THE EVOLUTION OF A BMP5 ENHANCER IN PRIMATES

A thesis submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

by
Kelly Eileen Droney
August, 2011
Thesis written by
Kelly Eileen Droney
B.S., B.A. The George Washington University, 2003
M.A. Kent State University, 2011

Approved by
Chi-hua Chiu ________________________, Advisor
Richard Meindl ________________________, Chair, Department of Anthropology
John R.D. Stalvey ________________________, Dean, College of Arts and Sciences
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... v
LIST OF TABLES .............................................................................................................. vi
ACKNOWLEDGEMENTS .................................................................................................. vii
ABSTRACT ..................................................................................................................... viii

Chapter

I. INTRODUCTION ........................................................................................................... 1
   Rib Curvature in Primates ......................................................................................... 5
   Developmental Genetic Basis for Morphological Evolution ......................... 7
   Bone Morphogenetic Protein 5 (BMP5) ............................................................... 14
   Hypothesis ............................................................................................................. 20

II. MATERIALS AND METHODS .................................................................................... 22
   PCR, Cloning, and Sequencing bmp5 Enhancer “A” in Primates ..................... 22
   Phylogenetic Analysis ............................................................................................ 31
   MATCH™ Analysis ................................................................................................. 31

III. RESULTS .................................................................................................................. 32
   Phylogenetic Tree Analysis .................................................................................. 32
   MATCH™ Analysis ................................................................................................. 34
   Bone Morphogenetic Protein (BMP5) ................................................................. 38
IV. DISCUSSION ...........................................................................................................39

V. CONCLUSIONS ......................................................................................................44

APPENDIX A. Genomic Alignment of Bmp5 Enhancer Region .............................47

APPENDIX B. Reconstructed Phylogenetic Tree .....................................................62

REFERENCES ...........................................................................................................64
LIST OF FIGURES

1. Vertebral Column in hominoids and its effect on thorax shape..........................4
2. Example of Non-coding Region’s Effect on Coding Region ............................9
3. Different Models of Gene Evolution Increase the Diversity of Gene Function ...........................................................13
4. Bmp5 Enhancer Locus in Mouse ..................................................................23
5. Map of Primers ............................................................................................25
6. Predicted Transcription Factor Binding Sites ............................................36
List of Tables

1. Summary of Study Species ......................................................................................12
2. List of *bmp5* enhancer region for species with mapped genomes .................26
3. Table of Primers ..................................................................................................30
4. Known Function of Transcription Factors or Proteins .....................................37
Acknowledgements

If patience is a virtue, I have found many virtuous individuals through this process. Firstly, I am very appreciative of the opportunity to complete this project with my advisor, Chi-hua Chiu. The resources and time that it takes to complete a laboratory-based study are great. Thank you for allowing me to learn under your wing in an advanced research facility. The resources and time that it takes to endure graduate school are also great. Thank you for supporting me through the ups and downs of the past two years. I would like to thank my husband, Nathan Olsson, for moving away from the ocean and the mountains so that we could pursue this endeavor together. You have taught me just how patient and kind love really is. I owe you a lifetime of surfing and rock-climbing. Committee members, you are the most “virtuous” of all those involved with my thesis. Thank you for telling me that there was something there on the paper and helping to make it readable. I felt encouraged from one end of that learning curve to the other.

To my fellow colleagues in the trenches, thank you for listening. A good chat in Lowry Hall does a body good. The camaraderie will be sorely missed. To the family, both near and far, thank you for believing in my abilities even when I did not. Thank you for encouraging me to do something that makes me happy, not something that makes me a lot of money. I’m doing it.

Kelly Droney Olsson
11 July 2011   Kent, Ohio
ABSTRACT

Reorganization of the primate bauplan allowed for changes to locomotor patterns. Evidence of such bauplan reorganization includes modifications in the shape of the thoracic cage, vertebral column invagination and repositioning of the scapulae. Specific to the rib cage, Guenther et.al (2008) functionally defined a cis-regulatory sequence of bone morphogenetic protein 5 (bmp5) that modulates its expression in mouse rib anlaga and influences rib curvature. I PCR amplified, cloned, and sequenced the noncoding DNA sequence in several primate taxa that were not previously publicly available including gibbon, squirrel monkey, and spider monkey and aligned these to of human, chimpanzee, orangutan, rhesus monkey, and common marmoset sequences. I confirmed orthology of the enhancer among the species of this study through phylogenetic analysis. Using a program that predicts transcription factor binding sites and, I traced the progression of putatively functional nucleotide substitutions in this bmp5 enhancer throughout primate phylogeny that potentially alter transcription factor binding in key taxa including human, gibbon, and spider monkey. While the enhancer is conserved, there are other species-specific nucleotide changes to the enhancer sequence which could have an unknown function, such as influencing the binding stability of TFs (e.g. nucleotide neighbor effects). The specific phenotypic effect of this bmp5 enhancer on primate rib curvature is not yet known. This study demonstrates the power of experimental and bioinformatic approaches to investigating big questions in human evolution and paves the way for future functional studies to directly relate genetic and phenotypic evolution of the primate thorax.
CHAPTER ONE: INTRODUCTION

The modern human skeleton reflects changes to and reorganization of various regions of the primate bauplan as an adaptation to bipedality. Evidence for such reorganization appears early in human evolutionary history. Investigating the nature and timing of the first appearance of these skeletal changes is hampered both by an incomplete hominid fossil record and by a lack of understanding about their genetic and developmental background. Not all skeletal changes are likely to have occurred simultaneously and thereby not likely to be found in any single fossil; and, as expected, analysis of available fossils indicates that there was a sequence of changes that have led to modern human form. While some anthropologists (Guy et al. 2005; Zollikofer et al. 2005) have argued that the position of the foramen magnum in *Sahelanthropus tchadensis* is a correlate of bipedality, suggesting that bipedality arose six to seven million years ago, others have attributed this trait to changes in neural development and not locomotion (Biegert 1963; Ruth 2010; Richmond and Jungers 2008). *Orrorin tugenensis* femora possess some characters, such as a long femoral neck and a human-like obturator externus groove, that indicate that it was at least partially bipedal six million years ago (Pickford et al. 2002; Richmond and Jungers 2008). Until recently, many anthropologists agreed that the genus *Australopithecus* provided the earliest, absolute evidence of bipedality at 3 to 4 million years before present (Stanford, Allen, Antón, 2008), as it shows an anteroposteriorly and craniocaudally shortened pelvis, more
human-like limb proportions, a nonopposable first pedal ray, and a more human-like thoracic cage. (Johanson et al. 1982; Haile-Selassie et al. 2010). The recently analyzed 4.4 million year old *Ardipithecus ramidus* is unique in that it possesses some intermediate anatomical features suggesting that bipedality occurred in a gradual, stepwise process (Lovejoy et al. 2009a,b,c).

The hominid fossil record demonstrates that bipedality is associated with changes in the position of the vertebral column within the thorax, with attendant changes to rib curvature and shape. The Last Common Ancestor (LCA) of hominoids had an invaginated vertebral column: as the vertebral column moves more ventrally within the body cavity the vertebrae become embedded more ventrally in the rib cage (Jellema et al. 1993; McCollum et al. 2009; Lovejoy 2005). Vertebral column invagination is controlled by an embryogenetic mechanism that also posterolateralized the pectoral girdle resulting in a more lateral-facing glenoid. The resulting posterior translation of the shoulder girdle was an adaptation to arboreal quadrumanual scrambling (Lovejoy et al 2009a). It has been hypothesized that hominids evolved from a LCA that carefully climbed Lovejoy et al. (2009a). The ancestors of African apes likely reduced the length of the vertebral column as an adaptation to vertical climbing and suspension as a way to increase abdominal rigidity (Lovejoy et al. 2009a). All hominoids (e.g., *Hylobates lar, the African apes*) have glenoid cavities that face more posterolaterally than those of above branch quadrupeds and vertical clinging and leaping primates (i.e. *Colobus guerza, Macaca mulatta, Saimiri sciureus, Callithrix jacchus*). Their thoraxes are expected to reflect the associated structural differences.
In habitually bipedal hominids, such as *Australopithecus afarensis*, *Homo erectus*, and *Homo sapiens*, the lumbar region of the vertebral column is not reduced and exhibits a unique lordosis. The vertebral column invagination, in concert with pronounced lumbar lordosis, causes the head and trunk to become repositioned over the hip joints (See Figure 1; McCollum et al 2009; Lovejoy et al. 2009b,c). For example, specimens from *Australopithecus afarensis* (Johanson et al. 1982; Haile-Selassie et al. 2010), *Australopithecus africanus* (Sanders 1998) and *Ardipithecus ramidus* (“Ardi”) (Lovejoy et al. 2009a) show evidence of lordotic changes to the vertebral column with coincident changes of thoracic cage position and shape.

Rib curvature and thoracic form should reflect corresponding differences in the genome, and species diversity is largely due to the genetic regulatory mechanisms that operate during early development (Carroll et al. 2001). Rib development in vertebrates is almost certainly modulated by a conserved set of regulatory genes, including the gene for bone morphogenetic protein 5 (*bmp5*) (based on Guenther et al. 2008). Therefore, I hypothesize that variation in *bmp5* regulatory sequences contributes to differences in rib curvature. As a first step in testing this hypothesis, I have sequenced and analyzed the evolution of a “rib curvature enhancer” of the *bmp5* locus (enhancer “A”, defined in mouse; Guenther et al. 2008) in a comparative range of primates including humans, apes, Old World monkeys, and New World monkeys, and report the results of my work here.
FIGURE 1 - Vertebral column invagination in hominoids and its effect on thorax shape

The superior view of the thoracic cage depicting cross-sectional shapes of the thoraxes of an Old World Monkey (top), chimpanzee (middle) and modern human (bottom). Note that the overall thoracic shape changes as invagination increases, particularly the lateral curvature of the rib. The dorsal projection of the rib also increases as the vertebral column becomes further imbedded in the thoracic cage.

Redrawn with permission from McCollum et al. 2009; redrawn from Schultz 1961
I. Rib Curvature Differences in Primates

Many investigators have studied thoracic cage variation among living primates (Chan 1997; Harrison 1986; Kagaya 2008, 2009; Larsen 1998; Schultz 1961; Stern et al. 1980; etc). They have shown that the thoraces of living hominoids (chimpanzee, gorilla, orangutan, and gibbon) are distinct from those of non-hominoid primates in being relatively broad and comparatively shallow with more strongly curved ribs.

There are two corresponding features in the upper thorax of all hominoids: (1) the dorsal portion of the rib protrudes posteriorly, as is expected with a more ventrally positioned vertebral column (Schultz 1941, 1942, 1956), and (2) the sternal end of the rib projects more medially (Kagaya et al. 2008:89). Traits 1 and 2 are claimed to produce a mechanical advantage in orthograde and/or suspensory locomotion. Kagaya et al. (2008) concluded that while Pan exhibited a “classic” hominoid-like thoracic cage; it still displays a non-hominoid posterior depth suggesting that its vertebral column is not invaginated to the same degree as in other hominoids. Additionally, they confirmed that the “barrel-shaped” thoracic cage in Hylobates (gibbon) can be distinguished from the “funnel-shaped” form in other hominoids (Schultz 1961; Schmid 1983). In suspensory Ateles (spider monkeys), Kagaya et al. (2008) found that features 1 and 2 were only minimally present. The forelimb morphology of Ateles is characterized as functionally convergent with extant apes, which is not surprising given that Ateles, Pongo and Hylobates rely on significant suspensory locomotion (Larsen et al. 1998). Therefore, Ateles is unique in demonstrating traits indicative of a suspensory forelimb yet lacking
the dorsal projection and medial orientation of the sternal end of ribs that is present in other suspensory hominoids. This contradicts the observation that suspensory primates, i.e. *Hylobates* and *Ateles*, have similar thoracic anatomy.

Kagaya et al. (2009) also confirmed quantitatively that hominoids exhibit more pronounced caudal obliquity in the lower ribs compared to Old and New World monkeys. Stern et al. (1980) proposed that this feature keeps the scapula of the weight-bearing forelimb on the dorsal surface of the trunk, suggesting reliance on the serratus anterior during propulsion. The position of maximum thoracic breadth, which coincides with the most inferior attachment of the serratus anterior, is more caudally located in *Hylobates* and *Pongo*. Kagaya et al. (2009) speculated that a combination of positional and forelimb suspensory behaviors have contributed to increased obliquity of the lower ribs in these species.

In comparison to extant apes, modern humans have cranio-caudally elongated thoraxes (Haile-Selassie et al. 2010), and an increased thoracic diameter due to a different degree of rib curvature in 1) each rib and 2) particular regions of each rib itself. Modern humans also exhibit uniquely angled rib bodies relative to the costal attachment with the vertebral column, indicating a more invaginated vertebral column (Haile-Selassie et al. 2010). Jellema et al. (1993) quantified the curvature of each rib in modern humans and chimpanzees and demonstrated morphological adaptations for bipedality in humans, such as a pronounced dorsal projection of the proximal portion of each rib (See Figure 1) and a distinct change in the orientation of each rib in lateral view.
II. Developmental Genetic Basis for Morphological Evolution

Within the genome it is useful to isolate two different categories of nucleotide sequences: protein coding and non-coding. Coding sequences specify the protein or proteins (in the case of alternative splicing) that carry out essential cellular functions. Slight alterations in coding sequences via mutation (e.g. point mutation, recombination) are typically under strong negative directional selection. Hence, coding sequences are typically under very strong purifying selection (Ka/Ks <1), where Ka is the number of nonsynonymous nucleotide substitutions and Ks is the number of synonymous substitutions.

The majority of virtually any genome is composed of noncoding DNA. Its full range of functions is still not well understood. Cis-regulatory DNA (e.g. enhancers, promoters) is one functional class of noncoding DNA that influences the transcription (expression) of gene products in temporal and/or spatial contexts (Edelman 1988; 1986a,b). For example, some human genes are ubiquitously expressed; their protein products are made throughout ontogeny and/or in all or several cell types. In contrast some other genes, e.g. homeoboxes, are expressed only in very specific spatiotemporal domains of the early embryo and are otherwise inactive for most of ontogeny (Carroll et al. 2001; Chiu and Hamrick 2002). The spatial and temporal ‘instructions’ for gene transcription and protein production are carried out by cis-regulatory sequences. While there is strong negative selection maintaining protein coding sequences, changes to the cis-regulatory sequences are more readily tolerated even though they can have major
effects within species (Carroll et al. 2001). Additionally, *cis*-regulatory sequences often have pleiotropic effects; therefore, changes to these sequences may result in a number of phenotypic changes.

In eukaryotes, the typical architecture of *cis*-regulatory sequences is hierarchical. At the most basic level, transcription factor binding sites (TFBS) are typically 8-10 base-pair-long DNA sequences that are specifically bound by transcription factor proteins (See Figure 2). At the second level, groups of TFBS are clustered in a relatively short span of DNA sequence (e.g. 100 base pairs or less), forming a *cis*-regulatory ‘module’. At the highest level, several *cis*-regulatory ‘modules’ modulate a given genomic locus (gene) and provide instructions on the quantity of protein that is produced and where and when it is produced in any given cell. For example, as an extreme simplification, a protein product can be made both in the brain and in the limb. One *cis*-regulatory module specifies ‘brain’ and a second *cis*-regulatory module specifies ‘limb’. Additional *cis*-regulatory modules will specify ‘when’ and ‘how much’ protein is produced in a specific spatial context (Chiu and Hamrick 2002; Berg et al. 2004).

There are very few defined enhancers for regulatory genes and identifying these sequences using phylogenetic or transgenic methods is tedious and problematic, and often described as “finding a needle in a haystack” (Carroll et al. 2001; Carroll 2005; Chiu and Hamrick 2002; Sholtis and Noonan 2010). The biggest bottleneck in evolutionary and developmental biology (evo-devo) research is locating and defining *cis*-regulatory sequences. The location of *cis*-regulatory sequences in the genome is varied and unpredictable with respect to the gene or gene(s) they modulate. For example,
FIGURE 2 – Example of Non-coding Region’s Effect on Coding Region

In this figure, the coding sequence is at the far right of the genetic sequence, depicted by the vertical bars containing it. The promoter region is highlighted in light green, and is positioned relatively close to the coding region. The two transcription factors (green ovals) are bound to the enhancer region (yellow box) that is located a greater distance from the promoter and coding sequence. Once the transcription factors have bound and engaged the enhancer region, the promoter region can be activated, thereby causing transcription of the specific protein.
they can lie thousands of base pairs away from their target coding (or other downstream noncoding) region, and occasionally even on a different chromosome. In fact, enhancers are known to act upon gene expression regions more than one million base pairs away (Levin 2010). Even if an educated guess as to their possible location turns out to be reasonably accurate, locating them is still a time-consuming task. In *Drosophila*, for example, one known enhancer of the *dpp* coding sequence is located 30 kilobases away from the promoting region of the gene it modulates (DiLeone et al. 2000). DiLeone et al. (1998, 2000) attempted to locate *bmp5* regulatory regions using large scale bacterial artificial chromosome clones (BAC clones). Despite their thorough study, they were unable to identify the particular *bmp5* enhancer analyzed in my study. A complicating problem in genetic regulatory studies is that genes and their associated regulatory regions are rarely in one-to-one relationships, and are instead usually parts of complex networks of interactions and signaling molecules that are involved in many, sometimes unrelated, pathways. Even if several coding and noncoding sequences within such a complex can be identified, their definition does not necessarily provide information with respect to such complex interactions and functions.

A great deal of research has focused on the evolution of *cis*-regulatory sequences (Carroll et al. 2001, 2005, 2007; Visel et al. 2009; Berg et al. 2004; Bird et al. 2007; Sholtis and Noonan 2010; King and Wilson 1975) (See Figure 3). Evo-devo research has shown that the genetic toolkit (i.e. protein products of developmental control genes) underlying the diversity of body plans is surprisingly small (Carroll 2005). Furthermore, small changes to the genetic toolkit often result in morphological changes among species
via lineage-specific modifications of *cis*-regulatory sequences (Chiu and Hamrick 2002).

It has long been recognized that the amount of phenotypic variation between chimpanzees and humans is not entirely explained by the small amount of genetic divergence in protein coding sequences present in the two species (King and Wilson 1975). Instead, evolutionary changes in phenotype are more likely to be the consequence of the mechanisms controlling gene expression (Carroll 2005) (See Figure 3).

One way to trace the evolution of a trait, such as rib curvature and shape, is to compare gene expression and rib development in closely related taxa. Due to extreme conservation of the genetic toolkit of patterning genes, model organisms such as mouse, chick and frog can provide some insight into gene expression patterns and functions in development and their correlation with body plan evolution across species. One of the most classical examples of this procedure was described by Burke et al. 1995. Although the *Hox* C6 and *Hox* C8 protein coding sequences are nearly 100% conserved in mouse and chick, the spatio-temporal expression (transcription) patterns of these genes are different in each taxon and are partly responsible for different trunk lengths (i.e. numbers of thoracic vertebrae) in mammals and birds. However, the *Hox* C6 and *Hox* C8 enhancers responsible for expression in the anlagen of trunk (thoracic) vertebrae have not yet been found in mouse vs. chick. Consequently, the proximate DNA sequence differences are still not known.

A small handful of feasible and powerful approaches has been developed to locate *cis*-regulatory sequences. Hardison (2000) described a wide range protocol for identifying regulatory sequences among multiple species: Conserved Noncoding (CNC)
<table>
<thead>
<tr>
<th>Species</th>
<th>Common Name</th>
<th>Locomotor pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mus musculus</em></td>
<td>Mouse</td>
<td>Quadrupedalism</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>Human</td>
<td>Bipedalism</td>
</tr>
<tr>
<td><em>Pan troglodytes</em></td>
<td>Chimpanzee</td>
<td>Knuckle-walking and Vertical Climbing</td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>Orangutan</td>
<td>Suspensory/Brachiation</td>
</tr>
<tr>
<td><em>Hylobates lar</em></td>
<td>Gibbon</td>
<td>Suspensory/Brachiation</td>
</tr>
<tr>
<td><em>Macaca mulatta</em></td>
<td>Macaque</td>
<td>Terrestrial/Arboreal Quadrupedalism</td>
</tr>
<tr>
<td><em>Saimiri sciureus</em></td>
<td>Squirrel monkey</td>
<td>Arboreal Quadrupedalism</td>
</tr>
<tr>
<td><em>Ateles paniscus</em></td>
<td>Spider monkey</td>
<td>Suspensory, Brachiation</td>
</tr>
<tr>
<td><em>Callithrix jacchus</em></td>
<td>Marmoset</td>
<td>Arboreal Quadrupedalism</td>
</tr>
</tbody>
</table>
FIGURE 3 - Different Modes of Gene Evolution Increase the Diversity of Gene Function

Adapted from Carroll 2005

The function of the original gene (A) with one *regulatory element* (circle) and a pair of *exons* (rectangles) can be expanded and diversified in several ways.

(B) Diversification of *cis-regulatory elements* by various means (B1: expansion/duplication, B2: point mutation, B3: rearrangement) can alter the number of tissues in which the gene is active. For example, in B2 expansion and mutation may cause increase/decrease in protein production in two new tissues. In B3, rearrangement may dictate activity of gene in new tissue.

(C1) Depiction of purifying selection on the protein-coding and *cis*-regulatory elements
sequence analysis. Use of this approach alone is by no means exhaustive as it only highlights a small number of cis-regulatory sequences in the genome. Visel et al. (2009) detailed the use of chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) as a means of identifying enhancer activity in specific tissues during development. ChIP-seq is unique in that it is possible to locate enhancer regions that are not under evolutionary conservation, thereby potentially detecting species-specific regulatory sequences. Embryo electroporation has also been successful in identifying large groups of enhancers in chick (Kamachi et al. 2009; Uchikawa et al. 2004) but this approach is not possible in primates.

III. Bone Morphogenetic Protein 5 (BMP5)

BMP5 is a member of the larger protein superfamily TGF-β (transforming growth factor β). It is one of several bone morphogenetic proteins originally defined based on the ability to stimulate cartilage and bone formation contributing to the size, shape, and patterning within condensations, anlagen, and physes (Urist 1965; Urist and Strates 1971 Lovejoy et al. 2003, DiLeone et al. 1998). BMPs induce bone formation suggesting that these proteins are key players in embryonic bone development, repair and growth (Rosen and Theis 1992 for review). Purified BMP5 can stimulate the formation of new cartilaginous and skeletal growth when implanted under the skin of adult mice (King et al. 1994). It is likely that the BMP5 protein would be constructed similarly in amino acid sequence or folding pattern between mouse and human. King et al. (1994) demonstrated
that mouse BMP5 showed 92% identity to the human BMP5 in the pre-pro-domain and 96% identity in the mature signaling regions (Burt 1992).

Postcranial bone formation, and related tendon and muscle attachment, is controlled by a set of genes that encode a variety of transcription factors and cell-signaling molecules (Lovejoy et al. 2003). Skeletal development begins with the formation of mesenchymal cell condensations which differentiate into chondrocytes that express specific molecular markers (Lovejoy et al. 2003). A complex regulatory feedback loop acts to maintain and control the timing of chondrocyte maturation (Vortkamp et al. 1996, Vortkamp 2000). BMPs are known to function in conjunction with the Indian hedgehog (IHH)/Parathyroid Hormone-related Peptide (PTHrP) loop (Urist 1997), a negative feedback loop that limits cell maturation and chondrocyte proliferation throughout ontogeny (Lanske et al. 1996, Vortkamp et al. 1996). The growth rate within physes is regulated by BMPs and FGFs (Fibroblast Growth Factors): BMPs upregulate bone growth by increasing the size and rate of mitotically active regions of proliferative chondrocytes whereas FGFs have opposite effects on bone growth (Minina et al. 2001). Minina et al. (2002) showed that these two signaling factors act in an antagonistic, measured manner suggesting that the differential growth within and between physes is likely controlled by cis-regulation of BMPs and FGFs.

Mutations to some BMPs can block the formation of certain skeletal features indicating that BMPs are necessary for proper skeletal development (Kingsley et al. 1992; Storm et al. 1994; Luo et al. 1995; Thomas et al. 1996). Mutations in Bmp5, Bmp6 or Bmp7 in mouse have been shown to result in mild skeletal phenotypes (Dudley et al. 2002).
indicating that *Bmp* genes regulate bone development in a highly redundant way. Bone develops in response to the level of BMP present, suggesting that the regulation, and not the presence or absence of BMP, affects the shape and size of a skeletal feature (Volek-Smith and Urist 1996). Even very slight changes to the DNA sequence within regulatory regions, like those of BMPs and FGFs, can have substantial effects on the phenotype (Minina et al. 2001, 2002).

Many of the twenty BMPs currently known are unique in that they operate in extremely specific anatomical features, rather than in developing cartilage and bone in general (DiLeone et al. 1998). DiLeone et al. (1998) identified and described several *cis*-regulatory regions of BMP5 in a roughly 150 kilobase-long sequence located downstream of the BMP5 coding region in mouse. Their data suggest that the vertebrate skeleton is constructed as the aggregate of these independent domains, each one controlled by distinctive regulatory elements controlling expression in explicit locations (DiLeone et al. 1998). Following this study, it is likely that a *bmp5 cis*-regulatory region not only modulates cartilage and bone formation but that it does so in precise anatomical structures, such as ribs.

Using functional analysis in mice, Guenther et al. (2008) located the *bmp5 cis*-regulatory elements that function within very discrete portions of skeletal structure, specifically, ribs and nasal cartilages (Guenther et al. 2008). To create a curved skeletal feature, it is necessary to increase bone through differential growth rates on the structure’s opposite sides. Guenther et al. (2008) identified and described a restricted
domain along the lateral surface of the developing rib that affects the shape of the dorsal portion of the rib. The study demonstrates that the lateral perichondrium is controlled by a specific enhancer sequence: a 1069 bp core sequence at the 3’ end of exon 4 called \textit{bmp5} enhancer “A” (Guenther et al. 2008).

The \textit{bmp5} enhancer in mouse regulates highly detailed patterns of bone deposition: changes in rib length and the circumference correspond to independent regions of the enhancer. It has been suggested that somite resegmentation plays a role in establishing such growth patterns in developing mouse ribs (Guenther et al. 2008, figure 7). Resegmentation of somites helps establish the lineage domains that make it possible to independently control cartilage and bone growth in specific rib surface domains (Verbout 1976; Guenther et al. 2008). Guenther et al. (2008) demonstrated that the distinction in somite origin leads to lateral-proximal specific bone modifications. Huang et al. (2000) showed that the lateral edges of rib shafts are derived from cells in the anterior portion of somites. Therefore, the developmental division in the rib bone corresponds with differential growth in these regions. Guenther et al. (2008) determined that the multiple enhancers present in BMP genes provide a mechanism for linking lineage domains to actual sites of bone growth in ribs. For example, anterior somite transcription factors could serve as a mechanism for controlling \textit{BMP5} expression in the lateral perichondrium. Posterior somite expression could provide another a mechanism for regulating \textit{BMP5} expression in rib head and necks, and in the anterior, medial, and posterior perichondrial domains along the rib shaft.
During development, the rib diameter increases transversely more rapidly than sagittally. Human rib cross-sections are flatter mediolaterally than ape ribs (Aiello and Dean 1990). As human rib cross sections are unique, I hypothesize that a human-specific genetic sequence contributes to this phenotypic outcome. Alterations to BMP gene regulatory elements offer a means for evolutionary modifications to skeletal structures. Although null mutations in BMP genes often have pleiotropic defects (DiLeone 1998, 2000; Hahn 1996; King et al. 1994), changes to specific regulatory sequences would affect particular skeletal structures. It would be possible to adaptively modify vertebrate anatomy while preserving viability and fitness.

My initial screening of genomic databases demonstrated that the bmp5 enhancer “A” is found within primates including humans, chimpanzees, and rhesus macaque. Due to the level of conservation, it is likely that the region is necessary to the functionality of bmp5 in these primates in much the same way as is demonstrated in mouse. It would be useful to know the extent of sequence variation in this rib enhancer in primates as a first step to investigating the relationship between variations in enhancer sequences and spatiotemporal expression of bmp5.

In this study, I confirmed orthology of non-coding DNA sequences in primate taxa that were not previously publicly available in database: gibbon, squirrel monkey, and spider monkey. I aligned these sequences to orthologous sequences of human, chimpanzee, orangutan, rhesus monkey, and common marmoset. It was important to select a variety of primates with different locomotor patterns and relatedness according to commonly accepted phylogenies (See Appendix B for example). Mus musculus, the key
species in defining the importance of this enhancer in rib curvature (Guenther et al. 2008), was the outgroup of this study. As the apes (hominoids) are the most closely related living primates, *Pan, Pongo, Hylobates*, and *Homo bmp5* enhancer regions were sequenced. The Old World Monkey, *Macaca mulatta*, and the New World Monkeys, *Saimiri sciureus* and *Ateles paniscus* were also sequenced.

The *Ateles paniscus* sequence was of particular interest because the invagination of the postcervical spine has altered the thoracic cross-sectional form in atelines (McCollum et al. 2009). Additionally, spider monkey (*Ateles paniscus*) demonstrates lumbar lordosis of the vertebral column (Lovejoy et al. 2009a). Genetic conservation underlying the common vertebral column and thoracic cage morphology of *Ateles* and hominids, two distantly related primates, may have evolved using similar modifications. Molecular confirmation of similar *bmp5* enhancer regions among primates with similar thoracic anatomy would support the hypothesis that *bmp5* is a contributor to that particular thoracic cage development and patterning. Alternatively, spider monkey (*Ateles paniscus*) and hominids could have convergently acquired similar vertebral column anatomy through different evolutionary pathways. If the *bmp5* enhancer region is unique in each species, yet the thoracic cage morphology is similar, it is possible that the sequence plays a part in rib curvature patterning, but it is not a major contributor to phenotypic outcomes of the thoracic cage.

Experimental and bioinformatic approaches to investigating major questions in human evolution can be powerful and may pave the way for future functional studies to
directly link genetic and phenotypic evolution of primate rib curvature and overall thorax shape. This project provides information about the genetics of many primate species (especially in three previously unpublished gene sequences), the suspected role of the *bmp5* enhancer region in rib development in primates. Information presented here may shed light on our understanding of the evolution and the dynamic relationship between structure, function and regulatory genetics.

**HYPOTHESIS**

I hypothesize that rib development in mouse and human is directed by a conserved set of regulatory genes, but that lineage specific substitution in enhancer sequences accumulating in primate phylogeny and culminating in human-specific mutations may correspond to the dynamic changes in the hominid thoracic cage associated with bipedality. The variation within the *bmp5* regulatory elements, not the BMP5 protein itself, is much more likely to account for phenotypic change to primate skeletal features.

Enhancers, one class of functional non-coding DNA, specify transcription factor (TF) binding that initiates or silences genes in specific tissues and times. The *bmp5* enhancer that specifically influences rib curvature in mouse (Guenther et al. 2008) is located in the intron immediately downstream of *bmp5* exon 4 and drives *bmp5* expression in the perichondrium adjacent to the lateral aspect of the ribs. Firstly, I had to
confirm that the primate sequences I analyzed were in fact orthologous to the mouse 

*bmp5* enhancer using sequence analysis and comparison and phylogenetic reconstruction. I sought to trace the progression of nucleotide substitutions in this enhancer that accumulated in primate phylogeny. The fixation of lineage-specific mutations of enhancers can be a proximate mechanism underlying morphological divergence. It is possible that functionality in regulatory regions is the result of lineage specific duplication resulting in accelerated rates of evolution (Ohno 1970; Force et al. 1999). My comparative analysis of a defined *bmp5* enhancer in primates demonstrates that there are lineage-specific nucleotide sequences accumulating throughout primate phylogeny that alter TF binding profiles.
CHAPTER TWO: MATERIALS AND METHODS

The \textit{bmp5} enhancer examined in this study, “A”, is located in the intron immediately downstream of \textit{bmp5} exon 4 and drives \textit{bmp5} expression in the perichondrium adjacent to the lateral aspect of the ribs (See Figure 4). A second enhancer, “B”, is located 100 kb downstream of the \textit{bmp5} coding region and drives \textit{bmp5} expression in the perichondrium of the medial aspect of the ribs (Guenther et al. 2008).

\textbf{I. Polymerase Chain Reaction (PCR), Cloning, and Sequencing \textit{bmp5} Enhancer “A” in Primates}

In my first set of experiments, I sequenced and analyzed \textit{bmp5} enhancer “A” in primates as functionally characterized in Guenther et al. (2008). The authors did not provide the enhancer “A” sequence nor did they enter it into GenBank. To locate the mouse \textit{bmp5} enhancer “A” sequence in the mouse genome, I used the PCR primer sequences published in Guenther et al. 2008 to search the most recent mouse genome sequence (July 2007) using the UCSC genome browser (http://genome.ucsc.edu/). The mouse \textit{bmp5} enhancer “A” is on mouse chromosome 9 and is 1127 base pairs (bp) in length (Table 2; See Alignment: Appendix A).

I used the mouse \textit{bmp5} enhancer “A” sequence to extract the defined \textit{bmp5} enhancer “A” in publicly available genome sequences of human (\textit{Homo sapiens}), common chimpanzee (\textit{Pan troglodytes}), orangutan (\textit{Pongo pygmaeus}), rhesus macaque
Distinct enhancers control growth in very specific regions of developing bones. The straight line represents the *bmp5* locus. The filled circles represent enhancers that correspond to unique skeletal anatomy. The green and blue circles are those enhancers known to affect curvature of the rib in mouse. The green circle (enhancer “A”) is shown to be functionally necessary and sufficient for rib growth in mouse. “The growth on the lateral edge of rib surfaces promotes rib curvature and expansion of the thoracic cavity”, as demonstrated with the green color in the circle and in the lateral rib (Guenther et al. 2008 p8).
(Macaca mulatta), and common marmoset (Callithrix jacchus) (See Table 1 for genus-species and common names). To account for the possibility that the bmp5 enhancer “A” is longer in primates than in mouse, I extracted a sequence of approximately 1800 bp for each taxon that contained the best match to the mouse bmp5 enhancer “A” (See Figure 5). I aligned these sequences in Geneious™ and designed PCR primers in conserved regions in order to PCR amplify enhancer “A” in additional primates (Hylobates lar, Ateles paniscus, Saimiri sciureus) that are not represented in the public database (See Figure 5).

The Exon 4 forward primer, from the 5’ to 3’ direction, is as follows:

GAA AAT GAC AGA TCA TAT AGC

The Exon 4 reverse primer, from the 3’ to the 5’ direction is as follows:

TTC AAC ACT GAA TAA ACT CC
The nucleotide sequence examined for primates was approximately 1850 bp long. The confirmed mouse enhancer region is 1069 bp long. All primers depicted here were sequenced in areas of conservation via comparison to publicly available sequences. The nucleotide composition of each primer is available in Table 3.
TABLE 2 - List of *bmp5* enhancer region for species with mapped genomes

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>Chromosome</th>
<th>Length (bp)</th>
<th>Starting nucleotide position</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo sapiens</em></td>
<td>6</td>
<td>1829</td>
<td>55,637,783</td>
</tr>
<tr>
<td><em>Pan troglodytes</em></td>
<td>6</td>
<td>1829</td>
<td>57,034,811</td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>6</td>
<td>1830</td>
<td>56,337,198</td>
</tr>
<tr>
<td><em>Macaca mulatta</em></td>
<td>4</td>
<td>1829</td>
<td>55,412,917</td>
</tr>
<tr>
<td><em>Callithrix jacchus</em></td>
<td>9</td>
<td>1832</td>
<td>57,059,243</td>
</tr>
<tr>
<td><em>Mus musculus</em></td>
<td>9</td>
<td>1127</td>
<td>75,723,337</td>
</tr>
</tbody>
</table>
PCR products were amplified using the following reaction conditions:

**Step 1:** Denaturation: 95°C for 5 minutes

**Step 2:** Denaturation: 95°C for 1 minute

**Step 3:** Primer annealing: 50°C for 1 minute

**Step 4:** Extension: 72°C for 2 minutes

**Step 5:** Repeat steps 2-4 thirty times

**Step 6:** Final Extension: 72°C for 10 minutes

**Step 7:** 10°C indefinitely (until samples are removed from thermocycler)

The MgCl$_2$ concentration for each reaction was 3.5mM.

PCR products, including positive (human DNA) and negative (no DNA) controls were visualized using agarose gel-electrophoresis in 1X TAE buffer. The 1X TAE buffer was diluted from 50X TAE stock composed of 121g Tris base dissolved in 250mL distilled water mixed with 28.6mL of acetic acid and 50.0mL of 0.5M EDTA (pH of 8.0); distilled water was added until the solution totaled 500.0mL. PCR products of the correct size were purified using Qiagen MinElute™ Gel Extraction Kit. Step 13 of this protocol was amended to elute the DNA in 10 μL of sterile water. Ligation was done overnight at 16°C using the following 20 μL reaction: 10 μL Ligase Buffer, 8 μL purified PCR product, 1μL T4 Ligase, 1μL pGem-T vector. JM109 competent cells were thawed on ice for 5 minutes after which 2.0μL of ligation product was added. The JM109 competent
cells and ligation product rested for 20-30 minutes, were then heatshocked for 45 seconds in a 42°C water bath, and returned to ice for 2-3 minutes before shaking for one hour at 37°C with selective media. The JM109 cells with inserts were grown in incubator overnight on agar plates treated with Ampicillin and X-gal (modified β-galactose). Resulting colonies were subjected to blue-white selection. White colonies contained the appropriate genetic insert; the insert disrupts enzyme production for β-galactose digestion which results in white color. Blue colonies are able to digest β-galactose, therefore they do not contain the insert of interest. A minimum of ten white colonies per species per PCR fragment were screened for the appropriate insert size using whole colony PCR. White colonies were scraped from agar plates with pipette tips and placed into PCR reaction tubes for amplification under the PCR conditions initially used to amplify the inserts. Colony PCR products, including positive (human DNA) and negative (no DNA) controls, were visualized using agarose gel-electrophoresis in 1X TAE buffer.

Colonies with inserts of the correct size (see Table 2 for exact PCR product sizes) were grown in selective media and purified using the Qiagen QIAprep® Spin Miniprep Kit. Approximately 300 ng of each purified clone was then sent for sequencing (at Ohio University Genomics Facility) using the T7 and Sp6 primers, which have a binding site in the pGEM®-T vector. Through our sequencing procedures, it was possible to obtain between 500-700 accurate nucleotide readings per sample. I requested sequencing in two directions simultaneously, from 5’ to 3’ (using T7 primer) and 3’ to 5’ (using Sp6 primer). After initial sequencing, clonal sequences for each species were used to design new “sequencing primers” that would signal the beginning of the sequencing process at a
nucleotide further downstream than the original starting base pair (See Figure 2 and Table 3), thereby extending sequence reading beyond the original 500-700 bp. This process, called “walking”, was repeated for the positive and negative strands twice resulting in full sequences on both strands of the DNA.

I obtained complete sequences for six Ateles paniscus and Hylobates lar clones, and five Saimiri sciureus clones. The sequences for each clone were compared within species to determine a consensus species sequence. In cases of single nucleotide discrepancy, the most common nucleotide in that position among the clones was selected. Consensus sequences were included in a multi-species alignment (See Appendix A). I confirmed that the consensus sequences are correct and orthologous using BLAT (UCSC Genome Browser). An analysis that includes this bmp5 enhancer for Neanderthal would be extremely helpful in further describing the hominid specific changes, if any, to the regulatory region. While some portions of the Neanderthal genome have been made available to the public, they are not searchable nor are they known to be relevant to the thoracic cage. Therefore, Neanderthal was not included in this study.
## TABLE 3 – Table of Primers

<table>
<thead>
<tr>
<th></th>
<th>Forward Primer Sequence</th>
<th>Reverse Primer Sequence</th>
<th>Species/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCR Forward</strong></td>
<td>5’ GAA AAT GAC AGA TCA TAT AGC 3’</td>
<td>Conserved (present in all)</td>
<td></td>
</tr>
<tr>
<td><strong>PCR Reverse</strong></td>
<td>5’ TTC AAC ACT GAA TAA ACT CC 3’</td>
<td>Conserved (present in all)</td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer 1 Forward</td>
<td>5’ ATG ACT CTG GCA GAG G 3’</td>
<td><em>Hylobates lar</em></td>
<td></td>
</tr>
<tr>
<td>Primer 1 Reverse</td>
<td>5’ CCT TTC AAC ATA ACG G 3’</td>
<td><em>Hylobates lar</em></td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer 2 Forward</td>
<td>5’ GAA ATC ATC AAC GTC C 3’</td>
<td><em>Saimiri sciureus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5’ CTG AAA TCA TCA AAG TC 3’</td>
<td><em>Ateles paniscus</em></td>
<td></td>
</tr>
<tr>
<td>Primer 2 Reverse</td>
<td>5’ AAT CAG GCA AAC ATG C 3’</td>
<td><em>Saimiri sciureus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5’ TCC TTA AAG CTG TGT C 3’</td>
<td><em>Ateles paniscus</em></td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer α Forward</td>
<td>5’ TGT CTA TAA GAC TTA ACC 3’</td>
<td><em>Ateles paniscus, Saimiri sciureus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5’ AAG TTA CGT GAT TTC TGC 3’</td>
<td><em>Hylobates lar</em></td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer β Reverse</td>
<td>5’ GTC TTA TAG ACA CAA ACC 3’</td>
<td><em>Ateles paniscus, Saimiri sciureus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5’ CGG AGA CTT GAA ATG CC 3’</td>
<td><em>Hylobates lar</em></td>
<td></td>
</tr>
</tbody>
</table>
**Phylogenetic Analysis**

Phylogenetic analysis (see Appendix B) was used to confirm the orthology of the *bmp5* enhancer “A” in primates and mouse. *Bmp5* enhancer “A” gene trees were constructed using three Geneious™ tree building formats: Jukes-Cantor; Hasegawa, Kishino and Yano (HKY); Tamura-Nei; and Bayesian methods (Drummond et al. 2010) and PAUP* (Swofford 2003). In all cases, mouse was selected as the outgroup species and bootstrapping analysis was done for 2000 replications. The resulting summary phylogenetic tree is presented in Appendix B.

**MATCH™ Analysis**

In the next experiment, I screened the *bmp5* enhancer “A” sequence of each taxon (See Alignment Appendix A) for potential transcription factor binding sites using an algorithmic prediction program MATCH™ (http://www.gene-regulation.com/pub/regulation.com/pub/programs.html). The summary of the transcription factor binding sites and corresponding results are shown in Figure 6.
CHAPTER THREE: RESULTS

In this study I PCR amplified, cloned, and sequenced \textit{bmp5} enhancer “A” sequences for three primates \textit{Hylobates lar}, \textit{Saimiri sciureus}, and \textit{Ateles paniscus} and analyzed these in comparison to additional publicly available sequences from primates (\textit{Homo sapiens}, \textit{Pan troglodytes}, \textit{Pongo pygmaeus}, \textit{Macaca mulatta}, and \textit{Callithrix jacchus}) and mouse.

Phylogenetic Tree Analysis

The phylogenetic analysis confirms orthology of the \textit{bmp5} enhancer “A” sequence in all species of this study. Using the PAUP* Phylogenetic Tree Building program, I reconstructed a phylogeny of \textit{bmp5} enhancer “A” (See Appendix B). With heuristic search settings, the most parsimonious tree agrees with other currently accepted primate phylogenies (Goodman et al. 1998, See Appendix B). The Neighbor Joining tree constructed for this same enhancer region confirms the branching pattern of the parsimony tree. The Catarrhines (\textit{Homo sapiens}, \textit{Pan troglodytes}, \textit{Pongo pygmaeus}, \textit{Hylobates lar}, \textit{Macaca mulatta}) separate from the Platyrrhines (\textit{Callithrix jacchus}, \textit{Saimiri sciureus}, \textit{Ateles paniscus}). Human and chimp are shown to be sister taxa while orangutan (\textit{Pongo pygmaeus}) and gibbon (\textit{Hylobates lar}) are shown to be the next most closely related taxa, respectively. Rhesus macaque (\textit{Macaca mulatta}) is the most basal of
the Catarrhines. Marmoset (*Callithrix jacchus*) and squirrel monkey (*Saimiri sciureus*) were grouped together while spider monkey (*Ateles paniscus*) was placed just outside this subgroup.

The program was run with 2000 bootstrap replications and a 50% majority rule was enforced. By convention, 70% of bootstrapping replications is considered an acceptable and reliable value for defining a relationship. Under these conditions, the grouping for marmoset and squirrel monkey was only 51%. The grouping of squirrel monkey, spider monkey, and common marmoset (*Saimiri sciureus*, *Ateles paniscus*, and *Callithrix jacchus* respectively) was 86%. The Catarrhines were all placed more reliably into their positions on the tree with 100% frequency at all branching divisions with the exception of a 99% frequency of the Rhesus macaque and ape split. The homoplasy index, excluding uninformative characters, was very low: 0.2030, indicating that convergent changes to this enhancer region are unlikely. The retention index expresses the “fraction of apparent synapomorphy in the character that is retained as synapomorphy on the tree” (Farris, 1989:418) and can be written as:

\[ R = \frac{(g-s)}{(g-m)} \]

where “g” denotes the greatest amount of change that the character may require on any tree. The index ranges between 0 and 1, with 0 indicating maximum homoplasy possible and 1 indicating that all similarities in tip taxa are apomorphic (Farris, 1989:407). The retention index for this tree is 0.8373, which confirms that the similarity among tip taxa is not indicative of convergence, but instead homology. There is still roughly 27% of the
similarity among terminal taxa that is not explained through homology. The 27% of undescribed homology in the phylogenetic reconstruction is notable in light of known phenotypic characteristics and molecular relatedness among primates.

MATCH™ Analysis

Using MATCH™ (Goessling et al. 2001), a program that predicts transcription factor binding sites, I obtained a list of predicted transcription factor binding sites for each species. The entire ~1830 bp sequence of each taxon was submitted to the MATCH™ program. The mouse sequence begins roughly 150 bp from the start and ends roughly 500 bp from the end of this ~1830 bp region (See Figure 5 and Alignment Appendix A). Figure 6 illustrates my findings for the predicted transcription factor (TF) binding sites for the ~1830 base pair region of the bmp5 enhancer “A” in all taxa in this study.

It is interesting to note the conservation and variation among the predicted transcription factor binding sites of bmp5 enhancer “A”. Table 4 highlights a number of known functions for these transcription factors and proteins. The list is not exhaustive, but represents the functions that were well-documented in published literature. All of the transcription factor binding sites highlighted in this study were predicted at over 85% likelihood with the clear majority of predictions at over 95% likelihood.
All hominoids except gibbons share four predicted TF binding sites: AP-1, Dorsal, HLF, and CRP. Gibbons, which lack the latter, have an additional, unique predicted TF, Abaa (See Figure 6). Interestingly, orangutan exhibits only two of these four sites. *Pongo pygmaeus* differs from the other apes in its absence of FOXD3 and HNF-4 binding sites and the presence of an Elf-1 binding site. *Homo sapiens, Pan troglodytes, Pongo pygmaeus, and Ateles paniscus* share the combination of binding sites for CRP and AP-1. Remarkably, the CRP and AP-1 binding site combination is lacking in *Hylobates lar, Macaca mulatta, and Saimiri sciureus*. The TF binding sites in this *bmp5* enhancer region in *Ateles* are relatively conserved and similar to those of hominoids. TF site differences are more pronounced among New World (squirrel monkey, spider monkey, common marmoset) and Old World (rhesus macaque) primates.

The outgroup mouse shared transcription factor binding sites with the other primates of this study but did not maintain the same location or combination of TF binding sites. Mouse shares with these primates the binding sites for HNF-4 (*Hylobates lar, Pan troglodytes, Homo sapiens*); FOXD3 (*Homo sapiens, Pan troglodytes, Hylobates lar, Macaca mulatta, Callithrix jacchus*); and Elf-1 (*Pongo pygmaeus*). Mouse is unique in the predicted binding sites for Nkx2.5, HSF, and CDP-CR1.
Figure 6 – Predicted Transcription Factor Binding Sites

<table>
<thead>
<tr>
<th>Species</th>
<th>Transcription Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>A, C, D, H, N</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>A, C, D, N</td>
</tr>
<tr>
<td>Pongo pygmaeus</td>
<td>A, C, D, H, N, E</td>
</tr>
<tr>
<td>Hylobates lar</td>
<td>A, C, D, N</td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>A, C, D</td>
</tr>
<tr>
<td>Saimiri sciuereus</td>
<td>C, A, D</td>
</tr>
<tr>
<td>Ateles paniscus</td>
<td>M, A, C</td>
</tr>
<tr>
<td>Callicithis jacech</td>
<td>R, E, F</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>F, H, R</td>
</tr>
</tbody>
</table>

Key for Transcription Factors
A = Ap-1 (Activator Protein 1)  
AB = AbaA  
C = CRE-BP1 (Cyclic AMP Response Element – BP1)  
CD = CDP-CR1 (CCAAT Displacement Protein –CR1)  
CR = CRP (C-reactive Protein)  
D = Dorsal  
E = Evi-1 (Ecotropic Viral Integration Site 1 4)  
EL = ELF-1(E74-Like Factor 1)  
F = FOXD3 (Forkhead Box D3)  
H = HNF-1 (Hepatocyte Nuclear Factor 1)  
HL = HLF (Hepatic Leukemia Factor)  
HS = Heat Shock Factor  
M = Mat1-Mc (Mat1-Mc protein)  
N = Nxx2-5 (NKK Homeodomain 2-5)  
O = Oct-1 (Octamer -1)  
P = Pax-4 (Paired Homeodomain)  
R = Freac-4 (Human Forkhead 4)
**TABLE 4 - Known Function of Transcription Factors or Proteins**

<table>
<thead>
<tr>
<th>Transcription Factor/Protein</th>
<th>Function/Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbaA^^</td>
<td>In <em>Aspergillus</em>, establish regulatory feedback loops, continued developmental inducton (Andrianopoulos and Timberlake 1994)</td>
</tr>
<tr>
<td>CRE-BP1</td>
<td>Activate and regulate transcription of target genes (Coven et al. 1998) (Matsuda et al. 1991)</td>
</tr>
<tr>
<td>CDP-CR1*</td>
<td>Recognition motif in gene expression for motor and sensory neurons (Uemura et al. 2005)</td>
</tr>
<tr>
<td>CRP</td>
<td>Cell homeostasis and apoptosis (Szalai et al. 2005), Inflammatory and immune response (Black et al. 2004)</td>
</tr>
<tr>
<td>Evi-1*</td>
<td>Proliferation and maintenance of hematopoietic stem cells (Goyama and Kurokawa 2010)</td>
</tr>
<tr>
<td>Elf-1</td>
<td>Regulate differentiation in epidermal structures, cuticle morphogenesis (Calero-Nieto et al. 2010; Bray and Kafatos 1991)</td>
</tr>
<tr>
<td>FOXD3</td>
<td>Cell migration and differentiation early embryogenesis (Javidan and Schilling 2004) Regulates dorsal mesoderm development (Chang and Kessler 2010)</td>
</tr>
<tr>
<td>HNF-4</td>
<td>Hepatocyte differentiation and function (Watt et al. 2003). Ligand-dependent steroid family protein, activate transcription (Sladek et al. 1990)</td>
</tr>
<tr>
<td>HLF-1</td>
<td>Upregulate leukemia oncogenes (Look 1997). Liver differentiation and development. Regulation of HNF-1. (Cereghini 1996)</td>
</tr>
<tr>
<td>Mat1-MC**</td>
<td>Regulation and transcriptional activation of HMG-box and promotes DNA-binding Ste-11(Kjaerulff et al. 1997)</td>
</tr>
<tr>
<td>Nkx2-5*</td>
<td>Establish dorsal-ventral polarity in neural development (McMahon 2000). Development of cardiac musculature (Chen and Schwartz 1995)</td>
</tr>
<tr>
<td>Oct-1^</td>
<td>Promotes transcription of cell-specific immunoglobulin genes (Petryniak et al. 1990)</td>
</tr>
<tr>
<td>Pax4^</td>
<td>Differentiation of pancreatic endocrine cells, embryonic cell specification (Mansouri et al. 1999). Organogenesis, cell self-renewal, migration of precursor cells, resistance to apoptosis (Lang et al. 2007)</td>
</tr>
<tr>
<td>Freac-4^</td>
<td>Bends DNA to allow for preferential DNA-binding (Pierrou et al. 1994)</td>
</tr>
</tbody>
</table>

* transcription factor is predicted in mouse only
** transcription factor is predicted in *Ateles paniscus* only
^ transcription factor is predicted in *Callithrix jacchus* only
^^ transcription factor is predicted in *Hylobates lar* only
**Bone Morphogenetic Protein (BMP5)**

I analyzed the *coding* sequence for the functional protein BMP5 in the species publicly available through the UCSC Genome Browser: *Homo sapiens, Pan troglodytes, Pongo pygmaeus, Macaca mulatta, Callithrix jacchus*, and the outgroup mouse. The coding sequence is 497 amino acids long, including start and stop codons. Overall, the coding region among these primates is highly conserved. There is a nucleotide insertion in the *Pongo pygmaeus* sequence and *Macaca mulatta*, each unique, that results in a potential shift of the open reading frame which is likely the result of a sequencing error. The *Pongo pygmaeus* coding sequence also contains an additional segment of 37 amino acids between amino acids 403 to 440. These 37 amino acids are lacking in the other primates and mouse. It will be interesting to re-sequence the *bmp5* locus in these taxa, especially to compare and contrast *bmp5* expression and function in orangutan and other primates, and also to test if BMP5 protein is present.
CHAPTER FOUR: DISCUSSION

The *bmp5* enhancer “A” sequence in mouse (Figure 4) influences curvature of rib anlage (Guenther et al. 2008). Since rib curvature varies in primates with differing locomotor strategies, I wanted to investigate the molecular evolution of this known functional enhancer throughout primate phylogeny. Guenther et al. (2008) did not publish the *bmp5* enhancer “A” sequences (which in mouse is 1127 base pairs long). I searched GenBank with used PCR primer sequences specific for the mouse enhancer (Guenther et al. 2008) to locate the full sequence in mouse. I blasted this sequence against publicly available primate genome sequences to isolate the *bmp5* enhancer “A” sequence. I obtained the *bmp5* enhancer sequence in additional primates with unpublished genomes using PCR with primers that target conserved regions of the enhancer. Table 2 lists the enhancer sequences and their length, in base pairs. It is possible that *bmp5* enhancer “A” sequence(s) identified in primates are the result of a lineage specific duplication(s) (see Figure 3), which could potentially lead to novel or accelerated rates of nucleotide sequence evolution (Ohno, 1970; Force et al. 1998). Sequence similarity, proximity to the *bmp5* gene proper region (Figure 2, Figure 4), and phylogenetic analysis (Appendix B) confirm their orthology.

As illustrated in Appendix A, the *bmp5* enhancer “A” sequence is overall highly conserved in primates. I used MATCH™ analysis to identify predicted TF binding sites for each species (Figure 6, Table 4). The four predicted TF binding sites, yet not the entirety of the enhancer “A” sequence, are identical in human and chimpanzee. The
absence of one of the four TF binding sites in gibbon suggests that the combination of TFs may not affect thorax shape. Because the predicted TF binding sites in chimpanzee and human are the same and the thoraxes of each are quite different, the lineage-specific changes to TF binding profiles may have little to do with overall thoracic shape.

While the enhancer is conserved, there are other species-specific nucleotide changes to the enhancer sequence which could have an unknown function, such as influencing the binding stability of TFs (e.g. nucleotide neighbor effects). The pace and rate of accumulated random mutations can be estimated using the Neutral Theory of Evolution (Kimura 1983). Conversely, if a genetic sequence is conserved with high fidelity, it has likely been regularly “edited” by purifying selection. If the bmp5 enhancer is so conserved, it should contain fewer mutations than those expected by chance. I found that 87.1% of the total 1127 nucleotides found in the orthologous mouse and human enhancer sequences are conserved, meaning that 982 nucleotides are the same in mouse and human. These lineages are estimated to have diverged between 65 and 85 million years ago (Foote 1999). The additive evolutionary time for the primates examined in this study totals over 200 million years [e.g., Chimpanzee and human are estimated a share a recent common ancestor of approximately 7-8 million years; therefore, humans and chimpanzees have evolved away from each other for a total of 14-16 million years (Suwa et al. 2007)]. In the bmp5 enhancer presented here, a much lower than expected number of mutations under the Neutral Model has occurred indicating that its function is strongly conserved. The entire enhancer sequence, including the predicted
TF binding sites, and flanking and interleaving nucleotides, are likely to be under negative selection (Ka/Ks<1).

The findings presented here suggest that this bmp5 enhancer in and of itself does not play a critical role in rib curvature. The authors defined other areas having potential effects on rib curvature, but did not publish the sequence nor provide ways to locate it (Guenther et al. 2008). A region downstream of exon 7 of this bmp5 enhancer was suggested to affect rib curvature (Guenther et al. 2008); however, Guenther et al. were unable to specify the sequence driving changes to the ribs (Kingsley D, personal communication). Consequently, we could not investigate other regions that were identified as potentially relevant to rib curvature. It is also likely that there are functional bmp5 enhancers contributing to rib curvature, in the bmp5 gene and others, still as yet to be discovered.

While some of the underlying genetic sequences that pattern the ribs during embryonic development in mouse have been functionally defined (Guenther et al. 2008), the specific phenotypic effect of this bmp5 enhancer on primate rib curvature is not yet known. There are obstacles to connecting these two elements of rib curvature more definitively for primates. With more knowledge of enhancer evolution and regulatory sequence function and access to primate embryonic material, many more tests could be completed to accurately identify the connection between genetics and phenotypic outcomes of this bmp5 enhancer region.
Further studies for this region could be completed to yield more definitive results about TF binding. An Electrophoretic Mobility Shift Analysis (EMSA) would confirm that the predicted transcription factors bind to the enhancer region. In an EMSA procedure, the predicted TFs and their antibodies bind to the TF binding sites. With the attached compounds, these sequence segments do not travel as readily through a gel as a genomic sequence without bound compounds. Such testing would help to confirm the level of conservation and functionality of the enhancer region. The table of transcription factors (see Table 4) highlights known TF functions. At present, none of the TFs are directly linked to rib formation and development. An EMSA study would confirm the TF’s presence in rib curvature development and define a new function of the TF.

Additionally, I would like to test the potential function of the human, chimpanzee, rhesus macaque, and spider monkey orthologous sequences using transgenic mouse studies. It would be interesting to see if the combination of TFs affects the functionality of this region in vivo. If all functionality were recovered through replacement of human, chimp, rhesus macaque, and spider monkey (species having the AP-1 and CRP TF combination), but not through orangutan, gibbon, or marmoset, then the AP-1 and CRP TF cassette would be isolated as a contributor to rib curvature during development. Without confirming TF binding, a mouse transgenic study would be helpful in outlining the progression of rib development in mouse. To date, there is no seriated map of mouse rib development available.
It has long been recognized that differential deposition on the lateral surface of ribs (and corresponding resorption on the medial surface) must underlie the expansion and ultimate shape of ribs and thoracic cavities (Bateman 1954). However, prior to my study, few sequences have been identified that are capable of driving rib domain expression. I have identified a specific enhancer sequence, including potential transcription factors, within a domain of rib expression. With this information, it is now possible to functionally test the transcription factors and enhancer in subsequent experiments.
CHAPTER FIVE: CONCLUSIONS

Reorganization of the primate bauplan (e.g. changes to thoracic cage shape, scapular relocation, vertebral column invagination) has played an integral role in locomotor adaptations. A known, causal genetic and developmental link is lacking between skeletal reorganization and bipedal locomotion, a hallmark of hominid evolution. However, a genetic and developmental link has been demonstrated specific to rib curvature and potentially the shape of the thoracic cage in the functionally defined enhancer sequence of \( \text{bmp5} \) in mouse (Guenther et al. 2008). There are few examples of regulatory sequences affecting the phenotypic outcomes exclusively (Chiu et al. 2004; Gompel et al. 2009; Wittkopp et al. 2002).Remarkably, there is definitive evidence of a specific regulatory region causing phenotypic changes in rib curvature and thoracic cage shape in mouse anlagen and influences rib curvature (Guenther et al. 2008). The \( \text{bmp5} \) enhancer is known to be involved in rib development, a morphological feature that is a major correlate of primate locomotor pattern. Chiu et al. (1997), among only a few others, were able to demonstrate that modifications of a regulatory region have led to phenotypic alterations in primates that in turn can be used to trace phylogenetic relationships. Any similarities or dissimilarities to the orthologous \( \text{bmp5} \) enhancer region in primates would be exceptional in regulatory genetic studies because of its known function. As such, it was worthwhile to examine potential nucleotide changes in the sequence described by Guenther et al. 2008 and reconstruct a phylogeny to determine the evolution of the region in the context of known speciation events in primates.
Successfully tracing the enhancer’s evolutionary history makes it particularly valuable in understanding primate evolution.

I PCR amplified, cloned, and sequenced the functional defined bmp5 enhancer sequence (Guenther et al. 2008) in Hylobates lar, Ateles paniscus, and Saimiri sciureus and aligned these to corresponding sequences of Homo sapiens, Pan troglodytes, Pongo pygmaeus, Macaca mulatta, Callithrix jacchus, and mouse. The phylogenetic analysis confirms that the sequence I analyzed is the orthologous bmp5 enhancer identified in mouse (Guenther et al. 2008). Using MATCH™ (Goessling et al. 2001), a transcription factor binding site prediction program, and phylogenetic analysis, I traced the progression of putatively functional nucleotide substitutions in this bmp5 enhancer that potentially alter transcription factor binding sites throughout primate phylogeny including key taxa human, gibbon, and spider monkey. My analysis demonstrates that while there are species-specific nucleotide differences, human, gibbon, and chimpanzee sequences share a core of predicted binding sites that is different from that predicted in monkeys and mouse (outgroup). This study is not definitive in uniting the phenotypic and genetic differences among these species as further functional testing is required. Furthermore, the genetic changes within this enhancer region do not correspond to the known skeletal changes in thoracic cage morphology.

As described by Guenther et al. (2008) there are other regions of this particular enhancer that may contribute to rib curvature such as the area downstream of exon 7. However, it is extremely difficult to identify the specific functional sequence. It is possible that other enhancers contribute to rib curvature, perhaps even those not affiliated
with *bmp5* but with other proteins involved in the PTHrP loop. To fully understand the factors influencing rib curvature during development at the genetic level more information is necessary. One way to begin analyzing the key genetic elements affecting rib curvature would be to confirm the TF binding sites known to the area presented here using EMSA. Any species-specific TF binding combinations, such as AP-1 and CRP found in human, chimp, rhesus macaque, and spider monkey, could be functionally confirmed through transgenic studies. Also, more information is necessary to create a seriated map of rib development in model species. Such studies would greatly help in identifying other genetic regions that may potentially contribute to rib curvature.

Evo-devo studies focus on the genetic and developmental mechanisms that result in morphological changes. The subsequent comparative analysis and phylogenetic construction can reveal a great deal of information regarding a probable evolutionary pathway for species that are lacking in the fossil record. Evo-devo studies that focus on functional noncoding sequences are likely to contribute in a unique way to the study of hominid evolution for this reason. However, there are significant challenges to studying embryonic life of primates. The conclusions of this study can be used to develop future analyses of the specific operation of the coding and binding of regulators to defined regulatory regions and the phenotypic implications. This study demonstrates the power of experimental and bioinformatic approaches to fundamental issues in comparative evolutionary studies of primate and human evolution.
<table>
<thead>
<tr>
<th></th>
<th>M. musculus</th>
<th>Callithrix jacchus</th>
<th>Saimiri sciureus</th>
<th>A. paniscus</th>
<th>M. mulatta</th>
<th>Homo sapiens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M. musculus</th>
<th>Callithrix jacchus</th>
<th>Saimiri sciureus</th>
<th>A. paniscus</th>
<th>M. mulatta</th>
<th>Homo sapiens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B

Accepted Primate Phylogeny

- Mus musculus
- Callithrix jacchus
- Saimiri sciureus
- Ateles paniscus
- Homo sapiens
- Pan troglodytes
- Pongo pygmaeus
- Hylobates lar
- Macaca mulatta
Reconstructed Phylogeny using PAUP* (Swofford et. al 2003)
REFERENCES


