1-1-2010

Evolution Of Lactate Dehydrogenase Genes In Primates, With Special Consideration Of Nucleotide Organization In Mammalian Promoters

Zack Papper
Wayne State University,

EVOLUTION OF LACTATE DEHYDROGENASE GENES IN PRIMATES, WITH SPECIAL CONSIDERATION OF NUCLEOTIDE ORGANIZATION IN MAMMALIAN PROMOTERS

by

ZACK PAPPER

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2010

MAJOR: MOLECULAR BIOLOGY AND GENETICS (Evolution)

_____

Advisor Date

_____

_____

_____

DEDICATION

This work, and the educational endeavors behind it, are dedicated to Dr. Renee Papper and Dr. Solomon Papper. They have taught me that a great mind is developed through humility and respect, securing my permanent status as a student.

ACKNOWLEDGMENTS

This page is not sufficient to describe my appreciation for my advisor, Dr. Derek E. Wildman. He supported me through many hurdles, and I can only hope to some day pay back his kindness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE—INTRODUCTION

A myriad of genes are involved in cellular metabolism, and the coordination of transporters, enzymes, transcription factors, and many others enables consistent, life-sustaining energy production. Specific, derived phenotypes have been associated with changes to energetics, such as hummingbird hovering (Welch and Suarez 2007) and primate encephalization (Mink et al. 1981; Syner and Goodman 1966). Two genes, lactate dehydrogenase A (*LDHA*) and B (*LDHB*), are the primary components of the lactate dehydrogenase enzyme, LDH. This enzyme bridges aerobic and anaerobic metabolism. The products of these genes have changed in relative abundance in the brain of primates, with a shift from primarily LDHA in strepsirrhines to predominantly LDHB in anthropoids (Goodman et al. 1969; Syner and Goodman 1966). The previous studies on primate LDH focused on proteins.

This thesis is a study of the evolution of the *LDHA* and *LDHB* genes in primates, with two distinct objectives. In the first project, we aimed to characterize the evolution of the coding sequences, hypothesizing that modifications to the gene products could alter the metabolic contributions of the resulting enzyme. In the second project, we aimed to identify the *cis* elements in the promoters of these two genes that may contribute to the changes in relative abundance previously described in anthropoid primate brains. These two projects were carried out by use of comparative genomics, for which we acquired orthologous sequence data across a number of primate species to determine

both the conserved and derived elements within and across different primate clades.

The study of *LDHA* and *LDHB* promoter regions led us to consider general characteristics of *cis* elements that are conserved across diverse taxa (e.g. width or composition biases), and we discovered that very little is known regarding the makeup of conserved *cis* elements. This led to a third project exploring the nucleotide composition of conserved *cis* elements in the promoters of mammals. Finally, a fourth project was conducted addressing the history of placental growth hormone expression in primates. For clarity, the research projects are organized starting with the study on *cis* element composition across mammals (chapter two), followed by the study on the evolution of *LDHA* and *LDHB* gene promoters (chapter three), the results of the study on the evolution of *LDHA* and *LDHB* coding regions (chapter four), and finally the published study on the history of placental growth hormone expression in primates (chapter five). This first chapter will serve as an introduction and background to these projects.

BACKGROUND

The catalog of molecular elements involved in gene regulation is immense, including *cis*-regulatory elements, trans-acting factors, enhancers, insulators, histone modifications, cytosine methylation, and small RNAs, amongst others (Cheng and Blumenthal 2010; Weake and Workman 2010; Zhang 2009). The endogenous expression of any one gene likely involves contributions from many of these factors, with numerous combinations of regulatory elements acting

on any one gene. This abundance impedes our ability to understand regulatory networks, since each gene is typically approached as an isolated unit. Consequently, extraordinary costs in both time and money are required in order to determine the mechanisms involved in regulation, and these costs become further amplified when evaluating context-dependent expression such as embryonic versus adult, male versus female, one tissue versus another, etc. The challenges faced when studying gene regulation are different than those when studying protein-coding genes, primarily because of the universal utility of the genetic code. No such code exists for regulatory elements (Pabo and Sauer 1992). Degeneracy in elements bound by transcription factors (TFs), multiple TFs binding the same target sequence, elements functioning as activators in one context and as repressors in others, all contribute to making a reliable, universal regulatory code unfeasible. Despite these complications, there is reason to suggest that some boundaries exist in limiting regulatory network organizations. Evolutionary constraints on all of these elements (*cis* elements, TFs, methylation, etc.) may restrict nucleotide sequences bound by TFs, molecular design of TFs (such as zinc-finger and homeodomain), level of methylation, and so on. Such constraints would prove invaluable to understanding and ultimately predicting regulatory networks without exhaustive experimental analyses. In addition, constraints themselves, as barriers to modifications, reveal the boundaries of molecular evolution, providing valuable insights into the possibilities, and restrictions in the evolution of gene regulation.

We hypothesized that such evolutionary constraints do exist in defining regulatory networks. Although all elements in a regulatory network may be reasonable targets, we decided that promoter regions are an ideal starting point for detecting regulatory constraint.

Promoter regions have been studied for decades, and, as the name itself implies, they often include *cis*-regulatory elements that are critical for endogenous expression of the downstream gene (Eron and Block 1971; Ippen et al. 1968). *Cis* elements include non-bound functional elements involved in maintaining local topological features (Parker et al. 2009), short RNAs (Zhang 2009), sites targeted for methylation (De Bustos et al. 2009), and transcription factor binding sites (TFBS) (Dierks et al. 1983; Lifton et al. 1978).

While promoters do not have a defined size, they do have a defined location, residing immediately upstream of the transcriptional start site (TSS) of an adjacent gene. A gene TSS can be related across species, a necessary feature when attempting to detect evolutionary constraints. As the number of sequenced genomes increases, in addition to transcriptomes (Maeda et al. 2006; Okamoto et al. 2010; Zhao et al. 2010) and improvements in gene annotations, comparisons of promoter landscapes become more viable for evaluating the nucleotide features shared across species. In fact, comparative genomics has been widely used for discovery of *cis*-regulatory elements since the introduction of phylogenetic footprinting more than twenty years ago (Tagle et al. 1988). By searching for conserved sequences (footprints) within nonconserved landscapes, such as promoters, those elements conserved across species are likely

maintained due to a functional role. Similar approaches have been used, and modified, to produce numerous programs aimed at detecting *cis*-regulatory elements (Berezikov et al. 2007; Kechris and Li 2008; Satija et al. 2009). While the success of this approach in detecting elements has been demonstrated at both the individual locus (Tagle et al. 1988) as well as genome-wide (Ortiz-Barahona et al. 2010), no study to date has evaluated the nucleotide composition of these elements across gene promoters.

With whole genome sequences, annotated TSSs for each gene within the genome, and the proven utility of comparative genomics, it is feasible to construct a database of the conserved elements across specific taxonomic groups. Upon acquisition of conserved elements, we can ask whether there exist constraints on the sequences that make up this database, such as nucleotide sequence biases across motifs, biases between the sense and antisense strands, or biases at specific positions within conserved elements. If no such evolutionary constraints exist, then the makeup of conserved elements would be subject to the individual nucleotide compositions across the genomes (adenine, cytosine, guanine, and thymine percentages). If, however, certain nucleotide strings are favored over others, this suggests that there exist evolutionary constraints in the organization of regulatory elements. There are likely species-specific elements that would not adhere to the organizational structure of elements conserved across broader taxonomic sampling, but these would indicate changes to regulatory networks during the evolutionary history of that species.

Special considerations are necessary in such a study. Since these elements are short, the quality of a sequenced genome can have a significant impact on detecting conserved elements. In addition, annotation of TSS has the obvious impact of dictating the regions of comparison across species. Genome quality and gene annotation quality suggest that only high-quality, well-annotated genomes should be included for evaluation. As a result, the quality threshold, determined by the individual investigator, dictates which taxonomic groups can be evaluated based on which genomes are considered sufficient. For instance, there is only one lizard genome available, and it is poorly annotated for genes (Sanger et al. 2008). As a consequence, evaluating regulatory constraints within squamates (lizards and snakes) cannot be done using this approach at present. Considering these factors, we chose placental mammals, a taxonomic group that fits these criteria, as well as a group relevant to our research as a whole. Within placental mammals, we determined that seven genomes (human, chimpanzee, mouse, rat, cow, dog, and opossum) have sufficient sequencing coverage as well as gene annotation for TSS to be confident in acquiring reliable data to test whether evolutionary constraints exist in conserved elements across this taxonomic clade. This study is the subject of chapter two.

Evolution, versus conservation, of *cis* elements

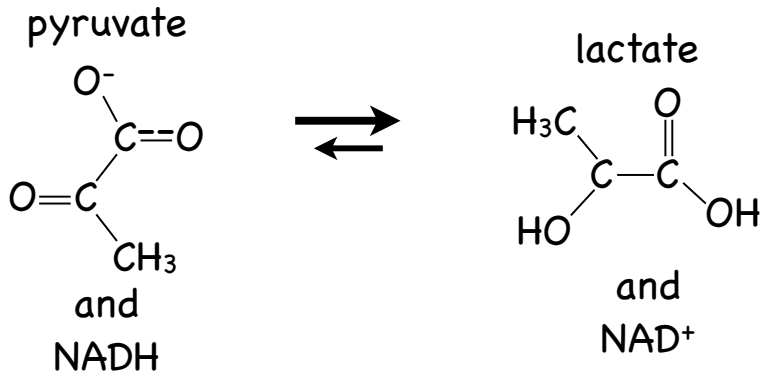
In the previously described project, we use evolutionary conservation to reveal constraints on the organization of *cis* elements across placental mammals. However, morphological change is often likely a result of changes to regulatory

networks (Carroll 2005; King and Wilson 1975). Such changes include loss of pre-existing elements (Mummidi et al. 2000), gain of novel elements (Komiyama et al. 2010; Mummidi et al. 2000), or modifications to conserved elements (Egli et al. 2009; Mummidi et al. 2000). Further understanding of the evolution of regulatory networks requires an evaluation of how *cis* elements have changed within a singular environment (such as a promoter), and how these changes impact the regulation of the affected gene. As noted above, this type of comprehensive determination of a regulatory network, even for a single gene, is a complex, laborious, and costly enterprise. Much like the enormous utility of model organisms based on the collective gain of independent investigations, choosing genes for which extensive knowledge has already been acquired helps facilitate the determination of a more complete picture of the regulatory network, and evolution thereof, for that target gene. For a second study on the evolution of regulatory networks, we chose two genes that have been studied extensively for over 50 years, including publications on their expression in varying tissues in primates (Goodman et al. 1969; Koen and Goodman 1969; Syner and Goodman 1966). Lactate dehydrogenase A (*LDHA*) and B (*LDHB*) are the targets of a more taxonomic-specific evaluation of the evolution of promoter sequences, detailed in chapter three. These genes are involved in metabolism, and their presence in diverse tissues across mammalian species, and beyond, have been thoroughly explored. The function of these genes, conserved and diverse expression profiles, and the potential impact of regulatory changes on primate evolution are discussed below.

Lactate dehydrogenase

Lactic acid was first characterized by Carl Wilhelm Scheele in 1780 (Robergs et al. 2004). Following its isolation, it was used as a cure-all, treating lupus, epidermal problems due to fungi, tuberculosis, nasal problems, as well as many others (Browning 1886). Lactic acid was first discovered from sour milk, from which it derives its name. Subsequent experiments identified this acid in muscle and blood, but the focus of research revolved around its prevalence following fermentation (Robergs et al. 2004). The “lactic acidosis” concept, suggests that the accumulation of lactic acid in muscles leads to acidosis and subsequent pain felt in these tissues following exercise (Robergs et al. 2004). Lactate accumulation is based on the conversion of pyruvate, the end byproduct of glycolysis, to lactate (Fig.1), a reaction catalyzed by the enzyme lactate dehydrogenase (LDH) (Baumberger et al. 1933; Dawson et al. 1964). This forward reaction is much more favored by the enzyme, but, as the name suggests, this is not the reaction from which the enzyme acquired its name. Marjory Stephenson (1928) first isolated the enzyme, from bacteria, with reasonable purity, and tested its ability to dehydrogenate lactate, thereby producing pyruvate (Stephenson 1928). The common name lactate dehydrogenase has been maintained, since all oxidoreductases, when appropriate, have been given the direction of dehydrogenase as the common name (Wain et al. 2002). The systematic name is lactate:NAD+ oxidoreductase (Wain et al. 2002).

Figure 1. Enzymatic reactions catalyzed by lactate dehydrogenase

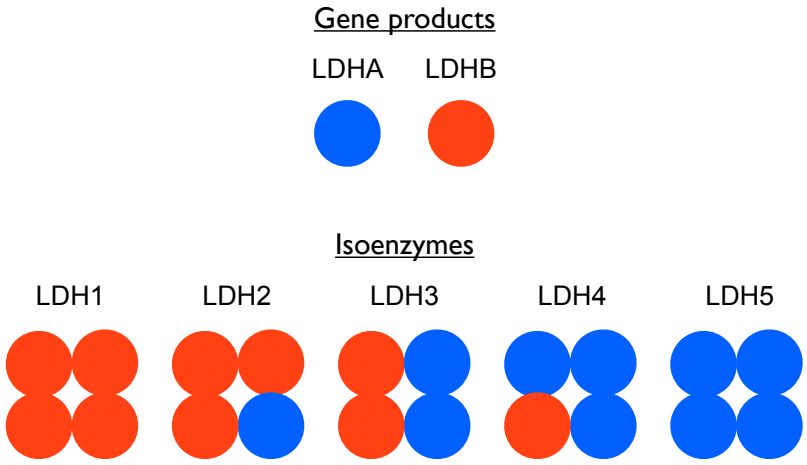


Size difference between arrows indicate the greater affinity for the forward reaction (pyruvate and NADH converted to lactate and NAD+).

LDH is a highly conserved enzyme that has been studied for nearly a century. It is present in eukaryotes, Eubacteria, and Archaea (Goodman et al. 1969; Klenk et al. 1997; Madern 2002; Stephenson 1928), demonstrating that organisms prior to the divergence of the three kingdoms of life also had gene(s) that encode this enzyme. A single copy *LDH* gene underwent a duplication event, resulting in the *LDHA* and *LDHB* genes that are common to all vertebrates. It remains unclear whether the duplication event took place in prochordates or jawless vertebrates (Li et al. 2002). Regardless, it has been proposed that the duplication likely occurred around 500 million years ago (Li et al. 2002). At the amino acid level, mammals share at least 89% and 80% sequence identity for LDHA and LDHB, respectively (Holmes and Goldberg 2009).

The LDH enzyme is a tetramer, composed of varying ratios of the two primary genes, *LDHA* and *LDHB* (Fig.2). The ratios are dictated by the abundance of each LDH protein, as the tetramer is known to assemble randomly (Markert 1963a). There are five different LDH isoenzymes, LDH1-LDH5 (Fig.2). An enzyme composed entirely of LDHA (4 LDHA) is called LDH5, whereas all LDHB (4 LDHB) is called LDH1, with LDH2 (3LDHB:1LDHA), LDH3 (2LDHB:2LDHA), and LDH4 (3LDHA:1LDHB) as heterotetramers (Koen and Goodman 1969; O'Brien et al. 2007).

Figure 2. LDH gene products and composition of LDH isoenzymes



The velocities of the forward and reverse reactions vary significantly across isoenzymes LDH1-LDH5 (Dawson et al. 1964; Nisselbaum et al. 1964; Vesell and Bearn 1961), with LDH4 and LDH5 rapidly converting pyruvate to lactate, and reducing NADH to NAD+ (Nisselbaum et al. 1964; Vesell and Bearn 1961). In contrast, LDH1-LDH3 are much slower at converting pyruvate to lactate

(Nisselbaum et al. 1964; Vesell and Bearn 1961), are inhibited by high levels of pyruvate (Bishop et al. 1972; Dawson et al. 1964; Latner et al. 1966), and have a greater affinity for the reverse reaction of converting lactate to pyruvate. In fact, LDH1-LDH3 require 3-5 times the lactate concentration than LDH4 and LDH5 in order to reach the same reaction velocity (Vesell and Bearn 1961). These differences in NADH oxidation have significant metabolic consequences (Greiner et al. 1994). The electrons from the coenzyme NADH enter the mitochondria via the glycerol-phosphate or malate-aspartate shuttle, and are essential for mitochondrial oxidation (Greiner et al. 1994). LDH outcompetes these shuttles in NADH oxidation in the cytosol (Greiner et al. 1994), and the higher the rate of oxidation, the less oxygen there is available for aerobic metabolism. Because different tissues have varying metabolic demands, much research has focused on the tissue distribution differences between these genes. An overview of mammalian LDH expression profiles in skeletal muscle, liver, and heart is shown in Table 1.

Euarchontoglires *LDHA* and *LDHB* Tissue Distributions

LDHA is also known as *LDH-M*, or muscle type, due to its prevalence in skeletal muscle, while *LDHB* was known as *LDH-H*, or heart type, due to its prevalence in heart muscle (Dawson et al. 1964; Syner and Goodman 1966). Tissue distribution analyses have been conducted on many different mammalian species, with striking similarities. Some of these studies evaluate the levels of LDHA and LDHB proteins, while others determine the presence of the LDH

isoenzymes, LDH1-LDH5 (Fig.2). Because the LDH enzyme assembles randomly, the presence of specific isoenzymes reflects the concentrations of the individual LDH proteins. In humans, LDH5 (all LDHA) is the primary isoenzyme in liver and skeletal muscle, with a minor presence of LDH4 and LDH3 (Latner and Skillen 1964; Vesell 1961; Vesell and Bearn 1961). LDH1 (all LDHB) is the primary isoenzyme in heart, with a minor presence of LDH2 and LDH3 (Latner and Skillen 1964; Vesell 1961; Vesell and Bearn 1961). The bonnet macaque, an Old World monkey, *Macaca radiata*, shows the same isoenzyme preference in heart for LDH1 and in skeletal muscle for LDH5 (Goodman et al. 1969). This pattern was found for other primate species, including a New World monkey *Saimiri sciureus* (Goodman et al. 1969; Koen and Goodman 1969), as well as three strepsirrhine species (Goodman et al. 1969; Koen and Goodman 1969). Within a sister clade to primates, Scandentia, *Tupaia glis* was found to share the same pattern (Goodman et al. 1969).

In the Order Rodentia, the same patterns emerge. In rat liver, LDH5 is the near-exclusive isoenzyme, as well as the primary isoenzyme in skeletal muscle, with significant portions of LDH4, and minimal amounts of LDH1-LDH3 (Beebee and Carty 1982). In heart, LDH1 is the primary isoenzyme, with significant amounts of LDH2 and minimal amounts of LDH3-LDH5 (Beebee and Carty 1982).

In rabbit, a lagomorph, LDH5 is almost exclusively the isoenzyme present in skeletal muscle, while LDH5 and LDH4 are the primary isoenzymes in liver (Nisselbaum and Bodansky 1959; Plagemann et al. 1960). LDH1 is the primary

isoenzyme in heart (Dawson et al. 1964; Nisselbaum and Bodansky 1959; Plagemann et al. 1960).

Table 1. Conserved relative abundance of *LDHA*:*LDHB* in varying tissues across mammals

Mammalian clade	**Skeletal Muscle**	**Liver**	**Heart**
Anthropoids	*LDHA* (Goodman et al. 1969; Latner and Skillen 1964)	*LDHA* (Koen and Goodman 1969; Latner and Skillen 1964)	*LDHB* (Goodman et al. 1969; Latner and Skillen 1964)
Strepsirrhines	*LDHA* (Goodman et al. 1969)	*LDHA* (Koen and Goodman 1969)	*LDHB* (Goodman et al. 1969)
Rodents	*LDHA* (Beebee and Carty 1982)	*LDHA* (Beebee and Carty 1982)	*LDHB* (Beebee and Carty 1982)
Lagomorphs	*LDHA* (Nisselbaum and Bodansky 1959; Plagemann et al. 1960)	*LDHA* (Nisselbaum and Bodansky 1959; Plagemann et al. 1960)	*LDHB* (Nisselbaum and Bodansky 1959; Plagemann et al. 1960)
Carnivores	*LDHA* (Milne and Doxey 1987)	*LDHA* (Milne and Doxey 1987)	*LDHB* (Milne and Doxey 1987)
Artiodactyls	*LDHA* (Charpentier and Goutefongea 1964; Hinks and Masters 1965)	*LDHA* (Hinks and Masters 1964)	*LDHB* (Hinks and Masters 1965)

Mammalian clade	Skeletal Muscle	Liver	Heart
Afrotherians	*LDHA* (Goodman et al. 1969)	*N/A*	*LDHB* (Goodman et al. 1969)

Listed mammalian clades are represented by species mentioned in the text. The relative expression profiles depict which gene, *LDHA* or *LDHB*, is expressed or found in greater abundance in the listed tissue. Citations are not exhaustive, see text for full list.

Laurasiatheria *LDHA* and *LDHB* Tissue Distributions

In pig, LDH5 is the primary isoenzyme in muscle (Charpentier and Goutefongea 1964), and skeletal muscle (Hinks and Masters 1965), while LDH1 is the primary isoenzyme in heart (Hinks and Masters 1965). In cow LDH5 is the primary isoenzyme in liver (Hinks and Masters 1964).

In dog, LDH5 is the primary isoenzyme in liver and skeletal muscle, with some LDH4 and LDH3 in liver, and some presence of LDH1-LDH4 in skeletal muscle (Milne and Doxey 1987). LDH1 is the primary isoenzyme in heart, with some LDH2 and LDH3 (Milne and Doxey 1987).

In the masked shrew, *Sorex cinereus*, LDH1 is the near-exclusive isoenzyme in heart, whereas LDH3-LDH5 are the primary isoenzymes in skeletal muscle (Goodman et al. 1969).

Afrotheria *LDHA* and *LDHB* Tissue Distributions

Finally, data from the afrotherian elephant shrew, *Elephantalus brachyrhynchus,* show a similar pattern, with primarily *LDHB* expressed in the heart and primarily *LDHA* expressed in skeletal muscle (Goodman et al. 1969).

LDH in the Brain

In contrast with the conserved expression profiles found in other tissues, the expression profile of these two genes in brain shows marked variation. Humans were found to have primarily LDHB in the brain (Lowenthal et al. 1961; Nisselbaum and Bodansky 1961) and these findings were extended to anthropoid primates (Goodman et al. 1969; Koen and Goodman 1969; Syner and Goodman 1966). Tarsiers were found to have near equal levels of LDHA and LDHB in the brain (Goodman et al. 1969), and strepsirrhine primates displayed primarily LDHA in the brain (Goodman et al. 1969). Additional studies demonstrated that LDH1 is the only isoenzyme present in human neurons, and that detection of other isoenzymes in human brain is due to the presence of astrocytes and synaptosols (Bittar et al. 1996). In rodents, rats show primarily LDHA in the brain, with almost no presence of LDH1 in neurons (O'Brien et al. 2007). In horses, LDHA transcript levels were highest in the brain (Echigoya et al. 2009). In rabbit, all five isoenzymes were detected, with enrichment of LDH1-3 (Plagemann et al. 1960), although cell-type specific assays revealed that while LDHB is the primary isoenzyme in neurons, there is also an abundance of LDHA (Gorelikov and Savel'ev 2008). In sheep, all five isoenzymes were detected, with greater presence of LDHB (Lowenthal et al. 1961).

Encephalization and Energetics

There are significant differences in brain sizes throughout mammals. The encephalization quotient (EQ) is a measurement used to distinguish brain sizes across groups of animals of varying sizes, and is measured as a ratio of brain mass or volume to body surface area (Jerison 1977). Rather than measuring body surface area, however, this value is calculated by transforming body mass or volume, often by placing the value to the power of 2/3 (Jerison 1977). The resulting measurements identify certain mammals as exceptionally encephalized relative to mammals as a whole, including haplorrhine primates (tarsiers+anthropoids) (Joffe and Dunbar 1998), Cetacea (dolphins+whales; (Marino et al. 2003), and elephants (Shoshani et al. 2006). Mammalian brain expansion has occurred primarily in the neocortex (Finlay and Darlington 1995). Since the brain is composed of varying cell types, such as neurons, astrocytes, microglia, and others, an expansion of the tissue as a whole could be the result of the increase of specific cell types, specific brain regions, or both (Sherwood et al. 2006). Recent studies, however, have determined that the glia:neuron ratios in anthropoid primates are similar, although humans, and to a lesser extent chimpanzees, have a higher ratio than do macaques (Sherwood et al. 2006). Based on this finding, the increase in brain size is an expansion of neocortical volume, rather than a specific increase in cell type (Sherwood et al. 2006). In addition, the higher ratio of neurons in macaques supports the argument of gained *LDHB* expression in neurons, since Old World monkeys were found to

have slightly higher *LDHB* expression than hominids in the brain (Goodman et al. 1969).

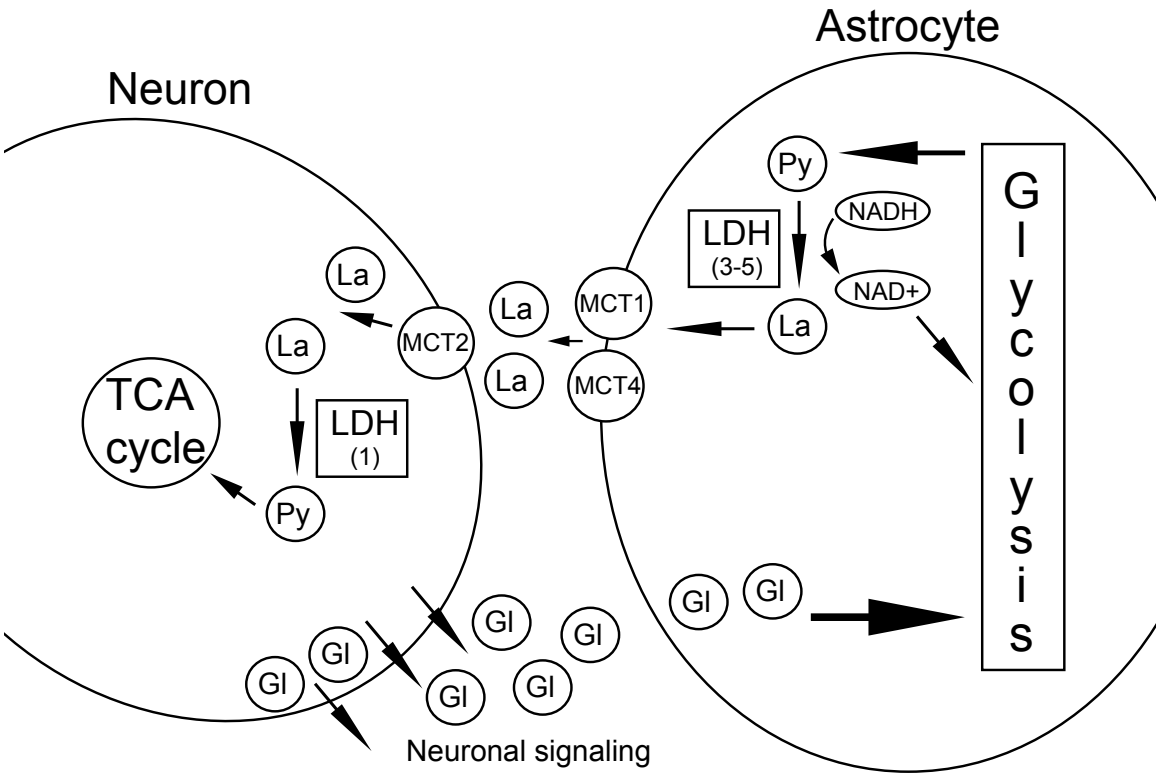
The brain is an expensive tissue, in terms of energetic demands (Aiello and Wheeler 1995). Brain tissue in humans functions at nine times the metabolic rate of the body as a whole (Aiello and Wheeler 1995), and this metabolic rate is comparable to the rates of other mammals (Aiello and Wheeler 1995). The notable increase in brain size in anthropoid primates has led to the hypothesis that such an increase would require a significant increase in energetic demands (Aiello and Wheeler 1995; Syner and Goodman 1966). Two general mechanisms could be involved in supplying the increased energy for an encephalized brain; either total body metabolism has also increased disproportionate to body size such that it could maintain the increase of energetic costs or a change in the allocation of total body metabolism could provide a greater percentage of total body metabolism to the brain (Armstrong 1983; Mink et al. 1981). In addressing the former, it has been shown that anthropoid primates do not have an increase in basal metabolic rate (BMR) or total body metabolism beyond that predicted by their size relative to other mammals (Armstrong 1983). The second possibility, increased energy allocation, has been demonstrated in anthropoid primates, with a greater percentage of total body metabolism dedicated to the brain (10-20%) than the vertebrate mean of 5.3% (Aiello and Wheeler 1995; Mink et al. 1981). Therefore, to compensate for the increased demands associated with encephalization, there has been a concomitant increase in the allocation of resources to the anthropoid brain (Armstrong 1983).

Brain energetics and lactate

The significance of *LDHA* and *LDHB* expression in different mammalian brains cannot be simplified as indicators of anaerobic versus aerobic metabolism, respectively. There has been demonstrated a complex interplay between neurons and astrocytes, as well as the use of aerobic metabolism and aerobic glycolysis (glyocolysis in the presence of sufficient oxygen). Neurons have the greatest energetic demands of the cells in the brain (Hyder et al. 2006), and their greatest requirement comes following neuronal signaling (Hyder et al. 2006). Signaling involves the release of glutamate, which stimulates the release of lactate from neighboring astrocytes (Pellerin and Magistretti 1994). The transport of lactate into and out of cells is carried out through monocarboxylate transporters (MCTs; (Pierre and Pellerin 2005). Astrocytes express primarily MCT1 (Pierre et al. 2000) and MCT4 (Bergersen et al. 2002), both of which have low affinity for lactate (Hertz and Dienel 2005). Neurons primarily express MCT2 (Bergersen et al. 2002; Pierre et al. 2000) which has significantly higher affinity for lactate, requiring less than $1/5^{th}$ available lactate as the other transporters to reach the same reaction velocity (Hertz and Dienel 2005). As a result, neurons preferentially uptake the lactate released from neighboring astrocytes. This shuttle of lactate upon stimulation by glutamate has been termed the astrocyte-neuron lactate shuttle hypothesis (ANLSH) and is depicted in Figure 3 (Pellerin and Magistretti 1994).

LDH plays a critical part in this shuttle in both astrocytes and neurons. The production of lactate in astrocytes is carried out through aerobic glycolysis, which is glycolysis despite the presence of sufficient oxygen for aerobic metabolism (Vander Heiden et al. 2009). The end byproduct of glycolysis, pyruvate, is converted to lactate by LDH prior to export (Pellerin et al. 2007). Upon uptake by neurons, lactate is converted back to pyruvate by LDH, and then enters the tricarboxylic acid (TCA) cycle (Pellerin et al. 2007). The differences in substrate affinity and turnover between LDH isoenzymes suggests that changes in the expression of *LDHA* and *LDHB* in neurons and/or astrocytes could contribute to changes in metabolic output. A greater shift towards lactate as a post-signaling fuel for neurons is one potential mechanism by which an increase in the allocation of energetic resources could have occurred during primate evolution.

Figure 3. Astrocyte-neuron lactate shuttle.



Gl, glutamate; Py, pyruvate; La, lactate; MCT, monocarboxylate transporters. Glutamate released upon neuronal signaling enters neighboring astrocytes and is converted to pyruvate via glycolysis. This pyruvate is converted to lactate via LDH, where isoenzymes responsible for this conversion in human are depicted in parentheses. This reaction produces NAD+ that maintains glycolysis in astrocytes. The lactate in astrocytes is shuttled out by MCT1 and MCT4, and taken up by neurons by MCT2. The lactate is converted back to pyruvate via LDH1, and the pyruvate then enters the tricarboxylic acid (TCA) cycle.

Regulation of *LDHA* and *LDHB*

The conservation of tissue-specific expression profiles for *LDHA* and *LDHB*, as well as the change in expression patterns in the brain, suggest that the *LDHA* and *LDHB* promoters harbor conserved and diverse *cis*-regulatory elements involved in coordinating these expression profiles. Multiple mechanisms are known to regulate expression levels for these genes, including

CpG methylation (Alcivar et al. 1991; Leiblich et al. 2006; Maekawa et al. 2002; Maekawa et al. 2003), a distal repressor (Chung et al. 1995), and transcript half-life (Jungmann and Kiryukhina 2005; Tian et al. 1998). However, extensive research has identified specific *cis*-regulatory elements in the *LDHA* promoter that are critical for expression (Jungmann et al. 1998; Semenza et al. 1996; Shim et al. 1997; Short et al. 1994). The promoter of *LDHB* has received little attention, although hypermethylation has been shown to eliminate transcription (Leiblich et al. 2006; Maekawa et al. 2003).

We hypothesized that within these gene promoters exist conserved elements that are primarily responsible for the tissue-specific expression profiles common amongst mammals. In addition, the derived expression profiles that emerged during primate evolution could be a result of *cis* element changes in the promoters of these genes. Chapter three is a study on the *LDHA* and *LDHB* gene promoters. We identify *cis* elements conserved across mammals, many of which have been previously characterized as critical for proper endogenous expression. We also describe elements gained during primate evolution, focusing on those elements gained during stem anthropoid evolution. These elements include targets that could inhibit *LDHA* expression in neurons, as well as gain neuronal *LDHB* expression, thereby contributing to the expression profile shifts of these genes in the brain during stem anthropoid evolution.

In addition to characterizing *cis* elements in these gene promoters, we determine the evolution of epigenetic modifications to conserved CpG sites, hypothesizing that *cis*-regulatory evolution is more dynamic than strictly

nucleotide substitutions to TFBS. We discover that methylation patterns are common between human and dwarf lemur (a strepsirrhine), but that the dwarf lemur displays greater levels of conserved CpG methylation. These results suggest differential use of cytosine methylation in the regulation of *LDHA* and *LDHB* between human and the dwarf lemur.

***LDHA* and *LDHB* structural modifications**

An additional mechanism by which LDH could affect the allocation of resources in the anthropoid brain is through changes in the LDHA and LDHB proteins. We hypothesized that protein-coding changes in *LDHA*, *LDHB*, or both during primate evolution, primarily during stem anthropoid evolution, occurred concomitant with the expression profile changes. Such changes could further alter the properties of the LDH enzyme, through more rapid production of lactate (for release from astrocytes), a greater rate of production of pyruvate (for use in neurons), or changes in the assembly of the LDH tetremer. We analyzed the coding regions for *LDHA* and *LDHB* in 16 primate species and three non-primate mammals. Surprisingly, we found very little change in these genes during primate evolution, with the exception of a significant increase in amino acid replacements on the lineage leading to the mouse lemur, *Microcebus murinus*. This species is known to enter a dormant state during winter, termed torpor, during which they greatly reduce energy output and metabolism (Giroud et al. 2010). We discuss the implications of our findings in the context of the unique metabolic requirements of the mouse lemur in chapter four.

The fifth chapter involves a study on tissue-specific expression profiles in primates. We discovered expression of growth hormone genes in the placenta of a Spider monkey. The results of this study were published in *The Proceedings of the National Academy of Sciences* in 2009 (Papper et al. 2009). Briefly, this study determined that New World monkeys express growth hormone genes in the placenta, similar to expression profiles found in anthropoids, despite the expansion of the single copy growth hormone gene independently in these two clades. The primary findings in this study are that placental expression of growth hormone genes likely predates the divergence of catarrhines and platyrrhines, and the history of these genes in prenatal growth and development are likely to be much more ancient than previously thought. While this project was not part of my dissertation proposal, the study required a significant amount of time and effort and is a contribution to scientific knowledge.

CHAPTER TWO—TRAVERSING MAMMALIAN PROMOTER LANDSCAPES: SIZE, STRUCTURE, AND COMPOSITION OF CONSERVED ELEMENTS

INTRODUCTION

A major advancement in deciphering the messages in DNA came in the 1960's, following decades of research, when Nirenberg, Khorana, and others, cracked the genetic code (Crick et al. 1957; Nirenberg and Matthaei 1961; Nishimura et al. 1965). At the same time, genes were found to be regulated at the level of transcription (Jacob and Monod 1961), and the promoter was identified as a primary region responsible for this regulation (Eron and Block 1971; Ippen et al. 1968). Within the promoter, specific sequences were discovered critical for initiating transcription, including the TATA box (Lifton et al. 1978), CCAAT box (Dierks et al. 1983), GC box (Gidoni et al. 1984), and others. The search for regulatory elements has led to the development of numerous *in vivo, in vitro*, and *in silico* approaches (Cleary et al. 1972; Harbison et al. 2004; Solomon et al. 1988; Tagle et al. 1988).

Despite extensive research focusing on gene regulation, very few universal rules or characteristics have been described for the regulatory motifs involved in transcription. A gene promoter provides a reasonable starting point for asking general questions about regulatory features. The promoter can be considered a molecular landscape, in which functional and/or structural elements reside. This landscape differs from genic, intronic, or other chromosomal landscapes (Maston et al. 2006). If universal characteristics exist in promoters, they would be evolutionarily conserved, and comparative genomics techniques

are able to detect conserved sequences within variable DNA stretches. The presence of identical nucleotide strings across species can be explained as conserved regulatory motifs (Tagle et al. 1988), alternative functional elements (Margulies and Birney 2008), topographical features (Parker et al. 2009), segments of small, non-coding RNAs (Zhang 2009), while some non-functional sites will be conserved due to random chance. Detection and characterization of these conserved elements produces a layout of the size, structure, and composition of nucleotide strings in comparable molecular landscapes.

In this study, we identify over 200,000 conserved elements in the promoter regions of more than 9,000 genes across seven mammalian species, and these motifs have 100% sequence identity across all of these species. We characterize the molecular landscape of mammalian promoters and find they have elevated G+C content common to a small portion of the genome as a whole. CG and TA are the least frequent dinucleotides whereas CA and TG dinucleotides are often the most prevalent, string biases also found within coding regions. We have assembled a database of putative functional motifs within mammalian gene promoters. By studying the features of these motifs, we discover that base compositions change with motif width, and that evolutionary conservation favors specific nucleotide strings over others. Individual base frequencies at specific motif positions suggest conserved structural arrangements. Frequencies of most nucleotide strings on the sense and antisense strands are not statistically distinct, suggesting that the strands exert near equal selective strength. Finally, we identify the promoters with the most abundant conserved elements in

mammalian genomes, and find highly significant functional relationships between the genes under control of these promoters.

For the purposes of the current study, we defined “conserved motifs” as those sharing 100% identity across all species included in the study. While we recognize that such stringency eliminates many similar, yet not identical functional elements, this study aims to help characterize the basic framework of mammalian promoter elements rather than identify all of them.

RESULTS

Nucleotide Composition

We extracted 1200bp of upstream gene flanking region for 10,056 orthologous Ensembl genes from human, chimpanzee, mouse, rat, cow, dog, and opossum. We considered this region as the putative promoter, and the assembly of the seven putative promoters for an individual gene is referred to as an orthologous set. Motif discovery from these orthologous sets found 222,275 conserved motifs across eight motif widths, ranging from 5-12 nucleotides. The average number of conserved motifs sharply declines with motif width, from nearly 16 conserved 5mers/orthologous set to 0.07 12mers/orthologous set (Table 1). The G+C content across all 10,056 orthologous sets is 50.34%, with an AT content of 49.66% (Appendix A1). Within conserved motifs, adenosine and thymine content have a positive correlation with motif width, from 23.5% and 22.9% respectively in motifs of width 5 to 25.7% and 27.4% respectively in motifs of width 12 (Appendices A1 and A2). In contrast, cytosine and guanine content

have a negative correlation with motif width, from 26.8% for both in motifs of width 5 to 23.6% and 23.2% respectively in motifs of width 12 (Appendices A1 and A2).

Table 1. Prevalence of conserved motifs of variable widths

Total # of orthologous sets = 9,398	# of sets* with motifs of listed width	Total # of motifs of listed width	# of motifs/set* of listed width	Total # of unique motifs of listed width
5mers	9380	149297	15.89	919
6mers	9058	55008	5.85	3267
7mers	3806	9699	1.03	4570
8mers	1403	3422	0.36	2806
9mers	819	1911	0.20	1768
10mers	565	1298	0.14	1274
11mers	435	937	0.10	936
12mers	340	703	0.07	703

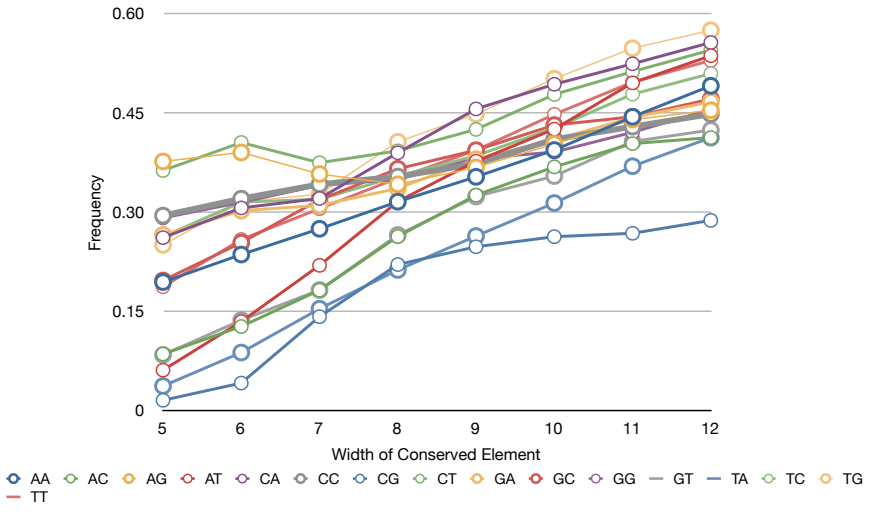
*set refers to orthologous gene set

Dinucleotide Composition

From a total of 149,297 conserved 5mers, we recovered only 919 unique motifs of the 1,024 possible combinations of five nucleotides (i.e. 4^5; Table 1). In evaluating the 105 missing motifs, we found that 100 contained CG dinucleotides. In addition to 105 missing unique 5mers, there are 18 orthologous sets that lack any conserved 5mers. We found that these sets contain high N content in at least one species, and often contain long dinucleotide repeats.

Previous studies identified a bias against CG and TA dinucleotides in coding regions, and a bias in favor of CA and TG (Ohno 1988). With the role of methylation on transcription initiation, we tested whether such biases persisted in the conserved motifs, and whether these biases changed with the width of the motif. With the frequency of each dinucleotide within the pool of conserved motifs for each width, we find a positive correlation between dinucleotide frequency and motif width for all 16 dinucleotides, although the slopes of the increases vary (Fig.1 and Appendix A3). As seen in coding regions, CG and TA dinucleotides are the least frequent across all widths (Fig.1), whereas CA and TG are the most common in widths 9-12, and are two of the top three dinucleotides within motifs of width 8 (Fig.1).

Figure 1. Dinucleotide frequencies for each conserved motif width



Dinucleotides represented by colors as shown below graph. X-axis indicates element widths (5-12) and y-axis indicates frequency.

Trinucleotide Composition

Previous research evaluating nucleotide neighbors surrounding methylated CpG sites in the human genome identified that the CG dinucleotide is most often found neighbored by cytosines and guanines, both preceding and following the dinucleotide (Clay et al. 1995). Using a genomewide approach, we tested whether such neighboring biases exist within the conserved motifs by evaluating nucleotide frequencies preceding and following all possible dinucleotides, such as NAA and AAN. This analysis includes 32 dinucleotide neighbors based on the 16 possible dinucleotides and a variable position preceding and following each dinucleotide. We find that only four out of the 32 dinucleotide neighbors are not statistically distinct from the expected (chi-square test with $p > 0.10$ for all four; Appendix A4). These four (NAA, TTN, CAN, NTG) include two reverse complement pairs. Surprisingly, in the case of NAA and TTN, in which the N can create a TA dinucleotide within the trinucleotide (Appendix A4), this does not deviate the frequency from the expected, demonstrating that while TA is selected against in most trinucleotide settings (Appendix A4), it is stable following a T or before an A.

With 28 dinucleotides plus a neighbor statistically distinct from the expected, we looked into the neighboring nucleotides that generated this statistical difference. Table 2 shows the 32 individual nucleotides, and their dinucleotide neighbor, whose contributions to their trinucleotide chi-square value

(Table S4) was sufficient for making the group as a whole statistically significant (chi-square > 7.82). 17 of these biases include greater than expected frequencies, 11 of which generate a dinucleotide repeat (e.g. AT-**T** or **G**-GC; Table 2). Adenines precede and follow adenines while thymines precede and follow thymines, so long as the other nucleotide is its complement (**A**-AT/TA-**A** or **T**-TA/AT-**T**, respectively; Table 2). Guanines precede and follow guanines if the other nucleotide is either cytosine or adenine (**G**-GC/CG-**G** or **G**-GA/AG-**G**; Table 2). Cytosines don't fit this pattern as well, preceding and following cytosines if the other nucleotide is a thymine (**C**-CT or TC-**C**) or following a cytosine if the other nucleotide is a guanine (GC-**C**; Table 2). Of the six remaining high frequency biases, there are two reverse complement pairs, **T**-GT/AC-**A** and CT-**G**/**C**-AG (Table 2). Those four, in addition to the remaining two (TC-**A** and **C**-AC), all result in the generation of either CA or TG, the two most frequent dinucleotides (Table 2).

15 of the 32 neighboring biases include less than expected frequencies, and 10 of these generate either a CG or TA dinucleotide, the least frequent dinucleotides in the dataset (Table 2). In generating the CG dinucleotide, a cytosine preceding a guanine is infrequent when the guanine is neighbored by either an adenine or thymine (**C**-GA or **C**-GT; Table 2). Likewise, a guanine following a cytosine is infrequent when the cytosine is neighbored by an adenine or thymine (AC-**G** or TC-**G**) but also if neighbored by another cytosine (CC-**G**; Table 2). A similar pattern is observed for the generation of the TA dinucleotide. A thymine preceding an adenine is infrequent if the adenine is neighbored by a

cytosine (**T**-AC), a guanine (**T**-AG), or a thymine (**T**-AT). An adenine following a thymine is infrequent if the thymine is neighbored by a cytosine (CT-**A**) or a guanine (GT-**A**; Table 2).

Of the remaining five low frequency neighbors, three involve neighbors of CG or TA, with adenine infrequently following the CG dinucleotide (CG-**A**; Table 2) and guanine or cytosine infrequently following the TA dinucleotide (TA-**C** or TA-**G**; Table 2). The remaining two show a low frequency for cytosine following the AT dinucleotide (AT-**C**) and guanine preceding the TT dinucleotide (**G**-TT; Table 2).

Table 2. Highest and lowest nucleotide-dinucleotide neighbors

Highest frequency nucleotide-dinucleotide neighbors	Lowest frequency nucleotide-dinucleotide neighbors
TA-**A** (42.4%; $p < 0.001$)	**C-**GA (6.8%; $p < 0.001$)
T-TA (39.8%; $p < 0.001$)	TC-**G** (9.9%; $p < 0.001$)
G-GA (38.8%; $p < 0.001$)	CG-**A** (11.2%; $p < 0.001$)
T-GT (37.7%; $p < 0.010$)	AC-**G** (13.2%; $p < 0.005$)
TC-**A** (36.1%; $p < 0.001$)	**T-**AG (13.3%; $p < 0.001$)
AT-**T** (35.9%; $p < 0.005$)	CT-**A** (13.6%; $p < 0.001$)
CG-**G** (35.1%; $p < 0.010$)	**C-**GT (13.7%; $p < 0.005$)
AC-**A** (33.9%; $p < 0.050$)	TA-**G** (15.0%; $p < 0.010$)
C-AC (32.9%; $p < 0.025$)	**T-**AC (15.3%; $p < 0.001$)
TC-**C** (32.4%; $p < 0.005$)	CC-**G** (15.7%; $p < 0.025$)
C-AG (32.2%; $p < 0.025$)	GT-**A** (16.6%; $p < 0.025$)
G-GC (31.7%; $p < 0.025$)	TA-**C** (16.8%; $p < 0.050$)
C-CT (31.7%; $p < 0.010$)	AT-**C** (17.6%; $p < 0.050$)
A-AT(31.6%; $p < 0.025$)	**G-**TT (17.7%; $p < 0.025$)
CT-**G** (31.3%; $p < 0.010$)	**T-**AT (19.1%; $p < 0.025$)
AG-**G** (31.1%; $p < 0.050$)	
GC-**C** (30.5%; $p < 0.050$)	

Bold letters are the variable position neighboring the tested dinucleotide. Frequency of neighbor is shown in parentheses, followed by p-value associated with the specified nucleotides deviation from expected. Degrees of freedom = 3.

Positional Effects

Describing the nucleotide frequencies at each position for each motif width presents certain difficulties. Although the motifs within each width are non-overlapping, between motif widths there are an unknown number of overlapping motifs. Consequently, nucleotide frequencies at the first position of motifs of width 5 may not always correspond to the first position of motifs of width 6. This likely reduces the signals of nucleotide biases at individual positions when considering the dataset as a whole.

Only one nucleotide bias is found consistent across all motif widths, with adenosine the least frequent base at position two (Appendix A5). The range of frequency for this nucleotide across all motif widths (20.0-22.7%) does not overlap the frequency ranges of any of the remaining three nucleotides (24.3-29.2%; Appendix A5). No other positions from one through nine have a nucleotide frequency that is the minimum or maximum across all widths, or a frequency range that does not overlap the frequency range of other nucleotides for that position (Appendix A5). We don't consider positions 10-12, because these positions are shared by so few widths. We did, however, consider the last position across all widths, but found no consistent nucleotide bias (Appendices A5 and A6).

Making Sense of Strands

We hypothesized that there are selective differences between the sense and antisense strands of mammalian promoters. We evaluated the frequencies of reverse complement trinucleotides, using the Wilcoxon Signed-Rank Test to identify statistical differences between reverse complement pairs (Appendix A7). Of the 32 reverse complement pairs, only six are statistically different in prevalence across all motif widths. GAT is more common than ATC and CAT is more common than ATG. CGG is more common than CCG while TCG is more common than CGA. CTA is more common than TAG and ATT is more common than AAT (Appendix A7).

Highly Conserved Promoters

We evaluated the orthologous sets for each motif width independently, characterizing the gene promoters with the most abundant conserved elements. We find that not all widths are well suited to answer this question. For instance, there are over 800 orthologous sets with over 22 conserved 5mers. In contrast, there are only 51 orthologous sets with four or more conserved 12mers. Due to these differences between motif widths, we chose to include only those motif widths for which the average number of motifs/orthologous set is less than 1 (widths 8-12, Table 1). We sought the gene promoters from each width that contained the most conserved motifs, setting the threshold at no fewer than 50 gene promoters per width. Using this criterion, we found 67 gene promoters with four or more 11mers, 64 promoters with 5 or more 10mers, 69 promoters with 6 or more 9mers, and 71 promoters with 8 or more 8mers. We eliminated

redundancy between datasets, resulting in a total of 93 unique gene promoters (Appendix A8). We then looked at these gene promoters to determine whether they overlapped with neighboring gene UTRs, exons, or introns, resulting in potential false positives. This approach eliminated 31 gene promoters, resulting in a final list of 60 gene sets (Table 3).

Table 3. List of the 60 genes with the most abundant highly conserved promoter elements

Gene	Human Ensembl ID	Criteria	Gene	Human Ensembl ID	Criteria
PHOX2B	ENSG00000109132	13 12mers	RAB11A	ENSG00000103769	4 12mers
HOXA2	ENSG00000105996	10 12mers	LIN28A	ENSG00000131914	4 12mers
EVX2	ENSG00000174279	10 12mers	B3GNT1	ENSG00000174684	4 12mers
HOXB1	ENSG00000120094	8 12mers	HOXB7	ENSG00000120087	4 12mers
DUSP6	ENSG00000139318	8 12mers	TBX19	ENSG00000143178	4 12mers
ESM1	ENSG00000164283	8 12mers	NFKBIA	ENSG00000100906	5 11mers
CDX2	ENSG00000165556	7 12mers	SPI1	ENSG00000066336	5 11mers
HOXD10	ENSG00000128710	6 12mers	VGF	ENSG00000128564	5 11mers
HOXB9	ENSG00000170689	6 12mers	MSTN	ENSG00000138379	5 11mers
OVOL2	ENSG00000125850	6 12mers	PREX2	ENSG00000046889	4 11mers
SMAD6	ENSG00000137834	6 12mers	FBXO32	ENSG00000156804	4 11mers
FGF6	ENSG00000111241	5 12mers	SOCS7	ENSG00000174111	4 11mers
DDIT4	ENSG00000168209	5 12mers	BAZ1B	ENSG00000009954	4 11mers
KCTD5	ENSG00000167977	5 12mers	SLC7A11	ENSG00000151012	4 11mers
ID3	ENSG00000117318	5 12mers	ACCN4	ENSG00000072182	4 11mers
DSG4	ENSG00000175065	5 12mers	NIPBL	ENSG00000164190	5 10mers
VSX2	ENSG00000119614	5 12mers	PI15	ENSG00000137558	5 10mers
MAF	ENSG00000178573	5 12mers	EDN1	ENSG00000078401	5 10mers
PLXNC1	ENSG00000136040	5 12mers	ARHGEF6	ENSG00000129675	5 10mers
C7ORF55	ENSG00000164898	5 12mers	CYR61	ENSG00000142871	5 10mers
TBR1	ENSG00000136535	5 12mers	BSX	ENSG00000188909	8 9mers
CTGF	ENSG00000118523	5 12mers	C1QL1	ENSG00000165985	6 9mers
MNT	ENSG00000070444	5 12mers	SHC4	ENSG00000185634	6 9mers
PRMT5	ENSG00000100462	5 12mers	ATOH7	ENSG00000179774	6 9mers
KDM6A	ENSG00000147050	4 12mers	EMILIN1	ENSG00000138080	10 8mers
HOXA5	ENSG00000106004	4 12mers	FOXA1	ENSG00000129514	9 8mers
DBX1	ENSG00000109851	4 12mers	COLQ	ENSG00000206561	8 8mers
ITGBL1	ENSG00000198542	4 12mers	NR0B2	ENSG00000131910	8 8mers
LOX	ENSG00000113083	4 12mers	IRX5	ENSG00000176842	8 8mers
DRGX	ENSG00000165606	4 12mers	ATP2A2	ENSG00000174437	8 8mers

Gene symbols and Ensembl IDs based on human.

Functional annotation clustering identified five clusters under the highest classification stringency for the 60 gene sets with enrichment scores greater than 2 (Table 4), nine clusters with enrichment scores greater than 1 (Appendix A9). These clusters include regulation of metabolic processes (enrichment score = 8.46), DNA-dependent transcription and RNA metabolic process (enrichment score = 7.93), regulation of transcription (enrichment score = 7.53), homeobox genes (enrichment score = 7.41), and insulin-like growth factor binding/growth factor binding (enrichment score = 2.45) (Dennis et al. 2003; Huang da et al. 2009).

DISCUSSION

In this study, we utilized comparative genomics and discovered 222,275 conserved elements in the promoters of over 9,000 genes in seven mammalian species. We hypothesized that the composition and organization of these elements is evolutionarily constrained, and these constraints would produce biases that could be detected through a genomewide analysis. Briefly, there are four principal findings in this study. Mammalian promoters maintain a high G+C content relative to the genome as a whole. Numerous nucleotide string biases are observed, with CG and TA infrequent dinucleotides in conserved elements while CA and TG are the most frequent dinucleotides, characteristics common to coding regions. Surprisingly, the sense and antisense strands appear to have the same selective strength. In addition, genes with the greatest abundance of

conserved elements are primarily transcription factors and growth factor binding proteins.

Although transcriptional regulation has been studied for decades, little description has focused on the framework and evolution of promoter landscapes. Nucleotide composition, including base frequencies, string biases, sense and antisense differences, and evolutionary conservation define the molecular features of these landscapes. Novel elements that are putatively involved in gene regulation are discovered with great frequency (Georgiev et al. 2010; Xie et al. 2009), so much so that the number of transcriptional variables apparently approaches infinity. Without a basic understanding of the nucleotide organization of these molecular regions, each gene is typically approached as a completely unique transcriptional unit (Laflamme et al. 2010; Srivastava et al. 2010). In this study we utilized evolutionary genomic techniques to better understand the conserved nucleotide features of the mammalian promoter.

Table 4. DAVID results identifying over-represented functional categories (enrichment score > 2)

Annotation Cluster	Enrichment Score	Number of Genes	Genes
Regulation of metabolic process	8.46	31	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EDN1 EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT MSTN NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2

Annotation Cluster	Enrichment Score	Number of Genes	Genes
Regulation of transcription, DNA-dependent	7.93	29	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2
Regulation of transcription	7.53	30	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT MSTN NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2
Homeodomain-related	7.41	10	CDX2 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 IRX5 PHOX2B VSX2
Insulin-like growth factor binding	2.45	3	CTGF CYR61 ESM1

Mammalian genomes have an overall G+C content ranging from ~38% (opossum) to ~42% (mouse), comprised of five different isochores, mega-base stretches of varying G+C content (Bernardi 1993; Costantini et al. 2009). Isochores L1, L2, H1, and H2 include stretches of G+C content below 50%. Only isochores H2 and H3 include regions of DNA with G+C content above 50%, and account for only 15% of the total genomic sequence (Costantini et al. 2009). We found that the mammalian promoter has a G+C content of 50.34%, placing it within the G+C content range of isochore H2, regions that make up only ~11% of the genome as a whole (Costantini et al. 2009). We can thus propose that natural selection has acted to preserve a relatively high G+C content in mammalian

promoters. Although the identification of genomic isochores is based on sequential genomic stretches, our study includes over 11 megabases of promoter sequence for each species, and consequently may also be considered as a genomic feature analogous to isochores.

Elevated G+C content is found despite the CG dinucleotide being the least frequently conserved dinucleotide, regardless of motif width (Fig. 1). This may be compensated for, in part, by the patterns of sequence conservation of CG neighbors. Previous research has shown that CG dinucleotides are often found within stretches of relatively high G+C content (Clay et al. 1995). Here we demonstrate that CG dinucleotides are not just present within such domains, but these domains are maintained through mammalian evolution, suggesting that G/C boundaries either function for conservation of this dinucleotide, or are selected for in conjunction with CG dinucleotides. While these results do not speak to the functional consequences of these dinucleotides within conserved elements, they improve our ability to predict their biological significance when coupled with broader genomic features such as CpG islands and methylation states (Hsieh et al. 2009).

The most common dinucleotides (CA and TG) agree with similar findings that were based on coding regions (Ohno 1988). One explanation for this is that CG dinucleotides, upon methylation, can be converted to CA or TG (Coulondre et al. 1978; Wang et al. 1998). Since methylation can be prevented through steric hindrance, conservation of motifs containing CG dinucleotides would strongly

suggest the binding of factors that help prevent methylation, rather than conservation of methylation patterns.

An alternative explanation is that CA and TG dinucleotides are locally advantageous in functional regions due to their structural bistability (Kato 1999). These dinucleotides, in response to HMG proteins, have been shown to induce localized folding (Churchill et al. 1995), and are biochemically the most flexible dinucleotides (Travers 2004). Since the inferred functions of CA/TG dinucleotides differ between coding and promoter regions, their abundance in both may be the result of structural advantages gained due to the biochemical properties that make these dinucleotides more open and available for binding trans-acting factors (Ross et al. 2001; Travers 2004).

Nucleotide neighbor analyses reveal not only that the CG dinucleotides often reside within C/G borders, but that the TA dinucleotides often reside within A/T borders. In fact, dinucleotide repeats (AA, GG, TT) are favored by A, G, and T when they are bordered by their complementary base (Table 2). This pattern would tend to create A+T rich regions and C+G rich regions, except that the dinucleotide cytosine repeat favors a thymine neighbor (Table 2). This deviation from the other three bases may be essential for maintaining a heterogenous population of bases within short stretches in the promoter region. The other neighbor biases show composition preferences across conserved motifs and allow us to distinguish between those motifs composed of frequently occurring neighboring sequences from those with unlikely neighbors. Since some transcription factor binding sites (TFBS) have shown variability at specific

positions within the binding site, whereas others have not, we would predict that the less likely the compositional arrangement within a given TFBS, the greater the contribution of those positions to the function of that motif.

The frequencies of all dinucleotides increase as the width of the conserved motifs increases, with all dinucleotides except CG present in over 40% of the conserved 12mers (Fig. 1). This observation suggests that longer sequence conservation favors more diverse motifs. Moreover, our analyses identified zero 12mers composed entirely of one nucleotide. This finding suggests that long strings of the same base are either selected against in mammalian promoters, or at the very least such strings lack a functional role and therefore are unconstrained by natural selection.

Across all motif widths, adenine is the least frequent nucleotide at position 2. This was the only signal we detected across all motif widths. Although some characteristics were found common across a subset of motif widths (for example, guanine was the most frequent base at position 1 across widths 8-12), the overlap between motifs of different widths no doubt reduced signals of motif organization. A motif of width 5 may be found in a motif of width 10, but the corresponding positions are likely to be different. Future work on motif sequence construction would need to reconcile position issues across motifs of different widths.

Of 32 trinucleotide reverse complement pairs, only six show significant differences in abundance across conserved elements. Three of the six are underrepresented in the data set as a whole since they contain either CG or TA

(Appendix A6). This suggests that there is very little difference in selective pressure between the sense and antisense strands in the promoter. This is surprising, since we would assume that the transcriptional machinery would be directed differently to the antisense strand, resulting in a composition bias between the sense and antisense strands. These results may suggest that the signals for differentiating the sense and antisense strands are not present in cis-regulatory elements, *per se*, but are coordinated through bound transcription factors or present within the 5' UTR.

In addition to the nucleotide composition of the promoter landscape, we discovered genes with the greatest abundance of highly conserved elements. These genes are annotated to functional clusters, including transcription regulation, regulation of metabolism, and insulin-like growth factor binding proteins. We are not surprised to find an overrepresentation of genes responsible for generalized biological functions. The description of the genes under control of highly conserved promoters provides a framework for the evolution of key regulatory networks, and we propose that the cluster of transcription factors that have promoters with large stretches of 100% conserved sequences are putative early regulators of cellular organization. Indeed, many of the genes we found to have large stretches of conserved sequences have been shown to be involved in early development, and these genes include several of the Hox genes (Favier and Dolle 1997) as well as Dbx1, Vsx2, and Evx2 (Pierani et al. 2001) and Phox2b (Huber and Ernsberger 2006). These proposed networks involved in

gene regulation include those transcription factors that bind these elements and the downstream targets of the genes identified in this study.

Cis-regulatory changes are often considered the primary contributors to morphological evolution, rather than changes at the protein level (Carroll 2005; King and Wilson 1975; Prager and Wilson 1975). Understanding the functional roles of the highly conserved promoters, upstream of transcription factors and growth factor binding proteins, represents a significant challenge for future research. Our findings would predict that these genes retain specific, highly conserved expression profiles, most likely in early development and/or cellular organization, and that mutations in these regions would likely have gross, deleterious consequences. We also note that the conserved elements may not function as regulatory motifs, but may have other, as-yet unappreciated biological functions.

Our analysis is limited to those elements maintained throughout mammals. This approach eliminates those elements that have been modified during mammalian evolution, yet maintain their functional role (Wang et al. 2007), thereby ignoring changes in selection within a given taxon (Sabeti et al. 2006).

In conclusion, this study has provided a bird's eye view of the features of mammalian promoters, from molecular landscapes to the shape of conserved elements. This view enhaces our ability to predict not only functional elements, but also the contributions of individual nucleotides within those elements. Further research can extend these findings, and determine the extent to which generalized attributes of conserved elements can be applied to individual

elements and individual promoter landscapes. Developing tools that allow greater predictive power when evaluating a single element or regulatory locus will help make sense of the near-limitless possibilities of gene regulation, and help future work. Ultimately, we find that, despite the complexity, mammalian promoters have features common across species and these features are consistent and informative.

MATERIALS AND METHODS

Orthologous Promoter Assembly

We extracted the orthologous gene Ensembl IDs from Ensembl 55 for human (GRCh37), chimp (CHIMP2.1), mouse (NCBIM37), rat (RGSC3.4), cow (Btau_4.0), dog (CanFam_2.0), and opossum (monDom5) (Birney et al. 2004). The orthologous set including human, chimp, cow, mouse, and rat totaled 1,921,403, while the set for human, dog, and opossum totaled 20,529. We reduced this set by eliminating all sets that contained any Ensembl gene IDs found elsewhere in our total dataset, reducing the orthologous sets to 10,056. We extracted 1200bp flank gene (upstream), listed as the region immediately upstream of the longest transcript for each species. Using the putative orthologous IDs, we used cdbfasta v0.92 to extract the promoters in fasta format, compiling 10,056 individual files. These files were used for subsequent analyses.

Identification of Conserved Motifs

We used the program Meme v.4.2.0 (Bailey and Elkan 1994) to identify motifs common to all species in the dataset. We used the parameters -oops, to

constrain the search to those common to all species, widths ranging from 5 to 12, each run independently, with a predefined search for 30 motifs per width, restricted to non-overlapping motifs. Only those motifs that were identical across all species were included in this study. Following the motif search, we eliminated those gene sets that included a gene flank with more than 840 Ns in the sequence data. This reduced the gene sets to 9,398. We determined the nucleotide frequency across the entire data set, each motif width independently, and at each individual position within motif widths.

Dinucleotide Analyses

We compiled the conserved motifs for each width, scanning the motifs for the presence or absence of each dinucleotide within each motif width, calculating the frequency of the presence of each dinucleotide relative to the total number of motifs within each width. Regression analyses identified whether the dinucleotide frequencies changed in relation to motif width.

Trinucleotide Analyses

We identified the number of motifs starting and ending with each dinucleotide (NAA, AAN, NAC, etc) across all widths. Chi-square analysis was conducted, with the expected value for each trinucleotide based on the total number of each dinucleotide set (e.g. AAA, CAA, GAA, TAA as the set for the trinucleotide NAA) multiplied by the corresponding nucleotide frequency for that width (Appendix A1). The degrees of freedom was 3 for all analyses. Using the

width with the lowest chi-square value for each individual trinucleotide, a conservative approach, the nucleotides within each trinucleotide with the greatest deviation from expected were determined if their individual contribution to the chi-square value was greater than 7.82 ($p<0.05$ for degrees of freedom of 3). Reverse complement trinucleotide frequencies were compared using the Wilcoxon Signed-Rank test (Appendix A7).

Highly Conserved Promoter Discovery

We used the conserved motifs from widths 8-12 to identify the genes regulated by the most conserved promoters. We discovered 51 genes with four or more 12mers, and set this value as the minimum number of genes necessary from each width. Using the UCSC genome browser (Kent et al. 2002), and human reference, we determined whether the upstream flanking region for these genes overlapped the intron, exon, or UTR of a neighboring gene. The human gene symbols for the 60 genes from Table 2 were run through DAVID functional annotation clustering under highest classification stringency and default options (Dennis et al. 2003; Huang da et al. 2009).

CHAPTER THREE—EVOLUTION OF *LDHA* AND *LDHB* GENE PROMOTERS IN PRIMATES

INTRODUCTION

Encephalization, the increase in brain size relative to body size, is a hallmark of anthropoid primate evolution (Allman 1999; Martin 1990a). Compared to strepsirrhine primates, haplorrhine primates (tarsiers + anthropoids) have significantly larger than expected brain sizes relative to body mass (Joffe and Dunbar 1998), in the most extreme case the human brain is estimated to be over 6-fold larger than expected (Jerison 1973). Based on rates of oxygen consumption, anthropoid brains utilize between 10-20% of total body metabolism (Aiello and Wheeler 1995; Mink et al. 1981), while the vertebrate mean is 5.3% (Mink et al. 1981). The energetic cost of encephalization in anthropoids has been met by an increase in the allocation of resources to that tissue (Armstrong 1983; Mink et al. 1981), rather than an overall increase in basal metabolic rate (Armstrong 1983; Elliott 1948).

The primary consumers of energy in the brain are neurons (Hyder et al. 2006), and lactate is one source of neuronal fuel provided by neighboring astrocytes (Pellerin and Magistretti 1994). Glutamate, released upon neuronal signaling, has been shown to stimulate neighboring astrocytes to produce and release lactate (O'Brien et al. 2007; Pellerin and Magistretti 1994). Lactate is then transported into the neuron (McKenna et al. 1998). Lactate production in astrocytes and subsequent use in neurons is facilitated by the enzyme lactate dehydrogenase (LDH), which interconverts pyruvate and lactate (Adams et al.

1973; Dawson et al. 1964). In astrocytes, LDH converts pyruvate (the final product of glycolysis) to lactate (O'Brien et al. 2007). In neurons, LDH converts lactate back to pyruvate, which can then enter the tricarboxylic acid cycle (O'Brien et al. 2007).

The LDH enzyme is a tetramer, which assembles from the products of two genes, *LDHA* and *LDHB* (Markert 1963a). LDH enzymes with three or four subunits of LDHA (termed LDH4 and LDH5, respectively) rapidly convert pyruvate to lactate, and are not inhibited by high levels of pyruvate (Bishop et al. 1972; Dawson et al. 1964; Vesell 1961). In humans, isoenzyme LDH5 is prevalent in astrocytes (Bittar et al. 1996). LDH enzymes composed of two to four LDHB subunits (LDH3-LDH1, respectively) have a much slower turnover of pyruvate to lactate, are inhibited by high levels of pyruvate (Dawson et al. 1964; Vesell 1961), and have a higher rate of turnover for the reverse reaction of converting lactate to pyruvate (Cahn et al. 1962; O'Brien et al. 2007). In humans, *LDHB* is the only LDH gene that is expressed in neurons (Bittar et al. 1996).

Previous research has found a change in the isoenzyme concentrations in primate brains. In strepsirrhine primates, LDH3-LDH5 are the primary isoenzymes found in the brain (Goodman et al. 1969; Syner and Goodman 1966). LDH2-LDH4 are the most prevalent in tarsiers, while anthropoids show primarily LDH1 and LDH2 in the brain (Goodman et al. 1969; Koen and Goodman 1969; Syner and Goodman 1966). The isoenzymes prevalent in anthropoid brains (LDH1 and LDH2), in contrast with those in strepsirrhine brains, could support greater use of lactate for oxidative phosphorylation in

neurons, as one means by which energy allocation to the brain may have increased in this clade.

The LDH tetramers randomly assemble based on the relative abundance of LDHA and LDHB (Markert 1963b). Therefore, the ratio of LDHA and LDHB can be used to determine the prevalence of each isoenzyme, and vice versa (Markert 1963b). The changes in isoenzyme patterns found in anthropoid brains could be a result of gained *LDHB* expression, decreased or lost *LDHA* expression, or both. We hypothesized that *cis*-regulatory elements in the promoter of *LDHA*, *LDHB*, or both, changed or were gained during stem anthropoid evolution, resulting in changes to brain expression. These tissue-specific changes would have subsequently been fixed in descendant lineages, while retaining the expression profiles in other tissues that remain conserved across mammals.

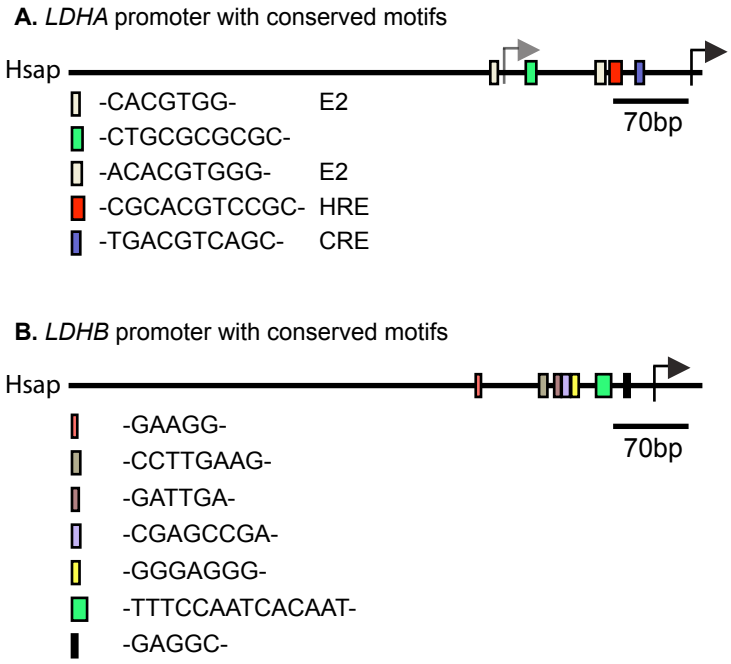
In this study, we characterize the evolution of the *LDHA* and *LDHB* gene promoters in primates. In order to identify conserved nucleotide strings (footprints, i.e. *cis*-regulatory elements) within stretches of noncoding promoter sequences we used a comparative genomic technique called phylogenetic footprinting (Tagle et al. 1988), We also examined those footprints unique to select primate clades who express LDH isoenzymes in differing proportions to one another in the brain. Finally, we evaluate the methylation status of conserved CpG sites in two primate species, hypothesizing that differences in cytosine methylation may account for differences in expression.

RESULTS & DISCUSSION

Promoter Distinction

Reexamination of genomic databases revealed a discrepancy between the transcription start site (TSS) of the *LDHA* reference sequence (canonical non-coding exon 1 and all seven coding exons; NM_005566) and mRNA/cDNA found in GenBank. The human reference sequence implies a TSS 179nt upstream of an alternative TSS (accession numbers AK296667 and AK130587; Fig.1A), and this alternative TSS is the one found (within 5nt) in rat, orangutan, and chimpanzee (Appendix B1). Therefore, we consider the TSS to be chr11: 18,416,108 (GRCh37/hg19 Assembly). For the subsequent analyses, we use the nucleotide region upstream of our TSS as the putative *LDHA* promoter for human, as well as the other species included in this study.

Figure 1. Mammalian conserved elements in the promoters of *LDHA* and *LDHB*



Conserved elements in the *LDHA* promoter. Human used as reference, gray arrow indicates reference sequence TSS (see text), black arrow indicates alternative TSS, used in this study (see text). Sequence of elements next to corresponding colored box, with known identifier of known TFBS, if known. (B) Conserved elements in the *LDHB* promoter. Human used as reference, black arrow indicates TSS. Sequence of elements next to corresponding colored box.

Phylogenetic Footprinting

Our study includes 15 mammals, 12 primate species and three non-primate mammals. The primate species represent four major clades, including apes (*Homo sapiens, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus,* and *Hylobates lar),* Old World monkeys (*Colobus guereza, Trachypithecus obscurus, Mandrillus leucophaeus)*, New World monkeys (*Ateles geoffroyi, Saimiri spp.*), and strepsirrhines (*Microcebus murinus* and *Cheirogaleus medius* for *LDHA/Hapalemur griseus* for *LDHB*; Figures 2 and 3).

In order to identify general features of the promoters of *LDHA* and *LDHB,* we interrogated the sequences and identified the most evolutionarily conserved nucleotide strings among our study taxa. Each element is given a footprint identifier, FPa and FPb, depending upon its presence in the *LDHA* or *LDHB* promoter, respectively (Table 1). For *LDHA*, we found five motifs with 100% conservation across all species, with widths ranging from 7-11nt (Fig.1A and Table 1). Three of these sites have been previously characterized in the rat promoter through experimental analyses (Jungmann et al. 1998; Shim et al. 1997; Short et al. 1994) and one has been characterized in the mouse promoter (Semenza et al. 1996). These four sites are all *cis*-regulatory elements and include two E-boxes (-ACACGTGGG- and -CACGTGG-), bound by the c-Myc/Max heterodimer transcription factor, an hypoxia response element (HRE; -CGCACGTCCGC-), bound by hypoxia inducible factor 1 (Hif-1), and a cyclic-AMP response element (CRE; -TGACGTCAGC-), bound by the CRE-binding protein (Table 1) (Jungmann et al. 1998; Semenza et al. 1996; Shim et al. 1997; Short et al. 1994). The conservation of these elements, in both sequence and spatial organization, as depicted in Figure 1A, suggests the functional importance of these regulatory responses in *LDHA* transcription.

The remaining motif is a 10mer that partially overlaps with an Sp1 site (Jungmann et al. 1998), but includes an uncharacterized domain. This domain includes three CpG repeats, -CGCGCG-, which are completely conserved in mammals (Fig.1A). No human matches in Transfac (Matys et al. 2006) or the literature could be found for this element without removal of at least four bases.

As a putative target for cytosine methylation, this motif may be involved in transcriptional regulation via methylation. It is also noteworthy that the first exon and proximal promoter of *LDHA* fall within a 758bp CpG island, and all conserved elements include at least one CpG dinucleotide (Table 1).

The *LDHB* promoter has not been characterized experimentally, but footprinting identified seven conserved motifs across mammals, ranging from 5-14nt (Fig.1B and Table 1). A 5nt element, -GAGGC-, is located 23 nucleotides upstream of the human TSS. This element matches the basic transcription element (BTE), shown to be necessary for high expression levels, but insufficient alone for activating transcription (Sogawa et al. 1993; Yanagida et al. 1990).

Table 1. Elements conserved across mammals in the promoters of *LDHA* and *LDHB*

***LDHA* motifs**	**Element ID**	**Position relative to Human TSS**	**Putative Transcription Factor Binding Site**	**Putative Transcription Factor**
CACGTGG	FPa1	-179	E-Box	c-MYC (Shim 1997)
CTGCGCGCGC	FPa2	-144	Unknown	Unknown
ACACGTGGG	FPa3	-80	E-Box	c-MYC (Shim 1997)
CGCACGTCCGC	FPa4	-65	HRE	HIF-1 (Semenza 1996)
TGACGTCAGC	FPa5	-42	CRE	CRE-BP (Short 1994)
***LDHB* motifs**		**Position relative to Human TSS**	**Putative Transcription Factor Binding Site**	**Putative Transcription Factor**
GAAGG	FPb1	-155	Cardiac-specific sequence (CSS;Dhar 1997)	Cardiac-specific sequence binding protein 2 (CSSBP2; Dhar 1997)
CCTTGAAG	FPb2	-94	Unknown	SF1 (Sandhoff 1998)
GATTGA	FPb3	-86	Cardiac-specific (Truter 1992)	Unknown

CGAGCCGA	FPb4	-77	Unknown	Unknown
GGGAGGG	FPb5	-69	GA Box (Bossone)	MAZ (Bossone)
TTTCCAATCACAAT	FPb6	-39	CCAAT box(es)	CCAAT box binding proteins
GAGGC	FPb7	-23	BTE (Yanagida 1990)	Unknown

The 14nt element, -TTTCCAATCACAAT- (Fig.1B), contains two tandem CCAAT boxes, -CCAAT- followed by -CACAAT- (Table 1). CCAAT/enhancer-binding proteins (C/EBPs) are considered key regulators of transcription due to their high levels of expression in diverse tissues (e.g. liver, lung, and adipose) and the number of processes in which they play critical roles, such as differentiation, metabolism, inflammatory response, and others (Ramji and Foka 2002).

The GA box (-GGGAGGG-; Fig.1B) has been shown to be a target of the MAZ transcription factor in the *c-myc* promoter (Bossone et al. 1992). MAZ has been shown to activate muscle-specific expression profiles, including skeletal and cardiac (Himeda et al. 2008). Although *LDHA* is the primary LDH gene in skeletal muscles (Beebee and Carty 1982; Latner and Skillen 1964; Milne and Doxey 1987; Vesell and Bearn 1961), MAZ has been shown to restrict muscle cell-type expression as well (Himeda et al. 2008), so this element may function in heart *LDHB* expression.

A fourth motif, -CCTTGAAG- (Fig.1B), is composed of a critical element (CCTTGA) for high-affinity binding of steroidogenic factor 1 (SF1) (Sandhoff et al. 1998). SF1 can activate transcription without cofactors (Sandhoff et al. 1998) and in humans it has widespread central nervous system (CNS) expression, in

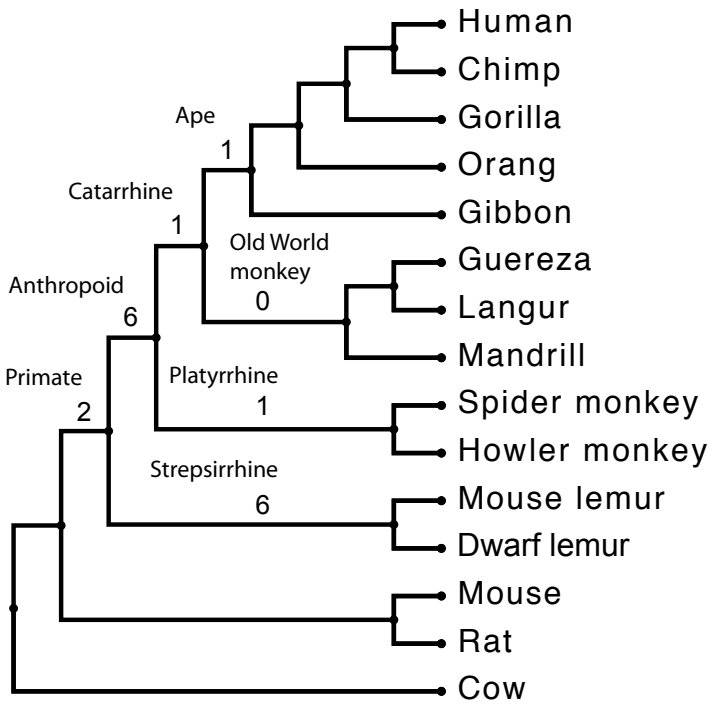
contrast with strict hypothalamus expression in mouse (Ramayya et al. 1997). This element provides one putative target for modifying brain expression of *LDHB*, based on changes in transcription factor expression.

For the fifth motif, -GATTGA- (Fig.1B), the reverse complement has been identified as a TFBS in the promoter of the pro-α2(V) collagen gene (Truter et al. 1992). Interestingly, COL5A2 has a strict tissue-specific expression profile in the heart (Table 1), a tissue in which *LDHB* is the primary LDH gene across mammals (Beebee and Carty 1982; Latner and Skillen 1964; Milne and Doxey 1987; Vesell and Bearn 1961). In conjunction with this, the sixth motif, -GGAAG- (Fig.1B), has been shown to inhibit expression in skeletal muscle of the cardiac myosin light chain 2 (*MLC2*) gene (Dhar et al. 1997). Similarly, *LDHA* is the primary LDH gene expressed in skeletal muscles across mammals, so this element may function to restrict *LDHB* expression. These results provide at least two conserved motifs matching those found in heart-specific genes, with one motif involved in activating expression (GATTGA), and another restricting expression (GGAAG).

The seventh motif, -CGAGCCGA-, contains two CpG sites, conserved across mammals (Fig. 1B). The function of the BTE element described previously has been shown to require other GC-rich elements, and this is the only other conserved, GC-rich element in this promoter (Table 1), therefore it may function in conjunction with the putative BTE element (Table 1). Alternatively, while not tandem repeats as seen in the *LDHA* promoter (Fig.1A), this motif provides

putative cytosine methylation targets as a conserved regulatory mechanism. As with *LDHA*, the 1st exon and proximal promoter fall within a CpG island.

Figure 2. *Cis*-elements gained during primate evolution in the promoter of *LDHA*



Phylogenetic tree with mammals included in this study. Values on branches indicate number of *cis*-element gains in the *LDHA* promoter. Branch names given above numbers.

Differential phylogenetic footprinting

Within primates, we evaluate differential footprints (conserved sequences within a subset of taxa) on seven branches of our phylogenetic tree (Fig.2A),

including stem primates, anthropoids, strepsirrhine, platyrrhine, catarrhine, Old World monkey, and ape (Fig.2). The elements found in the *LDHA* promoter are listed in Appendix B2, and those in the *LDHB* promoter are listed in Appendix B3, with identifiers dFPa or dFPb applied to *LDHA* or *LDHB* elements, respectively (Table 2 and Appendices B2 and B3). For *LDHA*, we found 17 elements, with two on the primate, six on the anthropoid, six on the strepsirrhine, one on the platyrrhine, one on the catarrhine, none on the Old World monkey, and one on the ape stem (Fig.2 and Appendix B2). For *LDHB*, we found 11 elements, with none on the primate, two on the anthropoid, four on the strepsirrhine, three on the platyrrhine, none on the catarrhine, one on the Old World monkey, and one on the ape stem (Fig.2 and Appendix B3).

Of the 12 elements conserved across all mammals, nine reside within 100nt of the TSS (Table 1). In contrast, of the 28 elements unique to different primate clades, only one of them is within 100nt of the TSS, that being the modification of the known Sp1 site, dFPa12 (Appendices B2 and B3). These results may reflect the stringency used to detect differential footprints, but may also indicate the importance of the intra-element spacing depicted in Figure 1. Novel elements gained between highly conserved elements may infringe on combinatorial relationships, forcing new elements to be gained at more distal sites.

Table 2. Anthropoid-specific elements in the *LDHA* and *LDHB* promoter

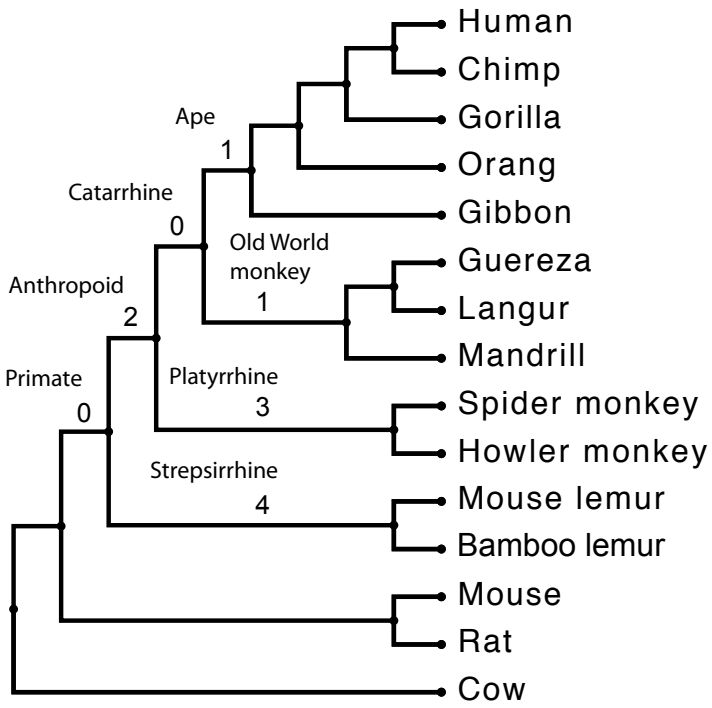
***LDHA* anthropoid element**	**Element ID**	**Position** (relative to human TSS)	**Putative transcription**

			factor
CGTCC	dFPa8	-254	GR (Ray 1990)
GCAGTC	dFPa10	-171	AP1 (Rylski 2009)
CCACCCC	dFPa11	-98	Sp1 (Jungmann 1998)
***LDHB* anthropoid element**	**Element ID**	**Position** (relative to human TSS)	**Putative transcription factor**
GACCAGCT	dFPb6	-180	MtEBP (Mt element-binding protein; Suzuki 1995)

The evidence that LDH1 is the primary isoenzyme in human neurons (Bittar et al. 1996), whereas LDH5 is the primary isoenzyme in rat neurons (O'Brien et al. 2007) suggests that there is a loss of neuronal *LDHA* expression. This is further supported by the conservation of the CRE site in the *LDHA* promoter, a site shown to upregulate many genes in neurons in response to Ca^{2+} influx (Sanchez-Munoz et al. 2010). Of the six elements common to anthropoids in the *LDHA* promoter (Table S1), we identify three that are putative contributors to downregulating neuronal expression, dFPa8, dFPa10, and dFPa11 (Table 2). The reverse complement of dFPa8, -GGACG-, is a putative target of the glucocorticoid receptor (GR), and binding of this factor to the interleukin-6 promoter downregulates expression (Ray et al. 1990). GRs have been shown to antagonize cAMP-induced transcriptional activation in neurons (Diaz-Gallardo et al. 2010). The strength of Cre-mediated transcription, in conjunction with the activation of CREB upon Ca^{2+} influx (Sanchez-Munoz et al. 2010) may require

such a repressor to downregulate, or completely eliminate *LDHA* transcription in neurons.

Figure 3. *Cis*-elements gained during primate evolution in the promoter of *LDHB*



Phylogenetic tree with mammals included in this study. Values on branches indicate number of *cis*-element gains in the *LDHB* promoter. Branch names given above numbers.

The reverse complement of dFPa10, -GACTGC-, is a putative target of AP1 (Minth and Dixon 1990), which can function as a transcriptional repressor in neurons (Rylski et al. 2009). The third element, dFPa11, is a modification of a previously described Sp1 binding site (Jungmann et al. 1998). Sp1 is near absent

in mature neurons, but Sp4 is prevalent (Mao et al. 2009). Since these two zinc-finger transcription factors bind similar, GC-rich elements (Mao et al. 2009), the modification of this element may reduce or eliminate Sp4 specificity in neurons. This trend appears to have been continued during stem ape evolution, when the other Sp1 site was changed in a similar manner (Appendix B3).

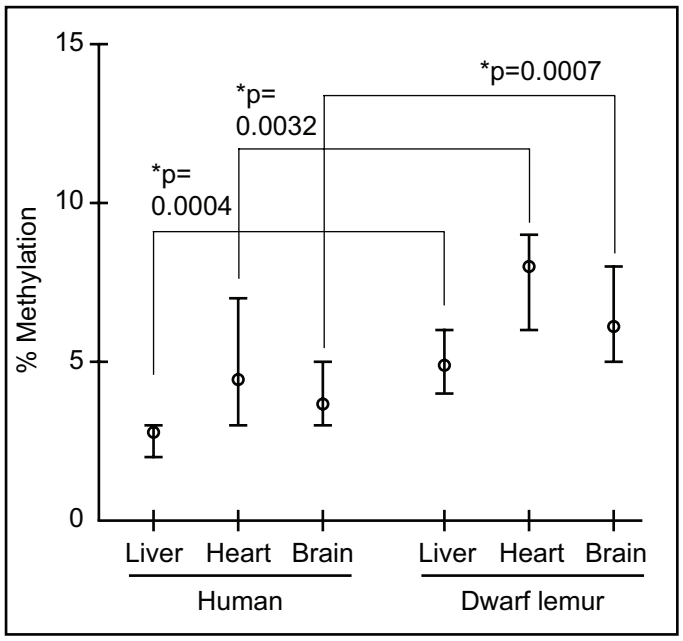
Of the two elements common to anthropoids in the *LDHB* promoter (Table 2), we identify only one, dFPb6 (-GACCAGCT-), as a potential upregulator of *LDHB* in neurons. This element is a putative oxidative phosphorylation element, important for the coordination of mitochondrial and nuclear elements of the electron transport chain (ETC) (Suzuki et al. 1991). Coordinating oxidative phosphorylation with the upregulation of *LDHB* fits the model in which LDH isoenzymes composed primarily of LDHB subunits support aerobic metabolism. Based on these findings, we suggest that both *LDHA* downregulation and *LDHB* upregulation have led to the changes in LDH isoenzymes in primate brains.

Epigenetic Modifications

In the context of evolutionary conservation, footprints in variable landscapes suggest not just functional importance, but similar or identical functional roles. Such an assumption may not hold true if the conserved sequences have different epigenetic modifications across species. Conserved CpG sites with variable methylation patterns across species have already been demonstrated in primates. The cell cycle-related kinase (CCRK) gene promoter displayed variable methylation amongst four catarrhine species in addition to

differences in the expression of this gene (Farcas et al. 2009). To further evaluate the evolution of these two gene promoters, we measured the epigenetic modifications to conserved CpG sites in the promoters of *LDHA* and *LDHB* in human and dwarf lemur (a strepsirrhine). We acquired genomic DNA from three tissues (liver, heart, and brain) representing different LDH conditions. Across mammals, *LDHA* is the primary LDH gene expressed in liver, and *LDHB* in heart (Beebee and Carty 1982; Latner and Skillen 1964; Milne and Doxey 1987; Vesell and Bearn 1961), whereas the relative abundance of these two gene products in the brain differs between these two species, as previously described.

Figure 3. Human and dwarf lemur methylation levels of conserved CpG sites in the promoter of *LDHA*



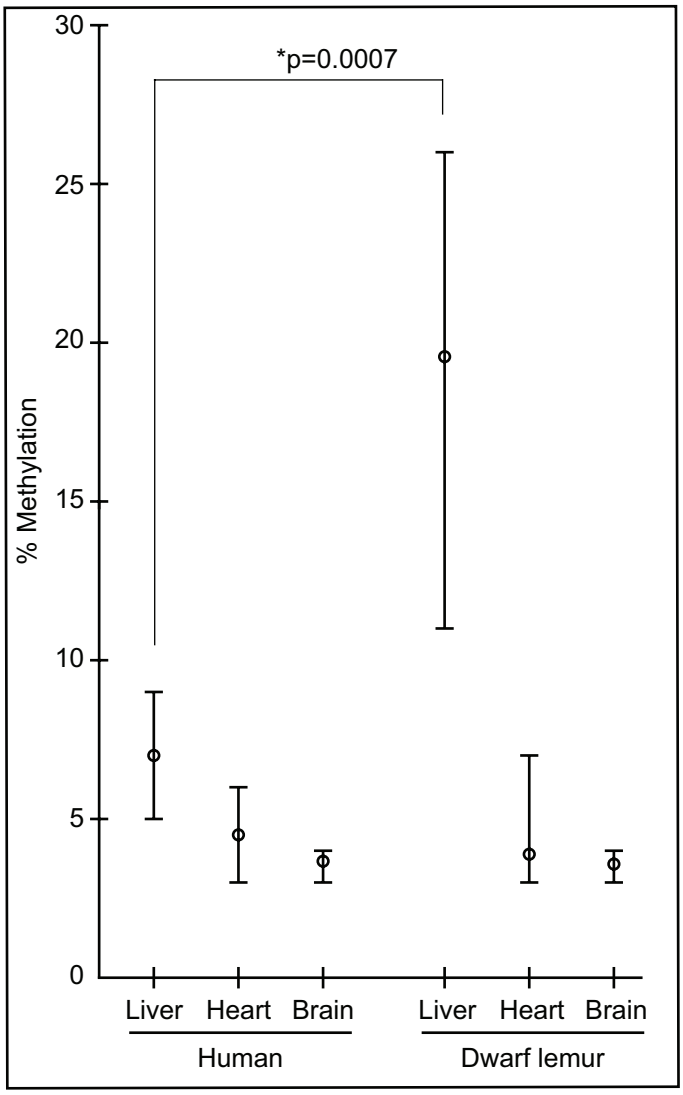
Open circles indicate mean across all three CpG sites, with horizontal bars above and below indicating range. Mann-Whitney tests were conducted to determine differences in methylation between human and dwarf lemur for equivalent tissue, with p-values given.

For *LDHA*, we tested the three CpG sites in promoter element FPa3, -CGCGCG-, shown in Table 1. For *LDHB*, we tested the two CpG sites in promoter element FPb3, -CGAGCCGA- (Table 1), as well as a single, conserved CpG site 10bp downstream of this element. Our results reveal a methylation pattern for both species, in which for *LDHA*, liver is the least methylated, brain is intermediate, and heart is the most methylated tissue (Fig.3). For *LDHB*, brain is the least methylated, followed by heart, while liver is the most methylated (Fig.4). In contrast with these patterns, however, the levels of methylation vary between these species.

Across all three human tissues (heart, liver, and brain), we detect low signals of methylation (mean <5%) for the three CpG sites in the conserved *LDHA* motif (Fig.3). For the dwarf lemur there is a greater disparity in methylation levels, with a mean of <5% methylation in liver, 6.1% in brain, and 8% in heart (Fig.3). The mean methylation levels for human and dwarf lemur are statistically distinct in all three tissues (Mann-Whitney w-tailed test; Fig.3).

Similar to *LDHA*, for *LDHB* we find low signals of methylation (mean <5%) for the three conserved CpG sites in human for both heart and brain, and a mean of 7% methylation for liver (Fig.4). For the dwarf lemur, *LDHB* shows <5% mean methylation in both heart and brain, but 19.6% methylation in liver (Fig.4). Only liver methylation levels were statistically distinct between human and dwarf lemur (Mann-Whitney 2-tailed test; Fig.4).

Figure 4. Human and dwarf lemur methylation levels of conserved CpG sites in the promoter of *LDHB*



Open circles indicate mean across all three CpG sites, with horizontal bars above and below indicating range. Mann-Whitney tests were conducted to determine differences in methylation between human and dwarf lemur for equivalent tissue, with p-values given when significant.

The greater methylation of *LDHA* CpG sites in the brain of dwarf lemur is surprising, since this would suggest lower levels of expression for this gene than in human. We consider two interpretations of these results. They could suggest

that the prevalence of isoenzymes composed primarily of *LDHB* in the brain of anthropoids is a consequence of an increase in expression of *LDHB*, rather than a decrease of *LDHA*. Alternatively, the levels of *LDHA* promoter methylation in dwarf lemur brain (6.1%; Figure 3) may have little to no impact on expression, and may not be a reasonable indicator of transcriptional repression. Further research is necessary to determine whether such methylation levels could greatly reduce expression of the downstream gene.

Conclusion

In this study, we discover conserved elements in the promoters of *LDHA* and *LDHB*. These elements likely contribute to the expression profiles common across mammals, as well as the profiles unique to certain taxonomic clades. We propose that during stem anthropoid evolution, modifications in both gene promoters resulted in a decrease or loss of *LDHA* expression, and the gain of *LDHB* expression.

Our results suggest that *LDHA* expression may be reduced or eliminated in neurons in anthropoid primates either by GR antagonism of CREB transcriptional activation, AP1 repression, modification of Sp4 binding site, or a combination of these elements. Although AP1 repression is possible, this TF is abundant in many other tissues, and we suspect more widespread loss of *LDHA* transcription would be evident. The modified Sp1 (and by association Sp4) site is the only differential footprint within 100nt of the TSS (Appendices B2 and B3), and Sp4 is prevalent in neurons. However, the further modification during stem

ape evolution would suggest an even more pronounced discrepancy between *LDHA* and *LDHB* expression in ape brains. Based on isoenzyme concentrations, apes, Old World monkeys, and New World monkeys have similar ratios of LDHA:LDHB (Goodman et al. 1969; Koen and Goodman 1969; Syner and Goodman 1966), so further decrease in *LDHA* expression in apes would require a decrease in *LDHB* expression as well. The gain of a GR site, as a potential antagonist of CREB activation, is an appealing mechanism for downregulation of *LDHA* expression, mitigating a strong activator element. We consider the modified Sp1 site, as well as the gained GR site as the two most likely candidates for reduction, if not loss of *LDHA* expression in neurons of anthropoid primates.

For *LDHB*, we identify two potential targets for gain of neuronal expression. The first is the gain of SF1 expression in human neurons, in which all mammals share a putative target binding site. We do not yet know when in primate evolution SF1 gained cortical neuronal expression, but we would hypothesize that this expression profile was gained during stem anthropoid evolution. Such a modification could have profound impacts on gene expression profiles in the neurons of anthropoid primates beyond metabolism and requires further investigation.

The *LDHB* element gained during anthropoid evolution is a putative oxidative phosphorylation element. This would imply a more global transcriptional effect rather than a neuron-specific gain in expression. However, due to the abundance of conserved elements in the *LDHB* promoter, this gene appears to

be tightly regulated, so such an element could be suppressed in other tissues. Despite such a possibility, we are inclined to consider the gain of SF1 neuronal expression as the more likely candidate for gaining *LDHB* neuronal expression during anthropoid evolution.

Further research is necessary to determine whether such elements are, in fact, *cis*-regulatory elements, rather than sequences necessary for local topology (Parker et al. 2009), portions of small RNAs (Zhang 2009), retain other, unknown functions (Margulies and Birney 2008), or have no function but remain conserved due to chance. These results do, however, provide putative targets of anthropoid-specific regulatory features that may play important roles in neuronal expression of these genes, and consequently in brain energetics. The differences in methylation levels at conserved CpG sites further suggest evolutionary shifts in the use of epigenetic modifications for transcriptional regulation. Evolution of methylation is one mechanism by which regulatory evolution can occur without modification to the spacing and organization of conserved elements, such as that found for both *LDHA* and *LDHB* (Fig.1), and should be further investigated. As a result, further examination of the diverse factors subject to regulatory evolution requires broader comparative approaches and should include comparative methylation, since little is known to what extent adaptive evolution is acting on these modifications. Understanding the entire regulatory network of a single gene is a monumental task. By developing a greater understanding of how these networks evolve, we can better understand how they ultimately impact the phenotype.

MATERIALS AND METHODS

Sequencing and Phylogenetic Footprinting

We PCR amplified the putative promoter for both *LDHA* and *LDHB* in 12 primate species, including *Homo sapiens, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus, Hylobates lar, Colobus guereza, Trachypithecus obscurus, Mandrillus leucophaeus, Ateles geoffroyi, Saimiri spp.*, *Cheirogaleus medius* for *LDHA*, and *Hapalemur griseus* for *LDHB*. Amplicons were gel-purified and ligated into pGem T-Easy vectors (Promega) either by overnight ligation at 4°C or one hour at room temperature. Vectors were transformed by heat shock (42°C) into DH5-α chemically competent cells (Invitrogen) and grown for 1.5 hours at 36°C in LB. Cells were grown on LB plates with X-gal, IPTG, and ampicillin, and grown overnight at 36°C. Positive colonies were selected and grown overnight in liquid LB with ampicillin, followed by plasmid extraction using Qiagen Miniprep kits (Qiagen) following manufacturer's protocols. Amplicons were sequenced at Michigan State University (RTSF) using T7 and Sp6 primers.

Chromatograms were visualized, aligned, and manually curated in Sequencher (Ann Arbor), with subsequent alignment in Geneious (Drummond et al. 2009). Phylogenetic footprinting was done using Meme v.4.2.0 (Bailey and Elkan 1994), constraining search to elements common across all species, non-overlapping, independently run for element widths 5-12nt, with a predefined search for 30 motifs per width, with manual curation. Differential phylogenetic footprinting was done as described above, except for use of a parameter allowing

for elements to be shared by a subset of species. Results were verified and further described by manual curation, based on the criteria of elements fixed across the subset analyzed, from widths 5-12nt with a maximum of 67% identity of the fixed element with the aligned elements of outgroup taxa. These elements were further constrained by eliminating those elements found in outgroup taxa within 50nt of the differential footprint. For instance, a 5mer fixed across all apes in this study would require two differences in the string in all other taxa, and could not be found within 50nt in the other taxa.

Bisulfite Conversion and Pyrosequencing

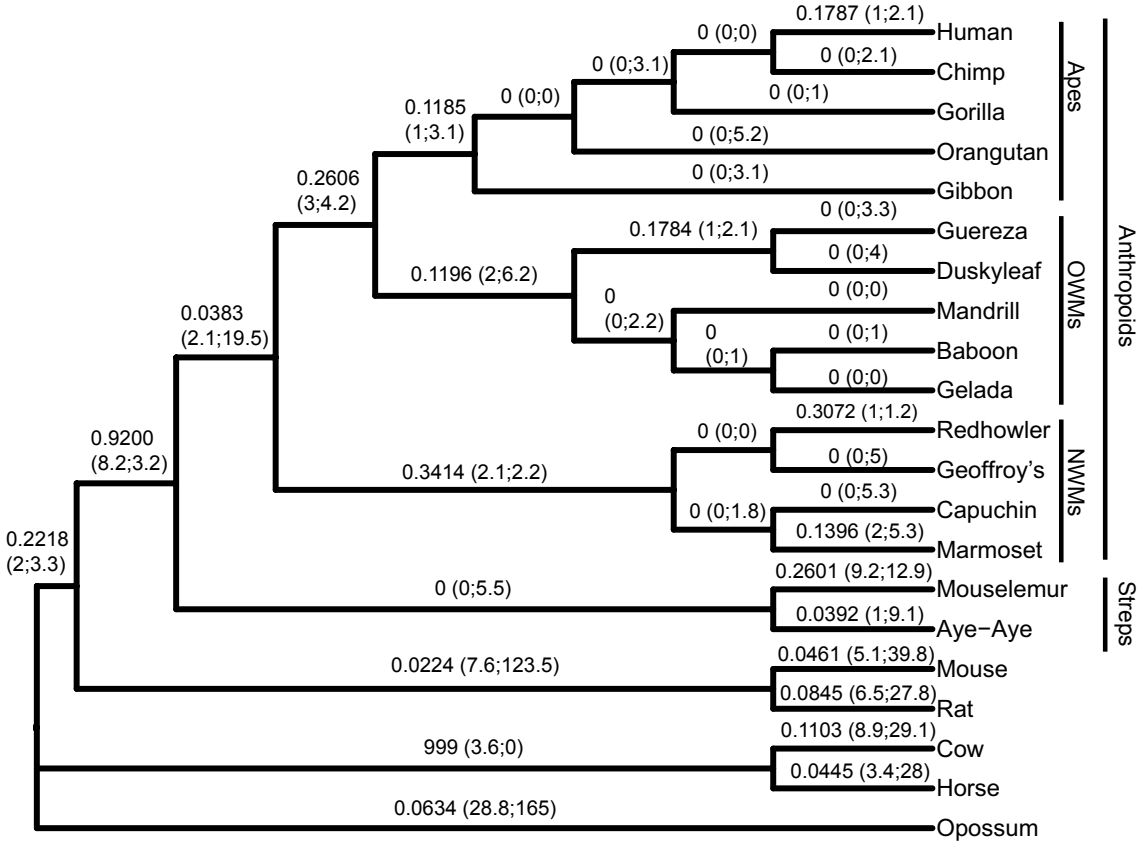
Liver, heart, and cortical brain tissue was acquired for a human (Proteogenex) and a strepsirrhine primate, *Cheirogaleus medius* (Duke University Lemur Center). DNA and RNA were extracted using TRIzol reagent (Invitrogen) followed by RNeasy kit (Qiagen) according to manufacturers protocol. DNA was bisulfite converted using Bisulfite Conversion Kit (Qiagen) according to manufacturer's protocol. Amplification and sequencing primers were designed by Pyromark Assay Design software (Qiagen), and primers and amplification conditions are provided in Supplemental Methods. Bisulfite amplification was done using Epitect Kit (Qiagen), with a gradient annealing temperature. Amplifications were verified by gel electrophoresis and pyrosequencing was done using Qmark24 according to manufacturers protocol (Qiagen). Two - three independent amplifications were done at different annealing temperatures and subsequently pyrosequenced. Individual methylation

percentages at each CpG site were used to calculate statistical significance between tissues of the same type (Mann-Whitney 2-tailed test).

CHAPTER FOUR—STASIS IN CHANGE CHARACTERIZE METABOLIC ADAPTATIONS IN THE EVOLUTION OF *LDHA* AND *LDHB* CODING REGIONS IN PRIMATES

The shift in brain expression profiles of *LDHA* and *LDHB* that distinguishes anthropoid from strepsirrhine primates led us to consider whether amino acid replacements that modified the structure and/or function of the individual proteins took place during primate evolution. Since the isoenzymes previously described have different rates of substrate turnover depending upon the composition of the LDH tetramer, we conjectured that modifications to the coding sequence could have altered these rates and changed the metabolic functions of the isoenzymes. We could not reasonably hypothesize on the impacts of the change in expression without first evaluating the effects of such modifications on the enzymatic properties of LDH. Therefore, we conducted a comparative sequence analysis on the coding regions of *LDHA* and *LDHB* across different primate clades, and tested whether either, or both of these genes underwent significant modifications during anthropoid evolution.

Figure 1. Inferred nonsynonymous and synonymous nucleotide substitutions in *LDHA* during primate evolution



Free ratio model (codeml model 1) ω values on mammalian species tree, along with ML estimates of nonsynonymous (N*dN); synonymous substitutions (S*dS) on each branch. OWMs, Old World monkeys; NWMs, New World monkeys; Streps, strepsirrhines. Scientific and common names are given in Appendix C1.

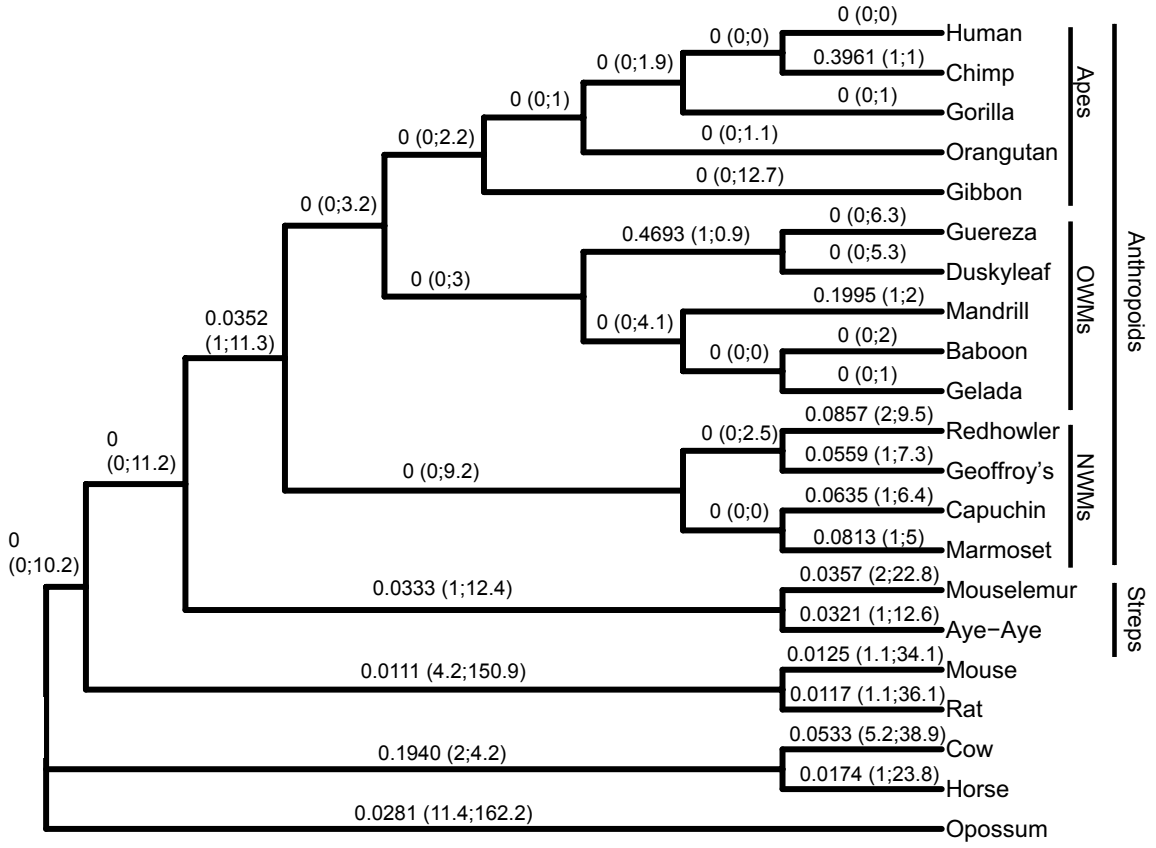
Both genes are composed of eight exons, seven of which are coding. The coding sequence of *LDHA* is 999nt (human NM_005566) whereas the coding sequence of *LDHB* is 1005nt (NM_002300). We amplified all seven coding exons for both *LDHA* and *LDHB* in fourteen species, representing four major primate clades: strepsirrhines, New World monkeys, Old World monkeys, and apes (Figures 1 and 2, Appendix C1). To test whether adaptive evolution occurred in either or both of these genes during anthropoid evolution, we evaluated the per site ratio of nonsynonymous (dN) to synonymous (dS) substitutions on each

branch of our mammalian tree for both *LDHA* and *LDHB* (codeml model 1; Figures 1 and 2, respectively). The signal for adaptive evolution is generally considered an ω value > 1. Only one branch on either tree, stem laurasiatheria (cow + horse) for *LDHA,* displays an ω value consistent with adaptive evolution (999, Fig. 1). However, this branch is unprotected at the base of the tree, and the ω value is based on an absence of synonymous substitutions (3.6 nonsynonymous, 0 synonymous substitutions; Fig. 1). This scenario is likely a result of shifting synonymous substitutions to the opossum branch (165 synonymous substitutions; Fig. 1). The average ω value across the entire tree for *LDHA* is 0.0703 and for *LDHB* is 0.0257, both strong signals of purifying selection. The stem anthropoid branch does not exhibit signals of adaptive evolution in *LDHA* (0.0383) or *LDHB* (0.0352).

Despite the absence of signals of adaptive evolution, we further explored whether selection pressures differ between anthropoid primates and the other mammals in our study. We reasoned that an elevated ω value on either the anthropoid stem or across anthropoids as a whole, relative to the other mammals, could suggest differences in selection pressures, and implicate adaptive modifications to one or both of these genes in anthropoids. We conducted three tests for both *LDHA* and *LDHB* (Table 1). First, we assigned one ω value to the anthropoid stem and another ω value to all other branches on the tree (branch-based 1; Table 1). Second, we assigned one ω value to all anthropoids, including the anthropoid stem, and another ω value to all other branches (branch-based 2; Table 1). Third, we assigned one ω value to all

anthropoids, including the anthropoid stem, another ω value to all strepsirrhines, including the strepsirrhine stem, and a third ω value to all other branches (branch-based 3; Table 1). We did not find any statistical support for a difference in selective pressures between anthropoid branches and all other branches for either *LDHA* or *LDHB*, regardless of the test (Table 1). Based on these findings, there is no evidence for adaptive evolution of either of these genes during anthropoid evolution.

Figure 2. Inferred nonsynonymous and synonymous nucleotide substitutions in *LDHB* during primate evolution



Free ratio model (codeml model 1) ω values on mammalian species tree, along with ML estimates of nonsynonymous (N*dN); synonymous substitutions (S*dS)

on each branch. OWMs, Old World monkeys; NWMs, New World monkeys; Streps, strepsirrhines. Scientific and common names are given in Appendix C1.

We do note, however, that the mouse lemur (*Microcebus murinus*) terminal branch displayed a strikingly high number of inferred nonsynonymous substitutions in *LDHA* (9.2) relative to the rest of the tree (Fig. 1). Based on this observation of an apparently exceptional number of substitutions, we decided to further test whether there was a statistical difference between the rate of substitution on this branch versus the rest of the tree. We first assigned one ω value to the mouse lemur terminal branch and another ω value to the rest of the tree. Then we assigned one ω value to the mouse lemur terminal branch, another ω value to all other primate branches, and a third ω value to all other branches (Table 2). Both approaches fit the data significantly better than the fixed ω model (Table 2), providing evidence that *LDHA* evolved adaptively on this branch. Since the ω value on the mouse lemur branch (0.2601; Fig. 1) is less than 1, we cannot rule out a reduction in functional constraint, with a rate approaching neutrality.

Table 1. Model tests of *LDHA* and *LDHB* evolution in anthropoid primates

Model-*LDHA*	**Anthropoid stem ω**	**All anthropoids ω**	**Strep ω**	**Non-primate mammal ω**	**Likelihood (-ln L)**	**2Δln L**	**p***
Fixed ω	0.0703	0.0703	0.0703	0.0703	-3545.42	N/A	N/A
Free ratio	0.0383	Variable	Variable	Variable	-3513.28	64.28	*P*<0.005
Branch-based 1(Anthropoid stem/All others)	0.0416	0.0716	0.0716	0.0716	-3545.15	0.54	*P*>0.25
Branch-based 2(All anthropoids /All others)	0.0676	0.0676	0.0709	0.0709	-3545.40	0.04	*P*>0.75
Branch-based 3(All	0.0673	0.0673	0.1494	0.0650	-3543.43	3.98	*P*>0.10

anthropoids /strepsirrhines/All others)							
Model-*LDHB*	**Anthropoid stem ω**	**All anthropoids ω**	**Strep ω**	**Non-primate mammal ω**	**Likelihood (-ln L)**	**2Δln L**	**p***
Fixed ω	0.0257	0.0257	0.0257	0.0257	-3396.66	N/A	N/A
Free ratio	0.0352	Variable	Variable	Variable	-3381.38	30.56	*P*>0.75
Branch-based 1(Anthropoid stem/All others)	0.0337	0.0255	0.0255	0.0255	-3396.63	0.06	*P*>0.75
Branch-based 2(All anthropoids /All others)	0.0363	0.0363	0.0234	0.0234	-3396.11	1.10	*P*>0.25
Branch-based 3(All anthropoids /strepsirrhines/All others)	0.0363	0.0363	0.0291	0.0226	-3396.01	1.30	*P*>0.50

Branch-based models (codeml model 2) list in parentheses the compared branches separated by "/." p* value is based on χ^2 test, with free-ratio and branch-based models compared with fixed model (codeml model 0). Strep, strepsirrhines; N/A, not applicable.

In order to further explore the amount of change on the mouse lemur terminal branch, we reconstructed the amino acid changes to both *LDHA* and *LDHB* on each branch of the primate tree, with the number of inferred nonsynonymous substitutions depicted in Appendices C2 and C3, respectively (marginal reconstruction with posterior probabilities). For *LDHA*, 19 branches (of 31 primate branches) do not exhibit any nonsynonymous substitutions (61%), six underwent one substitution (19%), three underwent two substitutions (10%), two underwent three substitutions (6%), and, surprisingly, one underwent eight substitutions (3%; Fig.1).

Table 2. Model tests for mouse lemur *LDHA* evolution

Model-*LDHA*	**Mouse**	**Other**	**Non-primate**	**Likelihood**	**2Δln L**	**p***

	lemur	**primates**	**mammals**	**(-ln L)**		
Fixed ω	0.0703	0.0703	0.0703	-3545.42	N/A	N/A
Branch-based (Mouse lemur/All other branches)	0.2508	0.0651	0.0651	-3541.80	7.24	*P*<0.01
Branch-based (Mouse lemur/Other primates/Non-primate mammals)	0.2493	0.0833	0.0598	-3541.12	8.6	*P*<0.025

Branch-based models (codeml model 2) list in parentheses the compared branches separated by "/." p* value is based on χ^2 test, with free-ratio and branch-based models compared with fixed model (codeml model 0). N/A, not applicable.

For *LDHB*, 20 branches do not exhibit any nonsynonymous substitutions (65%), nine underwent one substitution (29%), and two underwent two substitutions (6%; Fig.2). With a range of 0-2 substitutions across all primate branches, and only one substitution on the stem anthropoid branch (Fig.2), there is no evidence of significant nonsynonymous changes in *LDHB* within primates. This strict purifying selection of *LDHB* is not surprising, considering its association with oxidative phosphorylation, and the dependence upon this metabolic pathway in mammals.

LDHA has a range of 0-8 nonsynonymous substitutions, although 30 of the 31 primate branches have a range of 0-3 nonsynonymous substitutions (Fig.1). In order to determine whether eight substitutions is significantly different than the number of substitutions found on all other primate branches, we calculated whether any statistical outliers exist in the number of nonsynonymous substitutions for either *LDHA* or *LDHB* in primates (Grubbs' test; Appendix C4).

We find that the mouse lemur branch for *LDHA* represents the only statistical outlier for either *LDHA* or *LDHB* (Appendix C4).

We conducted branch-sites tests for positive selection on the mouse lemur terminal branch to determine whether specific codons evolved adaptively in *LDHA* on this branch (Appendix C5). The model, allowing for a subset of codons to evolve with an ω value greater than one, did not fit the data significantly better than the null model of an ω value = 1 for all codons (Appendix C5). In addition, of the eight significant nonsynonymous substitutions on this branch, four were found to have evolved adaptively, although none of the four had posterior probabilities > 95% (Appendix C5). Therefore, we cannot say that individual codons, or segments of *LDHA* evolved adaptively on the mouse lemur terminal branch.

These results provide two important points of interest. First, we provide evidence that neither *LDHA* or *LDHB* evolved adaptively during stem anthropoid evolution, or on descendent anthropoid lineages. We demonstrate that the ω values, either on the anthropoid stem or across all anthropoid branches, are not significantly different than the ω values found on the other branches of the tree (Table 1). Despite changes in the expression profiles during that evolutionary time period, there appear to be little to no changes to the structure of the LDH gene products in anthropoid primates. These results support the inferences made in Chapter Three regarding the impact on metabolism due to a shift from LDHA to LDHB in anthropoid primate brains.

Second, only one primate branch, the terminal mouse lemur branch (*LDHA*), shows significant nonsynonymous substitutions relative to the rest of the primates (Table 2 and Appendix C4). In trying to understand why these changes may have occurred, we note that this species shares a number of geographic, dietary, and behavioral traits with the aye-aye (*Daubentonia madagascariensis*), a species for which we do not find such substitutions in either *LDHA* or *LDHB* (Figures 1 and 2). Both species are found in Madagascar, with overlapping geographic distributions (Mittermeier et al. 2008). Both are omnivorous (Dammhahn and Kappeler 2010; Kaufman et al. 2005) and nocturnal (Giroud et al. 2010; Perry et al. 2007), suggesting that the modifications in mouse lemur *LDHA* are not likely associated with these characteristics.

The mouse lemur is, however, one of the few known primates that are heterothermic, with seasonal adjustments in energy intake and expenditure (Giroud et al. 2010). During the winter months, faced with food scarcity, this species drastically reduces caloric intake, between 40-80% food deprivation, and enters a state of torpor, or temporary hibernation (Giroud et al. 2010). Mouse lemur torpor involves a decrease of daily energy expenditure between 21-47%, depending upon the length of day and the level of food deprivation (Giroud et al. 2010), and survival at 20-30% of the normothermic metabolic rate (Schmid and Speakman 2000). This species utilizes lipid stores during torpor, rather than using protein (Giroud et al. 2010). The reproductive season immediately follows the food shortage winter, and male mouse lemurs with heavier weight (greater muscle mass) have higher reproductive rates (Giroud et al. 2010).

Since *LDHA* is the primary LDH gene expressed in skeletal muscles (Beebee and Carty 1982; Goodman et al. 1969; Latner and Skillen 1964; Milne and Doxey 1987), and has the greatest propensity for promoting glycolysis rather than aerobic metabolism (Greiner et al. 1994; Vesell 1961), we propose that during the evolution of mouse lemurs adaptive changes to the *LDHA* gene were positively selected in order to reduce glycolytic rates in skeletal muscles during periods of torpor. The use of lipid stores, coupled with food deprivation, would require oxidative phosphorylation for efficient use of lipid mass, rather than the inefficient extraction from glycolysis. Such a transition, from glycolytic enzymes to oxidative phosphorylation enzymes has been demonstrated during transition to torpor in the arctic ground squirrel (*Spermophilus parryii*) (Yan et al. 2008). We note that we did not detect evidence of adaptive evolution at individual codons in this study. However, this does not rule out the possibility that the modifications to *LDHA* on the mouse lemur terminal branch have adaptive functions.

The free-ratio model (codeml model 1) results shown in Figures 1 and 2 provide valuable information from which hypotheses can be developed. Having confirmed the absence of adaptive evolution of *LDHA* and *LDHB* in anthropoid primates, these analyses revealed significant changes in *LDHA* on the mouse lemur terminal branch (Fig.1). This terminal branch, however, includes four strepsirrhine families (Megaladapidae, Cheirogaleidae, Lemuridae, and Indriidae) (Schulke and Ostner 2007). Only genera in the family Cheirogaleidae are known to enter torpor (Schulke and Ostner 2007), therefore a study including species from each of these families would help resolve when the eight substitutions

occurred since the split of the mouse lemur and aye-aye lineages. If the substitutions are associated with torpor, we would expect to find most, if not all of these substitutions occurring following the split of Cheirogaleidae from the other families.

Many mammalian clades include species that utilize torpor as a means for coping with seasonal changes in food and/or water availability (e.g. marsupials, carnivores, rodents, and primates) (Melvin and Andrews 2009). This phenotypic convergence is likely achieved through variable molecular mechanisms, but there may be molecular convergence as well. Such convergence has been discovered between echolocating bats and dolphins, in modifications to the auditory system gene, *Prestin* (Liu et al. 2010). We suggest that other, heterothermic mammalian species may show similar modifications to *LDHA* as a means to shift from glycolysis to oxidative phosphorylation during periods of torpor. Future research may help identify whether these modifications to *LDHA* in mouse lemur impact the glycolytic rate in this species. In addition, comparing the mouse lemur *LDHA* sequence with other, heterothermic mammals may reveal a convergent molecular mechanism by which metabolism is adjusted to deal with seasonal environmental changes in resources.

MATERIALS AND METHODS

We PCR-amplified all seven coding exons for both *LDHA* and *LDHB* in *Gorilla gorilla, Pongo pygmaeus, Hylobates lar, Papio anubis, Theropithecus gelada, Mandrillus leucophaeus, Colobus guereza, Trachypithecus obscurus,*

Alouatta seniculus, Ateles geoffroyi, Callithrix jacchus, Cebus apella, Microcebus murinus, and Daubentonia madagascarensis (Appendix C1). Amplification and sequencing primers, as well as amplification conditions for each exon in each species are given in Appendices C6-C8. Bands of appropriate sizes were extracted and purified using QIAquick gel extraction kit (Qiagen) according to manufacturer's instructions. Clean bands were directly sequenced, whereas non-specific amplicons were ligated overnight at 4˚C or for 1 hour at room temperature in pGem T-Easy vectors (Promega). Vector-inserted amplicons were transformed by heat shock (42˚C) into DH5-α chemically-competent cells (Invitrogen), grown for 1.5 hours at 36˚C in LB, then grown overnight at 36˚C on LB plates made from 1L ddH2O, 25g LB, 15g agar, 5 mL 0.5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG), 128 μg X-Gal, and 100μg ampicillin. Positive colonies were selected and grown overnight at 36˚C in LB with ampicillin (100μg/mL). Plasmids extracted using Spin Miniprep Kit (Qiagen) using standard protocols.

Amplicons were sequenced, either directly or in vector at the Applied Genomics Technology Center at Wayne State University and the Research Technology Support Facility at Michigan State University. Chromatograms were visualized and aligned contigs assembled in Sequencher v4.7 (Gene Codes Corporation).

Additional mammalian and non-mammalian *LDHA* and *LDHB* sequences were acquired from GenBank, including human, chimp, mouse, rat, cow, horse, and opossum (Appendix C1). Individual exons were manually aligned and

concatenated in MacClade v4.08 (Maddison 2000). Marginal reconstruction of sequences using the primate (Goodman et al. 1998) and mammalian tree (Springer et al. 2003) were run using PAML 4.4 (Yang 1997), with one ω model (model 0), starting ω = 1, cleandata = 1, and RateAncestor = 1. Only those substitutions with likelihood values > 0.700 were considered significant. Branch-sites tests, model A, were evaluated relative to likelihood value from model A, fixed omega = 1 (Zhang et al. 2005). Grubbs' tests for statistical outliers were conducted online using QuickCalcs (GraphPad Software).

CHAPTER FIVE—ANCIENT ORIGIN OF PLACENTAL GROWTH HORMONE IN PRIMATES

Introduction

Mammalian species vary in terms of their rates of growth and development; for example, the normal length of gestation in mice is ≈20 days compared with 280 days in humans. Similarly, animals such as horses and cows walk shortly after being born, yet human infants require nearly a year of postnatal development before they reach this milestone. It is well appreciated that the actions of hormones, particularly growth hormones (GHs), shape the differences in rates of growth and development among species via the actions of the somatotrophic axis (Gluckman and Pinal 2002). Human disorders, including reduced stature and delayed sexual maturity, can result when the normal actions of GHs are disrupted (Dattani and Preece 2004; Zhou et al. 1997).

Humans belong to the group of primates called Anthropoidea, which can be further subdivided into catarrhines (Old World monkeys and apes, including humans) and platyrrhines (New World monkeys). Most anthropoids are characterized by prolonged gestation and delayed rates of maturation, with many anthropoid species having large brains relative to their body size (Allman 2000; Simpson 1945). These features have been advanced as the basis for increased social complexity and cognitive capacity in primates (Allman 2000; Martin 1990b; Simpson 1945). The genetic basis of these characteristic anthropoid primate phenotypes is unknown; however, fetal development depends on access to maternal resources during pregnancy. Indeed, it has recently been shown that

hemochorial placentation seen in anthropoids is associated with steeper brain-body allometry, faster prenatal brain growth, and slower prenatal body growth (Elliot and Crespi 2008). Moreover, it has been proposed that fetal acquisition of resources from the mother is mediated by peptides secreted by the placenta (Crespi and Semeniuk 2004; Haig 1996). Interestingly, there are several molecules uniquely produced by the placentas of anthropoid primates, including CG (Maston and Ruvolo 2002), siglecs (Brinkman-Van der Linden et al. 2007), and galectins (Than et al. 2009). Furthermore, placental GHs and placental lactogens have been implicated in fetal acquisition of maternal resources during anthropoid pregnancies (Haig 2008). Thus, study of the evolutionary history of genes uniquely shared among anthropoids can illuminate important aspects of human pregnancy and development.

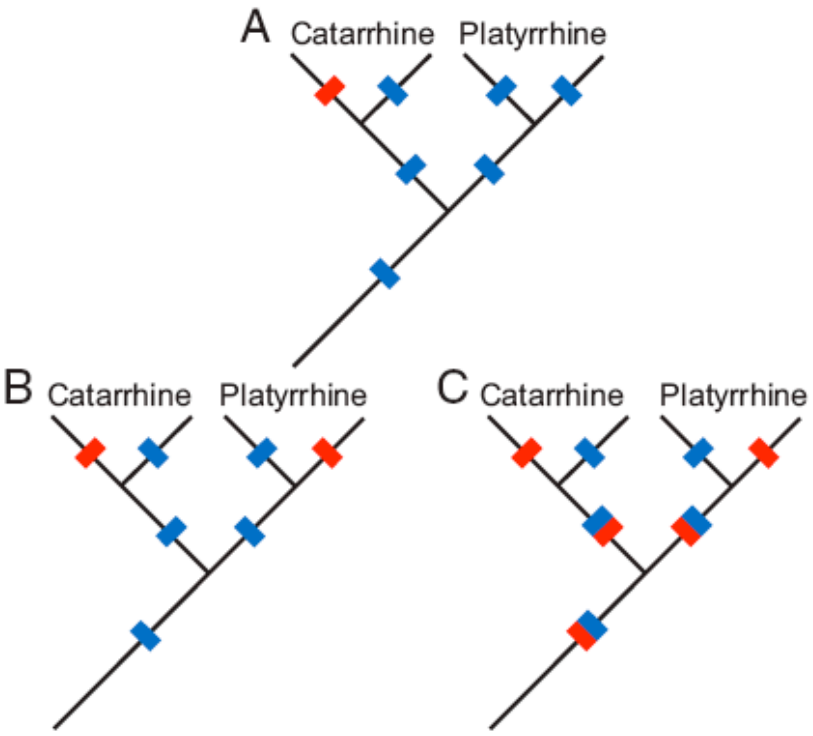
A cluster of 5 paralogous genes on human chromosome 17 (q23.3) encodes GHs and placental lactogens/chorionic somatomammotropins (CSHs). Similar clusters of paralogous genes have been found in all anthropoid species examined to date, although it has been shown that the platyrrhine and catarrhine gene clusters emerged independently via the tandem duplication process (Chen et al. 1989; Revol De Mendoza et al. 2004; Wallis and Wallis 2002). Most other mammal species have a single gene that encodes GH. Moreover, placental lactogens in nonanthropoids are derived from the prolactin gene family rather than the GH family (Goffin et al. 1996). Genes in the human (*GH2, CSH1, CSH2*, and *CSHL1*) (Chen et al. 1989) and rhesus macaque (Golos et al. 1993) clusters are transcribed in the placenta. These placenta-expressed genes play diverse

roles during pregnancy, from mediating trophoblast invasion (Lacroix et al. 2005) to regulating maternal resource availability for the developing fetus (Fleenor et al. 2005). Circulating placental GH serum concentrations have been associated with human pregnancy complications, including fetal growth restriction (McIntyre et al. 2000), impaired uteroplacental circulation (Schiessl et al. 2007), and preeclampsia (Mittal et al. 2007). The human gene *GH1* is expressed only in the pituitary, as is GH found in other mammals. As such, human *GH1* is assumed to retain the ancestral function of GH (Chen et al. 1989; Gonzalez Alvarez et al. 2006; Revol De Mendoza et al. 2004).

To evaluate GH evolution in mammals more systematically, it is necessary to know whether platyrrhine genes encoding GHs are also expressed in the placenta. Therefore, we isolated cDNA from the placenta of a platyrrhine Spider monkey and looked for GH sequences. Furthermore, we sought to examine the strength at which natural selection has acted on the platyrrhine and catarrhine genes. We predicted that if platyrrhine genes were not expressed in placenta, it is unlikely that the last common ancestor (LCA) of anthropoids would have possessed a single gene that was expressed in both the placenta and pituitary. Instead, we reasoned that if platyrrhine GH genes were not expressed placentally, it is only during catarrhine evolution that the ability to mediate physiological exchange through placental expression of GHs would have emerged (Fig. 1A). Conversely, if we found that these genes were expressed in the Spider monkey placenta, the implication would be that placental expression was gained convergently in both groups (Fig. 1B) or that placental expression

preceded the independent series of gene duplications in catarrhines and platyrrhines (Fig. 1C). Finally, studies of natural selection's effects on protein coding genes can be used to identify candidate sites of functionally important amino-acid residues. Adaptive changes in genes related to the immune system have been shown to affect host pathogen interactions (Wlasiuk et al. 2009), and it is possible that adaptive evolution in placental proteins similarly affects maternal-fetal interactions.

Figure 1. Scenarios for the evolution of GH expression in the placenta



Blue rectangles represent pituitary expression, and red rectangles represent placental expression. (A) GH genes gained placenta expression in catarrhines after divergence from platyrrhines. (B) Parallel evolution resulted in independently derived placenta expression in catarrhines and platyrrhines. (C) The LCA of anthropoids expressed GH in the placenta.

Results and Discussion

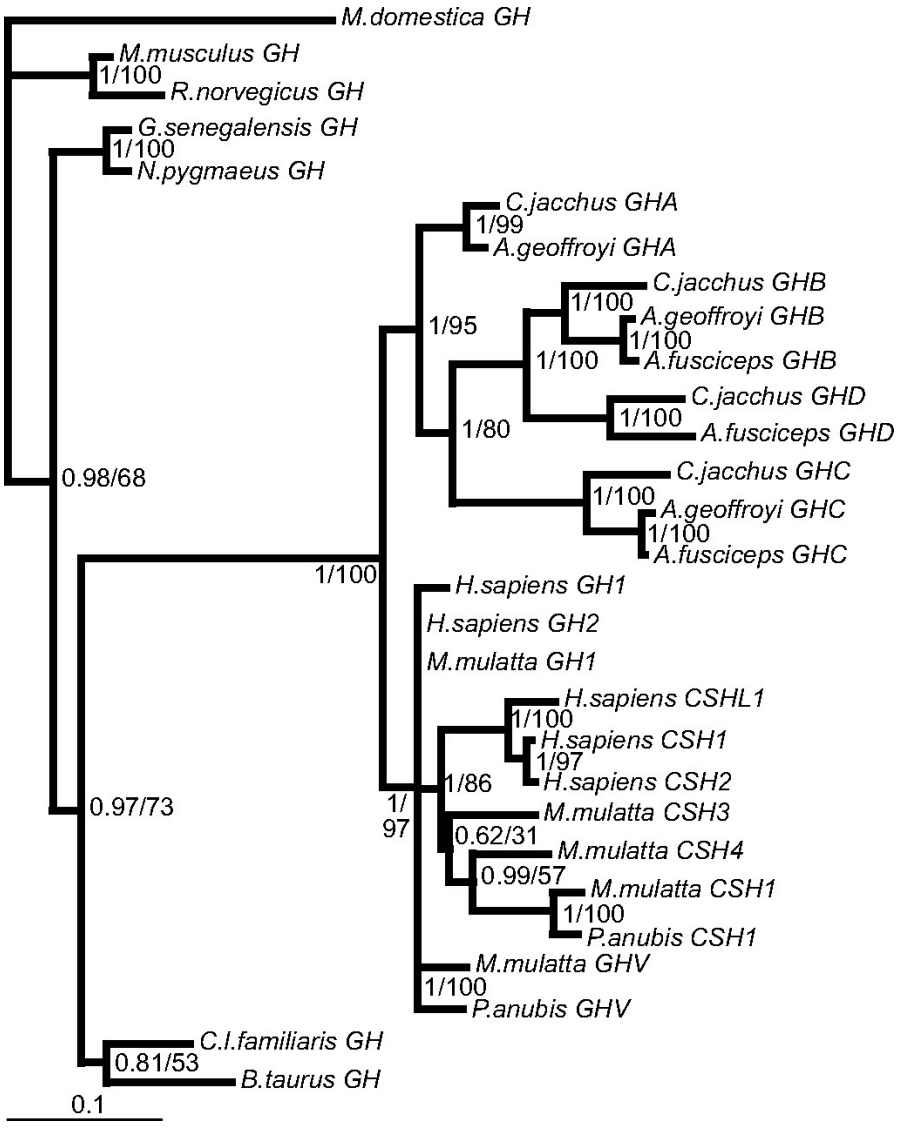
Placental Transcripts and Characterization of GH Genes.

As in the human, macaque, and baboon, GH genes are transcribed in the placenta of platyrrhines. Using RT-PCR, we amplified, cloned, and sequenced 10 distinct transcripts from at least 3 different genes from placental tissue of the Spider monkey [*Ateles fusciceps*; Appendix D1], for a total of 208 individual clones (Appendix D2). Comparison of these previously unreported cDNA sequences with previously reported Spider monkey genomic DNA sequences revealed that *GHB* (i.e., *GH2*, AF374235) and *GHC* (i.e., AY435434) (Revol De Mendoza et al. 2004) are transcribed in the placenta. The *GHB* transcripts are rare (2/208 = 1%). In contrast, *GHC* transcripts are relatively abundant (107/208 = 51%). In addition to these previously described genes, we identified an abundantly transcribed (99/208 = 48%) GH gene, *GHD* (EU935080; Appendix C2). We found no evidence that the pituitary-expressed platyrrhine *GHA* (i.e., *GH1*) (Liu et al. 2001) is transcribed in the Spider monkey placenta. To infer intron-exon boundaries for the placentally transcribed New World monkey genes, we compared our transcripts with the previously sequenced marmoset genomic GH gene cluster (Wallis and Wallis 2002).

A complete description of the splicing patterns is provided in Appendix D Supplementary Text and depicted in Appendix D1. In summary, both *GHC* and *GHD* are alternatively spliced. Vertebrates share a canonical 5-exon organization of GH. Two transcript variants retain intron 4, similar to variants found in human placenta (*hGH2*) and testes (*hCSH1*) (MacLeod et al. 1992; Untergasser et al.

2000), as well as in the cow pituitary *cGH* (Hampson and Rottman 1987). The human variants encode membrane-bound proteins (Cooke et al. 1988; Untergasser et al. 2000) and are known to increase their expression during human pregnancy up to parturition (Untergasser et al. 2000).

Figure 2. Phylogenetic tree of GH genes



The tree was inferred using MrBayes v.3.1. Branch lengths were scaled to the percentage of nucleotide substitutions. Nodes were labeled with Bayesian posterior probability/ML bootstrap values. Common names and accession numbers are listed in Appendix D2, and ML methods are provided in Appendix D Supplemental Text.

Phylogenetic Inference.

Fig. 2 depicts the optimal Bayesian tree derived from the multiple sequence alignment of mammalian GH-related sequences (ln L = -6,177.60). Accession numbers, gene symbols, and taxon abbreviations are shown in Appendix D2. The anthropoid GH genes cluster together, with the platyrrhine GH genes falling in one clade and the catarrhine GH genes falling in another clade. Confirming previous studies (Chen et al. 1989; Gonzalez Alvarez et al. 2006; Revol De Mendoza et al. 2004; Wallis and Wallis 2002), our results show that platyrrhine and catarrhine GH clusters are likely the products of an independent series of duplications in each of these 2 major anthropoid clades and that a single GH gene existed at the time of the LCA of anthropoid primates. We refer to platyrrhine paralogous genes *GHA* and *GHB* and catarrhine paralogs *GH1* and *GH2* rather than having *GH1* and *GH2* genes in both clades. We continue use of the platyrrhine gene symbol *GHC* (Appendix D2). *GHD* is a previously undescribed gene.

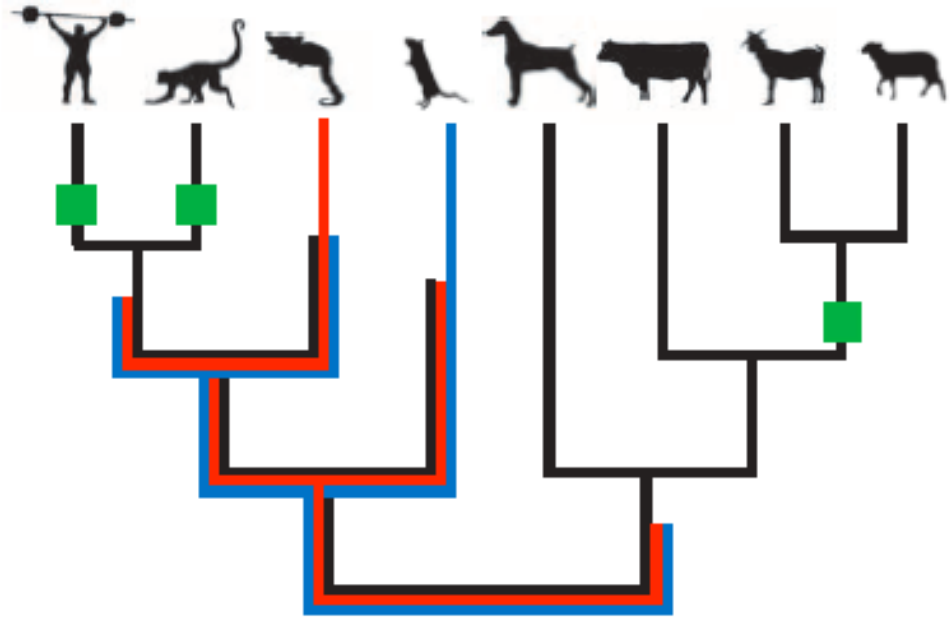
Within catarrhines, the only well-resolved clades are the clustering of CSH genes (i.e., the clade containing, *Macaca mulatta CSH1, Homo sapiens CSH1*, and related sequences) and, within this clade, the subclade of human CSH genes (i.e., *H. sapiens CSH1, CSH2*, and *CSHL1*). Gene conversion has

occurred in catarrhines (Chen et al. 1989; Revol De Mendoza et al. 2004), and this could explain the lack of resolution observed in this part of the tree.

In platyrrhines, the relations among GH genes are well resolved (Fig. 2). The placenta-expressed GH genes (i.e., *GHB, GHC, GHD*) form a clade to the exclusion of the pituitary expressed GH gene, *GHA*. Within the placenta-expressed genes, the sequences from *GHB* and *GHD* cluster together to the exclusion of those from *GHC*.

Outside of the anthropoid clade, the gene and species trees are incongruent. Although anthropoid primates are monophyletic, we were unable to recover monophyletic primate and Euarchontoglires clades. Instead, the clade consisting of cow and dog (i.e., Laurasiatheria) was found to be the sister group of anthropoids. This Laurasiatheria + Anthropoidea clade was next joined by a clade of strepsirrhine primates (loris and galago), and ultimately joined by the rodent clade. The gene tree is significantly better than the species trees (Appendix D3 and D4).

Figure 3. Gain and loss of GH genes in placental mammals



The gene tree and species tree were reconciled (Goodman et al. 1979). At least 2 gene duplications occurred before the time of the LCA of Boreoeutheria, and at least 7 subsequent gene losses occurred in descendant lineages. Three boreoeutherian GH genes are depicted in black, blue, and red. Truncated lines represent gene loss. Green boxes indicate placental expression. Additional gene duplications and losses that occurred in anthropoid primates (i.e., Fig.2) are not shown. Images depict (*Left* to *Right*) human, Spider monkey, galago (strepsirrhine), rat, dog, cow, goat, and sheep.

We reconciled the gene and species trees according to the methods outlined by Goodman et al. (Goodman et al. 1979), and this reconciliation requires at least 2 gene duplications and 7 gene loss events early in placental mammalian history (Fig. 3). In this scenario, 3 GH paralogs existed at the time of the LCA of Boreoeutheria (i.e., the LCA of the primates, rodents, carnivores, and bovids included in our study). One of these copies is maintained in anthropoids and laurasiatherians, another is maintained in rodents, and the third is maintained in strepsirrhines. The addition of 2 gene gains and 7 gene losses

results in a tree with an identical length as that of the species tree. These findings do not unambiguously favor either the gene or species tree; as such, we undertook all analyses of adaptive evolution on both tree topologies. We do note, however, that an independent piece of evidence supporting the scenario outlined in Fig. 3 is the presence of intron 4-containing transcript variants in anthropoids and artiodactyls (Hampson and Rottman 1987), variants not found in other mammals.

The possibility of multiple GH genes in the boreoeutherian LCA raises unique questions regarding the evolution of anthropoid GH genes. Rather than gene losses, gene conversions could have resulted in multiple GH copies that are indistinguishable from one another. We can, however, feel confident that no significant gene conversions occurred among the New World monkey placental GH coding sequences. If there had been, the GHs of each New World monkey genus would group together before joining the other GHs.

Table 1. ω values and significance tests for different models of GH evolution

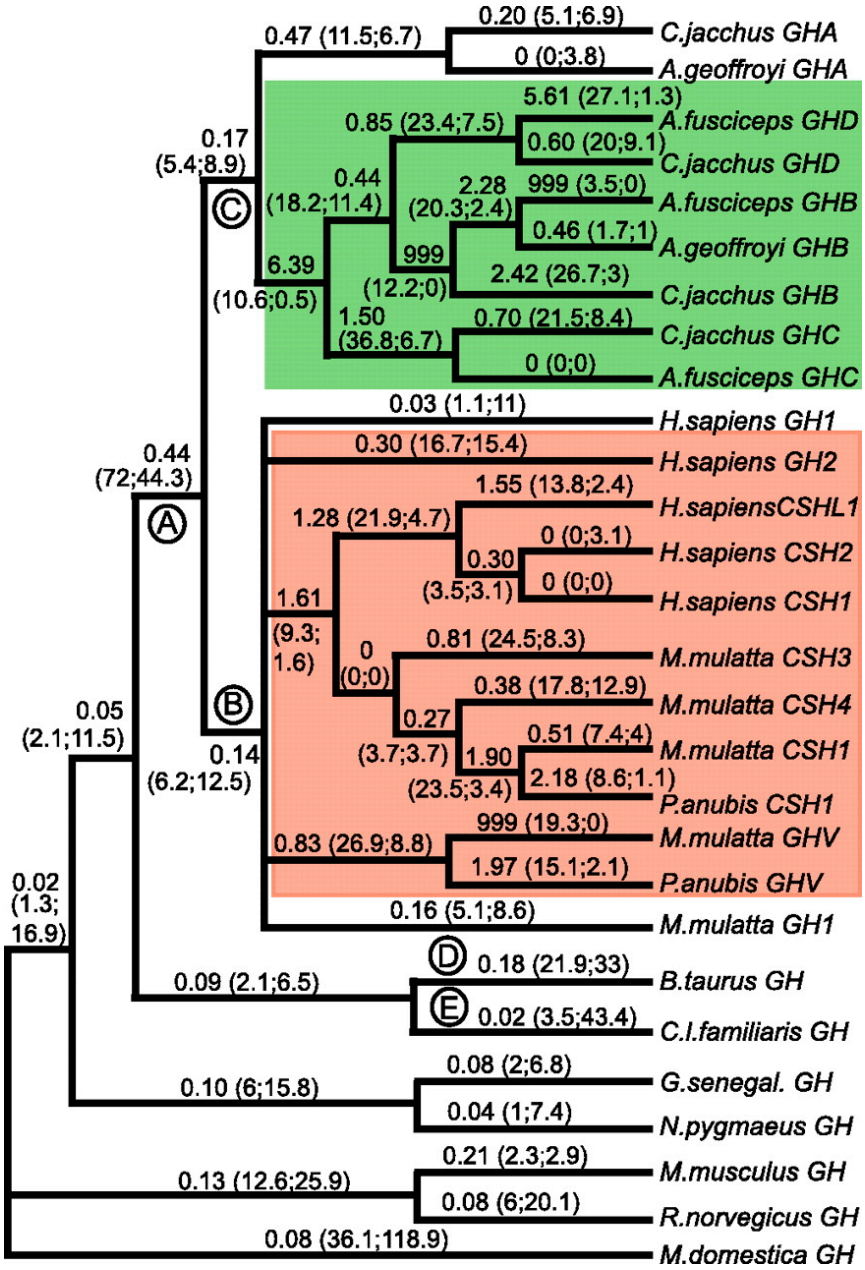
Model	Catarrhine placental GHs ω_{cpl}	Platyrrhine placental GHs ω_{ppl}	Pituitary GHs ω_{pi}	Likelihood (-ln L)	2Δln L	P^*
Fixed ω	0.36	0.36	0.36	-5,799.23	N/A	N/A
Free ratio	Variable	Variable	Variable	-5,641.56	315.33	$1.39E^{-40}$
Branch-based with 2 ω values	0.95	0.95	0.13	-5,699.81	198.84 2ω:1ω	$3.75E^{-45}$
Branch-based with 3 ω values	0.79	1.16	0.13	-5,698.20	3.22 3ω:2ω	0.07

*P value based on χ^2 test, with free-ratio and branch-based models (model 2, 2ω values) compared with fixed ω model (1ω value) and branch-based model (model 2, 3ω values) compared with branch-based model (2ω values). N/A, not applicable.

Placental Expression and Selection.

To test the hypothesis that GH genes underwent molecular adaptations during primate evolution, we analyzed the per site ratio of nonsynonymous (dN) to synonymous (dS) substitutions on each branch of the optimal Bayesian tree. Overall, GH genes exhibit slight signatures of purifying selection. The background ratio of dN to dS substitutions per site ω value is 0.36. However, the free ratio model (ln L = -5,641.56), which assumes independent ω values for each branch, fits the data significantly better than the fixed ω model (χ^2 P = 1.39 x 10^{-40}; Table 1), indicating significant variation in ω values across the different branches (Fig. 4), and provides evidence for positive selection (Fig. 4). The 14 branches exhibiting signals for positive selection have ω values ranging from 1.28 (stem human CSHs) to 999 (2 platyrrhine branches and 1 catarrhine branch). Remarkably, all branches exhibiting ω values >1 are on branches leading to and including placenta-expressed GH genes. In contrast, the branches leading to and including pituitary-expressed GH genes have relatively low ω values. Moreover, our results challenge previous interpretations that consider human *CSHL1* a pseudogene (Chen et al. 1989; Cooke et al. 1988; Revol De Mendoza et al. 2004) because of a high ω value (1.55) and ≈14 inferred dN substitutions without a single nonsense or frameshift substitution.

Figure 4. Adaptive evolution in GH genes



The free ratio model (codeml model 1) ω values of the Bayesian gene tree and the ML estimates of the number of dN (N*dN); dS (S*dS) substitutions are shown along each branch. Placenta-expressed catarrhine GH genes and their ancestral lineages are boxed in salmon, and placenta-expressed platyrrhine GH genes and their ancestral lineages are boxed in green. Branches A-E were used to test

hypotheses regarding divergence times (see the text). Values of 999 indicate branches with only dN substitutions, and values of 0.01 indicate branches with only dS substitutions. Scientific names and accession numbers are listed in Appendix D2.

To study differences in selection pressures between placenta and pituitary-expressed GH genes further, we conducted likelihood ratio tests comparing a one-ratio model to an alternative model, assigning one ω value (ω_{pl}) to the internal and terminal branches of the placenta-expressed GH genes (i.e., both the green- and salmon-shaded lineages in Fig. 4) and another ω value (ω_{pi}) to the internal and terminal branches of the pituitary expressed GH genes. Using this approach, placental genes and their ancestral lineages had a ω_{pl} value of 0.95, a value over 7 times greater than that assigned to the pituitary-expressed branches (ω_{pi} = 0.13). This model (model 2, ln L = -5,699.81) fits the data significantly better ($P < 0.001$) than the 1-ratio model (model 0, ln L = -5,799.23; Table 1). This finding provides some evidence that branches included in the ω_{pl} group evolved adaptively; however, because ω_{pl} <1 in this model, we cannot rule out a relaxation of functional constraint.

We implemented a further test to distinguish selection pressures between placenta expressed GH genes in catarrhines and platyrrhines. In this test, we assigned one ω value to internal and terminal branches of catarrhine placental GH genes (ω_{cpl}; salmon shading in Fig. 4), another ω value to internal and terminal branches of platyrrhine placental GH genes (ω_{ppl}; green shading in Fig. 4), and yet another ω value to all other GH branches (ω_{pi}; no shading in Fig. 4). Catarrhine placental GH genes and their lineages had the ω_{cpl} value of 0.79,

platyrrhine placental GH genes and their lineages had the ω_{ppl} value of 1.16, and all other GH branches had the ω_{pi} value of 0.13. This branch-based model (model 2 with 3 ω values, ln L = -5,698.20) is not significantly better than the branched-based model with 2 ω values (P = 0.07), suggesting that the selective forces acting on placenta-expressed GH genes are similar in platyrrhines and catarrhines (Table 1).

Rapid dN Substitution in GH on the Branch Descending to the LCA of Anthropoids.

The branch leading to the LCA of anthropoids (branch A in Fig. 4) does not exhibit signals of positive selection (ω = 0.44) even though one-quarter of the translated amino acids was replaced. This is attributable to the concomitant high number of inferred dS substitutions (S*dS = 44.3; Fig. 4). To explore this further, we evaluated the rates of change on the phylogenetic tree for both dN and dS rates (Table 2). Our rationale for this procedure was that the dS rates should more closely reflect neutral expectations, and thus should vary less between branches than dN rates on a substitution/site/year basis. We calculated these rates using the arrangements depicted in both the gene tree (Fig. 4) and the species tree (Appendix D5). In addition to the branch leading to the LCA of anthropoids, we examined the branch leading to the catarrhine LCA (branch B), the branch leading to the platyrrhine LCA (branch C), the cow terminal branch (branch D), and the dog terminal branch (branch E). The estimated amounts of

evolutionary time for each of these branches as well as the inferred substitution rates are listed in Table 2.

Table 2. Rates of nonsynonymous and synonymous substitutions/site/year on key branches

Branch leading to: (branches A-E in Fig.3 and Appendix D2)	Inferred branch time (in million years)	dN	dS	dN/year X 10^{-9}	dS/year X 10^{-9}
Species tree					
Anthropoid LCA (A)	23 (Goodman 1998)	0.1520	0.3793	6.61	16.49
Catarrhine LCA (B)	15 (Goodman 1998)	0.0130	0.0976	0.87	6.51
Platyrrhine LCA (C)	15 (Goodman 1998)	0.0110	0.0634	0.73	4.23
Cow (D)	82 (Springer 2003)	0.0457	0.2357	0.56	2.87
Dog (E)	82 (Springer 2003)	0.0066	0.3415	0.08	4.16
Gene tree					
Anthropoid LCA (A)	54 (Goodman 1998, Spinger 2003)	0.1483	0.3341	2.75	6.19
Catarrhine LCA (B)	15 (Goodman 1998)	0.0128	0.0945	0.85	6.30
Platyrrhine LCA (C)	15 (Goodman 1998)	0.0111	0.0667	0.74	4.45
Cow (D)	82 (Springer 2003)	0.0452	0.2486	0.55	3.03
Dog (E)	82 (Springer 2003)	0.0073	0.3272	0.09	3.99

Inferred divergence dates are from (Goodman et al. 1998) and (Springer et al. 2003).

The dS substitution rates along the species tree range from 2.87–6.51 substitutions/site/year X 10^{-9} for branches B–E. On the species tree, branch A encompasses ≈23 million years from the time of the LCA of primates (63 mya) to the time of the LCA of the anthropoids (40 mya). The dS substitution rate on this

branch is 16.49 dS substitutions/site/year X 10^{-9} (Table 2), which is significantly faster than the rates for the other 4 branches (Student's *t* test with Bonferroni correction, $P < 0.005$). On the gene tree, branch A encompasses ≈54 million years from the time of the LCA of Laurasiatheria and anthropoids (≈94 mya) to the LCA of anthropoids (40 mya). The dS substitution rate on this branch is 6.19 dS substitutions/site/year X 10^{-9}, a rate that is not significantly different from the rates on the other 4 branches (Student's *t* test, $P > 0.2$; Table 2). This supports our reconciliation method by placing the age of this branch at the LCA of Boreoeutheria (Fig. 2).

Functional Consequences of Amino-Acid Replacements.

The primary mechanism by which human GH genes regulate resource availability is through endocrine regulators of fetal growth and development, such as the IGF system (Fleenor et al. 2005; Gluckman and Pinal 2002; McIntyre et al. 2000). We note that in contrast to GH genes from nonprimate mammals, human GH genes function via interactions with both GH receptor and prolactin receptor (PRLR) (Peterson and Brooks 2004). GH1 has been shown to regulate both IGF-1 and IGF-2 postnatally (Fleenor et al. 2005; Rodriguez et al. 2007). GH treatment results in an increase in IGF-2 secretion in human fetal hepatocytes (Goodyer et al. 2001), and GH2 levels correlate with maternal IGF-1 levels starting in mid-gestation (Chellakooty et al. 2004). In addition, PRLR, which can bind GH2 and the CSHs, has been shown to regulate IGF-2 expression during gestation (Viengchareun et al. 2008). Moreover, PRLR signaling is essential for

implantation in mice (Ormandy et al. 1997), and GH2 has been shown to increase extravillous cytotrophoblast invasiveness (Lacroix et al. 2005). Previous evolutionary studies have suggested that the gain of placental expression was coincident with the acquisition of GH-PRLR activation (Goffin et al. 1996; Haig 1993). At least 8 amino-acid replacements essential for human GH-PRLR binding, including Q18H, A25F, I45F+L, T62S, G63N, D65E, K167R, and Y176F (Cunningham and Wells 1991; Peterson and Brooks 1997), occurred on the branch leading to the anthropoid LCA (branch A in Fig. 4 and Appendix D5). The coincident adoption of GH-PRLR activation and placental expression could provide a way for the anthropoid fetus to obtain greater access to maternal nutrients by inducing maternal insulin resistance (Haig 1993), especially during the prolonged gestations (Martin 1990b) as suggested by the maternal-fetal conflict hypothesis (Haig 1993).

Implications of This Study.

In this study, we sequenced GH-like transcripts from the placenta of the Brown-Headed Spider monkey, *A. fusciceps*. Thus, all anthropoids (i.e., catarrhines and platyrrhines) express GH genes placentally. We identified 10 distinct transcripts from at least 3 different genes. The major findings of this study are that (*i*) multiple platyrrhine GH genes are transcribed in the placenta, (*ii*) there is evidence that placenta-expressed GH genes have been subjected to positive selection in both platyrrhines and catarrhines, and (*iii*) pituitary expressed anthropoid GH genes have been constrained by purifying selection.

In addition, we provide evidence based on gene-species tree reconciliation and dS substitution rates suggesting the possibility that anthropoid primates and laurasiatherians share a GH gene copy, whereas strepsirrhine primates and rodents each maintain separate paralogous genes (Fig. 3). The GH family is similar to the CG (CGs) family in that both families include placenta-expressed hormones that are only found in anthropoids. However, CG evolution appears to be less complicated than that of the anthropoid GHs, because the evidence for duplication of CG from its luteinizing hormone progenitor likely occurred between 58 and 40 mya (Maston and Ruvolo 2002).

In the present study, we propose that in addition to the gain of PRLR binding (Haig 1993; Haig 2008), placental expression potentially existed at the time of the LCA of extant anthropoids. At least 8 amino-acid replacements that occurred on the lineage leading to the LCA of anthropoid primates could have conferred the ability for anthropoid GHs and CSHs to bind PRLR, thus enabling GH signaling at the maternal-fetal interface. PRLR is expressed on the maternal side of the maternal-fetal interface (Jones et al. 1998). Taken together, these findings suggest that the LCA of anthropoids could use GH-PRLR signaling at the maternal-fetal interface and that this ability has been maintained in descendant lineages by subfunctionalization after gene duplication. That there are more than 2 duplicates in both platyrrhines and catarrhines suggests that the single-copy ancestral anthropoid gene had other as yet undescribed functions that were subsequently subfunctionalized or that some of the more recent gene duplicates have gained previously undescribed functions unique to platyrrhines

and catarrhines, respectively. In contrast to the pituitary expressed GH genes, the placental GH genes have a much higher rate of dN substitution. The relatively ancient origin of placental expression, combined with the complicated history of gene gain and loss in mammals, suggests that the GH gene family has a longer history involving maternal-fetal interactions and prenatal growth than has been previously described.

Materials and Methods

Nucleotide Extraction.

Villous tissue was dissected from membranes, and total RNA was isolated using TRIzol Reagent (Invitrogen) followed by the RNeasy Kit (Qiagen) according to the manufacturers' recommendations. mRNA was isolated from totalRNAusing the MicroPoly(A) Purist Kit (Applied Biosystems). cDNA libraries were constructed using the SMART cDNA Library Construction Kit (Clontech), and DNA was isolated from transformed clones using the DirectPrep96 Miniprep Kit (Qiagen).

Amplification of Placental Transcripts.

We used 3' and 5' RACE-ready cDNA from villous and membranous tissue of the placenta of the Brown-Headed Spider monkey (*A. fusciceps*) as well as from the placenta of the Olive baboon (*Papio anubis*). Purified products were ligated overnight at 4 °C into pGEM T-Easy vectors (Promega), transformed by heat shock (42 °C) into DH5α chemically competent cells from Invitrogen, and grown on LB plates made from 1 L of ddH20, 25 g of LB, 15 g of agar, 5mL of

0.5-mM IPTG, 128 μg of X-Gal, and 100 μg of ampicillin. Positive colonies were selected and grown for 12–16 h at 36 °C in 3 mL of LB/ampicillin (100 μg/mL) liquid medium. Plasmids were extracted using the Spin MiniPrep Kit (Qiagen) according to the manufacturer's instructions.

Sequence Assembly, Alignment, and Consensus Sequence Construction.

Cloned products were sent to the Research Technology Support Facility at Michigan State University for sequencing. Chromatograms were imported into Sequencher v4.6 (Gene Codes Corporation). The reads from 5' and 3' RACE sequences overlapped by ≈400 bp. Consensus sequences for *GHB*, *GHC*, and *GHD* were constructed based on majority rule at each nucleotide position. The number of colonies sequenced for each transcript type is listed in Appendix D6. Sequences have been deposited in GenBank: EU935072-EU935081 (Spider monkey) and FJ041322 FJ041323 (Anubis baboon).

We aligned our individual full-length transcripts from *A. fusciceps* (*GHB*, *GHC*, and *GHD*), 2 previously undescribed GH transcripts isolated from Olive baboon (*P. anubis*) placenta, and publicly available sequences (Appendix D2). The marmoset cluster has been characterized genomically (Wallis and Wallis 2006), and the putative orthologous relations between genes from this cluster and the *Ateles* GH gene transcripts were identified via BLAST (Tatusova and Madden 1999). Alignments of nucleotide sequences were visualized, and reading frame integrity was checked using MacClade v4.08 (Maddison and Maddison 2000). The alignment is included in Appendix D8.

Phylogenetic Inference.

Phylogenetic trees were inferred with MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) using the canonical transcripts for each GH gene and species. We used MrModeltest v2.3 (Nylander 2004) to estimate the best-fit model for the sequences. Based on the Akaike Information Criterion, a SYM + γ model was selected with γ-distribution shape parameter α = 1.6030, an R matrix (1.0959, 5.0778, 1.0169, 1.4816, and 3.7177), and equal base frequencies. One cold chain and 3 hot chains were run simultaneously for 1 million generations, with sampling every 100 generations; the initial 2,500 samples were discarded as burnin, and convergence between chains was checked.

Branch-Based Tests of Positive Selection.

PAML 3.15 (Yang 1997) was used to investigate selection pressures (i.e., dN/dS or ω) among lineages. This ratio indicates purifying selection, neutral evolution, or positive selection when $\omega<1$, $\omega = 1$, and $\omega>1$, respectively (Yang 1997). Unable to amplify *GHA* transcripts from Spider monkey placental cDNA, previously published marmoset and Spider monkey *GHA* sequences represented platyrrhine pituitary-expressed GH (Liu et al. 2001). Likelihood values were calculated 3 times per model, with different starting values for ω(0.5, 1, and 2). Alternative models were compared by likelihood ratio tests, and models were considered significantly different if $P < 0.05$ (Chen et al. 2008). Please refer to Appendix D Supplemental Text for ancestral reconstruction methods, and Appendix D7 for amino acid replacements on stem anthropoid (branch A).

Substitution Rate Analyses.

We calculated rates of dS and dN substitutions on branches of both the gene and species trees. We used the branch leading to the LCA of anthropoids, the LCA of catarrhines, the LCA of platyrrhines, and the cow and dog terminal branches (branches A–E, respectively, in Fig. 4 and Appendix D1). Divergence times were from Goodman et al. (Goodman et al. 1998) for primate branches and from Springer et al. (Springer et al. 2003) for the other mammalian branches. We used the dS and dN values from the PAML model 1 output. Rates are reported as (substitutions/site/year) x 10^{-9}. Differences among rates were tested with the Student's *t* test (2-sample, 1-tailed) assuming unequal variance.

APPENDIX A

Appendix A1. Nucleotide frequencies for total dataset and within conserved motifs of each width

	A	C	G	T
Total dataset	24.59	25.11	25.23	25.07
Width 5	23.54	26.80	26.77	22.89
Width 6	23.51	25.91	25.71	24.87
Width 7	23.28	26.01	26.01	24.70
Width 8	23.63	25.62	25.57	25.18
Width 9	24.19	25.33	25.23	25.25
Width 10	24.87	25.08	24.04	26.02
Width 11	25.46	23.90	23.73	26.91
Width 12	25.68	23.66	23.22	27.44

Appendix A2. Linear regression of individual nucleotides across conserved motif widths

Nucleotides	Slope significance (p-value)	t-value	slope	r^2
A	p<0.001	6.01	0.36	0.86
C	p<0.001	8.65	-0.42	0.93
G	p<0.001	9.12	-0.49	0.93
T	p<0.001	7.56	0.55	0.90

Degrees of freedom is 6

Appendix A3. Linear regression of dinucleotide frequency and motif width

Dinucleotide	Slope significance (p-value)	t-value	slope	r^2
AA	p<0.001	56.273	0.042	0.391
AC	p<0.001	13.815	0.051	0.489
AG	p<0.05	2.452	0.011	0.350
AT	p<0.001	21.756	0.069	0.461
CA	p<0.001	18.484	0.044	0.430
CC	p<0.001	24.874	0.022	0.392
CG	p<0.001	6.489	0.041	0.410
CT	p<0.001	6.482	0.026	0.370
GA	p<0.001	21.517	0.029	0.355
GC	p<0.001	11.369	0.039	0.352
GG	p<0.001	19.851	0.022	0.335
GT	p<0.001	17.561	0.051	0.356
TA	p<0.001	51.983	0.054	0.381
TC	p<0.001	17.416	0.034	0.371
TG	p<0.001	22.755	0.048	0.369
TT	p<0.001	33.983	0.048	0.373

Degrees of freedom is 6

Appendix A4. Dinucleotides and observed frequency of preceding and following nucleotide

Dinucleotides with variable neighbors	p-value	Minimum chi-square value (width)
NAA	p>0.250	2.80 (width 11)
AAN	p<0.025	11.33 (width 11)
NAC	p<0.001	28.55 (width 12)
ACN	p<0.001	26.60 (width 12)
NAG	p<0.001	40.44 (width 12)
AGN	p<0.001	18.75 (width 12)
NAT	p<0.001	26.39 (width 10)
ATN	p<0.001	30.09 (width 11)
NCA	p<0.010	11.82 (width 12)
CAN	p>0.750	1.02 (width 9)
NCC	p<0.025	9.94 (width 12)
CCN	p<0.001	19.90 (width 12)
NCG	p<0.001	16.96 (width 12)
CGN	p<0.001	37.06 (width 12)
NCT	p<0.001	29.36 (width 12)
CTN	p<0.001	38.67 (width 12)
NGA	p<0.001	81.51 (width 12)
GAN	p<0.05	7.88 (width 11)
NGC	p<0.001	16.90 (width 12)
GCN	p<0.010	11.38 (width 9)
NGG	p<0.010	12.28 (width 12)
GGN	p<0.005	13.77 (width 12)
NGT	p<0.001	32.08 (width 12)
GTN	p<0.001	18.90 (width 12)
NTA	p<0.001	29.97 (width 12)
TAN	p<0.001	62.97 (width 11)
NTC	p<0.025	11.20 (width 12)
TCN	p<0.001	66.33 (width 12)
NTG	p>0.100	4.65 (width 10)
TGN	p<0.025	10.43 (width 11)
NTT	p<0.005	14.02 (width 11)
TTN	p>0.950	0.23 (width 11)

Degrees of freedom is 3.

Appendix A5. Nucleotide frequencies at individual positions across all widths

	Width 5	Width 6	Width 7	Width 8	Width 9	Width 10	Width 11	Width 12
A1	0.222	0.222	0.226	0.222	0.222	0.240	0.237	0.230
C1	0.325	0.294	0.265	0.262	0.251	0.222	0.211	0.209
G1	0.205	0.214	0.253	0.259	0.276	0.275	0.276	0.290
T1	0.248	0.270	0.256	0.258	0.252	0.263	0.275	0.270
A2	0.207	0.215	0.213	0.227	0.214	0.200	0.221	0.219
C2	0.274	0.269	0.259	0.253	0.266	0.277	0.253	0.243
G2	0.269	0.249	0.264	0.258	0.255	0.268	0.247	0.246
T2	0.250	0.267	0.264	0.262	0.265	0.255	0.280	0.292
A3	0.238	0.222	0.222	0.245	0.264	0.271	0.257	0.262
C3	0.265	0.259	0.261	0.266	0.248	0.246	0.247	0.233
G3	0.264	0.260	0.258	0.246	0.237	0.213	0.236	0.230
T3	0.234	0.259	0.258	0.243	0.252	0.270	0.260	0.275
A4	0.258	0.243	0.231	0.240	0.248	0.263	0.263	0.273
C4	0.271	0.266	0.265	0.264	0.276	0.250	0.237	0.230
G4	0.273	0.253	0.254	0.244	0.231	0.227	0.221	0.238
T4	0.198	0.238	0.249	0.252	0.245	0.261	0.280	0.259
A5	0.253	0.254	0.244	0.228	0.248	0.263	0.268	0.250
C5	0.205	0.254	0.264	0.261	0.259	0.259	0.256	0.265
G5	0.327	0.268	0.254	0.249	0.240	0.216	0.215	0.216
T5	0.214	0.224	0.239	0.262	0.253	0.262	0.261	0.269
A6		0.255	0.250	0.229	0.238	0.253	0.237	0.253
C6		0.213	0.262	0.255	0.238	0.250	0.255	0.230
G6		0.298	0.261	0.267	0.259	0.237	0.222	0.229
T6		0.235	0.227	0.249	0.266	0.260	0.286	0.287
A7			0.243	0.241	0.245	0.236	0.269	0.272
C7			0.245	0.249	0.229	0.236	0.201	0.236
G7			0.277	0.267	0.267	0.248	0.248	0.191
T7			0.236	0.243	0.259	0.280	0.283	0.302
A8				0.259	0.238	0.247	0.267	0.262
C8				0.240	0.262	0.229	0.220	0.209
G8				0.255	0.268	0.260	0.240	0.248
T8				0.245	0.231	0.263	0.273	0.282
A9					0.262	0.245	0.244	0.245
C9					0.252	0.291	0.272	0.255
G9					0.237	0.237	0.238	0.242
T9					0.250	0.227	0.245	0.259
A10						0.270	0.271	0.263
C10						0.249	0.240	0.235
G10						0.223	0.231	0.223
T10						0.258	0.258	0.279
A11							0.267	0.279
C11							0.237	0.250
G11							0.238	0.208

	Width 5	Width 6	Width 7	Width 8	Width 9	Width 10	Width 11	Width 12
T11							0.258	0.263
A12								0.273
C12								0.243
G12								0.226
T12								0.257

First column includes nucleotide followed by position within width

Appendix A6. Nucleotide frequencies at the last position for all widths

	Width 5	Width 6	Width 7	Width 8	Width 9	Width 10	Width 11	Width 12
A	0.253	0.255	0.243	0.259	0.262	0.270	0.267	0.273
C	0.205	0.213	0.245	0.240	0.252	0.249	0.237	0.243
G	0.327	0.298	0.277	0.255	0.237	0.223	0.238	0.226
T	0.214	0.235	0.236	0.245	0.250	0.258	0.258	0.257

Appendix A7. Reverse complement analysis of trinucleotides

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	AAA	13613	W+=7
6mers	AAA	5871	W-=29
7mers	AAA	1160	p<=0.1484
8mers	AAA	413	NOT SIGNIFICANT
9mers	AAA	255	
10mers	AAA	208	
11mers	AAA	165	
12mers	AAA	144	
5mers	TTT	13394	
6mers	TTT	6543	
7mers	TTT	1337	
8mers	TTT	475	
9mers	TTT	283	
10mers	TTT	215	
11mers	TTT	174	
12mers	TTT	147	
5mers	AAC	2381	W+=12
6mers	AAC	1524	W-=24
7mers	AAC	384	p<=0.4609
8mers	AAC	211	NOT SIGNIFICANT
9mers	AAC	154	
10mers	AAC	132	
11mers	AAC	105	
12mers	AAC	85	
5mers	GTT	2101	
6mers	GTT	1707	
7mers	GTT	435	
8mers	GTT	227	
9mers	GTT	145	
10mers	GTT	134	
11mers	GTT	111	
12mers	GTT	87	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	AAG	10064	W+=7
6mers	AAG	4666	W-=29
7mers	AAG	880	p<=0.1484
8mers	AAG	320	NOT SIGNIFICANT
9mers	AAG	207	
10mers	AAG	159	
11mers	AAG	136	
12mers	AAG	109	
5mers	CTT	9586	
6mers	CTT	5168	
7mers	CTT	959	
8mers	CTT	390	
9mers	CTT	243	
10mers	CTT	192	
11mers	CTT	159	
12mers	CTT	134	
5mers	AAT	2923	W+=0
6mers	AAT	2395	W-=36
7mers	AAT	704	p<=0.007812
8mers	AAT	334	SIGNIFICANT
9mers	AAT	225	
10mers	AAT	187	
11mers	AAT	169	
12mers	AAT	151	
5mers	ATT	2946	
6mers	ATT	2469	
7mers	ATT	792	
8mers	ATT	390	
9mers	ATT	287	
10mers	ATT	222	
11mers	ATT	184	
12mers	ATT	167	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	ACA	4934	W+=13
6mers	ACA	2334	W-=23
7mers	ACA	537	p<=0.5469
8mers	ACA	291	NOT SIGNIFICANT
9mers	ACA	230	
10mers	ACA	179	
11mers	ACA	137	
12mers	ACA	111	
5mers	TGT	4662	
6mers	TGT	2535	
7mers	TGT	573	
8mers	TGT	297	
9mers	TGT	207	
10mers	TGT	181	
11mers	TGT	147	
12mers	TGT	124	
5mers	ACC	2754	W+=28
6mers	ACC	1689	W-=8
7mers	ACC	477	p<=0.1953
8mers	ACC	236	NOT SIGNIFICANT
9mers	ACC	172	
10mers	ACC	125	
11mers	ACC	110	
12mers	ACC	92	
5mers	GGT	2727	
6mers	GGT	1790	
7mers	GGT	434	
8mers	GGT	221	
9mers	GGT	162	
10mers	GGT	120	
11mers	GGT	108	
12mers	GGT	91	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	ACG	42	W+=8
6mers	ACG	231	W-=28
7mers	ACG	185	p<=0.1953
8mers	ACG	135	NOT SIGNIFICANT
9mers	ACG	95	
10mers	ACG	72	
11mers	ACG	61	
12mers	ACG	43	
5mers	CGT	51	
6mers	CGT	225	
7mers	CGT	195	
8mers	CGT	153	
9mers	CGT	106	
10mers	CGT	71	
11mers	CGT	58	
12mers	CGT	45	
5mers	ACT	2863	W+=24.5
6mers	ACT	1931	W-=11.5
7mers	ACT	435	p<=0.3828
8mers	ACT	217	NOT SIGNIFICANT
9mers	ACT	143	
10mers	ACT	124	
11mers	ACT	95	
12mers	ACT	81	
5mers	AGT	2959	
6mers	AGT	1943	
7mers	AGT	430	
8mers	AGT	201	
9mers	AGT	131	
10mers	AGT	100	
11mers	AGT	83	
12mers	AGT	69	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	AGA	14367	W+=7
6mers	AGA	5296	W-=29
7mers	AGA	779	p<=0.1484
8mers	AGA	256	NOT SIGNIFICANT
9mers	AGA	143	
10mers	AGA	115	
11mers	AGA	95	
12mers	AGA	73	
5mers	TCT	14131	
6mers	TCT	5865	
7mers	TCT	864	
8mers	TCT	301	
9mers	TCT	163	
10mers	TCT	139	
11mers	TCT	114	
12mers	TCT	87	
5mers	AGC	7862	W+=15
6mers	AGC	3991	W-=21
7mers	AGC	711	p<=0.7422
8mers	AGC	270	NOT SIGNIFICANT
9mers	AGC	165	
10mers	AGC	125	
11mers	AGC	101	
12mers	AGC	91	
5mers	GCT	7103	
6mers	GCT	3922	
7mers	GCT	745	
8mers	GCT	322	
9mers	GCT	187	
10mers	GCT	161	
11mers	GCT	132	
12mers	GCT	118	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	AGG	18423	W+=8
6mers	AGG	7525	W-=28
7mers	AGG	1344	p<=0.1953
8mers	AGG	415	NOT SIGNIFICANT
9mers	AGG	246	
10mers	AGG	179	
11mers	AGG	139	
12mers	AGG	105	
5mers	CCT	17810	
6mers	CCT	7844	
7mers	CCT	1379	
8mers	CCT	463	
9mers	CCT	291	
10mers	CCT	223	
11mers	CCT	170	
12mers	CCT	133	
5mers	ATA	971	W+=25
6mers	ATA	919	W-=11
7mers	ATA	323	p<=0.3828
8mers	ATA	184	NOT SIGNIFICANT
9mers	ATA	129	
10mers	ATA	120	
11mers	ATA	105	
12mers	ATA	86	
5mers	TAT	803	
6mers	TAT	886	
7mers	TAT	350	
8mers	TAT	179	
9mers	TAT	141	
10mers	TAT	113	
11mers	TAT	98	
12mers	TAT	79	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	ATC	1103	W+=1
6mers	ATC	1165	W-=35
7mers	ATC	326	p<=0.01562
8mers	ATC	170	SIGNIFICANT
9mers	ATC	120	
10mers	ATC	98	
11mers	ATC	90	
12mers	ATC	76	
5mers	GAT	1126	
6mers	GAT	1159	
7mers	GAT	377	
8mers	GAT	221	
9mers	GAT	155	
10mers	GAT	124	
11mers	GAT	103	
12mers	GAT	84	
5mers	ATG	1817	W+=1
6mers	ATG	1648	W-=35
7mers	ATG	480	p<=0.01562
8mers	ATG	291	SIGNIFICANT
9mers	ATG	202	
10mers	ATG	161	
11mers	ATG	133	
12mers	ATG	105	
5mers	CAT	1904	
6mers	CAT	1780	
7mers	CAT	478	
8mers	CAT	299	
9mers	CAT	217	
10mers	CAT	167	
11mers	CAT	142	
12mers	CAT	113	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	CAA	3933	W+=7
6mers	CAA	2316	W-=29
7mers	CAA	555	p<=0.1484
8mers	CAA	274	NOT SIGNIFICANT
9mers	CAA	205	
10mers	CAA	153	
11mers	CAA	128	
12mers	CAA	107	
5mers	TTG	3784	
6mers	TTG	2818	
7mers	TTG	676	
8mers	TTG	362	
9mers	TTG	236	
10mers	TTG	191	
11mers	TTG	154	
12mers	TTG	125	
5mers	CAC	4971	W+=16
6mers	CAC	2245	W-=12
7mers	CAC	609	p<=0.8125
8mers	CAC	352	NOT SIGNIFICANT
9mers	CAC	227	
10mers	CAC	159	
11mers	CAC	128	
12mers	CAC	103	
5mers	GTG	5031	
6mers	GTG	2437	
7mers	GTG	589	
8mers	GTG	283	
9mers	GTG	193	
10mers	GTG	145	
11mers	GTG	128	
12mers	GTG	98	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	CAG	20268	W+=15
6mers	CAG	8061	W-=21
7mers	CAG	1162	p<=0.7422
8mers	CAG	375	NOT SIGNIFICANT
9mers	CAG	233	
10mers	CAG	176	
11mers	CAG	148	
12mers	CAG	114	
5mers	CTG	19109	
6mers	CTG	7975	
7mers	CTG	1183	
8mers	CTG	415	
9mers	CTG	246	
10mers	CTG	184	
11mers	CTG	162	
12mers	CTG	131	
5mers	CCA	13053	W+=14
6mers	CCA	5810	W-=22
7mers	CCA	1059	p<=0.6406
8mers	CCA	404	NOT SIGNIFICANT
9mers	CCA	252	
10mers	CCA	187	
11mers	CCA	153	
12mers	CCA	126	
5mers	TGG	12228	
6mers	TGG	5885	
7mers	TGG	1008	
8mers	TGG	414	
9mers	TGG	277	
10mers	TGG	199	
11mers	TGG	163	
12mers	TGG	136	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	CCC	13738	W+=24
6mers	CCC	5828	W-=4
7mers	CCC	1222	p<=0.1094
8mers	CCC	393	NOT SIGNIFICANT
9mers	CCC	226	
10mers	CCC	167	
11mers	CCC	111	
12mers	CCC	84	
5mers	GGG	13436	
6mers	GGG	5660	
7mers	GGG	1188	
8mers	GGG	405	
9mers	GGG	226	
10mers	GGG	141	
11mers	GGG	106	
12mers	GGG	86	
5mers	CCG	890	W+=0
6mers	CCG	761	W-=36
7mers	CCG	430	p<=0.007812
8mers	CCG	235	SIGNIFICANT
9mers	CCG	136	
10mers	CCG	105	
11mers	CCG	81	
12mers	CCG	64	
5mers	CGG	975	
6mers	CGG	856	
7mers	CGG	499	
8mers	CGG	262	
9mers	CGG	162	
10mers	CGG	111	
11mers	CGG	91	
12mers	CGG	72	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	CGA	56	W+=0
6mers	CGA	128	W-=36
7mers	CGA	78	p<=0.007812
8mers	CGA	68	SIGNIFICANT
9mers	CGA	43	
10mers	CGA	36	
11mers	CGA	20	
12mers	CGA	23	
5mers	TCG	69	
6mers	TCG	137	
7mers	TCG	92	
8mers	TCG	71	
9mers	TCG	52	
10mers	TCG	47	
11mers	TCG	41	
12mers	TCG	40	
5mers	CGC	851	W+=3
6mers	CGC	804	W-=25
7mers	CGC	498	p<=0.07812
8mers	CGC	268	NOT SIGNIFICANT
9mers	CGC	162	
10mers	CGC	126	
11mers	CGC	83	
12mers	CGC	65	
5mers	GCG	942	
6mers	GCG	874	
7mers	GCG	560	
8mers	GCG	288	
9mers	GCG	185	
10mers	GCG	118	
11mers	GCG	79	
12mers	GCG	65	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	CTA	611	W+=36
6mers	CTA	729	W-=0
7mers	CTA	192	p<=0.007812
8mers	CTA	134	SIGNIFICANT
9mers	CTA	99	
10mers	CTA	83	
11mers	CTA	67	
12mers	CTA	57	
5mers	TAG	600	
6mers	TAG	669	
7mers	TAG	180	
8mers	TAG	118	
9mers	TAG	84	
10mers	TAG	77	
11mers	TAG	57	
12mers	TAG	47	
5mers	CTC	16460	W+=28
6mers	CTC	6657	W-=8
7mers	CTC	1174	p<=0.1953
8mers	CTC	378	NOT SIGNIFICANT
9mers	CTC	208	
10mers	CTC	162	
11mers	CTC	128	
12mers	CTC	97	
5mers	GAG	16652	
6mers	GAG	6470	
7mers	GAG	1088	
8mers	GAG	313	
9mers	GAG	192	
10mers	GAG	117	
11mers	GAG	96	
12mers	GAG	84	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	GAA	12299	W+=7
6mers	GAA	5519	W-=29
7mers	GAA	978	p<=0.1484
8mers	GAA	351	NOT SIGNIFICANT
9mers	GAA	220	
10mers	GAA	168	
11mers	GAA	136	
12mers	GAA	122	
5mers	TTC	12043	
6mers	TTC	6087	
7mers	TTC	1062	
8mers	TTC	386	
9mers	TTC	236	
10mers	TTC	190	
11mers	TTC	150	
12mers	TTC	131	
5mers	GAC	1966	W+=4.5
6mers	GAC	1410	W-=31.5
7mers	GAC	376	p<=0.05469
8mers	GAC	196	NOT SIGNIFICANT
9mers	GAC	150	
10mers	GAC	126	
11mers	GAC	103	
12mers	GAC	77	
5mers	GTC	1965	
6mers	GTC	1473	
7mers	GTC	389	
8mers	GTC	233	
9mers	GTC	175	
10mers	GTC	122	
11mers	GTC	104	
12mers	GTC	92	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	GCA	6568	W+=13
6mers	GCA	3032	W-=23
7mers	GCA	609	p<=0.5469
8mers	GCA	259	NOT SIGNIFICANT
9mers	GCA	168	
10mers	GCA	130	
11mers	GCA	103	
12mers	GCA	82	
5mers	TGC	5853	
6mers	TGC	2999	
7mers	TGC	650	
8mers	TGC	301	
9mers	TGC	172	
10mers	TGC	143	
11mers	TGC	112	
12mers	TGC	92	
5mers	GCC	8560	W+=12
6mers	GCC	4088	W-=16
7mers	GCC	1010	p<=0.8125
8mers	GCC	395	NOT SIGNIFICANT
9mers	GCC	237	
10mers	GCC	199	
11mers	GCC	147	
12mers	GCC	115	
5mers	GGC	8507	
6mers	GGC	4102	
7mers	GGC	1060	
8mers	GGC	409	
9mers	GGC	261	
10mers	GGC	184	
11mers	GGC	143	
12mers	GGC	115	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	GGA	16130	W+=3
6mers	GGA	6965	W-=25
7mers	GGA	1338	p<=0.07812
8mers	GGA	438	NOT SIGNIFICANT
9mers	GGA	273	
10mers	GGA	207	
11mers	GGA	158	
12mers	GGA	131	
5mers	TCC	16499	
6mers	TCC	7409	
7mers	TCC	1370	
8mers	TCC	466	
9mers	TCC	268	
10mers	TCC	203	
11mers	TCC	166	
12mers	TCC	131	
5mers	GTA	394	W+=4.5
6mers	GTA	527	W-=23.5
7mers	GTA	170	p<=0.1094
8mers	GTA	110	NOT SIGNIFICANT
9mers	GTA	78	
10mers	GTA	67	
11mers	GTA	64	
12mers	GTA	55	
5mers	TAC	406	
6mers	TAC	533	
7mers	TAC	209	
8mers	TAC	107	
9mers	TAC	88	
10mers	TAC	74	
11mers	TAC	64	
12mers	TAC	48	

Width	Trinucleotide	Total number	Wilcoxon Signed-Rank Test
5mers	TAA	2350	W+=28
6mers	TAA	2008	W-=8
7mers	TAA	661	p<=0.1953
8mers	TAA	324	NOT SIGNIFICANT
9mers	TAA	211	
10mers	TAA	180	
11mers	TAA	161	
12mers	TAA	146	
5mers	TTA	2095	
6mers	TTA	1826	
7mers	TTA	680	
8mers	TTA	298	
9mers	TTA	216	
10mers	TTA	179	
11mers	TTA	158	
12mers	TTA	131	
5mers	TCA	5327	W+=19
6mers	TCA	3067	W-=17
7mers	TCA	656	p<=0.9453
8mers	TCA	351	NOT SIGNIFICANT
9mers	TCA	265	
10mers	TCA	207	
11mers	TCA	171	
12mers	TCA	146	
5mers	TGA	5590	
6mers	TGA	3203	
7mers	TGA	672	
8mers	TGA	345	
9mers	TGA	237	
10mers	TGA	185	
11mers	TGA	151	
12mers	TGA	111	

Appendix A8. List of the 91 genes with the most abundant highly conserved promoter elements.

Gene	Human Ensembl ID	Criteria	Gene	Human Ensembl ID	Criteria
PHOX2B	ENSG00000109132	13 12mers	SLC39A7	ENSG00000112473	4 12mers
HOXA2	ENSG00000105996	10 12mers	IL21	ENSG00000138684	4 12mers
EVX2	ENSG00000174279	10 12mers	PSENEN	ENSG00000205155	4 12mers
HOXB1	ENSG00000120094	8 12mers	FOSB	ENSG00000125740	4 12mers
DUSP6	ENSG00000139318	8 12mers	RAB24	ENSG00000169228	4 12mers
ESM1	ENSG00000164283	8 12mers	NFKBIA	ENSG00000100906	5 11mers
CDX2	ENSG00000165556	7 12mers	SPI1	ENSG00000066336	5 11mers
FUT11	ENSG00000196968	7 12mers	VGF	ENSG00000128564	5 11mers
HOXD10	ENSG00000128710	6 12mers	MSTN	ENSG00000138379	5 11mers
HOXB9	ENSG00000170689	6 12mers	GANAB	ENSG00000089597	5 11mers
OVOL2	ENSG00000125850	6 12mers	PREX2	ENSG00000046889	4 11mers
SMAD6	ENSG00000137834	6 12mers	FBXO32	ENSG00000156804	4 11mers
MAB21L1	ENSG00000180660	6 12mers	SOCS7	ENSG00000174111	4 11mers
ROGDI	ENSG00000067836	6 12mers	BAZ1B	ENSG00000009954	4 11mers
LBX1	ENSG00000138136	6 12mers	SLC7A11	ENSG00000151012	4 11mers
SCAMP2	ENSG00000116521	6 12mers	ACCN4	ENSG00000072182	4 11mers
FGF6	ENSG00000111241	5 12mers	MT-ND2	ENSG00000198763	4 11mers
DDIT4	ENSG00000168209	5 12mers	MT-CO1	ENSG00000198804	4 11mers
KCTD5	ENSG00000167977	5 12mers	ELMO3	ENSG00000102890	4 11mers
ID3	ENSG00000117318	5 12mers	JUN	ENSG00000177606	4 11mers
DSG4	ENSG00000175065	5 12mers	ARV1	ENSG00000173409	4 11mers
VSX2	ENSG00000119614	5 12mers	NIPBL	ENSG00000164190	5 10mers
MAF	ENSG00000178573	5 12mers	PI15	ENSG00000137558	5 10mers
PLXNC1	ENSG00000136040	5 12mers	EDN1	ENSG00000078401	5 10mers
C7ORF55	ENSG00000164898	5 12mers	ARHGEF6	ENSG00000129675	5 10mers
TBR1	ENSG00000136535	5 12mers	CYR61	ENSG00000142871	5 10mers

Gene	Human Ensembl ID	Criteria	Gene	Human Ensembl ID	Criteria
CTGF	ENSG00000118523	5 12mers	PNO1	ENSG00000115946	5 10mers
MNT	ENSG00000070444	5 12mers	ATXN7L2	ENSG00000162650	5 10mers
PRMT5	ENSG00000100462	5 12mers	THBS3	ENSG00000169231	5 10mers
BHLHE40	ENSG00000134107	5 12mers	BSX	ENSG00000188909	8 9mers
IPO4	ENSG00000196497	5 12mers	AGER	ENSG00000204305	7 9mers
KDM6A	ENSG00000147050	4 12mers	C1QL1	ENSG00000165985	6 9mers
HOXA5	ENSG00000106004	4 12mers	SHC4	ENSG00000185634	6 9mers
DBX1	ENSG00000109851	4 12mers	ATOH7	ENSG00000179774	6 9mers
ITGBL1	ENSG00000198542	4 12mers	ATOH7	ENSG00000179774	6 9mers
LOX	ENSG00000113083	4 12mers	TSSK3	ENSG00000162526	6 9mers
DRGX	ENSG00000165606	4 12mers	CHPF	ENSG00000123989	6 9mers
RAB11A	ENSG00000103769	4 12mers	EMILIN1	ENSG00000138080	10 8mers
LIN28A	ENSG00000131914	4 12mers	EIF2B4	ENSG00000115211	10 8mers
B3GNT1	ENSG00000174684	4 12mers	FOXA1	ENSG00000129514	9 8mers
HOXB7	ENSG00000120087	4 12mers	SF4	ENSG00000105705	9 8mers
TBX19	ENSG00000143178	4 12mers	COLQ	ENSG00000206561	8 8mers
C1ORF124	ENSG00000010072	4 12mers	NR0B2	ENSG00000131910	8 8mers
CD68	ENSG00000129226	4 12mers	IRX5	ENSG00000176842	8 8mers
RAB2B	ENSG00000129472	4 12mers	ATP2A2	ENSG00000174437	8 8mers
HOXA11	ENSG00000005073	4 12mers	NUP188	ENSG00000095319	8 8mers
			ST7L	ENSG00000007341	8 8mers
			INTS7	ENSG00000143493	8 8mers

Gene symbols and Ensembl IDs based on human.

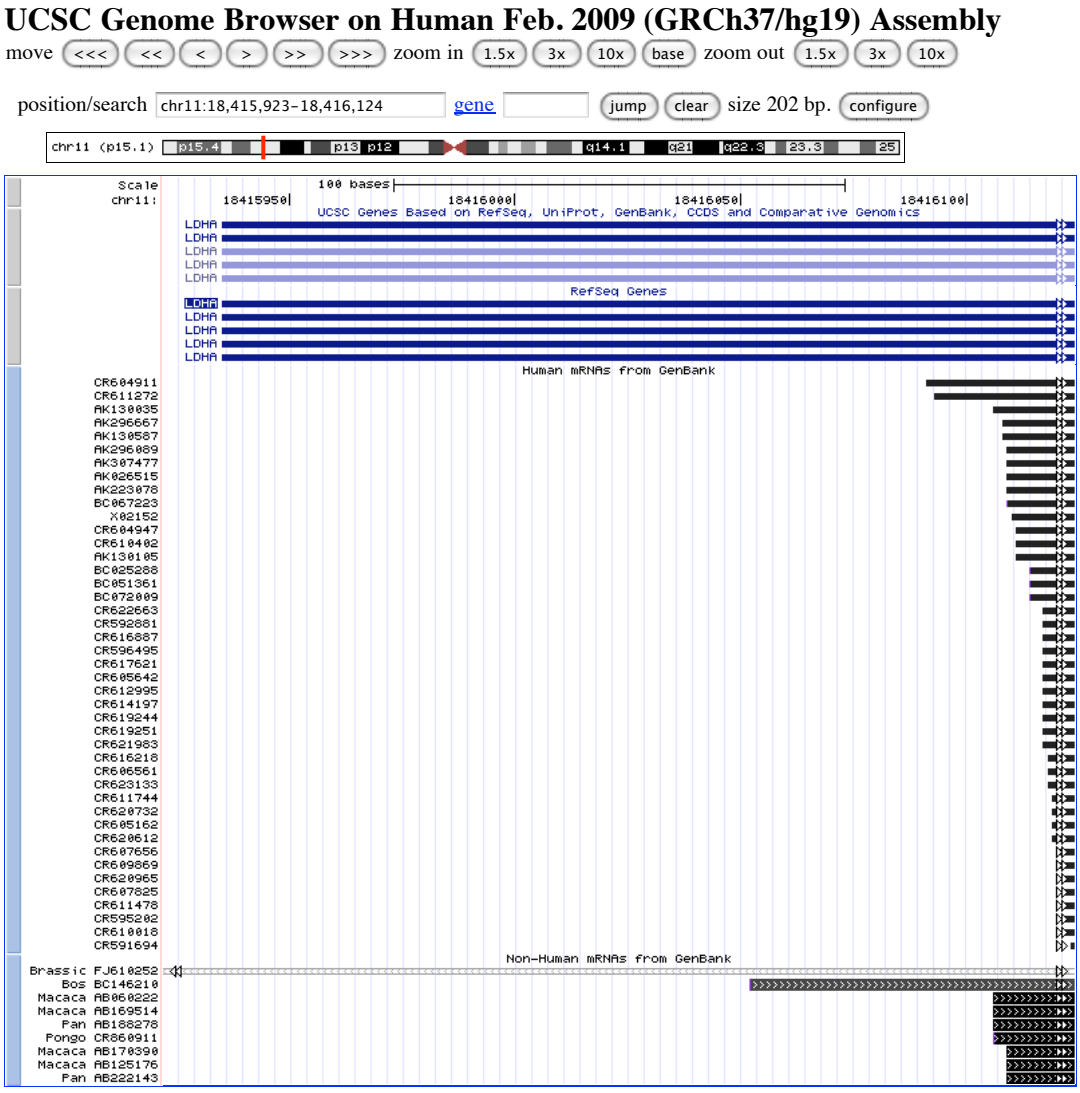
Appendix A9. DAVID results identifying over-represented functional categories (enrichment score > 1)

Annotation Cluster	Enrichment Score	Number of Genes	Genes
Regulation of metabolic process	8.46	31	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EDN1 EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT MSTN NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2
Regulation of transcription, DNA-dependent	7.93	29	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2
Regulation of transcription	7.53	30	ATOH7 BAZ1B BSX CDX2 DBX1 DRGX EVX2 FOXA1 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 ID3 IRX5 LIN28 MAF MNT MSTN NR0B2 OVOL2 PHOX2B PRMT5 RAB11A SMAD6 SPI1 TBR1 TBX19 VSX2
Homeodomain-related	7.41	10	CDX2 HOXA2 HOXA5 HOXB1 HOXB7 HOXB9 HOXD10 IRX5 PHOX2B VSX2
Insulin-like growth factor binding	2.45	3	CTGF CYR61 ESM1
Branching morphogenesis of a tube Morphogenesis of a branching structure Tube morphogenesis	1.7	3	CYR61 EDN1 FOXA1

Annotation Cluster	Enrichment Score	Number of Genes	Genes
Angiogenesis Blood vessel morphogenesis Anatomical structure formation	1.67	4	CTGF CYR61 EDN1 FGF6
Collagen triple helix repeat Collagen Phosphate transport Inorganic anion transport Anion transport	1.21	3	C1QL1 COLQ EMILIN1
Regulation of cell growth Cell growth Regulation of cell size Regulation of growth Cell morphogenesis Cellular structure morphogenesis	1.2	4	CTGF CYR61 ESM1 SOCS7

APPENDIX B

Appendix B1. Depiction of transcription start site variability for human *LDHA*.



Picture from University of California, Santa Cruz Genome Browser (Kent et al. 2002). Reference sequences for human *LDHA* are shown in blue bars, approximately 180nt upstream of other human RNAs, cDNAs (solid black boxes), as well as reference sequences of other closely related species (black boxes with arrow heads within them).

Appendix B2. *LDHA* promoter footprints gained during primate evolution

Branch	Element	Element ID	Position (relative to human TSS)
Strepsirrhine	GTGGG	dFPa1	-346
	GGGTGCC	dFPa3	-295
	TGCCTGT	dFPa4	-253
	CAGCTGG	dFPa5	-243
	CCTACG	dFPa6	-221
Anthropoid	TATTAC	dFPa7	-367
	CGTCC	dFPa8	-254
	CTGTGG	dFPa9	-218
	GCAGTC	dFPa10	-171
	CCACCCC	dFPa11	-98
	TTTCGT	dFPa12	-396
Platyrrhine	AAACTT	dFPa13	-354
Catarrhine	TGACTG	dFPa14	-418
Old World monkey	N/A		
Ape	CCACCC	dFPa15	-135
Primate	AAACAGGG	dFPa16	-334
	CAGATC	dFPa17	-314

N/A indicates absence of element on that branch

Appendix B3. *LDHB* promoter footprints gained during primate evolution

Branch	Element	Element ID	Position (relative to human TSS)
Strepsirrhine	GAGATA	dFPb1	-239
	GGCGTA	dFPb2	-196
	AAATAGCCGG	dFPb3	-180
	TTTGGG	dFPb4	-139
Anthropoid	GGATGG	dFPb5	-238
	GACCAGCT	dFPb6	-180
Platyrrhine	TCCGTT	dFPb7	-291
	CATTGA	dFPb8	-260
	CGGGC	dFPb9	-134
Catarrhine	N/A		
Old World monkey	ATTTTG	dFPb10	-262
Ape	Aluy	dFPb11	-302
Primate	N/A		

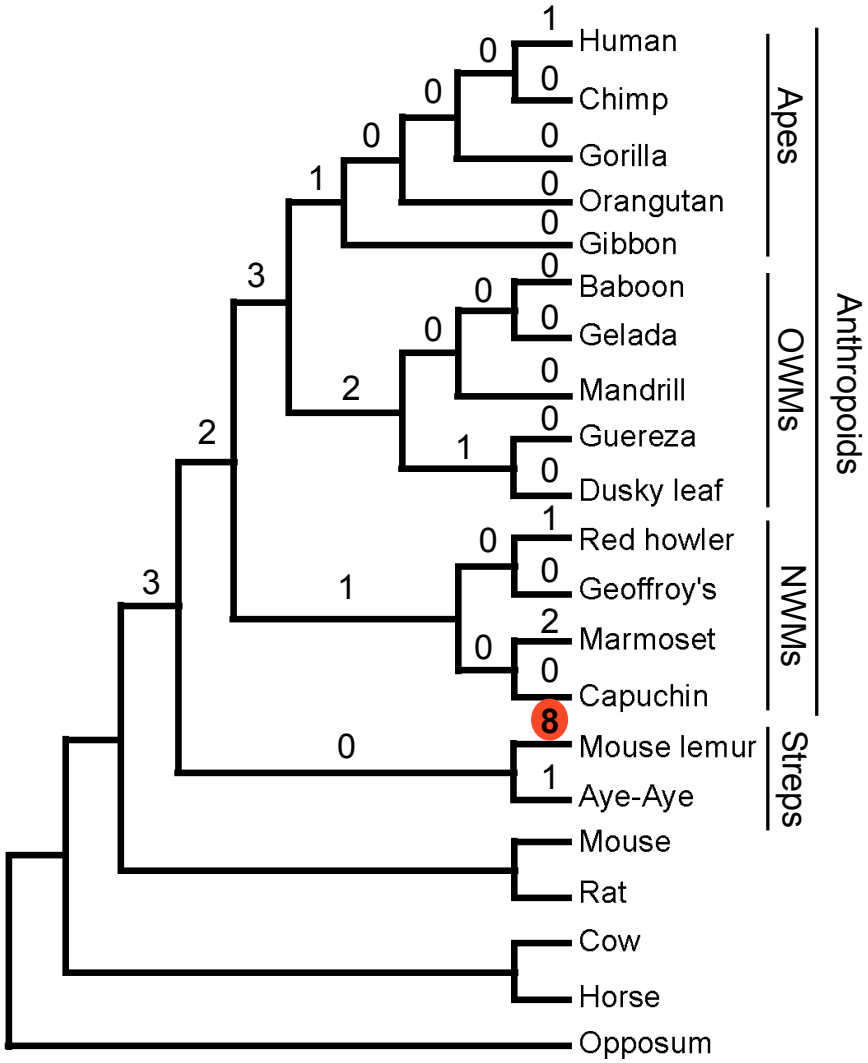
N/A indicates absence of element on that branch

APPENDIX C

Appendix C1. Scientific names and common names for the 22 species included in this study.

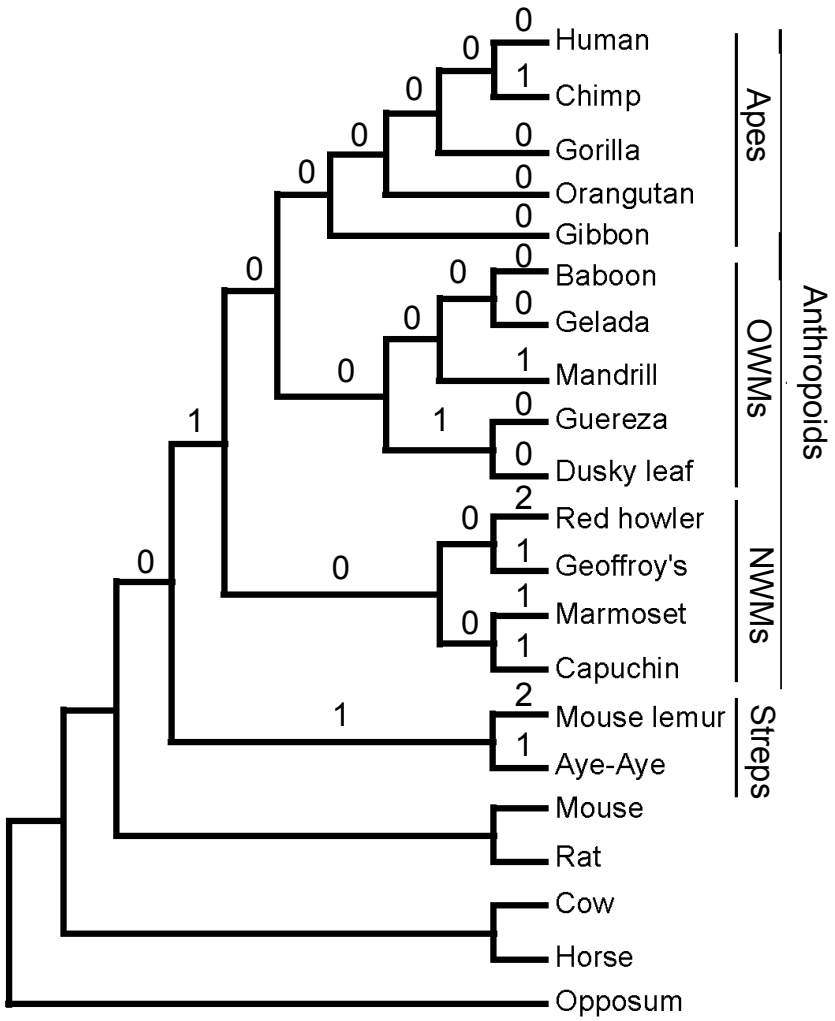
Scientific names	Common names
Homo sapiens	Human
Pan troglodytes	Common Chimpanzee
Gorilla gorilla	Gorilla
Pongo pygmaeus	Orangutan
Hylobates lar	White-handed Gibbon
Papio anubis	Olive Baboon
Theropithecus gelada	Gelada
Mandrillus leucophaeus	Mandrill
Colobus guereza	Guereza
Trachypithecus obscurus	Dusky leaf monkey
Alouatta seniculus	Red howler monkey
Ateles geoffroyi	Geoffroy's Spider monkey
Callithrix jacchus	Common marmoset
Cebus apella	Tufted Capuchin
Microcebus murinus	Gray Mouse Lemur
Daubentonia madagascarensis	Aye-Aye
Mus musculus	Mouse
Rattus norvegicus	Rat
Bos taurus	Cow
Equus caballus	Horse
Monodelphis domesticus	Opossum

Appendix C2. Inferred number of nonsynonymous substitutions in *LDHA* on primate branches



Marginal reconstruction of *LDHA* with number of nonsynonymous substitutions with posterior probabilities > 0.700. OWMs, Old World monkeys; NWMs, New World monkeys; Streps, strepsirrhines. Scientific and common names are given in Appendix C1.

Appendix C3. Inferred number of nonsynonymous substitutions in *LDHB* on primate branches



Marginal reconstruction of *LDHB* with number of nonsynonymous substitutions with posterior probabilities > 0.700. OWMs, Old World monkeys; NWMs, New World monkeys; Streps, strepsirrhines. Scientific and common names are given in Appendix C1.

Appendix C4. Grubbs' test for statistical outliers of nonsynonymous substitutions on primate branches for *LDHA* and *LDHB*.

Number of substitutions-*LDHA*	**Z value**	**p-value**
0	0.52	p>0.05
1	0.10	p>0.05
2	0.72	p>0.05
3	1.34	p>0.05
8	*4.44*	*p<0.05*
Number of substitutions-*LDHB*	**Z value**	**p-value**
0	0.68	p>0.05
1	0.94	p>0.05
2	2.55	p>0.05

Appendix C5. Branch-sites test for codon selection in *LDHA* on mouse lemur terminal branch

Model	P	Likelihood (-ln L)	2Δln L	p*	Parameters estimated	Inferred sites under adaptive evolution
Model A	4	-3481.52	0.04	p>0.75	p_0=0.88;p_1=0.057;p_{2a+2b}=0.059;ω_0=0.026;ω_1=1;*background* ω_{2a}=0.026; ω_{2b}=1;*foreground* ω_{2a}=ω_{2b}=1.534	94 (0.58); 180 (0.56); 257 (0.68); 265 (0.61)
Model A (fixed omega)	3	-3481.54	N/A	N/A	p_0=0.86;p_1=0.055;p_{2a+2b}=0.084;ω_0=0.026;ω_1=1;*background* ω_{2a}=0.026; ω_{2b}=1;*foreground* ω_{2a}=ω_{2b}=1	N/A

P, number of parameters; Differences in likelihood values (0.02) are not statistically significant. Numbers in parentheses indicate Bayesian posterior probability of adaptive evolution at that codon. N/A, not applicable.

Appendix C6. PCR protocols for each species and all seven coding exons for *LDHA*.

Gorila gorilla	z153		***Pongo pygmaeus***	z276	
LDHA1	Ldha1f2	Ldha1r1	**LDHA1**	Ldha1f2	Ldha1r1
58.1	1 min. x30	7 min.	58.1	1 min. x30	7 min.
LDHA2	Ldha23f1	LdhaCex3r1	**LDHA2**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, first	Ldha23f1	LdhaCex3r1	**LDHA3, first**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, second	Ldhacex3f1	Ldha34intr1	**LDHA3, second**	Ldhacex3f1	Ldha34intr1
51.6 OUTER	1.15 min. x30	5 min.	51.6 OUTER	1.15 min. x30	5 min.
LDHA3, second	Ldhacex3f1	Ldha34intr2	**LDHA3, second**	Ldhacex3f1	Ldha34intr2
56.8 INNER	1.15 min. x30	5 min.	56.8 INNER	1.15 min. x30	5 min.
LDHA4	Ldha45f1	Ldha45r2	**LDHA4**	Ldha45f1	Ldha45r2
54	1.45 x30	10 min.	54	1.45 x30	10 min.
LDHA5	Ldha45f1	Ldha45r2	**LDHA5**	Ldha45f1	Ldha45r2
54	1.45 x30	10 min.	54	1.45 x30	10 min.
LDHA6	Ldha67f1	Ldha67r2	**LDHA6**	Ldha67f1	Ldha67r2
54	1.45 min. x30	7 min.	54	1.45 min. x30	7 min.
LDHA7	Ldha67f1	Ldha67r2	**LDHA7**	Ldha67f1	Ldha67r2
54	1.45 min. x30	7 min.	54	1.45 min. x30	7 min.
Hylobates lar	z185		***Papio anubis***	z193	
LDHA1	Ldha1f2	Ldha1r1	**LDHA1**	Ldha1f2	Ldha1r1
58.1	1 min. x30	7 min.	58.1	1 min. x30	7 min.
LDHA2	Ldha23f1	LdhaCex3r1	**LDHA2**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, first	Ldha23f1	LdhaCex3r1	**LDHA3, first**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, second	Ldha23f1	LdhaE3ne	**LDHA3, second**	Ldhacex3f	Ldha34intr

		wr1		1	1
51.6	1.15 min	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHA3, second	Ldha23nwmf2	Ldha23r1	**LDHA3, second**	Ldhacex3f1	Ldha34intr2
51.6	1.15 min	5 min.	56.8	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHA4	Ldha45f1	Ldha45r2	**LDHA4**	Ldha45f1	Ldha45r2
54	1.45	10 min.	54	1.45	10 min.
	x30			x30	
LDHA5	Ldha45f1	Ldha45r2	**LDHA5**	Ldha45f1	Ldha45r2
54	1.45	10 min.	54	1.45	10 min.
	x30			x30	
LDHA6	Ldha67f1	Ldha67r2	**LDHA6**	Ldha67f1	Ldha67r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
	x30			x30	
LDHA7	Ldha67f1	Ldha67r2	**LDHA7**	Ldha67f1	Ldha67r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
	x30			x30	
Theropithecus gelada	z301		***Mandrillus leucophaeus***	z190	
LDHA1	Ldha1f2	Ldha1r1	**LDHA1**	Ldha1f2	Ldha1r1
58.1	1 min.	7 min.	58.1	1 min.	7 min.
	x30			x30	
LDHA2	Ldha23f1	LdhaCex3r1	**LDHA2**	Ldha23f1	LdhaCex3r1
54	1.45 min.	10 min.	54	1.45 min.	10 min.
	x30			x30	
LDHA3, first	Ldha23f1	LdhaCex3r1	**LDHA3, first**	Ldha23f1	LdhaCex3r1
54	1.45 min.	10 min.	54	1.45 min.	10 min.
	x30			x30	
LDHA3, second	Ldhacex3f1	Ldha34intr1	**LDHA3, second**	Ldhacex3f1	Ldha34intr1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHA3, second	Ldhacex3f1	Ldha34intr2	**LDHA3, second**	Ldhacex3f1	Ldha34intr2
56.8	1.15 min.	5 min.	56.8	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHA4	Ldha45f1	Ldha45r2	**LDHA4**	Ldha45f1	Ldha45r2
54	1.45	10 min.	54	1.45	10 min.
	x30			x30	
LDHA5	Ldha45f1	Ldha45r2	**LDHA5**	Ldha45f1	Ldha45r2
54	1.45	10 min.	54	1.45	10 min.
	x30			x30	
LDHA6	Ldha67f1	Ldha67r2	**LDHA6**	Ldha67f1	Ldha67r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
	x30			x30	

LDHA7	Ldha67f1	Ldha67r2	**LDHA7**	Ldha67f1	Ldha67r2
54	1.45 min. x30	7 min.	54	1.45 min. x30	7 min.
Colobus guereza	z183		***Trachypithecus obscurus***	z181	
LDHA1	Ldha1f2	Ldha1r1	**LDHA1**	Ldha1f2	Ldha1r1
58.1	1 min. x30	7 min.	58.1	1 min. x30	7 min.
LDHA2	Ldha23f1	LdhaCex3r1	**LDHA2**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, first	Ldha23f1	LdhaCex3r1	**LDHA3, first**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, second	Ldhacex3f1	Ldha34intr1	**LDHA3, second**	Ldhacex3f1	Ldha34intr1
51.6 OUTER	1.15 min. x30	5 min.	51.6 OUTER	1.15 min. x30	5 min.
LDHA3, second	Ldhacex3f1	Ldha34intr2	**LDHA3, second**	Ldhacex3f1	Ldha34intr2
56.8 INNER	1.15 min. x30	5 min.	56.8 INNER	1.15 min. x30	5 min.
LDHA4	Ldha45f1	Ldha45r2	**LDHA4**	Ldha45f1	Ldha45r2
54	1.45 x30	10 min.	54	1.45 x30	10 min.
LDHA5	Ldha45f1	Ldha45r2	**LDHA5**	Ldha45f1	Ldha45r2
54	1.45 x30	10 min.	54	1.45 x30	10 min.
LDHA6	Ldha67f1	Ldha67r2	**LDHA6**	Ldha67f1	Ldha67r2
54	1.45 min. x30	7 min.	54	1.45 min. x30	7 min.
LDHA7	Ldha67f1	Ldha67r2	**LDHA7**	Ldha67f1	Ldha67r2
54	1.45 min. x30	7 min.	54	1.45 min. x30	7 min.
Alouatta seniculus	z231		***Ateles geoffroyi***	z254	
LDHA1	Ldha1f2	Ldha1r2	**LDHA1**	Ldha1f2	Ldha1r2
51.6	1 min. x30	7 min.	51.6	1 min. x30	7 min.
LDHA2	Ldha23f1	LdhaCex3r1	**LDHA2**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.
LDHA3, first	Ldha23f1	LdhaCex3r1	**LDHA3, first**	Ldha23f1	LdhaCex3r1
54	1.45 min. x30	10 min.	54	1.45 min. x30	10 min.

LDHA3, second	Ldha23f1	LdhaE3newr1	**LDHA3, second**	Ldhacex3f1	Ldha34intr1
51.6	1.15 min	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHA3, second	Ldha23nwmf2	Ldha23r1	**LDHA3, second**	Ldhacex3f1	Ldha34intr2
51.6	1.15 min	5 min.	56.8	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHA4	Ldha45f1	Ldha45r2	**LDHA4**	Ldha45f1	Ldha45r2
51.6	1.45 min.	7 min.	54	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHA4	Ldha45f2	Ldha45nwmr1	**LDHA4**	Ldha45f2	Ldha45r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHA5	Ldha45f1	Ldha45r2	**LDHA5**	Ldha45f1	Ldha45r2
51.6	1.45 min.	7 min.	54	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHA5	Ldha45f2	Ldha45nwmr1	**LDHA5**	Ldha45f2	Ldha45r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHA6	Ldha67f2	Ldha67r2	**LDHA6**	Ldha67f1	Ldha67r2
51.6	1.45 min.	10 min.	54	1.45 min.	7 min.
	x30			x30	
LDHA7	Ldha67f2	Ldha67r2	**LDHA7**	Ldha67f1	Ldha67r2
51.6	1.45 min.	10 min.	54	1.45 min.	7 min.
	x30			x30	
Cebus apella	z248		***Callithrix jacchus***	z118	
LDHA1	Ldha1f2	Ldha1r1	**LDHA1**	Ldha1f2	Ldha1r1
51.6	1 min.	7 min.	51.6	1 min.	5 min.
	x30		INNER	x30	
			LDHA1	Ldha1f2	Ldha1r2
LDHA2	Ldha23f1	LdhaCex3r1	55.4	1 min.	7 min.
54	1.45 min.	10 min.	OUTER	x30	
	x30				
			LDHA2	Ldha23f1	LdhaCex3r1
LDHA3, first	Ldha23f1	LdhaCex3r1	54	1.45 min.	10 min.
54	1.45 min.	10 min.		x30	
	x30				
			LDHA3, first	Ldha23f1	LdhaCex3r1
LDHA3, second	Ldha23f1	LdhaE3newr1	54	1.45 min.	10 min.
51.6	1.15 min	5 min.		x30	
OUTER	x30				
LDHA3, second	Ldha23nwmf2	Ldha23r1	**LDHA3, second**	Ldha23f1	LdhaE3newr1
51.6	1.15 min	5 min.	51.6	1.15 min	5 min.
INNER	x30		OUTER	x30	
			LDHA3, second	Ldha23nwmf2	Ldha23r1
LDHA4	Ldha45f1	Ldha45r2	51.6	1.15 min	5 min.

51.6	1.45 min.	7 min.	INNER	x30	
OUTER	x30				
LDHA4	Ldha45f2	LdhaCex5R1	**LDHA4**	Ldha45f1	Ldha45r2
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x35		OUTER	x30	
			LDHA4	Ldha45f2	Ldha45r2
LDHA5	Ldha45f3	Ldha45r2	54	1.45 min.	7 min.
51.6	1.30 min.	7 min.	INNER	x30	
OUTER	x30				
LDHA5	Ldha45f2	Ldha45nwmr1	**LDHA5**	Ldha45f1	Ldha45r2
51.6	1.30 min.	7 min.	54	1.45 min.	7 min.
	x30		OUTER	x30	
			LDHA5	Ldha45f2	Ldha45r2
LDHA6	Ldha67f1	Ldha67r1	54	1.45 min.	7 min.
51.6	1.45 min.	7 min.	INNER	x30	
OUTER	x30				
LDHA6	Ldha67f2	Ldha67r2	**LDHA6**	Ldha6Fv5	Ldha3utr1
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
INNER	x30			x30	
LDHA7	Ldha67f1	Ldha67r1	**LDHA7**	Ldha6Fv5	Ldha3utr1
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30			x30	
LDHA7	Ldha67f2	Ldha67r2			
51.6	1.45 min.	7 min.			
INNER	x30				
Daubentonia madagascarensis	z217		***Microcebus murinus***	z167	
LDHA1	LdhaMmur1f1	LdhaMmur1r1	**LDHA1**	LdhaMmur1f1	LdhaMmur1r1
54	2.5 min.	7 min.	54	2.5 min.	7 min.
	x30			x30	
LDHA2	Ldha23nwmf2	LdhaCex3r1	**LDHA2**	Ldha23nwmf2	LdhaCex3r1
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
	x30			x30	
LDHA3	LdhaMmur3f1	LdhaMmur3r1	**LDHA3**	LdhaMmur3f1	LdhaMmur3r1
54	2.30 min.	7 min.	54	2.30 min.	7 min.
	x30			x30	
LDHA4	LdhaMmur4f1	LdhaMmur5r1	**LDHA4**	LdhaMmur4f1	LdhaMmur5r1
51.6	2.30 min.	7 min.	51.6	2.30 min.	7 min.
	x30			x30	
LDHA5	LdhaMmur4f1	LdhaMmur5r1	**LDHA5**	LdhaMmur4f1	LdhaMmur5r1
51.6	2.30 min.	7 min.	51.6	2.30 min.	7 min.
	x30			x30	
LDHA6	LdhaMmur6f1	LdhaMmur7r1	**LDHA6**	LdhaMmur6f1	LdhaMmur7r1
60	2.30 min.	7 min.	60	2.30 min.	7 min.
	x30			x30	

	LDHA7	Ldha67f2	Ldha67r2	**LDHA7**	LdhaMmur 6f1	LdhaMmur 7r1
	51.6	1.45 min. x30	7 min.	60	2.30 min. x30	7 min.

Protocol for each species contained within each box. Scientific name given in top left corner of each box. Number following LDHA indicates exon. Names to the right of exon number are primer names, see Appendix C8 for corresponding sequence. Number below exon indicates annealing temperature, followed by extension time, followed by final extension time. ‘x’ followed by a number, generally 30, indicates number of cycles. For those exons in which ‘OUTER’ and ‘INNER’ are found below annealing temperature, a nested PCR approach was used, and both PCR protocols are necessary for amplification.

Appendix C7. PCR protocols for each species and all seven coding exons for *LDHB*.

Gorilla gorilla	z153		***Pongo pygmaeus***	z276	
LDHB1	Ldhb1f1	Ldhb1r1	**LDHB1**	Ldhb1f1	Ldhb1r1
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB1	Ldhb1f2	Ldhb1r2	**LDHB1**	Ldhb1f2	Ldhb1r2
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
INNER	x30		INNER	x30	
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f1	Ldhb2r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB2	Ldhb2f2	Ldhb2r2	**LDHB2**	Ldhb2f2	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB3	Ldhb34f2	Ldhb34r2	**LDHB3**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB4	Ldhb34f1	Ldhb34r1	**LDHB4**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB4	Ldhb34f2	Ldhb34r2	**LDHB4**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r3	**LDHB5**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r3	**LDHB6**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB7	Ldhb67f1	Ldhb67r1	**LDHB7**	Ldhb67f1	Ldhb67r1
51.6	1.30 min.	7 min.	51.6	1.30 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB7	Ldhb67f2	Ldhb67r2	**LDHB7**	Ldhb67f2	Ldhb67r2
51.6	1.30 min.	7min.	51.6	1.30 min.	7min.
INNER	x30		INNER	x30	
Hylobates lar	z185		***Papio anubis***	z193	
LDHB1	Ldhb1f1	Ldhb1r1	**LDHB1**	Ldhb1f1	Ldhb1r1
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB1	Ldhb1f2	Ldhb1r2	**LDHB1**	Ldhb1f2	Ldhb1r2

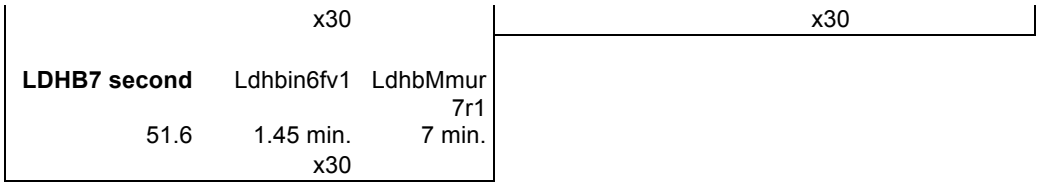
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
INNER	x30		INNER	x30	
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f3	Ldhb2r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB2	Ldhb2f2	Ldhb2r2	**LDHB2**	Ldhb2f3	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB3	Ldhb34f2	Ldhb34r2	**LDHB3**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB4	Ldhb34f1	Ldhb34r1	**LDHB4**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB4	Ldhb34f2	Ldhb34r2	**LDHB4**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r3	**LDHB5**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r3	**LDHB6**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB7	Ldhb67f1	Ldhb67r1	**LDHB7**	Ldhb67f1	Ldhb67r1
51.6	1.30 min.	7 min.	51.6	1.30 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB7	Ldhb67f2	Ldhb67r2	**LDHB7**	Ldhb67f2	Ldhb67r2
51.6	1.30 min.	7min.	51.6	1.30 min.	7min.
INNER	x30		INNER	x30	
Theropithecus gelada	z301		***Mandrillus leucophaeus***	z190	
LDHB1	Ldhb1f1	Ldhb1r1	**LDHB1**	Ldhb1f1	Ldhb1r1
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB1	Ldhb1f2	Ldhb1r2	**LDHB1**	Ldhb1f2	Ldhb1r2
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
INNER	x30		INNER	x30	
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f1	Ldhb2r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB2	Ldhb2f2	Ldhb2r2	**LDHB2**	Ldhb2f2	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.

INNER	x30		INNER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB3	Ldhb34f2	Ldhb34r2	**LDHB3**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB4	Ldhb34f1	Ldhb34r1	**LDHB4**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB4	Ldhb34f2	Ldhb34r2	**LDHB4**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r3	**LDHB5**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r3	**LDHB6**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB7	Ldhb67f1	Ldhb67r1	**LDHB7**	Ldhb67f1	Ldhb67r1
51.6	1.30 min.	7 min.	51.6	1.30 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB7	Ldhb67f2	Ldhb67r2	**LDHB7**	Ldhb67f2	Ldhb67r2
51.6	1.30 min.	7min.	51.6	1.30 min.	7min.
INNER	x30		INNER	x30	
Colobus guereza			***Trachypithecus obscurus***	z181	
LDHB1	Ldhb1f2	Ldhb1r1	**LDHB1**	Ldhb1f2	Ldhb1r1
51.6	1.30 min.	5 min.	51.6	1.30 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB1	Ldhb1f3	Ldhb1r2	**LDHB1**	Ldhb1newf1	Ldhb1newr1
55.4	1 min	2 min	55.4	1.30 min	5 min
INNER	x30		INNER	x30	
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f1	Ldhb2r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB2	Ldhb2f2	Ldhb2r2	**LDHB2**	Ldhb2f2	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB3	Ldhb34f2	Ldhb34r2	**LDHB3**	Ldhb34f2	Ldhbin4r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	

LDHB4	Ldhb34f1	Ldhb34r1	**LDHB4**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB4	Ldhb34f2	Ldhb34r2	**LDHB4**	Ldhb34f2	Ldhb34r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r3	**LDHB5**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r3	**LDHB6**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB7	Ldhbin6fv1	Ldhb67r2	**LDHB7**	Ldhb67f1	Ldhb67r1
54	1.30 min.	7 min.	51.6	1.30 min.	7 min.
	x30		OUTER	x30	
			LDHB7	Ldhb67f2	Ldhb67r2
			51.6	1.30 min.	7min.
			INNER	x30	
Alouatta seniculus	z231		***Ateles geoffroyi***	z254	
LDHB1	Ldhb1f3	Ldhb1r2	**LDHB1**	Ldhb1NWMF1 v1	Ldhb1Nwmr1
51.6	1.15 min.	7 min.	48	1.15 min.	5 min
OUTER	x30			x35	
LDHB1	Ldhb1nwmf1	Ldhb1nwm r1			
55	1.15 min.	5 min.	**LDHB2**	Ldhb2f1	Ldhb2r1
INNER	x30		51.6	1.15 min.	5 min.
			OUTER	x30	
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f2	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		INNER	x30	
LDHB2	Ldhb2f2	Ldhb2r2			
51.6	1.15 min.	5 min.	**LDHB3**	Ldhb34f1	Ldhb34r1
INNER	x30		51.6	1.15 min.	5 min.
			OUTER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f2	Ldhbin4r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		INNER	x30	
LDHB3	Ldhb34f2	Ldhbin4r2			
51.6	1.15 min.	5 min.	**LDHB4 first**	Ldhb34f2	Ldhb34r2
INNER	x30		51.6	1.30 min.	7 min.
			OUTER	x30	
LDHB4 first	Ldhb34f1	Ldhb34r1	**LDHB4 first**	Ldhb34f3	Ldhbin4r2
51.6	1.15 min.	5 min.	55.4	1.30 min.	7 min.
OUTER	x30		INNER	x30	
LDHB4 first	Ldhbin4f1	Ldhb34r1			
54	1.45 min.	7 min.	**LDHB4 second**	Ldhb34f1	Ldhb34r1
INNER	x30		51.6	1.15 min.	5 min.

			OUTER	x30	
LDHB4 second	Ldhb34f1	Ldhb34r1	**LDHB4 second**	Ldhb34f2	Ldhbin4r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		INNER	x30	
LDHB4 second	Ldhb34f2	Ldhbin4r2			
51.6	1.15 min.	5 min.	**LDHB5**	Ldhb56f2	Ldhb56r2
INNER	x30		51.6	1.45 min.	7 min.
			OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r3
51.6	1.45 min.	7 min.	54	1.45 min.	7 min.
OUTER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r3			
54	1.45 min.	7 min.	**LDHB6**	Ldhb56f2	Ldhb56r2
INNER	x30		51.6	1.45 min.	7 min.
			OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r3
51.6	1.45 min.	7 min.	54	1.45 min.	7 min.
OUTER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r3			
54	1.45 min.	7 min.	**LDHB7**	Ldhb67f1	Ldhb67r1
INNER	x30		51.6	1.30 min.	7 min.
			OUTER	x30	
LDHB7	Ldhb67f1	Ldhb67r1	**LDHB7**	Ldhb67f2	Ldhb67r2
51.6	1.30 min.	7 min.	51.6	1.30 min.	7min.
OUTER	x30		INNER	x30	
LDHB7	Ldhb67f2	Ldhb67r2			
51.6	1.30 min.	7min.			
INNER	x30				
Cebus apella	z248		***Callithrix jacchus***	z118	
LDHB1	Ldhb1NWMF1v1	Ldhb1Nwmr1	**LDHB1**	Ldhb1NWMF1v1	LdhbMmur1r1
48	1.15 min. x35	5 min	51.6	1.15 min. x30	5 min
LDHB2	Ldhb2f1	Ldhb2r1	**LDHB2**	Ldhb2f1	Ldhb2r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB2	Ldhb2f2	Ldhb2r2	**LDHB2**	Ldhb2f2	Ldhb2r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB3	Ldhb34f1	Ldhb34r1	**LDHB3**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB3	Ldhb34f2	Ldhbin4r2	**LDHB3**	Ldhb34f2	Ldhbin4r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB4 first	Ldhb34f1	Ldhb34r1	**LDHB4 first**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	
LDHB4 first	Ldhbin4f1	Ldhb34r1	**LDHB4 first**	Ldhbin4f1	Ldhb34r1
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB4 second	Ldhb34f1	Ldhb34r1	**LDHB4 second**	Ldhb34f1	Ldhb34r1
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
OUTER	x30		OUTER	x30	

LDHB4 second	Ldhb34f2	Ldhbin4r2	**LDHB4 second**	Ldhb34f2	Ldhbin4r2
51.6	1.15 min.	5 min.	51.6	1.15 min.	5 min.
INNER	x30		INNER	x30	
LDHB5	Ldhb56f2	Ldhb56r2	**LDHB5**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB5	Ldhb56f2	Ldhb56r3	**LDHB5**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB6	Ldhb56f2	Ldhb56r2	**LDHB6**	Ldhb56f2	Ldhb56r2
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB6	Ldhb56f2	Ldhb56r3	**LDHB6**	Ldhb56f2	Ldhb56r3
54	1.45 min.	7 min.	54	1.45 min.	7 min.
INNER	x30		INNER	x30	
LDHB7	Ldhb67f1	Ldhb67r1	**LDHB7**	Ldhb67f1	Ldhb67r1
51.6	1.30 min.	7 min.	51.6	1.30 min.	7 min.
OUTER	x30		OUTER	x30	
LDHB7	Ldhb67f2	Ldhb67r2	**LDHB7**	Ldhb67f2	Ldhb67newr1
51.6	1.30 min.	7min.	51.6	1.30 min.	7min.
INNER	x30		INNER	x30	
Daubentonia madagascarensis	z217		***Microcebus murinus***	z167	
LDHB1	Ldhb1newstrepf1	LdhbMmur1r1	**LDHB1**	LdhbMmur1f1	LdhbMmur1r1
54	1 min	5 min	55.4	2.30 min.	7 min.
	x30			x30	
LDHB2	LdhbMmur2f1	LdhbMmur2r1	**LDHB2**	LdhbMmur2f1	LdhbMmur2r1
55.4	2.30 min.	7 min.	55.4	2.30 min.	7 min.
	x30			x30	
LDHB3	LdhbMmur3f1	Ldhb3r1	**LDHB3**	LdhbMmur3f1	LdhbMmur3r1
55.4	2 min.	7 min.	55.4	2.30 min.	7 min.
	x30			x30	
LDHB4	LdhbMmur4f1	LdhbMmur5r1	**LDHB4**	Ldhb3plusintron34F	Ldhbin4r2
56.7	2.30 min.	7 min.	51.6	1.30 min.	5 min.
	x30			x30	Plus Ensembl Blat
LDHB5	Ldhb56f2	Ldhbin6rv1	**LDHB5**	Ldhb56f2	Ldhbin6rv1
51.6	1.45 min.	7 min.	51.6	1.45 min.	7 min.
	x30			x30	
LDHB6 first	Ldhb56f2	Ldhbin6rv1	**LDHB6**	LdhbMmur6f1	LdhbMmur7r1
51.6	1.45 min.	7 min.	56.7	2 min	7 min
	x30			x30	
LDHB6 second	Ldhbin6fv1	LdhbMmur7r1	**LDHB7**	LdhbMmur6f1	LdhbMmur7r1
51.6	1.45 min.	7 min.	56.7	2 min	7 min



Protocol for each species contained within each box. Scientific name given in top left corner of each box. Number following LDHB indicates exon. Names to the right of exon number are primer names, see Appendix C8 for corresponding sequence. Number below exon indicates annealing temperature, followed by extension time, followed by final extension time. ‘x’ followed by a number, generally 30, indicates number of cycles. For those exons in which ‘OUTER’ and ‘INNER’ are found below annealing temperature, a nested PCR approach was used, and both PCR protocols are necessary for amplification.

Appendix C8. Primer sequences for both *LDHA* and *LDHB* exon amplifications.

LDHA	
Ldha1f2	CTAAGGTATGGGCCTTCAC
Ldha1r1	GTCCCTTCTCACTTGAATCC
Ldha1r2	CATTGGTAGTCAATCCATGC
LdhaMmur1f1	GAGGCTATACTTACACCCAAAC
LdhaMmur1r1	GAACTGGGTCTCAAGCCTGAC
Ldha23f1	CCAGTGTTCTGCTTCCACATC
LdhaCex3r1	GCAACTTGCAGTTCGGGCTG
Ldha23nwmf2	GTAGCAGAACTCTCCTATGC
LdhaCEx3f1	CTCCAAGCTGGTCATTATCAC
Ldha34intr1	GGCTACTCCAGATATATGGC
Ldha34intr2	GGATCTTGATATTGATCATGGC
LdhaE3newr1	GCACGTGCCTGTAATCC
Ldha23r1	ATGCAGTCAAAAGCCTCAC
LdhaMmur3f1	CTTGGCTTGACCCTTTCACAC
LdhaMmur3r1	GCTCCAACTATGTTCAGAGG
Ldha45f1	GCCATGATCAATATCAAGATCC
Ldha45r2	GCCACCAAGTTGAAAGCTTCC
Ldha45f2	GCCATATATCTGGAGTAGCC
Ldha45nwmr1	GAACTCCTGAGCTCAAGTGATCC
LdhaCex5r1	GCAACATTCATTCCACTCC
LdhaMmur4f1	CCTCTGAACATAGTTGGAGC
LdhaMmur5r1	CTCAGGATTATCTCAAGTGG
Ldha45f3	CAGGTGTGAGTTGTCATGC
Ldha67f1	CTTGTGATTGGTCAAGCAAGG
Ldha67r2	CTGTGTTCCCTATAGTGAC
Ldha67f2	GTATTCTGCTGGTGTTTGG
Ldha67r1	CATGCTAGTCTTGAACTCC
Ldha6Fv5	CCCTTACCTATGGTTTCC
Ldha3utr1	GGACTAGGCATGTTCAGTGAAGG
LdhaMmur6f1	CCACTTGAGATAATCCTGAG
LdhaMmur7r1	GTGCAGCATTGGCAGTGGTG
LDHB	
Ldhb1f1	GGTGAGATCTAAGCTCACTGC
Ldhb1r1	GAACATCATGTGTCTCCTGATGG
Ldhb1f2	CCTGAGTAGTTAGGACTACAGG
Ldhb1r2	GGACATTCTGCAGAACTCACC
Ldhb1f3	CAAAGTGTTGGGATTACAGG
Ldhb1newf1	GATTACAGGCATGAGCCACC
Ldhb1newr1	GTGGGATTACATCTACAACTTC
Ldhb1nwmf1	CTAGTAGACTATCATTAGGCTG
Ldhb1nwmr1	GTGTACACATGCTGAACTTCTC
Ldhb1nwmf1v1	GCAGATACATGTGTAATG
LdhbMmur1r1	TGGGTTGAGGTGAAAACAGC
Ldhb1newstrepf1	GCAACAGTAACCTTTCAAGG
LdhbMmur1f1	CATGGAATCATAGGCTGAAG
Ldhb2f1	CCTCAAGGATGGACACATTGC
Ldhb2r1	CTGCCAACACCAGATTCACC
Ldhb2f2	CTCTTGAACTAAGCTGGTTTCC
Ldhb2r2	CAGTGATCACTTACGGCTAGC

Ldhb2f3	GACCTTTGGTGCTTAGCC
LdhbMmur2f1	CCATTGCCCAAGGTTGTTTC
LdhbMmur2r1	GCCAACACTATATTCACCAG
Ldhb34f1	GCAGTGTAGCTCAAGGGAG
Ldhb34r1	CATGCCTTTGGATGTGATGC
Ldhb34f2	GAAGCTTATGTATGGTTAGGC
Ldhb34r2	GGACACTATTGAGACAGTTG
Ldhbin4r2	CTTTACCACTTGAGTCGCCATG
LdhbMmur3f1	TCAGTGTAGCCCAAGGAAG
LdhbMmur3r1	CTACTATGCTCACCAACATGC
Ldhbin4f1	GACATTCTTACRTATGTTACCTG
Ldhb34f3	CTGCAGGAGTCCGTCAGCAAG
LdhbMmur4f1	CCTTACCTTCCTTGTGTGCAC
LdhbMmur5r1	CTGGGGCAACTGGGATAGCAAG
Ldhb3plusintron34f	GTGGTTTCCAACCCAGGTATTG
Ldhb56f2	GCAGTTCCAGTTGGTATTCAG
Ldhb56r2	CCTTGYTGTCCAAGACTTC
Ldhb56r3	CCTTGTTGACCAAGACTTCTG
Ldhbin6rv1	GATACTGGATGAATCCTGG
LdhbMmur7r1	GTTCAAGGGCTCAGACTGCAAG
LdhbMmur6f1	GCTCTGATTTCAGTGATCTC
Ldhbin6fv1	GTGCCTATGAAGTCATCAAGC
Ldhb67f1	CARTGGCAAGTGGGACCTC
Ldhb67r1	GTCTACCTAAGTTTCCCTG
ldhb67f2	CAAGTGGGACCTCACAATTGAG
Ldhb67r2	GTGTGGTAAGAATGACTTGG
Ldhb67newr1	GCAAACTGCGATCCATGTAC

APPENDIX D

Supplementary Text

Supplementary Results

Both *GHC* and *GHD* share a similar splice variant that fails to excise canonical intron 4 (EU935075 and EU935081, respectively; Appendix D1A). The remaining splice variants are found only for *GHC* transcripts. Two of these splice variants (*GHC* transcript variant 3, EU935077, and transcript variant 4, EU935078) result in frameshifts coding for early to mid-nonsense substitutions. *GHC* transcript variant 5 (EU935079) does not include the canonical exon 4, resulting in an ORF of 55 fewer amino acids (Appendix D1B). In addition, 2 distinct *GHC* transcripts (*GHC* transcript variant 1b, EU935074, and *GHC* transcript variant 2b, EU935076) differ from the other *GHC* transcripts at 6 and 5 nucleotide sites, respectively (likely attributable to an incomplete *GHC* transcript variant 2b; Appendix D1). Two of these differing sites are within the canonical coding region and are dN substitutions, 2 are within the canonical intron 4 (one substitution results in a missense replacement and the other in a nonsense replacement), and 2 are within the 3' untranslated region (Appendix D1). Our transcript data cannot distinguish whether these differing sites represent variable sites within a single gene (i.e., heterozygotic sites) or whether the 2 distinct transcripts result from recent gene duplication. Conservatively, we currently consider these 2 transcripts as originating from a single gene. We used the more conserved of the 2 putative haplotypes for subsequent phylogenetic and molecular evolution analyses. Appendix D2 lists the transcript variants and their accession numbers.

The gene and species trees differ by 9 maximum parsimony (MP) steps. MP and maximum likelihood (ML) topology tests (Appendix D3 and D4) show that the gene tree (i.e., Fig. 2) is a significantly better fit of the data than the species tree (i.e., Appendix D5; $P < 0.05$). We reconstructed ancestral states on lineages using the min-f and ML methods in PAUP* (Swofford 2002). Significantly more sites support the grouping of anthropoids and laurasiatherians than support a monophyletic primate clade ($P = 0.031$, 1-tail Fisher's exact test). MP ancestral state reconstructions on the gene tree show that 14 of 1,057 total steps support an anthropoid + laurasiatherian clade. On the species tree, only 5 of 1,066 steps support a strepsirrhine + anthropoid clade. Moreover, ML reconstructions on the gene tree indicate 12 substitutions on the branch leading to the LCA of anthropoids and laurasiatherians. Zero ML inferred substitutions support the anthropoid + strepsirrhine grouping in the species tree, with only 2 ML inferred substitutions supporting the rodent + primate grouping.

SI Methods

Primers were designed using previously sequenced genes from the Black-Handed Spider monkey (*Ateles geoffroyi*; AF374234, AF374235, and AY435434). These primers were used for both the Spider monkey and Olive baboon (R-5'RACE-CCAGTCTGGGGGCTGCCATCTTCC- and F-3'RACE2-CCAGGTGGCCTTTGACACCTACCAG-). The 5' and 3' RACE protocols were followed according to the manufacturer's instructions for the SMART RACE cDNA Amplification Kit (Clontech), with final volumes of 25 μL. RACE reactions

were performed in an Eppendorf thermal cycler under the following conditions: 25–30 cycles of 94 °C for 30 s and annealing temperatures of either 65 °C or 68 °C for 30 s and 72 °C for 3 min. Amplified products were run on 1% agarose gels, plugged, and purified using the Gel Extraction Kit (Qiagen). Additional amplification primers for the Spider monkey were designed based on sequencing results aimed at targeting specific GH genes and are as follows:

F-3'race1B-CCAGCTGGCCTTTGACATCTACCAG-,

R-5'race1B-CAGTCTGGGGGCTGACCTCTTCC-,

GH Bridge 3'RACE-F1-CTCCAGGCCTTTCTCTACACCATG-,

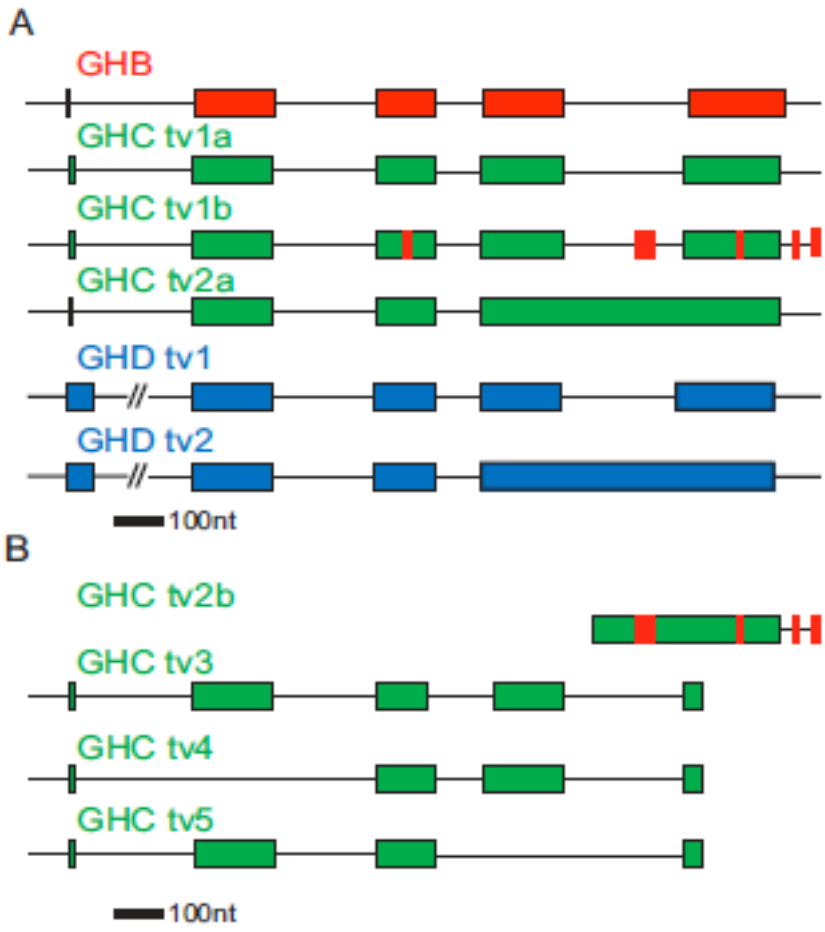
AF34 5'RACE-R1-CTCTAGGCTGGATTTTGCCAGCAC-,

GH PARTIAL 5'RACE-R1-CAAGTGCTTGGACACCGTTACCTC-

We isolated plasmids with standard protocols using QIAprep Spin Miniprep and DirectPrep 96 Miniprep Kits (Qiagen).

ML trees and nodal supports were inferred with Paup* v4.0b10 (Swofford 2002) using the canonical transcripts for each GH-related gene and species. We used Modeltest v3.7 (Posada and Crandall 1998) to estimate the best-fit model for the sequences. Based on the Akaike Information Criterion, a SYM $+ \gamma$ model was selected with γ-distribution shape parameter α = 1.6030, an R matrix (1.0959, 5.0778, 1.0169, 1.4816, and 3.7177), and equal base frequencies. The ML criterion, with 500 bootstrap replicates, starting trees obtained via stepwise addition with as-is addition sequence, and nearest-neighbor interchange algorithm, was used.

Appendix D1. GH transcripts from Spider monkey placenta



Boxes indicate exons, horizontal lines indicate introns, and red vertical lines indicate variable sites. Forward slash indicates a longer genetic distance than can be illustrated to scale. A marmoset gneomic sequence (Wallis and Wallis 2006) was used as a reference to identify intron/exon boundaries. (A) Full-length transcripts for *GHB*, *GHC*, and *GHD*. (B) Partial *GHC* transcripts.

Appendix D2. List of all species and GH genes included in phylogenetic reconstruction

Table S2. List of all species and GH genes included in phylogenetic reconstruction

Species and gene symbol	Common name and gene symbol	Accession no.
Homo sapiens GH1	Human *GH1*	NM_000515
Homo sapiens GH2	Human *GH2*	NM_002059
Homo sapiens CSH1	Human *CSH1*	NM_001317
Homo sapiens CSH2	Human *CSH2*	NM_020991
Homo sapiens CSHL1	Human *CSHL1*	NM_022579
Macaca mulatta GH1	Macaque *GH1*	NM_001042738
Macaca mulatta GHV	Macaque *GHV*	NM_001042739
Macaca mulatta CSH1	Macaque *CSH1*	NM_001047156
Macaca mulatta CSH3	Macaque *CSH3*	NM_001047153
Macaca mulatta CSH4	Macaque *CSH4*	NM_001047155
Papio anubis GHV	Olive baboon *GHV*	FJ041322 (this study)
Papio anubis CSH1	Olive baboon *CSH1*	FJ041323 (this study)
Ateles geoffroyi GHA	Black-Handed Spider monkey *GH1*	AF374234
Ateles geoffroyi GHB	Black-Handed Spider monkey *GH2*	AF374235
Ateles geoffroyi GHC	Black-Handed Spider monkey *GHC*	AY435434
Ateles fusciceps GHB	Brown-Headed Spider monkey *GHB*	EU935072 (this study)
Ateles fusciceps GHC transcript variant 1a	Brown-Headed Spider monkey *GHC* transcript variant 1a	EU935073 (this study)
Ateles fusciceps GHC transcript variant 1b	Brown-Headed Spider monkey *GHC* transcript variant 1b	EU935074 (this study)
Ateles fusciceps GHC transcript variant 2a	Brown-Headed Spider monkey *GHC* transcript variant 2a	EU935075 (this study)
Ateles fusciceps GHC transcript variant 2b	Brown-Headed Spider monkey *GHC* transcript variant 2b	EU935076 (this study)
Ateles fusciceps GHC transcript variant 3	Brown-Headed Spider monkey *GHC* transcript variant 3	EU935077 (this study)
Ateles fusciceps GHC transcript variant 4	Brown-Headed Spider monkey *GHC* transcript variant 4	EU935078 (this study)
Ateles fusciceps GHC transcript variant 5	Brown-Headed Spider monkey *GHC* transcript variant 5	EU935079 (this study)
Ateles fusciceps GHD transcript variant 1	Brown-Headed Spider monkey *GHD* transcript variant 2	EU935080 (this study)
Ateles fusciceps GHD transcript variant 2	Brown-Headed Spider monkey *GHD* transcript variant 2	EU935081 (this study)
Callithrix jacchus GHA	Common Marmoset *GHA*	AJ297563
Callithrix jacchus GHB	Common Marmoset *GHB*	AM235212
Callithrix jacchus GHC	Common Marmoset *GHC*	AM235212
Callithrix jacchus GHD	Common Marmoset *GHD*	AM235212
Galago senegalensis GH	Senegal Bushbaby *GH*	AF292938
Nycticebus pygmaeus GH	Pygmy Slow Loris *GH*	AJ297562
Mus musculus GH	Mouse *Gh*	NM_008117
Rattus norvegicus GH	Rat *GH*	NM_001034848
Bos taurus GH	Cow *GH*	NM_180996
Canis lupus familiaris GH	Dog *GH*	NM_001003168
Monodelphis domestica GH	Gray Short-Tailed Opposum *GH*	NM_001032993

Appendix D3. Parsimony tests of alternative topologies

Tree	Length	Kishino-Hasegawa test				Templeton test			
		Difference	SD (difference)	*t*	*P**	Rank sums †	*N*	*z*	*P*‡
Species	1,066	9	4.34912	2.0694	0.0389	140.0	19	-2.0647	0.0389
Gene	1,057	Best				Best			

*Probability of getting a more extreme t value under the null hypothesis of no difference between the 2 trees (2-tailed test).
†Wilcoxon signed-rank test statistic is the smaller of the absolute values of the 2 rank sums.
‡Approximate probability of getting a more extreme test statistic under the null hypothesis of no difference between the 2 trees (2-tailed test).

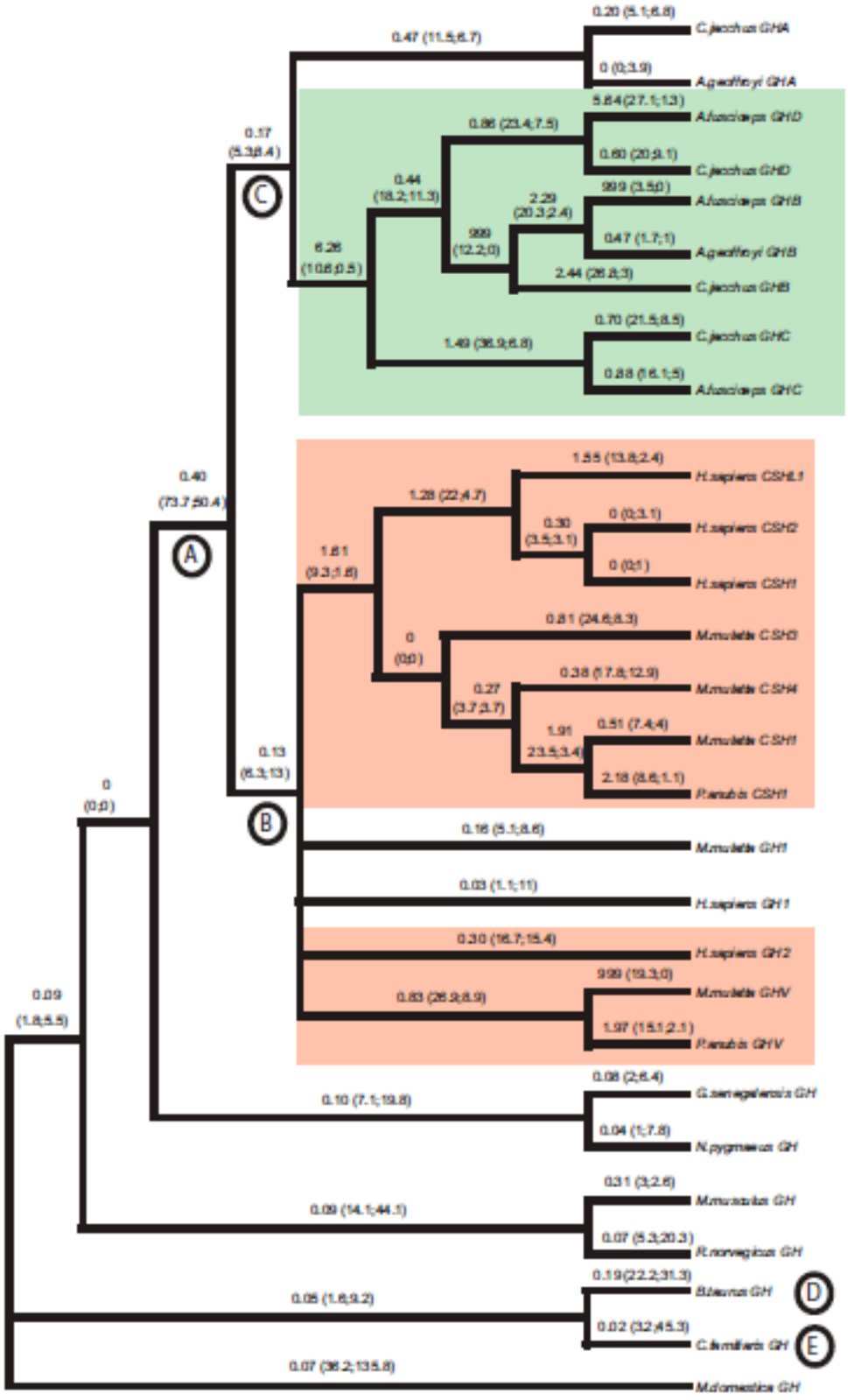
Appendix D4. Likelihood tests of alternative topologies

Shimodaira-Hasegawa test

Tree	-ln L	Diff –ln L	*P*
Species	6,260.82919	13.16114	0.020
Gene	6,247.66805	Best	

Shimodaira-Hasegawa test using RELL bootstrap (1-tailed test). Number of bootstrap replicates is 1,000.

Appendix D5. Adaptive evolution in GH-related genes



The free ratio model (codeml model 1) was used to calculate ω values on each branch of the species tree. The ω values, and the ML estimates fo the number of dN (N*dN) and dS (S*dS) substitutions are shown along each branch. Placentally expressed catarrhine GH-related genes and their ancestral lineages are boxed in salmon, and placentally expressed platyrrhine GH-related genes and their ancestral lineages are boxed in green. Branches A-E were used to test hypotheses regarding divergence times (see the text). Appendix D2 contains a list of scientific names and accession numbers for the sequences used in this study.

Appendix D6. Number of colonies sequenced for each platyrrhine GH transcript variant for each placental tissue type

	Total no. colonies	Membranous	Villous
GHB	2	0	2
GHC transcript variant 1a	52	32	20
GHC transcript variant 1b	39	16	23
GHC transcript variant 2a	10	0	10
GHC transcript variant 2b	2	0	2
GHC transcript variant 3	1	1	0
GHC transcript variant 4	2	1	1
GHC transcript variant 5	1	1	0
GHD transcript variant 1	87	45	42
GHD transcript variant 2	12	1	11

Appendix D7. Inferred amino-acid replacements in the GH gene on the branch leading to the LCA of anthropoids (branch A in Fig.4)

Amino-acid position	Ancestral state	Derived state	Posterior probability (ancestral state/derived state)
3	P	L	1/1
4	M	I	1/1
8	S	R	1/1
11	A	D	1/1
14	V	M	1/1
18	Q	H	1/1
19	H	R	1/1
25	A	F	1/1
29	K	Q	1/1
33	R	E	1/1
38	E	K	1/1
39	G	E	1/1
41	R	K	1/1
45	I	F + L	1/1
48	A	P	0.97/1
50	A	T	1/1
51	A	S	1/1
52	F	L	1/1
57	T	S	1/1
62	T	S	1/0.78
63	G	N	1/0.88
65	D	E	1/0.97
67	A	T	1/1
70	R	K	1/1
72	D	N	1/1
78	F	I	1/1
88	G	E	1/1
94	S	R	1/0.92
95	R	S	1/1
98	T	A	1/1
103	F	Y	1/1
105	T	A	1/0.84
108	R	S	1/1
123	A	T	1/1
126	R	G	1/1
127	E	R	1/1
135	A	T	0.96/1
139	L	F	1/1
144	D	S	1/1
150	L	S	1/1
151	R	H	1/1
152	S	N	1/1
164	S	Y	1/1
167	K	R	0.99/1
170	L	M	1/1
171	H	D	1/1
173	A	V	1/1
176	Y	F	1/1
179	V	I	1/1
180	M	V	1/1
181	K	Q	1/1
184	F	S	1/1
187	S	G	1/1
190	A	G	1/1

Reconstruction is based on marginal reconstruction using PAML v3.15 (Yang 1997). Amino-acid position, ancestral state, derived state on branch leading to LCA of anthropoids, and posterior probabilities (ancestralA.A./derivedA.A.) are listed. Position 45 includes the I45F replacement as well as a leucine (L) insertion.

Appendix D8. Alignment of GH sequences

#NEXUS

BEGIN DATA;

DIMENSIONS NTAX=29 NCHAR=729;

FORMAT DATATYPE=DNA MISSING=? GAP=- ;

MATRIX

HsapGH1 ----------ATGGCTACAGGCTCC---
CGGACGTCCCTGCTCCTGGCTTTTGGCCTGCTCTGCCTGCCCTGGCTTCAAGAGGGCAGT
GCCTTCCCAACCATTCCCTTATCCAGGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTC
TGCACCAGCTGGCCTTTGACACCTACCAGGAGTTTGAA----------
GAAGCCTATATCCCAAAGGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTC
TGTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAACAGAAATCCAAC
CTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC
CTCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGAC
CTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGCAG
CCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAAACTCACACAA
CGATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAA
GGTCGAGACATTCCTGCGCATCGTGCAGTGCCGC---
TCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

HsapGH2 ----------ATGGCTGCAGGCTCC---
CGGACGTCCCTGCTCCTGGCTTTTGGCCTGCTCTGCCTGTCCTGGCTTCAAGAGGGCAGT
GCCTTCCCAACCATTCCCTTATCCAGGCTTTTTGACAACGCTATGCTCCGCGCCCGTCGCC
TGTACCAGCTGGCATATGACACCTATCAGGAGTTTGAA----------
GAAGCCTATATCCTGAAGGAGCAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTC
TGCTTCTCAGAGTCTATTCCAACACCTTCCAACAGGGTGAAAACGCAGCAGAAATCTAACC
TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCATGGCTGGAGCCCGTGCAGCTCC
TCAGGAGCGTCTTCGCCAACAGCCTGGTGTATGGCGCCTCGGACAGCAACGTCTATCGCC
ACCTGAAGGACCTAGAGGAAGGCATCCAAACGCTGATGTGGAGGCTGGAAGATGGCAGC
CCCCGGACTGGGCAGATCTTCAATCAGTCCTACAGCAAGTTTGACACAAAATCGCACAAC
GATGACGCACTGCTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAG
GTCGAGACATTCCTGCGCATCGTGCAGTGCCGC---
TCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

HsapCSH1 ----------ATGGCTCCAGGCTCC---
CGGACGTCCCTGCTCCTGGCTTTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGGCTGGTG
CCGTCCAAACCGTTCCGTTATCCAGGCTTTTTGACCACGCTATGCTCCAAGCCCATCGCGC
GCACCAGCTGGCCATTGACACCTACCAGGAGTTTGAA----------
GAAACCTATATCCCAAAGGACCAGAAGTATTCATTCCTGCATGACTCCCAGACCTCCTTCT
GCTTCTCAGACTCTATTCCGACACCCTCCAACATGGAGGAAACGCAACAGAAATCCAATC
TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCGAGTCGTGGCTGGAGCCCGTGCGGTTCC
TCAGGAGTATGTTCGCCAACAACCTGGTGTATGACACCTCGGACAGCGATGACTATCACC
TCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGACGGCAGC
CGCCGGACTGGGCAGATCCTCAAGCAGACCTACAGCAAGTTTGACACAAACTCGCACAAC
CATGACGCACTGCTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAG
GTCGAGACATTCCTGCGCATGGTGCAGTGCCGC---
TCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

HsapCSH2 ----------ATGGCTGCAGGCTCC---

CGGACGTCCCTGCTCCTGGCTTTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGGCTGGTG
CCGTCCAAACCGTTCCGTTATCCAGGCTTTTTGACCACGCTATGCTCCAAGCCCATCGCGC
GCACCAGCTGGCCATTGACACCTACCAGGAGTTTGAA----------
GAAACCTATATCCCAAAGGACCAGAAGTATTCATTCCTGCATGACTCCCAGACCTCCTTCT
GCTTCTCAGACTCTATTCCGACACCCTCCAACATGGAGGAAACGCAACAGAAATCCAATC
TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCGAGTCGTGGCTGGAGCCCGTGCGGTTCC
TCAGGAGTATGTTCGCCAACAACCTGGTGTATGACACCTCGGACAGCGATGACTATCACC
TCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGACGGCAGC
CGCCGGACTGGGCAGATCCTCAAGCAGACCTACAGCAAGTTTGACACAAACTCACACAAC
CATGACGCACTGCTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAG
GTCGAGACATTCCTGCGCATGGTGCAGTGCCGC---
TCTGTAGAGGGTAGCTGTGGCTTCTAG [654]

HsapCSHL1 ----------ATGGCTGCAGGCTCC---
CGGACGTCCCTGCTCCTGGCTTTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGGCTGGTG
CCGTCCAAACCGTTCCCTTATCCAGGCTTTTTAAAGAGGCTATGCTCCAAGCCCATCGCGC
ACACCAGCTGGCCATTGACACCTACCAGGAGTTT---
ATAAGCTCTTGGGGAATGGAAGCCTATATCACAAAGGAACAGAAGTATTCATTCCTGCATG
ACTCCCAGACCTCCTTCTGCTTCTCAGACTCTATTCCGACATCCTCCAACATGGAGGAAAC
GCAGCAGAAATCCAACTTAGAGCTGCTCCACATCTCCCTGCTGCTCATCGAGTCGCGGCT
GGAGCCCGTGCGGTTCCTCAGGAGTACCTTCACCAACAACCTGGTGTATGACACCTCGGA
CAGCGATGACTATCACCTCCTAAAGGACCTAGAGGAAGGCATCCAAATGCTGATGGGGAG
GCTGGAAGACGGCAGCCACCTGACTGGGCAGACCCTCAAGCAGACCTACAGCAAGTTTG
ACACAAACTCGCACAACCATGACGCACTGCTCAAGAACTACGGGCTGCTCCACTGCTTCA
GGAAGGACATGGACAAGGTCGAGACATTCCTGCGCATGGTGCAGTGCCGC---
TCTGTGGAGGGCAGCTGTGGCTTCTAG [669]

MmulGH1 ----------ATGGCTGCAGGCTCC---
CGGACATCCCTGCTCCTGGCTTTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGGGCAGT
GCCTTCCCAACCATTCCCTTATCCAGGCTTTTTGACAACGCTATGCTCCGCGCCCATCGCC
TGCACCAGCTGGCCTTTGACACCTACCAGGAGTTTGAA----------
GAAGCCTATATCCCAAAGGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTC
TGCTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAGCAGAAATCCAAC
CTAGAGCTGCTCCGCATCTCCCTGCTTCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCC
TCAGGAGTGTGTTTGCCAACAGCCTGGTGTATGGCACCTCGTACAGTGACGTCTATGACCT
CTTAAAGGACCTAGAGGAAGGCATCCAAACACTGATGGGGAGGCTGGAAGATGGCAGCT
CCCGGACTGGACAGATCTTCAAGCAGACCTACAGCAAGTTTGACACAAACTCACACAACA
ATGATGCACTGCTCAAGAACTACGGGCTGCTCTATTGCTTCAGGAAGGACATGGACAAGA
TCGAGACATTCCTGCGCATTGTGCAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

MmulGHV ----------ATGGCTGCAGGCTCC---
TGGACATGCCTAATTCTGGCTATTGCCCTGCTCTGCCTACCTTGGCTTCAAGAGGGCAGTG
CCTTCCCAACCATTCCCTTATCCTGGCTTTTTAACACCGCTGTGTTCCGCGCCCATCACCTG
CACAAGCTGGCATTTGACACGTACCCGAAGTTTGAA----------
GAAGCCTATATCCCAAAGGAACAGAAGTATTCATTCCTGCGCAACCCCCAGACCTCCCTC
TGCTTCTCAGAGTCTATTCCGACACCCTCCAACAAGGAGGAAACACAGCAGAAATCCAAC
CTAGAGCTGCTCCACATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC
CTCAGGAGTGTGTTTGCCAACCACCTGGTGCATACCAACTCGAACTTCGACATCTATCTCT
ACCTAAAGAAACTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGACGGCAGC
CCCCGGACTGGGCAGATCTTCAAGGAGACCTACAGCAAGTATGACACAAACTCGCACAAC
GATGACACACTGCTCAAGAACTACAGGCTGCTCTACTGCTTCAGGAAGGACATGAACAAG
GTCGAGACATTCCTGCGCACCGTGCGGTGCCGC---
GCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

MmulCSH1 ----------ATGGCTGCAGGCTCC---

CGGACATCCCTGCTCCTGGTTTTTGCCCTCCTCTGCCTGCCTTGGCTTCAAGAGGGCGGTG
GGTTCCCAAGCGTACCCTTATCCAGACTTTTTGACCATGCTATGATCCAAGCCCATCGCCT
GCACCAACTGGCCTTTGACACCTACCAGGAGTTTGAA----------
GAAGCCTATATCCCAAAGGAAAAGAAGCATTCGCTCATGGAGAACCCCCAGGCCTCCTTC
TGCTTCGCAGACTCTATTCCCACACCCTCCAGCTTGGAGGAAACGCAGCAGAAATCCAAC
CTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC
CTCAGTAGTGTGTTTGCCAACAACCTGTTGCATCACACCTCGGACAGCGACGTCCATGAC
CTCCTAAAGGACCTAGAGGAAGGCATCCAAACGTTGATGTGGAGGCTGGAAGATGGCAG
CCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTTGACACACACTCACAGAA
CGATGACTCACTGCTCAAGAACTACAGGCTGCTCCACTGCTTCAGGAAGGACATGGATAT
GGTCGAGACATTCCTGCGCATGGTGCAGTGCCGC---
GCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

MmulCSH3 ----------ATGGCTGCAGGCTCC---
CGGACCTCCCAGATCCTGGCTTTTGCCCTGCTCTGCCGGCCATGGCTTCAAGAGGCCGGT
GCCGTCCAATCCGTTCCCTTATCCAGGTTTTTTGACAACGCTATGCTCCACGCCCATCGCC
TGCACCAGCTGGCCTTTGACACCTACCAGAAGTTTGAA----------
GAAGCCTATATCCCAAAGGAAAATAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTCT
GCTTCTCAGAGTCTATTCCCACACCCTCCAACATGGAGGAAAGGCAGCAGAAATCTAACC
TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCT
CAGGAGTGTGTTTGCCAACAGGCTAGTGTGTGGCACCTCGGACAGCGACGTCTATGACCT
CCTAAAAAGCCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGCAGCC
CCAGGACTGGGCAGATCTTCAAGGCGACCTACATCAAGTTTGACACAAACTCACACAACA
ATGATGCACTGCTCGAGAACTACAGGCTGCTCCACTGCTTCAGGAGGAACATGGATAAGG
TCGAGACATTCCTGCGCATGGCGCAGTGCCGC---
TCTGTGGAGGGCAGCTGTGGCTTCTAG [654]

MmulCSH4 ----------ATGGCTGCAGGCTCC---
CAGACGTCCCTGCTCCTGGCTTTTGCCCTCCTCTGCCTGCCTTGGCTTCAAGAGGCTGGTA
GGGTCCCAAGCGTACCCTTATCCAGGCTTTTTGACAACGCTATGATCCACGCCTATCGCCT
GCACCGGCTGGCCTTTGACACCTACCAGGAGTTTGAA----------
GAAGCCTATATCCCAAAGCAACAGAAGTATTCATTCCTGCGGAAACCCCAGACTCCTCTCT
GGCTCTCAGAGTCTATTCCCACACCCTCCAACATGGAGGAAACACAACAGAAATCCAACC
TAGAGCTGCTCCACATCTCCCTACTGCTCATCCAGTCCTGGCTGGAGCCCGTGCAGTTCCT
CAGGAGTGTGTTTGCCAACAGGCTGGTGTATGGCACCTCAGACAGCGACGTCTATGACCT
CCTAAAAAGCCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGTAGCC
CCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCACGTTTGACACAAACTCGCACAATG
ATGACACACTGCTCAAGAACTACTGGCTGCTCCACTGCTTCAGGAAGGACATGGACAAGG
TCGAGACATTCCTGCGCATGGTGCAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

PanuGHV ----------ATGGCTGCAGGCTCC---
GGGACATGCCTAATTCTGGCTATTGCCCTGCTCTGCCTACCTTGGCCTCAAGAGGGCAGT
GCCTTCCCAACCATTCCCTTATCCTGGCTTTTTAACAGCGCTATACTCCACGCCCATCACCT
GCACAAGCTGGCATTTGAAACGTACCAGAAGTTTGAA----------
GAAGCCTATATCTCAAAGGAACAGAAGTATTCATTCCTGCGCAACCCCCAGACCTCCTTCT
GCTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACGCAGCAGAAATCCAACC
TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCT
CAGGAGTGTGTTTACCAACAAGCTGGTGTATGGCGCCTCGAACTTCAACGTCTATCTCTAC
CTAGCGAACCTAGAGGAAGGCATCCAAACGCTGATGTGGAGGCTGGACGACGGCAGTCC
CCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTTGACACAAACTCGCACAACGA
TGACACACTGCTCAAGAACTACTGGCTGCTCTACTGCTTCAGGAAGGACATGAACAAGGT
CGAGACATTCCTGCGCACCGTGCAGTGCCGC---GCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

PanuCSH1 ----------ATGGCTGCAGGCTCC---

CAGACATCCCTGCTCCTGGTTTTTGCCCTCGTCTGCCTGCCTTGGCTTCAAGAGGCTGGTA
GGGTCCCAAGCGTACCCTTATCCAGGCTTTTTGACCACGCTATGATCCAAGCCCATCGCCT
GCACCAACTGGCCTTTGACACCTACCAGGAGTTTGAA----------
GAAGCCTATATCCCAAAGGAAAAGAAGCATTCGATCATGGAGAACCCCGAGGCCTCCTAC
TGCTTCGCAGACTCTATTCCCACACCCTCCAACTTGGAGGAAACGCAGCAGAAATCCAAC
CTAGAACTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC
CTCAGTAGTGTGTTTGCCAACAACCTGGTGCATCACAGCTCGGACAGCGACGTCCATGAC
CTCCTAAAGGACCTAGAGGAAGGCATCGAAACGTTGATGGGGAGGCTGGAAGATGGCAG
CCCCCRGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTTGACACACACTCACAGAA
CGATGACTCACTGCTCAAGAACTACGGGCTGCTCCACTGCTTCAGGAAGGACATGGACAT
GGTCGAGACATTCCTGCGCATGGTGCAGTGCCGC---
ACTGTGGAGGGCAGCTGTGGCTTCTAG [654]

CjaccGHA ----------ATGGCTGCAGGCTCC---
TGGACATCCCTGCTCCTGGCTTTCACCCTGCTCTGCCTGCCCCAGCTTCGAGAGGCCGGT
GCCTTCCCAACAATTCCCTTATCCAGGCTTTTGGACAATGCTATGCTCCGTGCCCATCGCC
TGCACCAGCTGGCCTTTGACACCTACCAGGAGTTCGAA----------
GAAGCCTATATCCCGAAGGAACAGAAGTATTCCTTCCTGCAGAACCCCCAGACCTCCCTC
TGCTTCTCAGAGTCTATCCCGACACCCGCCAGCAAAAAGGAAACTCAGCAGAAATCCAAC
CTAGAGCTGCTCCGCATGTCCCTGCTGCTCATCCAGTCCTGGTTCGAGCCTGTGCAGTTCC
TCAGAAGTGTCTTCGCCAACAGCCTATTGTATGGTGTCTCAGACAGCGACGTCTATGAGTA
CCTGAAGGACCTAGAGGAAGGCATCCAAACTCTGATGGGGAGGCTGGAAGATGGCAGCC
CCCGGACTGGGGAGATCTTCATGCAGACCTACAGGAAGTTTGACGTCAACTCGCAGAACA
ATGATGCACTGCTCAAGAACTACGGGCTGCTCTACTGCTTCCGGAAGGACATGGACAAGG
TCGAGACGTTCCTGCGCATTGTGCAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

CjaccGHB ----------ATGGCTGCAGGCTCC---
CGGATGTCCCTGCTCATGGCTTTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGACTGGTG
CCCTTCCAAGAATTCCCTTATCCAGGCTTTTTGGTGACGCTATGCTCCGTGCCCGTCAGCT
GCACCATCTGGCCTTGGAAACCTACCGGGAGTTGGAA----------
AAAAATTGTGTCCCGAAGGAACAGAAGTATTTCTTCCTGCGTAACCCCGAGACCTTCGTCT
GCTTCTCAGAGTCTATCCCAACACCCTTCCACAAGGAGGAAATGCTGGGAAAATCCAACG
TAGAGCTGCTCCACATCTCCCTGCTGCTCATCCAGTCCTGGCTTGAACCCATGCAGCGCCT
CGGCAGTATCTTTGCCAACAGCCAACTGCATAGCATCGTGAACACTGATGTCTATGAGTAC
CTGAAGGACCTAGAGGAAGGCATCCAAACTCTGACGGGGAGGCTGGAAGATGGCAGCCC
CCAGACTGGGGAGATCTTCAGGCAGACCTACAGCAAGTTTGACAGAAGCCTGCACAACG
ATGACACACTGCTCAAGAACTACTGGCTGCTCTTCTGCTTCCGGAAGGACATGAGCAAGG
TCGAGACATTCCTGCGCATTGTGCAGTGCCAC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

CjaccGHC ----------ATGGCTGCAGTCTCC---
CCGGCATCCCTGCTCCTGACTTTCACCCTGCTCTGCCTGCCCTGGCTTCGAGAGGCCGGT
GCCTTCCAATCAATTCCCTTATCCTCGCTTTATGACTATGCTGTGATCCGCGCCTATCGCCT
CAACCACCTGGCCTTTGACATCTACCAGAAGTTTGAA----------
GAAGCTCGTAGCCCGAAGGAACAGAAGAATTTCTTCCAGTTTAACGCCAGGACCTCCCTT
TGCTTCTCAGCGTCTGTCCCAACACCCACTAACAGAAAGGAAACTCTGCAGAAATCTAAC
CTAGAGCTGCTCCAAAACTCCCTGCTGCTCATCCAGATGTGGCTCAAGCCCATGCAGTCC
CTCAACAGTGTCTTTGCCAACAGCCAACAGCATAGTGTCTCAAACAGCTTCATCTATGAGT
ACCTGAAGGATCTAGAGGAAGTCATCCAAACTCTGATGGGGAGGCTGGAAGATGGCAGC
CCCTGGACTGGGGAGATCTTCAGGCAGACCTACAGCAAGTTTGACAGAAACCTGCACAAC
GATGATGCAGTGCTCAAGAACTATGGGCTGCTCTACTGCTTCCGGAGAGACATGAACAAG
GTCGCAACATTCCTGCGCATTGTGAAGTGCCGT---
GCTGTGGAGGGCAGCTGTGGCCTCTAg [654]

CjaccGHD

ATGGGAGGAAGTGCTTGTGGTCCGTGTCTGTTGCCGGGATTTCTGTTTCTT---------
GGCTCC---
CAGATGTCCCTGCTCCTGGCTTTCATCCTGCTTTGCCTGCCCTGGCTTCAAGAGACTGGTG
CCCTTCCACGCATTCCCTTATCCAGGCTTTTGGAAGACGCTGTGCTCCATGCCCATCAGCT
GCATCAGCTGGCCTTTGACACCTACCAGGAGTTGGAA----------
GAAGATTGTATCCCAAATGAACAGAAGTATTTCTTCCTGCAGAACCCCAAGACCTCCCTCT
GCTTCTCAGAGTCTATCCTGACACCCTCAAACAAGGAGGAAATGCTGGCGAAATCTAACC
TAGAGCTGCTCCACATCTCCCTGCTGTTCATCCAGTCCTGGCTCAAACCCGTGCAGCTCCT
CAGGGGTGTCTTTGCCAACAGTGAACTGCATAGGGTCTCGAACACCAACATCTATGAGTC
CCTGAAGGACCTAGAAGAAGGCATTCAAACTGTGATGGGGAGGCTGGAAGAGGTCAGCC
ACCAGACTGGGGAGATCATCAAGCAAAATTGCAGCATGTTTGACACAAACTTGCACAACA
ATGACACACTGCTCAAGAGCTACTGGCTGCTCTATTACTTCAGGAAGGACATGCACAAGG
TCAAGACATTCCTGCGCATTGTGCAGTGCCTC---TCTGTGGAGGGCCTCTGTGGCCTCTAG
[696]

AgeofGHA ----------ATGGCTGCAGGCTCC---
CGGACATCCCTGCTCCTGGCTTTCACCCTGCTCTGCCTGCCCCAGCTTCAAGAGGCTGGT
GCCTTCCCAACAATTCCCTTATCCAGGCTTTTGGACAATGCTATGCTCCGCGCCCATCGCC
TGCACCAGCTGGCCTTTGACACCTACCAGGAGTTCGAA----------
GAAGCCTATATCCCGAAGGAACAGAAGTATTCCTTCCTGCAGAACCCCCAGACCTCCCTC
TGCTTCTCAGAGTCTATCCCGACACCCGCCAGCAAAAAGGAAACTCAGCAGAAATCCAAC
CTAGAGCTGCTCCGCATCTCCCTGCTGCTTATCCAGTCCTGGTTCGAGCCTGTGCAGTTCC
TCAGGAGTGTCTTCGCCAACAGCCTATTGTATGGCGTCTCAGACAGCGATGTCTACGAGTA
CCTGAAGGACCTAGAGGAAGGCATCCAAACTCTGATGGGGAGGCTGGAAGATGGCAGCC
CCCAGACTGGGGAGATCTTCAGGCAGACCTACAGGAAGTTTGACATAAACTCGCAAAACA
ATGACGCATTGCTCAAGAACTACGGGCTGCTCTACTGCTTCCGGAAGGACATGGACAAGG
TCGAGACGTTCCTGCGCATTGTGCAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

AgeofGHB ----------ATGGCTGCAGGCTCC---
CGGATGTCCCTGCTCCTGACTCTTGCCCTGCTCTGCCTGCCCTGGCTTCAAGAGACTGGTG
CCTTTCCAAGAATTCCCTTATCCAGGCTTTTTGGTGACGCTATGCTCCGTGCCCATCAGCT
GCACCAGGTGGCCTTTGACACCTACCAGGAGTTGGAA----------
GAAAATTGTATCCCGAAGAAACAGAAGTATTTCTTCCTGCGTAACCCCAAGAACTTCCTCT
GCTTCTCAGAGTCTATCCCGACACCCTTCAACAAGGAGGAAGTGCTGGCAAAATCCAGCC
TAGAGCTGCTCCACATCTCCCTGCTGCTCATCCAGTCCTGGCTCGAACCCGTGCAACCCCT
CGGCGGTGTCTTTGCCAACAGCCAACGGCATAACATCTCAAACACCGATGTCTACGAGTA
CCTGAAGGACCTAGAGGAAGGCATCCAAATTCTGACGTGGAGGCTGGAAGATGGCAGCC
CCCAGACTGGGGAGATCTTCAGGCAGACCTACAGGAAGTTTGACAGAAACTCACACAACG
ATGACGCACTGCTCAAGAACTACTGGCTGCTCTTCTGCTTCCGGAAGGACATGAACAAGG
TCGAGACATTCCTGCACATTGTGCAGTGCCGC---TCTCTGCAGGACAGCTGTGGCTTCTAG
[654]

AgeofGHC ----------ATGGCTGCAGTCTCC---
CGGGCATCCCTGCTCCTGACTTTCGCCCTGCTCTGCCTGCCCTGGCTTCGGGAGGCTGGT
GCCTTCCCAGCAATTCCCTTATCCTCGCTTTACGACTATGCTATGATCCGCGCCTATCGCCT
GAACCAGCTGGCCTTTGACATCTACCAGAAGTTTGAA----------
GAAGCTCGTAGCCCGAAGGAACAGAAGGATTTCTTGCAGCATAAAGCCAGGACCTCCCTT
TGCTTCTCAGCGTCTGTCCCAACACCCACTAACAGAAAGGAAACTCTGCAGCAATCTAAC
GTAGAGCTGCTCCGCAACTCCCTGCTGCTCATCCAGTTGTGGTTCAAGCCCGTGCAGGTCT
TCAGCAGTGTCTTTGCCAACAGCCAACTGCATGGTGTCTCACACAGCTTCATCTATGAGTA
CCTGAAGGACCTAGAGGAAGTCATCCAAACTCTGATCAGGAAGTTGGAAGATGGCAGCCC
CCGGACTGGGGACATCTTCGGGCAGACCTACAGCAAGTTTGACACAAACTTGCACAAGGA
TGACACACTGCTCAAG----------
---------- [552]

AfuscGHB ----------ATGGCTGCAAGCTCC---
CGGATGTCCCTGCCCCTGACTCTTACCCTGCTCTGCCTGCCCTGGCTTCAAGAGACTGGTG
CCTTTCCAAGAATTCCCTTATCCAGGCTTTTTGGTGACGCTATGCTCCGTGCCCATCAGCT
GCACCAGGTGGCCTTTGACACCTACCAGGAGTTGGAA----------
GAAAATTGTATCCCGAAGAAACAGAAGTAYTTCTTCCTGCGTAACCCCAAGAACTTCCTCT
GCTTCTCAGAGTCTATCCCRACRCCCTTCAACAAGGAGGAAGTGCTGGCAAAATCCAGCC
TAGAGCTGCTCCACATCTCCCTGCTGCTCATCCAGTCCTGGCTCGAACCCGTGCAACCCCT
CAGCGGTGTCTTTGCCAACAGCCAACGGCATAACATCTCAAACACCGATGTCTACGAGTA
CCTGAAGGACCTAGAGGAAGGCATCCAAATTCTGACGAGGAGGCTGGAAGATGGCAGCC
CCCAGACTGGGGAGATCTTCAGGCAGACCTACAGGAAGTTTGACAGAAACTCGCACAAC
GATGACGCACTGCTCAAGAACTACTGGCTGCTCTTCTGCTTCCGGAAGGACATGAACAAG
GTCGAGACATTCCTGCACATTGTGCAGTGCCGC---
TCTCTGCAGGACAGCTGTGGCTTCTAG [654]

AfuscGHC ----------ATGGCTGCAGTCTCC---
CGGGCATCCCTGCTCCTGACTTTCACCCTGCTCTGCCTGCCCTGGCTTCGGGAGGCTGGT
GCCTTCCCAGCAATTCCCTTATCCTCGCTTTACGACTATGCTATGATCCGCGCCTATCGCCT
GAACCAGCTGGCCTTTGACATCTACCAGAAGTTTGAA----------
GAAGCTCGTAGCCCGAAGGAACAGAAGGATTTCTTGCAGCATAAAGCCAGGACCTCCCTT
TGCTTCTCAGCGTCTGTCCCAACACCCACTAACAGAAAGGAAACTCTGCAGCAATCTAAC
GTAGAGCTGCTCCGCAACTCCCTGCTGCTCATCCAGTTGTGGTTCAAGCCCGTGCAGGTCT
TCAGCAGTGTCTTTGCCAACAGCCAACTGCATGGTGTCTCACACAGCTTCATCTATGAGTA
CCTGAAGGACCTAGAGGAAGTCATCCAAACTCTGATCGGGAGGTTGGAAGATGGCAGCC
CCCGGACTGGGGACATCTTCGGGCAGACCTACAGCAAGTTTGACACAAACTTGCACAAGG
ATGACACACTGCTCAAGAACTACGCGCTGCTGTATTGCTTCCGGAGAGACATGGACAAGG
TCGCGACATTCCTGCGCATTGTGAAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCTTCTAG
[654]

AfuscGHD
ATGGGAAGAAGCGCTTGTGGTCCGTGTCTGTTCCCGGGATTTCTGCTTTTT---------
GGCTCC---
CAGATGTCCCTGCTCCTGGCTTTTACCCTGCTCTGCCTGCCCTGGCTTCAAGAGACTGGTG
CCCTTCCAGGCGTTCCCTTACCCAAGCTTTTAGAAGACACTGTGTTCCGTGCCCATCAGCT
GCACCAGGTGGCCTTTGACACCTACCAGGAGTTGGAA----------
GAAGATTGTATCCCGAATGAACAGAAGTACTTCTTCCTGCAGAACTCCAAGCCCTCCCTCT
GCTTCTCAGGGTCTATGCCGATACCCTCCAACAAGGAGGAAACACTGGCAAAATCCAACC
CAGAGCTGCTCCACATCTCCCTGCTGCTCTCCCAGTCCTGGCTCGAACCCGTGCAGCTCCT
CAGCAGTGTCTTTGCCAACAGTGAACTGCATAGCGTCTTGAACACCAACGTCTATGAGTTA
CTGAAGGACCTAGAAGAACGCATTCAAACTCTGATAGGGAGGCTGGAAGAGGTCAGCCC
CCAGACTGGGGAGATCATCAAGCAGAATTGCAGCATGTTTGACACAAACTCGCACAACGA
TGACACACTGCTCAAGAGCTACAGGCTGCTCTATTACTTCAGGAACGACATGCACAAGGT
TGAGACATTCCTGAACATTGTGAAGTGCCGC---TCTGTGGAGGGCAGCTGTGGCCTCTAG
[696]

GsenegGH ----------
ATGGCCACAGGCTCTCACACCACCACCCTGCTCCTGGCTGTGGCCCTGCTCGGCCTGCCC
TGGCCTCAGGAGGCTGGTGCCTTTCCGGCCATGCCTTTGTCCAGCCTGTTCGCCAACGCT
GTGCTCCGTGCCCAGCACCTGCACCAACTGGCTGCTGACACTTACAAGGAGTTTGAG-------
----------CGTGCCTACATCCCGGAGGGACAGCGATATTCCATC---
CAGAACACCCAGGCTGCCTTCTGCTTCTCAGAGACCATCCCGGCCCCCACGGGCAAGGA
CGAAGCCCAGCAGAGATCTGACATGGAGCTGCTCCGCTTCTCGCTGCTGCTCATCCAGTC
GTGGCTCGGGCCTGTGCAGCTCCTCAGCAGGGTCTTCACCAACAGCCTGGTGCTCGGAAC
CTTGGACCGA---
GTCTATGAGAAACTGAAGGACCTGGAGGAAGGCATCCAAGCCCTGATGCGGGAGCTGGA
AGATGGCAGCCCCCGGGTGGGGCAGATCCTCAAGCAAACCTACGACAAGTTTGACACCA
ACCTGCGCAGCGACGACGCACTGCTCAAGAACTACGGGCTGCTCTCCTGCTTCAAGAAGG

ACCTGCACAAGGCTGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTTGTGGAAAGCA
GCTGTGCCTTCTAG [654]

NpygmGH ----------
ATGGCCACAGGCTCTCACACCGCCACCCTGCTCCTGGCTGTAGCCCTGCTCGGCCTGCCC
TGGCCTCAGGAGGCTGGTGCCTTTCCGGCCATGCCTTTGTCCAGCCTGTTTGCCAACGCTG
TGCTCCGAGCCCAGCACCTGCACCAGCTGGCTGCTGATACTTACAAGGAGTTTGAG---------
---------CGTGCCTACATCCCGGAGGGACAGCGATATTCCATC---
CAGAACGCCCAGGCTGCCTTCTGCTTCTCAGAGACCATCCCGGCCCCCACGGGCAAGGA
CGAGGCCCAGCAGAGATCCGACATGGAGCTGCTCCGCTTCTCGCTGCTGCTCATCCAGTC
GTGGCTCGGGCCTGTGCAGCTGCTCAGCAGGGTCTTCACCAACAGCCTGGTACTCGGAAC
CTCGGACCGA---
GTCTATGAGAAACTGAAGGACCTGGAGGAAGGCATCCAAGCCCTGATGCGGGAGCTGGA
AGATGGCAGCCCCCGGGTGGGACAGATCCTCAAGCAAACCTACGACAAGTTTGACACCA
ACCTGCGCAGTGACGACGCACTGCTTAAGAACTACGGGCTGCTCTCTTGCTTCAAGAAGG
ACCTGCACAAGGCTGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTTGTGGAAAGCA
GCTGTGCCTTCTAG [654]

MmusGH ----------ATGGCTACAGACTCT---
CGGACCTCCTGGCTCCTGACCGTCAGCCTGCTCTGCCTGCTCTGGCCTCAGGAGGCTAGT
GCTTTTCCCGCCATGCCCTTGTCCAGTCTGTTTTCTAATGCTGTGCTCCGAGCCCAGCACC
TGCACCAGCTGGCTGCTGACACCTACAAAGAGTTCGAG----------
CGTGCCTACATTCCCGAGGGACAGCGCTATTCCATT---
CAGAATGCCCAGGCTGCTTTCTGCTTCTCAGAGACCATCCCGGCCCCCACAGGCAAGGAG
GAGGCCCAGCAGAGAACCGACATGGAATTGCTTCGCTTCTCGCTGCTGCTCATCCAGTCA
TGGCTGGGGCCCGTGCAGTTCCTCAGCAGGATTTTCACCAACAGCCTGATGTTCGGCACC
TCGGACCGT---
GTCTATGAGAAACTGAAGGACCTGGAAGAGGGCATCCAGGCTCTGATGCAGGAGCTGGA
AGATGGCAGCCCCCGTGTTGGGCAGATCCTCAAGCAAACCTATGACAAGTTTGACGCCAA
CATGCGCAGCGACGACGCGCTGCTCAAAAACTATGGGCTGCTCTCCTGCTTCAAGAAGGA
CCTGCACAAAGCGGAGACCTACCTGCGGGTCATGAAGTGTCGCCGCTTTGTGGAAAGCAG
CTGTGCCTTCTAG [651]

RnorGH ----------ATGGCTGCAGACTCT---
CAGACTCCCTGGCTCCTGACCTTCAGCCTGCTCTGCCTGCTGTGGCCTCAAGAGGCTGGT
GCTTTACCTGCCATGCCCTTGTCCAGTCTGTTTGCCAATGCTGTGCTCCGAGCCCAGCACC
TGCACCAGCTGGCTGCTGACACCTACAAAGAGTTCGAG----------
CGTGCCTACATTCCCGAGGGACAGCGCTATTCCATT---
CAGAATGCCCAGGCTGCGTTCTGCTTCTCAGAGACCATCCCAGCCCCCACCGGCAAGGAG
GAGGCCCAGCAGAGAACTGACATGGAATTGCTTCGCTTCTCGCTGCTGCTCATCCAGTCA
TGGCTGGGGCCCGTGCAGTTTCTCAGCAGGATCTTTACCAACAGCCTGATGTTTGGTACCT
CGGACCGC---
GTCTATGAGAAACTGAAGGACCTGGAAGAGGGCATCCAGGCTCTGATGCAGGAGCTGGA
AGACGGCAGCCCCCGTATTGGGCAGATCCTCAAGCAAACCTATGACAAGTTTGACGCCAA
CATGCGCAGCGATGACGCTCTGCTCAAAAACTATGGGCTGCTCTCCTGCTTCAAGAAGGA
CCTGCACAAGGCAGAGACCTACCTGCGGGTCATGAAGTGTCGCCGCTTTGCGGAAAGCA
GCTGTGCTTTCTAG [651]

ClupGH ----------ATGGCTGCAAGCCCT---
CGGAACTCTGTGCTCCTGGCCTTCGCCTTGCTCTGCCTGCCCTGGCCTCAGGAGGTGGGC
GCCTTCCCGGCCATGCCCTTGTCCAGCCTGTTTGCCAACGCCGTGCTCCGGGCCCAGCAC
CTGCACCAACTGGCTGCCGACACCTACAAAGAGTTTGAG----------
CGGGCGTACATCCCCGAGGGACAGAGGTACTCCATC---
CAGAACGCGCAGGCCGCCTTCTGCTTCTCGGAGACCATCCCGGCCCCCACGGGCAAGGA
CGAGGCCCAGCAGCGATCCGACGTGGAGCTGCTCCGCTTCTCCCTGCTGCTCATCCAGTC
GTGGCTCGGGCCCGTGCAGTTTCTCAGCAGGGTCTTCACCAACAGCCTGGTGTTCGGCAC

CTCAGACCGA---
GTCTACGAGAAGCTCAAGGACCTGGAGGAAGGCATCCAAGCCCTGATGCGGGAGCTGGA
AGATGGCAGTCCCCGGGCCGGGCAGATCCTGAAGCAGACCTACGACAAGTTTGACACGA
ACCTGCGCAGTGACGATGCGCTGCTTAAGAACTACGGGCTGCTCTCCTGCTTCAAGAAAG
ACCTGCATAAGGCCGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTCGTGGAAAGCA
GCTGTGCCTTCTAG [651]

BtauGH ----------ATGGCTGCAGGCCCC---
CGGACCTCCCTGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCCTGGACTCAGGTGGTGGGC
GCCTTCCCAGCCATGTCCTTGTCCGGCCTGTTTGCCAACGCTGTGCTCCGGGCTCAGCAC
CTGCATCAGCTGGCTGCTGACACCTTCAAAGAGTTTGAG----------
CGCACCTACATCCCGGAGGGACAGAGATACTCCATC---
CAGAACACCCAGGTTGCCTTCTGCTTCTCTGAAACCATCCCGGCCCCCACGGGCAAGAAT
GAGGCCCAGCAGAAATCAGACTTGGAGCTGCTTCGCATCTCACTGCTCCTCATCCAGTCG
TGGCTTGGGCCCCTGCAGTTCCTCAGCAGAGTCTTCACCAACAGCTTGGTGTTTGGCACCT
CGGACCGT---
GTCTATGAGAAGCTGAAGGACCTGGAGGAAGGCATCCTGGCCCTGATGCGGGAGCTGGA
AGATGGCACCCCCCGGGCTGGGCAGATCCTCAAGCAGACCTATGACAAATTTGACACAAA
CATGCGCAGTGACGACGCGCTGCTCAAGAACTACGGTCTGCTCTCCTGCTTCCGGAAGGA
CCTGCATAAGACGGAGACGTACCTGAGGGTCATGAAGTGCCGCCGCTTCGGGGAGGCCA
GCTGTGCCTTCTAg [651]

MonodelphisGH ----------ATGGCTCCAGGTATG---
CGAGTCTGTCTTTTGCTCCTCATCGCC---
TTCACCTTGCTGGGGCCACAGAGGGCTGCTGCCTTCCCAGCCATGCCTCTGTCCAGCCTC
TTTGCCAACGCTGTACTCCGTGCCCAACATCTGCACCAGCTGGTTGCTGACACCTACAAG
GAGTTTGAA----------CGAACCTACATTCCAGAGGCTCAGAGACATTCCATC---
CAGAGTACCCAGACAGCTTTCTGTTTCTCTGAGACCATCCCAGCTCCCACTGGCAAGGAT
GAGGCCCAGCAGAGATCTGATGTGGAGTTGCTTCGTTTTTCCCTTCTGCTTATCCAGTCTT
GGCTCAGCCCTGTACAGTTCCTCAGCAGAGTCTTCACCAACAGCCTGGTCTTTGGTACCTC
AGACCGT---
GTCTATGAGAAGCTGAGGGATCTGGAAGAGGGGATCCAGGCTCTCATGCAGGAGCTGGA
AGATGGCAGTTCAAGAGGTGGTCTGGTCCTCAAGACAACCTATGACAAATTTGATACCAA
CCTACGCAGTGATGAGGCACTGCTCAAGAATTATGGACTGCTCTCCTGCTTCAAGAAGGA
CCTGCACAAAGCTGAGACCTACCTCCGGGTCATGAAGTGCCGCCGTTTTGTGGAGAGCAG
TTGTGCCTTCTGA [648]

;

END;

REFERENCES

Adams, MJ, Buehner, M, Chandrasekhar, K, Ford, GC, Hackert, ML, Liljas, A, Rossmann, MG, Smiley, IE, Allison, WS, Everse, J et al.(1973) Structure-function relationships in lactate dehydrogenase. *Proc Natl Acad Sci U S A* **70:** 1968-1972.

Aiello, LC , Wheeler, P(1995) The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution. *current anthropology* **36:** 199.

Alcivar, AA, Trasler, JM, Hake, LE, Salehi-Ashtiani, K, Goldberg, E , Hecht, NB(1991) DNA methylation and expression of the genes coding for lactate dehydrogenases A and C during rodent spermatogenesis. *Biol Reprod* **44:** 527-535.

Allman, J. 2000. *Evolving Brains*. Scientific American Library - W.H.Freeman, New York.

Allman, JM. 1999. *Evolving Brains*. Scientific American Library.

Armstrong, E(1983) Relative brain size and metabolism in mammals. *Science* **220:** 1302-1304.

Bailey, TL , Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36. AAAI Press.

Baumberger, JP, Jurgensen, JJ , Bardwell, K(1933) The Coupled Redox Potential of the Lactate-Enzyme-Pyruvate System. *J Gen Physiol* **16:** 961-976.

Beebee, TJ , Carty, DS(1982) Purification and radioimmunoassay of rat lactate dehydrogenase A and B subunits. *Biochem J* **205:** 313-320.

Berezikov, E, Guryev, V , Cuppen, E(2007) Exploring conservation of transcription factor binding sites with CONREAL. *Methods Mol Biol* **395:** 437-448.

Bergersen, L, Rafiki, A , Ottersen, OP(2002) Immunogold cytochemistry identifies specialized membrane domains for monocarboxylate transport in the central nervous system. *Neurochem Res* **27:** 89-96.

Bernardi, G(1993) The vertebrate genome: isochores and evolution. *Mol Biol Evol* **10:** 186-204.

Birney, E, Andrews, TD, Bevan, P, Caccamo, M, Chen, Y, Clarke, L, Coates, G, Cuff, J, Curwen, V, Cutts, T et al.(2004) An overview of Ensembl. *Genome Res* **14:** 925-928.

Bishop, MJ, Everse, J , Kaplan, NO(1972) Identification of lactate dehydrogenase isoenzymes by rapid kinetics. *Proc Natl Acad Sci U S A* **69:** 1761-1765.

Bittar, PG, Charnay, Y, Pellerin, L, Bouras, C , Magistretti, PJ(1996) Selective distribution of lactate dehydrogenase isoenzymes in neurons and astrocytes of human brain. *J Cereb Blood Flow Metab* **16:** 1079-1089.

Bossone, SA, Asselin, C, Patel, AJ , Marcu, KB(1992) MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A* **89:** 7452-7456.

Brinkman-Van der Linden, EC, Hurtado-Ziola, N, Hayakawa, T, Wiggleton, L, Benirschke, K, Varki, A , Varki, N(2007) Human-specific expression of Siglec-6 in the placenta. *Glycobiology* **17:** 922-931.

Browning, W(1886) IV. Lactic Acid as a Destroyer of Pathogenic Tissue. *Ann Surg* **3:** 231-232.

Cahn, RD, Zwilling, E, Kaplan, NO , Levine, L(1962) Nature and Development of Lactic Dehydrogenases: The two major types of this enzyme form molecular hybrids which change in makeup during development. *Science* **136:** 962-969.

Carroll, SB(2005) Evolution at two levels: on genes and form. *PLoS Biol* **3:** e245.

Charpentier, J , Goutefongea, R(1964) Electrophoretic Characteristics of Lactic Dehydrogenase in Normal and Exudative Pig Muscle. *Nature* **201:** 1325-1326.

Chellakooty, M, Vangsgaard, K, Larsen, T, Scheike, T, Falck-Larsen, J, Legarth, J, Andersson, AM, Main, KM, Skakkebaek, NE , Juul, A(2004) A longitudinal study of intrauterine growth and the placental growth hormone (GH)-insulin-like growth factor I axis in maternal circulation: association between placental GH and fetal growth. *J Clin Endocrinol Metab* **89:** 384-391.

Chen, C, Opazo, JC, Erez, O, Uddin, M, Santolaya-Forgas, J, Goodman, M, Grossman, LI, Romero, R , Wildman, DE(2008) The human progesterone receptor shows evidence of adaptive evolution associated with its ability to act as a transcription factor. *Mol Phylogenet Evol* **47:** 637-649.

Chen, EY, Liao, YC, Smith, DH, Barrera-Saldana, HA, Gelinas, RE , Seeburg, PH(1989) The human growth hormone locus: nucleotide sequence, biology, and evolution. *Genomics* **4:** 479-497.

Cheng, X , Blumenthal, RM(2010) Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. *Biochemistry* **49:** 2999-3008.

Chung, KC, Huang, D, Chen, Y, Short, S, Short, ML, Zhang, Z , Jungmann, RA(1995) Identification of a silencer module which selectively represses cyclic AMP-responsive element-dependent gene expression. *Mol Cell Biol* **15:** 6139-6149.

Churchill, ME, Jones, DN, Glaser, T, Hefner, H, Searles, MA , Travers, AA(1995) HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J* **14:** 1264-1275.

Clay, O, Schaffner, W , Matsuo, K(1995) Periodicity of eight nucleotides in purine distribution around human genomic CpG dinucleotides. *Somat Cell Mol Genet* **21:** 91-98.

Cleary, PP, Campbell, A , Chang, R(1972) Location of promoter and operator sites in the biotin gene cluster of Escherichia coli. *Proc Natl Acad Sci U S A* **69:** 2219-2223.

Cooke, NE, Ray, J, Emery, JG , Liebhaber, SA(1988) Two distinct species of human growth hormone-variant mRNA in the human placenta predict the expression of novel growth hormone proteins. *J Biol Chem* **263:** 9001-9006.

Costantini, M, Cammarano, R , Bernardi, G(2009) The evolution of isochore patterns in vertebrate genomes. *BMC Genomics* **10:** 146.

Coulondre, C, Miller, JH, Farabaugh, PJ , Gilbert, W(1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274:** 775-780.

Crespi, B , Semeniuk, C(2004) Parent-offspring conflict in the evolution of vertebrate reproductive mode. *Am Nat* **163:** 635-653.

Crick, FH, Griffith, JS , Orgel, LE(1957) Codes without Commas. *Proc Natl Acad Sci U S A* **43:** 416-421.

Cunningham, BC , Wells, JA(1991) Rational design of receptor-specific variants of human growth hormone. *Proc Natl Acad Sci U S A* **88:** 3407-3411.

Dammhahn, M , Kappeler, PM(2010) Scramble or contest competition over food in solitarily foraging mouse lemurs (Microcebus spp.): New insights from stable isotopes. *Am J Phys Anthropol* **141:** 181-189.

Dattani, M , Preece, M(2004) Growth hormone deficiency and related disorders: insights into causation, diagnosis, and treatment. *Lancet* **363:** 1977-1987.

Dawson, DM, Goodfriend, TL , Kaplan, NO(1964) Lactic Dehydrogenases: Functions of the Two Types Rates of Synthesis of the Two Major Forms Can Be Correlated with Metabolic Differentiation. *Science* **143:** 929-933.

De Bustos, C, Ramos, E, Young, JM, Tran, RK, Menzel, U, Langford, CF, Eichler, EE, Hsu, L, Henikoff, S, Dumanski, JP et al.(2009) Tissue-specific variation in DNA methylation levels along human chromosome 1. *Epigenetics Chromatin* **2:** 7.

Dennis, G, Jr., Sherman, BT, Hosack, DA, Yang, J, Gao, W, Lane, HC , Lempicki, RA(2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** P3.

Dhar, M, Mascareno, EM , Siddiqui, MA(1997) Two distinct factor-binding DNA elements in cardiac myosin light chain 2 gene are essential for repression of

its expression in skeletal muscle. Isolation of a cDNA clone for repressor protein Nished. *J Biol Chem* **272:** 18490-18497.

Diaz-Gallardo, MY, Cote-Velez, A, Charli, JL , Joseph-Bravo, P(2010) A rapid interference between glucocorticoids and cAMP-activated signalling in hypothalamic neurones prevents binding of phosphorylated cAMP response element binding protein and glucocorticoid receptor at the CRE-Like and composite GRE sites of thyrotrophin-releasing hormone gene promoter. *J Neuroendocrinol* **22:** 282-293.

Dierks, P, van Ooyen, A, Cochran, MD, Dobkin, C, Reiser, J , Weissmann, C(1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* **32:** 695-706.

Drummond, AJ, Ashton, B, Cheung, M, Heled, J, Kearse, M, Moir, R, Stones-Havas, S, Thierer, T , Wilson, A. 2009. Geneious v4.7.

Echigoya, Y, Sato, T, Itou, T, Endo, H , Sakai, T(2009) Molecular characterization and expression pattern of the equine lactate dehydrogenase A and B genes. *Gene* **447:** 40-50.

Egli, RJ, Southam, L, Wilkins, JM, Lorenzen, I, Pombo-Suarez, M, Gonzalez, A, Carr, A, Chapman, K , Loughlin, J(2009) Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum* **60:** 2055-2064.

Elliot, MG , Crespi, BJ(2008) Placental invasiveness and brain-body allometry in eutherian mammals. *J Evol Biol* **21:** 1763-1778.

Elliott, KA(1948) Metabolism of brain tissue slices and suspensions from various mammals. *J Neurophysiol* **11:** 473-484.

Eron, L , Block, R(1971) Mechanism of initiation and repression of in vitro transcription of the lac operon of Escherichia coli. *Proc Natl Acad Sci U S A* **68:** 1828-1832.

Farcas, R, Schneider, E, Frauenknecht, K, Kondova, I, Bontrop, R, Bohl, J, Navarro, B, Metzler, M, Zischler, H, Zechner, U et al.(2009) Differences in DNA methylation patterns and expression of the CCRK gene in human and nonhuman primate cortices. *Mol Biol Evol* **26:** 1379-1389.

Favier, B , Dolle, P(1997) Developmental functions of mammalian Hox genes. *Mol Hum Reprod* **3:** 115-131.

Finlay, BL , Darlington, RB(1995) Linked regularities in the development and evolution of mammalian brains. *Science* **268:** 1578-1584.

Fleenor, D, Oden, J, Kelly, PA, Mohan, S, Alliouachene, S, Pende, M, Wentz, S, Kerr, J , Freemark, M(2005) Roles of the lactogens and somatogens in perinatal and postnatal metabolism and growth: studies of a novel mouse model combining lactogen resistance and growth hormone deficiency. *Endocrinology* **146:** 103-112.

Georgiev, S, Boyle, AP, Jayasurya, K, Ding, X, Mukherjee, S , Ohler, U(2010) Evidence-ranked motif identification. *Genome Biol* **11:** R19.

Gidoni, D, Dynan, WS , Tjian, R(1984) Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* **312:** 409-413.

Giroud, S, Perret, M, Stein, P, Goudable, J, Aujard, F, Gilbert, C, Robin, JP, Le Maho, Y, Zahariev, A, Blanc, S et al.(2010) The grey mouse lemur uses season-dependent fat or protein sparing strategies to face chronic food restriction. *PLoS ONE* **5:** e8823.

Gluckman, PD , Pinal, CS(2002) Maternal-placental-fetal interactions in the endocrine regulation of fetal growth: role of somatotrophic axes. *Endocrine* **19:** 81-89.

Goffin, V, Shiverick, KT, Kelly, PA , Martial, JA(1996) Sequence-function relationships within the expanding family of prolactin, growth hormone, placental lactogen, and related proteins in mammals. *Endocr Rev* **17:** 385-410.

Golos, TG, Durning, M, Fisher, JM , Fowler, PD(1993) Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. *Endocrinology* **133:** 1744-1752.

Gonzalez Alvarez, R, Revol de Mendoza, A, Esquivel Escobedo, D, Corrales Felix, G, Rodriguez Sanchez, I, Gonzalez, V, Davila, G, Cao, Q, de Jong, P, Fu, YX et al.(2006) Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. *Gene* **380:** 38-45.

Goodman, M, Czelusniak, J, Moore, GW, Romero-Herrera, AE , Matsuda, G(1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. *Systematic Zoology* **28:** 132-163.

Goodman, M, Porter, CA, Czelusniak, J, Page, SL, Schneider, H, Shoshani, J, Gunnell, G , Groves, CP(1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* **9:** 585-598.

Goodman, M, Syner, FN, Stimson, CW , Rankin, JJ(1969) Phylogenetic changes in the proportions of two kinds of lactate dehydrogenase in primate brain regions. *Brain Res* **14:** 447-459.

Goodyer, CG, Figueiredo, RM, Krackovitch, S, De Souza Li, L, Manalo, JA , Zogopoulos, G(2001) Characterization of the growth hormone receptor in human dermal fibroblasts and liver during development. *Am J Physiol Endocrinol Metab* **281:** E1213-1220.

Gorelikov, PL , Savel'ev, SV(2008) Lactate dehydrogenase isoenzymes in sympathetic neurons and satellite gliocytes in normal conditions and in blockade of nicotinic cholinoreceptors. *Neurosci Behav Physiol* **38:** 817-820.

Greiner, EF, Guppy, M , Brand, K(1994) Glucose is essential for proliferation and the glycolytic enzyme induction that provokes a transition to glycolytic energy production. *J Biol Chem* **269:** 31484-31490.

Haig, D(1993) Genetic conflicts in human pregnancy. *Q Rev Biol* **68:** 495-532.

Haig, D(1996) Placental hormones, genomic imprinting, and maternal-fetal communication. *Journal of Evolutionary Biology* **9:** 357-380.

Haig, D(2008) Placental growth hormone-related proteins and prolactin-related proteins. *Placenta* **29 Suppl A:** S36-41.

Hampson, RK , Rottman, FM(1987) Alternative processing of bovine growth hormone mRNA: nonsplicing of the final intron predicts a high molecular weight variant of bovine growth hormone. *Proc Natl Acad Sci U S A* **84:** 2673-2677.

Harbison, CT, Gordon, DB, Lee, TI, Rinaldi, NJ, Macisaac, KD, Danford, TW, Hannett, NM, Tagne, JB, Reynolds, DB, Yoo, J et al.(2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99-104.

Hertz, L , Dienel, GA(2005) Lactate transport and transporters: general principles and functional roles in brain cells. *J Neurosci Res* **79:** 11-18.

Himeda, CL, Ranish, JA , Hauschka, SD(2008) Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Mol Cell Biol* **28:** 6521-6535.

Hinks, M , Masters, CJ(1964) Developmental Changes in Ruminant Lactate Dehydrogenase. *Biochemistry* **3:** 1789-1791.

Hinks, M , Masters, CJ(1965) The Multiple Forms of Ovine Lactate Dehydrogenase. *Aust J Biol Sci* **18:** 154-159.

Holmes, RS , Goldberg, E(2009) Computational analyses of mammalian lactate dehydrogenases: human, mouse, opossum and platypus LDHs. *Comput Biol Chem* **33:** 379-385.

Hsieh, F, Chen, SC , Pollard, K(2009) A Nearly Exhaustive Search for CpG Islands on Whole Chromosomes. *Int J Biostat* **5:** 1.

Huang da, W, Sherman, BT , Lempicki, RA(2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57.

Huber, K , Ernsberger, U(2006) Cholinergic differentiation occurs early in mouse sympathetic neurons and requires Phox2b. *Gene Expr* **13:** 133-139.

Huelsenbeck, JP , Ronquist, F(2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754-755.

Hyder, F, Patel, AB, Gjedde, A, Rothman, DL, Behar, KL , Shulman, RG(2006) Neuronal-glial glucose oxidation and glutamatergic-GABAergic function. *J Cereb Blood Flow Metab* **26:** 865-877.

Ippen, K, Miller, JH, Scaife, J , Beckwith, J(1968) New controlling element in the Lac operon of E. coli. *Nature* **217:** 825-827.

Jacob, F , Monod, J(1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3:** 318-356.

Jerison, HJ. 1973. *Evolution of the Brain and Intelligence*. Academic Press, New York.

Jerison, HJ(1977) The theory of encephalization. *Ann N Y Acad Sci* **299:** 146-160.

Joffe, TH , Dunbar, RIM(1998) Tarsier brain component composition and its implications for systematics. *Primates* **39:** 211-216.

Jones, RL, Critchley, HO, Brooks, J, Jabbour, HN , McNeilly, AS(1998) Localization and temporal expression of prolactin receptor in human endometrium. *J Clin Endocrinol Metab* **83:** 258-262.

Jungmann, RA, Huang, D , Tian, D(1998) Regulation of LDH-A gene expression by transcriptional and posttranscriptional signal transduction mechanisms. *J Exp Zool* **282:** 188-195.

Jungmann, RA , Kiryukhina, O(2005) Cyclic AMP and AKAP-mediated targeting of protein kinase A regulates lactate dehydrogenase subunit A mRNA stability. *J Biol Chem* **280:** 25170-25177.

Kato, M(1999) Structural bistability of repetitive DNA elements featuring CA/TG dinucleotide steps and mode of evolution of satellite DNA. *Eur J Biochem* **265:** 204-209.

Kaufman, JA, Ahrens, ET, Laidlaw, DH, Zhang, S , Allman, JM(2005) Anatomical analysis of an aye-aye brain (Daubentonia madagascariensis, primates: Prosimii) combining histology, structural magnetic resonance imaging, and diffusion-tensor imaging. *Anat Rec A Discov Mol Cell Evol Biol* **287:** 1026-1037.

Kechris, K , Li, H(2008) c-REDUCE: incorporating sequence conservation to detect motifs that correlate with expression. *BMC Bioinformatics* **9:** 506.

Kent, WJ, Sugnet, CW, Furey, TS, Roskin, KM, Pringle, TH, Zahler, AM , Haussler, D(2002) The human genome browser at UCSC. *Genome Res* **12:** 996-1006.

King, MC , Wilson, AC(1975) Evolution at two levels in humans and chimpanzees. *Science* **188:** 107-116.

Klenk, HP, Clayton, RA, Tomb, JF, White, O, Nelson, KE, Ketchum, KA, Dodson, RJ, Gwinn, M, Hickey, EK, Peterson, JD et al.(1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390:** 364-370.

Koen, AL , Goodman, M(1969) Lactate dehydrogenase isozymes: qualitative and quantitative changes during primate evolution. *Biochem Genet* **3:** 457-474.

Komiyama, H, Aoki, A, Tanaka, S, Maekawa, H, Kato, Y, Wada, R, Maekawa, T, Tamura, M , Shiroishi, T(2010) Alu-derived cis-element regulates tumorigenesis-dependent gastric expression of GASDERMIN B (GSDMB). *Genes Genet Syst* **85:** 75-83.

Lacroix, MC, Guibourdenche, J, Fournier, T, Laurendeau, I, Igout, A, Goffin, V, Pantel, J, Tsatsaris, V , Evain-Brion, D(2005) Stimulation of human trophoblast invasion by placental growth hormone. *Endocrinology* **146:** 2434-2444.

Laflamme, K, Owen, AN, Devlin, EE, Yang, MQ, Wong, C, Steiner, LA, Garrett, LJ, Elnitski, L, Gallagher, PG , Bodine, DM(2010) Functional analysis of a novel cis-acting regulatory region within the human ankyrin (ANK-1) gene promoter. *Mol Cell Biol*.

Latner, AL, Siddiqui, SA , Skillen, AW(1966) Pyruvate inhibition of lactate dehydrogenase activity in human tissue extracts. *Science* **154:** 527-529.

Latner, AL , Skillen, AW(1964) Lactate Dehydrogenase Isoenzymes in Foetal and Neonatal Tissues. *J Embryol Exp Morphol* **12:** 501-509.

Leiblich, A, Cross, SS, Catto, JW, Phillips, JT, Leung, HY, Hamdy, FC , Rehman, I(2006) Lactate dehydrogenase-B is silenced by promoter hypermethylation in human prostate cancer. *Oncogene* **25:** 2953-2960.

Li, YJ, Tsoi, SC, Mannen, H , Shoei-lung Li, S(2002) Phylogenetic analysis of vertebrate lactate dehydrogenase (LDH) multigene families. *J Mol Evol* **54:** 614-624.

Lifton, RP, Goldberg, ML, Karp, RW , Hogness, DS(1978) The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harb Symp Quant Biol* **42 Pt 2:** 1047-1051.

Liu, JC, Makova, KD, Adkins, RM, Gibson, S , Li, WH(2001) Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. *Mol Biol Evol* **18:** 945-953.

Liu, Y, Cotton, JA, Shen, B, Han, X, Rossiter, SJ , Zhang, S(2010) Convergent sequence evolution between echolocating bats and dolphins. *Curr Biol* **20:** R53-54.

Lowenthal, A, Van Sande, M , Karcher, D(1961) Heterogeneity of lactic and malic dehydrogenase in serum, cerebrospinal fluid, and brain extracts in man and sheep. *Ann N Y Acad Sci* **94:** 988-995.

MacLeod, JN, Lee, AK, Liebhaber, SA , Cooke, NE(1992) Developmental control and alternative splicing of the placentally expressed transcripts from the human growth hormone gene cluster. *J Biol Chem* **267:** 14219-14226.

Maddison, D , Maddison, W. 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Inc, Sunderland, MA.

Maddison, DaM, W. 2000. MacClade 4:Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA.

Madern, D(2002) Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J Mol Evol* **54:** 825-840.

Maeda, N, Kasukawa, T, Oyama, R, Gough, J, Frith, M, Engstrom, PG, Lenhard, B, Aturaliya, RN, Batalov, S, Beisel, KW et al.(2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet* **2:** e62.

Maekawa, M, Inomata, M, Sasaki, MS, Kaneko, A, Ushiama, M, Sugano, K, Takayama, J , Kanno, T(2002) Electrophoretic variant of a lactate dehydrogenase isoenzyme and selective promoter methylation of the LDHA gene in a human retinoblastoma cell line. *Clin Chem* **48:** 1938-1945.

Maekawa, M, Taniguchi, T, Ishikawa, J, Sugimura, H, Sugano, K , Kanno, T(2003) Promoter hypermethylation in cancer silences LDHB, eliminating lactate dehydrogenase isoenzymes 1-4. *Clin Chem* **49:** 1518-1520.

Mao, XR, Moerman-Herzog, AM, Chen, Y , Barger, SW(2009) Unique aspects of transcriptional regulation in neurons--nuances in NFkappaB and Sp1-related factors. *J Neuroinflammation* **6:** 16.

Margulies, EH , Birney, E(2008) Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes. *Nat Rev Genet* **9:** 303-313.

Marino, L, Uhen, MD, Pyenson, ND , Frohlich, B(2003) Reconstructing cetacean brain evolution using computed tomography. *Anat Rec B New Anat* **272:** 107-117.

Markert, CL(1963a) Lactate Dehydrogenase Isozymes: Dissociation and Recombination of Subunits. *Science* **140:** 1329-1330.

Markert, CL(1963b) Lactate Dehydrogenase Isozymes: Dissociation and Recombination of Subunits. *Science* **140:** 1329-1330.

Martin, RD. 1990a. *Primate Origins and Evolution*. Chapman and Hall, London, UK.

Martin, RD. 1990b. *Primate Origins and Evolution: A Phylogenetic Reconstruction*. Princeton University Press, London.

Maston, GA, Evans, SK , Green, MR(2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7:** 29-59.

Maston, GA , Ruvolo, M(2002) Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol* **19:** 320-335.

Matys, V, Kel-Margoulis, OV, Fricke, E, Liebich, I, Land, S, Barre-Dirrie, A, Reuter, I, Chekmenev, D, Krull, M, Hornischer, K et al.(2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34:** D108-110.

McIntyre, HD, Serek, R, Crane, DI, Veveris-Lowe, T, Parry, A, Johnson, S, Leung, KC, Ho, KK, Bougoussa, M, Hennen, G et al.(2000) Placental growth hormone (GH), GH-binding protein, and insulin-like growth factor axis in normal, growth-retarded, and diabetic pregnancies: correlations with fetal growth. *J Clin Endocrinol Metab* **85:** 1143-1150.

McKenna, MC, Tildon, JT, Stevenson, JH, Hopkins, IB, Huang, X , Couto, R(1998) Lactate transport by cortical synaptosomes from adult rat brain: characterization of kinetics and inhibitor specificity. *Dev Neurosci* **20:** 300-309.

Melvin, RG , Andrews, MT(2009) Torpor induction in mammals: recent discoveries fueling new ideas. *Trends Endocrinol Metab* **20:** 490-498.

Milne, EM , Doxey, DL(1987) Lactate dehydrogenase and its isoenzymes in the tissues and sera of clinically normal dogs. *Res Vet Sci* **43:** 222-224.

Mink, JW, Blumenschine, RJ , Adams, DB(1981) Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. *Am J Physiol* **241:** R203-212.

Minth, CD , Dixon, JE(1990) Expression of the human neuropeptide Y gene. *J Biol Chem* **265:** 12933-12939.

Mittal, P, Espinoza, J, Hassan, S, Kusanovic, JP, Edwin, SS, Nien, JK, Gotsch, F, Than, NG, Erez, O, Mazaki-Tovi, S et al.(2007) Placental growth hormone is increased in the maternal and fetal serum of patients with preeclampsia. *J Matern Fetal Neonatal Med* **20:** 651-659.

Mittermeier, RA, Ganzhorn, JU, Konstant, WR, Glander, K, Tattersall, I, Groves, CP, Rylands, AB, Hapke, A, Ratsimbazafy, J, Mayor, MI et al.(2008) Lemur Diversity in Madagascar. *International Journal of Primatology* **29:** 1607-1656.

Mummidi, S, Bamshad, M, Ahuja, SS, Gonzalez, E, Feuillet, PM, Begum, K, Galvis, MC, Kostecki, V, Valente, AJ, Murthy, KK et al.(2000) Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. *J Biol Chem* **275:** 18946-18961.

Nirenberg, MW , Matthaei, JH(1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* **47:** 1588-1602.

Nishimura, S, Jones, DS , Khorana, HG(1965) Studies on polynucleotides. 48. The in vitro synthesis of a co-polypeptide containing two amino acids in alternating sequence dependent upon a DNA-like polymer containing two nucleotides in alternating sequence. *J Mol Biol* **13:** 302-324.

Nisselbaum, JS , Bodansky, O(1959) Reactions of lactic dehydrogenase from various rabbit organs with antirabbit muscle lactic dehydrogenase. *J Biol Chem* **234:** 3276-3280.

Nisselbaum, JS , Bodansky, O(1961) Reactions of human tissue lactic dehydrogenases with antisera to human heart and liver lactic dehydrogenases. *J Biol Chem* **236:** 401-404.

Nisselbaum, JS, Packer, DE , Bodansky, O(1964) Comparison of the Actions of Human Brain, Liver, and Heart Lactic Dehydrogenase Variants on Nucleotide Analogues and on Substrate Analogues in the Absence and in the Presence of Oxalate and Oxamate. *J Biol Chem* **239:** 2830-2834.

Nylander, JAA. 2004. MrModeltest. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

O'Brien, J, Kla, KM, Hopkins, IB, Malecki, EA , McKenna, MC(2007) Kinetic parameters and lactate dehydrogenase isozyme activities support possible lactate utilization by neurons. *Neurochem Res* **32:** 597-607.

Ohno, S(1988) Universal rule for coding sequence construction: TA/CG deficiency-TG/CT excess. *Proc Natl Acad Sci U S A* **85:** 9630-9634.

Okamoto, OK, Janjoppi, L, Bonone, FM, Pansani, AP, da Silva, AV, Scorza, FA , Cavalheiro, EA(2010) Whole transcriptome analysis of the hippocampus: toward a molecular portrait of epileptogenesis. *BMC Genomics* **11:** 230.

Ormandy, CJ, Camus, A, Barra, J, Damotte, D, Lucas, B, Buteau, H, Edery, M, Brousse, N, Babinet, C, Binart, N et al.(1997) Null mutation of the prolactin receptor

gene produces multiple reproductive defects in the mouse. *Genes Dev* **11:** 167-178.

Ortiz-Barahona, A, Villar, D, Pescador, N, Amigo, J , del Peso, L(2010) Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction. *Nucleic Acids Res* **38:** 2332-2345.

Pabo, CO , Sauer, RT(1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61:** 1053-1095.

Papper, Z, Jameson, NM, Romero, R, Weckle, AL, Mittal, P, Benirschke, K, Santolaya-Forgas, J, Uddin, M, Haig, D, Goodman, M et al.(2009) Ancient origin of placental expression in the growth hormone genes of anthropoid primates. *Proc Natl Acad Sci U S A* **106:** 17083-17088.

Parker, SC, Hansen, L, Abaan, HO, Tullius, TD , Margulies, EH(2009) Local DNA topography correlates with functional noncoding regions of the human genome. *Science* **324:** 389-392.

Pellerin, L, Bouzier-Sore, AK, Aubert, A, Serres, S, Merle, M, Costalat, R , Magistretti, PJ(2007) Activity-dependent regulation of energy metabolism by astrocytes: an update. *Glia* **55:** 1251-1262.

Pellerin, L , Magistretti, PJ(1994) Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization. *Proc Natl Acad Sci U S A* **91:** 10625-10629.

Perry, GH, Martin, RD , Verrelli, BC(2007) Signatures of functional constraint at aye-aye opsin genes: the potential of adaptive color vision in a nocturnal primate. *Mol Biol Evol* **24:** 1963-1970.

Peterson, FC , Brooks, CL(1997) Identification of a motif associated with the lactogenic actions of human growth hormone. *J Biol Chem* **272:** 21444-21448.

Peterson, FC , Brooks, CL(2004) Different elements of mini-helix 1 are required for human growth hormone or prolactin action via the prolactin receptor. *Protein Eng Des Sel* **17:** 417-424.

Pierani, A, Moran-Rivard, L, Sunshine, MJ, Littman, DR, Goulding, M , Jessell, TM(2001) Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1. *Neuron* **29:** 367-384.

Pierre, K , Pellerin, L(2005) Monocarboxylate transporters in the central nervous system: distribution, regulation and function. *J Neurochem* **94:** 1-14.

Pierre, K, Pellerin, L, Debernardi, R, Riederer, BM , Magistretti, PJ(2000) Cell-specific localization of monocarboxylate transporters, MCT1 and MCT2, in the adult mouse brain revealed by double immunohistochemical labeling and confocal microscopy. *Neuroscience* **100:** 617-627.

Plagemann, PG, Gregory, KF , Wroblewski, F(1960) The electrophoretically distinct forms of mammalian lactic dehydrogenase. 1. Distribution of lactic dehydrogenase. 1. Distribution of lactic dehydrogenases in rabbit and human tissue. *J Biol Chem* **235:** 2282-2287.

Posada, D , Crandall, KA(1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817-818.

Prager, EM , Wilson, AC(1975) Slow evolutionary loss of the potential for interspecific hybridization in birds: a manifestation of slow regulatory evolution. *Proc Natl Acad Sci U S A* **72:** 200-204.

Ramayya, MS, Zhou, J, Kino, T, Segars, JH, Bondy, CA , Chrousos, GP(1997) Steroidogenic factor 1 messenger ribonucleic acid expression in steroidogenic and nonsteroidogenic human tissues: Northern blot and in situ hybridization studies. *J Clin Endocrinol Metab* **82:** 1799-1806.

Ramji, DP , Foka, P(2002) CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J* **365:** 561-575.

Ray, A, LaForge, KS , Sehgal, PB(1990) On the mechanism for efficient repression of the interleukin-6 promoter by glucocorticoids: enhancer, TATA box, and RNA start site (Inr motif) occlusion. *Mol Cell Biol* **10:** 5736-5746.

Revol De Mendoza, A, Esquivel Escobedo, D, Martinez Davila, I , Saldana, H(2004) Expansion and divergence of the GH locus between spider monkey and chimpanzee. *Gene* **336:** 185-193.

Robergs, RA, Ghiasvand, F , Parker, D(2004) Biochemistry of exercise-induced metabolic acidosis. *Am J Physiol Regul Integr Comp Physiol* **287:** R502-516.

Rodriguez, S, Gaunt, TR , Day, IN(2007) Molecular genetics of human growth hormone, insulin-like growth factors and their pathways in common disease. *Hum Genet* **122:** 1-21.

Ronquist, F , Huelsenbeck, JP(2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

Ross, ED, Hardwidge, PR , Maher, LJ, 3rd(2001) HMG proteins and DNA flexibility in transcription activation. *Mol Cell Biol* **21:** 6598-6605.

Rylski, M, Amborska, R, Zybura, K, Michaluk, P, Bielinska, B, Konopacki, FA, Wilczynski, GM , Kaczmarek, L(2009) JunB is a repressor of MMP-9 transcription in depolarized rat brain neurons. *Mol Cell Neurosci* **40:** 98-110.

Sabeti, PC, Schaffner, SF, Fry, B, Lohmueller, J, Varilly, P, Shamovsky, O, Palma, A, Mikkelsen, TS, Altshuler, D , Lander, ES(2006) Positive natural selection in the human lineage. *Science* **312:** 1614-1620.

Sanchez-Munoz, I, Sanchez-Franco, F, Vallejo, M, Fernandez, A, Palacios, N, Fernandez, M , Cacicedo, L(2010) Activity-dependent somatostatin gene expression is regulated by cAMP-dependent protein kinase and Ca2+-calmodulin kinase pathways. *J Neurosci Res* **88:** 825-836.

Sandhoff, TW, Hales, DB, Hales, KH , McLean, MP(1998) Transcriptional regulation of the rat steroidogenic acute regulatory protein gene by steroidogenic factor 1. *Endocrinology* **139:** 4820-4831.

Sanger, TJ, Losos, JB , Gibson-Brown, JJ(2008) A developmental staging series for the lizard genus Anolis: a new system for the integration of evolution, development, and ecology. *J Morphol* **269:** 129-137.

Satija, R, Novak, A, Miklos, I, Lyngso, R , Hein, J(2009) BigFoot: Bayesian alignment and phylogenetic footprinting with MCMC. *BMC Evol Biol* **9:** 217.

Schiessl, B, Strasburger, CJ, Bidlingmaier, M, Gutt, B, Kirk, SE, Oberhoffer, R , Friese, K(2007) Role of placental growth hormone in the alteration of maternal arterial resistance in pregnancy. *J Reprod Med* **52:** 313-316.

Schmid, J , Speakman, JR(2000) Daily energy expenditure of the grey mouse lemur (Microcebus murinus): a small primate that uses torpor. *J Comp Physiol B* **170:** 633-641.

Schulke, O , Ostner, J(2007) Physiological ecology of cheirogaleid primates: variation in hibernation and torpor. *Acta ethologica* **10:** 13-21.

Semenza, GL, Jiang, BH, Leung, SW, Passantino, R, Concordet, JP, Maire, P , Giallongo, A(1996) Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. *J Biol Chem* **271:** 32529-32537.

Sherwood, CC, Stimpson, CD, Raghanti, MA, Wildman, DE, Uddin, M, Grossman, LI, Goodman, M, Redmond, JC, Bonar, CJ, Erwin, JM et al.(2006) Evolution of increased glia-neuron ratios in the human frontal cortex. *Proc Natl Acad Sci U S A* **103:** 13606-13611.

Shim, H, Dolde, C, Lewis, BC, Wu, CS, Dang, G, Jungmann, RA, Dalla-Favera, R , Dang, CV(1997) c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proc Natl Acad Sci U S A* **94:** 6658-6663.

Short, ML, Huang, D, Milkowski, DM, Short, S, Kunstman, K, Soong, CJ, Chung, KC , Jungmann, RA(1994) Analysis of the rat lactate dehydrogenase A subunit gene promoter/regulatory region. *Biochem J* **304 (Pt 2):** 391-398.

Shoshani, J, Kupsky, WJ , Marchant, GH(2006) Elephant brain. Part I: gross morphology, functions, comparative anatomy, and evolution. *Brain Res Bull* **70:** 124-157.

Simpson, G(1945) The principles of classification and a classification of mammals. *Bulletin of the American Museum of Natural History* **85:** 1-350.

Sogawa, K, Imataka, H, Yamasaki, Y, Kusume, H, Abe, H , Fujii-Kuriyama, Y(1993) cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2. *Nucleic Acids Res* **21:** 1527-1532.

Solomon, MJ, Larsen, PL , Varshavsky, A(1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53:** 937-947.

Springer, MS, Murphy, WJ, Eizirik, E , O'Brien, SJ(2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* **100:** 1056-1061.

Srivastava, K, Srivastava, A , Mittal, B(2010) DNMT3B -579 G>T Promoter Polymorphism and Risk of Gallbladder Carcinoma in North Indian Population. *J Gastrointest Cancer*.

Stephenson, M(1928) On lactic dehydrogenase: A cell-free enzyme preparation obtained from bacteria. *Biochem J* **22:** 605-614.

Suzuki, H, Hosokawa, Y, Nishikimi, M , Ozawa, T(1991) Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system. *J Biol Chem* **266:** 2333-2338.

Swofford, DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Syner, FN , Goodman, M(1966) Differences in the lactic dehydrogenases of primate brains. *Nature* **209:** 426-428.

Tagle, DA, Koop, BF, Goodman, M, Slightom, JL, Hess, DL , Jones, RT(1988) Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol* **203:** 439-455.

Tatusova, TA , Madden, TL(1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174:** 247-250.

Than, NG, Romero, R, Goodman, M, Weckle, A, Xing, J, Dong, Z, Xu, Y, Tarquini, F, Szilagyi, A, Gal, P et al.(2009) A primate subfamily of galectins expressed at the maternal-fetal interface that promote immune cell death. *Proc Natl Acad Sci U S A* **106:** 9731-9736.

Tian, D, Huang, D, Short, S, Short, ML , Jungmann, RA(1998) Protein kinase A-regulated instability site in the 3'-untranslated region of lactate dehydrogenase-A subunit mRNA. *J Biol Chem* **273:** 24861-24866.

Travers, AA(2004) The structural basis of DNA flexibility. *Philos Transact A Math Phys Eng Sci* **362:** 1423-1438.

Truter, S, Di Liberto, M, Inagaki, Y , Ramirez, F(1992) Identification of an upstream regulatory region essential for cell type-specific transcription of the pro-alpha 2(V) collagen gene (COL5A2). *J Biol Chem* **267:** 25389-25395.

Untergasser, G, Hermann, M, Rumpold, H, Pfister, G , Berger, P(2000) An unusual member of the human growth hormone/placental lactogen (GH/PL) family, the testicular alternative splicing variant hPL-A2: recombinant expression revealed a membrane-associated growth factor molecule. *Mol Cell Endocrinol* **167:** 117-125.

Vander Heiden, MG, Cantley, LC , Thompson, CB(2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324:** 1029-1033.

Vesell, ES(1961) Significance of the heterogeneity of lactic dehydrogenase activity in human tissues. *Ann N Y Acad Sci* **94:** 877-889.

Vesell, ES , Bearn, AG(1961) Isozymes of lactic dehydrogenase in human tissues. *J Clin Invest* **40:** 586-591.

Viengchareun, S, Servel, N, Feve, B, Freemark, M, Lombes, M , Binart, N(2008) Prolactin receptor signaling is essential for perinatal brown adipocyte function: a role for insulin-like growth factor-2. *PLoS One* **3:** e1535.

Wain, HM, Bruford, EA, Lovering, RC, Lush, MJ, Wright, MW , Povey, S(2002) Guidelines for human gene nomenclature. *Genomics* **79:** 464-470.

Wallis, OC , Wallis, M(2002) Characterisation of the GH gene cluster in a new-world monkey, the marmoset (Callithrix jacchus). *J Mol Endocrinol* **29:** 89-97.

Wallis, OC , Wallis, M(2006) Evolution of growth hormone in primates: the GH gene clusters of the New World monkeys marmoset (Callithrix jacchus) and white-fronted capuchin (Cebus albifrons). *J Mol Evol* **63:** 591-601.

Wang, DG, Fan, JB, Siao, CJ, Berno, A, Young, P, Sapolsky, R, Ghandour, G, Perkins, N, Winchester, E, Spencer, J et al.(1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:** 1077-1082.

Wang, QF, Prabhakar, S, Chanan, S, Cheng, JF, Rubin, EM , Boffelli, D(2007) Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons. *Genome Biol* **8:** R1.

Weake, VM , Workman, JL(2010) Inducible gene expression: diverse regulatory mechanisms. *Nat Rev Genet* **11:** 426-437.

Welch, KC, Jr. , Suarez, RK(2007) Oxidation rate and turnover of ingested sugar in hovering Anna's (Calypte anna) and rufous (Selasphorus rufus) hummingbirds. *J Exp Biol* **210:** 2154-2162.

Wlasiuk, G, Khan, S, Switzer, WM , Nachman, MW(2009) A history of recurrent positive selection at the toll-like receptor 5 in primates. *Mol Biol Evol* **26:** 937-949.

Xie, X, Rigor, P , Baldi, P(2009) MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **25:** 167-174.

Yan, J, Barnes, BM, Kohl, F , Marr, TG(2008) Modulation of gene expression in hibernating arctic ground squirrels. *Physiol Genomics* **32:** 170-181.

Yanagida, A, Sogawa, K, Yasumoto, KI , Fujii-Kuriyama, Y(1990) A novel cis-acting DNA element required for a high level of inducible expression of the rat P-450c gene. *Mol Cell Biol* **10:** 1470-1475.

Yang, Z(1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13:** 555-556.

Zhang, C(2009) Novel functions for small RNA molecules. *Curr Opin Mol Ther* **11:** 641-651.

Zhang, J, Nielsen, R , Yang, Z(2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* **22:** 2472-2479.

Zhao, S, Shao, C, Goropashnaya, AV, Stewart, NC, Xu, Y, Toien, O, Barnes, BM, Fedorov, VB , Yan, J(2010) Genomic analysis of expressed sequence tags in American black bear Ursus americanus. *BMC Genomics* **11:** 201.

Zhou, Y, Xu, BC, Maheshwari, HG, He, L, Reed, M, Lozykowski, M, Okada, S, Cataldo, L, Coschigamo, K, Wagner, TE et al.(1997) A mammalian model for Laron syndrome produced by targeted disruption of the mouse growth hormone receptor/binding protein gene (the Laron mouse). *Proc Natl Acad Sci U S A* **94:** 13215-13220.

ABSTRACT

EVOLUTION OF LACTATE DEHYDROGENASE GENES IN PRIMATES, WITH SPECIAL CONSIDERATION OF NUCLEOTIDE ORGANIZATION IN MAMMALIAN PROMOTERS

by

ZACK PAPPER

August 2010

Advisor: Dr. Derek E. Wildman

Major: Molecular Medicine and Genetics

Degree: Doctor of Philosophy

Concomitant with an increase in brain volume and mass, the allocation of energetic resources to the brain increased during stem anthropoid evolution, leading to humans. One mechanism by which this allocation may have occurred is through greater use of lactate as a neuronal fuel. Both the production of lactate, and conversion to pyruvate for use in aerobic metabolism, are catalyzed, in part, by the tetrameric enzyme lactate dehydrogenase (LDH). The two primary LDH genes, *LDHA* and *LDHB*, confer different rates of substrate turnover to the LDH enzyme, and these rates lend to the argument that *LDHA* supports anaerobic while *LDHB* supports aerobic metabolism. The expression profiles of these proteins shifted during primate evolution, with LDHA and LDHB the primary LDH proteins expressed in strepsirrhine and anthropoid brains, respectively. We demonstrate that this expression shift does not coincide with changes to protein structure.

Previous research has identified critical *cis*-regulatory elements within the *LDHA* promoter, demonstrating that transcriptional regulation is critical for proper expression of this gene. In this thesis, we characterize the promoters of *LDHA* and *LDHB* in primates, in order to determine the elements responsible for the expression shifts in brain during primate evolution. We identify motifs conserved across mammals, likely responsible for the common expression profiles. We also identify elements that were gained during different periods of primate evolution. Anthropoid-specific elements in the *LDHA* promoter include a modification of a known Sp1 site, as well as two putative repressor elements. Anthropoid-specific elements in the *LDHB* promoter include an oxidative phosphorylation element, which may coordinate aerobic metabolism pathways. In addition, both promoters have CpG sites conserved across mammals, which led us to hypothesize that species-specific and/or tissue-specific epigenetic modifications may have also changed during primate evolution. We conduct a cross-tissue, cross-species methylation analysis, and determine that CpG methylation patterns across tissues appear similar between human and dwarf lemur; however, methylation levels across species vary.

AUTOBIOGRAPHICAL STATEMENT

Zack Papper
Center for Molecular Medicine & Genetics
Wayne State University School of Medicine
540 E. Canfield
Detroit, MI 48201
313.577.0086
zpapper@gmail.com

EDUCATION

Ph.D., Molecular Medicine & Genetics, Wayne State University School of Medicine, Detroit, MI
Evolution of LDHA and LDHB in Primates
(Dissertation), Derek E. Wildman (Advisor)
Projected Defense, August 2010

2002 Certificate of Emergency Medical Technician, Oakland Community College, Auburn Hills, MI

2000 B.S., Biology, University of Michigan, Ann Arbor, MI

RESEARCH EXPERIENCE

Evolution of *LDHA* and *LDHB* coding regions and promoters in primates
Center for Molecular Medicine & Genetics, Wayne State University, Detroit, MI
(NSF-funded grant/Sigma Xi grant)
Manuscript in Prep

Traversing the mammalian promoter landscape
Center for Molecular Medicine & Genetics, Wayne State University, Detroit, MI
(NSF-funded grant)
Manuscript in Submission

DNA extraction from bone samples
Center for Molecular Medicine & Genetics, Wayne State University, Detroit, MI
Experimental design, tissue decontamination, DNA extraction, validation through DNA sequencing

Ancient origin of placental expression in the growth hormone genes of anthropoid primates
Center for Molecular Medicine & Genetics, Wayne State University, Detroit, MI
(NICHD-funded grant)
Manuscript published

PUBLICATIONS

2009 Papper, Z, NM Jameson, R Romero, AL Weckle, P Mittal, K Benirschke, J Santolaya-Forgas, M Uddin, D Haig, M Goodman, DE Wildman. Ancient origin of placental expression in the growth hormone genes of anthropoid primates. (2009) *Proc Natl Acad Sci USA*. 106(40):17083-8.