

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
21 August 2008 (21.08.2008)

PCT

(10) International Publication Number  
**WO 2008/100313 A1**

(51) International Patent Classification:  
C12Q 1/68 (2006.01) G01N 33/68 (2006.01)

(74) Agents: VIKSNINS, Ann S. et al.; P.O. Box 111098, Saint Paul, Minnesota 55111-1098 (US).

(21) International Application Number:  
PCT/US2007/062134

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:  
14 February 2007 (14.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **RE-  
GENTS OF THE UNIVERSITY OF MINNESOTA**  
[US/US]; 450 McNamara Alumni Center, 200 Oak Street  
S.E., Minneapolis, Minnesota 55455-2070 (US).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants and  
(72) Inventors: **VALBERG, Stephanie J.** [CA/US]; 100 Woodridge Lane, Lino Lakes, Minnesota 55014 (US). **MCCUE, Molly E.** [US/US]; 605 Harriet Avenue, #602, Shoreview, Minnesota 55126 (US). **MICKELSON, James R.** [US/US]; 1891 Merrill Street, Roseville, Minnesota 55108 (US).

**Published:**  
— with international search report  
— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau



**WO 2008/100313 A1**

(54) Title: METHOD OF DETECTING EQUINE POLYSACCHARIDE STORAGE MYOPATHY

**Figure 3.**

PSSM Horses  
(SEQ ID NO:4)

1	11	21	31
IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFVHGHFYG

All other species  
(SEQ ID NO:5)

IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFVGRGHFYG
------------	------------	------------	------------

(57) Abstract: The present invention relates to diagnosing Polysaccharide Storage Myopathy (PSSM) disease in equines.

## METHOD OF DETECTING EQUINE POLYSACCHARIDE STORAGE MYOPATHY

### Background of the Invention

5 Polysaccharide Storage Myopathy (PSSM) is a debilitating muscle disease in many and diverse breeds of horses. Previous data indicates that approximately 10% of Quarter Horses and 36% of Belgian draft horses are affected. Clinical signs vary, but can range from muscle atrophy and progressive weakness in Draft horse breeds, to acute post-exercise muscle cramping and cell damage in Quarter Horses and other  
10 breeds. All forms of PSSM in horses are highly associated with deposits of an abnormal polysaccharide in skeletal muscle fibers that are demonstrated by histochemical staining of muscle biopsies. PSSM is also characterized by as much as four times the normal level of glycogen in skeletal muscle. Mutations in genes of glucose and glycogen metabolism are known to cause various types of glycogen  
15 storage diseases (glycogenoses) in humans and animal species, of which several histologically resemble PSSM. However, none of these genes appear to be responsible for equine PSSM.

The current diagnosis of PSSM in horses is based on clinical signs of muscle cramping or progressive atrophy (depending on the breed), often with elevated  
20 serum levels of muscle enzymes, combined with the histopathology finding of abnormal polysaccharide in thin sections cut from skeletal muscle biopsies.

Muscle biopsies are invasive, require skilled veterinary personnel to collect, are relatively expensive for the owner, and take a skilled muscle histopathologist to interpret. Further, although the muscle biopsy analysis has been a highly reliable  
25 diagnostic tool, it is not now 100% specific or sensitive, and can never hope to be.

Therefore, despite the foregoing, there is a need in the art for additional diagnostic tests for diagnosing PSSM in horses.

### Summary of the Invention

30 The present invention provides a method for detecting the presence of a biomarker associated with equine Polysaccharide Storage Myopathy (PSSM). In

one embodiment of the invention, the method involves obtaining a physiological sample from a horse, wherein the sample comprises nucleic acid, and determining the presence of the biomarker. As used herein, the phrase "physiological sample" is meant to refer to a biological sample obtained from a mammal that contains nucleic acid. For example, a physiological sample can be a sample collected from an individual horse, such as including, but not limited to, *e.g.*, a cell sample, such as a blood cell, *e.g.*, a lymphocyte, a peripheral blood cell; a sample collected from the spinal cord; a tissue sample such as cardiac tissue or muscle tissue, *e.g.*, cardiac or skeletal muscle; an organ sample, *e.g.*, liver or skin; a hair sample, *e.g.*, a hair sample with roots; and/or a fluid sample, such as blood.

Examples of breeds of affected horse include, but are not limited to, Quarter Horses, Percheron Horses, Paint Horses, Draft Horses, Warmblood Horses, or other related or unrelated breeds. The phrase "related breed" is used herein to refer to breeds that are related to a breed, such as Quarter Horse, Draft Horse, or Warmblood Horse. Such breeds include, but are not limited to stock breeds such as the American Paint horse, the Appaloosa, and the Palomino. The term "Draft Horse" includes many breeds including but not limited to Clydesdale, Belgian, Percheron, and Shire horses. The term "Warmblood" is also a generic term that includes a number of different breeds. "Warmblood" simply distinguishes this type of horse from the "cold bloods" (draft horses) and the "hot bloods" (Thoroughbreds and Arabians). The method of the present invention also includes horses of crossed or mixed breeds.

The term "biomarker" is generally defined herein as a biological indicator, such as a particular molecular feature, that may affect or be related to diagnosing or predicting an individual's health. For example, in certain embodiments of the present invention, the biomarker comprises a mutant equine glycogen synthase enzyme 1 (*GSY1*) gene, such as a polymorphic allele of *GSY1* has a substitution of G to A at nucleotide 926 in exon 6. The *GSY1* gene encodes an enzyme having an R (arginine) to H (histidine) substitution at amino acid residue 309.

"Oligonucleotide probe" can refer to a nucleic acid segment, such as a primer, that is useful to amplify a sequence in the *GSY1* gene that is complementary

to, and hybridizes specifically to, a particular sequence in *GYSI*, or to a nucleic acid region that flanks *GYSI*.

As used herein, the term “nucleic acid” and “polynucleotide” refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues.

A “nucleic acid fragment” is a portion of a given nucleic acid molecule. Deoxyribonucleic acid (DNA) in the majority of organisms is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. The term “nucleotide sequence” refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers.

The terms “nucleic acid,” “nucleic acid molecule,” “nucleic acid fragment,” “nucleic acid sequence or segment,” or “polynucleotide” may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene, *e.g.*, genomic DNA, and even synthetic DNA sequences. The term also includes sequences that include any of the known base analogs of DNA and RNA.

In one embodiment of the present invention, the method also involves contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid. “Amplifying” utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or

ligase chain reaction, LCR), strand displacement amplification, nucleic acid sequence-based amplification, and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. Reagents and hardware for conducting PCR are commercially available. For example, in certain embodiments of the present invention, exon 6 of the equine glycogen synthase enzyme 1 gene (also referred to as *GSY1*), or a portion thereof, may be amplified by PCR. In another embodiment of the present invention, at least one oligonucleotide probe is immobilized on a solid surface.

The methods of the present invention can be used to detect the presence of a biomarker associated with equine Polysaccharide Storage Myopathy (PSSM) in a horse such as a foal, *e.g.*, a neonatal foal or an aborted foal, one of a breeding pair of horses, *e.g.*, the potential dam and/or sire, or any horse at any stage of life. The horse can be alive or dead.

Further provided by the present invention is a method for diagnosing Polysaccharide Storage Myopathy (PSSM) in a horse, the method involving obtaining a physiological sample from the horse, wherein the sample comprises nucleic acid; and detecting the presence of a biomarker in the sample, wherein the presence of the biomarker is indicative of the disease. One embodiment of the method further involves contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid. For example, in one embodiment, exon 6 of equine glycogen synthase enzyme 1 or a portion thereof is amplified, for example, by polymerase chain reaction, strand displacement amplification, ligase chain reaction, amplification methods based on the use of Q-beta replicase and/or nucleic acid sequence-based amplification. In one embodiment of the method, the biomarker contains an equine glycogen synthase enzyme 1 gene having a G to A substitution at nucleotide 926 in exon 6 of the equine glycogen synthase enzyme 1 gene, or a gene encoding a glycogen synthase enzyme having an R to H substitution at amino acid residue 309. The method can be used to detect PSSM in a horse.

Further provided by the present invention is a kit comprising a diagnostic test for detecting the presence of equine PSSM in a horse comprising packaging

material, containing, separately packaged, at least one oligonucleotide probe capable of forming a hybridized nucleic acid with *GYS1* and instructions means directing the use of the probe in accord with the methods of the invention.

5

### **Brief Description of Drawings**

**Figure 1.** Normal Equine *GYS1* Coding DNA Sequence (SEQ ID NO:1). Exon 6 is indicated in bold. The site of a G to A mutation site at nucleotide position 926 is underlined. This region of sequence is expanded below in Figure 2.

10 **Figure 2.** *GYS1* Exon 6 and Flanking DNA Sequence from Normal (SEQ ID NO:2) and PSSM Horses (SEQ ID NO:3). Exon 6 in these equine *GYS1* DNA sequences contains positions 33 – 150. At position 135 a G in the normal horse sequence is replaced by an A in the PSSM horse sequence. This changes the underlined three base codon from one coding for an arginine (CGT) to one coding for a histidine (CAT).

15 **Figure 3.** Glycogen Synthase Amino Sequences Encoded by Exon 6 of the *GYS1* Genes. Species included in the analysis are Human, Control Horse, Chimpanzee, Canine, Bovine, Mouse, Rat, Pig, and Zebrafish. All species have identical amino acid sequences in this region of the skeletal muscle glycogen synthase protein (SEQ ID NO:4), which represents the 39 amino acids encoded by  
20 nucleotide positions 33 – 150 in the DNA sequences of Figure 2. However, PSSM horses have a histidine (H) at amino acid position 34 in this exon (H) (SEQ ID NO:5), while all other species have an arginine (R). This codon represents number 309 in the complete coding sequence.

**Figure 4.** Horse *GYS1* Intron 5, Exon 6, and Intron 6 genomic DNA  
25 sequence from which PCR primers to amplify the PSSM *GYS1* mutation would be most appropriately derived (SEQ ID NO:6). Exon 6 is indicated in bold.

**Figure 5.** The entire *GYS1* coding nucleotide sequence in Figure 1 was translated to give this amino acid sequence (SEQ ID NO:9). The site of the R to H mutation at codon 309 is underlined.

30

### Detailed Description of the Invention

Horses affected with Polysaccharide Storage Myopathy (PSSM) are typically heterozygous for the affected gene.

An "allele" is a variant form of a particular gene. For example, the present invention relates, *inter alia*, to the discovery that some alleles of the *GYS1* gene cause PSSM in horses. A "GYS1 allele" refers to a normal allele of the *GYS1* locus as well as an allele carrying a variation(s) that predispose a horse to develop PSSM. The coexistence of multiple alleles at a locus is known as "genetic polymorphism." Any site at which multiple alleles exist as stable components of the population is by definition "polymorphic." An allele is defined as polymorphic if it is present at a frequency of at least 1% in the population. A "single nucleotide polymorphism (SNP)" is a DNA sequence variation that involves a change in a single nucleotide.

The terms "protein," "peptide" and "polypeptide" are used interchangeably herein.

The invention encompasses isolated or substantially purified nucleic acid compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule is a DNA molecule that, by human intervention, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule may exist in a purified form or may exist in a non-native environment. For example, an "isolated" or "purified" nucleic acid molecule, or portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Fragments and variants of the disclosed nucleotide sequences and proteins or

partial-length proteins encoded thereby are also encompassed by the present invention.

By "fragment" or "portion" of a sequence is meant a full length or less than full length of the nucleotide sequence encoding, or the amino acid sequence of a polypeptide or protein. As it relates to a nucleic acid molecule, sequence or  
5 segment of the invention when linked to other sequences for expression, "portion" or "fragment" means a sequence having, for example, at least 80 nucleotides, at least 150 nucleotides, or at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means, for example, at least 9, 12, 15, or at least 20,  
10 consecutive nucleotides, *e.g.*, probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the nucleic acid molecules of the invention.

Alternatively, fragments or portions of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments or portions of a nucleotide sequence may range from at  
15 least about 6 nucleotides, about 9, about 12 nucleotides, about 20 nucleotides, about 50 nucleotides, about 100 nucleotides or more.

A "variant" of a molecule is a sequence that is substantially similar to the sequence of the native molecule. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical  
20 amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed  
25 mutagenesis that encode the native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have in at least one embodiment 40%, 50%, 60%, to 70%, *e.g.*, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, *e.g.*, 81%-84%, at least 85%, *e.g.*, 86%, 87%, 88%, 89%, 90%, 91%, 92%,  
30 93%, 94%, 95%, 96%, 97%, to 98%, sequence identity to the native (endogenous) nucleotide sequence.

“Synthetic” polynucleotides are those prepared by chemical synthesis.

“Recombinant DNA molecule” is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook and Russell  
5 (2001).

The term “gene” is used broadly to refer to any segment of nucleic acid associated with a biological function. Genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA, functional RNA, or a specific protein,  
10 such as glycogen synthase enzyme 1, including its regulatory sequences. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters. In  
15 addition, a “gene” or a “recombinant gene” refers to a nucleic acid molecule comprising an open reading frame and including at least one exon and (optionally) an intron sequence. The term “intron” refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

“Naturally occurring,” “native” or “wild type” is used to describe an object  
20 that can be found in nature as distinct from being artificially produced. For example, a nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified in the laboratory, is naturally occurring. Furthermore, “wild-type” refers to the normal gene, or organism found in nature without any known mutation.

25 A “mutant” glycogen synthase enzyme 1 (GYS1) refers to the protein or fragment thereof that is encoded by a *GYS1* gene having a mutation, *e.g.*, such as might occur at the GYS1 locus. A mutation in one GYS1 allele may lead to enhanced or increased enzymatic activity in a horse heterozygous for the allele. Increased enzymatic activity can be determined by methods known to the art.  
30 Mutations in GYS1 may be disease-causing in a horse heterozygous for the mutant GYS1 allele, *e.g.*, a horse heterozygous for a mutation leading to a mutant gene

product such as a substitution mutation in exon 6 of GYS1, such as that designated herein as G926A.

“Somatic mutations” are those that occur only in certain tissues, *e.g.*, in liver tissue, and are not inherited in the germline. “Germline” mutations can be found in any of a body’s tissues and are inherited. The present *GYSEI* mutation is a germline mutation.

“Homology” refers to the percent identity between two polynucleotides or two polypeptide sequences. Two DNA or polypeptide sequences are “homologous” to each other when the sequences exhibit at least about 75% to 85% (including 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, and 85%), at least about 90%, or at least about 95% to 99% (including 95%, 96%, 97%, 98%, 99%) contiguous sequence identity over a defined length of the sequences.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) “reference sequence,” (b) “comparison window,” (c) “sequence identity,” (d) “percentage of sequence identity,” and (e) “substantial identity.”

(a) As used herein, “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, “comparison window” makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (see the World Wide Web at [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length  $W$  in the query sequence, which either match or satisfy some positive-valued threshold score  $T$  when aligned with a word of the same length in a database sequence.  $T$  is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters  $M$  (reward score for a pair of matching residues; always  $> 0$ ) and  $N$  (penalty score for mismatching residues; always  $< 0$ ). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity  $X$  from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum

probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, less than about 0.01, or even less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. When using BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. See the World Wide Web at [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov). Alignment may also be performed manually by visual inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by a BLAST program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence

identity is used in reference to proteins, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%; at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%; at least 90%, 91%, 92%, 93%, or 94%; or even at least 95%.

96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking  
5 into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, or at least 80%, 90%, or even at least 95%.

Another indication that nucleotide sequences are substantially identical is if  
10 two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified  
15 herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two  
20 nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%,  
25 76%, 77%, 78%, or 79%; at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%; or at least 90%, 91%, 92%, 93%, or 94%; or even at least 95%, 96%, 97%, 98% or 99% sequence identity to the reference sequence over a specified comparison window. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second  
30 peptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm  
5 program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under  
10 stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces  
15 minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as  
20 Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors  
25 being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the  $T_m$  can be approximated from the equation of Meinkoth and Wahl:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$$

where  $M$  is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of  
30 formamide in the hybridization solution, and  $L$  is the length of the hybrid in base pairs.  $T_m$  is reduced by about  $1^\circ\text{C}$  for each 1% of mismatching; thus,  $T_m$ ,

hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the  $T_m$  can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point ( $T_m$ ); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point ( $T_m$ ); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point ( $T_m$ ). Using the equation, hybridization and wash compositions, and desired T, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes. Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 1X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long probes (*e.g.*, >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the

particular hybridization assay indicates detection of a specific hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical.

This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum  
5 codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the  $T_m$  for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, *e.g.*, hybridization in 50% formamide,  
10 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45%  
15 formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of  
20 one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides of the invention may be altered in various ways  
25 including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. Guidance as to appropriate amino acid substitutions that do not affect biological  
30 activity of the protein of interest are well known in the art. Conservative

substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the polypeptides of the invention encompass naturally-occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired activity. The deletions, insertions, and substitutions of the polypeptide sequence encompassed herein are not expected to produce radical changes in the characteristics of the polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations."

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms."

5 A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous nucleic acid molecule. Thus, "transformed," "transgenic," and "recombinant" refer to a host cell or organism into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome generally known in the art. Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" cells have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal cells that have not been through the transformation process.

"Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically includes sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter that initiates transcription only

when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

Such expression cassettes will have the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Such an expression  
5 cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The transcriptional cassette will include in the 5'-3' direction of  
10 transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source.

15 The terms "heterologous DNA sequence," "exogenous DNA segment" or "heterologous nucleic acid," each refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example,  
20 the use of single-stranded mutagenesis. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous  
25 polypeptides.

A "homologous" DNA sequence is a DNA sequence that is naturally associated with a host cell into which it is introduced.

"Genome" refers to the complete genetic material of an organism.

30 "Coding sequence" refers to a DNA or RNA sequence that codes for a specific amino acid sequence and excludes the non-coding sequences. For example, a DNA "coding sequence" or a "sequence encoding" a particular polypeptide, is a

DNA sequence which is transcribed and translated into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory elements. The boundaries of the coding sequence are determined by a start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (*e.g.*, mammalian) DNA, and even synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence. It may constitute an "uninterrupted coding sequence," *i.e.*, lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA that is contained in the primary transcript but that is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA.

"Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

The term "regulatory sequence" is art-recognized and intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are known to those skilled in the art. It should be understood that the design of the expression vector may depend on such factors

as the choice of the host cell to be transfected and/or the amount of fusion protein to be expressed.

The term DNA "control elements" refers collectively to promoters, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell. Not all of these control sequences need always be present in a recombinant vector so long as the desired gene is capable of being transcribed and translated.

A control element, such as a promoter, "directs the transcription" of a coding sequence in a cell when RNA polymerase binds the promoter and transcribes the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones having a population of daughter cells containing the exogenous DNA.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other, *e.g.*, an arrangement of elements wherein the components so described are configured so as to perform their usual function. For example, a regulatory DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (*i.e.*, that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or

antisense orientation. Control elements operably linked to a coding sequence are capable of effecting the expression of the coding sequence. The control elements need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed  
5 sequences can be present between a promoter and the coding sequence and the promoter can still be considered "operably linked" to the coding sequence.

"Transcription stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as polyadenylation signal sequences, capable of terminating transcription. Examples include the 3' non-regulatory regions of  
10 genes encoding nopaline synthase and the small subunit of ribulose biphosphate carboxylase.

"Translation stop fragment" or "translation stop codon" or "stop codon" refers to nucleotide sequences that contain one or more regulatory signals, such as one or more termination codons in all three frames, capable of terminating  
15 translation. Insertion of a translation stop fragment adjacent to or near the initiation codon at the 5' end of the coding sequence will result in no translation or improper translation. The change of at least one nucleotide in a nucleic acid sequence can result in an interruption of the coding sequence of the gene, *e.g.*, a premature stop codon. Such sequence changes can cause a mutation in the polypeptide encoded by  
20 a *GYS1* gene. For example, if the mutation is a nonsense mutation, the mutation results in the generation of a premature stop codon, causing the generation of a truncated GYS polypeptide.

#### Nucleic Acids of the Invention

25 Sources of nucleotide sequences from which the present nucleic acid molecules can be obtained include any prokaryotic or eukaryotic source. For example, they can be obtained from a mammalian, such as an equine, cellular source. Alternatively, nucleic acid molecules of the present invention can be obtained from a library, such as the CHORI-241 Equine BAC library or the BAC  
30 library developed at INRA, Centre de Recherches de Jouy, Laboratoire de

Génétique biochimique et de Cytogénétique, Département de Génétique animale, 78350 Jouy-en-Josas Cedex, France.

As discussed above, the terms "isolated and/or purified" refer to *in vitro* isolation of a nucleic acid, *e.g.*, a DNA or RNA molecule from its natural cellular environment, and from association with other components of the cell, such as nucleic acid or polypeptide, so that it can be sequenced, replicated, and/or expressed. For example, "isolated nucleic acid" may be a DNA molecule that is complementary or hybridizes to a sequence in a gene of interest, *i.e.*, a nucleic acid sequence encoding an equine glycogen synthase enzyme, and remains stably bound under stringent conditions (as defined by methods well known in the art). Thus, the RNA or DNA is "isolated" in that it is free from at least one contaminating nucleic acid with which it is normally associated in the natural source of the RNA or DNA and in one embodiment of the invention is substantially free of any other mammalian RNA or DNA. The phrase "free from at least one contaminating source nucleic acid with which it is normally associated" includes the case where the nucleic acid is reintroduced into the source or natural cell but is in a different chromosomal location or is otherwise flanked by nucleic acid sequences not normally found in the source cell, *e.g.*, in a vector or plasmid.

As used herein, the term "recombinant nucleic acid," *e.g.*, "recombinant DNA sequence or segment" refers to a nucleic acid, *e.g.*, to DNA, that has been derived or isolated from any appropriate cellular source, that may be subsequently chemically altered *in vitro*, so that its sequence is not naturally occurring, or corresponds to naturally occurring sequences that are not positioned as they would be positioned in a genome that has not been transformed with exogenous DNA. An example of preselected DNA "derived" from a source would be a DNA sequence that is identified as a useful fragment within a given organism, and which is then chemically synthesized in essentially pure form. An example of such DNA "isolated" from a source would be a useful DNA sequence that is excised or removed from said source by chemical means, *e.g.*, by the use of restriction endonucleases, so that it can be further manipulated, *e.g.*, amplified, for use in the invention, by the methodology of genetic engineering.

Thus, recovery or isolation of a given fragment of DNA from a restriction digest can employ separation of the digest on polyacrylamide or agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. Therefore, "recombinant DNA" includes completely synthetic DNA sequences, semi-synthetic DNA sequences, DNA sequences isolated from biological sources, and DNA sequences derived from RNA, as well as mixtures thereof.

Nucleic acid molecules having base substitutions (*i.e.*, variants) are prepared by a variety of methods known in the art. These methods include, but are not limited to, isolation from a natural source (in the case of naturally occurring sequence variants) or preparation by oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared variant or a non-variant version of the nucleic acid molecule.

#### Nucleic Acid Amplification Methods

According to the methods of the present invention, the amplification of DNA present in a physiological sample may be carried out by any means known to the art. Examples of suitable amplification techniques include, but are not limited to, polymerase chain reaction (including, for RNA amplification, reverse-transcriptase polymerase chain reaction), ligase chain reaction, strand displacement amplification, transcription-based amplification, self-sustained sequence replication (or "3SR"), the Q $\beta$  replicase system, nucleic acid sequence-based amplification (or "NASBA"), the repair chain reaction (or "RCR"), and boomerang DNA amplification (or "BDA").

The bases incorporated into the amplification product may be natural or modified bases (modified before or after amplification), and the bases may be selected to optimize subsequent electrochemical detection steps.

Polymerase chain reaction (PCR) may be carried out in accordance with known techniques. See, *e.g.*, U.S. Patent Numbers 4,683,195; 4,683,202; 4,800,159; and 4,965,188. In general, PCR involves, first, treating a nucleic acid sample (*e.g.*,

in the presence of a heat stable DNA polymerase) with one oligonucleotide primer for each strand of the specific sequence to be detected under hybridizing conditions so that an extension product of each primer is synthesized that is complementary to each nucleic acid strand, with the primers sufficiently complementary to each strand  
5 of the specific sequence to hybridize therewith so that the extension product synthesized from each primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer, and then treating the sample under denaturing conditions to separate the primer extension products from their templates if the sequence or sequences to be detected are  
10 present. These steps are cyclically repeated until the desired degree of amplification is obtained. Detection of the amplified sequence may be carried out by adding to the reaction product an oligonucleotide probe capable of hybridizing to the reaction product (*e.g.*, an oligonucleotide probe of the present invention), the probe carrying a detectable label, and then detecting the label in accordance with known  
15 techniques. Where the nucleic acid to be amplified is RNA, amplification may be carried out by initial conversion to DNA by reverse transcriptase in accordance with known techniques.

Strand displacement amplification (SDA) may be carried out in accordance with known techniques. For example, SDA may be carried out with a single  
20 amplification primer or a pair of amplification primers, with exponential amplification being achieved with the latter. In general, SDA amplification primers comprise, in the 5' to 3' direction, a flanking sequence (the DNA sequence of which is noncritical), a restriction site for the restriction enzyme employed in the reaction, and an oligonucleotide sequence (*e.g.*, an oligonucleotide probe of the present  
25 invention) that hybridizes to the target sequence to be amplified and/or detected. The flanking sequence, which serves to facilitate binding of the restriction enzyme to the recognition site and provides a DNA polymerase priming site after the restriction site has been nicked, is about 15 to 20 nucleotides in length in one embodiment. The restriction site is functional in the SDA reaction. The  
30 oligonucleotide probe portion is about 13 to 15 nucleotides in length in one embodiment of the invention.

Ligase chain reaction (LCR) is also carried out in accordance with known techniques. In general, the reaction is carried out with two pairs of oligonucleotide probes: one pair binds to one strand of the sequence to be detected; the other pair binds to the other strand of the sequence to be detected. Each pair together  
5 completely overlaps the strand to which it corresponds. The reaction is carried out by, first, denaturing (*e.g.*, separating) the strands of the sequence to be detected, then reacting the strands with the two pairs of oligonucleotide probes in the presence of a heat stable ligase so that each pair of oligonucleotide probes is ligated together, then separating the reaction product, and then cyclically repeating the  
10 process until the sequence has been amplified to the desired degree. Detection may then be carried out in like manner as described above with respect to PCR.

In one embodiment of the invention, each exon of the *GYS1* gene is amplified by PCR using primers based on the known sequence. The amplified exons are then sequenced using automated sequencers. In this manner, the exons of  
15 the *GYS1* gene from horses suspected of having PSSM in their pedigree are sequenced until a mutation is found. Examples of such mutations include those in exon 6 of the *GYS1* DNA. For example, one mutation is the G to A substitution at nucleotide base 926 in exon 6. Using this technique, additional mutations causing equine PSSM can be identified.

20 According to the diagnostic method of the present invention, alteration within the wild-type *GYS1* locus is detected. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons,  
25 frameshift mutations or amino acid substitutions. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the *GYS1* gene product, or to a decrease in mRNA stability or translation efficiency. PSSM is a disease caused by a  
30 point mutation at nucleic acid 926. Horses predisposed to or have PSSM only need to have one mutated allele.

Diagnostic techniques that are useful in the methods of the invention include, but are not limited to direct DNA sequencing, PFGE analysis, allele-specific oligonucleotide (ASO), dot blot analysis and denaturing gradient gel electrophoresis, and are well known to the artisan.

5 There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. Another approach is the single-stranded conformation polymorphism assay (SSCA). This method does not  
10 detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments that have shifted mobility on SSCA  
15 gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE), heteroduplex analysis (HA) and chemical mismatch cleavage (CMC). Once a mutation is known, an allele specific detection approach such as allele  
20 specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation. Such a technique can utilize probes which are labeled with gold nanoparticles to yield a visual color result.

Detection of point mutations may be accomplished by molecular cloning of the *GYS1* allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic  
25 DNA preparation from equine tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a mutant allele: 1) single stranded conformation analysis (SSCA); 2) denaturing gradient gel electrophoresis (DGGE); 3) RNase  
30 protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6)

allele-specific PCR. For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular *GYSI* mutation. If the particular mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used. Insertions and deletions of genes can also be  
5 detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Other techniques for detecting insertions and deletions as known in the art can be used.

10 In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band that migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of  
15 mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and  
20 wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its  
25 mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe that is complementary to the horse wild-type *GYSI* gene coding sequence. The riboprobe and either  
30 mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A that is able to detect some

mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the *GYS1* mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the *GYS1* mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

10 In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. With either riboprobes or DNA probes, the cellular mRNA or DNA that might contain a mutation can be amplified using PCR before hybridization.

15 Nucleic acid analysis *via* microchip technology is also applicable to the present invention.

DNA sequences of the *GYS1* gene that have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the *GYS1* gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the *GYS1* gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the *GYS1* gene. Hybridization of allele-specific probes with amplified *GYS1* sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tissue as in the allele-specific probe.

25 Alteration of *GYS1* mRNA expression can be detected by any technique known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type *GYS1* gene.

Alteration of wild-type *GYS1* genes can also be detected by screening for alteration of wild-type GYS1 protein, or a portion of the GYS1 protein. For example, monoclonal antibodies immunoreactive with GYS1 (or to a specific portion of the GYS1 protein) can be used to screen a tissue. Lack of cognate antigen would indicate a mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant *GYS1* gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered GYS1 protein can be used to detect alteration of wild-type *GYS1* genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used that detect GYS1 biochemical function. Finding a mutant *GYS1* gene product indicates alteration of a wild-type *GYS1* gene.

Mutant *GYS1* genes or gene products can be detected in a variety of physiological samples collected from a horse. Examples of appropriate samples include a cell sample, such as a blood cell, *e.g.*, a lymphocyte, a peripheral blood cell; a sample collected from the spinal cord; a tissue sample such as cardiac tissue or muscle tissue, *e.g.*, cardiac or skeletal muscle; an organ sample, *e.g.*, liver or skin; a hair sample, especially a hair sample with roots; a fluid sample, such as blood.

The methods of diagnosis of the present invention are applicable to any equine disease in which *GYS1* has a role. The diagnostic method of the present invention is useful, for example, for veterinarians, Breed Associations, or individual breeders, so they can decide upon an appropriate course of treatment, and/or to determine if an animal is a suitable candidate as a broodmare or sire.

#### **Oligonucleotide Probes**

As noted above, the method of the present invention is useful for detecting the presence of a polymorphism in equine DNA, in particular, the presence of a G to A nucleotide substitution at position 926 in exon 6 of the coding sequence of equine *GYS1* (SEQ ID NO:1). This substitution results in the replacement of an arginine (R) amino acid at codon 309 by a histidine (H) in the glycogen synthase protein (SEQ ID NO:9).

Primer pairs are useful for determination of the nucleotide sequence of a particular *GYS1* allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the *GYS1* gene in order to prime amplifying DNA synthesis of the *GYS1* gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the *GYS1* coding sequences, *i.e.*, the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular *GYS1* mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

The first step of the process involves contacting a physiological sample obtained from a horse, which sample contains nucleic acid, with an oligonucleotide probe to form a hybridized DNA. The oligonucleotide probes that are useful in the methods of the present invention can be any probe comprised of between about 4 or 6 bases up to about 80 or 100 bases or more. In one embodiment of the present invention, the probes are between about 10 and about 20 bases.

The primers themselves can be synthesized using techniques that are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines that are commercially available. Given the sequence of the *GYS1* coding sequence as set forth in SEQ ID NO:1, design of particular primers is well within the skill of the art.

Oligonucleotide probes may be prepared having any of a wide variety of base sequences according to techniques that are well known in the art. Suitable bases for preparing the oligonucleotide probe may be selected from naturally occurring nucleotide bases such as adenine, cytosine, guanine, uracil, and thymine; and non-naturally occurring or "synthetic" nucleotide bases such as 7-deaza-guanine, 8-oxo-guanine, 6-mercaptopguanine, 4-acetylcytidine, 5-(carboxyhydroxyethyl)uridine, 2'-O-methylcytidine, 5-carboxymethylamino-methyl-2-thioridine, 5-carboxymethylaminomethyluridine, dihydrouridine, 2'-O-methylpseudouridine,  $\beta$ ,D-galactosylqueosine, 2'-O-methylguanosine, inosine, N6-isopentenyladenosine, 1-methyladenosine, 1-methylpseudouridine, 1-methylguanosine, 1-methylinosine, 2,2-dimethylguanosine, 2-methyladenosine, 2-

methylguanosine, 3-methylcytidine, 5-methylcytidine, N6-methyladenosine, 7-methylguanosine, 5-methylaminomethyluridine, 5-methoxyaminomethyl-2-thiouridine,  $\beta$ ,D-mannosylqueosine, 5-methoxycarbonylmethyluridine, 5-methoxyuridine, 2-methylthio-N6-isopentenyladenosine, N-((9- $\beta$ -D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine, N-((9- $\beta$ -D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine, uridine-5-oxyacetic acid methylester, uridine-5-oxyacetic acid, wybutoxosine, pseudouridine, queosine, 2-thiocytidine, 5-methyl-2-thiouridine, 2-thiouridine, 2-thiouridine, 5-Methyluridine, N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine, 2'-O-methyl-5-methyluridine, 2'-O-methyluridine, wybutosine, and 3-(3-amino-3-carboxypropyl)uridine. Any oligonucleotide backbone may be employed, including DNA, RNA (although RNA is less preferred than DNA), modified sugars such as carbocycles, and sugars containing 2' substitutions such as fluoro and methoxy. The oligonucleotides may be oligonucleotides wherein at least one, or all, of the internucleotide bridging phosphate residues are modified phosphates, such as methyl phosphonates, methyl phosphonotioates, phosphoroinorpholidates, phosphoropiperazidates and phosphioramidates (for example, every other one of the internucleotide bridging phosphate residues may be modified as described). The oligonucleotide may be a "peptide nucleic acid" such as described in Nielsen *et al.*, Science, 254, 1497-1500 (1991).

The only requirement is that the oligonucleotide probe should possess a sequence at least a portion of which is capable of binding to a known portion of the sequence of the DNA sample.

It may be desirable in some applications to contact the DNA sample with a number of oligonucleotide probes having different base sequences (*e.g.*, where there are two or more target nucleic acids in the sample, or where a single target nucleic acid is hybridized to two or more probes in a "sandwich" assay).

The nucleic acid probes provided by the present invention are useful for a number of purposes. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the *GYS1* gene or mRNA using other techniques.

### Hybridization Methodology

The DNA (or nucleic acid) sample may be contacted with the oligonucleotide probe in any suitable manner known to those skilled in the art. For example, the DNA sample may be solubilized in solution, and contacted with the oligonucleotide probe by solubilizing the oligonucleotide probe in solution with the DNA sample under conditions that permit hybridization. Suitable conditions are well known to those skilled in the art. Alternatively, the DNA sample may be solubilized in solution with the oligonucleotide probe immobilized on a solid support, whereby the DNA sample may be contacted with the oligonucleotide probe by immersing the solid support having the oligonucleotide probe immobilized thereon in the solution containing the DNA sample.

### Example 1: Method of Detecting a DNA Mutation Associated with Equine Polysaccharide Storage Myopathy

The present invention relates to mutations in the *GYS1* gene and their use in the diagnosis of PSSM, the diagnosis of predisposition to PSSM, and to the detection of a mutant *GYS1* allele in a horse.

The present inventors discovered a mutation in the equine *GYS1* gene (encoding the skeletal muscle glycogen synthase enzyme) that is present in many populations of PSSM affected horses studied to date. This was possible by first deriving the protein-encoding DNA sequence of the equine *GYS1* gene from mRNA isolated from skeletal muscle of both an affected and a control horse. In both horses the sequence length from the start codon (ATG) to the stop codon (TAA) was 2,214 bases (Figure 1) and would code for a protein of 737 amino acids. The only difference between the PSSM and control horse sequences was a G to A base substitution in exon 6 at nucleotide position 926.

The DNA sequence difference at position 926 of the *GYS1* coding sequence present in skeletal muscle mRNA was subsequently confirmed in the genomic DNA of several horses. An expanded view of exon 6 with its flanking intron sequence from genomic DNA is shown in Figure 2. Figure 2 also shows that the change from

a G to A in the DNA sequence causes the replacement of an arginine (R) amino acid at codon 309 by a histidine (H) in the glycogen synthase protein. Thus, this mutation may be referred to as the G926 to A926 DNA mutation or the R309 to H309 amino acid mutation. The normal alleles of this gene may be referred to as G926, R or R309, and the mutant alleles as A926, H or H309.

To date, no other mutations in the *GYS1* gene have been shown to cause a glycogen storage disease in humans or animal species. The related *GYS2* gene, encoding the liver form of glycogen synthase that is expressed in non-muscle tissues, has several known mutations that lead to a deficiency in this enzyme and fasting hypoglycemia. However, unlike the *GYS2* mutations that greatly reduce the activity of the glycogen synthase enzyme and are inherited in a recessive manner, the PSSM horse muscle *GYS1* mutation does not reduce the glycogen synthase activity. Rather, it appears to result in an increased glycogen synthase activity and be inherited in a dominant fashion (see Table 1 below). This region of the muscle glycogen synthase amino acid sequence contained in exon 6 is highly conserved throughout the animal kingdom, lending support to its mutation in PSSM horses being a causative mutation (Figure 3).

The inventors have found the *GYS1* R to H mutation in PSSM-affected Quarter Horses, Draft horses, and Warmbloods (Table 1), and it is likely to extend to even more breeds of horses. Approximately 80% of the Quarter Horses and Belgian Draft Horses diagnosed with PSSM by the muscle biopsy method thus far are either homozygous (have two copies of the H allele; H/H) or are heterozygous (an H and an R allele; R/H). PSSM horses with the *GYS1* H allele can be of either sex, and this is consistent with, but does not prove an autosomal dominant inheritance. Only 4% of Quarter Horses and 14% of the Belgian Draft Horses with negative biopsy results were heterozygous. The inventors believe this in large part reflects the less than 100% accuracy of the current diagnostic method, but could also reflect an incomplete penetrance; *i.e.*, carriers of the H allele may not always develop disease symptoms due to other genetic and environmental factors.

30

**Table 1. GYS1 Genotype Frequencies in PSSM and Control Horses of Different Breeds**

Genotype	PSSM QH	Control QH	PSSM Belgian	Control Belgian	PSSM Warmblood	Control Warmblood
R/R	18	85	4	29	1	4
R/H	67	4	28	5	3	0
H/H	4	0	4	0	0	0

5 That approximately 20% of the Quarter Horses and 11% of Belgians with abnormal polysaccharide in muscle biopsies and clinical signs of PSSM do not carry the GYS1 H allele is suggestive that there may be other causes of PSSM. In other words, the GYS1 mutation appears to explain most, but not all cases of equine PSSM, and there is likely to be another gene responsible for a non-GYS1 form of PSSM that will be need to be the subject of additional investigation.

10 The inventors have determined the GYS1 genotype frequency in random populations of horses obtained from samples submitted for the purposes of breed registration requirements. Hair root samples were taken from every 10<sup>th</sup> submission to ensure even distribution across the US. Table 2 indicates that the GYS1 mutation is very prevalent in four major breeds examined to date, but not yet in 15 Thoroughbreds. The GYS1 genotype distribution in Quarter Horses and Paint Horses is similar at 6-7% heterozygous with few homozygotes for the H allele. However, approximately 42% of Percherons are heterozygous and 14% are homozygous for the H allele. Since the GYS1 H allele appears to be dominant we 20 predict that approximately 7% of all Quarter Horses and Paint horses, 36% of all Belgians and 56% of all Percherons are actually genetically susceptible to PSSM.

**Table 2. GYS1 Genotype Frequencies in Random Sample Populations of Different Breeds**

25

Genotype	Quarter Horses	Paint Horses	Belgian	Percheron	Thoroughbred
R/R	313 (93%)	180 (92%)	20 (61%)	22 (44%)	96 (100%)
R/H	21 (6%)	14 (7%)	13 (26%)	21 (42%)	0 (0%)
H/H	1 (<1%)	1 (<1%)	5 (10%)	7 (14%)	0 (0%)

The nearly complete DNA sequence of the horse *GYS1* gene (Horse *GYS1* Intron 5, Exon 6, and Intron 6; Figure 4) was assembled from sequences deposited into the NCBI trace sequence archive by the Broad Institute sequencing center during their recent equine whole genome shotgun sequencing project (SEQ ID NO:6). Introns and exons of the horse *GYS1* gene sequence were then predicted from the homologous *GYS1* exon sequences of other mammals. Intron 5 in this sequence comprises bases 1 – 471. Exon 6 in this sequence is highlighted and comprises bases 472 – 589. Intron 6 in this sequence comprises bases 590 – 886. The G to A mutation in exon 6 that causes the R to H amino acid mutation at codon 309 is underlined and is at base 574.

Using the *GYS1* sequence, PCR primers are developed that can amplify the PSSM *GYS1* mutation. For example, a PCR primer pair that has been successfully and reliably used to amplify this region from isolated horse DNA samples lies in introns 5 and 6 and the sequence locations are also underlined (Figure 4). These sequences are 5'-TGAAACATGGGACCTTCTCC-3' (SEQ ID NO:7) and 5'-AGCTGTCCCCTCCCTTAGAC-3' (SEQ ID NO:8). Many other primer pairs are also possible.

Using the above PCR primers to amplify the region, the genotype of any horse (G/G, G/A or A/A for the DNA sequence, and R/R, R/H, and H/H for the amino acid sequence) can be obtained. In this method the restriction enzyme *HypCH4* V cuts the *GYS1* H allele at the exon 6 site (base 574), as well as at an intronic site 100 bp distant present in both the R and H alleles that serves to monitor enzyme efficiency. The products are separated by agarose gel electrophoresis and visualized by ethidium bromide staining under ultraviolet light. Many other methods of detecting the G or A nucleotide at this position of the horse *GYS1* sequence are possible.

DNA testing based on the present invention now provides veterinarians and veterinary pathologists with a means to more accurately determine if a horse with clinical signs of PSSM has the heritable and most common form of disease that can be specifically attributed to this *GYS1* gene mutation. All that is needed are a tissue

sample containing the individual's DNA (typically hair root or blood) and appropriate PCR and sequence analysis technology to detect the G to A single nucleotide change. Such PCR primers are based in exon 6 and its flanking intron sequences as depicted in Figure 2, sequences nearby this region depicted in Figure 1, or in other DNA sequence from introns of this gene.

Also, DNA testing provides owners and breeders with a means to determine if any horse can be expected to produce offspring with this form of PSSM. An H/H horse would produce an affected foal 100% of the time, while an H/R horse would produce an affected foal 50% of the time when mated to an R/R horse. Mating of H/H and H/R horses would produce an affected foal 75% of the time. Breeding programs could incorporate this information in the selection of parents that could eventually reduce and even eliminate this form of PSSM in their herds.

All publications are incorporated by reference herein, as though individually incorporated by reference. The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the scope of the invention.

## WHAT IS CLAIMED IS:

1. A method for detecting the presence of Polysaccharide Storage Myopathy (PSSM) in a horse, comprising identifying in a nucleic acid sample from a horse nucleotide 926 of SEQ ID NO:1, wherein the presence of an adenine (A) nucleotide at nucleotide 926 in one or both alleles is indicative of the horse being predisposed to or has PSSM.
2. The method of claim 1, further comprising contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid.
3. The method of claim 2, wherein exon 6 of equine glycogen synthase enzyme 1 or a portion thereof is amplified.
4. The method of claim 2, wherein the amplification of the hybridized nucleic acid is carried out by polymerase chain reaction, strand displacement amplification, ligase chain reaction, or nucleic acid sequence-based amplification.
5. The method according to claim 2, wherein at least one oligonucleotide probe is immobilized on a solid surface.
6. A method for detecting the presence of an Polysaccharide Storage Myopathy (PSSM) allele in a horse, comprising identifying in a nucleic acid sample from the horse nucleotide 926 of SEQ ID NO:1, wherein the presence of an adenine (A) nucleotide at nucleotide 926 in one or both alleles is indicative of the horse being predisposed to or has PSSM.
7. The method of claim 6, further comprising contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid.

8. The method of claim 7, wherein exon 6 of equine glycogen synthase enzyme 1 or a portion thereof is amplified.
9. The method of claim 7, wherein the amplification of the hybridized nucleic acid is carried out by polymerase chain reaction, strand displacement amplification, ligase chain reaction, or nucleic acid sequence-based amplification.
10. The method according to claim 7, wherein at least one oligonucleotide probe is immobilized on a solid surface.
11. A method for detecting the presence of a biomarker associated with equine Polysaccharide Storage Myopathy (PSSM), comprising determining the presence of the biomarker in a physiological sample from a horse, wherein the sample comprises nucleic acid.
12. The method of claim 11, further comprising contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid.
13. The method of claim 12, wherein exon 6 of equine glycogen synthase enzyme 1 (*GYS1*) or a portion thereof is amplified.
14. The method of claim 12, wherein the amplification of the hybridized DNA is carried out by polymerase chain reaction, strand displacement amplification, ligase chain reaction, or nucleic acid sequence-based amplification.
15. The method according to claim 12, wherein at least one oligonucleotide probe is immobilized on a solid surface.

16. The method of claim 11, wherein the biomarker comprises an equine *GYS1* gene having an A at nucleotide 926 in exon 6.
17. The method of claim 11, wherein the *GYS1* gene encodes an enzyme having an H at amino acid residue 309.
18. A method for diagnosing Polysaccharide Storage Myopathy (PSSM) in a horse comprising detecting the presence of a biomarker in a physiological sample from the horse the sample, wherein the presence of the biomarker is indicative of the disease.
19. The method of claim 18, wherein the sample comprises nucleic acid.
20. The method of claim 18, further comprising contacting the sample with at least one oligonucleotide probe to form a hybridized nucleic acid and amplifying the hybridized nucleic acid.
21. The method of claim 20, wherein exon 6 of glycogen synthase enzyme 1 (*GYS1*) or a portion thereof is amplified.
22. The method of claim 21, wherein the amplification of the hybridized DNA is carried out by polymerase chain reaction, strand displacement amplification, ligase chain reaction, or nucleic acid sequence-based amplification.
23. The method of claim 18, wherein the biomarker comprises an equine *GYS1* gene has an A at nucleotide 926 in exon 6.
24. The method of claim 23, wherein the *GYS1* gene encodes an enzyme has an H at amino acid residue 309.
25. The method of claim 18, wherein the biomarker is a *GYS1* protein.

26. The method of claim 25, wherein the GYS1 protein an enzyme has an H at amino acid residue 309.

**Figure 1.**

```

1   ATGCCTCTAA ACCGCACTTT GTCCATGTCC TCACTGCCAG GACTGGACGA
51  TTGGGAGGAT GAATTTCGACC TGGAGAATAC AGTGCTCTTC GAGGTGGCCT
101 GGGAGGTGGC CAACAAGGTG GGTGGCATCT ACACGGTGCT ACAGACGAAG
151 GCGAAGGTGA CAGGGGATGA ATGGGGCGAC AACTACTACC TGGTGGGACC
201 ATACACGGAG CAAGGCGTGA GGACCCAGGT GGAGCTGCTC GAGCCCCCAA
251 CCCC GGCCCT GAAGAGGACG CTGGACTCCA TGAACAGCAA GGGCTGCAAG
301 GTGTATTTTCG GCGCTGGCT GATCGAGGGG GGCCCCCTGG TGGTGTCTCT
351 GGATGTGGGG GCCTCAGCCT GGGCCCTGGA ACGCTGGAAG GGAGAGCTTT
401 GGGACACCTG CAACATCGGG GTGCCCTGGT ACGACCGTGA GGCCAACGAC
451 GCCGTCTTTT TTGGCTTCCT CACCACCTGG TTCTTGGGTG AGTTCCTGGC
501 CCAGAGCGAG GAGAAGCCAC ATGTGGTTGC ACACTTCCAC GAGTGGTTGG
551 CGGGCATCGG GCTCTGCCTG TGCCGTGCCG GCGGCTGCC TGTGGCTACA
601 ATCTTCACCA CCCACGCCAC GCTGCTGGGG CGATACCTGT GTGCCGGTGC
651 TGTGGACTTC TACAACAACC TGGAGAATTT CAACGTGGAC AAGGAAGCTG
701 GTGAGAGGCA AATTTATCAC CGTTACTGCA TGGAGCGGGC GGCAGCCCAC
751 TGCCTCAGC TCTTCACTAC CGTGTCCAG ATCACCGCCA TTGAGGCTCA
801 GCACCTACTC AAGAGGAAAC CAGATATCGT GACCCCAAT GGACTGAATG
851 TGAAGAAGTT CTCTGCCATG CATGAGTTCC AGAACCTCCA TGCTCAGAGC
901 AAGGCCCGAA TCCAGGAGTT TGTGCGTGGC CATTTTTTATG GGCACCTGGA
951 CTTCAACTTG GATAAGACCC TGTATTTCTT TATCGCCGGC CGCTACGAGT
1001 TCTCCAACAA GGGGGCTGAC GTCTTCTTGG AGGCCTTGGC CCGGCTCAAC
1051 TATCTGCTCA GAGTAAATGG CAGCGAGCAG ACGGTGGTCG CCTTCTTCAT
1101 CATGCCGGCT CGGACCAACA ACTTCAACGT GGAAACCCTC AAGGGGCAAG
1151 CCGTGCACAA GCAGCTCTGG GATACGGCGA ACACAGTGAA GGAGAAGTTC
1201 GGGAGGAAGC TTTACGAATC CCTGCTGGTT GGGAGCCTCC CGGACATGAA
1251 CAAGATGCTG GACAAGGAGG ATTTCACTAT GATGAAGAGA GCCATCTTTG
1301 CCACGCAGCG GCAGTCTTTT CCCCCTGTGT GCACCCACAA TATGCTGGAC
1351 GACTCCTCAG ACCCTATCCT GACCACCATC CGTCGAATCG GCCTCTTCAA
1401 TAGTAGTGCT GACAGGGTCA AGGTGATTTT CCACCCAGAG TTCTCTCTCT
1451 CCACGAGCCC CCTGCTCCCC GTGGACTATG AGGAGTTTGT CCGTGGCTGC
1501 CACCTTGGGG TTTTCCCCTC CTA CTATGAG CCTTGGGGCT ACACACCAGC
1551 TGAGTGCACG GTTATGGGCA TCCCAGTAT CTCCACCAAC CTCTCCGGCT
1601 TCGGCTGCTT CATGGAGGAA CACATCGCAG ACCCCTCAGC TTACGGCATC
1651 TACATTCTGG ACCGGCGGTT CCGCAGCCTG GATGATTCCT GCTCGCAGCT
1701 TACCTCCTTC CTCTACAGCT TCTGCCAGCA GAGCCGGCGG CAGCGCATCA
1751 TCCAGCGGAA CCGCACGGAG CGCCTCTCCG ACCTTCTGGA CTGGAAATAC
1801 CTAGGCCGGT ACTATATGTC CGCGCGCCAC ATGGCGCTGG CCAAGGCCTT
1851 TCCAGAACAT TTCACCTACG AGCCCCGCGA GGCTGATGCG ACCCAGGGCT
1901 ACCGCTACCC ACGGCCTGCA TCGGTGCCTC CGTCGCCCTC ACTGTCACGA
1951 CACTCGAGCC CGCACCAGAG CGAGGACGAG GAGGAGCCCC GGGACGTGCC
2001 GCCCGATGAA GACAGTGAGC GCTACGACGA GGACGAGGAG GCCGCCAAGG
2051 ACCGGCGCAA CATCCGCGCC CCGGAGTGGC CGCGTCGCGC CTCCTGCACC
2101 TCTTCCACGA GCGGGAGCAA GCGCGGCTCG GTGGACACGG GGCCCTCCAG
2151 CTCGCTCAGC ACCCCAGCG AGCCCTCAG CCCCAGCAG TCCTTGGGCG
2201 AGGAGCGCAA CTAA (SEQ ID NO:1)
    
```

**Figure 2.**

	1	11	21	31	41
Normal	TTGAAACATG	GGGCCTTCTC	CCCCATGCCT	AGATATCGTG	ACCCCCAATG
PSSM	TTGAAACATG	GGGCCTTCTC	CCCCATGCCT	AGATATCGTG	ACCCCCAATG
	51	61	71	81	91
Normal	GACTGAATGT	GAAGAAGTTC	TCTGCCATGC	ATGAGTTCCA	GAACCTCCAT
PSSM	GACTGAATGT	GAAGAAGTTC	TCTGCCATGC	ATGAGTTCCA	GAACCTCCAT
	101	111	121	131	141
Normal	GCTCAGAGCA	AGGCCCGAAT	CCAGGAGTTT	GTGCGTGGCC	ATTTTTATGG
PSSM	GCTCAGAGCA	AGGCCCGAAT	CCAGGAGTTT	GTGCATGGCC	ATTTTTATGG
	151	161	171	181	191
Normal	GTATGTGGGC	CAGATACCCA	GGTCTTGAGA	GAGGTGGGGG	TTGGGTGCCC
PSSM	GTATGTGGGC	CAGATACCCA	GGTCTTGAGA	GAGGTGGGGG	TTGGGTGCCC
	201	211	221	231	
Normal	AGACTCCCGG	GTCTAAGGGG	GGGACAGCTA	A (SEQ ID NO:2)	
PSSM	AGACTCCCGG	GTCTAAGGGG	GGGACAGCTA	A (SEQ ID NO:3)	

**Figure 3.**

	1	11	21	31
PSSM Horses	IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFV <u>H</u> GHFYG
(SEQ ID NO:4)				
All other species	IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFV <u>R</u> GHFYG
(SEQ ID NO:5)				

**Figure 4.**

```

1   GTGGGGCATA GGAGGGGGTC CCGAGACATG AATGAGGTAG TTGACGGTCT
51  ACGCCAAGGC CCGTACCAAT TTTCTTATTG TAAAGAGCAG AGAACTACAA
101 CTCCCAGAAG ACCCTAAGTT GATGGCCTGC ATATGAGGCT GCCGGGCTGT
151 ACCCTAGAGG CTAATGGGAG GTGTAGTTTC TTTGACATTC ATTTTCAGAAA
201 AGGGAAAATT ATTTTATTGG CTAAGGTGGA TATTGGTCTT CTGAAAGAGG
251 ATGGAGGTTT ACTGGATTTT GTTTTTGAGA AAAGCAGTGT TAGAACCAGT
301 TATTCTTCCC AGGCTATAGC ATCAATTCTC AATGGATGTG AGAACTACAA
351 CTCCCAGAAG GCCTCTGGCT GGCCAGCTCC TTAATATTCC TGGAATTACC
401 TTCATTTATT CATTCAACCA CATTGATCTG AATGAAGAAG TGAAACATGG
451 GACCTTCTCC CCCATGCCTA GATATCGTGA CCCCCAATGG ACTGAATGTG
501 AAGAAGTTCT CTGCCATGCA TGAGTTCCAG AACCTCCATG CTCAGAGCAA
551 GGCCCGAATC CAGGAGTTTG TGCGTGGCCA TTTTTATGGG TATGTGGGCC
601 AGATACCCAG GTCTTGAGAG AGGTGGGGGT TGGGTGCCCA GACTCCCGGG
651 TCTAAGGGAG GGGACAGCTG GGGGCTCAGA CTCTTGAGTT CCTGAGCGTC
701 AGAACAATGG CTGAGTATTG CAGGAGATCC CCGTTTAAGG AGAAAAGAGG
751 CACAAGTCCC ATGAGTCCCA TGATGCCTCT GTACTGTCTA AGATGTTCTT
801 TGCCCTGTG GCTTCTGGGG TTTGTAGTTT TGAGCCAGGG CTAGAGGGTG
851 AGGATCTTGG CTTACCCCTG TCTTGTGGCT TTTTAG (SEQ ID NO:6)
    
```

**Figure 5.**

```

1   MPLNRTLMS SLPGLDDWED EFDLENTVLF EVAWEVANKV GGIYTVLQTK
51  AKVTGDEWGD NYYLVGPYTE QGVRTQVELL EPPTPALKRT LDSMNSKGCK
101 VYFGRWLIEG GPLVVLLDVG ASAWALERWK GELWDTCNIG VPWYDREAND
151 AVLFGFLTTW FLGEFLAQSE EKPHVVAHFH EWLAGIGLCL CRARRLPVAT
201 IFTTHATLLG RYLCAGAVDF YNNLENFNVD KEAGERQIYH RYCMERAAAH
251 CTHVFTTVSQ ITAIEAQHLL KRKPDI VTPN GLNVKKFSAM HEFQNLHAQS
301 KARIQEFVRG HFYGHLDLNL DKTLYFFIAG RYEF SNKGAD VFLEALARLN
351 YLLRVNGSEQ TVVAFFIMPA RTNNFNVELL KGQAVRKQLW DTANTVKEKF
401 GRKLYESLLV GSLPDMNKML DKEDFTMMKR AIFATQRQSF PPVCTHNMLD
451 DSSDPILTTI RRIGLFNSSA DRVKVIFHPE FLSSTSP LLP VDYEEFVRGC
501 HLGVFPSYYE PWGYTPAECT VMGIPSISTN LSGFGCFMEE HIADPSAYGI
551 YILDRRFRSL DDSCSQLTSE LYSFCQQSRR QRIIQNRNTE RLSDLLDWKY
601 LGRYYMSARH MALAKAFPEH FTYEPREADA TQGYRYPRPA SVPPSPSLSR
651 HSSPHQSEDE EEPDRVPPDE DSERYDEDEE AAKDRRNIRA PEWPRRASCT
701 SSTSGSKRGS VDTGPSSSLS TPSEPLSPAS SLGEERN-
    
```

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2007/062134

## A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. *C12Q 1/68* (2006.01) *G01N 33/68* (2006.01)

US. Cl. 435/6 436/86 436/63

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 WPIDS, CA, Medline, Biosis, JAPIO with keywords: polysaccharide storage myopathy, equine, glycogen synthase enzyme 1, EC 2.4.1.11, and abbreviations, plurals and synonyms.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VALBERG, S. J. ET AL. 'Familial basis of exertional rhabdomyolysis in Quarter Horse-related breeds' American Journal of Veterinary Research (1996) Volume 57, No. 3, pages 286-290. See page 286 column 1, 'Conclusions'; page 289 paragraph 2 of the 'Discussion'; page 290, last 12 lines of paragraph 2.	
A	MCDIARMID, A. 'Possible familial basis to equine polysaccharide storage myopathy' The Veterinary Record (2005) Volume 156, Number 3, pages 95-96. See page 95, paragraph 4.	



Further documents are listed in the continuation of Box C



See patent family annex

## \* Special categories of cited documents:

"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
17 August 2007

Date of mailing of the international search report  
**06 FEB 2008**

Name and mailing address of the ISA/US  
Mail Stop PCT, Attn: ISA/US  
Commissioner for Patents  
P.O. Box 1450, Alexandria, Virginia 22313-1450  
Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774

