313-22.

Dekkers J.C.M. (2004) Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science* **82**, E313-E28.

Dekkers J.C.M. & Hospital F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**, 22-32.

Derman K.D. & Noakes T.D. (1994) Comparative aspects of exercise physiology. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine.* (ed. by Hodgson DR & Reuben RJ), pp. 13-25. W.B. Saunders, Philadelphia.

Dierks C., Löhring K., Lampe V., Wittwer C., Drögemüller C. & Distl O. (2007) Genome-wide search for markers associated with osteochondrosis in Hanoverian warmblood horses. *Mammalian Genome* **18**, 739-47.

Diesterbeck U.S., Hertsch B. & Distl O. (2007) Genome-wide search for microsatellite markers associated with radiologic alterations in the navicular bone of Hanoverian warmblood horses. *Mammalian Genome* **18**, 373-81.

Dionne I.J., Garant M.J., Nolan A.A., Pollin T.I., Lewis D.G., Shuldiner A.R. & Poehlman E.T. (2002) Association between obesity and a polymorphism in the bold beta-adrenoceptor gene (Gly389Arg *ADRB1*) in Caucasian women. *International Journal of Obesity* **26**, 633-9.

Downes G.B. & Gautam N. (1999) The G protein subunit gene families. *Genomics* **62**, 544-52.

Druzhevskaya A.M., Ahmetov, II, Astratenkova I.V. & Rogozkin V.A. (2008) Association of the *ACTN3* R577X polymorphism with power athlete status in Russians. *European Journal of Applied Physiology* **103**, 631-4.

Eaton M.D. (1994) Energetics and performance. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine* (ed. by Hodgson DR & Rose RJ), pp. 49-61. W.B. Saunders, Philadelphia.

Eaton M.D., Evans D.L., Hodgson D.R. & Rose R.J. (1995) Maximal accumulated oxygen deficit in thoroughbred horses. *Journal of Applied Physiology* **78**, 1564-8.

Eggleston-Stott M.L., DelValle A., Bautista M., Dileanis D., Wictum E. & Bowling A.T. (1997) Nine equine dinucleotide repeats at microsatellite loci *UCDEQ136*,

UCDEQ405, UCDEQ412, UCDEQ425, UCDEQ437, UCDEQ467, UCDEQ487, UCDEQ502 and *UCDEQ505*. *Animal Genetics* **28**, 370.

Eggleston-Stott M.L., DelValle A., Bautista M., Dileanis S. & Wictum E. (1999) Twelve equine dinucleotide repeats at microsatellite loci *UCDEQ304, UCDEQ380, UCDEQ387, UCDEQ411, UCDEQ439, UCDEQ440, UCDEQ455, UCDEQ457, UCDEQ464, UCDEQ465, UCDEQ482* and *UCDEQ497*. *Animal Genetics* **30**, 69-70.

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992) Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23**, 133-42.

Es G. (2003) The nature and identification of quantitative trait loci: a community's view. *Nature Reviews Genetics* **4**, 911.

Evans D.L. (1994) Training Thoroughbred racehorses. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine.* (ed. by Hodgson DR & Rose RJ), pp. 393-7. W.B. Saunders, Philadelphia.

Evans D.M. & Cardon L.R. (2004) Guidelines for genotyping in genomewide linkage studies: single-nucleotide–polymorphism maps versus microsatellite maps. *The American Journal of Human Genetics* **75**, 687-92.

Ewen K.R. & Matthews M.E. (1994a) *VIAS-H17* and *VIAS-H34*: two new polymorphic equine microsatellite loci. *Animal Genetics* **25**, 63.

Ewen K.R. & Matthews M.E. (1994b) *VIAS-H39*, an equine tetranucleotide microsatellite repeat polymorphism. *Animal Genetics* **25**, 433.

Fadista J., Bendixen C., Thomsen B. & Holm L.-E. (2008) A snapshot of copy number variation in the pig genome. In: *ISAGXXXI*, p. Poster 2092, Amsterdam, The Netherlands.

Falconer D.S. & Mackay T.F.C. (1996a) Heritability. In: *Introduction to Quantitative Genetics* (ed. by Falconer DS & Mackay TFC), pp. 160-83. Longman, New York.

Falconer D.S. & Mackay T.F.C. (1996b) Quantitative trait loci. In: *Introduction to Quantitative Genetics* (ed. by Falconer DS & Mackay TFC), pp. 356-78. Longman, New York.

Farber C.R. & Medrano J.F. (2004) Identification of putative homology between horse microsatellite flanking sequences and cross-species ESTs, mRNAs and genomic sequences. *Animal Genetics* **35**, 28.

Feitosa M.F., Gaskill S.E., Rice T., Rankinen T., Bouchard C., Rao D.C., Wilmore J.H., Skinner J.S. & Leon A.S. (2002) Major gene effects on exercise ventilatory threshold: the HERITAGE Family Study. *Journal of Applied Physiology* **93**, 1000-6.

Feuk L., Carson A.R. & Scherer S.W. (2006) Structural variation in the human genome. *Nature Reviews Genetics* **7**, 85-97.

Flint J. & Mott R. (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics* **2**, 437-45.

Gaffney B. & Cunningham E.P. (1988) Estimation of genetic trend in racing performance of thoroughbred horses. *Nature* **332**, 722-4.

Gagnon J., Province M.A., Bouchard C., Leon A.S., Skinner J.S., Wilmore J.H. & Rao D.C. (1996) The HERITAGE family study: quality assurance and quality control. *Annals of Epidemiology* **6**, 520-9.

Gayagay G., Yu B., Hambly B., Boston T., Hahn A., Celermajer D.S. & Trent R.J. (1998) Elite endurance athletes and the ACE I allele–the role of genes in athletic performance. *Human Genetics* **103**, 48-50.

George L.A., Miller L.M., Valberg S.J. & Mickelson J.R. (1998) Fourteen new polymorphic equine microsatellites. *Animal Genetics* **29**, 469.

Georges M. (2007) Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual Review of Genomics and Human Genetics* **8**, 131-62.

Glazier A.M., Nadeau J.H. & Aitman T.J. (2002) Finding genes that underlie complex traits. *Science* **298**, 2345.

Glowatzki-Mullis M.L., Muntwyler J., Pfister W., Marti E., Rieder S., Poncet P.A. & Gaillard C. (2006) Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics* **37**, 33.

Godard S., Vaiman D., Oustry A., Nocart M., Bertaud M., Guzylack S., Meriaux J.C., Cribiu E.P. & Guerin G. (1997) Characterization, genetic and physical mapping analysis of 36 horse plasmid and cosmid-derived microsatellites. *Mammalian Genome* **8**, 745-50.

Goh G., Raudsepp T., Durkin K., Wagner M.L., Schäffer A.A., Agarwala R., Tozaki T., Mickelson J.R. & Chowdhary B.P. (2007) High-resolution gene maps of horse

chromosomes 14 and 21: Additional insights into evolution and rearrangements of HSA5 homologs in mammals. *Genomics* **89**, 89-112.

Gonzalez N.C., Kirkton S.D., Howlett R.A., Britton S.L., Koch L.G., Wagner H.E. & Wagner P.D. (2006) Continued divergence in VO_2 max of rats artificially selected for running endurance is mediated by greater convective blood O_2 delivery. *Journal of Applied Physiology* **101**, 1288-96.

Green P., Falls K. & Crooks S. (1990) Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, MO.

Grobet L., Royo Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Menissier F. & Massabanda J. (1997) A deletion in the bovine myostatin gene causes the double–muscled phenotype in cattle. *Nature Genetics* **17**, 71-4.

Grundberg E., Brandstrom H., Ribom E.L., Ljunggren O., Mallmin H. & Kindmark A. (2004) Genetic variation in the human vitamin D receptor is associated with muscle strength, fat mass and body weight in Swedish women. *European Journal of Endocrinology* **150**, 323.

Guerin G., Bailey E., Bernoco D., Anderson I., Antczak D.F., Bell K., Binns M.M., Bowling A.T., Brandon R., Cholewinski G., Cothran E.G., Ellegren H., Forster M., Godard S., Horin P., Ketchum M., Lindgren G., McPartlan H., Meriaux J.C., Mickelson J.R., Millon L.V., Murray J., Neau A., Roed K., Ziegle J. & et al. (1999) Report of the International Equine Gene Mapping Workshop: male linkage map. *Animal Genetics* **30**, 341-54.

Guérin G., Bailey E., Bernoco D., Anderson I., Antczak D.F., Bell K., Biros I., Bjornstad G., Bowling A.T., Brandon R., Caetano A.R., Cholewinski G., Colling D., Eggleston M., Ellis N., Flynn J., Gralak B., Hasegawa T., Ketchum M., Lindgren G., Lyons L.A., Millon L.V., Mariat D., Murray J., Neau A., Roed K., Sandberg K., Skow L.C., Tammen I., Tozaki T., Van Dyk E., Weiss B., Young A. & Ziegle J. (2003) The second generation of the International Equine Gene Mapping Workshop half-sibling linkage map. *Animal Genetics* **34**, 161-8.

Guillet C., Auguste P., Mayo W., Kreher P. & Gascan H. (1999) Ciliary neurotrophic factor is a regulator of muscular strength in aging. *Journal of Neuroscience* **19**, 1257-62.

Gustafson A.L., Tallmadge R.L., Ramlachan N., Miller D., Bird H., Antczak D.F., Raudsepp T., Chowdhary B.P. & Skow L.C. (2003) An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenetic and Genome Research* **102**, 189-95.

Gustafson-Seabury A., Raudsepp T., Goh G., Kata S.R., Wagner M.L., Tozaki T., Mickelson J.R., Womack J.E., Skow L.C. & Chowdhary B.P. (2005) High-resolution RH map of horse chromosome 22 reveals a putative ancestral vertebrate chromosome. *Genomics* **85**, 188-200.

Hanson R.W. & Hakimi P. (2008) Born to run; the story of the PEPCK-Cmus mouse. *Biochimie* **90**, 838-42.

Harrison S.P. & Turrion-Gomez J.L. (2006) Mitochondrial DNA: An important female contribution to Thoroughbred racehorse performance. *Mitochondrion* **6**, 53-66.

Henderson K.K., Wagner H., Favret F., Britton S.L., Koch L.G., Wagner P.D. & Gonzalez N.C. (2002) Determinants of maximal O_2 uptake in rats selectively bred for endurance running capacity. *Journal of Applied Physiology* **93**, 1265-74.

Hintz R.L. (1980) Genetics of performance in the horse. *Journal of Animal Science* **51**, 582.

Hirota K., Mashima S., Tozaki T., Sakagami M., Mukoyama H. & Miura N. (1997) Sequence tagged sites on horse chromosomes. *Archivos de Zootecnia* **46**, 3-7.

Hirota K., Tozaki T., Mashima S. & Miura N. (2001) Cytogenetic assignment and genetic characterization of the horse microsatellites, *TKY4-18, TKY20, TKY22-24, TKY30-41* derived from a cosmid library. *Animal Genetics* **32**, 160.

Hopman T.J., Han E.B., Story M.R., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci *COR001-COR020*. *Animal Genetics* **30**, 225-6.

Howlett R.A., Gonzalez N.C., Wagner H.E., Fu Z., Britton S.L., Koch L.G. & Wagner P.D. (2003) Genetic models in applied physiology: selected contribution: Skeletal muscle capillarity and enzyme activity in rats selectively bred for running endurance. *Journal of Applied Physiology* **94**, 1682-8.

Huygens W., Thomis M.A., Peeters M.W., Aerssens J., Janssen R., Vlietinck R.F. & Beunen G. (2004) Linkage of myostatin pathway genes with knee strength in humans. *Physiological Genomics* **17**, 264-70.

Huygens W., Thomis M.A.I., Peeters M.W., Aerssens J., Vlietinck R. & Beunen G.P. (2005) Quantitative Trait Loci for human muscle strength: linkage analysis of myostatin pathway genes. *Physiological Genomics* **22**, 390-7.

Jamshidi Y., Montgomery H.E., Hense H.W., Myerson S.G., Torra I.P., Staels B., World M.J., Doering A., Erdmann J. & Hengstenberg C. (2002) Peroxisome proliferator-

activated receptor a gene regulates left ventricular growth in response to exercise and hypertension. *Circulation* **105**, 950-5.

Kakoi H., Tozaki T., Hirota K., Mashima S., Kurosawa M. & Miura N. (2000) Ten equine microsatellite loci: *TKY25, TKY26, TKY27, TKY28, TKY29, TKY267, TKY268, TKY269, TKY270* and *TKY271*. *Animal Genetics* **31**, 68-9.

Karlsson E.K., Baranowska I., Wade C.M., Hillbertz N., Zody M.C., Anderson N., Biagi T.M., Patterson N., Pielberg G.R. & Kulbokas Iii E.J. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* **39**, 1321-8.

Karlsson E.K. & Lindblad-Toh K. (2008) Leader of the pack: gene mapping in dogs and other model organisms. *Nature Reviews Genetics* **9**, 713.

Koch L.G. & Britton S.L. (2001) Artificial selection for intrinsic aerobic endurance running capacity in rats. *Physiological Genomics* **5**, 45-52.

Kong A. & Cox N.J. (1997) Allele-sharing models: LOD scores and accurate linkage tests. *The American Journal of Human Genetics* **61**, 1179-88.

Kruglyak L. (1997) The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* **17**, 21-4.

Lander E. & Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241-7.

Langlois B. (1980) Heritability of racing ability in thoroughbreds—A review. *Livestock Production Science* **7**, 591-605.

Langlois B. (1996) A consideration of the genetic aspects of some current practices in Thoroughbred horse breeding. *Annales de Zootechnie* **45**, 41-51.

Langlois B. & Blouin C. (2004) Practical efficiency of breeding value estimations based on annual earnings of horses for jumping, trotting, and galloping races in France. *Livestock Production Science* **87**, 99-107.

Langlois B., Blouin C. & Tavernier A. (1996) Nouveaux résultats d'estimation de l'héritabilité des gains en course des pur-sang en France. *Genetics Selection Evolution* **28**, 275-83.

Lanouette C.M., Chagnon Y.C., Rice T., Perusse L., Muzzin P., Giacobino J.P., Gagnon J., Wilmore J.H., Leon A.S. & Skinner J.S. (2002) Uncoupling protein 3 gene is

associated with body composition changes with training in HERITAGE study. *Journal of Applied Physiology* **92**, 1111-8.

Laub M., Hvid-Jacobsen K., Hovind P., Kanstrup I.L., Christensen N.J. & Nielsen S.L. (1993) Spleen emptying and venous hematocrit in humans during exercise. *Journal of Applied Physiology* **74**, 1024-6.

Leamy L.J., Pomp D. & Lightfoot J.T. (2008) An epistatic genetic basis for physical activity traits in mice. *Journal of Heredity* **99**, 639-46.

Lear T.L., Brandon R., Piumi F., Terry R.R., Guerin G., Thomas S. & Bailey E. (2001) Mapping of 31 horse genes in BACs by FISH. *Chromosome Research* **9**, 261-2.

Lee E.-J., Raudsepp T., Kata S.R., Adelson D., Womack J.E., Skow L.C. & Chowdhary B.P. (2004) A 1.4-Mb interval RH map of horse chromosome 17 provides detailed comparison with human and mouse homologues. *Genomics* **83**, 203-15.

Leeb T., Vogl C., Zhu B., de Jong P.J., Binns M.M., Chowdhary B.P., Scharfe M., Jarek M., Nordsiek G., Schrader F. & Blocker H. (2006) A human-horse comparative map based on equine BAC end sequences. *Genomics* **87**, 772-6.

Lightfoot J.T. (2006) Experimentally evolving exercise endurance: one step at a time. *Journal of Applied Physiology* **101**, 1277.

Lightfoot J.T., Turner M.J., Debate K.A. & Kleeberger S.R. (2001) Interstrain variation in murine aerobic capacity. *Medicine & Science in Sports & Exercise* **33**, 2053.

Lightfoot J.T., Turner M.J., Knab A.K., Jedlicka A.E., Oshimura T., Marzec J., Gladwell W., Leamy L.J. & Kleeberger S.R. (2007) Quantitative trait loci associated with maximal exercise endurance in mice. *Journal of Applied Physiology* **103**, 105.

Lightfoot J.T., Turner M.J., Pomp D., Kleeberger S.R. & Leamy L.J. (2008) Quantitative trait loci for physical activity traits in mice. *Physiological Genomics* **32**, 401.

Lindblad-Toh K., Wade C.M., Mikkelsen T.S., Karlsson E.K., Jaffe D.B., Kamal M., Clamp M., Chang J.L., Kulbokas E.J., 3rd, Zody M.C., Mauceli E., Xie X., Breen M., Wayne R.K., Ostrander E.A., Ponting C.P., Galibert F., Smith D.R., DeJong P.J., Kirkness E., Alvarez P., Biagi T., Brockman W., Butler J., Chin C.W., Cook A., Cuff J., Daly M.J., DeCaprio D., Gnerre S., Grabherr M., Kellis M., Kleber M., Bardeleben C., Goodstadt L., Heger A., Hitte C., Kim L., Koepfli K.P., Parker H.G., Pollinger J.P., Searle S.M., Sutter N.B., Thomas R., Webber C., Baldwin J., Abebe A., Abouelleil A., Aftuck L., Ait-Zahra M., Aldredge T., Allen N., An P., Anderson S., Antoine C., Arachchi H., Aslam A., Ayotte L., Bachantsang P., Barry A., Bayul T., Benamara M., Berlin A., Bessette D., Blitshteyn B., Bloom

T., Blye J., Boguslavskiy L., Bonnet C., Boukhgalter B., Brown A., Cahill P., Calixte N., Camarata J., Cheshatsang Y., Chu J., Citroen M., Collymore A., Cooke P., Dawoe T., Daza R., Decktor K., DeGray S., Dhargay N., Dooley K., Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Egbiremolen O., Elong R., Falk J., Farina A., Faro S., Ferguson D., Ferreira P., Fisher S., FitzGerald M., Foley K., Foley C., Franke A., Friedrich D., Gage D., Garber M., Gearin G., Giannoukos G., Goode T., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N., Hagopian D., Hagos B., Hall J., Healy C., Hegarty R., Honan T., Horn A., Houde N., Hughes L., Hunnicutt L., Husby M., Jester B., Jones C., Kamat A., Kanga B., Kells C., Khazanovich D., Kieu A.C., Kisner P., Kumar M., Lance K., Landers T., Lara M., Lee W., Leger J.P., Lennon N., Leuper L., LeVine S., Liu J., Liu X., Lokyitsang Y., Lokyitsang T., Lui A., Macdonald J., Major J., Marabella R., Maru K., Matthews C., McDonough S., Mehta T., Meldrim J., Melnikov A., Meneus L., Mihalev A., Mihova T., Miller K., Mittelman R., Mlenga V., Mulrain L., Munson G., Navidi A., Naylor J., Nguyen T., Nguyen N., Nguyen C., Nicol R., Norbu N., Norbu C., Novod N., Nyima T., Olandt P., O'Neill B., O'Neill K., Osman S., Oyono L., Patti C., Perrin D., Phunkhang P., Pierre F., Priest M., Rachupka A., Raghuraman S., Rameau R., Ray V., Raymond C., Rege F., Rise C., Rogers J., Rogov P., Sahalie J., Settipalli S., Sharpe T., Shea T., Sheehan M., Sherpa N., Shi J., Shih D., Sloan J., Smith C., Sparrow T., Stalker J., Stange-Thomann N., Stavropoulos S., Stone C., Stone S., Sykes S., Tchuinga P., Tenzing P., Tesfaye S., Thoulutsang D., Thoulutsang Y., Topham K., Topping I., Tsamla T., Vassiliev H., Venkataraman V., Vo A., Wangchuk T., Wangdi T., Weiand M., Wilkinson J., Wilson A., Yadav S., Yang S., Yang X., Young G., Yu Q., Zainoun J., Zembek L., Zimmer A. & Lander E.S. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-19.

Lindgren G. (2000) Ten equine dinucleotide microsatellite repeats *HTG18-19, HTG22-24, HTG26-27, HGT29-30* and *HTG32*. *Animal Genetics* **31**, 145.

Lindgren G., Persson H. & Ellegren H. (1999) Five equine dinucleotide microsatellite loci *HTG17, HTG20, HTG21, HTG28* and *HTG31*. *Animal Genetics* **30**, 70-1.

Lindgren G., Sandberg K., Persson H., Marklund S., Breen M., Sandgren B., Carlsten J. & Ellegren H. (1998) A primary male autosomal linkage map of the horse genome. *Genome Research* **8**, 951.

Liu G., Van Tassell C.P., Sonstegard T.S., Alexander L.J., Keele J.W., Matukumalli L., Smith T. & Gasbarre L. (2007) Bovine copy number variation and its implication in animal health. In: *International Symposium on Animal Genomics for Animal Health* p. 32, Paris, France.

Locke M.M., Penedo M.C., Bricker S.J., Millon L.V. & Murray J.D. (2002) Linkage of the grey coat colour locus to microsatellites on horse chromosome 25. *Animal Genetics* **33**, 329-37.

Lorentzon M., Lorentzon R. & Nordström P. (2001) Vitamin D receptor gene polymorphism is related to bone density, circulating osteocalcin, and parathyroid hormone in healthy adolescent girls. *Journal of Bone and Mineral Metabolism* **19**, 302-7.

MacArthur D.G. & North K.N. (2005) Genes and human elite athletic performance. *Human Genetics* **116**, 331-9.

MacArthur D.G., Seto J.T., Raftery J.M., Quinlan K.G., Huttley G.A., Hook J.W., Lemckert F.A., Kee A.J., Edwards M.R. & Berman Y. (2007) Loss of *ACTN3* gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nature Genetics* **39**, 1261.

Mackay T.F. (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, 303-39.

Mardis E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-41.

Mariat D., Oustry-Vaiman A., Cribiu E.P., Raudsepp T., Chowdhary B.P. & Guerin G. (2001) Isolation, characterization and FISH assignments of horse BAC clones containing type I and II markers. *Cytogenetic and Genome Research* **92**, 144-8.

Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25**, 19-23.

Marti E., Breen M., Fischer P., Swinburne J. & Binns M.M. (1998) Six new cosmid derived and physically mapped equine dinucleotide repeat microsatellites. *Animal Genetics* **29**, 236.

Mashima S., Kakoi H. & Tozaki T. (1999) TKY101: a highly polymorphic equine dinucleotide repeat locus. *Animal Genetics* **30**, 163.

Mashima S., Tozaki T., Swinburne J., Kakoi H., Binns M. & Miura N. (2001) Characterization of equine microsatellite loci, *TKY102-TKY112*. *Animal Genetics* **32**, 117-9.

Maynard Smith J. & Haigh J. (1974) The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23-35.

Maziade M., Roy M.A., Chagnon Y.C., Cliche D., Fournier J.P., Montgrain N., Dion C., Lavallee J.C., Garneau Y. & Gingras N. (2005) Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. *Molecular Psychiatry* **10**, 486-99.

McCarroll S.A., Kuruvilla F.G., Korn J.M., Cawley S., Nemesh J., Wysoker A., Shapero M.H., de Bakker P.I.W., Maller J.B., Kirby A., Elliott A.L., Parkin M., Hubbell E., Webster T., Mei R., Veitch J., Collins P.J., Handsaker R., Lincoln S., Nizzari M., Blume J., Jones K.W., Rava R., Daly M.J., Gabriel S.B. & Altshuler D. (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics* **40**, 1166-74.

McGivney B.A., Eivers S., Katz L. & Hill E. (2008) Equine cDNA microarrays and the transcriptomic response of thoroughbred skeletal muscle to exercise. In: *Plant & Animal Genome XVI*, San Diego, CA, USA.

McPherron A.C., Lawler A.M. & Lee S.J. (1997) Regulation of skeletal muscle mass in mice by a new TGF-p superfamily member. *Nature* **387**, 83-90.

McPherron A.C. & Lee S.J. (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences* **94**, 12457.

Meyer A.H., Valberg S.J., Hillers K.R., Schweitzer J.K. & Mickelson J.R. (1997) Sixteen new polymorphic equine microsatellites. *Animal Genetics* **28**, 69.

Mickelson J., McCue M., Bannasch D., Penedo C., Bailey E., Binns M., Distl O., Guerin G., Hasegawa T., Hill E., Leeb T., Lindgren G., Roed K., Swinburne J., Tozaki T., Vaudin M. & Wade C. (2009) EquineSNP50 beadchip evaluation of diverse breeds of horses. In: *Plant & Animal Genomes XVII Conference*, San Diego, CA, U.S.A.

Mickelson J.R., Wagner M.L., Goh G., Wu J.T., Morrison L.Y., Alexander L.J., Raudsepp T., Skow L.C., Chowdhary B.P., Swinburne J.E. & Binns M.M. (2004) Thirty-five new equine microsatellite loci assigned to genetic linkage and radiation hybrid maps. *Animal Genetics* **35**, 481-4.

Mickelson J.R., Wu J.T., Morrison L.Y., Swinburne J.E., Binns M.M., Reed K.M. & Alexander L.J. (2003) Eighty-three previously unreported equine microsatellite loci. *Animal Genetics* **34**, 71-4.

Milenkovic D., Oustry-Vaiman A., Lear T.L., Billault A., Mariat D., Piumi F., Schibler L., Cribiu E. & Guerin G. (2002) Cytogenetic localization of 136 genes in the horse: comparative mapping with the human genome. *Mammalian Genome* **13**, 524-34.

Momozawa Y., Takeuchi Y., Tozaki T., Kikusui T., Hasegawa T., Raudsepp T., Chowdhary B.P., Kusunose R. & Mori Y. (2005) Sequence, detection of polymorphisms and radiation hybrid mapping of the equine catechol-o-methyltransferase gene. *Animal Genetics* **36**, 190.

Montgomery H.E., Marshall R., Hemingway H., Myerson S., Clarkson P. & Dollery C. (1998) Human gene for physical performance. *Nature* **393**, 221-2.

Moore G.E., Shuldiner A.R., Zmuda J.M., Ferrell R.E., McCole S.D. & Hagberg J.M. (2001) Obesity gene variant and elite endurance performance. *Metabolism* **50**, 1391-2.

Moran C.N., Yang N., Bailey M.E.S., Tsiokanos A., Jamurtas A., MacArthur D.G., North K., Pitsiladis Y.P. & Wilson R.H. (2007) Association analysis of the *ACTN3* R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. *European Journal of Human Genetics* **15**, 88-93.

Morison I.M., Ramsay J.P. & Spencer H.G. (2005) A census of mammalian imprinting. *Trends in Genetics* **21**, 457-65.

Mosher D.S., Quignon P., Bustamante C.D., Sutter N.B., Mellersh C.S., Parker H.G. & Ostrander E.A. (2007) A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genetics* **3**, e79.

Murphie A.M., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci *COR021-COR040*. *Animal Genetics* **30**, 235-7.

Murphy W.J., Davis B., David V.A., Agarwala R., Schäffer A.A., Pearks Wilkerson A.J., Neelam B., O'Brien S.J. & Menotti-Raymond M. (2007) A 1.5-Mb-resolution radiation hybrid map of the cat genome and comparative analysis with the canine and human genomes. *Genomics* **89**, 189-96.

Musilova P., Kubickova S., Vychodilova-Krenkova L., Kralik P., Matiasovic J., Hubertova D., Rubes J. & Horin P. (2005) Cytogenetic mapping of immunity-related genes in the domestic horse. *Animal Genetics* **36**, 507-10.

Myerson S., Hemingway H., Budget R., Martin J., Humphries S. & Montgomery H. (1999) Human angiotensin I-converting enzyme gene and endurance performance. *Journal of Applied Physiology* **87**, 1313-6.

Myerson S.G., Montgomery H.E., Whittingham M., Jubb M., World M.J., Humphries S.E. & Pennell D.J. (2001) Left ventricular hypertrophy with exercise and *ACE*

gene insertion/deletion polymorphism a randomized controlled trial with losartan. *Circulation* **103**, 226-30.

Nazarov I.B., Woods D.R., Montgomery H.E., Shneider O.V., Kazakov V.I., Tomilin N.V. & Rogozkin V.A. (2001) The angiotensin converting enzyme I/D polymorphism in Russian athletes. *European Journal of Human Genetics* **9**, 797-801.

Niemi A.K. & Majamaa K. (2005) Mitochondrial DNA and *ACTN3* genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics* **13**, 965-9.

Oetting W.S., Lee H.K., Flanders D.J., Wiesner G.L., Sellers T.A. & King R.A. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* **30**, 450-8.

Oki H., Sasaki Y. & Willham R.L. (1995) Genetic parameter estimates for racing time by restricted maximum likelihood in the Thoroughbred horse of Japan. *Journal of Animal Breeding and Genetics* **112**, 146–50.

Penedo M.C., Millon L.V., Bernoco D., Bailey E., Binns M., Cholewinski G., Ellis N., Flynn J., Gralak B., Guthrie A., Hasegawa T., Lindgren G., Lyons L.A., Roed K.H., Swinburne J.E. & Tozaki T. (2005) International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogenetic and Genome Research* **111**, 5-15.

Perrocheau M., Boutreux V., Chadi S., Mata X., Decaunes P., Raudsepp T., Durkin K., Incarnato D., Iannuzzi L., Lear T.L., Hirota K., Hasegawa T., Zhu B., de Jong P., Cribiu E.P., Chowdhary B.P. & Guérin G. (2006) Construction of a medium-density horse gene map. *Animal Genetics* **37**, 145-55.

Perusse L., Rankinen T., Rauramaa R., Rivera M.A., Wolfarth B. & Bouchard C. (2003) The human gene map for performance and health-related fitness phenotypes: the 2002 update. *Medicine & Science in Sports & Exercise* **35**, 1248-64.

Pillai S.G., Chiano M.N., White N.J., Speer M., Barnes K.C., Carlsen K., Gerritsen J., Helms P., Lenney W. & Silverman M. (2006) A genome-wide search for linkage to asthma phenotypes in the genetics of asthma international network families: evidence for a major susceptibility locus on chromosome 2p. *European Journal of Human Genetics* **14**, 307-16.

Rabon-Stith K.M., Hagberg J.M., Phares D.A., Kostek M.C., Delmonico M.J., Roth S.M., Ferrell R.E., Conway J.M., Ryan A.S. & Hurley B.F. (2005) Vitamin D

receptor FokI genotype influences bone mineral density response to strength training, but not aerobic training. *Experimental Physiology* **90**, 653-61.

Rankinen T., An P., Perusse L., Rice T., Chagnon Y.C., Gagnon J., Leon A.S., Skinner J.S., Wilmore J.H. & Rao D.C. (2002a) Genome-wide linkage scan for exercise stroke volume and cardiac output in the HERITAGE family study. *Physiological Genomics* **10**, 57-62.

Rankinen T., Bray M.S., Hagberg J.M., Perusse L., Roth S.M., Wolfarth B. & Bouchard C. (2006) The human gene map for performance and health-related fitness phenotypes: the 2005 update. *Medicine & Science in Sports & Exercise* **38**, 1863-88.

Rankinen T., Perusse L., Gagnon J., Chagnon Y.C., Leon A.S., Skinner J.S., Wilmore J.H., Rao D.C. & Bouchard C. (2000a) Angiotensin-converting enzyme ID polymorphism and fitness phenotype in the HERITAGE Family Study. *Journal of Applied Physiology* **88**, 1029-35.

Rankinen T., Perusse L., Rauramaa R., Rivera M.A., Wolfarth B. & Bouchard C. (2002b) The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Medicine & Science in Sports & Exercise* **34**, 1219-33.

Rankinen T., Perusse L., Rauramaa R., Rivera M.A., Wolfarth B. & Bouchard C. (2001) The human gene map for performance and health-related fitness phenotypes. *Medicine & Science in Sports & Exercise* **33**, 855.

Rankinen T., Rice T., Leon A.S., Skinner J.S., Wilmore J.H., Rao D.C. & Bouchard C. (2002c) G protein {beta} 3 polymorphism and hemodynamic and body composition phenotypes in the HERITAGE Family Study. *Physiological Genomics* **8**, 151-7.

Rankinen T., Wolfarth B., Simoneau J.A., Maier-Lenz D., Rauramaa R., Rivera M.A., Boulay M.R., Chagnon Y.C., Perusse L. & Keul J. (2000b) No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. *Journal of Applied Physiology* **88**, 1571-5.

Raudsepp T. & Chowdhary B.P. (2008) The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. *Cytogenetic and genome research* **121**, 102-9.

Raudsepp T., Fronicke L., Scherthan H., Gustavsson I. & Chowdhary B.P. (1996) Zoo-FISH delineates conserved chromosomal segments in horse and man. *Chromosome Research* **4**, 218-25.

Raudsepp T., Gustafson-Seabury A., Durkin K., Wagner M.L., Goh G., Seabury C.M., Brinkmeyer-Langford C., Lee E.J., Agarwala R., Stallknecht Rice E., Schaffer A.A., Skow L.C., Tozaki T., Yasue H., Penedo M.C.T., Lyons L.A., Khazanehdari K.A., Binns M.M., MacLeod J.N., Distl O., Guerin G., Leeb T., Mickelson J.R. & Chowdhary B.P. (2008a) A 4103 marker integrated map of the horse genome. *Cytogenetic and Genome Research* **122**, 28-36.

Raudsepp T., Kata S.R., Piumi F., Swinburne J., Womack J.E., Skow L.C. & Chowdhary B.P. (2002) Conservation of gene order between horse and human X chromosomes as evidenced through radiation hybrid mapping. *Genomics* **79**, 451-7.

Raudsepp T., Kijas J., Godard S., Guerin G., Andersson L. & Chowdhary B.P. (1999) Comparison of horse chromosome 3 with donkey and human chromosomes by cross-species painting and heterologous FISH mapping. *Mammalian Genome* **10**, 277-82.

Raudsepp T., Lee E.J., Kata S.R., Brinkmeyer C., Mickelson J.R., Skow L.C., Womack J.E. & Chowdhary B.P. (2004a) Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. *Proceedings of the National Academy of Sciences* **101**, 2386.

Raudsepp T., Otte K., Rozell B. & Chowdhary B.P. (1997) FISH mapping of the *IGF2* gene in horse and donkey-detection of homoeology with HSA11. *Mammalian Genome* **8**, 569-72.

Raudsepp T., Paria N. & Chowdhary B.P. (2008b) Sequencing-ready physical map of the horse Y chromosome Sequencing-ready physical map of the horse Y chromosome. In: *31st Conference of the International Society of Animal Genetics*, Amsterdam, The Netherlands.

Raudsepp T., Santani A., Wallner B., Kata S.R., Ren C., Zhang H.B., Womack J.E., Skow L.C. & Chowdhary B.P. (2004b) A detailed physical map of the horse Y chromosome. *Proceedings of the National Academy of Sciences,* **101**, 9321-6.

Ricard A., Bruns E. & Cunningham E.P. (2000) Genetics of performance traits. In: *The Genetics of the Horse* (ed. by Bowling AT & Ruvinsky A), pp. 411-38. CABI, UK.

Rico-Sanz J., Rankinen T., Rice T., Leon A.S., Skinner J.S., Wilmore J.H., Rao D.C. & Bouchard C. (2003) Quantitative trait loci for maximal exercise capacity phenotypes and their responses to training in the HERITAGE Family Study. *Physiological Genomics* **16**, 256-60.

Ridgway K.J. (1994) Training endurance horses. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine.* (ed. by Hodgson DH & Rose RJ), pp. 409-28. W.B. Saunders, Philadelphia.

Rijsdijk F.V. & Sham P.C. (2002) Estimation of Sib-Pair IBD sharing and multipoint polymorphism information content by linear regression. *Behavior Genetics* **32**, 211-20.

Rivera M.A., Dionne F.T., Simoneau J.A., PÉRusse L., Chagnon M., Chagnon Y., Gagnon J., Leon A.S., Rao D.C. & Skinner J.S. (1997) Muscle-specific creatine kinase gene polymorphism and VO2max in the HERITAGE Family Study. *Medicine & Science in Sports & Exercise* **29**, 1311.

Rivera M.A., PÉRusse L., Simoneau J.A., Gagnon J., Dionne F.T., Leon A.S., Skinner J.S., Wilmore J.H., Province M. & Rao D.C. (1999) Linkage between a muscle-specific CK gene marker and O2max in the HERITAGE Family Study. *Medicine & Science in Sports & Exercise* **31**, 698.

Roberts M.C., Murtaugh J., Valberg S.J., Mickelson J.R. & Alexander L.J. (2000) Twenty-one new equine dinucleotide repeat microsatellites. *Animal Genetics* **31**, 141.

Roed K.H., Midthjell L. & Bjornstad G. (1998) Eight new equine dinucleotide repeat microsatellites at the *NVHEQ26, NVHEQ29, NVHEQ31, NVHEQ40, NVHEQ43, NVHEQ90, NVHEQ98* and *NVHEQ100* loci. *Animal Genetics* **29**, 470.

Ron M. & Weller J.I. (2007) From QTL to QTN identification in livestock-winning by points rather than knock-out: a review. *Animal Genetics* **38**, 429.

Rose R.J. & Hodgson D.R. (1994) An overview of performance and sports medicine. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine.* (ed. by Hodgson DR & Rose RJ), pp. 3-11. W.B. Saunders, Philadelphia.

Roth S.M., Schrager M.A., Ferrell R.E., Riechman S.E., Metter E.J., Lynch N.A., Lindle R.S. & Hurley B.F. (2001) *CNTF* genotype is associated with muscular strength and quality in humans across the adult age span. *Journal of Applied Physiology* **90**, 1205-10.

Rothschild M.F., Hu Z. & Jiang Z. (2007) Advances in QTL Mapping in Pigs. *International Journal of Biological Sciences* **3**, 192.

Rupert J.L. (2003) The search for genotypes that underlie human performance phenotypes. *Comparative Biochemistry and Physiology, Part A* **136**, 191-203.

Ruth L.S., Hopman T.J., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci *COR041-COR060*. *Animal Genetics* **30**, 320-1.

Schaffer A.A., Rice E.S., Cook W. & Agarwala R. (2007) rh_ tsp_ map 3. 0: end-to-end radiation hybrid mapping with improved speed and quality control. *Bioinformatics* **23**, 1156.

Schuelke M., Wagner K.R., Stolz L.E., Hubner C., Riebel T., Komen W., Braun T., Tobin J.F. & Lee S.J. (2004) Myostatin mutation associated with gross muscle hypertrophy in a child. *New England Journal of Medicine* **350**, 2682.

Scott R.A., Moran C., Wilson R.H., Onywera V., Boit M.K., Goodwin W.H., Gohlke P., Payne J., Montgomery H. & Pitsiladis Y.P. (2005) No association between Angiotensin Converting Enzyme (ACE) gene variation and endurance athlete status in Kenyans. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **141**, 169-75.

Sham P.C., Cherny S.S., Purcell S. & Hewitt J.K. (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *The American Journal of Human Genetics* **66**, 1616-30.

Shiue Y.L., Bickel L.A., Caetano A.R., Millon L.V., Clark R.S., Eggleston M.L., Michelmore R., Bailey E., Guerin G., Godard S., Mickelson J.R., Valberg S.J., Murray J.D. & Bowling A.T. (1999) A synteny map of the horse genome comprised of 240 microsatellite and RAPD markers. *Animal Genetics* **30**, 1-9.

Slonim D., Kruglyak L., Stein L. & Lander E. (1997) Building human genome maps with radiation hybrids. *Proceedings of the First Annual International Conference on Computational Molecular Biology*, 277-86.

Snow D.H. & Valberg S.J. (1994) Muscle anatomy, physiology, and adaptations to exercise and training. In: *The Athletic Horse: Principles and Practice of Equine Sports Medicine.* (ed. by Hodgson DR & Rose RJ), pp. 145-79. W.B. Saunders, Philadelphia.

Sobczynska M. & Lukaszewicz M. (2004) Genetic parameters of racing merit of thoroughbred horses in Poland. *Journal of Animal Breeding and Genetics* **121**, 302-6.

Sonna L.A., Sharp M.A., Knapik J.J., Cullivan M., Angel K.C., Patton J.F. & Lilly C.M. (2001) Angiotensin-converting enzyme genotype and physical performance during US Army basic training. *Journal of Applied Physiology* **91**, 1355-63.

Stefansson H., Rujescu D., Cichon S., Pietilainen O.P.H., Ingason A., Steinberg S., Fossdal R., Sigurdsson E., Sigmundsson T., Buizer-Voskamp J.E., Hansen T., Jakobsen K.D., Muglia P., Francks C., Matthews P.M., Gylfason A., Halldorsson B.V., Gudbjartsson D., Thorgeirsson T.E., Sigurdsson A., Jonasdottir A., Jonasdottir A., Bjornsson A., Mattiasdottir S., Blondal T., Haraldsson M., Magnusdottir B.B., Giegling I., Moller H.-J., Hartmann A., Shianna K.V., Ge D., Need A.C., Crombie C., Fraser G., Walker N., Lonnqvist J., Suvisaari J., Tuulio-Henriksson A., Paunio T., Toulopoulou T., Bramon E., Di Forti M., Murray R., Ruggeri M., Vassos E., Tosato S., Walshe M., Li T., Vasilescu C., Muhleisen T.W., Wang A.G., Ullum H., Djurovic S., Melle I., Olesen J., Kiemeney L.A., Franke B., Sabatti C., Freimer N.B., Gulcher J.R., Thorsteinsdottir U., Kong A., Andreassen O.A., Ophoff R.A., Georgi A., Rietschel M., Werge T., Petursson H., Goldstein D.B., Nothen M.M., Peltonen L., Collier D.A., St Clair D. & Stefansson K. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-6.

Stefánsson S.E., Jónsson H., Ingvarsson T., Manolescu I., Jónsson H.H., Ólafsdóttir G., Pálsdóttir E., Stefánsdóttir G., Sveinbjörnsdóttir G. & Frigge M.L. (2003) Genomewide scan for hand osteoarthritis: A novel mutation in matrilin-3. *The American Journal of Human Genetics* **72**, 1448-59.

Svobodova S., Blouin C. & Langlois B. (2005) Estimation of genetic parameters of Thoroughbred racing performance in the Czech Republic. *Animal Research* **54**, 499-509.

Swinburne J., Gerstenberg C., Breen M., Aldridge V., Lockhart L., Marti E., Antczak D., Eggleston-Stott M., Bailey E., Mickelson J., Roed K., Lindgren G., von Haeringen W., Guerin G., Bjarnason J., Allen T. & Binns M. (2000a) First comprehensive low-density horse linkage map based on two 3-generation, full-sibling, cross-bred horse reference families. *Genomics* **66**, 123-34.

Swinburne J.E., Boursnell M., Hill G., Pettitt L., Allen T., Chowdhary B., Hasegawa T., Kurosawa M., Leeb T. & Mashima S. (2006) Single linkage group per chromosome genetic linkage map for the horse, based on two three-generation, full-sibling, crossbred horse reference families. *Genomics* **87**, 1-29.

Swinburne J.E., Lockhart L., Aldridge V., Marti E., Breen M. & Binns M.M. (2000b) Characterisation of 25 new physically mapped horse microsatellite loci: *AHT24++-48*. *Animal Genetics* **31**, 237-8.

Swinburne J.E., Turner A., Alexander L.J., Mickleson J.R. & Binns M.M. (2003) Characterization and linkage map assignments for 61 new horse microsatellite loci (*AHT49-109*). *Animal Genetics* **34**, 65.

Takahashi R., Yokoji H., Misawa H., Hayashi M., Hu J. & Deguchi T. (1994) A null mutation in the human *CNTF* gene is not causally related to neurological diseases. *Nature Genetics* **7**, 79-84.

Tallmadge R.L., Evans K.G., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999a) Equine dinucleotide repeat loci *COR081-COR100*. *Animal Genetics* **30**, 462-78.

Tallmadge R.L., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999b) Equine dinucleotide repeat loci *COR061-COR080*. *Animal Genetics* **30**, 462-6.

Tavernier A. (1988) Advantages of BLUP animal model for breeding value estimation in horses. *Livestock Production Science* **20**, 149-60.

Taylor R.R., Mamotte C.D.S., Fallon K. & van Bockxmeer F.M. (1999) Elite athletes and the gene for angiotensin-converting enzyme. *Journal of Applied Physiology* **87**, 1035-7.

Terry R.B., Archer S., Brooks S., Bernoco D. & Bailey E. (2004) Assignment of the appaloosa coat colour gene (LP) to equine chromosome 1. *Animal Genetics* **35**, 134-7.

Tolley E.A., Notter D.R. & Marlowe T.J. (1985) A review of the inheritance of racing performance in horses. *Animal Breeding Abstracts* **53**, 163-85.

Tozaki T., Kakoi H., Mashima S., Hirota K., Hasegawa T., Ishida N., Miura N., Choi-Miura N.H. & Tomita M. (2001a) Population study and validation of paternity testing for Thoroughbred horses by 15 microsatellite loci. *Journal of Veterinary Medical Science* **63**, 1191-7.

Tozaki T., Kakoi H., Mashima S., Hirota K., Hasegawa T., Ishida N., Miura N. & Tomita M. (2000a) The isolation and characterization of 18 equine microsatellite loci, TKY272-TKY289. *Animal Genetics* **31**, 149.

Tozaki T., Kakoi H., Mashima S., Hirota K., Hasegawa T., Ishida N., Miura N. & Tomita M. (2000b) The isolation and characterization of 34 equine microsatellite loci, TKY290-TKY323. *Animal Genetics* **31**, 234.

Tozaki T., Mashima S., Hirota K., Choi-Miura N.H.O. & Tomita M. (2001b) Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. *DNA Research* **8**, 33-45.

Tozaki T., Penedo M.C., Oliveira R.P., Katz J.P., Millon L.V., Ward T., Pettigrew D.C., Brault L.S., Tomita M., Kurosawa M., Hasegawa T. & Hirota K. (2004) Isolation, characterization and chromosome assignment of 341 newly isolated equine microsatellite markers. *Animal Genetics* **35**, 487-96.

Tozaki T., Sakagami M., Mashima S., Hirota K. & Mukoyama H. (1995) ECA-3: equine (CA) repeat polymorphism at chromosome 2p1. 3-4. *Animal Genetics* **26**, 283.

Tozaki T., Swinburne J., Hirota K., Hasegawa T., Ishida N. & Tobe T. (2007) Improved resolution of the comparative horse–human map: Investigating markers with in silico and linkage mapping approaches. *Gene* **392**, 181-6.

Tozaki T., Takezaki N., Hasegawa T., Ishida N., Kurosawa M., Tomita M., Saitou N. & Mukoyama H. (2003) Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a european horse outgroup. *Journal of Heredity* **94**, 374.

van Deursen J., Heerschap A., Oerlemans F., Rultenbeek W., Jap P., ter Laak H. & Wieringa B. (1993) Skeletal muscles of mice deficient in muscle creatine kinase lack burst activity. *Cell* **74**, 621-31.

van Haeringen W.A., van de Goor L.H.P., van der Hout N. & Lenstra J.A. (1998) Characterization of 24 equine microsatellite loci. *Animal Genetics* **29**, 153.

Vega-Pla J.L., Garrido J.J., Dorado G. & de Andres-Cara D.F. (1996) Three new polymorphic equine microsatellites: *HLM2, HLM3, HLM5*. *Animal Genetics* **27**, 215.

Wade C.M., Gnerre S., Garber M., Lander E.S. & Lindblad-Toh K.A. (2008) Preliminary analysis of the equine genome sequence. *Plant & Animal Genome XVI,* **W141**, 34.

Wagner M.L., Goh G., Wu J.T., Morrison L.Y., Alexander L.J., Raudsepp T., Skow L.C., Chowdhary B.P. & Mickelson J.R. (2004a) Sixty-seven new equine microsatellite loci assigned to the equine radiation hybrid map. *Animal Genetics* **35**, 484-6.

Wagner M.L., Goh G., Wu J.T., Raudsepp T., Morrison L.Y., Alexander L.J., Skow L.C., Chowdhary B.P. & Mickelson J.R. (2004b) Radiation hybrid mapping of 63 previously unreported equine microsatellite loci. *Animal Genetics* **35**, 159-62.

Wagner M.L., Goh G., Wu J.T., Raudsepp T., Morrison L.Y., Alexander L.J., Skow L.S., Chowdhary B.P. & Mickelson J.R. (2004c) Radiation hybrid mapping of 75 previously unreported equine microsatellite loci. *Animal Genetics* **35**, 68.

Wagner M.L., Raudsepp T., Goh G., Agarwala R., Schaffer A.A., Dranchak P.K., Brinkmeyer-Langford C., Skow L.C., Chowdhary B.P. & Mickelson J.R. (2006) A 1.3-Mb interval map of equine homologs of HSA2. *Cytogenetic and Genome Research* **112**, 227-34.

Wagoner L.E., Laura L., Zengel P., Nancy McGuire B.S.N., Rathz D.A., Ii G.W.D. & Liggett S.B. (2002) Polymorphisms of the beta-adrenergic receptor predict exercise capacity in heart failure. *American Heart Journal* **144**, 840-6.

Wang Y.-X., Zhang C.-L., Yu R.T., Cho H.K., Nelson M.C., Bayuga-Ocampo C.R., Ham J., Kang H. & Evans R.M. (2004) Regulation of muscle fiber type and running endurance by PPARδ. *PLoS Biology* **2**, e294.

Ways J.A., Cicila G.T., Garrett M.R. & Koch L.G. (2002) A genome scan for loci associated with aerobic running capacity in rats. *Genomics* **80**, 13-20.

Ways J.A., Smith B.M., Barbato J.C., Ramdath R.S., Pettee K.M., DeRaedt S.J., Allison D.C., Koch L.G., Lee S.J. & Cicila G.T. (2007) Congenic strains confirm aerobic running capacity quantitative trait loci on rat chromosome 16 and identify possible intermediate phenotypes. *Physiological Genomics* **29**, 91.

Whittemore A.S. & Halpern J. (1994) A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118-27.

Wigginton J.E. & Abecasis G.R. (2005) PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445-7.

Willett P. (1981) *The Classic Racehorse*. Stanley Paul, London.

Williams A.G., Rayson M.P., Jubb M., World M., Woods D.R., Hayward M., Martin J., Humphries S.E. & Montgomery H.E. (2000) The *ACE* gene and muscle performance. *Nature* **403,** 614.

Williamson S.A. & Beilharz R.G. (1998) The inheritance of speed, stamina and other racing performance characters in the Australian Thoroughbred;[Vererbung von Geschwindigkeit, Ausdauer und anderen Rennleistungscharakteristika beim Australischen Vollblut]. *Journal of Animal Breeding and Genetics* **115**, 1-16.

Wilson A.J. & Rambaut A. (2008) Breeding racehorses: what price good genes? *Biology Letters* **4**, 173-5.

Wilson D.E., Willham R.L., Buttram S.T., Hoekstra J.A. & Luecke G.R. (1988) Genetics of racing performance in the American Quarter Horse: IV. Evaluation using a reduced animal model with repeated records. *Journal of Animal Science* **66**, 2817.

Wittwer C., Dierks C., Hamann H. & Distl O. (2008) Associations between candidate gene markers at a quantitative trait locus on equine chromosome 4 responsible for osteochondrosis dissecans in fetlock joints of South German Coldblood Horses. *Journal of Heredity* **99**, 125-9.

Wittwer C., Lohring K., Drogemuller C., Hamann H., Rosenberger E. & Distl O. (2007) Mapping quantitative trait loci for osteochondrosis in fetlock and hock joints and palmar/plantar osseus fragments in fetlock joints of South German Coldblood Horses. *Animal Genetics* **38**, 350.

Woehlke A., Nordsiek G., Scharfe M., Jarek M., Schrader F., Wrede J., Pook J.P., Zhu B., Jong P.J.D., Chowdhary B.P., Leeb T., Blöcker H. & Distl O. (2008) Development of a BAC-based physical map of the horse genome. In: *Plant & Animal Genome XVI*, San Diego, CA, USA.

Wolfarth B., Bray M.S., Hagberg J.M., Perusse L., Rauramaa R., Rivera M.A., Roth S.M., Rankinen T. & Bouchard C. (2005) The human gene map for performance and health-related fitness phenotypes: the 2004 update. *Medicine & Science in Sports & Exercise* **37**, 881-903.

Wolfarth B., Rivera M.A., Oppert J.M., Boulay M.R., Dionne F.T., Chagnon M., Gagnon J., Chagnon Y., PÉRusse L. & Keul J. (2000) A polymorphism in the alpha2a-adrenoceptor gene and endurance athlete status. *Medicine & Science in Sports & Exercise* **32**, 1709.

Womack J.E. (2005) Advances in livestock genomics: Opening the barn door. *Genome Research* **15**, 1699-705.

Woods D.R., Humphries S.E. & Montgomery H.E. (2000) The *ACE* I/D polymorphism and human physical performance. *Trends in Endocrinology & Metabolism* **11**, 416-20.

Yang N., MacArthur D.G., Gulbin J.P., Hahn A.G., Beggs A.H., Easteal S. & North K. (2003) *ACTN3* genotype is associated with human elite athletic performance. *The American Journal of Human Genetics* **73**, 627-31.

Yang F., Fu B., O'Brien P.C.M., Nie W., Ryder O.A. & Ferguson-Smith M.A. (2004) Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. *Chromosome Res* **12**, 65-76.

Yonan A.L., Alarcón M., Cheng R., Magnusson P.K.E., Spence S.J., Palmer A.A., Grunn A., Hank Juo S.H., Terwilliger J.D. & Liu J. (2003) A genome wide screen of 345

families for autism-susceptibility loci. *The American Journal of Human Genetics* **73**, 886-97.

APPENDIX A

Table A-1 Human candidate genes mapped in this study, primers and closest microsatellite in genome scanning panel

Gene symbol	Name	Chr No.	Closest microsatellite in panel	Distance to microsatellite (EquCab2)	Primers 5' > 3'	Size (bp)	Temp	$MgCl_2$ (mM)
GPR10 (aka PRLHR)	G protein-coupled receptor 10	1	1CA30	292 kb	F:GCTGATCGTGCTGCTCTACA R:AGCAGGTTGAAGACGTGCAG	717	58	1.5
AGT	Angiotensinogen	1	AGTms	196 kb	F:CAGGTGACCGGGTGTACATA R:GGAAGTGGACGTAGGTGTTGA	759	54	1.5
RYR2	Ryanodine receptor 2 (cardiac)	1	No ms adjacent		F:AGCTTTGAGTGCTACAGACATGG R:GCTGAGCTTTGGTAAGTGAACAG	182	58	1.5
LIPC	Lipase, hepatic	1	UCDEQ440	2.8 Mb	F:CCTCTTCATCGACTCCTTGCT R:GAGGCTCTTGCTCTTCTTGC	169	58	1.5
CYP19A1	Cytochrome P450, family 19	1	HMS15	2.3 Mb	F:TGAAGGTGATGCTGGTTTCA R:TGTTCGAGGCACTTTTCTGA	169	58	3
CPT2	Carnitine palmitoyltransferase II	2	No ms adjacent		F:CCAGATGGCCTTTCTGAGG R:CTGAAGCTCCGCAGCACT	170	58	1.5
FGA	Fibrinogen, A alpha polypeptide	2	TKY850	910 kb	F:GACTCTGTCCTTCGGGTTGA R:AGCCTCCTCCGTAGACTTCC	259	58	1.5
FGB	Fibrinogen, B beta polypeptide	2	TKY850	910 kb	F:TCGGAAATGACAGAATTAGCC R:GGTTTTCTCCCACCAGTTGA	213	50	2
UCP1	Uncoupling protein 1	2	UCP1ms	63 kb	F:AACAACAGAAGGCTTGACGG R:ACCAGGGCCTCCTTCATTAG	220	58	1.5
FABP2	Fatty acid binding protein 2	2	FABP2ms	60 kb	F:CGGTTAGACAACGGAAATGAA R:TCCAAGAACAATGCTCACTCC	572	58	1.5
HP	Haptoglobin	3	TKY1085	705 kb	F:CCCATCTGCCTACCTTCAAA R:TACCAGGTGTCGTCCTCCTC	320	58	1.5
PPARGC1A	Peroxisome proliferative activated receptor, gamma, coactivator 1, alpha	3	PPARGC1Ams	Intron 3	F:AGACCAGTGAACTGAGGGACA R:GATTTGGGTGGTGACACAGA	229	60	1.5
PGAM2	Phosphoglycerate mutase 2 (muscle)	4	TKY942	767 kb	F:CAAGGACGCAAAGATGGAAT R:GTGGTGGGATGTCAAAGGAG	236	55	1.5
IGFBP1	Insulin-like growth factor binding protein 1	4	TKY942	617 kb	F:GAGCAGCTCCTGGACAGTTT R:CCACTTTTTGATGTCGGAGA	123	58	3
PON1	Paraoxonase 1	4	PON1ms	82 kb	F:GGTGAACCATCCAGATTTTAAG R:GGCAGAAGTTTGTGTCTGATG	102	55	1.5
PON2	Paraoxonase 2	4	PON1ms	82 kb	F:ATGGTATCAGCACTTTCATAGACA R:AAGCTCATGTTTGATTGTTTTTAGA	678	55	1.5

Table A-1 (continued)

NPY	Neuropeptide Y	4	IL6ms	1.2 Mb	F:AAGCGACTGGGGTTGTCC R:GTGTCGCAGCGCTGAGTAGT	150	58	1.5
ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	5	TKY508	702 kb	F:CATGCTGGTGCTGTGTCTTT R:TCCATTGTTTTAACCTGACTGAA	300	58	1.5
APOA2	Apolipoprotein A-II	5	TKY521	1.1 Mb	F:GAGCTTTGGTTCGGAGACAG R:TGTCAGCTGCTCCTTTGACTT	586	58	1.5
ATP1A2	ATPase, Na+/K+ transporting, alph-2 polypeptide	5	TKY521	417 kb	F:CTGCCATCTCCTTGGCCTAT R:AGGTGAAGAAGCCACCCAGT	352	60	1.5
S100A1	S100 calcium binding protein A1	5	AHT050	387 kb	F:GTGGACAAGGTGATGAAGGAC R:TGTTACAGGCCACTGTGAGG	100	58	1.5
CASQ2	Calsequestrin 2	5	AMPD1ms	1.0 Mb	F:CCTTGAAGATGAATGAGGTTGA R:GGTGCTCTTTCACAAACTCCA	111	56	1.5
AMPD1	Adenosine monophosphate deaminase	5	AMPD1ms	79 kb	F:CATCATGGTGCTCAACAACC R:TTTTAAATTCAGGCCATGAGAGA	900	58	1.5
LEPR	Leptin receptor	5	LEPRms	11 kb	F:CCGTTGTTCGTCCTGATCTT R:AATGATAGGCAGTCCGAAGG	159	58	3
GNB3	Guanine nucleotide binding protein (G protein), beta polypeptide 3	6	GNB3ms	22 kb	F:CTCTGTAGCCAGGGCAGTGT R:GGAGGCTGTCACAGAAAAGC	173	66	1.5
VDR	Vitamin D (1,25-dihydroxyvitamin D3) receptor	6	VDRms	11 kb	F:AGGACCAGATCGTACTGCTGA R:AGTCTTGGTTGCCACAGGTC	108	56	1.5
UCP2	Uncoupling protein 2	7	UCP2ms	137 kb	F:GACCCCAAGCCTTCTACAAA R:CCTCTTCAGCTGCTCGTAGG	464	52	1.5
UCP3	Uncoupling protein 3	7	UCP2ms	158 kb	F:GATGAGCTTCGCCTCCATC R:GCCAAAATCCGGGTAATGAT	337	60	1.5
LDHA	Lactate dehydrogenase A	7	AHT019	942 kb	F:TGCACACTCCAAGCTGGTTA R:TGGATTGGAAACAACAAGCAA	160	58	1.5
PNMT	Phenylethanolamine N-methyltransferase	11	AHT076	813 kb	F:GTGTGTACAGCCAGCACGTC R:AGTGTGGTGATGTGGTCCAG	355	58	1.5
SGCA	Sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)	11	AHT076	4.0 Mb	F:AGGAAGTGTTGCCCTCGTC R:AGGAGAAGGGTGAGGCAGAG	540	66	1.5
CNTF	Ciliary neurotrophic factor	12	TKY404	1.2 Mb	F:AGCTTACCGTACCTTCCATGTTAT R:AGCTCTTGCAGCACCTTCAG	261	58	1.5

Table A-1 (continued)

SCGB1A1	Secretoglobin, family 1A, member 1	12	TKY404	2.0 Mb	F:CTGAAGACGCTGGTGGACTT R:CAGCTTCTAAATCCTAAGCACACA	450	58	1.5
PYGM	Phosphorylase, glycogen; muscle	12	UCDEQ411	54 kb	F:CCCTGACAGACCAGGAGAAG R:GCCAGCGCGAAGTAGTAGTC	160	66	1.5
ACTN3	Actinin, alpha 3	12	ACTN3ms	27 kb	F:GGAGGACTTTCGGGACTACC R:GAGCAGCCAGTCCTCATAGC	308	58	1.5
SERPINE1	Plasminogen activator inhibitor-1	13	TKY585	1.2 Mb	F:ACATGTTCAGGCCAAACCAG R:CGGTCACCTCGATCTTCACT	219	58	1.5
PPARG	Peroxisome proliferative activated receptor, gamma	16	AHT037	954 kb	F:GGGTGTCAGTTTCGCTCAGT R:CACCAAAAGGCTTTCTCAGG	235	62	3
SGCG	Gamma sarcoglycan	17	COR105	817 kb	F:AGTTGGTTCAGGGAACTTGG R:GATCCGTGGAAGATGCAGTT	163	58	2
CASR	Calcium-sensing receptor	19	TKY730	1.8 Mb	F:GGCACAATTGGATTTGCTTT R:CTCCCAGGTGAGCAGGTATAG	377	58	1.5
TNF	Tumor necrosis factor	20	UM011	2.2 Mb	F:CTTCTCGAACCCCAAGTGAC R:GAGACAGCTAAGCGGCTGAT	548	58	1.5
PLCG1	Phospholipase C gamma 1	22	SGCV01	292 kb	F:GCTTTGGCCACAATTTTCAT R:TGCCCCTAGTGAGAAGCTGT	242	58	1.5
CNTFR	Ciliary neurotrophic factor receptor	23	SGCV04	1.4 Mb	F:CCCAACACCTTCAATGTGACT R:TGATAGCTGTGGCATTGTGG	346	58	1.5
BDKRB2	Bradykinin receptor B2	24	COR024	2.3 Mb	F:ACTTCCTCATGCTGGTGAGC R:TCTGCATCTCGTTGTTACGC	336	58	1.5
ADRB3	Beta-3-adrenergic receptor	27	ADRB3ms	28 kb	F:GCAACCTGCTGGTAATCGTG R:CTCATGATGGGCGCAAAC	364	56	1.5
PPARA	Peroxisome proliferative activated receptor, alpha	28	UCDEQ425	991 kb	F:AGTGCACGTCCGTAGAGACC R:TCACTGTCGTCCAGTTCCAG	292	58	1.5
GK	Glycerol kinase	X	GKms	64 kb	F:CTTCTTATGGCTGCTACTTCGTC R:GCAATATGGCATTTATTGGTGA	613	58	4

Data published in (Raudsepp *et al.* 2008a)

Table A-2 Human candidate genes mapped by others

Gene symbol	Name	Chr No.	Location EquCab2 Mb	Human	Location Human Mar. 2006	Closest Microsatellite	Distance	Reference
ADRB1	Adrenergic, beta-1-, receptor	1	17	10q25.3	115.7	LEX039	1.6 Mb	EquCab2
ADRA2A	Alpha-2A-adrenergic receptor	1	19.6	10q25.2	112.8	LEX020	226 kb	EquCab2
ANG	Angiogenin monophosphate, Rnase A family,5	1	157.6	14q11.1-q11.2	20.2	1CA16	521 kb	EquCab2
DI01	Delodinase, iodothyronine, type I	2	5.55	1p33p32	54.1	No ms adjacent		EquCab2
LPL	Lipoprotein lipase	2	49	8p22	19.8	No ms adjacent		(Chowdhary *et al.* 2003)
CETP	Cholesteryl ester transfer protein, plasma	3	*9.2*	*16q21*	*55.5*	CETPms	197 kb	EquCab2
TK2	Thymidine kinase 2, mitochondrial	3	17	16q22q23.1	65.1	COR033*	3.54 Mb	EquCab2
IGFBP3	Insulinlike growth factor binding protein 3	4	16.3	7p13p12	45.9	TKY942	600 kb	EquCab2
IL6	Interleukin 6	4	54.4	7p21	22.7	IL6ms	2 kb	(Musilova *et al.* 2005)
CFTR	Cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)	4	74.5	7q31.2	116.9	HMS19	3.9 Mb	(Chowdhary *et al.* 2003)
LEP	Leptin	4	83.4	7q31.3	127.6	HMS09	1 Mb	(Caetano *et al.* 1999)
NOS3	Nitric oxide synthase 3 (endothelial cell)	4	102.6	7q36	150.3	NOS3ms	117 kb	(Perrocheau *et al.* 2006)
PFKM	Phosphofructokinase, muscle	6	65.77	12q13.3	46.8	COR070	80 kb	(Chowdhary *et al.* 2003)
DRD2	Dopamine receptor D2	7	*21.8*	*11q23*	*112.78*	APOC3ms	3.0 Mb	(Chowdhary *et al.* 2003)
APOC3	Apolipoprotein C-III	7	24.8	11q23.1-q23.2	116.2	APOC3ms	10 kb	(Chowdhary *et al.* 2003)
APOA1	Apolipoprotein AI	7	21.53	11q23	116.2	APOC3ms	3.3 Mb	EquCab2
SUR aka ABCC8	Sulfonylurea receptor	7	85.6	11p15.1	17.4	AHT019	Exon 11	(Raudsepp *et al.* 2008a)
COMT	Catechol-O-methyltransferase	8	*Un*	*22q11.21*	*18.3*	AHT005	218 kb	(Momozawa *et al.* 2005)

Table A-2 (continued)

LIPG	Lipase, endothelial	8	68.1	18q21.1	45.3	COR003	3.8 Mb	EquCab2
MC4R	Melanocortin 4 receptor	8	77.1	18q22	56.1	No ms adjacent		EquCab2
TGFB1	Transforming growth factor, beta 1	10	11.8	19q13.2	46.5	NVHEQ018	3.5 Mb	(Perrocheau *et al.* 2005)
APOE	Apolipoprotein E	10	15.5	19q13.31	50.1	NVHEQ018	145 kb	EquCab2
CKM	Creatine kinase, muscle	10	15.8	19q13.2-q13.3	50.5	NVHEQ018	515 kb	(Chowdhary *et al.* 2003)
ACE	Angiotensin I converting enzyme	11	15.8	17q23	58.9	ACEms	133 kb	(Milenkovic *et al.* 2002)
BRCA1	Breast cancer 1, earley onset	11	20	17q21	38.4	SGCV24	520 kb	(Raudsepp *et al.* 2008a)
COL1A1	Collagen, type I, alpha 1	11	25.9	17q21.3-q22.1	45.6	AHT076	4.0 Mb	(Raudsepp *et al.* 2008a)
ENO3	Enolase 3	11	49.5	17pter-p11	4.7	ACADVLms	449 kb	(Chowdhary *et al.* 2003)
ACADVL	Acyl coenzyme A dehydrogenase, very long chain	11	50.2	17p13-p11	7.06	ACADVLms	221 kb	EquCab2
IGF2	Insulin-like growth factor 2	12	30.6	11p15.5	2.11	COR058	2.6 Mb	(Raudsepp *et al.* 1997)
KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	12	30.8	11p15.5	2.4	AHT017	230 kb	EquCab2
ADRB2	Beta-2-adrenergic receptor	14	28.9	5q31-q32	148.18	ADRB2ms	38 kb	(Chowdhary *et al.* 2003)
NR3C1	Nuclear receptor subfamily 3, group C, member 1	14	*33.8*	*5p31*	*142.63*	UM032	5.2 Mb	(Goh *et al.* 2007)
AGTR1	Angiotensin II receptor, type 1	16	78.8	3q21q25	149.9	TKY936	5.1 Mb	EquCab2
BRCA2	Breast cancer 2, earley onset	17	11.5	13q12.3	31.78	COR007	4.8 Mb	(Chowdhary *et al.* 2003)
TTN	Titin	18	56.9	2q31	179	HTG17	973 kb	(Chowdhary *et al.* 2003)
GDF8(MSTN)	Growth differentiation factor 8	18	66.4	2q32.2	190.6	GDF8ms	6 kb	(Chowdhary *et al.* 2003)
MYLK	Myosin, light polypeptide kinase	19	31.9	3q21	124.8	TKY730	3.4 Mb	EquCab2
EDN1	Endothelin 1	20	12	6p24.1	12.39	AHT018	1.9 Mb	(Caetano *et al.* 1999)
HLA-A	Major histocompatibility complex, class I, A	20	29	6p21.3	30.01	UMNe056	28 kb	(Chowdhary *et al.* 2003)

Table A-2 (continued)

HIF1A	Hypoxia-inducible factor 1, alpha subunit	24	8.9	14q21-q24	61.23	No ms adjacent		(Chowdhary *et al.* 2003)
SLC25A4	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	27	24.9	4q35	186.3	TKY480	0.5 Mb	(Raudsepp *et al.* 2008a)
IGF1	Insulin-like growth factor I	28	26.1	12q22-q23	101.3	TKY872	2.4 Mb	(Chowdhary *et al.* 2003)
IL15RA	Interleukin 15 receptor, alpha.	29	*27.9*	*10p15.1*	*6.03*	L12.2	2.6 Mb	EquCab2
ADIPOR1	Adiponectin receptor 1	30	29.7	1q32	201.1	LEX075	2.83 Mb	EquCab2
ESR1	Estrogen receptor 1	31	15	6q25.1	152.1	TKY105	5.4 Mb	(Raudsepp *et al.* 2008a)
STS	Steryl-sulfatase precursor	X	*3.8*	*Xp22.32*	*7.14*	AHT099	729 kb	(Raudsepp *et al.* 2002)
PHKA1	Phosphorylase kinase, alpha 1 (muscle)	X	54.1	Xq13.1	71.7	UMNe402	1.57 Mb	(Raudsepp *et al.* 2004a)
PGK1	Phosphoglycerete kinase 1	X	58	Xq13	76.12	PGK1ms	99 kb	(Chowdhary *et al.* 2003)
LAMP2	Lysosomalassociated membrane protein 2	X	un	Xq24	119.4	No ms adjacent		(Raudsepp *et al.* 2004a)

EquCab2 refers to the Sep. 2007 Equus caballus draft genome assembly

Table A-3 Microsatellites that showed no amplification with the M13 primer

Marker	Chr No.	Size	Temp	$MgCl_2$ (mM)	Primers 5' > 3'	Comments	Reference
1CA20	1	194	50	1.5	F:TGGACAAAATGCAAAAGTCA R:TCCACTACACAGGAAAACGAA	F:In DNA element	(Chowdhary *et al.* 2003)
ASB012	1	171	65	3	F:TCAGCAATAGAAGCCAGCTCC R:TCCTATGGAGGTGACCTTCCC		(Breen *et al.* 1997)
TKY899	1	169	58	3	F:AGCAACAGAGTAATGCCAAG R:TAGGCGGGTTTTAAACATGG		(Tozaki *et al.* 2004)
UM041	1	105-115	58	2	F:TGCCCTTCCATGAACAGAC R:TCCCTCTCTCTCTCTCCTTCTC		(Swinburne *et al.* 2000a)
UMNe115	1	166	58	3	F:TCCTCCTACACTGGCCATATC R:TTTCCTATCGGAGTGCTTGC		(Wagner *et al.* 2004c)
TKY318	1	152-170	58	1.5	F:ACAGAAGTGGAGATGGTTTG R:CCACAGGTTCTACTGTCTTG		(Tozaki *et al.* 2000b)
UMNe336	2	119	58	1.5	F:ACCTTCTCTTTGGAAGGAAATG R:GAGACCTAAGCATGGAGCATG		(Mickelson *et al.* 2004)
UMNe374	2	194	50	3	F:AATTTCATATGGTTTCCATGCC R:TAAGCACTGCGTTAATGTTCTG		(Wagner *et al.* 2004c)
UMNe563	2	125	55	1.5	F:CTTTCACCCCAGCACTTCTC R:TTCTCCATTAAACACACGCG		(Wagner *et al.* 2004b)
AHT022	3	189-197	58	2	F:AAGCACAATGTGGOSGTTAG R:TCCACGTTCACACATACCTCA	No sequence available	(Swinburne *et al.* 2000a)
AHT013	4	142-146	58	2	F:CTTCCTCAGGTGCATAGGTTG R:TCATTAAAATACAACCTGCCCC		(Swinburne *et al.* 2000a)
TKY430	4	156	50	3	F:TTTTGGTTAGGCCTCTGTAC R:CCCACTCTCCTATAAACAGT	No sequence available	(Swinburne *et al.* 2000a)
TKY337	4	200	53	1.5	F:AGCAGGGTTTAATTACCGAG R:TAGATGCTAATGCAGCACAG		(Tozaki *et al.* 2001b)
TKY271	5	112-124	56	1.5	F:CAGTGAAGAGTGAATGGATA R:GCAAATGGCAGAATTTCCTT	F & R: In LINE	(Tozaki *et al.* 2001a)
UMNe481	5	130	55	1.5	F:ACATTGGGCCAGTTGCTTAC R:ATTGTTGGGACACCAGATTACC		(Wagner *et al.* 2004b)
TKY1113	6	206	58	2	F:GATCAGAGGGAGGAAAGAG R:AAGAGATGAACAGAATCAGTG		(Raudsepp *et al.* 2008a)
TKY305	7	128-134	55	1.5	F:CCTCCCTCATCATGTAAGTC R:TTATTCAGGGTTCTCCAGAG	F & R: In LTR	(Tozaki *et al.* 2000b)
HLM2	11	123-137	58	1.5	F:CCCACCTCCCCATCTCCCAACC R:AAGCCAGTTCCTCAGCCCCACC		(Vega-Pla *et al.* 1996)

Table A-3 (continued)

TKY626	11	237-239	56	3	F:CCAGCTGAGCCCATTTTAGA R:GGGAGGAGCTGAGAAGATTTG	R: In SINEs	(Tozaki *et al.* 2004)
AHT027	12	101-109	58	2	F:GAGCAAGTCTTCCTTCCACG R:ACGTGCACACCTGTACGTGT		(Swinburne *et al.* 2000b)
TKY499	12	122-178	55	1.5	F:GATGCCCTCTGGCTAGTGTT R:TAGAGACCCACCTGCTGCTT		(Tozaki *et al.* 2004)
UMNe331	12	158	64	1.5	F:TAGCCATGATATGGAAACAACC R:ACTGTCCATCCATGTTGCTG	F & R: In LINE	(Raudsepp *et al.* 2008a)
HBA1	13	170	50	2	F:CTGTCGAATCTGAGCGACCT R:GTGGATCGAGGGAGTGTCAG	No sequence available	(Chowdhary *et al.* 2003)
UMNe305	13	162	58	2	F:ACAGTTCACAGCGGCCTC R:GAGAGGGAGCAAGCCCTC		(Wagner *et al.* 2004c)
VHL161	13	160-164	58	2	F:GACCCAACATTGTATGTAAGCG R:TGTGAATAGATATAAGTTGCTTGC	F: In LTR	(van Haeringen *et al.* 1998)
TKY267	14	97-113	50	1.5	F:CCACTGCCAAATGAAACAAA R:CCACACATTTCAGGAAAGAA		(Kakoi *et al.* 2000)
VHL204	14	100	62	1.5	F:ACTGAAGTTGAGAATCATTAATGG R:ACTTCCTCGACATCCTTCCCT	F & R: In LINE	(van Haeringen *et al.* 1998)
COR081	15	102-114	58	1.5	F:AACTTTCGTGTCACTCCTACACT R:GCATCTCTTCCACATCAAGG		(Tallmadge *et al.* 1999a)
SGCV06	15	167	58	3	F:GGGCCTGGTTTTCCTTCTAA R:GCATTTGTGGCCTGTGTCATA	R: In DNA element	(Godard *et al.* 1997)
TKY861	15	241	58	1.5	F:GACTCCAGTGTGTTTAGAGG R:CAGGACTGCCACAGCTTC		(Raudsepp *et al.* 2008a)
HTG13	16	120	58	2	F:TTAGCACGGGGAGATCGGATCCTG R:GGTCTCCCTCTCCATTCACCCTGC		(Marklund *et al.* 1994)
NVHEQ216	16	142-144	50	4	F:TTGGACTACTGTATTTGGTC R:TGATCTCTCCATAGCATAAT		(Penedo *et al.* 2005)
COR101	18	232-272	58	2	F:CAATATAAGTGCACGCCTTC R:TGGACCTTGAGGGTATGATG	F: In SINE	(Swinburne *et al.* 2000a)
TKY322	18	203-224	54	3	F:TGCAAACACTTGTGAACTGC R:AACCTAGTGTAATTGCTACC		(Tozaki *et al.* 2000b)
UCD387	18	78-88	58	3	F:ACCCCCGCCCCAGCAC R:TGCCCCGTCATTCTGC		(Eggleston-Stott *et al.* 1997)
TKY538	19	151-165	58	1.5	F:TGGATGGAAGGGAAACAAGA R:TTGAAGAGGGTGGAGCAAAT		(Tozaki *et al.* 2004)
TKY817	19	155	54	3	F:TATAAACAAGGGCAGAGCGA R:GAGTGAATTCCCTGCTTCTC		(Tozaki *et al.* 2004)

Table A-3 (continued)

VHL137	20	149	57	1.5	F:CCCAAACATTTTTGACTGATGC R:TTATTTCTAAAGGGGTACGGCT		(van Haeringen *et al.* 1998)
21CA001	21	295-311	58	1.5	F:TCATGCTGAAATCAATGCAAA R:TTCCTTATGGCAACAACACG	F & R: In DNA element	(Raudsepp *et al.* 2008a)
HTG10	21	93	54	2	F:CAATTCCCGCCCCACCCCCGGCA R:TTTTTATTCTGATCTGTCACATTT		(Marklund *et al.* 1994)
HTG32	21	150-160	58	3	F:CCTGAAACCTCAGTAAACAGA R:TGTGGCTTTGGTGTGGAAAC		(Lindgren 2000)
SGCV14	21	188	58	3	F:CCCCAGTGGTTCCATTTAGATGT R:GGGGAGAGCATTTTGGTGA		(Godard *et al.* 1997)
TKY306	21	266-279	56	1.5	F:GTTTGTGGTGCTTTGTTAGC R:CTCTGCACTTGCTGAACATC	F: In LINE; R in LTR	(Tozaki *et al.* 2000b)
COR022	22	254-264	58	2	F:AAGACGTGATGGGAAATCAA R:AGAAAGTTTTCAAATGTGCCA		(Murphie *et al.* 1999)
UM012	24	100-110	58	3	F:GGAATTTAGACTATGACTGAGG R:GCCACCTGAACACTTTTAC		(Meyer *et al.* 1997)
UM005	26	207-222	50	1.5	F:CCCTACCTGAAATGAGAATTG R:GGCAAAAGATCAGGCCAT		(Meyer *et al.* 1997)
TKY425	28	109-125	58	1.5	F:CCTGGGTGTCGTGTGTTTTA R:TTCCTCTCTCCTGCCTCATC		(Tozaki *et al.* 2004)
TKY515	28	145-151	50	4	F:AGGCCAGCAGTGTTCCTCTA R:GGCTCAGGTACGTTCCCTTC	F:In LTR	(Tozaki *et al.* 2004)
TKY715	29	224-246	58	1.5	F:CAGTTTCACAGGAGAGAGAGTCC R:CTGGAGTCCCACCTCCAAC		(Tozaki *et al.* 2004)
VIAS-H39	29	154-158	56	2	F:AATGTGATTATAGCAGATAGGGTT R:CTATCCAATCTTCACAATCATGTA	F & R: In LTR	(Ewen & Matthews 1994b)
UMNe060	X	146-154	55	2	F:TGTGGCAGGAAAAACACATG R:CCATAATCCATGAGCCTATTCC	F & R: In LTR	(Roberts *et al.* 2000)

Table A-4 Markers not used due to excessive stutter or poor amplification

Marker	Chr No.	Size	No. alleles	Temp	$MgCl_2$ (mM)	Primers 5' > 3'	Comments	Reference
VHL134	1	157-163	3	58	1.5	F:CTGGGAACAGAATCAAGACTTG R:TAATATGCATGTATCTGATAGCTC	F & R: In LINE	(van Haeringen *et al.* 1998)
AHT009	4	~190	3	58	4	F:TAACCATGTCCCTGCAATGA R:TCAGAACTGTCCTTGTGAAAGC		(Raudsepp *et al.* 2008a)
UMNe477	5	259-263	4	58	1.5	F:GTCAAGGAACAGCTGAAGGTG R:TTCCTGAGCTAGGGTAGGAGC		(Raudsepp *et al.* 2008a)
TKY911	5	150-166	5	54	3	F:GATCTTTAGAATCAGCTTGTTG R:CTCGCCACGTTAGTTGATG	F & R: In LINE	(Tozaki *et al.* 2004)
TKY028	6	276	3	58	1.5	F:TTCAGCAGGGTCTCATGCCAC R:TTCGGCTCTGGTTCAAGAGG		(Kakoi *et al.* 2000)
AHT025	8	193-199	3	50	4	F:TCCCACATGCCACAACTAGA R:TTCCCCCTTGCTTTTAAAAC	F & R: In LTR	(Swinburne *et al.* 2000b)
COR056	8	201	4	58	1.5	F:AGATTCCAGGCATTAGGACC R:TCAGGGACAATCTTCCTCAAG	F & R: In LTR	(Ruth *et al.* 1999)
UM033	8	143-159	4	54	3	F:CATTGTCCTGAGCAAGTC R:CTATCCGTCAAGTGTTTC	F & R: In LTR	(George *et al.* 1998)
HTG04	9	129	3	58	1.5	F:CTATCTCAGTCTTGATTGCAGGAC R:CTCCCTCCCTCCCTCTGTTCTC		(Ellegren *et al.* 1992)
COR020	10	173-189	6	58	1.5	F:TCTCTACCGCAAGTGAAACC R:CTGAATTGTAGGACATCCCG	F & R: In LTR	(Hopman *et al.* 1999)
COR045	10	150	3	58	3	F:TCTCTACCGCAAGTGAAACC R:CTGAGCCCTTAACTTGTGGA	F & R: In LTR	(Ruth *et al.* 1999)
UMNe217	13	172-180	4	58	2	F:CTTTGAGTTCACCAGTTCTCCC R:AACCAAAAGGAACTTTGGTGG	F & R: In LTR	(Wagner *et al.* 2004c)
COR104	14	187-201	3	58	3	F:GGGAGTGTGTCCAGTTTGTC R:CCAGATAAAGCCCAAATCCT	F & R: In LINE	(Swinburne *et al.* 2000a)
HTG03	16	131-143	4	58	1.5	F:TAACCTGGGTGCAAAGCCACCCAT R:TCAGGGCCAATCTTCCTCAC	F: In SINE	(Ellegren *et al.* 1992)
I-18	16	94	3	58	1.5	F:CAACAAAGATGTTGCAAGGG R:TGTGCCTCTTGTCTCTTAGG	F & R: In LINE	(Marti *et al.* 1998)
LEX056	16	226-244	4	58	1.5	F:GACCTACAGGCCACTCATCAA R:GGCAGTTTCCTCCATCCTTA	F & R: In LINE	(Coogle & Bailey 1997)
TKY802	16	117-125	3	58	2	F:GGAGGAAAGAGCTGAATGGA R:TCATCCAGTTCACCCATCAG	F: In LTR; R: In SINE	(Tozaki *et al.* 2004)
TKY356	17	175-189	4	54	3	F:TCAATCACACGAGCCAATTC R:GATCATAGGATGATTAGCAC	F & R: In LTR	(Tozaki *et al.* 2001b)

Table A-4 (continued)

LEX054	18	182-194	3	58	1.5	F:TGCATGAGCCAATTCCTTAT R:TGGACAGATGACAGCAGTTC	F & R: In LTR	(Coogle & Bailey 1997)
HTG23	19	196-208	5	58	2	F:GTCCTTCAGAGTTGTCCCTG R:GGAGAACAACCTTGCCTGAG	F & R: In LTR	(Lindgren 2000)
I-12	19	178-180	3	54	2	F:AACTAAGCACGTCATACAAG R:CTTGTAGTTTTCGTTGTATAGC	F & R: In LTR	(Marti *et al.* 1998)
LEX035	19	272-288	3	58	1.5	F:CCCAGCATATCAAAGATGTT R:GCTCAGTGTACTTCAAGCAG	F & R: In LINE	(Coogle *et al.* 1997)
SGCV16	21	165-175	3	55	1.5	F:AATTCTCAAATGGTTCAGTGA R:CTCCCTCCCTTCCTTCTA	F & R: In LINE	(Godard *et al.* 1997)
LEX063*	23	257-265	3	58	1.5	F:CGGGGTGTGCATCTCTTAGG R:TGGCGAATGCTGAATCTGG	R: Tm too high	(Coogle & Bailey 1997)
AHT007	25	139-141	3	58	3	F:CCTTAGATCCGAGAAGGAGA R:GAAGCCTCACTCCATCCAGG	R: In SINE	(Swinburne *et al.* 2000a)
UCD405	25	286-292	3	54	2	F:ACCTCGTCTGGCTGTTGTAAG R:ACTTGCTGTGCGACTCTG	F & R: In DNA element	(Eggleston-Stott *et al.* 1997)
VHL20	30	105-114	3	58	1.5	F:CAAGTCCTCTTACTTGAAGACTAG R:AACTCAGGGAGAATCTTCCTCAG	F & R: In SINE	(Lindgren *et al.* 1998)
LEX022	X	107	6	58	1.5	F:AACATATCCATCGCCTCACA R:TGCAAATTCACTGAGAGTGG	F & R: In LINE	(Coogle *et al.* 1996b)

Table A-5 Microsatellites not used due to small number of alleles in the Thoroughbred

Marker	Chr No.	Size	No. alleles	Temp	$MgCl_2$ (mM)	Primers 5' > 3'	Comments	Reference
1CA18	1	140-142	2	50	2	F:AGAAAGCAAGTTCGCTAGATGG R:AGTTCCCAAGGAATGTGTGTG		(Swinburne *et al.* 2000a)
ASB041	1	155-167	2	58	1.5	F:AAAGTTCACTTAGTCCTTGG R:CCACCTGTTTGCACTTGC		(Swinburne *et al.* 2000a)
COR100	1	195-227	2	58	1.5	F:CCCAGAGGTTTCAGAGGG R:ATTCTAGGGCATATTATGACAA		(Tallmadge *et al.* 1999a)
LEX030	1	173-179	2	56	2	F:GGAGGGTGCAAGGTGCTA R:GGCAGGTCAGAAGGGACA		(Coogle *et al.* 1996a)
SGCV02	1	123	2	58	2	F:CCTTGAGTTGCACTTGGAGTC R:CTGCTCATATACCTGGGGATT		(Godard *et al.* 1997)
TKY015	1	149	2	58	1.5	F:GGATTTTAGAAGTACAGAGGG R:CATCCTACTGAGAACAATGCG		(Hirota *et al.* 2001)
UMNe429	1	271	2	58	3	F:ATGTAGTTCCCCCAGCCTG R:TCCCTCCCTCCACTCTCC	F: In LINE	(Wagner *et al.* 2004a)
VIAS-H34	1	144-160	2	50	1.5	F:TGAGTGTTTGCGTGTGTGTG R:TCCCGTCTCCTCTCTTGTTC		(Ewen & Matthews 1994a)
NVHEQ100	1	214-230	2	58	1.5	F:CCAAAGCAGAACATGTGAAGTT R:TGGCATAGATGTTAGCTCAGTGA	F & R: In SINE	(Roed *et al.* 1998)
TKY1142a	2	301	2	58	2	F:CTACATTGAACATCTATTGCTC R:AAGAATGCCCCCTCATATAG		(Raudsepp *et al.* 2008a)
UM007	2	125-159	2	58	1.5	F:GGGAATAGAGAAAGGTGAAG R:TTAGAGTTCCTGCTCCTCC		(Meyer *et al.* 1997)
UMNe205	2	187	2	58	3	F:TGGAGAAAAGGCTGATTCTAGG R:GCCATGGAAACCATGGAG	F & R: In LINE	(Mickelson *et al.* 2003)
VHL123	2	151	2	58	1.5	F:CCTCCTTCACAGTGAAGTGC R:GAGTATATAGCTCCAGACCTC		(van Haeringen *et al.* 1998)
COR005	3	198-200	1	58	1.5	F:CAATTCCTGGCATGCTGTAAG R:GATGCTCACTTCCATGAACC	F & R: In LINE	(Hopman *et al.* 1999)
UMNe158	3	146	2	50	3	F:AATTGAGAGCCAAGATGACACC R:GGCACCATTTGAGGAAGATG		(Wagner *et al.* 2004c)
LEX050	4	112-126	2	52	3	F:ATAGTCTGGGGTTAGGTAAGG R:TCTAGCCCAATGTAAATGC		(Coogle & Bailey 1997)
TKY210	4	240-280	2	52	1.5	F:CGGAAATGACCAGTTTGCTT R:AGGTTGCCACGGGATTAAGT	F: In LINE; R: In SINE	(Swinburne *et al.* 2000a)

Table A-5 (continued)

TKY833	4	149	2	58	1.5	F:TTTTGGTTCGAGTCCTTGGA R:GGGAAGTGGTTGTAGACAAG		(Tozaki *et al.* 2004)
UMNe199	4	124	2	58	1.5	F:TCCAGCAAATGGTAAGGTAGTG R:GATCTCTCTCCACCCTATACCA	F & R: In LINE	(Mickelson *et al.* 2003)
UMNe377	4	240	2	58	1.5	F:CAGACATGGCACCGCTTG R:AGGATATTGTGTTTCCCTCAGG	F: In SINE	(Wagner *et al.* 2004b)
UMNe404	4	166	2	58	3	F:TTGGAACTTTTAGCAAAGAACC R:GATCCATTCCCACATATGGC	F: In LINE	(Wagner *et al.* 2004b)
ASB029	4	138-152	2	56	3	F:CTGGCCCATAAAAAAACACTG R:TGTATGGTTGTCAGCTCAAACC	F & R: In LINE	(Raudsepp *et al.* 2008a)
TKY774	5	92	1	58	3	F:ACCATGCTCTTTCCAAACTCAA R:CGAGGTGGGTTGTACATTTA		(Tozaki *et al.* 2004)
COR023	5	269-275	2	50	2	F:CGTTTAGCACCTCTCATGAAC R:TCTTTGCAAAATAGGGCTTG		(Murphie *et al.* 1999)
NVHEQ122	5	203-207	2	58	1.5	F:GAGGAGGGTTGGCAGCAGAT R:CCAAGGGGGCACAAGACATA	F: In LINE; R: In SINE	(Penedo *et al.* 2005)
TKY041	5	160	2	56	2	F:TGGACAAGTGCTCTGTAATG R:TTCCATGAGTCTGGAGTTGG		(Hirota *et al.* 2001)
TKY1120	5	106	2	54	3	F:ACCTAAGTGTCCATCAATGC R:TCCTTCTTTATAAAGCTGCAC	F & R: In LINE	(Raudsepp *et al.* 2008a)
TKY1175	5	203	2	58	1.5	F:TTATCACCAGTTTCCAGAGC R:CTTATTCCACCCACTAATTCAC		(Tozaki *et al.* 2004)
TKY644	5	142	2	58	3	F:TAAGTGTCCGCAGACTTTAG R:TAGGACAAAGCACACTATGA		(Raudsepp *et al.* 2008a)
UMNe515	5	130	2	58	3	F:TCCAGAAACTCTACTCTTAGGG R:TGCCTTCATTCAACATTATG	F & R: In LINE	(Wagner *et al.* 2004b)
COR088	6	283-297	2	58	1.5	F:GAGCCAGCTTGTCGTGTATT R:GCACGAAGAAGAACCAAAGA		(Tallmadge *et al.* 1999a)
TKY030	6	203	2	58	1.5	F:CCAGTCCACTGCCCACCCAC R:GTGCATGGTGGGAGCTACTG		(Hirota *et al.* 2001)
TKY111	6	125 - 131	2	58	1.5	F:TATGGCGATTTCTGGTCTGTGTC R:GATGACAACACTGGGAAGAAAGAG		(Mashima *et al.* 2001)
SGCV28	7	157	2	58	1.5	F:CTGTGGCAGCTGTCATCTTGG R:CCCAATTCCAGCCCAGCTTGC		(Godard *et al.* 1997)
TKY690	7	193-199	2	58	1.5	F:GCGGATCAGGATCAACTAGC R:GGGCCCTGACAAGTTCCTAT	F: In LTR	(Tozaki et al. 2004)
UMNe074	7	139 - 143	2	58	3	F:CGATGGATGTGCTGTAAACG R:TGCTGCCTTCTCCCTCAC		(Roberts *et al.* 2000)

Table A-5 (continued)

UMNe100	7	183	2	58	3	F:CTGATGCAGAGGCATTTCTG R:CGTCAGCTGAGGCAACTATG		(Wagner *et al.* 2004c)
COR097	8	236-242	2	58	1.5	F:GGGATTTCTGAGATGCTGAA R:ATGGCTGGCTAGAGTTTGTG		(Tallmadge *et al.* 1999a)
HTG8	9	185-197	2	58	1.5	F:CAGGCCGTAGATGACTACCAATGA R:TTTTCAGAGTTAATTGGTATCACA		(Marklund *et al.* 1994)
HMS23	10	95	2	58	1.5	F:GATCCAATATTGTAAACCCCGCC R:CCTTCATAACCCTTATTGCAGCC		(Godard *et al.* 1997)
LEX062/LEX066	10	209-215	2	54	3	F:GCTCTCAGTAACCTCGATGTT R:ATTAAGGAGAAGGTGGAAAAGAC		(Coogle & Bailey 1999)
LEX068	11	162-174	2	58	1.5	F:AAATCCCGAGCTAAAATGTA R:TAGGAAGATAGGATCACAAGG	F: In SINE	(Coogle & Bailey 1999)
SGCV13	11	169-179	2	58	1.5	F:GGACTAAAGCCCAACCATCCAGC R:CTCACCAGTAAGGGGTTATGGGGC		(Godard *et al.* 1997)
TKY424	11	158-174	2	58	1.5	F:ATACAGGAGTGCGCTTTTCC R:AAACCATCCTCCACCTTTCC		(Tozaki *et al.* 2004)
TKY648	11	278-282	2	58	2	F:ACCCATCCATGCTGAAAAGA R:CGAAAGGTATTTGGTGTGTCTC		(Tozaki *et al.* 2004)
TKY954	11	123	2	54	3	F:AGCTCCTCAGGAATCTAATG R:GGACATGTTAACACCTCTCC	F & R: In DNA	(Raudsepp *et al.* 2008a)
TKY978	11	213	2	58	2	F:TGGCAGTCCAGTATCCATC R:GAACGTGGACACTTAACTGC	F: In SINE; R: Partially in LINE	(Raudsepp *et al.* 2008a)
COR009	12	122-130	2	56	2	F:CGGTGTTCTTATTTCATGGA R:GGAACGAAACACAGTGGAAC	F & R: In LINE	(Hopman *et al.* 1999)
UCD497	12	105-109	2	58	2	F:GTGGGAGGCAGCAGGAAC R:CCCCAGACACCGTGTGAT		(Eggleston-Stott *et al.* 1999)
TKY647	14	96	1	58	2	F:CTGGGAAGCAGCAGGAATAA R:GCTAATGGAAAGGCCACAGA		(Tozaki *et al.* 2004)
VHL209	14	91-103	1	58	1.5	F:TCTTACATCCTTCCATTACAACTA R:TGATACATATGTACGTGAAAGGAT		(van Haeringen *et al.* 1998)
TKY1114b	14	247	2	58	1.5	F:GTTTATTAGGTCACTGCTTCC R:GCAAGCTCATTGCCAAGAC		(Raudsepp *et al.* 2008a)
HTG018	14	160	2	58	2	F:CTGAAACCTCATTTTATACAG R:TACTAGAACACAGAAAGCCTA		(Lindgren 2000)
TFb	16	399	1	58	1.5	F:CTCCAGCTCACAGAATGGAC R:GAAACACTGGACTGCCTGC	F: In LINE; R: in SINE	(Raudsepp *et al.* 2008a)

Table A-5 (continued)

TKY445	16	298-308	2	58	1.5	F:CCTGGGCTAATTCCCATTTT R:CCAGCTCTTTGAAGGTAGCA		(Tozaki *et al.* 2004)
HMS41	17	113	2	58	1.5	F:AAAGTCTTCATTTGAAGTTTCCTAAG R:GACTAAGTAGATGAGATGTGTTTGG		(Godard *et al.* 1997)
NVHEQ79	17	179-197	2	58	1.5	F:ATTGCCTGTGCTGAGATGG R:GCAAATTGCCTCTGTATCACAC		(Bjornstad *et al.* 2000)
TKY546	17	244-248	2	58	1.5	F:TAGGATGGGGCACCAAGTAG R:CATTCCTAAGGGTGGAGCAC		(Tozaki *et al.* 2004)
NVHEQ024	17	161-163	2	58	2	F:CCACTGTGGAAAAGACTGAAAG R:TGTACTTCCTTGAAACCCAACA		(Guerin *et al.* 1999)
UMNe384	17	251	2	58	1.5	F:AGTCAGGAGAGACAGTGTAGGC R:TCATGGAAGCATTTCCCTAG	F & R: In LTR	(Raudsepp *et al.* 2008a)
AHT115	18	124-134	2	58	2	F:GAACCCAAGAGGAGTCCACA R:ACACACCTCAAGCAAAACACC		(Wagner *et al.* 2006)
COR096	18	314-326	2	58	1.5	F:CCCCTCTTTTGCTTGAGAAT R:GCGTGTATGTGAGGATTGAAG	R: In SINE	(Tallmadge *et al.* 1999a)
HTG28	18	180	2	58	3	F:AATCAACTAATATTAGGCCTCCT R:GAATACAGTTCTAGGGGCGT		(Lindgren *et al.* 1999)
TKY017	18	127	2	58	1.5	F:CAACTGTATGTTGACAGCACA R:CGGCCATATTAGGTTTATCTG	F: In LINE; R: In LINE	(Hirota *et al.* 2001)
UMNe180	18	148	2	58	2	F:TGGAAAATCCTCACAAACTGC R:TATATTTTCCTTTTGCGTGTGC		(Wagner *et al.* 2006)
LEX036	19	148-172	2	58	1.5	F:CCCAGCCTCTTCAATATCCA R:CCCTAAGGGGATTTAGCAGTT		(Coogle *et al.* 1997)
LEX073	19	241-251	2	58	1.5	F:CCAGCCATCCACTGGTAGAG R:GGGAAAAGGGGAACCTTCTA		(Bailey *et al.* 2000)
TKY004	19	87	2	58	3	F:TCTGTGTATGTGAGTGTATA R:GATCGAGGTATTCCTGAGTG		(Hirota *et al.* 1997)
DQA	20	260	1	50	1.5	F:CGACTCAGATGACCACATTG R:GGGGACACATACTGTTGGTAG		(Chowdhary *et al.* 2003)
LEX071	20	192-211	1	55	1.5	F:CTTTATTCTACTCTTTGGTCC R:CCGATATTTCACTGATTATT	R: In LINE	(Bailey *et al.* 2000)
COR050	20	283-297	2	58	1.5	F:TCTGTTGCCTTTATCCACAA R:ATGAAAACCCTGGGAATAGC		(Ruth *et al.* 1999)
HTG05	20	79-95	2	58	1.5	F:TGCTAAGCCTCAGCACATACA R:TGGAAATAAGGTTAGCAGGGATGC		(Ellegren *et al.* 1992)
LEX037	21	196-199	1	58	1.5	F:GGATTCCTCAACCTCCTAAA R:AGGGATAAGTGACCACCAC		(Coogle *et al.* 1997)

Table A-5 (continued)

COR016	22	184-203	2	58	1.5	F:CAGCTCAGTAGATGATTGTCCA R:GCAAAGACAAGGAGGTTAAGTT		(Hopman *et al.* 1999)
HMS47	22	202-207	2	58	1.5	F:CCTGCTGAGGACCTTGGAAGCT R:ATGTATTTTCAAGTCTAATATCTGCC		(Godard *et al.* 1997)
HTG21	22	132-144	2	58	1.5	F:ATTACTTCCTCCAGGTATCTCAG R:AGGCAGGGCTGGGAGACGT		(Lindgren *et al.* 1999)
COR001	22	123-137	2	58	1.5	F:GGCAGCATCCCACATAAAACAG R:GCTCACTATTACCACGATGATTGATTC	F: In SINE	(Hopman *et al.* 1999)
COR055	23	240-266	2	58	1.5	F:TAGTGACGCCTACGGATTTC R:CCCAAGAGGGCTTAGAAAGAG		(Ruth *et al.* 1999)
TKY441	23	190-210	2	56	2	F:TGTCCTCCAGGAGAGGGAAG R:TTTTTGAACGCTGATTGCAG		(Tozaki *et al.* 2004)
UM027	23	235-239	2	58	1.5	F:TGCAAGAATTGTGAGGGAC R:GTGCTCAGTTAGTGGTATTC	R: In SINE	(George *et al.* 1998)
COR025	24	172	1	58	1.5	F:ACAGAGCTGACTGCCTATGG R:TCCTCTTCTCAGGGAGACCT	F & R: In LINE	(Murphie *et al.* 1999)
COR018	25	246-278	2	58	1.5	F:AGTCTGGCAATATTGAGGATGT R:AGCAGCTACCCTTTGAATACTG	F: In LINE	(Hopman *et al.* 1999)
TKY275	26	144-154	2	54	3	F:TCTCAGTGGATATAACTAGC R:GAGATGGATACAGATAGAAG	F & R: In LINE	(Tozaki *et al.* 2001b)
UMNe186	27	149	1	58	1.5	F:TATAAAACCTCCCTGATTTGCC R:TGATAGAAGGAGCTAAGCCTGC		(Raudsepp *et al.* 2008a)
COR040	27	287-295	2	58	1.5	F:GTGTTGGGACACGAATGAAT R:AGGCTGTCTCCAGAGTCCTT	F: In SINE; R: In LINE	(Murphie *et al.* 1999)
HMS45	27	215	2	58	1.5	F:TGGTAAACTGTGCATGATTGG R:AAAGGAAGATTGGCAAACCA		(Godard *et al.* 1997)
COR021	29	202	2	58	1.5	F:CTGTAGCCAGCCCTGACAGT R:GATGGTGGGGTATTTGTCCA	Primers shows no match in draft sequence	(Murphie *et al.* 1999)
LEX018	29	228-242	2	54	3	F:TTTCATCACTTTCTGCTTCC R:TTCTCTTCCTTTGCTCATCCT		(Coogle *et al.* 1996b)
TKY913	29	176	1	58	2	F:TAATTTTGCGATGCCAGCTG R:AGGTAACAACCTGAACTTGC	F & R: In LTR	(Tozaki *et al.* 2004)
COR038	31	210-214	2	58	1.5	F:GCTGGAAAAGAGCAGTTTCA R:TGACATTAACTCCCGCATCT	F & R: In LINE	(Murphie *et al.* 1999)

Table A-5 (continued)

TKY368a	31	251-263	2	58	1.5	F:TTCAAGCCAGCAAAATTATAGC R:GTCAGGATAATTTTCAGCCC		(Tozaki *et al.* 2001b)
COR091	X	205	2	58	2	F:GGTGATTCAAGGTTAATGGC R:TGTATCTGTCCACAGCATGG		(Tallmadge *et al.* 1999a)
UMNe397a	X	126	2	58	2	F:TGTGGCTCCATCTCTCCAG R:TTTTCATGTCCCTAGGAAATTC		(Wagner *et al.* 2004b)
VHL81	X	162-174	2	58	1.5	F:CAACTATGTACTTTGGGGAGCT R:GTCCATGAAATTCTAGTTGTTGC	F: In SINE; R: In LINE	(van Haeringen *et al.* 1998)

Table A-6 Microsatellites checked for polymorphism in Thoroughbred but not used in panel

Marker	Chr No.	Size (bp)	No. alleles	Temp	$MgCl_2$ (mM)	Primers 5' > 3'	Comments	Reference
TKY577	1	140	3	58	1.5	F:CTGATGATCTGCTGCACAAG R:CAGCCCCTAATCACATAAAC	F: In LINE ; R: In SINE	(Raudsepp *et al.* 2008a)
ASB008	1	139-163	3	58	1.5	F:GACAACGTGGCAGCTCACTGCC R:GCAAGTAAGCCATATGTGCATGCG		(Breen *et al.* 1997)
AHT069	1	192-194	3	50	3	F:ACCGCTGGACCTTCCTTC R:ATCGTAGCATCCCTCACATACA		(Swinburne *et al.* 2003)
TKY659	1	129-141	3	58	1.5	F:TTCAAGCAGCAGTAGCAGGA R:CGAGGAGGGTTTTGTTCAAG		(Tozaki *et al.* 2004)
COR065	2	280-292	4	58	1.5	F:CAAAAGCACACACAAAGTGC R:TCCGGAAAGTGCAAAGTTAG	R: In LINE	(Tallmadge *et al.* 1999b)
UMNe236	2	156-180	3	52	4	F:CAGAACGTGCAAACTTAACTGC R:TTTGTTGAACTGACAAAATCCG	R: In SINE	(Mickelson *et al.* 2003)
AHT012	2	104-116	3	58	3	F:ACCCAAAGTCATGGGAATCA R:TTGTTGCCGACAACATGC		(Chowdhary *et al.* 2003)
TKY798	2	240	3	58	3	F:GAGCAGAAGGTACGAGAAGA R:AACTTAACCAGGCTGTTCTG		(Tozaki *et al.* 2004)
UMNe101	4	100	3	58	1.5	F:CATTTTAATCAACATTTTCCATCTG R:GCAATGCAGTGAGATGATGC		(Mickelson *et al.* 2003)
COR089	4	279-297	3	58	1.5	F:CCTGCCATAAATTTGTTTCC R:TCCCTACCTCATCTCCACAC		(Tallmadge *et al.* 1999a)
LEX033	4	182-197	3	58	1.5	F:TTTAATCAAAGGATTCAGTTG R:TTTCTCTTCAGGTGTCCTC		(Coogle *et al.* 1996a)
ASB022	4	158-172	4	58	1.5	F:GAGGAATGTGAAATACAGGAGG R:TTTGTGGTCTTCCGTGCACC		(Breen *et al.* 1997)
SGCV23	4	221-233	3	55	1.5	F:GGCTTAAGATATGGGTGAGTAAGG R:GCCCACCCTCTTACTTTTCTCAA		(Godard *et al.* 1997)
TKY698	4	250	5	58	1.5	F:TGTTGAGGCAAGGGTTCTTT R:CTCCATTGCCCACTCCTTAG		(Tozaki *et al.* 2004)
TKY223	4	170	4	52	1.5	F:GCAAGAAATCAACGCAAAAG R:CCGATTCAGTAGTTCAGGGATG	R: DNA element	(Raudsepp *et al.* 2008a)
LEX004	5	282-300	3	58	1.5	F:AATAGCAAATCTCCCACTTCA R:GTCCTCACAACCTCATCATAA		(Coogle *et al.* 1996a)
LEX069	5	142-170	3	58	1.5	F:TTTCTTTTTCCCACTTAAAGC R:TGGGACTTAGCAGTATGAAAC	F: In LTR	(Coogle & Bailey 1999)
TKY887	5	100	3	56	2	F:GAGAACTAGATGCCACCC R:TGTTGGAGTGTGTAGGCT		(Tozaki *et al.* 2004)

Table A-6 (continued)

TKY344	5	108	3	58	2	F:GTGTCCATCAATGGATGAAG R:CTTAAGGCTAAATAATATCCC	F & R: In LINE	(Tozaki *et al.* 2001b)
TKY025	6	100	3	58	1.5	F:AATCTCATGGCAGAATACCG R:GACTCTGGGAAGGGCTAAGG		(Kakoi *et al.* 2000)
NVHEQ082	6	134-148	3	58	1.5	F:TGTGGCAGCATCCCACAAAC R:CCTCCATTTTTGTCGGTTAGCG	F: In SINE	(Bjornstad *et al.* 2000)
TKY744	6	269-277	4	58	2	F:CCTGCCTTTGCTCTAGAAACC R:GGCTCCCAAGGGACTAGAAG		(Tozaki *et al.* 2004)
COR004	7	297-319	4	58	1.5	F:GAGTGTGACGGAGGACGA R:AGGGAGCAAACGCAAGAC		(Hopman *et al.* 1999)
ASB014	8	115-129	3	58	1.5	F:CTCCATGAATTCTCGCAGGTTGG R:CCATGGGCCATATGCACACTGC	R: In LTR	(Breen *et al.* 1997)
TKY452	9	266-278	3	58	1.5	F:TATTATGCCACCAGGCCAGT R:TGACCATGGTGAACCAGAGA	F: In MER1_type; R: SINE	(Tozaki *et al.* 2004)
TKY805	9	199	4	58	3	F:TGCCTTTTTCTCTCATCACC R:AGACTAGTCTGCAAGTTCAG	F: In SINE	(Tozaki *et al.* 2004)
TKY541	9	106-124	3	56	2	F:GAGGAGGGGCTTTCTCTCTC R:TGCAAAAAGCCAACATTTCC		(Tozaki *et al.* 2004)
UM037	9	114	3	55	1.5	F:TCATTTTATCCTCCACCTC R:AAAAGGGCGTAATATGG		(George *et al.* 1998)
ASB006	10	185-212	3	58	1.5	F:GGCACAGATGTTAGCTCAGC R:ATGGAACCAGCCTGGATTGC	F: In SINE: R: In SINE	(Breen *et al.* 1997)
COR048	10	178-186	3	58	1.5	F:GATTGGGATGCAAAGATGAG R:CAAGAGGATTGGGAACAAAGG	F: In LINE L2a	(Ruth *et al.* 1999)
SGCV8	12	126-143	4	50	1.5	F:GAGTTCATTCTTTTTCGTGGCTG R:GGAAACACCCTAAGTGTCCCTTG	F & R: In LINE	(Godard *et al.* 1997)
ASB037	13	131-145	3	58	1.5	F:CCTGCAACTTTTTCCCAGCC R:GGCAGATGTTAGCTCATGGC	F: In SINE	(Lindgren *et al.* 1998)
UM030	13	127 - 143	3	54	3	F:CCGTGAAGTCACAGACTTAG R:ACAGTTTCTACAACAAACTGA	F: In LTR; R: In DNA element	(George *et al.* 1998)
LEX047	14	237 - 245	3	54	3	F:TATAATAATGTGTCTTGGTGTG R:TGTTAATCAGGGTTCTCC	F: In LINE ; R: In LTR	(Coogle *et al.* 1997)
HTG11	14	110	3	58	4	F:CAATGATGGTACTTTGCATATTAA R:ATCGGCATGCACACTCATAGGTAG	F: In low comlexity region	(Marklund *et al.* 1994)
UM010	14	117-127	3	58	1.5	F:TACAGCCATTGGAAATCTAC R:CACCATTACATTTTCCCAG		(Meyer *et al.* 1997)
COR075	15	201-215	5	58	1.5	F:GCCCTAGTTAGCAACCAACA R:AAGATTGATTCCTCAGCACG		(Tallmadge *et al.* 1999b)

Table A-6 (continued)

HTG06	15	84-100	4	56	2	F:CCTGCTTGGAGGCTGTGATAAGAT R:GTTCACTGAATGTCAAATTCTGCT		(Ellegren *et al.* 1992)
UMNe277	16	158	3	58	1.5	F:AGGCGATGGTGACATCTTTC R:CTACACACGGGATAAATTCGC	F & R: In LINE	(Wagner *et al.* 2004a)
HMS20	16	116-140	3	58	1.5	F:TGGGAGAGGTACCTGAAATGTAC R:GTTGCTATAAAAAATTGTCTCCCTAC		(Godard *et al.* 1997)
TKY523	16	119-183	3	58	2	F:TGCACACCCATTCTAGCTCA R:GTGGCTCACTCCTCGCTTAC		(Tozaki *et al.* 2004)
COR011	16	288-298	3	58	1.5	F:CCTTCCGGTCTTTATTCACA R:GGTGGCTGGAGACACAATAG	F & R: In LINE	(Hopman *et al.* 1999)
TKY1128	17	252	3	58	1.5	F:TGGACATGTCACAAGAATCC R:ATGTTGGTCCTCATGCCTG	F: In MER1_type	(Raudsepp *et al.* 2008a)
LEX055	17	216-232	5	55	1.5	F:AGGGACACACAGGTGGTAG R:TGGATTTCCACTGTTACTTAT		(Coogle & Bailey 1997)
TKY287	17	224-240	4	58	3	F:ATCAGAGAACACCAAGAAGG R:TCTCTGCTATAGGTAAGGTC	F: In LINE	(Tozaki *et al.* 2000a)
TKY407	18	230-232	5	54	3	F:TCCTGCTGTGAGTTCCATGA R:CATCTTGTGCTGGGGATCTT	R: In LTR	(Tozaki *et al.* 2004)
COR092	19	190-206	4	58	1.5	F:GGCAAGAGCCAGGTATTTTC R:ACTGCTTGGACGAAACTGAG	R: In DNA element	(Tallmadge *et al.* 1999a)
COR029	20	222-228	3	58	1.5	F:CTAGAAGGGTTTCCCAAAGG R:TCGAGCTCCTGAAGAACATC		(Murphie *et al.* 1999)
HMS42	20	132-140	3	58	1.5	F:TAGATTTCTTAAGTGCCAATAGTGG R:GAACTGCTATAGATATACCTAACTC		(Godard *et al.* 1998)
TKY507	20	129-141	3	56	2	F:CACCTGCCTACAGTCCAAGC R:TTTGTGCTTAATGCCTTTGTG		(Tozaki *et al.* 2004)
LEX052	20	208-214	3	58	1.5	F:GGAACGGAAGAGTGTAGTTTT R:CATTTATTCATCAGCGATTTG	F: In LINE	(Coogle & Bailey 1997)
TKY421	22	270-292	4	58	1.5	F:CCTTGTAGGAGGCGAGTCAG R:GCCACTTCCTACCAATGCTC		(Tozaki *et al.* 2004)
AHT030	22	181-183	3	58	1.5	F:TCACCGCTCACCTTTTGAC R:CGTGCAGGTGTACATTTACATG	F & R: In LINE	(Swinburne *et al.* 2000b)
COR061	24	200-210	5	58	1.5	F:TTAAGAGTGGCAGACCGACT R:GAACGCTTCTTAAGTGGCAGA	F: In SINE	(Tallmadge *et al.* 1999b)
A-17	26	112-118	3	58	1.5	F:GTGGAGAGATAAAAGAAGATCC R:GGCCACAAGGAATGAACACAC		(Marti *et al.* 1998)
UM031	26	~200	4	58	1.5	F:GCTCAAACCAACCTTTCAAAAC R:TCAGGCCTTAAAACAGACACAC		(George *et al.* 1998)

Table A-6 (continued)

TKY828	27	186	3	58	3	F:CTCCTTCTGGCTCTACTATT R:GCATGGATTAAGGTGTATGC		(Tozaki *et al.* 2004)
TKY808	28	128	3	54	3	F:CCTGAGTGCTTTTGAAGTGG R:ATACTTTTTGCCAACTACAAAATA		(Tozaki *et al.* 2004)
TKY319	28	112-118	3	58	2	F:TATGCACGAGATTAAACGGG R:AAAGAAGTCAGATGAGCAGG		(Tozaki *et al.* 2001b)
AHT034	31	121-142	3	58	3	F:CTCAGGGCGAATGTTCCTC R:CCCCACCATGAGTCAAAAAC	R: In SINE	(Swinburne *et al.* 2000b)
UMNe107	X	143	3	55	1.5	F:TGCATATGTAGATGTATATAGGACAGG R:TTTCCCTACACTGGGACTGC		(Mickelson *et al.* 2003)
UM038	X	120-144	3	58	1.5	F:CAAGACAGAACAGAAGAAGAC R:ATATGGCTCGCTCCTAC	F & R: In LINE	(Godard *et al.* 1997)

Table A-7 Microsatellites dropped from panel due to deviation from Hardy-Weinberg equilibrium

Marker	Chr No.	Location EquCab2 Mb	Primers 5' > 3'	Size (bp)	No. Alleles	HET (%)	Temp	$MgCl_2$ (mM)	Comment	Reference
TKY558	1	126.49	F:TGGCGGATGCTAGATGTAGAG R:GCTGTGGAAATGACCCAGTT	232-236	3	33.8	58	1.5		(Tozaki et al. 2004)
CFTRms	4	74.65	F:GCATTTGGGCCACAATTTTA R:GCATCATGCTTTAGCTGTGG	234-276	7	56.2	58	1.5		
LEX023*	8	25.94	F:GTGGATGAAACAGGGAAGGA R:TGAAGGCAAAAACTTGCTGA	274-294	5	55.1	58	1.5	Excessive stutter peaks	(Coogle *et al.* 1996b)
TGFB1ms2	10	11.81	F:CTCATTGCGTCAGTCAGCAT R:CCAAACACACAGTGGGACAG	190-192	2	40.1	58	1.5		
ASB009	10	54.96	F:GTGCGCATGTATGTGCGTGCC R:ATTTCCACAAGGGACATGAGG	106-118	5	67.9	58	1.5		(Breen *et al.* 1997)
TKY693	13	37.55	F:CAGAGACTGCTGTCAGCTCCT R:CACAAATGCAGAACCCACAA	221-233	4	57.1	58	1.5	F: In DNA repeat; R: In SINE	(Tozaki *et al.* 2004)
B-8	15	21.78	F:TCCTCAGTCCTTTCTCATGC R:AGCTGAAGGCAATCTGTACC	96-116	7	84.5	58	1.5		(Marti *et al.* 1998)
TKY101	18	63.52	F:TCTGAAATACCGTGTGCCT R:TTCTGCCTCCCTCCAACTTT	218-238	6	52.4	58	2		(Mashima *et al.* 1999)
TKY606	19	18.76	F:ACATGCCAACTCACCAACTG R:TTCATCCTACGAGGGCTCAG	123-127	3	48.1	62	1.5		(Tozaki *et al.* 2004)
LEX031	21	36.12	F:CCCATTAAGAACTTTTCATCCTG R:GGCAAGCCCCACAAAATTAT	271-273	2	28.4	58	1.5	F & R: In LINE	(Coogle *et al.* 1996a)
TKY524*	24	9.79	F:AGTTGTGGCGTGCTTTCTAC R:TTGCACTTGAGCACTTAGTC	252-270	6	54.8	58	3	F: Tm too low; R: Tm too low	(Tozaki *et al.* 2004)
LEX032	24	36.11	F:CGTAGTAGGGTTTTGGGTCC R:TTGCGTTTCAATTTTTAATGAC	273-283	6	79.5	58	2		(Coogle *et al.* 1996a)
COR080	25	8.78	F:CGTGCTGCCAGAGGTAAATA R:ACTGAGATGAGGTTTGCTGC	218-222	3	43.2	58	1.5		(Tallmadge *et al.* 1999b)
HTG30	28	10.87	F:TCAAGGCAAATCTTTCCCAG R:GTAAAATAACAAGTTGTTCCAG	246-264	4	44.4	58	1.5	R: In SINE	(Lindgren 2000)
LEX025*	30	2.04	F:TTCTGTCCTTGCTCCTGCTT R:AATGATTTGCTGGCGAGAAC	292-296	3	59	58	1.5	F: Mismatch; R: Tm too low	(Coogle *et al.* 1996a)

HET% = percent heterozygosity

Markers in bold newly identified microsatellites from the draft equine sequence

Table A-8 Microsatellites dropped from panel during linkage analysis

Marker	Chr No.	Location EquCab2 Mb	Primers 5' > 3'	Size (bp)	No. Alleles	HET (%)	Temp	$MgCl_2$ (mM)	Comment	Reference
TKY530	1	90.67	F:ACAAAGCTGTGTGACAGACCA R:TCGTTTCTGCATCTCTTCCA	352-364	4	28.4	58	1.5	R: In LINE	(Tozaki *et al.* 2004)
COR053	1	182.81	F:AATTGACTGTGGAAGCCTTG R:GGCTGAGGAGTAAGCTGAAAG	195-217	4	74.4	58	3		(Ruth *et al.* 1999)
ASB018*	2	5.2	F:TTACATCAATGCAGGGCAAA R:CCTGCATTCAGTGAGGGAGT	273-291	6	64.3	58	1.5	R: Tm too low/High end self complementarity	(Breen *et al.* 1997)
TKY615*	2	14.06	F:GCTGGGGAGCACTTACAAAG R:CTGAAGCCTGTGGTTGGAAT	217-235	6	79.8	58	1.5	Excessive stutter peaks	(Tozaki *et al.* 2004)
LPLms3	2	49.3	F:GGACCCTAGTGGCAATGAAA R:CATGAGATGGGAGAGGAAGC	227-229	2	51.8	58	1.5		
A-14*	2	74.47	F:TGCTACCCTTAAAACTGGCAAT R:GTCATCACTACTCCCTACAC	152-160	3	65.5	58	1.5	F: Mismatch	(Marti *et al.* 1998)
UMNe236*	2	92.64	F:AAGATGCAATTAAACTCACCAGT R:TCCTCAGGCTGACACAGTTG	452-474	5	79.2	58	1.5	F: In SINE; R: High 3' stability	(Mickelson *et al.* 2003)
HTG02*	3	82.58	F:TGCATGAAGATCTCAATTACCC R:CCTTCCCTATGGAGGGAGTC	454-458	2	14.3	58	1.5	R: In SINE	(Ellegren *et al.* 1992)
CNTFms	7	18.9	F:TTCCCACTGCATATTCACCA R:TGTGGATGCTTTTCCATACC	250-254	3	53.6	58	1.5		
TKY461	7	26.62	F:ATGGCCCATCGTAAGAAACA R:GAGGGAGGAAGAAAGGAAGG	177-183	3	67.3	58	1.5		(Tozaki *et al.* 2004)
TKY775*	7	57.41	F:CAGTCGACAGCCAGATTTTCT R:TGTACAAGGGATGCACTGGA	161-171	3	45.2	58	1.5	R: In SINE, Tm too low/High 3' stability	(Tozaki *et al.* 2004)
ASB005*	9	71.66	F:GAGGAGCTCATGACCTGGAG R:CCCCATTCCTTTGTGGTAAA	217-223	3	73.8	58	1.5	F: Tm too high/High end self complementarity/High 3' stability; R: Tm too low	(Breen *et al.* 1997)
UCDEQ039*	11	7.39	F:CCGGCTAGAAGGGAAGTTCT R:GCAGGAACCCAGCACATTAT	289-309	4	45.8	58	1.5	R: High 3' stability	(Chowdhary et al. 2003)
TKY988*	11	26.22	F:GACTCCTGCCTGCTAACGTC R:TAACTGCGGTTTCCACTTCC	231-239	4	54.8	58	1.5	F: Tm too low; R: Tm too low/High end self complementarity, in SINE	(Tozaki *et al.* 2004)
COR014*	15	86.77	F:CCCAAGATTGATTCCTCAGC R:GTGGAAAGAGGGAATGAGCA	280-292	6	83.9	58	1.5	F: High 3' stability; R:Mismatch	(Hopman *et al.* 1999)
BRCA2ms	17	11.46	F:TGGAAAATTTCTTGGCTGCT R:TGGACAGGTGCCTTCAGATT	217-223	4	28	58	1.5		

Table A-8 (continued)

TKY572	22	36.31	F:GTGTAGCGCATGATCCAGTG R:CCCCAGCCTGACCTTCAT	122-128	4	72	58	1.5		(Tozaki *et al.* 2004)
VIASH-21	31	31.64	F:AAATGATAACGCCAAGTGCTCT R:ATGTGAGTGCCAGCTTGTGAT	263-265	2	17.9	58	1.5		(Swinburne *et al.* 2000a)

HET% = percent heterozygosity

Markers in bold newly identified microsatellites from the draft equine sequence

Table A-9 Final genome scanning panel

Marker	Chr No.	Location EquCab 2 Mb	cM	Primers 5' > 3'	Size (bp)	No. alleles	HET (%)	PIC 1	PIC 2	Temp	MgCl2 (mM)	Comment	Reference
HLM5	1	1.63	0	F:GCTGAAATCCTGTGGGTCTCCA R:CCTGGCTCCCTTGGTGGTCTGA	137-151	4	51.2	38.5	20.9	62	1.5		(Vega-Pla *et al.* 1996)
TKY711	1	7.66	9	F:GGAGCTCATATCCGGAGGTC R:GGAGGGGCTGTTTGTTGTTA	143-147	3	59.9	44.6	23.8	58	1.5		(Tozaki *et al.* 2004)
1CA30	1	14.02	26.9	F:TGGGGAGGGGTTGTTCTAG R:GCTCTCCACTGCATGCATAA	151-155	2	46.3	33.6	17.9	58	1.5		(Chowdhary *et al.* 2003)
TKY1016	1	15.77	36.9	F:CAATTGGTCCTGAGGTCAG R:ATTGGCGGTGGATGCTAAC	262-270	3	46.2	33.8	19.2	58	1.5	F: In SINE	(Tozaki *et al.* 2004)
LEX039	1	19.37	57.4	F:CCTCTGTCCCCACTACTCTC R:TTGATCTCCACTCCCAATG	201-207	2	39.5	31.7	26.5	58	1.5	F: In DNA repeat	(Coogle *et al.* 1997)
LEX020	1	20.59	57.4	F:GGAATAGGTGGGGGTCTGTT R:AGGGTACTAGCCAAGTGACTGC	225-233	3	36.4	33.3		58	1.5		(Coogle *et al.* 1996b)
TKY706*	1	65.93	78.7	F:CCCAAGAGTTCCGTGTCATT R:CCATGATGCTGATGGGTGTA	252-260	5	55.6	55.7	34.4	58	1.5	Excessive stutter peaks	(Tozaki *et al.* 2004)
AGTms	1	67.3	81.1	F:CGGCCTTAATCCACATCAGT R:AGCAGCTAAAGTGCCACTGC	275-279	2	27.8	25.2		58	1.5		
TKY670	1	84.52	81.1	F:GTTGCTTTGGGACCAGTCAT R:TCCAGCCCAGTCAGGTCTAT	169-173	3	11.1	10.9	32.2	58	1.5		(Tozaki *et al.* 2004)
AHT021/40	1	89.89	113.7	F:GCAAGTTCAGCACCTCCCT R:TTTATGACACCTGCTGAGAACG	241-257	6	77.8	68.2	35.5	58	1.5		(Chowdhary *et al.* 2003)
1CA32	1	100.66	118.8	F:AGTTACCAAATGTCGCATTGC R:TTCATCTGTAAAATGGGCAGG	116-122	3	51.2	42.6	31.3	58	2	R: In SINE	(Chowdhary *et al.* 2003)
1CA25	1	117.75	165.2	F:TCCAATTTTCCCCAATGGTA R:CTGCATTTTGACAATGGTGG	221-225	3	21	21.8	40.0	58	1.5	F & R: In SINE	(Chowdhary *et al.* 2003)
UCDEQ440	1	130.12	165.3	F:TGTTCGGACAGTGTGGAT R:GCAGGGTATGTGTGTGCT	123-129	4	64.8	58.4		58	2		(Eggleston-Stott *et al.* 1997)
HMS15*	1	136.85	165.3	F:TATCGTCCTAAGCCCGAAAA R:GGATGGGGAAAACTTTGGAT	338-352	4	51.9	47.7	40.3	58	1.5	R: Mismatch	(Godard *et al.* 1998)
UM026	1	150.24	177.4	F:CCCAAAATCAATTAGGTCTC R:ATCAGTTGCTCTCTACTTTTC	221-231	3	24.1	24.7	29.0	58	1.5		(George *et al.* 1998)
1CA16	1	157.18	183.2	F:TCACTGGGGGGTATATGCAT R:GATCCTACTCCACCTGAAGTGG	131-141	4	66.7	57.3	29.0	58	1.5		(Chowdhary *et al.* 2003)
TKY466	1	170.12	235.9	F:TGGAACACATTCCTCACCAG R:GTTCTCCTTCCACCCCAAAT	311-319	3	50.6	40.1	15.3	58	1.5		(Tozaki *et al.* 2004)

Table A-9 (continued)

TKY003	2	22.73	0	F:GGTTCACACAGGAGTCAGGGA R:CCTTCTGGTTTGCCTCGTCTC	179-185	4	55	52.9	27.5	62	1.5		(Tozaki *et al.* 1995)
ASB017	2	30.6	28.5	F:GAGGGCGGTACCTTTGTACC R:ACCAGTCAGGATCTCCACCG	113-135	6	78.40	71.1	35.9	60	1.5		(Breen *et al.* 1997)
TKY784	2	38.79	65.5	F:GATCAGTACTTTGCAAATGGATAAC R:GTAACTCCAAGGCTACGTTC	221-231	4	59.3	57.1	25.2	58	1.5	F:In SINE	(Tozaki *et al.* 2004)
ASB013	2	75.64	89.4	F:CTCTGAAAGAGCAGGATTGG R:GTCTTCTAAGTGGTAAGAGCC	143-157	4	44.4	39.4	26.1	58	1.5		(Breen *et al.* 1997)
TKY850	2	80.27	100.7	F:TGGTTTGCTGGTTTTGCCTC R:AGACATAGGTATTGACTGGG	165-181	4	72.8	66.6	39.3	58	2		(Tozaki *et al.* 2004)
UCP1ms	2	90.98	109.9	F:GTCTCCCTCCCTCAAACCTC R:GGGTTGCATACTGGATGGAC	155-177	3	58.8	38.9	32.3	58	1.5		
TKY497*	2	104.82	109.9	F:TGTTTGCACAGTGATAGTTTGAA R:TAGTGGCTGACTGTGCTCCA	453-465	5	57.4	39.7		58	1.5	F:In SINE	(Tozaki *et al.* 2004)
FABP2ms	2	108.16	125.8	F:TGCAGTGATGTCCTTTGGAA R:CCACTCGCTGTGGGAATACT	261-265	3	6.8	5.6	14.9	58	1.5		
TKY903	2	118.16	147.8	F:TCAAGTTCTTGCAATCCCAG R:CTGAGCCTCACAGATGAAC	233-239	4	53.4	38.7	12.7	58	3		(Tozaki *et al.* 2004)
AHT036*	3	2.94	0	F:TTGATCTTCCCTCCAAATGC R:CGAAGGCTTCCATATGTTAAAAA	329-339	5	72.2	63.2	28.4	58	1.5	R: High end self complementarity/3' stability	(Swinburne *et al.* 2000b)
CETPms	3	9.43	18.6	F:CAGCCATTGTGGTCTTGGTA R:CTGCTTACAGGGGGAAAGTG	246-256	4	64.8	59.3	31.7	58	1.5		
COR033	3	13.46	40.8	F:GAAGGGGCCCATTATTCATT R:TTCAAGGATATGTCCATTGGTG	257-287	7	82	75.3	42.7	58	1.5	R: in LTR	(Murphie *et al.* 1999)
TKY1085	3	22.55	47.4	F:ACCAAACTCAAGGTTAAGCTC R:TGAACTCTACCTTCAAGTGG	197-211	4	45.7	39.2	37.9	58	2		(Raudsepp *et al.* 2008a)
UCDEQ437	3	31.28	56.9	F:CTGTTCTGGGCAGGCTTCTCTA R:TTGCTGGCTTGGCTGGTC	195-207	6	74.5	68.0	44.6	60	1.5		(Eggleston-Stott *et al.* 1997)
LEX057	3	36.3	56.9	F:TGGTCCCCTAATCAAATCAGA R:ACGGCATCCCACATAAAATAG	177-187	5	32.7	28.0		58	1.5	R: In SINE	(Coogle & Bailey 1997)
ASB023*	3	79.27	75.4	F:GCAGGTGGAGGAGGTTTGTA R:CCCTGGTGGGTTAGATGAGA	193-219	6	82	76.8	43.2	58	1.5	F: Tm too low; R: Tm too low	(Lindgren *et al.* 1998)

Table A-9 (continued)

PPARGC 1Ams	3	100.8	88.8	F:GCGAGCTGTACCATAGCAAA R:GAGACAGCTCCCACCTATGC	333-337	3	51.2	37.3	23.5	58	1.5		
TKY439*	3	111.28	101	F:CATTTGAAAGGATCATTCACCA R:CAGCCTCTTACTTGACGCTTT	405-407	2	30.9	26.9	18.0	58	1.5	R: In SINE	(Tozaki *et al.* 2004)
AHT043*	4	2.91	0	F:CCATGACGGTCTTTGATCCT R:GGCCTTGGACTACTGCAGAC	303-319	5	62.3	51.7	32.0	58	1.5	F: High 3' stability	(Swinburne *et al.* 2000b)
TKY942	4	15.67	17	F:GAAGGGACCTTGTTCAGAAG R:TGACTGGAGCAGTGTGAAG	233-241	3	51.9	36.7	35.4	58	1.5		(Raudsepp *et al.* 2008a)
COR057	4	32.74	17	F:GGAGGAGAGGAAGAGAGTGG R:ATCCAGGGCTCTCCATAGTC	254-260	3	68.5	57.1		62	1.5	F & R: In LINE	(Ruth *et al.* 1999)
PON1ms	4	38.77	34.9	F:TTCCCAAGATGGTCCTGAAG R:TTCCCTATGGTGGTCTTCTCA	275-289	5	53.4	48.2	30.6	58	1.5		
IL6ms	4	54.41	57.3	F:AGAATTACCGAGGTGGCATC R:TCTGGGAGAGGTTTTCTTGG	169-171	2	40.7	33.6	25.9	58	1.5		
TKY830*	4	58.76	71.1	F:TCCTGATGCTGACAGTTTGC R:CACCATTGCAATCAAACTGG	396-400	3	60.5	55.4	35.1	58	1.5	F: Tm too low	(Tozaki *et al.* 2004)
HTG07	4	64.16	81.3	F:CCTGAAGCAGAACATCCCTCCTTG R:ATAAAGTGTCTGGGCAGAGCTGCT	138-146	4	61.7	47.5	33.4	58	1.5		(Marklund *et al.* 1994)
TKY552	4	65.14	81.3	F:CTAGAGGTGCCTTCCCAGAC R:ACCACCAAGACGAAAGGTGA	144-152	4	63	50.4		58	1.5	F & R: In LINE	(Tozaki *et al.* 2004)
HMS19	4	70.51	81.3	F:CTAACCAGCACAGAATGAATGGC R:TAAAAGAACAGTGGAGAGTAAAGTG	99-101	2	8.6	7.4		58	2	F & R: In LINE	(Godard *et al.* 1997)
HMS09	4	84.42	102.1	F:GCAACAGATATTAGCTCAGGGCC R:TGCTGTTCTTTTGCTGTGAAGGG	118-126	3	39.5	36.5	21.7	58	1.5	R: In SINE	(Godard *et al.* 1997)
NOS3ms	4	102.73	120.8	F:TTCCCCTTTGTCCACTTTTG R:TGTGGCGAAAGGCTTAAGAT	187-223	10	69.1	62.3	29.6	58	1.5		
TKY508	5	6.24	0	F:CAAAGCCCGCAGGATAAATA R:GCATTGCAAGTGCAGAAGTC	170-176	3	37	31.9	22.8	58	1.5	F: In DNA repeat	(Tozaki *et al.* 2004)
AHT024	5	12.16	0	F:TCCACTAATATCCCTCCACCC R:CTGTACCGCATGTGCAGC	214-218	3	29.6	26.2		62	1.5		(Swinburne *et al.* 2000b)
TKY731	5	17.26	13.9	F:GGGCACAGACCTACTCCACT R:GCAACAGCATCCCAGAATTT	253-269	3	39.5	36.1	21.7	58	1.5	R: Im SINE	(Tozaki *et al.* 2004)

Table A-9 (continued)

TKY521	5	36.99	49.5	F:TCTTCTCAGGATTTGGGAGGT R:CCCTTCTGAACGGCTTATGA	213-217	3	38.9	33.3	29.1	62	1.5		(Tozaki *et al.* 2004)
AHT050	5	44.28	56.2	F:GTTGGCTGGTTTTTGCATTT R:CGTACACACATTTTCACCCA	308-330	8	58.4	50.5	35.8	58	1.5		(Swinburne *et al.* 2003)
TKY673 *	5	52.26	62.3	F:TGTCTGGTTCCAGCTTAGGG R:AGCGTCACAAGGGAGCTTTA	363-375	3	69.8	56.1	34.1	58	1.5	F: Tm too low, mismatch	(Raudsepp *et al.* 2008a)
AMPD1ms	5	54.16	78	F:GAATCATTTCCCTTCCTCCA R:TCTCTGTGGGCCCCTTTAC	163-177	5	71.4	58.6	32.6	58	1.5		
UCDEQ304	5	56.67	87	F:CGCTTTCCTGCTGTCACC R:GAGGGACTGTGGGGGAGGT	111-129	4	64	50.3	30.1	58	1.5		(Eggleston-Stott *et al.* 1997)
LEX034	5	76.17	114.5	F:GCGGAGGTAAGAAGTGGTAG R:GGCCTAAGATGAGGGTGAA	261-271	4	75.3	67.9	38.3	58	2		(Coogle *et al.* 1997)
LEPRms	5	95.14	158.2	F:ACTTGCCCAGATGACTGCTT R:TGCCAACAAAATTTCTGAACC	276-280	3	66.2	54.6	21.6	58	1.5		
TKY543 *	6	6.24	0	F:AGGAAAGCGGTGTGGTACAG R:AGCACCTGAAATGTGGAAGG	153-157	3	30.2	26.9	12.1	58	1.5	R: Mismatch	(Tozaki *et al.* 2004)
TKY1001	6	11.87	26	F:TCTCAGAAGCCATCTGGAG R:ATCGATGCAGAACACGTGG	260-274	5	58	52.6	25.8	58	2		(Tozaki *et al.* 2004)
TKY556	6	19.55	39.3	F:GCAGAGGGTGAAGCCAGTAA R:TAATTCCTGCTTGTCGCAAA	195-229	7	52.5	51.8	31.8	58	3		(Tozaki *et al.* 2004)
GNB3ms	6	34.29	46.9	F:GATGAGGTGAAGTGGGGTGT R:GCTGGGAAAGGGGTTTTAAG	286-302	5	53.7	46.1	35.2	58	1.5		
VDRms	6	65.49	69.9	F:TCACTTCCTCAGGCTGGATT R:CCACCTTGTAGCTCCCTGAC	290-300	4	67.9	57.7	43.2	58	1.5		
COR070	6	65.85	71.5	F:CATCTGTTCCGTGGCATTA R:TTCAGGTGTGGGTTTTGAATC	288-314	9	77.8	79.8	45.7	62	1.5		(Tallmadge *et al.* 1999b)
TKY952 *	6	79.47	101	F:CCGTGTGTGTGTGTGTGTGT R:TTTTGTAGTTGCCTGGCTGA	293-305	5	64.2	57.5	26.1	58	3	F: High end self complementarity/ 3' stability; R: Tm too low, in MER1_type	(Tozaki *et al.* 2004)
TKY506	7	16.35	0	F:CTCCTTCATGCGTGAATCCT R:CAGTGACAGAAAACAGGATGG	187-195	3	46.9	35.8	26.3	58	1.5		(Tozaki *et al.* 2004)

Table A-9 (continued)

APOC3ms	7	24.88	3.5	F:AAGAGGCTGAAAGGGAAAGG R:TGTCACATGGCAAGTTGGTT	240-258	5	67.9	60.3	32.0	58	1.5		
UCP2ms3	7	70.02	25.8	F:CGGTGGGTTTTAGTGTTCCA R:ACTTAACCATCAGGCCAAGG	373-389	6	66.5	57.6	32.4	58	1.5		
TKY793	7	81.64	35.3	F:TCGACACAGATGTTAGCTCT R:TGCAATGCAATTTGCTTCCC	262-272	5	66	60.0	33.0	62	1.5	F: In SINE	(Tozaki *et al.* 2004)
AHT019	7	85.68	37.9	F:CATTTCTCTGGTGTATCTCCCA R:GGAATAGTCATAGTCCACGACC	165-167	2	51.2	37.5	30.8	58	1.5	F: In SINE	(Swinburne *et al.* 2000a)
TKY380*	7	92.4	69.6	F:GGCTACAATGGCAAGCAAAT R:CAAACAACACCATGCCAGAC	261-283	3	45.7	37.1	14.3	58	1.5	Excessive stutter peaks	(Tozaki *et al.* 2001b)
AHT005*	8	Un	0	F:GGGTCCAGTCCTTCTCTGCT R:GGGTCGTAGAGCTGCACAGT	337-347	5	74.1	64.0	25.7	58	1.5	R: Mismatches; F: Tm too high/3' stability	(Swinburne *et al.* 2000a)
UM034*	8	18.87	39.4	F:GTAGACGCCACCCTCATCTC R:CAGAGAGCAAGTGTGAG	227-243	2	26.3	32.2	29.0	58	1.5	F: Tm too low/High 3' stability; R: Too short	(George *et al.* 1998)
COR012*	8	46.41	61.6	F:ACCCATCACGTGAGTCTGGA R:AAAAATCTGCATGATTCTCTGGA	299-309	4	57.4	49.6	24.4	58	1.5	R: In LINE	(Hopman *et al.* 1999)
COR003	8	64.25	77.8	F:TAGGGAAACTCCTCAAAGCC R:GAAACCAAAACCTTCATCCA	211-225	4	49.4	45.1	25.9	58	1.5		(Hopman *et al.* 1999)
TKY932*	8	91.06	93.5	F:GGGGCCCACGTCTATATCTT R:GTGACCGAAGCTTGCTTTTC	383-389	4	61.1	60.1	32.6	58	1.5	F: Tm too low; R:High 3' stability	(Tozaki *et al.* 2004)
TKY457	9	16.72	0	F:GCCTCAAAGTTGGGTGAAAA R:ACTCCCTTTCCGAGATTGGT	272-284	5	56.8	54.1	37.4	58	1.5		(Tozaki *et al.* 2004)
COR008	9	18.91	5	F:TAAGTGCTGAGTCTGGGACC R:TGGTAGATAGCGTCTGGAGG	261-285	7	56.5	52.4	41.3	58	1.5	R: In SINE	(Hopman *et al.* 1999)
TKY533	9	22.77	13.1	F:CACTCTCCAGCTGGGTTAGC R:AGGTGATGGGCTGGATACAC	225-233	5	77.2	68.0	39.1	58	1.5		(Tozaki *et al.* 2004)
ASB004*	9	61.72	43	F:CCAGTGCTTCATCATCTGGA R:CCCCATTGTTCAAAGCAAAT	352-362	6	66	55.7	27.2	58	1.5	F: High 3' stability; R: Mismatch	(Breen *et al.* 1997)
LEX019	9	75.59	66.8	F:TTCCCTTTTCCTCACATCCT R:TTTTAGGTTCATCTATGTTGTTGC	177-181	3	54.9	50.0	25.5	60	1.5	F & R: In LINE	(Coogle *et al.* 1996b)

Table A-9 (continued)

NVHEQ 018	10	15.38	0	F:GGAGGAGACAGTGGCCCCAGTC R:GCTGAGCTCTCCCATCCCATCG	129-149	5	51.9	46.0	32.6	58	1.5		(Guerin *et al.* 1999)
SGCV30	10	19.34	8.6	F:ACTGGAGGGGTGAAACAGATTCA GA R:GGAAGGGAGGTCATCAGAA	174-182	5	77	69.6	41.6	58	1.5		(Godard *et al.* 1997)
LEX008*	10	20.38	10.6	F:TGACACTGGAAGCACACACA R:CGAAAAAGCCACTTGAGGTC	198-206	5	64.2	52.4	42.7	58	1.5	Excessive stutter peaks	(Coogle *et al.* 1996a)
NVHEQ 007*	10	36.37	28.9	F:GGGAGGATGCAACTGAAAAG R:AGCAAAAAGGGTGTGGTTGT	385-391	4	59.3	47.6	30.9	58	1.5	F: In SINE	(Guerin *et al.* 1999)
UM040	10	51.31	37.4	F:CTCTTGTACATGTCTCCTTGTC R:TACTTTCTCTCTTCCAAACC	264-272	3	57.4	51.7	28.5	58	3	R: In LINE	(George *et al.* 1998)
SGCV17	10	57.8	46.3	F:GGCCCAACGTCTATAGAAAGATG T R:CCCCCAAATGGCTATTTTCTAA	147-159	2	48.1	34.8	21.9	58	1.5		(Godard *et al.* 1997)
LEX009*	10	61.88	60.6	F:CTGCATGACTGAAAGCCGTA R:ATGGTCCATTGTGAGGGTGT	399-411	3	53.1	46.6	28.5	58	1.5	R: Mismatch	(Coogle *et al.* 1996a)
NVHEQ 067*	10	71.4	77.4	F:GCTCCCAGCTTTTTGTTGAG R:GGTAGCATGTGGCTTCCATT	392-400	2	38.3	33.0	16.3	58	1.5	F & R: In LINE	(Bjornstad *et al.* 2000)
TKY551*	11	1.82	0	F:CGTAACATACTGCCCATTCC R:GGGTGAGCACCTCCTCTACA	369-373	3	55.6	45.8	21.8	58	1.5	F: In LTR	(Tozaki *et al.* 2004)
TKY343*	11	12.99	18.1	F:ATGGGTGCTTCACCAGC R:CCCTGAGCTTGCCAATTTTA	366-380	6	67.9	65.3	37.2	58	1.5	R: In SINE	(Tozaki *et al.* 2001b)
ACEms1	11	15.71	27.7	F:TGTTTCCACCAGGAAACCTC R:ATCTTTCAGCCCACTTTGGA	384-412	4	57.4	52.7	32.2	58	1.5		
SGCV24*	11	19.53	38.8	F:TGTTTGCAGCTGGATCTTTG R:TGTGAAACGAGCAGGAAGTG	298-308	6	51.6	50.4	37.3	58	1.5	F: In LTR, high 3' stability; R: Tm too high	(Godard *et al.* 1997)
UMNe378	11	20.21	49.6	F:CTAGGGACTCTGGAAAGGGC R:CCACGTAACCCAAGTGTGTG	259-271	4	57.4	54.7	33.4	58	2		(Wagner *et al.* 2004b)
AHT076	11	21.92	61.1	F:ACTGAGCGGCCATGACTC R:GAGGATTGGTGGCAGATGTT	182-188	2	37.7	31.6	23.0	58	1.5	F: In SINE	(Swinburne *et al.* 2003)
ACADVLms3	11	50.02	98.4	F:AGTGATCAGGGCAGAAGGTG R:CCCCAAAGAACACTTGGCTA	270-284	4	43.7	42.6	20.3	58	1.5		
TKY683	12	3.56	0	F:CAATGGTGAAAATCTCTTTCCAG R:AGGTACAGCCCTGGTGTAAAA	179-193	6	69.8	63.2	27.3	58	1.5		(Tozaki *et al.* 2004)
SGCV10	12	9.5	19.5	F:CATCCATCCTTTCCAGCTCGATA TTC R:CAAGACCGTAACTCAGGAGCCC	193-201	4	66	65.0	41.9	58	1.5	F: In SINE	(Godard *et al.* 1997)
TKY404	12	20.14	35.8	F:TTGTCAGTGTGTGAGGAGATCA R:TGCAGAGCAGCAATTTCATT	190-198	4	66	57.5	39.6	58	1.5	F: In LINE	(Tozaki *et al.* 2004)

Table A-9 (continued)

UCDEQ 411	12	24.88	48.5	F:TGCTGCGAGTCAGTGGCT R:ATGCCTCACATCCCTGGGT	94-106	5	60	50.5	43.8	58	1.5		(Eggleston-Stott *et al.* 1997)
ACTN3 ms	12	26.55	51.9	F:CCTCACAAGGAACTGCAACA R:CTTCCTGTGTCCAGCAGTCA	365-375	3	46.9	39.1	45.6	58	1.5		
COR058*	12	27.94	54.2	F:ATCACTCTGGCAGCTGTGTG R:CAAATCCACCCCTCTCTTCA	240-260	8	80.1	79.3	47.3	58	1.5	Excessive stutter peaks	(Ruth *et al.* 1999)
AHT017	12	31.03	71.9	F:CCCCATAACCACAAGTGAGG R:GAAGTGGGAGAGTCGGTAAGG	128-148	7	69.1	62.2	34.0	58	1.5		(Swinburne *et al.* 2000a)
COR069	13	6.09	0	F:AGCCACCAGTCTGTTCTCTG R:AATGTCCTTTGGTGGATGAAC	292-300	4	45.7	42.7	20.8	58	1.5	F & R: In LINE	(Tallmadge *et al.* 1999b)
TKY585	13	7.84	25.8	F:GCAGCCTGAGGAAATGAAGT R:TTATGACCCCCACTCTCCTG	137-151	4	55.6	47.6	27.0	60	1.5		(Tozaki *et al.* 2004)
VHL047*	13	16.89	40.8	F:GCCTCTGCTGTGTTTCTTCC R:GCTGTGGTTACCAGGCAGAT	321-339	4	74.1	67.3	36.1	58	1.5	F & R: Mismathches	(van Haeringen *et al.* 1998)
ASB001	13	31.71	95.1	F:AGCAGAAACCCACTCAAGCC R:GCATAATACCCTCAAGGTC	175-183	3	55.6	52.4	18.6	58	1.5	F: In SINE	(Breen *et al.* 1997)
LEX043	14	16.14	0	F:CATTAAGCAACAAAAAGCATC R:GGAAAAGCATGACAAGACACT	259-263	3	30.2	25.8	9.5	58	1.5		(Coogle *et al.* 1997)
UM032	14	28.53	33.2	F:AAATGGTCAGCCTCTCCTC R:TGTCTCTCTAGTCCCACTCCTC	165-171	3	55.6	45.8	23.0	58	1.5		(Swinburne *et al.* 2000a)
ADRB2 ms	14	28.92	44.9	F:CTGTCTCCTCTTCCGAGTGC R:GACAAGCAGGAGAGGTGGAG	205-221	5	52.5	46.8	20.6	58	1.5		
TKY491	14	81.17	120.3	F:CCTCTTGGGACAGAGGACAG R:TCTCTCAGGAGCCTGTGTTG	253-269	5	77.8	69.4	35.7	58	2	F: In SINE	(Tozaki *et al.* 2004)
COR002	14	90	136.7	F:CTTGAGCACCCAGTAACACC R:CCAGGAATCTTCTCTACCGA	248-252	3	58.6	58.0	31.0	58	1.5	R: In SINE	(Hopman *et al.* 1999)
TKY565	15	31.64	0	F:GGGGCGTTAAGGCAGTAAG R:ATGCGCTTCAGCATCTCTTT	189-203	6	51.6	43.9	23.3	58	1.5		(Tozaki *et al.* 2004)
LEX046	15	39.36	24.7	F:ATAAGCCAATCCACTTTTCC R:ATTACCACCCCATTTCCTT	132-144	5	69.6	69.7	32.3	58	3		(Coogle *et al.* 1997)
TKY795	15	45.69	51.9	F:AGAGTAATGTGGTGGAGGAG R:TTTTGTGGCTAGGTTTTGGG	159-171	3	43.2	34.2	33.6	58	2		(Tozaki *et al.* 2004)
ASB019	15	58.31	51.9	F:GAGTTGGAGCTCAAGTCTGTC R:GTTTAGCAACTACAGCGTAGG	193-205	3	38.9	34.9		58	2		(Breen *et al.* 1997)

Table A-9 (continued)

AHT016	15	65.46	56.9	F:ATGTTGTGCAAATGGGATGA R:TGCCCATTGATTGATGATTG	151-169	5	69.6	63.4	37.7	58	1.5	F & R: In LINE	(Swinburne *et al.* 2000a)
15CA001	15	88.65	109	F:CTCTGTGCCAGACCCTGTTC R:GGTCTGCTCGAACGTCATTT	257-261	3	48.1	40.9	13.0	58	2	F & R: In LINE	(Penedo *et al.* 2005)
AHT037	16	3.91	0	F:ATTTCACCCACTCCCAACC R:TTTCGTTTGCGACAATATGG	226-238	3	67.9	55.7	26.3	58	1.5	F & R: In LINE	(Swinburne *et al.* 2000b)
AHT038	16	30.27	36.6	F:TTCATGGCCTTCAAAACTCC R:CCAGCTGGGGATACTTACCA	145-153	5	68.5	62.7	39.1	58	1.5	F & R: In LINE	(Swinburne *et al.* 2000b)
L15.2	16	51.93	53.3	F:GGGCAATGAAAGGTCTGACTATGAG R:AACAGTTTCTGCTTGTGCTGACAC	161-173	5	46.3	40.9	28.0	58	2		(Guerin *et al.* 1999)
TKY936	16	73.65	85.2	F:ACCACTGTACTGAATACTGG R:ACAAAGCATCTCCTCGAATAG	122-142	6	87	78.3	42.2	58	1.5	F & R: In LINE	(Tozaki *et al.* 2004)
COR105	17	4.76	0	F:TTTCCTCATTGCTTCCTGAG R:CCCAAGGTCTGTCTTGCTCTC	193-211	5	73.5	68.7	39.0	58	1.5		(Swinburne *et al.* 2000a)
COR007	17	6.6	11.4	F:GTGTTGGATGAAGCGAATGA R:GACTTGCCTGGCTTTGAGTC	185-195	4	65.8	64.2	38.9	62	1.5		(Hopman *et al.* 1999)
UCDEQ014	17	28.63	31.3	F:GCATTTGCTCACTGGCTAC R:ACTCCTCCACTCCCACCTA	147-153	4	51.9	64.2	24.1	58	1.5		(Eggleston-Stott *et al.* 1996)
COR032	17	41.42	41.5	F:GCCCTCTTAGAGCATTTTCC R:CAGAGATGGCTGGAGTAAGG	265-271	3	3.7	3.7	19.9	58	1.5		(Murphie *et al.* 1999)
LEX067	17	59.85	51.6	F:GTTGCTAAAATTGTTCCAGAC R:CCAATAAAAGGAATCACTGCT	227-233	5	67.9	59.5	36.9	58	3	F & R: In LINE	(Coogle & Bailey 1999)
HMS25*	17	61.87	55.7	F:GCAAACATAAAATATGCATGTCC R:TGTAAGGCTTGAGGCCAACT	130-134	3	48.8	40.5	33.4	58	3	F: High end self complementarity	(Godard *et al.* 1997)
TKY684	17	66.8	62.1	F:TTTGCAGGCTTTCTGTATTTTT R:TTCTGTTTCGTTTTCCCTGAA	247-257	6	72.8	69.2	35.5	58	1.5		(Tozaki *et al.* 2004)
TKY019	18	0.53	0	F:CTTCTGCTGATTCCTGAATG R:GGATCTCCTTAAATGGAACA	164-178	4	72.3	72.3	26.4	58	1.5	F & R: In LTR	(Kakoi *et al.* 1999)
SGCV07	18	26.36	34.3	F:GAATTTGAATGTATCTATTCTGAATG R:GTGAGTTTTCAAGCTGGCATATTC	151-161	4	55.6	55.6	22.3	58	1.5		(Godard *et al.* 1997)

Table A-9 (continued)

TKY545 *	18	43.78	56.4	F:CTGCCTTGGTGGGATTTCTA R:CCAAAAACACAGCAGAACGA	360-366	4	64.8	64.8	32.5	58	1.5	Excessive stutter peaks	(Tozaki *et al.* 2004)
HTG17	18	57.92	63.4	F:GCTATCCCTCCTGAGTCTTA R:AGGTAATTTGAAATAAAATACAC	147-151	3	63	63.0	27.7	58	2		(Lindgren *et al.* 1999)
GDF8ms2	18	66.5	86.4	F:CTTACCCAGCCCAGGTCATA R:CCATCCCAGTTTCGAGAGAA	331-341	2	6.2	6.2		58	1.5		
TKY016	18	66.8	86.4	F:GGTTATGGTTTGGTATCTGTC R:AAAACAATGGCTTCCTGGTCA	131-143	4	45.7	45.7	21.4	58	2		(Kakoi *et al.* 1999)
HLM3	18	74.94	111.4	F:GAAGGTAGAAAAGGAGGGCTAGAAC R:TCTAGAGGACCATTCTCTGGGCTGTG	133-145	4	73.5	73.5	28.6	58	1.5	F: In LTR	(Vega-Pla *et al.* 1996)
TKY730	19	35.34	0	F:CCTCCTTGAGTAACAGTCAC R:GACCTTATCAGTGCCCTTG	282-292	5	62.3	60.1	33.3	58	1.5		(Farber & Medrano 2004)
NVHEQ011*	19	44.11	11.3	F:CTGTGCAGGATACAGGGTGA R:TTACAAGGCCATGCCTATCC	307-313	3	61.1	49.8	31.7	58	2	F: Mismatch	(Guerin *et al.* 1999)
TKY783	19	58.23	38.8	F:TAGACAACTGACCAGTGCAA R:CTGGTTGCTCCTGTCTTAAA	161-173	5	67.3	62.2	33.5	58	1.5		(Tozaki *et al.* 2004)
AHT018	20	10.04	0	F:TTTTCCAGTGACTCTGAGTGTG R:GTTGTGGGAAAACTAGTCTGGC	187-195	4	45.7	38.0	22.7	58	2		(Swinburne *et al.* 2000a)
UMNe056	20	29.28	22	F:TCTGTCTGCAGCTAAAGAGGC R:GCGGGGTACATAAGACTGTAGC	204-220	6	69.1	69.2	48.1	58	1.5		(Roberts *et al.* 2000)
UM011*	20	33.5	29.7	F:TTAGGGGTTTTCCCATTTCC R:TGAAGCTGTTGGCATCAGAC	407-423	6	82.1	76.7	50.0	58	1.5	F: Tm too low; R: High 3' stability	(Meyer *et al.* 1997)
TKY547	20	40.84	35.9	F:TTGCTCAGGAAGCAAAGGAT R:AAAATGAGGCTTTCGCACAT	220-230	3	63.6	53.5	40.7	58	1.5		(Tozaki *et al.* 2004)
NVHEQ021	20	60.07	122.7	F:CCAGAACCTGGACTGAACAGTGTC R:GAATGTGCTTGATGCAGAAGAAGG	173-181	4	51.9	41.6	20.2	58	1.5		(Bjornstad *et al.* 2000)
COR073	21	20.25	0	F:GCCAAGACATGGAAACAATC R:GTTCTCAAGGTGCATCCCTA	206-218	6	68.3	66.9	34.9	58	1.5	F & R: In LINE	(Tallmadge *et al.* 1999b)
TKY671	21	26.39	11.5	F:AGGCAACATGAGAAGGCACA R:ATAGCACCTGTTCCCTGGAG	116-132	5	79.6	71.5	33.2	58	1.5	F: In LINE	(Tozaki *et al.* 2004)
TKY623	21	53.46	29.6	F:CAGTGTGGGTGGGCTTTATC R:ACCACTAGGGTGTGCATGTG	291-299	4	63.6	55.3	28.5	58	1.5	F & R: In LTR	(Tozaki *et al.* 2004)

Table A-9 (continued)

TKY308*	22	10.87	0	F:GGGAGAAGCAAGCACACACT R:CTGAGCATGAAGCGTCTGAG	260-286	5	50.9	46.4	30.1	58	1.5	F: High Tm/self complementarity/3' stability; R: Tm too low	(Tozaki *et al.* 2001b)
HTG14*	22	14.27	5	F:TGATGCAAATTTCCCAATGA R:CTATACGCCAGAGGGAGCAG	364-380	5	74.7	64.7	37.0	58	1.5	R: Tm too high	(Marklund *et al.* 1994)
SGCV01*	22	30.65	48.2	F:TTCCTCTTCCCAAGGGTTCT R:AGGATACGGATGAGCTGGTG	165-169	3	46.9	38.8	17.2	58	1.5	Excessive stutter peaks	(Godard *et al.* 1997)
SGCV19	22	47.77	93.1	F:GCCCCCACCTGCTCCACC R:GGGGCAAAGTGGAAATCC	156-162	4	60.5	54.4	28.1	58	1.5		(Godard *et al.* 1997)
AHT039*	23	4.79	0	F:CCATGCTCAGATGCTGAATG R:AAACAACGTTGGCTGAATCC	161-175	3	63.1	53.0	23.1	58	1.5	F: High 3' stability	(Swinburne *et al.* 2000b)
ASB039	23	20.91	27.2	F:ACAGCTGCCTGGATATGTGG R:GCAGAGAGAAATAGAGATGC	183-189	3	65.4	51.0	16.6	58	1.5		(Swinburne *et al.* 2000a)
TKY542	23	42.04	55	F:GGTGCCTAACCCAGATATGC R:CCCCATCAATTTCTGCTTTT	164-168	2	46.3	36.7	15.8	58	1.5	F: In LINE	(Tozaki *et al.* 2004)
SGCV04	23	51.9	66.5	F:CGACGCCTCCTCCTAAAC R:CAGCTGTGTGCCTTTGATTAT	223-229	2	21.6	17.4	13.4	58	1.5		(Godard *et al.* 1997)
TKY756	24	22.56	0	F:TGAAGGGAGCATCCGTTACT R:GGTGGATCTGCCCTCAGTAG	161-175	5	61.1	51.0	32.4	58	1.5		(Tozaki *et al.* 2004)
LEX074	24	34.01	28.7	F:AAGAGTGCTCCCGTGTG R:GACAATGCAGAACTGGGTAA	173-191	7	74.5	72.0	36.3	58	1.5		(Bailey *et al.* 2000)
COR024	24	41	33.7	F:CAAAAGTGATTGCCTTCGAT R:TTGGAAGCTGGGTGATTG	223-231	5	68.5	63.5	30.8	58	1.5		(Murphie *et al.* 1999)
UCDEQ464*	25	1.97	0	F:TCTGGAGAGGCTCTGAGGAA R:GAATTTGGGCAGGTGACAAT	440-446	3	30.2	26.9	10.0	58	1.5	R: Tm too low	(Eggleston-Stott *et al.* 1999)
NVHEQ043	25	31.06	46.1	F:TGACACAAGATAAAAGCCCCAGG R:GATTGGGAAAAGAGCACAGCC	161-175	5	69.8	60.4	24.3	60	1.5		(Roed *et al.* 1998)
TKY794	26	8.99	0	F:CAAACGCCCAGCAGAGTG R:AGTCTGCAACTAACTCTCAG	221-223	2	27.2	22.0	15.0	58	1.5		(Tozaki *et al.* 2004)
NVHEQ070	26	30.25	34.8	F:GCTGGTCAAGTCACACTGTG R:AACCTCACCCCAAGTTGTAT	206-216	4	63	56.9	30.1	58	2	F: In SINE	(Bjornstad *et al.* 2000)
COR071	26	19.05	74.7	F:CTTGGGCTACAACAGGGAATA R:CTGCTATTTCAAACACTTGGA	209-221	5	66	64.5	31.4	62	1.5		(Tallmadge *et al.* 1999b)

Table A-9 (continued)

COR031	27	1.37	0	F:CAATTGCCATTTGTTCCAGTG R:GCTTAAGAAACACCAGGCAG	222-234	4	75.3	66.8	42.3	58	1.5		(Murphie *et al.* 1999)
ADRB3 ms2	27	7.66	2.2	F:GCTGGGGAGACAGTTGTGTT R:AGCACTGCCCACCAGAAG	377-393	5	54.9	50.8	42.5	58	1.5		
UCDEQ 005*	27	14.04	11.3	F:CTTTGTGCTCTGAGGGGAAG R:TGCAGCACTTCCTCATGTTC	346-352	3	51.9	47.8	36.7	58	1.5	F: Mismatch; R: High 3' stability	(Eggleston-Stott *et al.* 1996)
TKY603 *	27	20.74	20.9	F:CTCATGGGATTGGGAGAAAA R:GATGCCTCGAACTAGCTTGC	294-298	2	40.1	32.2	30.4	58	1.5	F: Mismatch	(Tozaki *et al.* 2004)
TKY480	27	25.42	35.2	F:CAGAGGGCAGAGGATTTGTC R:AGGGACGAGGACCAAATGTA	264-278	6	66.7	60.3	37.8	58	2	F & R: In LTR	(Tozaki *et al.* 2004)
COR017	27	35.27	65.8	F:GAAGGCCTGAAGCATTTACA R:CGTAATGTTGACCAAACTTCA	263-275	6	65.4	57.6	25.1	62	1.5		(Hopman *et al.* 1999)
TKY872	28	28.64	0	F:TGGGGCTTTGGGAGACAT R:CAAGGCCAGAATTTCTTGAAGT	136-142	4	44.4	53.3	18.8	62	1.5	F: In DNA repeat	(Tozaki *et al.* 2004)
TKY299	28	33.86	11.5	F:TGAAGTTTGAGCCTGGACACC R:TGCAATGTCCTGGGAAATCC	116-132	4	51.3	51.1	28.6	62	1.5		(Tozaki *et al.* 2000b)
UCDEQ 425*	28	43.08	17.9	F:AGCACAGCTGCCTCGTTAAT R:GACCTGGTACCCTCCTCTCC	299-311	6	45.7	42.2	27.3	58	1.5	F: Tm too lowR: Tm too low	(Eggleston-Stott *et al.* 1997)
COR082	29	4.27	0	F:GCTTTTGTTTCTCAATCCTAGC R:TGAAGTCAAATCCCTGCTTC	239-247	4	71	69.9	38.3	58	1.5		(Tallmadge *et al.* 1999a)
COR027	29	22.22	40	F:CAGCTCTGCAATTTCTCCTC R:AATGACCAAGGCATTGAAAG	249-263	4	54.3	45.1	31.2	58	1.5		(Murphie *et al.* 1999)
L12.2	29	30.57	52.6	F:TCACACGAAACCAGTCACGGGAG R:ACAGACACTGCTGGAGTCTCATGG	149-163	4	75.2	68.1	41.6	58	1.5		(Guerin *et al.* 1999)
HMS18	30	11.4	0	F:CAACAATGAAAATTTGTCCTGTGC R:GTAAATGAGTAGACAATCATGAGG	188-192	3	34	45.1	6.7	58	1.5		(Godard *et al.* 1997)
LEX075	30	26.87	51	F:TCTGAAAAGTTGCAGTTTGAGAA R:TACAGTGTATTGGGGGCACA	229-237	5	49.4	68.1	10.0	58	1.5		(Swinburne *et al.* 2000a)
TKY668	31	2.86	0	F:GCTGTAATGTACCGCCTGGT R:TAGGCAGCTGCGATAAGACA	180-194	6	70.4	65.8	39.3	58	1.5		(Tozaki *et al.* 2004)
AHT033	31	0.6	6.1	F:CTGAGGGCGTAAGTCGAGTC R:GTTAATAGGAGCGGTTGTTTGG	173-183	5	69.1	62.0	39.0	58	2		(Swinburne *et al.* 2000b)
TKY105	31	20.76	55.4	F:TCGGGACAGGAAAGGAAGCT R:GGTATCCAGAATGAAAGACCCC	215-223	4	38.3	35.3	16.5	58	1.5		(Mashima *et al.* 2001)

Table A-9 (continued)

AHT099	X	4.55	0	F:TTCTTGGGCAGGGGATTT R:GGGGTGAGTTGGCTGTATTC	187-191	3	58.80	51.9	37.7	58	1.5		(Swinburne *et al.* 2003)
UMNe058	X	21.52	30.4	F:GATCCAAGACTTGAAGGTTAGC R:TTTCCTCACCATCCTCCTTGAC	164-186	5	57.80	60.6	37.2	58	2		(Roberts *et al.* 2000)
GKms	X	23.9	47.6	F:GTGTGCGTCTGTGTGTGTGT R:TCTGTCCACCAGGAAAGACA	315-323	6	64.60	52.6	51.3	58	1.5		
UMNe148	X	32.57	65.7	F:GATCAAACACTAGAATGTTCACAC R:CAGCTGTGAGGCAGAGACTG	113-125	6	70.30	72.1	46.3	58	1.5		(Mickelson *et al.* 2003)
UMNe402	X	55.67	87.3	F:AAGATGTGGCCTGTTTCAGG R:TTGATTCCTGGAGACTGATGG	259-263	3	48.00	42.0	32.3	58	1.5		(Wagner *et al.* 2004b)
PGK1ms	X	57.91	134.5	F:ACAGGGGGTTATCTGTGTGG R:TCCAGACCAAATCGAAACTATG	235-241	3	47.10	37.2	28.4	58	1.5		
LEX013	X	82.44	164	F:TGCTAGAGGAAGGGATAAAGG R:CTCTGCTCTTCCATTTCTTGC	142-146	3	54.40	47.2	20.8	58	3	R: In LTR	(Coogle *et al.* 1996a)
TKY020*	X	119.44	185.8	F:TTGCCACAGCAGTCTTTCAC R:CAGGTAGTGGAGCAGGGAAG	343-346	2	38.20	34.5	18.2	58	1.5	Excessive stutter peaks	(Hirota *et al.* 2001)

PIC 1 Percent information content, calculated by Genetic Power Calculator, (http://ibgwww.colorado.edu/~pshaun/gpc/mpic.html)
PIC 2 Percent information content calculated by MERLIN
HET% = percent heterozygosity
Markers in bold newly identified microsatellites from the draft equine sequence

Table A-10 Results of Linkage analysis for the autosomes. Location of markers rater than marker names given for each chromosome.

```
MERLIN 1.1.2 - (c) 2000-2007 Goncalo Abecasis

References for this version of Merlin:

   Abecasis et al (2002) Nat Gen 30:97-101        [original citation]
   Fingerlin et al (2004) AJHG 74:432-43          [case selection for association
studies]
   Abecasis and Wigginton (2005) AJHG 77:754-67   [ld modeling, parametric analyses]
   Fingerlin et al (2006) Gen Epidemiol 30:384-96 [sex-specific maps]
   Chen and Abecasis (2007) AJHG 81:913-26        [qtl association analysis, qtl
simulation]


The following parameters are in effect:
                     Data File :       ped09.dat (-dname)
                 Pedigree File :       ped09.ped (-pname)
            Missing Value Code :         -99.999 (-xname)
                      Map File :       ped09.map (-mname)
            Allele Frequencies : ALL INDIVIDUALS (-f[a|e|f|m|file])
                   Random Seed :          123456 (-r9999)

Data Analysis Options
         General : --error, --information, --likelihood, --model [param.tbl]
      IBD States : --ibd, --kinship, --matrices, --extended, --select
     NPL Linkage : --npl, --pairs, --qtl [ON], --deviates, --exp
      VC Linkage : --vc, --useCovariates, --ascertainment, --unlinked [0.00]
     Association : --infer, --assoc, --fastAssoc, --filter, --custom [cov.tbl]
     Haplotyping : --best, --sample, --all, --founders, --horizontal
   Recombination : --zero, --one, --two, --three, --singlepoint
       Positions : --steps, --maxStep, --minStep, --grid, --start, --stop
     LD Clusters : --clusters [], --distance, --rsq, --cfreq
          Limits : --bits [24], --megabytes, --minutes
     Performance : --trim, --noCoupleBits, --swap, --smallSwap
          Output : --quiet, --markerNames, --frequencies, --perFamily, --pdf,
                   --tabulate, --prefix [merlin]
      Simulation : --simulate, --reruns, --save, --trait []

Estimating allele frequencies... [using all genotypes]
   ...........................................................
Done estimating frequencies for 178 markers


Analysing Chromosome 1

Phenotype: Prize-loge [QTL] (21 families)
======================================================
          Pos   Zmean  pvalue    delta    LOD  pvalue
          min   -5.98     1.0   -0.324  -3.20     1.0
          max    6.88 0.00000    0.561   5.46 0.00000
        0.000    0.57     0.3    0.561   0.48    0.07
        9.000    0.99     0.2    0.561   0.97    0.02
       26.900    0.82     0.2    0.561   0.81    0.03
       36.900    0.69     0.2    0.561   0.63    0.04
       57.400    0.54     0.3    0.561   0.33    0.11
       78.700    0.55     0.3    0.359   0.21     0.2
       81.100    0.70     0.2    0.464   0.35    0.10
      113.700    0.69     0.2    0.389   0.27    0.13
      118.800    0.64     0.3    0.367   0.23    0.15
      165.200   -0.06     0.5   -0.035  -0.00     0.5
      165.300   -0.06     0.5   -0.038  -0.00     0.5
      177.400   -0.01     0.5   -0.013  -0.00     0.5
      183.200    0.04     0.5    0.045   0.00     0.5
      235.900   -0.09     0.5   -0.164  -0.02     0.6
```

Table A-10 (continued)

Analysing Chromosome 2
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.94	0.8	-0.324	-0.54	0.9
28.500	-0.29	0.6	-0.318	-0.09	0.7
65.500	-0.08	0.5	-0.129	-0.01	0.6
89.400	0.14	0.4	0.178	0.03	0.4
100.700	0.23	0.4	0.202	0.05	0.3
109.900	0.18	0.4	0.184	0.03	0.3
125.800	0.18	0.4	0.256	0.05	0.3
147.800	0.10	0.5	0.237	0.02	0.4

Analysing Chromosome 3
Phenotype: Prize-loge [QTL] (20 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.82	1.0	-0.324	-3.04	1.0
max	6.75	0.00000	0.561	5.21	0.00000
0.000	-0.66	0.7	-0.324	-0.30	0.9
18.600	-0.67	0.7	-0.324	-0.30	0.9
40.800	-0.10	0.5	-0.053	-0.00	0.6
47.400	0.68	0.2	0.397	0.26	0.14
56.900	0.95	0.2	0.542	0.52	0.06
75.400	0.78	0.2	0.420	0.34	0.11
88.800	0.26	0.4	0.319	0.08	0.3
101.000	0.51	0.3	0.561	0.34	0.11

Analysing Chromosome 4
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.27	0.6	-0.324	-0.09	0.7
17.000	-0.18	0.6	-0.161	-0.03	0.6
34.900	-0.00	0.5	-0.003	-0.00	0.5
57.300	0.05	0.5	0.050	0.00	0.5
71.100	0.07	0.5	0.076	0.01	0.4
81.300	-0.17	0.6	-0.214	-0.04	0.7
102.100	-0.26	0.6	-0.324	-0.10	0.7
120.800	0.09	0.5	0.151	0.01	0.4

Analysing Chromosome 5
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.33	0.6	-0.324	-0.12	0.8
13.900	-0.16	0.6	-0.160	-0.02	0.6
49.500	-0.25	0.6	-0.216	-0.05	0.7
56.200	-0.35	0.6	-0.282	-0.10	0.7
62.300	-0.36	0.6	-0.297	-0.10	0.8
78.000	-0.28	0.6	-0.267	-0.07	0.7
87.000	-0.29	0.6	-0.271	-0.08	0.7
114.500	-0.64	0.7	-0.324	-0.25	0.9
158.200	-0.39	0.7	-0.324	-0.17	0.8

Table A-10 (continued)

Analysing Chromosome 6
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.32	0.4	0.561	0.27	0.13
26.000	0.38	0.4	0.421	0.16	0.2
39.300	0.52	0.3	0.349	0.20	0.2
46.900	0.26	0.4	0.197	0.05	0.3
69.900	0.47	0.3	0.273	0.13	0.2
71.500	0.47	0.3	0.269	0.13	0.2
101.000	0.19	0.4	0.385	0.07	0.3

Analysing Chromosome 7
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.05	0.5	0.056	0.00	0.5
3.500	0.07	0.5	0.066	0.00	0.4
25.800	0.01	0.5	0.016	0.00	0.5
35.300	0.00	0.5	0.004	0.00	0.5
37.900	-0.04	0.5	-0.053	-0.00	0.5
69.600	-0.01	0.5	-0.036	-0.00	0.5

Analysing Chromosome 8
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.11	0.5	0.109	0.01	0.4
39.400	0.04	0.5	0.062	0.00	0.5
61.600	-0.13	0.6	-0.191	-0.02	0.6
77.800	0.27	0.4	0.516	0.13	0.2
93.500	-0.22	0.6	-0.255	-0.06	0.7

Analysing Chromosome 9
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.48	0.3	0.469	0.23	0.2
5.000	0.54	0.3	0.490	0.26	0.14
13.100	0.79	0.2	0.561	0.49	0.07
43.000	0.23	0.4	0.307	0.07	0.3
66.800	0.36	0.4	0.378	0.13	0.2

Analysing Chromosome 10
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.02	0.5	-0.019	-0.00	0.5
8.600	0.19	0.4	0.129	0.02	0.4
10.600	0.17	0.4	0.115	0.02	0.4
28.900	-0.10	0.5	-0.129	-0.01	0.6
37.400	-0.19	0.6	-0.226	-0.04	0.7
46.300	-0.28	0.6	-0.324	-0.12	0.8
60.600	-0.39	0.7	-0.324	-0.17	0.8
77.400	-0.39	0.7	-0.324	-0.20	0.8

Table A-10 (continued)

Analysing Chromosome 11
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.59	0.7	-0.324	-0.33	0.9
18.100	-0.51	0.7	-0.324	-0.22	0.8
27.700	-0.23	0.6	-0.227	-0.05	0.7
38.800	-0.16	0.6	-0.113	-0.02	0.6
49.600	-0.09	0.5	-0.088	-0.01	0.6
61.100	-0.18	0.6	-0.255	-0.05	0.7
98.400	0.18	0.4	0.255	0.05	0.3

Analysing Chromosome 12
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.49	0.7	-0.324	-0.23	0.8
19.500	-0.67	0.7	-0.324	-0.30	0.9
35.800	-0.40	0.7	-0.318	-0.13	0.8
48.500	-0.72	0.8	-0.324	-0.31	0.9
51.900	-1.09	0.9	-0.324	-0.56	0.9
54.200	-1.25	0.9	-0.324	-0.67	1.0
71.900	-0.47	0.7	-0.213	-0.11	0.8

Analysing Chromosome 13
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.91	0.2	0.561	0.80	0.03
25.800	1.09	0.14	0.561	0.76	0.03
40.800	1.04	0.15	0.561	0.67	0.04
95.100	0.16	0.4	0.450	0.07	0.3

Analysing Chromosome 14
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.45	0.3	0.561	0.45	0.08
33.200	0.69	0.2	0.561	0.58	0.05
44.900	0.84	0.2	0.561	0.71	0.04
120.300	0.33	0.4	0.211	0.07	0.3
136.700	0.49	0.3	0.372	0.18	0.2

Analysing Chromosome 15
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.07	0.5	0.130	0.01	0.4
24.700	0.49	0.3	0.496	0.26	0.14
51.900	0.27	0.4	0.307	0.08	0.3
56.900	0.37	0.4	0.373	0.14	0.2
109.000	0.24	0.4	0.333	0.08	0.3

Table A-10 (continued)

Analysing Chromosome 16
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.38	0.6	-0.324	-0.16	0.8
36.600	0.83	0.2	0.447	0.38	0.09
53.300	0.87	0.2	0.561	0.59	0.05
85.200	-0.20	0.6	-0.175	-0.03	0.7

Analysing Chromosome 17
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.07	0.5	-0.052	-0.00	0.6
11.400	-0.28	0.6	-0.173	-0.05	0.7
31.300	-0.12	0.5	-0.092	-0.01	0.6
41.500	-0.03	0.5	-0.024	-0.00	0.5
51.600	0.17	0.4	0.102	0.02	0.4
55.700	0.18	0.4	0.113	0.02	0.4
62.100	0.21	0.4	0.118	0.02	0.4

Analysing Chromosome 18
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.19	0.4	0.300	0.06	0.3
34.300	0.15	0.4	0.211	0.03	0.4
56.400	-0.53	0.7	-0.324	-0.25	0.9
63.400	-0.51	0.7	-0.324	-0.25	0.9
86.400	-0.39	0.7	-0.324	-0.17	0.8
111.400	-0.53	0.7	-0.324	-0.26	0.9

Analysing Chromosome 19
Phenotype: Prize-loge [QTL] (20 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.90	1.0	-0.324	-3.07	1.0
max	6.83	0.00000	0.561	5.27	0.00000
0.000	0.40	0.3	0.204	0.08	0.3
11.300	0.38	0.4	0.270	0.10	0.3
38.800	0.03	0.5	0.040	0.00	0.5

Analysing Chromosome 20
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.12	0.5	0.163	0.02	0.4
22.000	0.23	0.4	0.129	0.03	0.4
29.700	0.50	0.3	0.276	0.14	0.2
35.900	0.54	0.3	0.372	0.21	0.2
122.700	-0.30	0.6	-0.324	-0.15	0.8

Analysing Chromosome 21
Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.36	0.4	0.223	0.08	0.3
11.500	0.31	0.4	0.215	0.07	0.3
29.600	0.35	0.4	0.305	0.11	0.2

Table A-10 (continued)

Analysing Chromosome 22

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.23	0.6	-0.197	-0.04	0.7
5.000	-0.34	0.6	-0.253	-0.09	0.7
48.200	-0.38	0.6	-0.324	-0.16	0.8
93.100	-0.45	0.7	-0.324	-0.17	0.8

Analysing Chromosome 23

Phenotype: Prize-loge [QTL] (20 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.90	1.0	-0.324	-3.07	1.0
max	6.83	0.00000	0.561	5.27	0.00000
0.000	-0.62	0.7	-0.324	-0.32	0.9
27.200	-0.17	0.6	-0.248	-0.04	0.7
55.000	0.01	0.5	0.022	0.00	0.5
66.500	-0.06	0.5	-0.089	-0.00	0.6

Analysing Chromosome 24

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.11	0.5	-0.110	-0.01	0.6
28.700	-0.27	0.6	-0.251	-0.07	0.7
33.700	-0.19	0.6	-0.207	-0.04	0.7

Analysing Chromosome 25

Phenotype: Prize-loge [QTL] (19 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.66	1.0	-0.324	-2.88	1.0
max	6.82	0.00000	0.561	5.10	0.00000
0.000	0.68	0.2	0.561	0.53	0.06
46.100	-0.43	0.7	-0.324	-0.24	0.9

Analysing Chromosome 26

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.21	0.4	0.561	0.11	0.2
34.800	-0.26	0.6	-0.324	-0.09	0.7
74.700	-0.30	0.6	-0.324	-0.11	0.8

Analysing Chromosome 27

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.10	0.5	0.060	0.01	0.4
2.200	0.05	0.5	0.029	0.00	0.5
11.300	-0.28	0.6	-0.174	-0.05	0.7
20.900	-0.40	0.7	-0.304	-0.12	0.8
35.200	-0.62	0.7	-0.324	-0.25	0.9
65.800	-0.41	0.7	-0.324	-0.19	0.8

Table A-10 (continued)

Analysing Chromosome 28

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	0.61	0.3	0.561	0.40	0.09
11.500	0.52	0.3	0.352	0.20	0.2
17.900	0.51	0.3	0.356	0.19	0.2

Analysing Chromosome 29

Phenotype: Prize-loge [QTL] (21 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.98	1.0	-0.324	-3.20	1.0
max	6.88	0.00000	0.561	5.46	0.00000
0.000	-0.38	0.6	-0.324	-0.12	0.8
40.000	0.85	0.2	0.561	0.65	0.04
52.600	1.25	0.11	0.561	1.16	0.010

Analysing Chromosome 30

Phenotype: Prize-loge [QTL] (19 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.64	1.0	-0.324	-2.87	1.0
max	6.76	0.00000	0.561	5.06	0.00000
0.000	0.02	0.5	0.077	0.00	0.5
51.000	0.10	0.5	0.219	0.02	0.4

Analysing Chromosome 31

Phenotype: Prize-loge [QTL] (20 families)

Pos	Zmean	pvalue	delta	LOD	pvalue
min	-5.90	1.0	-0.324	-3.07	1.0
max	6.83	0.00000	0.561	5.27	0.00000
0.000	-0.55	0.7	-0.324	-0.24	0.9
6.100	-0.25	0.6	-0.198	-0.05	0.7
55.400	0.53	0.3	0.561	0.44	0.08

Table A-11 Results of Linkage analysis for the X chromosome

```
MERLIN 0.9.12b - (c) 2000-2003 Goncalo Abecasis
Modifications: CHROMOSOME-X

The following parameters are in effect:
                     Data File :          x.dat (-dname)
                 Pedigree File :          x.ped (-pname)
            Missing Value Code :        -99.999 (-xname)
                      Map File :          x.map (-mname)
            Allele Frequencies : ALL INDIVIDUALS (-f[a|e|f|file])
                   Random Seed :         123456 (-r9999)

Data Analysis Options
         General : --error, --information, --likelihood
      IBD States : --ibd, --kinship, --matrices
         Linkage : --npl, --pairs, --qtl [ON], --deviates
      VC Linkage : --vc, --useCovariates, --ascertainment, --unlinked [0.00]
     Haplotyping : --best, --sample, --all, --founders, --horizontal
   Recombination : --zero, --one, --two, --three, --singlepoint
       Positions : --steps, --maxStep, --minStep, --grid, --start, --stop
     Performance : --bits [24], --megabytes, --minutes, --trim
          Output : --quiet, --markerNames, --frequencies, --perFamily, --pdf
      Simulation : --simulate, --save
      Additional : --simwalk2, --swap

Estimating allele frequencies... [using all genotypes]
   AHT099 UMNe058 GKms UMNe148 UMNe402 PGK1ms LEX013 TKY020

Phenotype: Prize-loge [QTL] (18 families)
============================================================
                 Pos   Zmean  pvalue    delta    LOD  pvalue
                 min   -4.43     1.0   -0.358  -2.48     1.0
                 max    5.91 0.00000    0.729   5.38 0.00000
               0.000   -0.31     0.6   -0.261  -0.08     0.7
              30.400   -0.29     0.6   -0.254  -0.07     0.7
              47.600   -0.42     0.7   -0.230  -0.09     0.7
              65.700   -1.10     0.9   -0.358  -0.56     0.9
              87.300   -0.65     0.7   -0.358  -0.25     0.9
             134.500   -1.34     0.9   -0.358  -0.79     1.0
             164.000   -0.98     0.8   -0.358  -0.55     0.9
             185.800   -0.18     0.6   -0.225  -0.04     0.7
```

Table A-12 The 75 most significant results from 1000 simulated data sets produced by simulate option in Merlin, highlighted in green are the most significant 5%, shows that for the given data set LOD scored under 1.54 occur by chance more than 5% of the time.

	Zmean	pvalue	delta	LOD	pvalue
1	2.32	0.01	0.561	2.09	0.001
2	2.57	0.005	0.561	2.06	0.001
3	2.19	0.014	0.561	2.03	0.0011
4	2.17	0.015	0.561	1.92	0.0015
5	2.14	0.02	0.561	1.85	0.002
6	2.41	0.008	0.561	1.85	0.002
7	2.03	0.02	0.561	1.83	0.002
8	2	0.02	0.561	1.83	0.002
9	1.94	0.03	0.561	1.8	0.002
10	2.01	0.02	0.561	1.79	0.002
11	2.18	0.015	0.561	1.78	0.002
12	2.01	0.02	0.561	1.75	0.002
13	2.03	0.02	0.561	1.75	0.002
14	1.99	0.02	0.561	1.73	0.002
15	1.96	0.02	0.561	1.71	0.003
16	2.05	0.02	0.561	1.71	0.003
17	1.86	0.03	0.561	1.7	0.003
18	1.93	0.03	0.561	1.68	0.003
19	1.86	0.03	0.561	1.67	0.003
20	1.87	0.03	0.561	1.67	0.003
21	1.98	0.02	0.561	1.66	0.003
22	1.93	0.03	0.561	1.65	0.003
23	1.85	0.03	0.561	1.65	0.003
24	1.74	0.04	0.561	1.65	0.003
25	1.97	0.02	0.561	1.63	0.003
26	1.96	0.02	0.561	1.62	0.003
27	1.97	0.02	0.561	1.62	0.003
28	1.85	0.03	0.561	1.62	0.003
29	1.96	0.03	0.561	1.62	0.003
30	1.96	0.03	0.561	1.62	0.003
31	1.91	0.03	0.561	1.61	0.003
32	2.2	0.014	0.561	1.6	0.003
33	1.74	0.04	0.561	1.6	0.003
34	1.77	0.04	0.561	1.59	0.003
35	1.92	0.03	0.561	1.58	0.004
36	1.87	0.03	0.561	1.58	0.003
37	1.88	0.03	0.561	1.57	0.004
38	1.83	0.03	0.561	1.57	0.004
39	1.76	0.04	0.561	1.56	0.004
40	1.96	0.03	0.561	1.56	0.004
41	1.73	0.04	0.561	1.55	0.004
42	1.78	0.04	0.561	1.55	0.004
43	1.72	0.04	0.561	1.55	0.004
44	1.74	0.04	0.561	1.55	0.004
45	1.77	0.04	0.561	1.55	0.004
46	1.79	0.04	0.561	1.55	0.004
47	1.87	0.03	0.561	1.54	0.004
48	1.75	0.04	0.561	1.54	0.004
49	1.86	0.03	0.561	1.54	0.004
50	**1.96**	**0.03**	**0.561**	**1.54**	**0.004**
51	1.75	0.04	0.561	1.53	0.004
52	1.84	0.03	0.561	1.52	0.004
53	1.68	0.05	0.561	1.52	0.004
54	1.73	0.04	0.561	1.52	0.004
55	1.75	0.04	0.561	1.52	0.004
56	1.91	0.03	0.561	1.52	0.004
57	1.68	0.05	0.561	1.52	0.004
58	1.61	0.05	0.561	1.51	0.004
59	1.69	0.05	0.561	1.51	0.004
60	1.92	0.03	0.561	1.51	0.004
61	1.64	0.05	0.561	1.51	0.004
62	1.61	0.05	0.561	1.5	0.004
63	1.76	0.04	0.561	1.5	0.004
64	1.96	0.02	0.561	1.5	0.004
65	1.84	0.03	0.561	1.5	0.004
66	1.77	0.04	0.561	1.49	0.004
67	1.7	0.04	0.561	1.49	0.004
68	1.63	0.05	0.561	1.49	0.004
69	1.81	0.04	0.561	1.49	0.004
70	1.65	0.05	0.561	1.48	0.005
71	1.84	0.03	0.561	1.48	0.005
72	1.6	0.05	0.561	1.48	0.005
73	1.69	0.05	0.561	1.48	0.004
74	1.83	0.03	0.561	1.47	0.005
75	1.66	0.05	0.561	1.47	0.005

Table A-13 Results for family based association analysis carried out in QTDT using 66 microsatellites adjacent to candidate genes. P values are not displayed as no marker achieved statistical significance.

```
QTDT - Quantitative TDT 2.6.0
(c) 1998-2007 Goncalo Abecasis (goncalo@umich.edu)

This program implements tests described by
Abecasis et al, Am J Hum Genet 66:279-292 (2000)
Abecasis et al, Eur J Hum Genet 8:545-551 (2000)
and others

The following parameters are in effect:
           QTDT Data File :              adj.dat -dname
       QTDT Pedigree File :              adj.ped -pname
     QTDT IBD Status File :              adj.ibd -iname
       Missing Value Code :              -99.999 -xname
               Covariates :       USER SPECIFIED —c p|s|u|-})
        Association Model :           ORTHOGONAL —a a|d|f|m|o|p|r|t|w|-]
     Full Model Variances :         NOT MODELLED —v e|c|g|n|t|a|d|-})
     Null Model Variances :           NON SHARED —w e|c|g|n|t|a|d|-}
                          &            POLYGENIC
                          &             ADDITIVE
 Parent of Origin Effects :                 NONE -o[f|t|m|p|-]
 Monte-Carlo Permutations :                    0 -m9999
              Random Seed :               123456 -r9999
        Numeric Minimizer :      NELDER AND MEAD -n[f|n|p]
     Transmission Scoring :        FULL PEDIGREE -t[n|p]

Additional Options
 --dominance, --snp, --multi-allelic [ON], --deviates, --references,
 --exclude-founder-phenotypes, --p-values, --no-regress-tbl

Online documentation http://www.sph.umich.edu/csg/abecasis/QTDT
Comments, bugs: goncalo@umich.edu

The following models will be evaluated...
  NULL MODEL
     Means = Mu + B
 Variances = Ve + Vg + Va

  FULL MODEL
     Means = Mu + B + W
 Variances = Ve + Vg + Va

  Likelihood ratio statistic from Abecasis et al (AJHG, 2000)

Lumping alleles with frequencies of 0.05 or less...

Testing trait:                    Prize-loge
=============================================

Testing marker:                        1CA30
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.48     118   270.40    0.18          ( 123/124 probands)

Testing marker:                       LEX039
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.86     118   270.82    0.08          ( 123/124 probands)

Testing marker:                       LEX020
---------------------------------------------

Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   268.41     117   268.22    0.36          ( 123/124 probands)
```

Table A-13 (continued)

```
Testing marker:                         AGTms
---------------------------------------------

Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
All     119   270.71     118   270.51    0.40         ( 123/124 probands)

Testing marker:                        UCD440
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   269.84     115   269.33    1.03         ( 123/124 probands)

Testing marker:                         HMS15
---------------------------------------------


 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   267.27     113   266.54    1.46         ( 123/124 probands)

Testing marker:                         1CA16
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   269.51     116   269.51    0.00         ( 123/124 probands)

Testing marker:                        TKY850
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   270.43     116   270.42    0.03         ( 123/124 probands)

Testing marker:                        UCP1ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   267.55     116   267.55    0.00         ( 122/123 probands)

Testing marker:                       FABP2ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.57     119   270.57    0.00         ( 123/124 probands)

Testing marker:                        CETPms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     114   262.89     112   262.49    0.81         ( 120/121 probands)

Testing marker:                        COR033
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     111   262.18     107   261.68    1.00         ( 119/120 probands)

Testing marker:                       TKY1085
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     114   264.56     112   264.43    0.26         ( 120/121 probands)

Testing marker:                    PPARGC1Ams
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     115   264.24     114   264.01    0.46         ( 120/121 probands)
```

Table A-13 (continued)

```
Testing marker:                        TKY942
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.56     117   270.22    0.68          ( 123/124 probands)

Testing marker:                        PON1ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   264.65     115   264.56    0.20          ( 122/123 probands)

Testing marker:                         IL6ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.30     118   270.22    0.15          ( 123/124 probands)

Testing marker:                         HMS19
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.84     119   270.84    0.00          ( 123/124 probands)

Testing marker:                         HMS09
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.16     117   270.13    0.06          ( 123/124 probands)

Testing marker:                        NOS3ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   269.23     113   268.84    0.78          ( 123/124 probands)

Testing marker:                        TKY508
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.26     117   270.13    0.26          ( 123/124 probands)

Testing marker:                        TKY521
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   269.19     117   269.14    0.10          ( 123/124 probands)

Testing marker:                        AHT050
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   270.33     115   270.16    0.35          ( 123/124 probands)

Testing marker:                       AMPD1ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     114   264.89     112   264.79    0.21          ( 121/122 probands)

Testing marker:                        LEPRms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     111   253.54     109   253.10    0.88          ( 116/117 probands)

Testing marker:                        GNB3ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   270.46     114   269.92    1.08          ( 123/124 probands)
```

Table A-13 (continued)

```
Testing marker:                         VDRms
---------------------------------------------

Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   268.17     116   267.82    0.69           ( 123/124 probands)

Testing marker:                        COR070
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     114   266.13     112   265.77    0.72          ( 123/124 probands)

Testing marker:                       APOC3ms
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   270.69     113   269.97    1.44          ( 123/124 probands)

Testing marker:                       UCP2ms3
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     115   267.41     114   267.21    0.38          ( 122/123 probands)

Testing marker:                        AHT019
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   268.68     118   268.50    0.36          ( 123/124 probands)

Testing marker:                        AHT005
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   268.94     115   268.61    0.65          ( 123/124 probands)

Testing marker:                        COR003
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   270.14     114   269.87    0.53          ( 123/124 probands)

Testing marker:                       NVHEQ18
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.15     116   269.56    1.17          ( 123/124 probands)

Testing marker:                        ACEms1
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   270.83     115   270.45    0.77          ( 123/124 probands)

Testing marker:                        SGCV24
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     112   257.13     109   256.61    1.04          ( 120/121 probands)

Testing marker:                        AHT076
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.83     118   270.63    0.41          ( 123/124 probands)

Testing marker:                     ACADVLms3
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     115   265.49     114   265.46    0.06          ( 121/122 probands)
```

Table A-13 (continued)

Testing marker: TKY404

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	117	270.69	115	270.54	0.30		(123/124 probands)

Testing marker: UCD411

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	116	267.41	114	267.21	0.39		(122/123 probands)

Testing marker: ACTN3ms

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	118	271.00	117	270.71	0.57		(123/124 probands)

Testing marker: COR058

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	115	269.58	113	269.22	0.72		(123/124 probands)

Testing marker: AHT017

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	115	269.90	112	269.44	0.93		(123/124 probands)

Testing marker: TKY585

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	117	270.60	116	270.55	0.11		(123/124 probands)

Testing marker: UM032

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	118	266.81	116	266.56	0.49		(123/124 probands)

Testing marker: ADRB2ms

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	117	265.48	115	265.48	0.00		(123/124 probands)

Testing marker: TKY480

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	118	270.30	116	270.07	0.47		(123/124 probands)

Testing marker: AHT037

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	115	270.28	111	269.66	1.25		(123/124 probands)

Testing marker: TKY936

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	116	269.03	113	268.60	0.87		(123/124 probands)

Testing marker: COR105

Allele	df(0)	-LnLk(0)	df(T)	-LnLk(T)	ChiSq	p	
All	117	269.62	114	269.41	0.41		(123/124 probands)

Table A-13 (continued)

```
Testing marker:                       COR007
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p

    All     118   269.90     116   269.54    0.72           ( 123/124 probands)
Testing marker:                        HTG17
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.68     118   270.53    0.31           ( 123/124 probands)

Testing marker:                      GDF8ms2
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   269.58     114   269.39    0.37           ( 123/124 probands)

Testing marker:                       TKY730
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.49     117   270.26    0.47           ( 123/124 probands)

Testing marker:                       AHT018
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   270.22     114   269.98    0.48           ( 123/124 probands)

Testing marker:                      UMNe056
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     115   264.41     113   264.09    0.65           ( 123/124 probands)

Testing marker:                        UM011
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   268.96     117   268.81    0.29           ( 123/124 probands)

Testing marker:                       SGCV01
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     119   270.82     118   270.55    0.55           ( 123/124 probands)

Testing marker:                       SGCV04
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   268.57     114   268.03    1.08           ( 123/124 probands)

Testing marker:                       COR024
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   269.66     115   269.65    0.03           ( 123/124 probands)

Testing marker:                     ADRB3ms2
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     116   270.19     113   269.61    1.17           ( 123/124 probands)

Testing marker:                       TKY872
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   269.77     117   269.54    0.47           ( 123/124 probands)
```

Table A-13 (continued)

```
Testing marker:                      UCDEQ425
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   261.90     115   261.70    0.40          ( 123/124 probands)
Testing marker:                          L12.2
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     115   265.82     112   265.51    0.63          ( 122/123 probands)

Testing marker:                         LEX075
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     118   270.83     117   270.68    0.30          ( 123/124 probands)

Testing marker:                         TKY105
---------------------------------------------

 Allele   df(0) -LnLk(0)   df(T) -LnLk(T)   ChiSq       p
    All     117   267.52     115   267.23    0.58          ( 123/124 probands)


Run completed on Fri Jan 23 09:06:21 2009
66 tests carried out

The most significant result refers to:
    Trait: Prize-loge
   Marker: VDRms
    ChiSq: 0.694
  p-value: 0.4047

Overall Bonferroni significance level: 1.0000
```

APPENDIX B

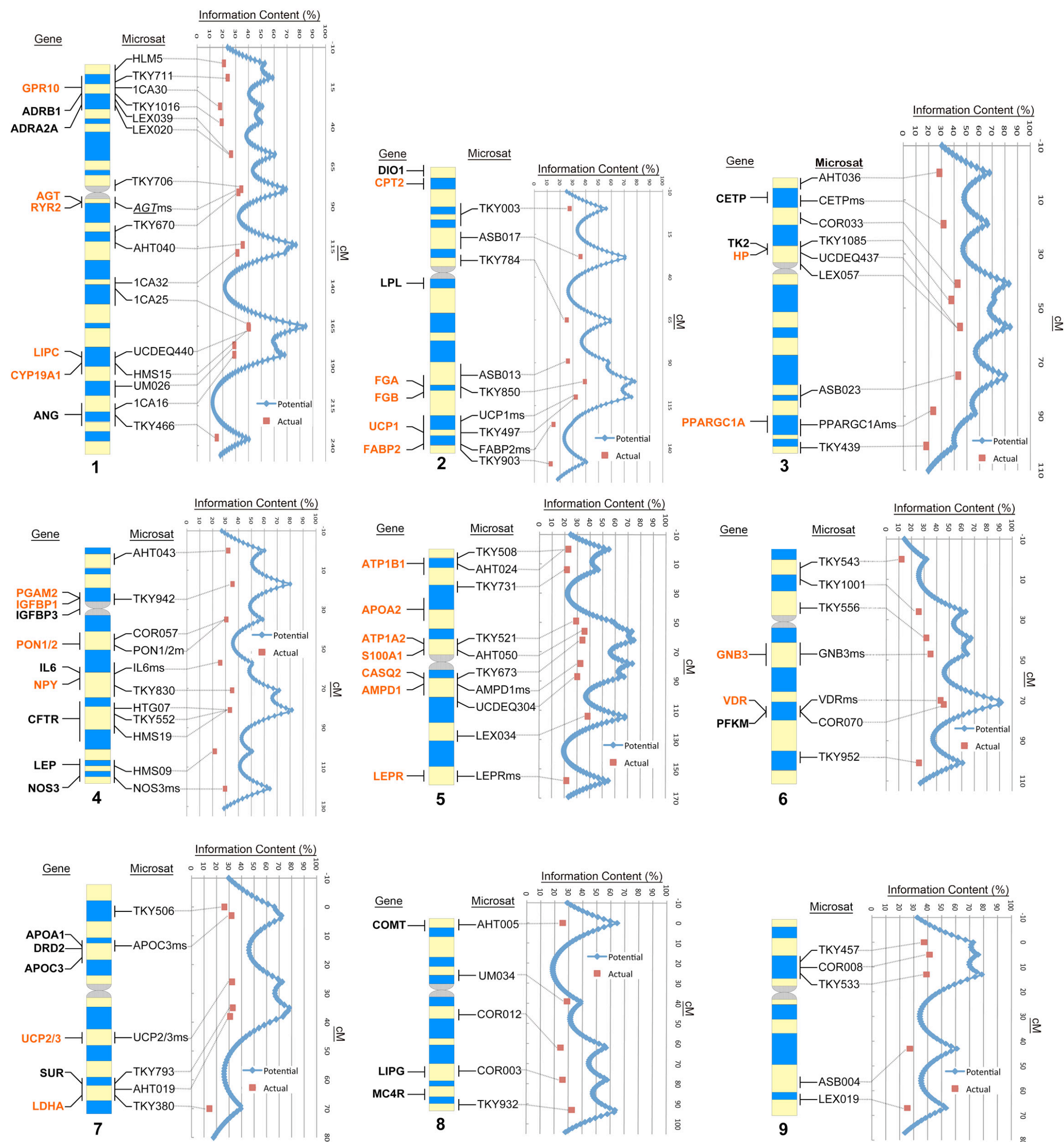

Figure B-1 Location of candidate genes and microsatellites used in the genome scanning panel. Blue line represents the multipoint information content (MPIC) calculation based on marker density and level of polymorphism for each marker (assumes genotypes for parents and offspring available). Red boxes represents the actual percent information content calculated in MERLIN and based on the sample population used in the study.

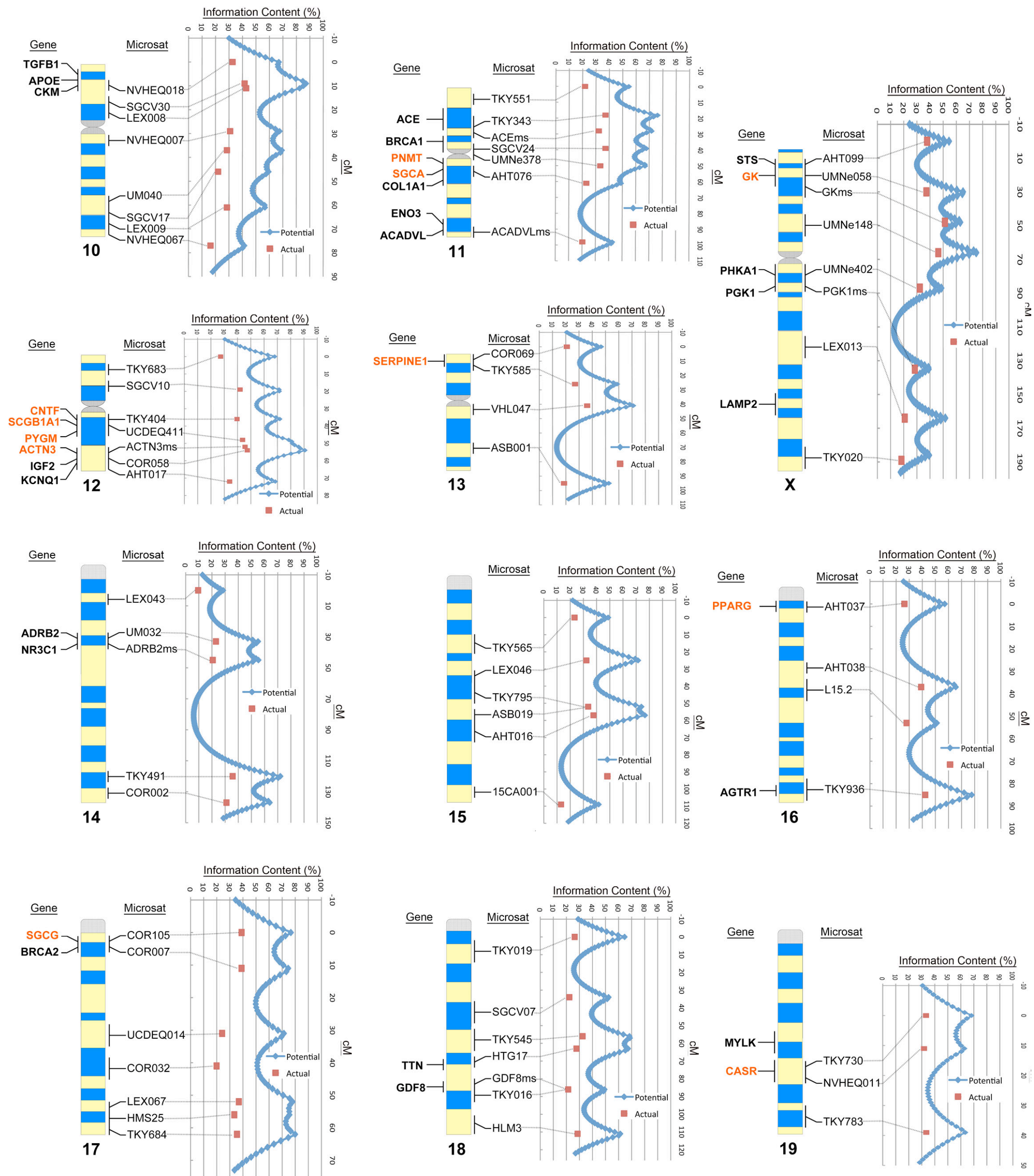

Figure B-1 (continued)

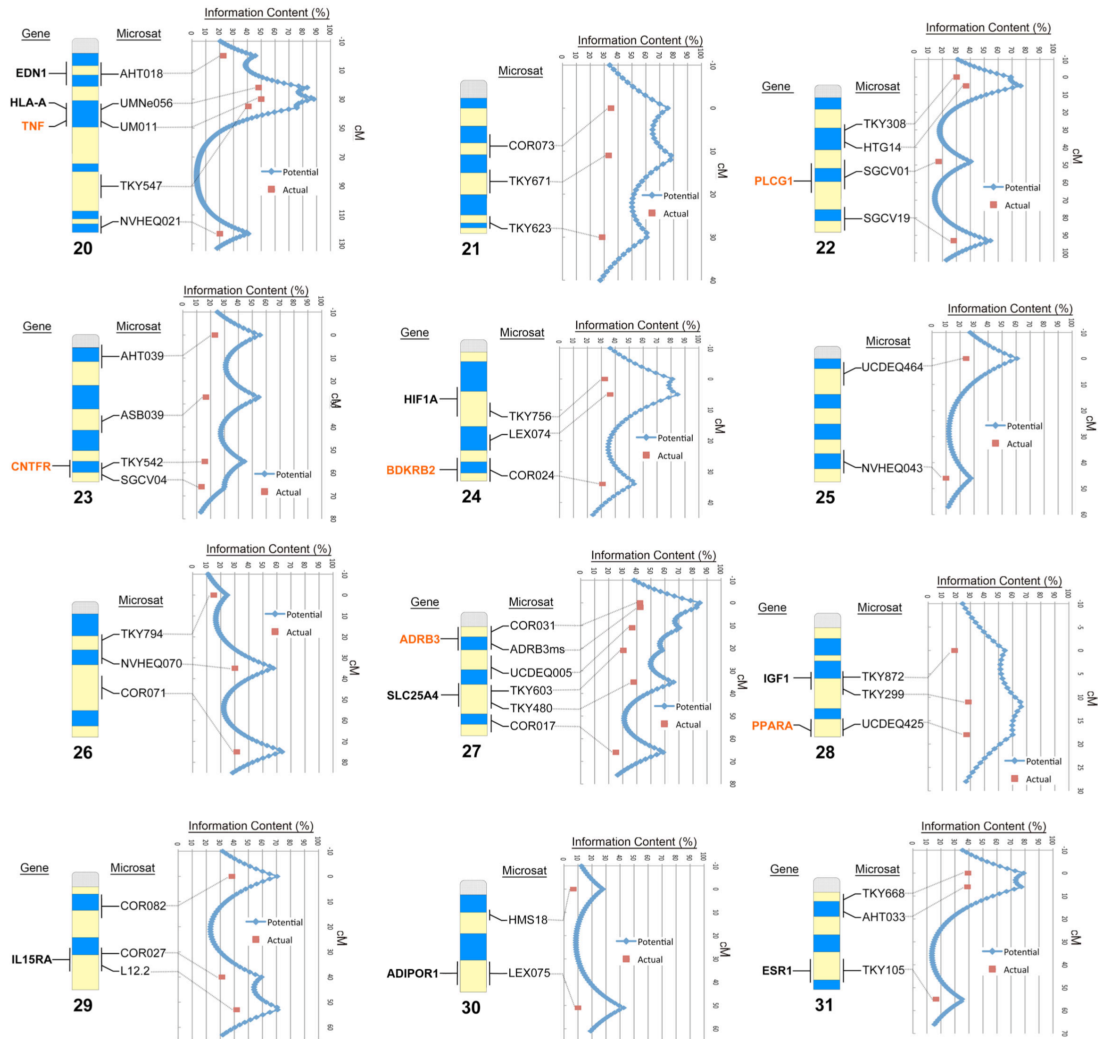

Figure B-1 (continued)

VITA

Name: Keith William Durkin

Address: Department of Veterinary Integrative Biosciences, Texas A & M University, Mail Stop 4458, College Station, TX 77843, USA.

Email Address: keithdurkin@gmail.com

Education: B.S., Equine Science, University of Limerick, Limerick, Ireland
Graduation date: August 2002

Ph.D., Genetics, Texas A&M University, College Station, Texas
Graduation date: May 2009

Publications:

Brito LFC, Sertich PL, Durkin K, Chowdhary BP, Turner RM, Greene LM, McDonnell S. (2008) Autosomic 27 trisomy in a standardbred colt. *Journal of Equine Veterinary Science* **28,** 431-6

Raudsepp T, Gustafson-Seabury A, Durkin K, Wagner ML, Goh G, Seabury CM, Brinkmeyer-Langford C, Lee EJ, Agarwala R, Stallknecht Rice E, Schaffer AA, Skow LC, Tozaki T, Yasue H, Penedo MCT, Lyons LA, Khazanehdari KA, Binns MM, MacLeod JN, Distl O, Guerin G, Leeb T, Mickelson JR, Chowdhary BP. (2008) A 4103 marker integrated map of the horse genome. *Cytogenetic and Genome Research* **122**, 28-36.

Goh G, Raudsepp T, Durkin K, Wagner ML, Schaffer AA, Agarwala R, Tozaki T, Mickelson JR, Chowdhary BP. (2007) High-resolution gene maps of horse chromosomes 14 and 21: Additional insights into evolution and rearrangements of HSA5 homologs in mammals. *Genomics* **89**, 89-112.

Perrocheau M, Boutreux V, Chadi S, Mata X, Decaunes P, Raudsepp T, Durkin K, Incarnato D, Iannuzzi L, Lear TL, Hirota K, Hasegawa T, Zhu B, de Jong P, Cribiu EP, Chowdhary BP, Guerin G. (2006) Construction of a medium-density horse gene map. *Animal Genetics* **37**, 145-55.