Polymorphism Identification and Association Analysis of *MC4R* and *POMC* Genes

Honors Thesis

Presented to the College of Agriculture and Life Sciences,

Department of Animal Sciences

Of Cornell University

In Partial Fulfillment of the Requirements for the

Research Honors Program

By

Caitlin D. Armstrong

December 2009

Dr. Samantha Brooks

Table of Contents

Section **Pages(s)**

Abstract

Identification of horses at high risk for developing laminitis, such as those predisposed to Equine Metabolic Syndrome (EMS), is an essential component of any equine management strategy. Pro-opiomelanocortin (*POMC)* and melanocortin 4-receptor (*MC4R)* are two genes expressed in neurons of the hypothalamus that are involved in the regulation of energy homeostasis in mammalian species. Re-sequencing of these two genes revealed a total of five synonymous SNPS; one in *MC4R* and four in *POMC*. The *MC4R* SNP showed an association with body condition score (BCS) and a significant association with light and dark shades of coat color. Three of the POMC SNPs were located in coding regions. All three were synonymous. The last *POMC* SNP was located upstream of the coding region of the gene, and is potentially regulatory. Horses were genotyped for the *MC4R* SNP using Fluorescence Resonance Energy Transfer (FRET) technology and melting curve analysis. New sequence data was generated for *POMC* and four distinct haplotypes were identified within the gene. Comparisons between the protein sequences of the two equine genes and other mammalian species show a high amount of evolutionary conservation. Additional work should focus on elucidating the exact nature of the link between the *MC4R* SNP and coat color, gathering data on more horses to determine if there is statistical significance between the *MC4R* SNP and BCS, as well as work to determine if the regulatory mutation in *POMC* affects transcript copy number in the hypothalamus. Either gene could affect adiposity in the horse, which is a known risk factor for both EMS and laminitis.

Acknowledgements

I would like to express my gratitude to Dr. Samantha Brooks for mentoring me as an undergraduate researcher. She patiently answered all my questions and guided me along, which was especially appreciated when we were working in the dark (one of the genes was missing part of the genomic sequence) and my primers repeatedly failed to work. A love of all things “horse” has been an integral part of who I am for as long as I can remember, and I feel extraordinarily lucky to have been able to do a research project in a subject area that has always been meaningful for me.

In addition, I would like to thank the other undergraduates in the laboratory, who have always been supportive of my work, and served as wonderful colleagues to bounce ideas off of and to help figure out my next step.

Next I would like to thank my friends and family, who have been behind me, no matter what endeavors I embark on.

Finally, I would like to thank the organizations that have provided the funding for my research. In particular, this work was funded in part by a grant by the USDA as well as a gift from the Animal Health Foundation.

Introduction:

A clear understanding of equine laminitis has been, and continues to be, a priority research area for the horse industry. The disease process is not well understood, although considerable progress has been made in the last 20 or so years. One clear predisposing factor to development of laminitis is Equine Metabolic Syndrome. EMS is characterized by regional adiposity, hyperinsulinemia, peripheral insulin resistance, and normal levels of adrenocorticotropic hormone (Walsh, McGowan, McGowan, Lamb, J, & Place, 2007). Studies using mouse models have uncovered key hormones, neurons, and signaling molecules that are involved in regulation of mammalian feed intake and energy expenditure. Two genes of interest, pro-opiomelanocortin (*POMC*) and melanocortin-4 receptor (*MC4R*) are involved in this pathway. POMC protein, produced by POMC neurons in the hypothalamus, undergoes posttranslational cleavage to produce α-melanocyte-stimulating hormone (α-MSH), which serves as a ligand for the MC4R. POMC protein production is stimulated by high circulating levels of insulin and leptin, which are signals of positive energy balance. Stimulation of the G-protein coupled receptor MC4R leads to decreased feed intake and increased energy expenditure. Knockout studies in mice, as well as association studies in *MC4R* and *POMC* variants in chickens, pigs, and humans, underscore the importance of functional MC4R protein and POMC protein for proper regulation of this pathway.

MC4R and *POMC* are therefore attractive candidate genes for study in horses. This work has several aims:

1) Characterization of polymorphisms in the two genes of interest.

2) Determination of coding regions.

3) Comparison of equine sequence with other mammalian species.

4) Determination of association, if any, between genotype and phenotype.

Literature Review

Laminitis is undoubtedly the most despised word in any horse person's vocabulary. Anyone who has spent time with horses knows that laminitis is not a well characterized disease process. Research in this area has revealed a myriad of trigger factors, including ingestion of lush spring pasture (Geor, 2009), induction of hyperinsulinemia (Asplin, Sillence, Pollit, & McGowan, 2007), carbohydrate overload (Nourian, GI, van Eps, & Pollitt, 2007), and black walnut ingestion (Galey, Whiteley, Goetz, AR, CA, & VR, 1991). Equine Metabolic Syndrome (EMS), and increased adiposity in general, are highly correlated with the development of laminitis (Geor, 2008). EMS is characterized by hyperinsulinemia, normal levels of adrenocorticotropic hormone (ACTH), peripheral insulin resistance, and increased regional fat deposits, especially on the dorsal areas of the neck, withers, and back (Walsh, McGowan, McGowan, Lamb, J, & Place, 2007). Plasma insulin levels also correlate with severity of laminitis (Walsh, McGowan, McGowan, Lamb, J, & Place, 2007). Metabolic regulation appears to have a role in the laminitis disease process, but the exact nature of this association remains unclear.

Given that adiposity is a critical factor in the development of laminitis, it is logical to focus attention on the central pathway regulating adiposity and energy balance in mammals. Feed intake and energy expenditure are regulated by the hormones leptin, insulin, and ghrelin (Barsh & Schwartz, 2002). Figure 1 illustrates the roles that two genes of interest, melanocortin-4 receptor (*MC4R*) and pro-opiomelanocortin (*POMC*) have in this pathway. *POMC* is a large protein produced by *POMC* neurons. It undergoes posttranslational proteolytic cleavage to yield different products, including ACTH and α-melanocyte stimulating hormone (α-

MSH) (O'Rahilly, Yeo, & Sadaf Farooqi, 2004). As shown in Figure 1, the α-MSH protein is a ligand for *MC4R*, a G-protein coupled receptor (O'Rahilly, Yeo, & Sadaf Farooqi, 2004).

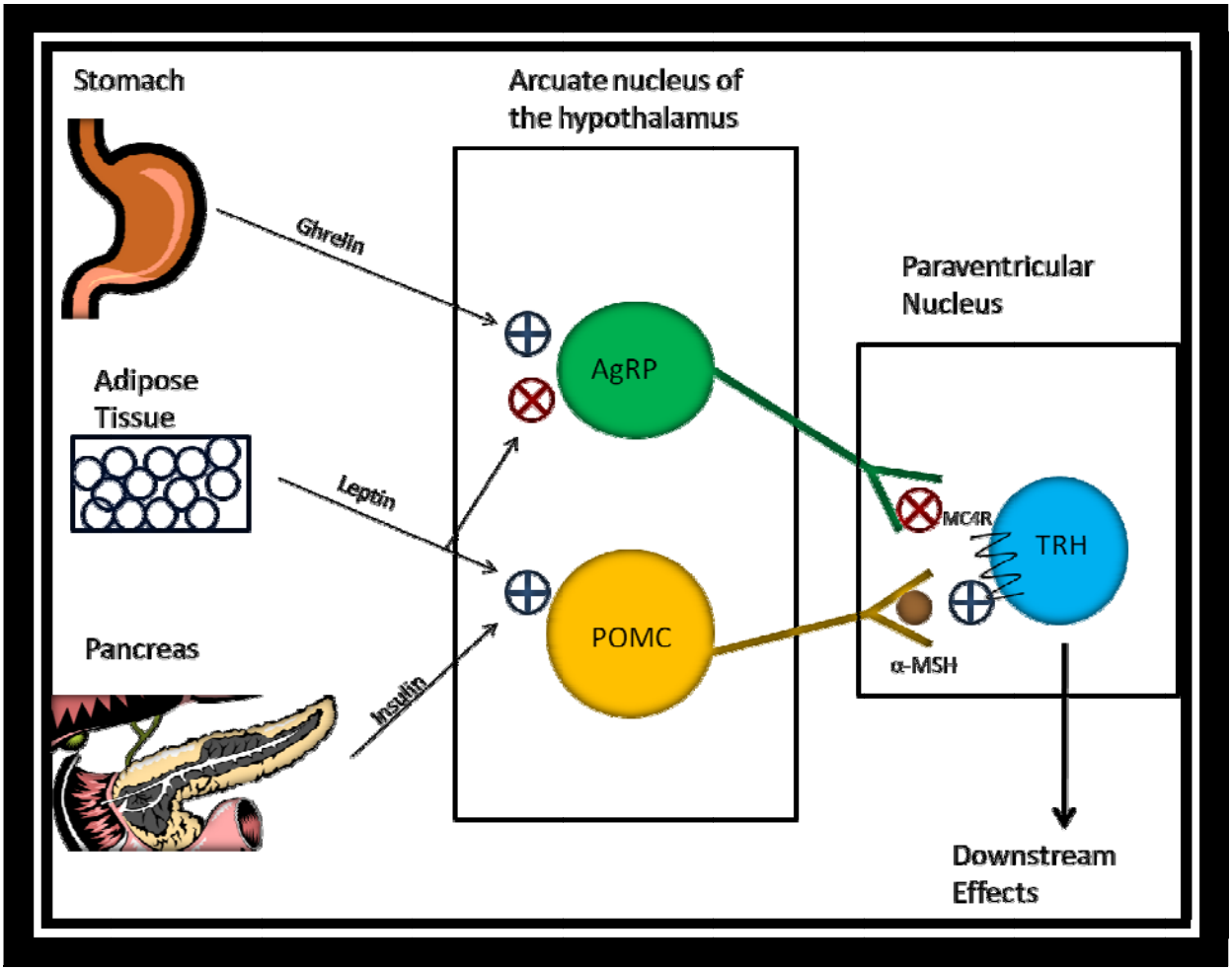


Figure 1: Signaling Pathway of *MC4R* and *POMC*.
Input coming into the system reflects the energy status of the animal. If the animal is in positive energy balance, then there will be high levels of insulin and leptin. POMC neurons will be stimulated, (indicated by the blue circle with a plus sign) and AgRP neurons will be inhibited (indicated by the red circle with the "x"). POMC neurons will then release α-MSH, a ligand for the MC4R on the secondary neuron. The secondary neuron will be stimulated, which will produce the downstream effects of decreased food intake and increased energy expenditure. If the animal is in negative energy balance, insulin and leptin levels will be low, and ghrelin levels will be high. AgRP neurons will be stimulated, and POMC neurons will be inhibited. The secondary neuron will be inhibited, and the animal will show the opposite downstream effects (Barsh & Schwartz, 2002), (O'Rahilly, Yeo, & Sadaf Farooqi, 2004).

Leptin, a small protein hormone, circulates in the bloodstream in proportion to the amount of fat stored in adipose tissue (Barsh & Schwartz, 2002). When circulating levels of leptin are high, this signals to the hypothalamus that the body has an excess of fat stores. The animal therefore reduces its feed intake and increases its energy expenditure. Insulin, released from the pancreas, circulates when blood glucose concentration is high. Similarly, when insulin levels are high, this signals that the animal has an excess of carbohydrate. Taken together, the two hormones reflect the relatively high amounts of adipose tissue reserves and carbohydrates present, both of which can be utilized by the various metabolic pathways to yield energy for the animal. The two hormones act on two classes of neurons located in the arcuate nucleus of the hypothalamus, which both express leptin receptors (Barsh & Schwartz, 2002). Insulin and leptin stimulate POMC neurons while simultaneously inhibiting Agouti-related peptide (AgRP) neurons. POMC neurons, in turn, signal for decreased feed intake via TRH neurons located in the paraventricular nucleus of the hypothalamus (Barsh & Schwartz, 2002). By stimulating the neurons that signal for the animal to stop eating (*POMC*) and inhibiting the neurons that signal for the animal to eat (*AgRP*), the net result is that the animal will reduce its feed intake (Barsh & Schwartz, 2002). When the opposite is true; that is, that the animal is thin, then leptin will be low. Low leptin stimulates AgRP neurons and inhibits POMC neurons, and the animal will increase its feed intake.

Another input into the POMC/AgRP signaling system is the hormone ghrelin. When ghrelin levels are high, the animal will also increase its feed intake (Barsh & Schwartz, 2002). Ghrelin, produced by the stomach, signals that the animal is hungry. Ghrelin stimulates *AgRP* neurons (Barsh & Schwartz, 2002), therefore inhibiting TRH neurons. Feed intake is increased

and energy expenditure is decreased. While this is not a focus of this paper, it should be noted that the signaling pathway has multiple inputs, beyond just insulin and leptin. Additionally, there are multiple secondary neurons in the brain which express MC4Rs and modulate feed intake and energy expenditure in addition to TSH neurons (Baskin, 2006).

The TRH released by the TRH neurons in the paraventricular nucleus stimulates the release of TSH by the anterior pituitary, which in turn stimulates thyroid hormone production and therefore metabolic rate (Barsh & Schwartz, 2002). Dopaminergic pathways, which direct motor activity as well as motivation and reward networks, regulate and utilize information from downstream targets in the hypothalamus. Integration of the various dopaminergic pathways has been postulated to regulate feeding behavior (Barsh & Schwartz, 2002). Therefore, when the animal is in a state of positive energy balance, insulin and leptin levels are high, and ghrelin levels are low. This stimulates *POMC* neurons while inhibiting *AgRP* neurons. TRH neurons are stimulated, and the net effect is that the animal decreases its feed intake while simultaneously increasing its energy expenditure. When the animal is in a state of negative energy balance, the opposite is true. This simple feedback loop, which originates in the hypothalamus, contributes to a multitude of other pathways which regulate complex behaviors and energy balance.

Additional research has been done on *MC4R* mutants in several species. An early study in mice (Huszar, et al., 1997) generated *MC4R* knockouts. The mutant mice showed distinct phenotypes. Knockout mice showed increased weight gain, linear growth, feed intake, blood insulin, and blood leptin levels. Heterozygous mice showed the same phenotypic changes, just to a lesser degree. Considering *MC4R's* role in energy homeostasis, it makes sense that knocking it out would yield a fat mouse that ate a lot. High levels of insulin and leptin were

trying to signal that the animal had an excess of energy, but the transducer of that signal to decrease appetite and stimulate energy utilization had been removed. The prevalence of *MC4R* variants in humans (Santoro & Cirillo, 2009) selected for the characteristic phenotype (obese, tall, family history of obesity) is 1.6% in one population. A study done with pigs (Kim, Larsen, Short, Plastow, & Rothschild, 2000) showed that polymorphisms in *MC4R* correlated with similar changes: backfat thickness increased, growth rate increased, and feed intake increased. In chickens, *MC4R* SNPS were further correlated with variations in body weight (Xuemei, Ning, Xuemei, Xingbo, Qingyong, & Xiuli, 2006).

Notably, α-MSH is also a ligand for another melanocortin receptor, *MC1R*, which is known to influence coat color in mice, horses, and cattle (Marklund, Moller, Sandberg, & Anderson, 1996). Mutations in the *MC1R* gene are responsible for coat colors that appear red, such as the chestnut color in horses. Normal signaling for production of the darker eumelanin is blocked by the absence of a functional receptor in the melanocyte. One study found (Krude, Biebermann, Luck, Horn, Brabant, & Gruters, 1998) several human subjects who had a mutation in *POMC* which prematurely stopped transcription of the *POMC* protein, including the portions that were later cleaved into ACTH and α-MSH. The resulting phenotype, as predicted, was red hair, obesity, and adrenal insufficiency. ACTH is a ligand for another melanocortin receptor, *MC2R*, located in the adrenal cortex of the adrenal gland (O'Rahilly, Yeo, & Sadaf Farooqi, 2004). Lack of α-MSH led to the red hair phenotype, as there was no signal for *MC1R* to stimulate eumelanin production. It also led to the development of obesity, since an important stimulatory input on *MC4R* had been removed. The absence of ACTH led to problems with adrenal function, as the necessary ligand for *MC2R* no longer existed. A clear link between

MC4R and POMC and an EMS-like phenotype has been established in other mammals. Therefore, we aim to characterize polymorphisms in two genes in the pathway, *POMC* and *MC4R*, and look for associations between genotype and phenotype.

Materials and Methods

MC4R

DNA from twelve horses from various breeds including the Morgan, Mustang, Arabian, and Quarter Horse breeds was obtained from samples previously banked at the Brooks Equine Genetics Laboratory. DNA was extracted using one of two protocols, one previously published (Brooks & Bailey, 2005) and a second using a slightly modified (modifications from D. Cook, personal communication) protocol from the QIAGEN Puregene Blood Core Kit B (QIAGEN Sciences, Kansas City, MO). Three sets of PCR primers (shown in Table 1) encompassing the human and bovine coding regions of *MC4R* were designed based on the current assembly of the equine genome (EqCabv2.0) from publicly available databases using Primer3 v. 0.4.0 (Rozen and Skaletsky, 2000) or Vector NTI (v10.3, Invitrogen, Carlsbad, CA.) All PCR reactions were run with Roche Diagnostics Faststart Taq DNA Polymerase and included buffers (Branchburg, NJ) using the Eppendorf Gradient MasterCycler (Hamburg, Germany). Optimization reactions were run first for each set of primers to determine the optimum annealing temperature and obtain a single product. Following optimization, a PCR was conducted in a total volume of 20 µl; half of each reaction was run on a gel to visualize the amount of DNA generated. The remaining half was used to prepare samples for sequencing as described on the Cornell University Life

Table I: *MC4R* Primers

MC4R-1	Forward: 5'-ACT GAA CTG CCT GCA TGG AAT ATC-3' Reverse: 5'-TTC ACA TTA AGA GGA TCG CTG TCC-3'
MC4R-3	Forward: 5'-CTC CCA GAG GGT AGC AAC AG-3' Reverse: 5'-CAC CAT GTT CTT CAC CAT GC-3'
MC4R-2	Forward: 5'-AAG ACT CCA ATC AAT ATG GTC AAG GTG-3' Reverse: 5'-AGC TTC ACA GCA TGT GAC ATG TG-3'

Sciences Core Laboratories Center website (DNA Sequencing and Genotyping: Frequently Asked Questions, 2009), and sequencing was performed at that facility. ContigExpress from the Vector NTI suite and Codoncode Aligner (v. 3.5.1 Codoncode Corporation, Dedham, MA) were used to align newly generated sequence data as well as sequence for *MC4R* from EquCabv2.0. Vector NTI and Codoncode Aligner were also used to translate coding regions.

To genotype the *MC4R* SNP, a melting curve assay was designed following the concept laid out in a previous paper (Nieto & Aleman, 2009). Figure 2 illustrates the melting curve assay strategy. Two primers were designed: 5'-GAA AAG GCT ACT CTG ATG (forward) and 5'-CCA CAA TCA CCA GAA TAT (reverse). Two probes were also designed: a sensor (5'-GGT GCT ATG AGC

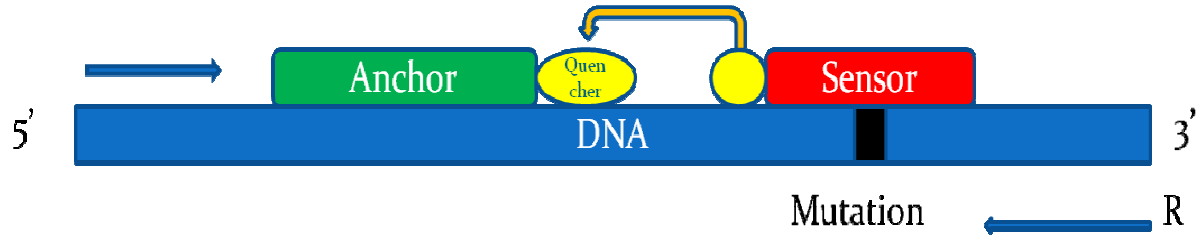


Figure 2: Melting curve Assay Concept.
PCR generated a 108-bp product. 1 µM anchor and 0.5 µM sensor probes were then added. After initial denaturation step at 95°C and annealing step at 40°C, the temperature was raised incrementally to 70°C. The sensor probe either had perfect complementarity to the DNA template, or had one mismatch at the SNP site. The mismatched sensor melts earlier than then perfectly matched sensor. Once melted, the fluorophore's fluorescence is no longer quenched by the nearby quencher on the anchor probe. This will be detected as an increase in overall fluorescence.

AAC TTT TTG TC) probe, labeled at its 3' end with a fluorophore, and an anchor probe (5'-GTT TGT GAC TCT GGG TGT CAT CAG), labeled at its 5' end with a quencher. Probes anneal to the template DNA so there is 11-bp separating them. PCR in a total reaction volume of 10 µl generated a 108-bp product. After completion of PCR, the probes were added at the following concentrations: 1 µM anchor and 0.05 µM sensor (20:1 anchor: sensor ratio). A melting curve program was run on the Eppendorf Mastercycler: 2 minutes at 95°C, 2 minutes at 40°C, then 15

minutes from 40°C ramping up to 70°C. The first derivative of the melting curve data was graphed in Microsoft Excel for analysis. Unrelated Arabian horses of various colors and body condition scores were chosen for genotyping. Body condition scores were assigned by a single observer (the author) based on previous definitions of each category (Henneke, Potter, Kreider, & Yeates, 1983). A BCS of 5 is considered ideal. Below 5, the animal is considered underweight, and above five, the animal is considered overweight.

POMC

The latest assembly, EquCabv2.0, (September 2007) was used for primer design. The UCSC Genome Browser (Kent, Sugnet, and Furey, 2002) was used to identify DNA for re-sequencing. Using alignment data from mRNAs from other species, three major regions of interest were identified. These included two exons, as well as a regulatory region upstream of the two exons termed the 5'UTR (untranslated region toward the 5' end of the gene). Primers were designed for each of these regions (Table II). A schematic of the gene, with relative primer locations, is shown in Figures 3.

Table II: *POMC* Primers

POMC 5'UF: chr15: 70782628 POMC5'UR: chr15: 70782805	Forward: 5'-GCC CGG CAG ATA TAT AAG GA-3' Reverse: 5'-AGG GCT CCC AGG GTA AGA G-3'
POMC1bF: chr15: 70785855 POMC1bR: chr15: 70786146	Forward: 5'-AAA AAG CCT TGA GCT CCA CA-3' Reverse: 5'-TCT GAT GCC CAA ACA AGA TG-3'
POMC2bF: chr15: 70787694 POMC2bR: chr15: 70787860	Forward: 5'-TAG AGA GGG CAC GGG GAA C-3' Reverse: 5'-GAC GTA CTT CCG GGG GTT CT-3'
POMC3b: chr15: 70788639	Reverse: 5'-TCA CCC AGT CAC CAT TCA AA-3'
POMCU1F: chr15: 70788033 POMCU1R: chr15: 70788486	Forward: 5'-GCA AGC GCT CCT ACT CCA T-3' Reverse: 5'-GAC TTC CGG GGA GAG CAG-3'
POMCU2F: chr15: 70787873 POMCU2R: chr15: 70788033	Forward: 5'-AGT ACG TCA TGG GCC ACT TC-3' Reverse: 5'-CAT GGA GTA GGA GCG CTT G-3'
POMCU3F: chr15: 70787433	Forward: 5'-TGG CCA TTT AAT TCC CAT GT-3'

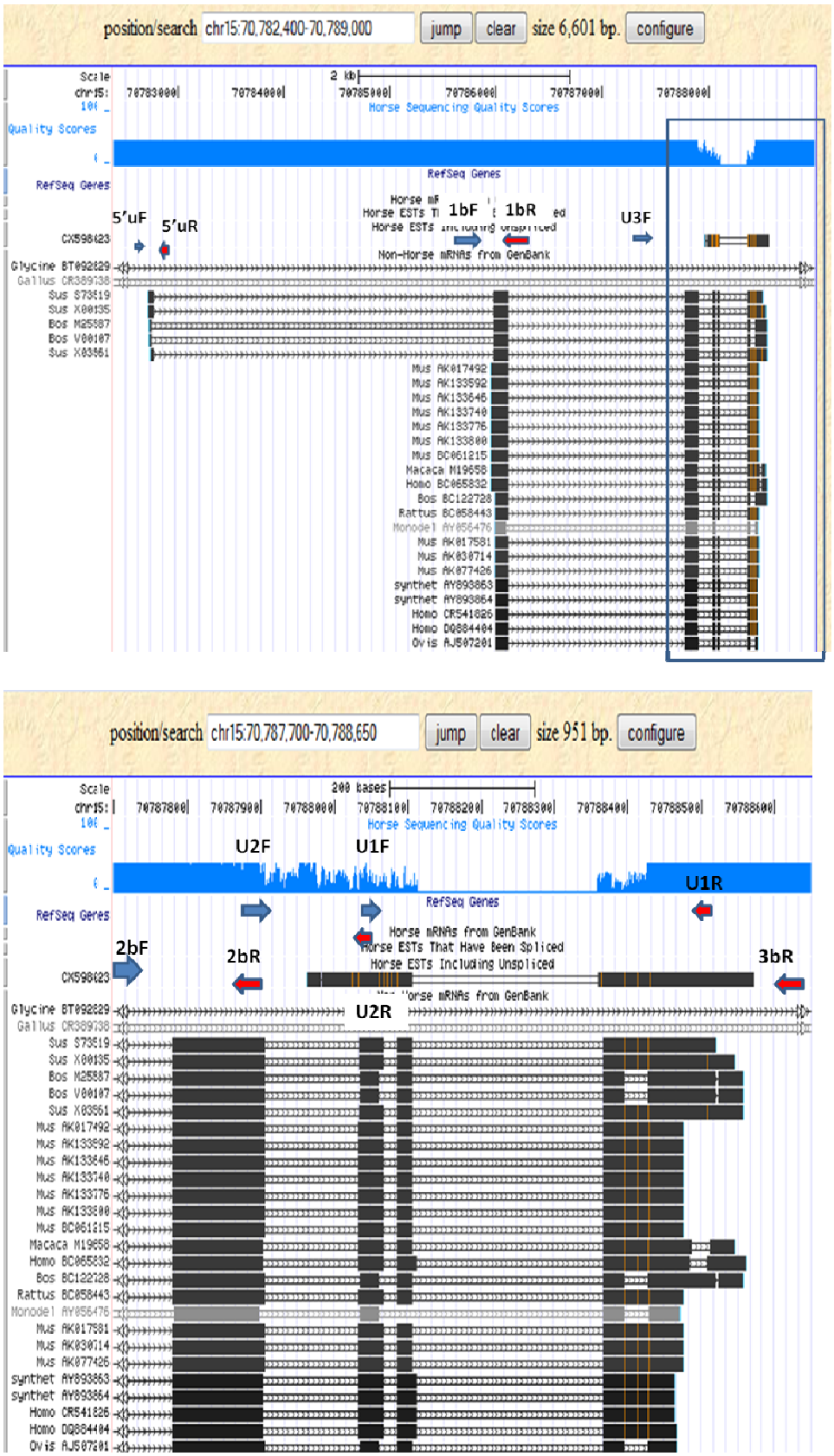


Figure 3: Relative Primer Locations in *POMC*.
The top screenshot from the UCSC Genome Browser (*http://genome.ucsc.edu)* shows the entire gene; the bottom screenshot shows a zoomed in region (denoted by boxed region in the top image). Note the area of low quality genomic sequence. Forward primers are denoted in blue; reverse in red.

The downstream region was complicated by two factors: high G/C content and lack of high quality sequence data in the assembly. The latter problem was resolved using published EST sequence data (horse EST CX598023). This data was used to generate additional primers (see Table II). POMCU1 primers were designed to flank the gap in the genome using EST sequence on either side. The forward primer of POMCU2 was designed from genomic sequence; the reverse was designed using the upstream portion of the EST.

The high G/C content of this area of the genome (about 75%) was the other complicating factor in getting specific amplification from the PCR reaction. This problem was resolved using the U1 primer set and a higher annealing temperature (62.5°C). DNA from nine horses was sequenced using this primer set. This still left a large amount of coding region to sequence, so two different combinations of the above primers were used: 2bF/U1R and 1bF/U1R. Optimization was redone in a 20 µl reaction, using 7.5, 8.0, or 8.5 µl G/C solution (Roche, Branchburg, NJ) per reaction over a temperature range from 50 to 60°C. Of the 8.5 µl of G/C solution reactions, the one performed at 51.9°C was prepared for sequencing. Five more horses were selected to sequence, and of those, only 3 amplified and generated usable data. A final primer was designed further upstream of the high G/C region, POMCU3 (see Table II). This forward primer was paired with the POMCU2 reverse primer, and optimized using the G/C solution, extra Taq, and extra initial denature time. Success from this led to the re-sequencing of six horses using the U3 forward primer, and completion of the re-sequencing of the majority of the transcribed regions of the gene.

Results

MC4R

Re-sequencing of about 1733 base pairs of the coding region of the gene revealed a single, synonymous SNP (Figure 4). Initial genotyping data using the new melting curve method revealed 25 clear A/A homozygotes, 11 clear A/G heterozygotes, and three clear G/G homozygotes. Figure 5 shows sample curves for each of the three genotypes.

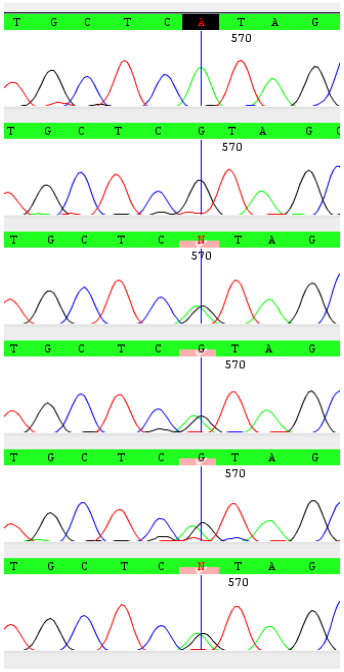


Figure 4: *MC4R* SNP Detection. The top horse is a A/A homozygote, and the next horse down is a G/G homozygote. The remaining four horses are A/G heterozygotes.

Using a G-dominance model (hypothetical dominance of the G allele over A), allele counts were calculated for dark- and light-shade horses (each color category included horses from chestnut and bay base colors). Light- and dark-shaded coat colors have been previously defined (Sponenberg, 1996). For example, dark-shade horses included the dark bay (almost black) and liver chestnut colors, while light-shade horses were bright bays and chestnuts. An examples of each color type is shown in Figure 6. Total numbers are shown in Figure 7. Using a Fisher's exact test with the JMP program (v8.0 SAS Institute Inc. Cary, NC), the one-tailed p-value for association of this SNP with light or dark shade equals 0.0358. There appears to be a trend between genotype at this SNP and BCS (JMP v.8.0, Figure 8); however, the small sample size (n=38) and the use of categorical assignments for weight instead of actual weight measurements may have prevented statistical significance from being obtained. Power analysis indicated a sample size of 89 horses would be sufficient using the current phenotype categories. To get accurate weight measurements for phenotype new samples would need to be recruited.

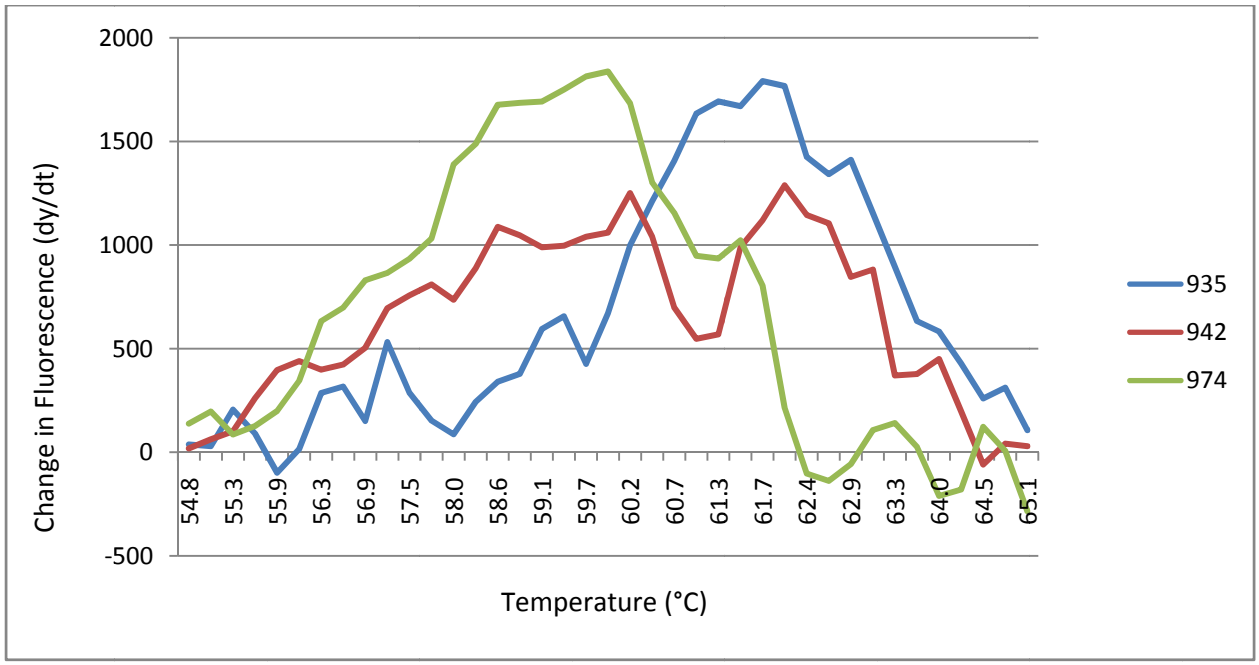


Figure 5: Sample First Derivative Curves for the 3 Genotypes
The first derivative of the fluorescence vs. temperature, as measured in the real-time PCR assay, for the three genotypes is shown. Blue shows A/A, red shows A/G, green shows G/G.

Figure 6: Light and Dark Shades
Clockwise from top left: dark bay, light bay, light chestnut, dark (liver) chestnut. Images adapted from http://www.whitehorseproductions.com/.

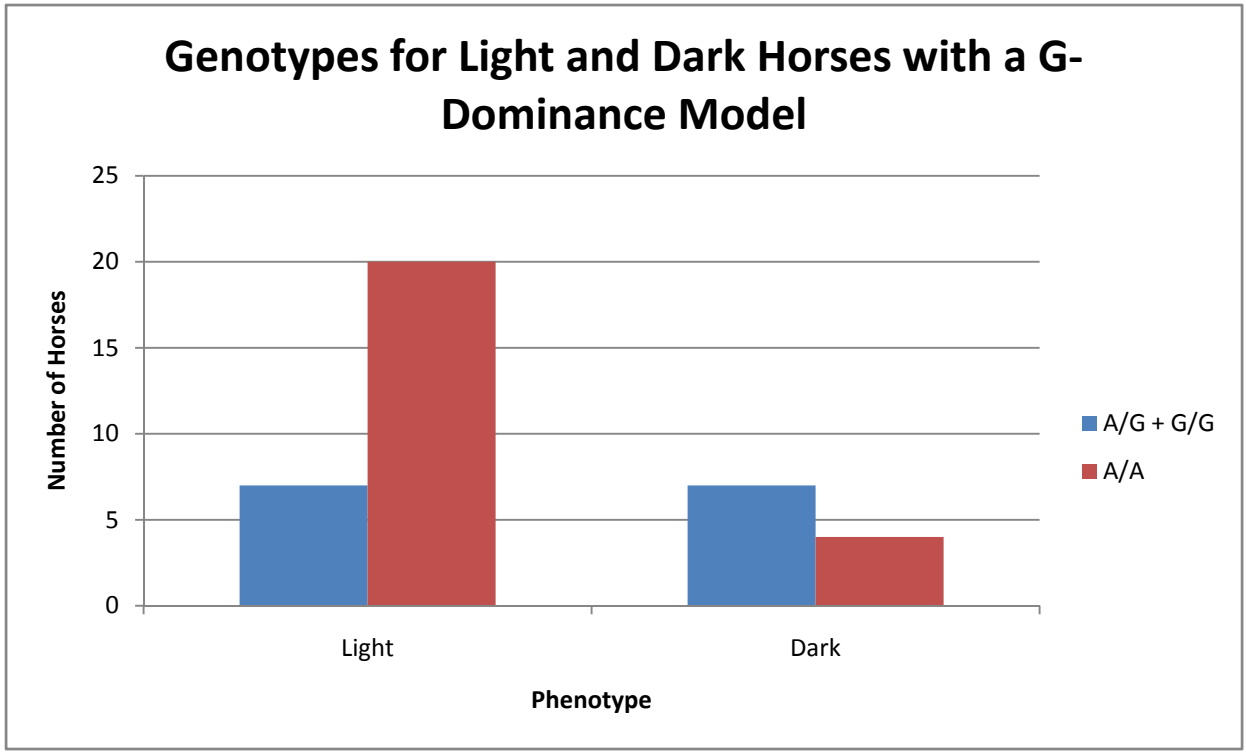


Figure 7: G-dominance Model.
Distribution of shade (light vs. dark) among these genotypic categories show a statistically significant association (p=0.0358)

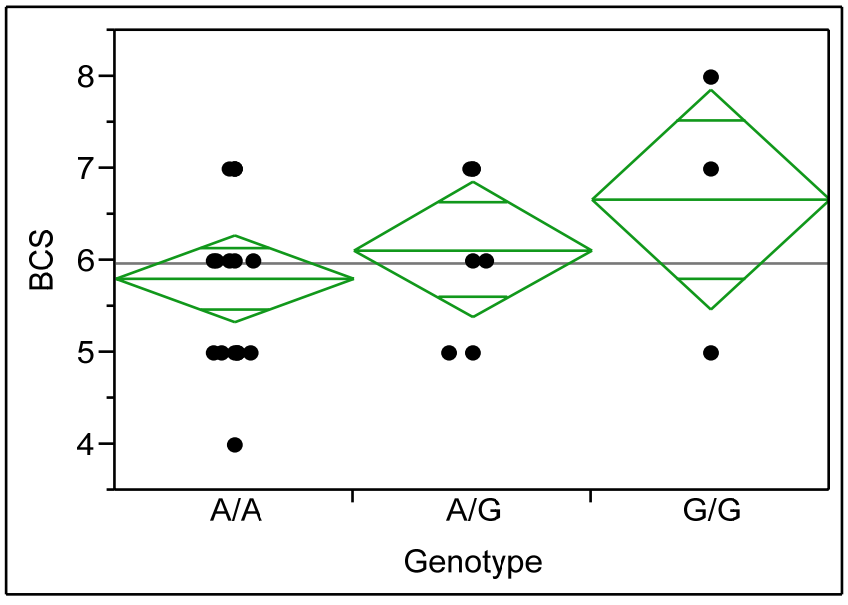


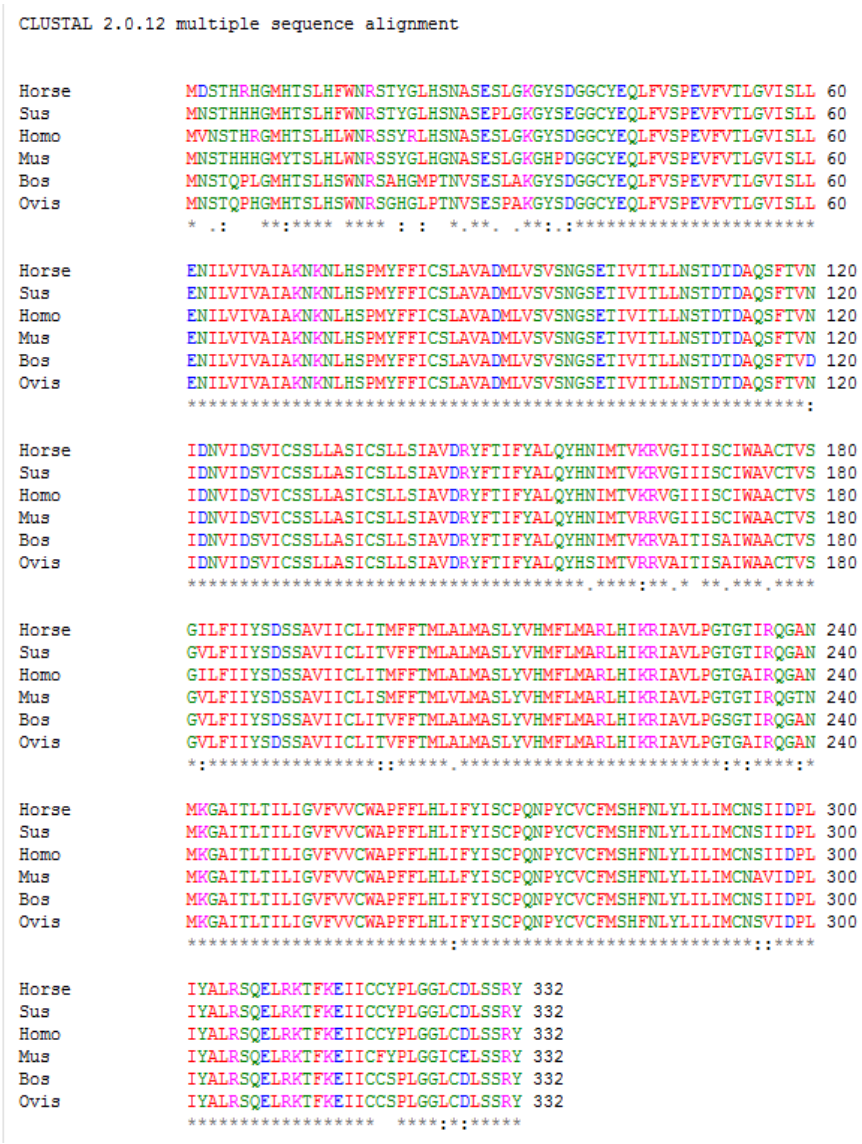
Figure 8: Trends in BCS Relative to Genotype.
Each dot represents an individual horse. The middle line of each diamond represents the mean of each genotype. The top and bottom peaks of the diamonds represent a 95% confidence interval for each genotype.

A model of the protein structure of the seven transmembrane domains of the G-protein coupled receptor was generated by Tusnády and Simon, 2001, and is shown in Figure 9.

```
seq  MDSTHRHGMH TSLHFWNRST YGLHSNASES LGKGYSDGGC YEQLFVSPEV    50
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOoooooo ooooooooooH

seq  FVTLGVISLL ENILVIVAIA KNKNLHSPMY FFICSLAVAD MLVSVSNGSE   100
pred HHHHHHHHHH HHHHHHHHHi iiiiiiiHHH HHHHHHHHHH HHHHHHoooo

seq  TIVITLLNST DTDAQSFTVN IDNVIDSVIC SSLLASICSL LSIAVDRYFT   150
pred oooooooooo oooooooooo ooooooHHHH HHHHHHHHHH HHHHHiiiii

seq  IFYALQYHNI MTVKRVGIII SCIWAACTVS GILFIIYSDS SAVIICLITM   200
pred iiiiiiiiii iiiiiHHHHH HHHHHHHHHH HHHHHHoooo ooHHHHHHHH

seq  FFTMLALMAS LYVHMFLMAR LHIKRIAVLP GTGTIRQGAN MKGAITLTIL   250
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiiiiiii iiHHHHHHHH

seq  IGVFVVCWAP FFLHLIFYIS CPQNPYCVCF MSHFNLYLIL IMCNSIIDPL   300
pred HHHHHHHHHH HHHHHHHooo oooooooooo oooooHHHHH HHHHHHHHHH

seq  IYALRSQELR KTFKEIICCY PLGGLCDLSS RY  332
pred HHHHiiiiii iiiiiiiiiI IIIIIIIIII II
```

Figure 9: Prediction of Tertiary Structure of *MC4R*
A representation of the receptor protein. "O" denotes the extracellular portion, "H" represents the helical transmembrane portion, and "I" represents the intracellular portion. From Tusnády and Simon, 2001.

An alignment using the ClustalW2 server (http://www.ebi.ac.uk/Tools/clustalw2/index.html, (Larkin, et al., 2007)) was generated between the predicted amino acid sequence of horse MC4R and 5 other mammalian species (Figure 10). The protein sequences for the other species is available from the NCBI database, with the following accession numbers: Ovis aries (NP_001119842); Bos taurus (NP_776535); Mus musculus (NP_058673); Sus scrofa (ABD28176); and Homo sapiens (NP_005903).



```
CLUSTAL 2.0.12 multiple sequence alignment


Horse           MDSTHRHGMHTSLHFWNRSTYGLHSNASESLGKGYSDGGCYEQLFVSPEVFVTLGVISLL 60
Sus             MNSTHHHGMHTSLHFWNRSTYGLHSNASEPLGKGYSEGGCYEQLFVSPEVFVTLGVISLL 60
Homo            MVNSTHRGMHTSLHLWNRSSYRLHSNASESLGKGYSDGGCYEQLFVSPEVFVTLGVISLL 60
Mus             MNSTHHHGMYTSLHLWNRSSYGLHGNASESLGKGHPDGGCYEQLFVSPEVFVTLGVISLL 60
Bos             MNSTQPLGMHTSLHSWNRSAHGMPTNVSESLAKGYSDGGCYEQLFVSPEVFVTLGVISLL 60
Ovis            MNSTQPHGMHTSLHSWNRSGHGLPTNVSESPAKGYSDGGCYEQLFVSPEVFVTLGVISLL 60
                * .:   **:**** **** : :  *.**. .**:.:********************** 

Horse           ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN 120
Sus             ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN 120
Homo            ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN 120
Mus             ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN 120
Bos             ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVD 120
Ovis            ENILVIVAIAKNKNLHSPMYFFICSLAVADMLVSVSNGSETIVITLLNSTDTDAQSFTVN 120
                ************************************************************: 

Horse           IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHNIMTVKRVGIIISCIWAACTVS 180
Sus             IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHNIMTVKRVGIIISCIWAVCTVS 180
Homo            IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHNIMTVKRVGIIISCIWAACTVS 180
Mus             IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHNIMTVRRVGIIISCIWAACTVS 180
Bos             IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHNIMTVKRVAITISAIWAACTVS 180
Ovis            IDNVIDSVICSSLLASICSLLSIAVDRYFTIFYALQYHSIMTVRRVAITISAIWAACTVS 180
                ************************************.****:**.* **.***.**** 

Horse           GILFIIYSDSSAVIICLITMFFTMLALMASLYVHMFLMARLHIKRIAVLPGTGTIRQGAN 240
Sus             GVLFIIYSDSSAVIICLITVFFTMLALMASLYVHMFLMARLHIKRIAVLPGTGTIRQGAN 240
Homo            GILFIIYSDSSAVIICLITMFFTMLALMASLYVHMFLMARLHIKRIAVLPGTGAIRQGAN 240
Mus             GVLFIIYSDSSAVIICLISMFFTMLVLMASLYVHMFLMARLHIKRIAVLPGTGTIRQGTN 240
Bos             GVLFIIYSDSSAVIICLITVFFTMLALMASLYVHMFLMARLHIKRIAVLPGSGTIRQGAN 240
Ovis            GVLFIIYSDSSAVIICLITVFFTMLALMASLYVHMFLMARLHIKRIAVLPGTGAIRQGAN 240
                *:****************::*****.************************:*:****:* 

Horse           MKGAITLTILIGVFVVCWAPFFLHLIFYISCPQNPYCVCFMSHFNLYLILIMCNSIIDPL 300
Sus             MKGAITLTILIGVFVVCWAPFFLHLIFYISCPQNPYCVCFMSHFNLYLILIMCNSIIDPL 300
Homo            MKGAITLTILIGVFVVCWAPFFLHLIFYISCPQNPYCVCFMSHFNLYLILIMCNSIIDPL 300
Mus             MKGAITLTILIGVFVVCWAPFFLHLLFYISCPQNPYCVCFMSHFNLYLILIMCNAVIDPL 300
Bos             MKGAITLTILIGVFVVCWAPFFLHLIFYISCPQNPYCVCFMSHFNLYLILIMCNSIIDPL 300
Ovis            MKGAITLTILIGVFVVCWAPFFLHLIFYISCPQNPYCVCFMSHFNLYLILIMCNSVIDPL 300
                *************************:****************************::**** 

Horse           IYALRSQELRKTFKEIICCYPLGGLCDLSSRY 332
Sus             IYALRSQELRKTFKEIICCYPLGGLCDLSSRY 332
Homo            IYALRSQELRKTFKEIICCYPLGGLCDLSSRY 332
Mus             IYALRSQELRKTFKEIICFYPLGGICELSSRY 332
Bos             IYALRSQELRKTFKEIICCSPLGGLCDLSSRY 332
Ovis            IYALRSQELRKTFKEIICCSPLGGLCDLSSRY 332
                ******************  ****:*:***** 
```

Figure 10: ClustalW2 Alignment of Horse MC4R protein with five other mammals. Stars indicate complete conservation, and colons/dots indicate partial conservation.

POMC

Re-sequencing of POMC yielded four SNPs, three in exons (synonymous) and one in the proposed 5'UTR (Figures 11 and 12).

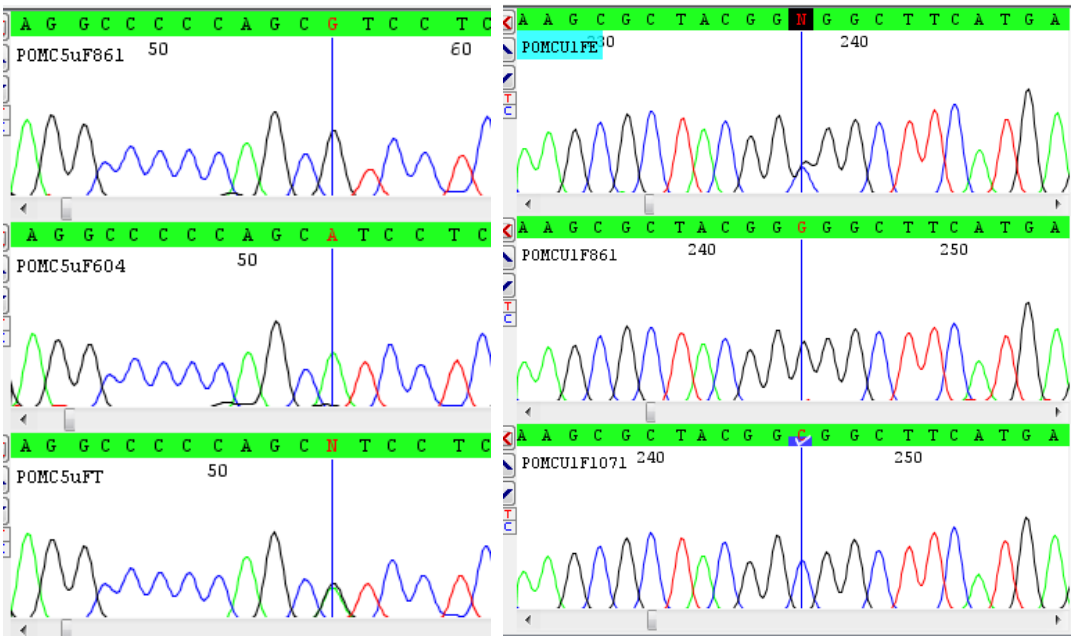


Figure 11: **Electropherogram alignments of SNP sequences in POMC**
SNPs at the 5'UTR (left) and gap-1 (right)

Table III: *POMC* Haplotypes

Haplotype	5'UTR	1b	Gap-1	Gap-2
A	A	G	C	G
A/B	A/G	G	C/G	G
B	G	G	G	G
C	A/G	C/G	G/C	T/G

Four distinct haplotypes were predicted (Table III). Once the complete sequence was generated, a predicted transcript could be constructed. The 1b region includes an ATG initiation codon and an intron/exon boundary, which was predicted using the Netgene2 server ((http://www.cbs.dtu.dk/services/NetGene2/, (Hebsgaard, Korning, Tolstrup, Engelbrecht, Rouze, & Brunak, 1996)). The gap region also includes an intron/exon boundary ((http://www.cbs.dtu.dk/services/NetGene2/, (Hebsgaard, Korning, Tolstrup, Engelbrecht, Rouze, & Brunak, 1996)), and proceeds through an open reading frame to a terminal stop

codon. Finally, the 5'UTR region has a exon/intron boundary

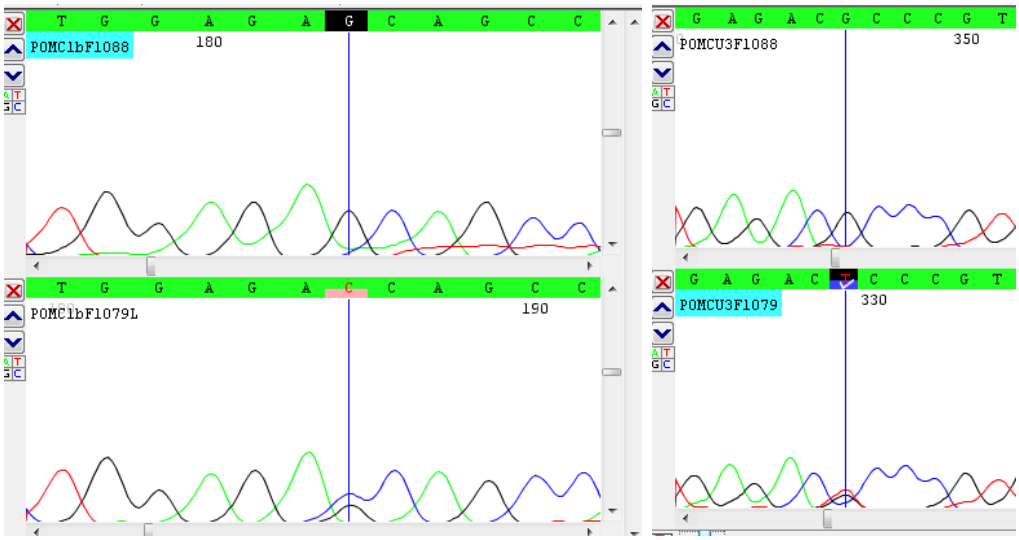


Figure 12: Electropherogram Alignments of Dam and Foal with Haplotype C in POMC
A single horse heterozygous at two loci; 1b on the left and gap-2 on the right. In contrast the dam of this horse is shown to be homozygous at both loci (top row).

((http://www.cbs.dtu.dk/services/NetGene2, (Hebsgaard, Korning, Tolstrup, Engelbrecht, Rouze, & Brunak, 1996)), which supports that it is expressed. Using this information, the resulting regions were translated using ExPASy Proteomics server ((http://ca.expasy.org/tools/dna.html, (Gasteiger, Gattiker, Hoogland, Ivany, Appel, & Bairoch, 2003)) and a 254 amino acid protein was generated. Protein sequences were aligned using the ClustalW2 server (http://www.ebi.ac.uk/Tools/clustalw2/index.html, (Larkin, et al., 2007) with published protein sequences for other mammals (Figure 13). The protein sequences for the other species are available from the NCBI database, with the following accession numbers: Homo sapiens (NP_000930); Sus scrofa (CAA24968); Bos Taurus (NP_776576); Ovis aries (NP_001009266); and Mus musculus (NP_032921).

The complete model for *POMC* is as follows. Two exons plus a 5'UTR make up the full mRNA, which then gets translated into a 254 amino acid protein.

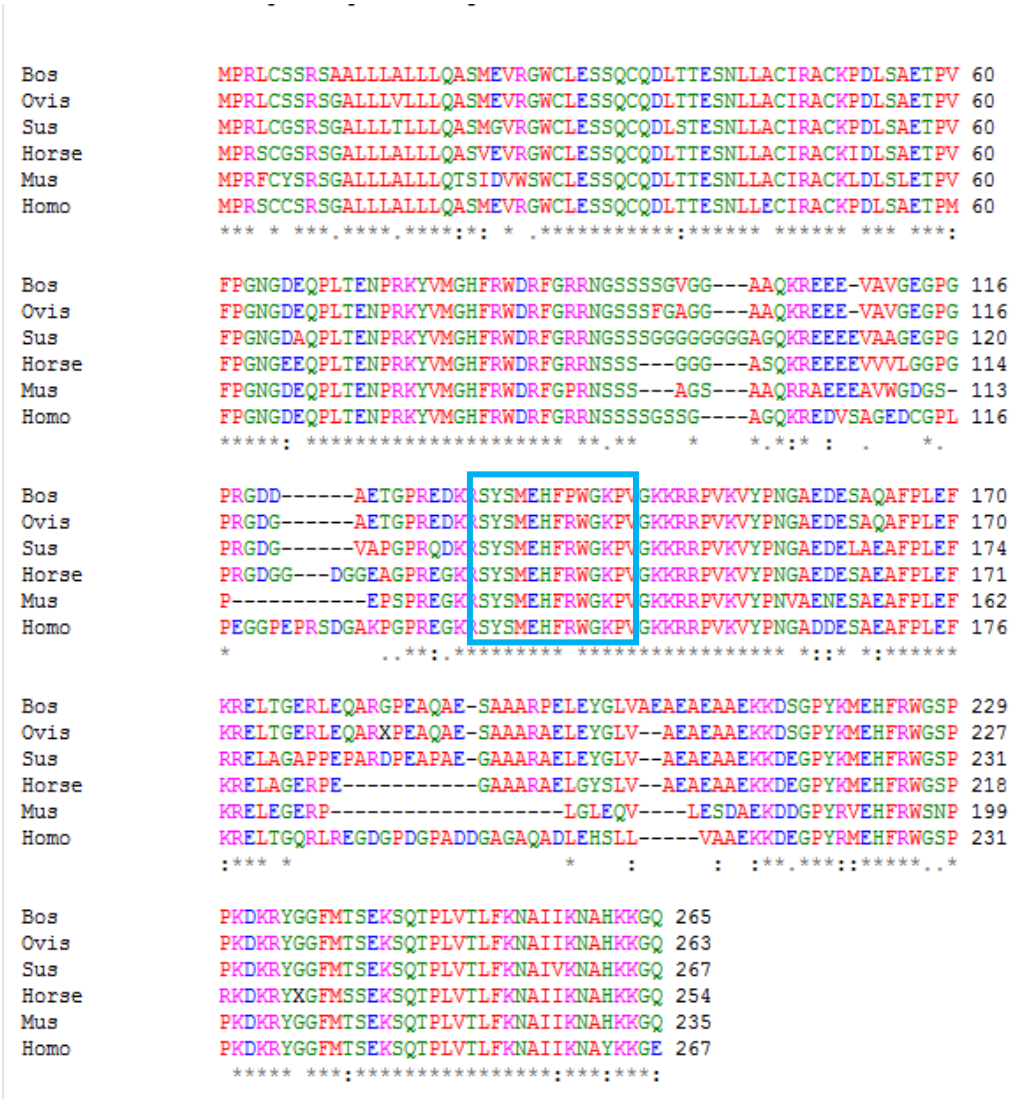


Figure 13: ClustalW2 Alignment of Horse POMC protein with five other mammals.
The blue box shows the amino acid sequence of α-MSH.
Stars indicate complete conservation, and colons/dots indicate partial conservation.

Discussion

Laminitis is a major concern to anyone involved in the horse industry. Identification of individual horses at risk for gaining excess weight should be a priority, so that preventative management can be started early before weight gain has begun and metabolic syndrome sets in. *MC4R* and *POMC* variants in other mammals show an increased risk of gaining excess weight (Huszar, et al., 1997) (Kim, Larsen, Short, Plastow, & Rothschild, 2000) (Krude, Biebermann, Luck, Horn, Brabant, & Gruters, 1998) (Santoro & Cirillo, 2009) (Xuemei, Ning, Xuemei, Xingbo, Qingyong, & Xiuli, 2006). These genes have not been well-characterized in the horse. This work has generated additional DNA sequences of the coding regions and some regulatory regions of these two genes. Five novel SNPS between the two genes have been identified, and using homology-based methods, gene models have been constructed. As can be seen in Figures 10 and 13 in the results section, among mammals, the protein sequence for both α-MSH and MC4R is conserved. The few deviations in MC4R that are observed are mainly in the predicted extracellular or intracellular tails of the protein (see Figure 9), which may have important species-specific signaling consequences. This pathway must have evolved early on in mammals, without much deviation, and is functionally important.

The pathways regulating feed intake and energy expenditure are central to controlling the weight of mammalian species (Barsh & Schwartz, 2002) (O'Rahilly, Yeo, & Sadaf Farooqi, 2004). While the SNPs identified do not themselves change amino acids in either of the two gene products observed, it is possible that they can serve as markers for nearby regulatory mutations that could be controlling phenotype. The significant association between the *MC4R* SNP and coat color merits additional study. It is likely in linkage disequilibrium with another nearby mutation which is affecting the phenotype. Association between BCS and the *MC4R*

SNP was not significant in this data set, due to the limitations outlined in the results section. Additional horses should be genotyped to determine if there is a significant association. It would not be surprising if there was an association. Coat color and adiposity are known to be related traits, and an association between the SNP and coat color has already been identified. The 5'UTR SNP identified in *POMC* may have a regulatory function and should be investigated further. Changes in the level of *POMC* mRNA would have a profound effect on the pathway it is involved in. Underexpression of POMC would lessen the stimulatory effect on the *MC4R*-expressing neurons (Barsh & Schwartz, 2002). The inhibitory effect of the *AgRP*-producing neurons would be felt more strongly, and this would lead to increased feeding behavior (Barsh & Schwartz, 2002). Any defect in the *MC4*-receptor itself would change its affinity for its ligand. Observations in other animals has shown that other loss of function mutations for MC4R (Huszar, et al., 1997), (Kim, Larsen, Short, Plastow, & Rothschild, 2000), which alter ligand affinity, lead to changes in phenotype. While coding sequence changes were not observed in this work, there was a significant association observed between the SNP and coat color and a trend observed between the SNP and BCS. The SNP may be in linkage disequilibrium with a regulatory mutation in nearby intronic sequence, and this should be investigated further to determine if expression of the *MC4R* gene is affected. There would then be a loss of feedback from the current energy state of the animal from leptin and insulin signaling pathways and control of feed intake (Barsh & Schwartz, 2002). In other words, even if the animal was satiated, the feedback loop would have been lost and the animal would continue to eat even though its caloric needs had been met. In turn, this would lead to an increase in weight gain and the development of an obese phenotype.

Further work needs to be done to better understand what correlation these SNP markers have with phenotype. In particular, work should focus on RNA expression in the hypothalamus of horses with the haplotypes identified in Table IV to determine if the 5'UTR *POMC* SNP has a regulatory function. Association of the MC4R SNP with phenotype should be looked at more closely for confirmation and to determine what nearby changes could have a causative effect.

Works Cited

Asplin, K. E., Sillence, M. N., Pollit, C. C., & McGowan, C. M. (2007). Induction of laminitis by prolonged hyperinsulinaemia in clinically normal ponies. *The Veterinary Journal , 174*, 530-535.

Barsh, G. S., & Schwartz, M. W. (2002). Genetic Approaches to Studying Energy Balance: Perception and Integration. *Nature Reviews: Genetics , 3*, 589-600.

Baskin, D. G. (2006). Single-Minded View of Melanocortin Signaling in Energy Homeostasis. *Endocrinology , 147* (10), 4539–4541.

Brooks, S., & Bailey, E. (2005). Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mammalian Genome , 16*, 893-902.

DNA Sequencing and Genotyping: Frequently Asked Questions. (2009). Retrieved from DNA Sequencing and Genotyping: http://cores.lifesciences.cornell.edu/brcinfo/?p=faq&f=1

Galey, F., Whiteley, H., Goetz, T., AR, K., CA, D., & VR, B. (1991). Black walnut (Juglans nigra) toxicosis: a model for equine laminitis. *Journal of Comparative Pathology , 104* (3), 313-326.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivany, I., Appel, R., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research , 31*, 3784-3788.

Geor, R. J. (2008). Metabolic Predispositions to Laminitis in Horses and Ponies: Obesity, Insulin Resistance and Metabolic Syndromes. *28* (12), 753-759.

Geor, R. J. (2009). Pasture-Associated Laminitis. *Veterinary Clinics of North America: Equine Practice , 25* (1), 39-50.

Hebsgaard, S., Korning, P., Tolstrup, N., Engelbrecht, J., Rouze, P., & Brunak, S. (1996). Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. *Nucleic Acids Research , 24* (17), 3439-3452.

Henneke, D., Potter, G., Kreider, J., & Yeates, B. (1983). Relationship between condition score, physical measurements and body fat percentage in mares. *Equine Veterinary Journal , 15* (4), 371-372.

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Qing Fang, Q., Berkemeier, L. R., et al. (1997). Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. *1997 , 88*, 131-141.

Kim, K., Larsen, N., Short, T., Plastow, G., & Rothschild, M. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake ratios. *Mammalian Genome , 11*, 131-135.

Krude, H., Biebermann, H., Luck, W., Horn, R., Brabant, G., & Gruters, A. (1998). Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genetics , 19*, 155-157.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., et al. (2007). ClustalW and ClustalX version 2. *Bioinformatics , 23* (21), 2947-2948.

Marklund, L., Moller, M. J., Sandberg, K., & Anderson, L. (1996). A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with chestnut coat color in horses. *Mammalian Genome , 7*, 895-899.

Nieto, J., & Aleman, M. (2009). A Rapid Detection Method for the Ryanodine Receptor 1 (C7360G) Mutation in Quarter Horses. *Journal of Veterinary Internal Medicine , 23* (3), 619 - 622.

Nourian, A., GI, B., van Eps, A., & Pollitt, C. (2007). Equine laminitis: ultrastructural lesions detected 24-30 hours after induction with oligofructose. *Equine Veterinary Journal , 39* (4), 360-364.

O'Rahilly, S., Yeo, G. S., & Sadaf Farooqi, S. I. (2004). Melanocortin receptors weigh in. *Nature Medicine , 10* (4), 351-352.

Santoro, N., & Cirillo, G. e. (2009). Prevalance of pathogenetic MC4R mutations in Italian children with early Onset obesity, tall stature and familial history of obesity. *BMC Medical Genetics , 10* (25).

Sponenberg, D. P. (1996). Equine Color Genetics. In D. P. Sponenberg, *Equine Color Genetics* (p. 18). Ames, Iowa: Iowa State University Press.

Walsh, D. M., McGowan, C. M., McGowan, T. W., Lamb, S. V., J, S. B., & Place, N. J. (2007). Equine Cushing's Disease/Equine Metabolic Syndrome: A Practitioner Field Study. *Fourth International Equine Conference on Laminitis and Diseases of the Foot.*

Xuemei, Q., Ning, L., Xuemei, D., Xingbo, Z., Qingyong, M., & Xiuli, W. (2006). The single nucleotide polymorphisms of chicken melanocortin-4 receptor (MC4R) and their association analysis with carcass traits. *Science in China Series C: Life Sciences , 49* (6), 560-566.