GENOTYPE-PHENOTYPE RELATIONSHIPS AND THE PATTERNING OF COMPLEX TRAITS AS EXEMPLIFIED IN THE MAMMALIAN DENTITION

A Dissertation in

Anthropology

by

Samuel Jacob Sholtis

© 2008 Samuel Jacob Sholtis

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

December, 2008
The dissertation of Samuel Jacob Sholtis was reviewed and approved* by the following:

Kenneth M. Weiss  
Evan Pugh Professor of Biological Anthropology and Genetics and Science, Technology and Society  
Dissertation Advisor  
Chair of Committee

Joan T. Richtsmeier  
Professor of Biological Anthropology

Alan Walker  
Evan Pugh Professor of Biological Anthropology and Biology

Cooduvalli S. Shashikant  
Associate Professor of Molecular and Developmental Biology

Kazuhiro Kawasaki  
Research Associate of Biological Anthropology  
Special Member

Nina G. Jablonski  
Professor and Head Department of Anthropology

*Signatures are on file in the Graduate School
ABSTRACT

The relationship between genes and traits is complex, involving the interaction among genes, environment, and chance. However, common practice in genetics treats this relationship as a straightforward one-to-one mapping from genotype to phenotype. The roots of this practice can be traced to Mendel who chose traits with a direct relationship between genetic variation and phenotypic variation in formulating his particulate theory of inheritance. It has been further solidified by the successes of modern genetics in identifying genes involved in many simple traits, such as rare human diseases. However, most traits are not simple and to understand complex traits it is necessary to decipher the developmental processes that occur between genes and traits. To date most attempts to build developmental models of traits have relied on natural or experimentally induced mutations that produce pathologic-scale phenotypic effects, but the raw material of most evolution is normal variation in traits found among members of a population. The goals of this dissertation, therefore, are to describe the cellular mechanisms and developmental processes that lead to a many-to-many genotype-phenotype relationship and to attempt to decipher the genetic and developmental processes that produce normal variation in traits using the mammalian dentition as a model system.

Two approaches to understanding the genotype-phenotype relationship are described and examples given of how both lead to a many-to-many relationship. First, cellular and genetic mechanisms, such as alternative splicing, DNA and chromatin modification, cellular gene choice, and gene regulation, which lead from DNA sequence to protein structure, are discussed. And, second, examples of variation in the genotype-phenotype relationship which can produce variable phenotypes from the same genetic information and stable phenotypes despite genetic variation are presented.
To examine how normal variation in complex repeated traits such as the mammalian dentition is produced two experimental approaches are taken. First, the AXB/BXA recombinant inbred mouse set is used to genetically map normal variation identified in cusp pattern between the teeth of C57BL/6J and A/J mice. Three chromosomal regions on mouse chromosomes 11, 13, and 19 reach the suggestive level of association with the trait. Within these regions candidate genes are identified by direct sequence comparison between the C57BL/6J and A/J mice. Two genes, Itga3 and Glis3, with non-synonymous substitutions and several genes with potential regulatory variation, including epiprofin/Sp6, may play a role in producing the variation identified. Additionally, identification of all genes in the mapping regions that are expressed in developing teeth using the GenePaint database suggests that traditional methods of identifying candidate genes are biased. In the second approach the expression of the signaling molecule Bmp4 was altered transgenically using the epithelium-specific enhancer of the Dlx2 gene. No overt phenotype was produced suggesting that tooth development is robust to over-expression of Bmp4. Examination of gene expression of downstream targets and antagonists of BMP4 suggests that a complex buffering mechanism is involved.

This dissertation highlights the complexities of the genotype-phenotype relationship and attempts to directly confront this complexity by developing methods to study normal variation in traits. The question of how normal variation is produced is complex and no simple answers are provided, but the importance of explaining complex common traits and the production of normal variation for biomedical purposes and our basic understanding of evolution means that we cannot hide from this complexity or pretend it does not exist.
TABLE OF CONTENTS

List of Figures..................................................................................................................viii

List of Tables......................................................................................................................x

Acknowledgments...........................................................................................................xi

Chapter 1: INTRODUCTION AND BACKGROUND.....................................................1
  Coming to grips with complexity..................................................................................3
  Reuniting development and evolution..........................................................................8
  Phenogenetics..............................................................................................................11
  Complex repeated traits and patterning processes......................................................12
  Teeth as a model for developmental patterning processes..........................................17
  Teeth and anthropology.............................................................................................19
  The anthropology of mice..........................................................................................22
  The development and patterning of the mammalian dentition.....................................25
  Summary.....................................................................................................................27
  Thesis organization.....................................................................................................28
  References..................................................................................................................29

Chapter 2: PHENOGENETICS: GENOTYPES, PHENOTYPES, AND VARIATION........36
  Introduction..................................................................................................................36
  Mechanism versus Variation......................................................................................40
  From genotype to phenotype: Mechanism...............................................................43
    A quick digression concerning DNA sequence: arbitrary and saturated.................43
    Pre-transcriptional mechanisms..............................................................................45
    Post-transcriptional mechanisms............................................................................49
  From genotype to phenotype: Variation....................................................................52
    A lexicographer’s nightmare: Canalization, Robustness, Plasticity, Polyphenism...53
    Developmental process: patterning repeated traits................................................57
    Gene regulation and the evolution of phenotypes......................................................61
    Phenogenetic drift: the role of chance in the evolution of genotype-phenotype relationship...........................................................................................................62
  Summary.....................................................................................................................64
  References..................................................................................................................66

Chapter 3: MAPPING NATURAL DENTAL VARIATION IN RECOMBINANT INBRED MICE.................................................................74
  Introduction..................................................................................................................74
  Materials and Methods..............................................................................................84
    Mice..........................................................................................................................84
    Inbred progenitor mice...........................................................................................84
<table>
<thead>
<tr>
<th>Chapter 4: TRANSGENIC ALTERATION OF BMP4 EXPRESSION IN THE DEVELOPING DENTITION OF THE MOUSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction .................................................................................................................................. 128</td>
</tr>
<tr>
<td>Transgenic alteration of dental patterning .................................................................................. 133</td>
</tr>
<tr>
<td>The Dlx2 epithelial enhancer ........................................................................................................ 135</td>
</tr>
<tr>
<td>Bmp4 in tooth development .......................................................................................................... 137</td>
</tr>
<tr>
<td>Downstream targets of BMP4 signaling .......................................................................................... 140</td>
</tr>
<tr>
<td>Pax9 .................................................................................................................................................. 140</td>
</tr>
<tr>
<td>Msx1 ................................................................................................................................................ 141</td>
</tr>
<tr>
<td>Msx2 ................................................................................................................................................ 141</td>
</tr>
<tr>
<td>Shh .................................................................................................................................................. 141</td>
</tr>
<tr>
<td>p21 ................................................................................................................................................... 142</td>
</tr>
<tr>
<td>BMP antagonists and tooth development ....................................................................................... 142</td>
</tr>
<tr>
<td>Follistatin ....................................................................................................................................... 143</td>
</tr>
<tr>
<td>Ectodin ........................................................................................................................................... 144</td>
</tr>
</tbody>
</table>

### Materials and Methods
- Mice ............................................................................................................................................... 145
- Reporter mice ................................................................................................................................. 145
Transgenic construct ................................................................. 146
Generation and genotyping of transgenic mice ...................... 147
Adult teeth .................................................................................. 149
Transgene expression ................................................................. 149
Histology .................................................................................. 150
In situ hybridization ................................................................. 150

Results .................................................................................... 152
Confirmation of Dlx2 enhancer activity .................................. 152
Production, genotyping, and phenotyping of Dlx2-Bmp4 mice ... 153
Transgene expression ................................................................. 155
Expression of downstream targets of BMP4 ......................... 157
Expression of BMP antagonists ............................................... 159
    Follistatin ........................................................................... 160
    Ectodin ............................................................................. 161
Summary ................................................................................. 161
References .............................................................................. 165

Chapter 5: CONCLUSIONS AND FUTURE DIRECTIONS .................. 172
Introduction .............................................................................. 172
Phenogenetics .......................................................................... 173
    Normal variation and dental patterning .............................. 174
    Recombinant inbred mapping ............................................. 174
    Over-expression of Bmp4 .................................................... 176
Phenogenetics: an update ......................................................... 177
    Mechanisms influencing the genotype-phenotype relationship 177
    Variation in the genotype-phenotype relationship ............... 180
Future directions ...................................................................... 182
    Recombinant inbred mapping ............................................. 183
    Dlx2-Bmp4 transgenics ....................................................... 185
Summary ................................................................................. 186
References .............................................................................. 188
LIST OF FIGURES

Figure 1.1: William Bateson and images from his *Material for the Study of Variation* showing variation in the number of repeated units..........................................................13

Figure 1.2: Patterning by combinatorial codes..............................................................15

Figure 1.3: Mammalian dental diversity. ........................................................................17

Figure 1.4: Variation in tooth row (A) and cusp (B) pattern in mammalian dentitions. ....18

Figure 1.5: Variation in tooth pattern among primates...................................................19

Figure 1.6: Upper dentition of the mouse. .................................................................23

Figure 1.7: Development of the dentition........................................................................26

Figure 2.1: Old (A) and new (B) views of the relationship between genotype and phenotype...37

Figure 2.2: Saturation: splice donor and acceptor sequence elements highlighted in a randomly selected sequence from the human genome.................................................................45

Figure 2.3: A basic reaction-diffusion model...............................................................60

Figure 3.1: Stages of tooth development.......................................................................75

Figure 3.2: Normal variation in teeth.........................................................................79

Figure 3.3: Breeding scheme for creating recombinant inbred (RI) mice by brother-sister mating........................................................................................................................................82

Figure 3.4: Photographs depicting the dentition of one C57BL/6J mouse.....................86

Figure 3.5: Four tooth variants identified between the progenitor strains of recombinant Inbred sets................................................................................................................92

Figure 3.6: 3D reconstruction of micro-CT scans of upper molars of C57BL/6J and A/J mice.........................................................................................................................93

Figure 3.7: Full genome interval mapping results for tooth trait in AXB/BXA recombinant inbred mice................................................................................................................95

Figure 3.8: Detail of mapping results for chromosomes 11, 13, and 19............................96

Figure 3.9: UCSC genome browser shot of Glis3 region on chromosome 19.................102
Figure 3.10: In situ hybridization showing Glis3 expression. .................................103
Figure 3.11: Strain distribution patterns of AXB/BXA RI mice showing shared chromosomal regions..............................................................104
Figure 3.12: Epiprofin/Sp6 knockout results. .......................................................108
Figure 3.13: UCSC Browser shot of epiprofin (Sp6) including 5 kb up and down stream. ......109
Figure 3.14: In situ hybridization showing expression of epiprofin/Sp6......................110
Figure 3.15: Tooth expression in GenePaint ...........................................................113
Figure 4.1: Similarities between feather and tooth patterning ..................................130
Figure 4.2: Computer simulation of tooth development ............................................131
Figure 4.3: Transgenic alteration of dental patterning .............................................134
Figure 4.4: Endogenous Dlx2 expression in developing tooth ..................................136
Figure 4.5: Expression of Dlx2-LacZ reporter construct during development ..........137
Figure 4.6: Endogenous expression of Bmp4 during tooth development ................138
Figure 4.7: Schematic showing relationship between interaction of Bmp4 and Fgf8 in the epithelium at ~E10.5 and future teeth ........................................139
Figure 4.8: Teeth of follistatin mutants .................................................................143
Figure 4.9: Range of phenotypic effects in teeth of ectodin knockout mice ..........144
Figure 4.10: Constructing the Dlx2-Bmp4 transgene ............................................147
Figure 4.11: Detail of transgenic construct showing restriction sites ......................148
Figure 4.12: X-gal staining showing β-galactosidase activity under control of the Dlx2 epithelial enhancer .........................................................152
Figure 4.13: Teeth of Dlx2-Bmp4 mice .................................................................154
Figure 4.14: Histological analysis of teeth ...............................................................154
Figure 4.15: Bmp4 expression .............................................................................156
Figure 4.16: Expression of genes downstream of Bmp4 .......................................158
Figure 4.17: Expression of BMP antagonists .......................................................160
LIST OF TABLES

Table 1.1: Human dental anomalies and mouse mutants. .................................................24
Table 3.1: Progenitor Inbred Strains. ...........................................................................84
Table 3.2: AXB/BXA Recombinant Inbred lines. .........................................................85
Table 3.3: Tooth Phenotype of AXB/BXA RI lines. ....................................................94
Table 3.4: Summary of Genes in 1 LOD interval of mapping peaks. .........................97
Table 3.5: Genes in 1 LOD intervals with link to tooth development. .........................99
Table 3.6: Genes with non-synonymous SNPs in 1 LOD intervals. .........................100
Table 3.7: Narrowed Chromosomal Intervals. .........................................................105
Table 3.8: High regulatory potential SNPs near tooth genes. .................................107
Table 3.9: Potential regulatory SNPs near epiprofin/Sp6 .......................................109
Table 3.10: Summary of GenePaint analysis. .........................................................112
Table 4.1: Dlx2-Bmp4 Transgenic Founder Mice. ...................................................153
ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my advisor Ken Weiss. When I returned to college after all but flunking out the first time he was willing to give me a job in his lab and let me do real work as an undergraduate. He also encouraged me to continue my education in graduate school, something I had not considered as a possibility. Ken has been a patient and encouraging mentor who seemed to know precisely when to push and when to let me muddle through. Most importantly he never failed to inspire me with his intellectual curiosity, something I hope I can emulate in my own career. I am also grateful to the other members of my dissertation committee: Joan Richtsmeier, Alan Walker, Cooduvalli Shashikant, and Kazuhiko Kawasaki. Each brought a unique perspective and generously gave of their time to help improve this dissertation. I owe a special debt to Kazz Kawasaki who taught me everything in the lab from how to hold a pipette to how to think about designing experiments. I would like to thank Anne Buchanan for her friendship and helping me struggle through many complex issues whether they were scientific or personal. I am also thankful to the members of the Weiss Lab, past and present who have provided assistance throughout my graduate and undergraduate career: Heather Lawson, Brian Metscher, Brian Lambert, Katie Schwabenbauer, and Julia Lee.

I made many great friends while at Penn State and the camaraderie of the graduate students in the anthropology department helped make many tough situations bearable. I would especially like to thank Jason De Leon, Eric Jones, Erick Rochette, Sharon Dewitte, Abigail Bigham, Kirk French, Laurel Pearson, Holly Dunsworth, and the members of the Wilcox Hotel for letting me in on all the fun.
I must also thank my parents, Joe and Maggie Sholtis, my siblings, Jason and Julie, my grandmothers, Geraldine Cunningham and Jean Sholtis, and my aunt LuAnn Reeves for belief in me and support that never waned as I struggled through many tough times.

Most importantly, I thank my wife, Jen, and my children, Jacob and Anja. More than anyone they suffered through the many difficult times and long hours that were required to complete this dissertation. Without their support and understanding this truly would not have been possible. It is impossible for me to express the amount of love, respect, and appreciation I have for them, and although I love my work the most important thing I do each day is go home to them.

Lastly, I am thankful to Dr. Timothy Ryan at Penn State for microCT scans of mouse teeth, Dr. Anthony Cox of the Wellcome Trust Sanger Center for A/J sequence data, Dr. Robert Williams at the University of Tennessee for AXB/BXA strain distribution patterns, Dr. John Rubenstein at the University of California San Francisco for the Dlx2 reporter construct, and Dr. Cheng-Ming Chuong at the University of South California for K14-Noggin mice. Chapter 2 first appeared in Variation: A central concept in biology, co-authored by Kenneth M. Weiss. We thank the editors Brian K. Hall and Benedikt Hallgrímsson for inviting the contribution.

This project was funded by the National Science Foundation (NSF award BCS-0343442) and the Penn State Research and Graduate Studies Office.
There is no fundamental difference in kind between the relation of genes to
differences in such structures as the webs of different species of spiders and the
differences in the arrangement their eyes.

Both are indirect.

Sewell Wright (1968) *Evolution and the Genetics of Populations, Volume 1: Genetic and Biometric Foundations*, University of Chicago Press, Chicago (p. 58)
Chapter 1

INTRODUCTION AND BACKGROUND

In the slightly over a century since the rediscovery of Gregor Mendel’s breeding experiments with pea plants (*Pisum sativum*) launched modern genetics, advances in understanding, technology, and the sheer quantity of comparative data (including whole genome sequences) have created a rigorous scientific field. Modern genetics is, of course, not a singular endeavor, but subsumes several subfields employing a wide variety of experimental approaches and theoretical perspectives that each attempt to explain the evolution of, and variation within, modern organisms. The notion that this explanation can be derived solely from genetic information is widespread among geneticists and the general public and has a long history but is false. As Richard Lewontin put it: “(i)f we had the complete DNA sequence of an organism and unlimited computational power, we could not compute the organism, because the organism does not compute itself from its genes” (Lewontin 2000 p. 17). Missing from a purely genetic explanation are the physical, environmental, and stochastic processes that together make up organismal development and lead, sometimes circuitously, from genes to traits.
Adding to the problem of explaining the evolution and development of traits is the traditional focus of genetic research on pathologic variation. Although unquestionably important on biomedical grounds, the study of pathologic variation gives of false picture of normal trait development and evolution. Pathologic traits are often the result of mutations of major effect and are generally rare and maladaptive. These traits therefore tend to be on the simple end of the gene to trait complexity spectrum, but more importantly they are not representative of the kinds of variation that play a major role in evolution.

The main goals of this work are therefore to shine light on the complexity of genotype-phenotype relationships in the construction of complex traits and to examine the genetic and developmental processes that produce normal variation. Three approaches are taken. The first is an analytical examination of the genetic and cellular mechanisms that lead from DNA sequence to protein structure and the ways variation can arise in the genotype-phenotype relationship (chapter 2). The next two chapters present examples of experimental methods to study normal variation in complex traits using the mammalian dentition as a model system. In chapter 3 natural variation identified between the teeth of C57BL/6J and A/J inbred mice is used to identify genes involved in the production of normal variation by genetically mapping the trait using the AXB/BXA recombinant inbred set. In chapter 4 the dynamic interaction model of tooth patterning, discussed below, is tested by transgenically altering the expression level of the bone morphogenetic protein 4 (Bmp4) gene, an important signaling molecule in tooth development, in as natural a way as possible. The dentition has played an important role in the development of biological anthropology, and variation in teeth is often used to make taxonomic inference, especially among the fossil remains of human ancestors. An additional motivation for the research presented is to inform interpretation of tooth variation in the fossil record by
attempting to uncover the type of genetic and developmental changes that bring about normal variation in teeth.

As will be seen the results of such research may not be as glamorous or definitive as finding a mutation responsible for a major developmental disorder, but this is an unavoidable consequence of attempting to confront the complex and subtle nature of minor natural variants in complex traits.

**COMING TO GRIPS WITH COMPLEXITY**

Evolution works, by chance or selection, through the differential proliferation of phenotypic variants and only indirectly on the genes that underlie them. This has been referred to as ‘evolution by phenotype’ by Weiss and Buchanan (Weiss and Buchanan 2003). As long as a trait is ‘good enough’ how it was built is unimportant for its success such that the same trait (or trait value) may have different genetic underpinnings in different individuals. So, in order to understand the evolution of a particular organism or trait we must grapple with the complexity that results from this indirect relationship between genes and traits. Only by building models of the developmental processes that produce a trait can we begin to see the different ways variation in the trait can occur and thus begin to understand its evolution. We must recognize, however, that the model is a stereotype and for any particular trait variant found in an individual a unique cause may be responsible.

Dissecting the connection between genes and traits is a major focus of modern genetics and one question traditionally asked in this regard is how variation at the level of DNA leads to
variation in traits? For much of the past century the answer was assumed to be irrelevant to evolution. It was believed by many that for each trait variant we should expect to find a corresponding genetic change, or ‘gene for’ that trait. Through historical happenstance the relationship between genes and traits was set up and treated as if it were one-to-one. But the production of a trait involves not only genes, but also their interactions with each other and the environment, and chance. The ‘gene for’ language so common in the scientific and popular genetics literature implies a level of genetic reductionism and determinism that belies the growing body of evidence to the contrary and veils complexity in the genotype-phenotype relationship that has been known for most of the century. If not actually admitting to believing it, or even fully recognizing the multigenic nature of most biological traits, geneticists have proceeded as if the relationship was straightforward with changes in traits being traceable to a single genetic cause which, in turn, lead to many false hopes for simple genetic answers to complex biomedical issues.

The historical rationalization for this goes back to Mendel himself who developed his particulate model of inheritance by choosing traits that behaved in a simple and predictable manner in crosses of purebred pea plants. These traits, such as seed coat color and texture, plant height and others bred true in the purebred lines (the offspring resembled the parents) and hybrids resembled one or the other parent, not an intermediate form. Without any knowledge of the genetic mechanisms underlying them Mendel had chosen those rare traits whose variation had a very direct relationship to genetic variation, at least in his pea plants. This brilliance allowed him to make his insights into the mechanism of inheritance but at the same time set up an expectation among future geneticists that there was a simple relationship between genes and traits for all traits. This view was solidified through the work of George Beadle and Edward
Tatum who interpreted the results of their experiments with irradiated bread mold (*Neurospora crassa*) as showing a one-to-one relationship between genes and enzymes (Beadle and Tatum 1941). Once Francis Crick and James Watson revealed the structure of DNA (Watson and Crick 1953) with the help of Rosalyn Franklin and the genetic code was worked out (Nirenberg *et al.* 1965) the one way flow of information from DNA to RNA to protein became dogma and it was tacitly and incorrectly extended to traits themselves. In fact, nearly all of these long-held ‘truths’ in biology have recently come under question. Whether or not modern geneticists take it seriously we now have direct evidence for a many-to-many genotype-phenotype relationship as was anticipated by such stalwarts of evolutionary biology as William Bateson (Bateson 1902), Ronald A. Fisher (Fisher 1918), Thomas. H. Morgan (Morgan 1917), and others, but ignored for most of the 20th century. Many specific examples of the evidence for a many-to-many relationship will be discussed in chapter 2.

Human diseases have been the most extensively studied in this light and research into their genetic basis is, for good reason, heavily funded. When diseases are simple enough (having high penetrance, large effect size, and straightforward inheritance) the mendelian model has produced success. The discovery of the genes responsible for many diseases (e.g. phenylketenuria, Tay Sachs, sickle-cell anemia, cystic fibrosis) has further strengthened the deterministic view of the gene-trait relationship. However, the usual textbook explanation offered for even these so-called simple traits masks a deeper level of complexity; often hundreds of alleles at a particular locus are known to affect phenotype and different mutations and different combinations of alleles produce varying levels of trait severity in an individual, not to mention the effects of secular trends in environmental variables (e.g. Scriven and Waters 1999; Weatherall 2001). These simple traits tend to be rare and debilitating and the genes ‘for’ them
only really inform us of one way development can go wrong. They tell us little of the normal processes that build traits or produce normal variation in them.

Most traits, including most human diseases, are not this simple and applying the same mendelian based model to complex traits has produced mixed results, at best. Complex traits are those whose development is influenced by many genes, often individually of small effect, and they often have a large environmental component. Also, unlike the traits chosen by Mendel and many of the diseases for which genes have been identified, complex traits tend to vary quantitatively in populations, to have low penetrance, and not to follow simple mendelian inheritance patterns in families. Until 1918 when R. A. Fisher (Fisher 1918) showed that for practical purposes continuous variation could be accounted for by the combinatorial effect of many mendelian genes there was much debate over whether continuously varying traits could be explained by mendelian inheritance. But knowing the genes could exist has not made finding them any easier. The genes affecting a continuous trait would individually be of small effect and therefore difficult to detect with traditional methods. But with large enough sample sizes similar methods can be attempted to approach complex, quantitative traits to those that were so successful in identifying genes for simple traits. Quantitative trait mapping and genome wide association studies identify chromosomal regions referred to as quantitative trait loci (QTLs) that are statistically associated with the trait. Usually there are several such associations, each on the order of megabases (Mb) in length containing the usual diversity of single nucleotide polymorphisms (SNPs), one to two thousand per Mb, and there has been little success identifying the specific mutations responsible in these studies.

An additional issue is that these studies are often not repeatable. The problem is manifold. First, for variation in many complex common traits such as heart disease, diabetes,
obesity, addiction, and intelligence any genetic component that we hope to discover may be completely swamped by environmental influences and these will surely differ among individuals in a population and between populations. In fact, in cases such as the examples just mentioned money spent to find genetic causes may be much better spent educating people to avoid environmental risk factors. But this requires the realization that the blame for the prevalence of such traits lies squarely on the shoulders of our own lifestyle and public policy choices, not on our genes (an easy scapegoat). The idea of a ‘magic bullet’ pill that allows us to avoid health issues while continuing to eat unhealthy diets, drink alcohol, or live sedentary lifestyles is difficult to combat.

Second, many complex traits are difficult to define. Examples include diseases like autism, attention deficit and hyperactivity disorder, and schizophrenia that have broad spectrums of effect size and syndromic effects that are variably present or absent in individuals. It is therefore difficult in many cases to know for sure whether two individuals diagnosed with a particular disorder really have the same trait. And finally, even if we could be sure that environmental factors and diagnoses are not at issue, because the production of complex traits involves many genes and developmental pathways, what appears to be the same trait at the phenotypic level can have many different genetic etiologies. Thus, to get a big enough sample size for case/control studies we are almost certainly grouping individuals with the same trait for different genetic reasons. This heterogeneity of causation in a sample will muddle the statistical signal and different samples will almost assuredly have a different mix of causative factors leading to the problems with repeatability. This final complication derives, at least partially, from a simple misplacement of the locus of selection. Selection acts on traits, not genes
themselves, so there is an inherent disconnect between genes and one of the major evolutionary forces that shape them.

Because of these complications the mendelian model is not appropriate for dissecting complex traits. Even if we could identify all the genes involved in the production of a complex trait it would not be a sufficient explanation of the trait; we need to understand the role those genes play in context including their interaction with other genes and proteins, the biochemical function of their coded protein, and the influence of the cellular and external environment. New approaches must be developed and the best will take seriously the complexity of the genotype-phenotype relationship by trying to build developmental models of how a trait is formed. This is a tall order, especially for disease traits that have late onset, but for many morphological traits the work of building models of development has been underway for close to a century, albeit mostly in a completely different research tract than evolutionary biology.

REUNITING DEVELOPMENT AND EVOLUTION

The study of development, an integral part of evolutionary biology at the beginning of the 20th century, was systematically eliminated from the study of evolution during the modern synthesis of the 1930s and 40s. Evolution became a problem purely of genetics, which at the time meant strictly inheritance. This made sense given the prevalence of a simple one-to-one view of the gene to trait relationship and was used to justify the reductionism that still dominates today. If there was a direct one-to-one relationship between genetic variation and trait variation there really would be no need to understand how a trait developed and thus the complexity of the
The genotype-phenotype relationship was ignored. The synthesis was built upon the results of early classical genetics most prominently in Thomas Morgan’s fly room at Columbia University and the brilliant work of Ronald Fisher, Sewell Wright and J. B.S. Haldane. Fisher, Wright, and Haldane were the main forces behind the development of the mathematical and statistical models that would become the theoretical basis for evolution and form what we now call population genetics. Evolution was redefined as change in allele frequency in populations over time and could be explained solely by the four forces of microevolution producing and influencing genetic variation: mutation, natural selection, gene flow, and genetic drift. Macroevolution, evolution at or above the level of the species, was simply the accumulation of the effect of these four microevolutionary forces.

The main currently active subfields of modern genetics have traditionally given little credence to development and the complexity of the genotype-phenotype relationship. Population genetics takes a mathematical or statistical approach to understanding genome evolution and changes in the frequency of DNA variants through time often with only a limited idea of how those changes influence phenotypes. Classical genetics, which is still practiced largely in the form of forward genetic screens, follows the inheritance of traits through crosses assuming a genetic basis if offspring resemble parents, the location of the genetic variant responsible can then be searched for using linkage analysis and positional cloning. Quantitative genetics has followed a similar path to classical genetics studying traits that vary continuously in populations. The result is usually abstract, producing heritability scores, the theoretical number of genes necessary to account for variation in a trait, or suggestions of pleiotropy based on similar mapping results for different traits, but studies often stop short of concrete causal relationships between actual DNA variants and trait variation. Developmental genetics focuses on
understanding the genes and networks involved in producing a trait, but the result is often a platonic stereotype of development not considering how variation in the process might affect trait variation or evolution.

Recently, however, there has been renewed interest in the role of development in evolution. The burgeoning field of evolutionary-developmental biology (evo-devo) takes as one of its main goals demonstrating the influence development has on the evolution of traits. Proponents of the evo-devo view rightly point out that evolution occurs through changes in the development of traits, which may or may not have changes in DNA as their root cause. The processes that produce traits occur during development and involve more than just genes.

All animals begin life as a fertilized egg, a single cell containing mitochondria and other organelles, and enough maternally derived RNA and proteins to kick start development and activate the egg’s own genome housed within its nucleus. As the egg begins to divide the daughter cells differentiate, establishing the main tissue layers, major body axes, and eventually all the different cell types and organs that make an adult organism. This differentiation from so simple a beginning has been one of the most vexing problems in biology because each cell contains essentially the same genetic information. Different cell-types and tissues are defined by the subset of that genetic information that is active. But how does the differentiation process begin? The key insight in resolving this conundrum came from fly genetics and was the realization that the egg is not a homogenous sack of protoplasm. The maternally-derived genes active in the fertilized egg are asymmetrically distributed such that at the first cell division each daughter cell receives a different complement of factors. Development continues as a hierarchical process and each adult tissue is the result of a history of asymmetric cell division
and differentiation. So, to explain variation in a trait and understand its evolution we must unearth the processes leading from genotype and phenotype buried in its developmental history.

A specific field for the study of the relationship between genotype and phenotype (phenogenetics) is essentially non-existent, but understanding this relationship would inform all other fields of genetics and is the focus of chapter 2.

**PHENOGENETICS**

With the realization of the many-to-many relationship between genes and traits the necessity for a more systematic effort to characterize the genotype-phenotype relationship has become apparent. Chapter 2 of this dissertation takes an analytical look at the ways that this relationship has been conceptualized, examining both the mechanisms that are responsible for producing traits from genetic, environmental, and developmental inputs and how variation in these processes influence variation in traits. This chapter was originally published in the volume *Variation: A Central Concept in Biology* edited by Benedikt Hallgrimsson and Brian K. Hall in 2005 (Hallgrimsson and Hall 2005). It catalogues many specific examples of the complexity of the genotype-phenotype relationship including such mechanisms as alternative splicing, RNA interference, and gene regulation, as well as ways in which variation in traits is influenced through developmental constraints or plasticity, patterning processes and phenogenetic drift. An update of this chapter is provided in chapter 5.

Complex, repeated or serially homologous traits are a special class of complex traits that have a long history in debates over the mechanism of inheritance and the tempo and mode of
evolution and are of particular interest to studies of genotype-phenotype relationships. An example is the mammalian dentition which will be discussed in great detail below, but first a general description of repeated traits, the developmental patterning processes that create them and why they are so interesting and important in evolutionary biology.

**COMPLEX REPEATED TRAITS AND PATTERNING PROCESSES**

Complex, repeated or serially homologous traits are characterized by an array of repeated units that each shares a common ancestor with all of the other units. Such traits are ubiquitous in nature and found at nearly all taxonomic levels. Examples include: leaves and flower parts in plants; segments, limbs and mouth parts in invertebrates; and vertebrae, limbs and teeth in vertebrates. They have been of interest in the study of evolution since before the turn of the last century and played a major role in the early 20th century debates between darwinians (biometricians) and mendelians over the ‘tempo and mode’ of evolution. In his 1894 book, *Material for the study of variation: treated with especial regard to discontinuity in the origin of species*, William Bateson, who coined the term genetics, documented numerous examples (e.g. figure 1.1), showing that these traits vary discontinuously and argued that they could not be explained by darwinian gradualism (Bateson 1894). Bateson showed that variation in these traits could be *meristic*, occurring through changes in the number of repeated units, or *homeotic*, occurring through change in the identity of an individual unit. Bateson’s book was written before the rediscovery of Mendel’s pea experiments so at the time he was unable to fit this discontinuous variation with a specific model of inheritance and he groped for an explanation
suggesting that variation in serially homologous traits was the result of the ‘intrinsic nature of organisms’ (p. 567). Although slightly mystical because Bateson did not have the knowledge of genetic and developmental patterning processes, this explanation anticipated more modern models that will be discussed below. But with the rediscovery of Mendel he became a staunch mendelian, even publishing a book entitled *Mendel’s principles of heredity: a defense* (Bateson 1902) and suggested that mendelian inheritance with its discrete factors of heredity could explain discontinuous variation.

Figure 1.1:  **William Bateson and images from his *Material for the Study of Variation* showing variation in the number of repeated units.** Bateson (A) studied discontinuous variation suggesting that examples such as an atelees monkey with a supernumerary premolar in the upper, but not lower jaw (B) and variation in vertebrae number in frogs (C) could not be explained by darwinian gradualism. Sources:  A: http://www.amphilsoc.org/library/images/genetics/bateson2.jpg,  B & C: (Bateson 1894)
As a general statement, serially homologous traits evolve through the duplication of the units sometimes followed by specialization or divergence (Weiss 1990; Weiss 1993; Weiss 2004; Sholtis and Weiss 2005). It is believed, and in many cases there is clear fossil evidence, that early in their evolution these traits consisted of an array of essentially identical units which differentiated morphologically and became specialized over evolutionary time. Duplication with specialization is one of the most predominant forms of evolutionary change found at nearly all levels of biological complexity from individual nucleotides, to codons, whole genes, chromosomes, cells, and entire organs. Understanding the genetic and developmental processes that pattern serially homologous traits is therefore essential to understanding this fundamental evolutionary process.

There are two main models for the developmental and genetic processes that pattern serially homologous traits. The first and most widely known can be referred to as a combinatorial code. The basic idea of such models is that there are overlapping patterns of expression of transcription factors (genes that influence the expression of other genes) in different regions or segments of a developing organism and it’s the complement of transcription factors active in a given region that determine its fate or identity. The most famous example of a combinatorial code patterning system is the hox or homeobox code involved in anterior/posterior patterning (figure 1.2A). It was first described in drosophila but has since been shown to be highly conserved among animals and although a bit more complex also patterns the anterior/posterior axis of mammals (e.g. Carroll et al. 2001). A similar code has been proposed for the patterning of tooth types along the mammalian jaw (figure 1.2B). Previous work in our lab has shown that the expression patterns of a group of homeobox transcription factors, the six mammalian Dlx (distal-less homeobox) genes, fit with a possible role in this model of dental
patterning (Zhao et al. 2000). This idea has been greatly expanded by Paul Sharpe and his colleagues at Guy’s Hospital in London incorporating the expression patterns of several other homeobox transcription factors into what they refer to as the “odontogenic homeobox code” (McCollum and Sharpe 2001a).

Figure 1.2: **Patterning by combinatorial codes.** The homeobox code for anterior-posterior patterning in drosophila (A, Carroll et al. 2001) relies on the overlapping expression patterns of Hox genes to establish segment identity and is conserved in mammals. A similar “odontogenic” homeobox code (B) has been hypothesized for the patterning of tooth types in the mammalian dentition (McCollum and Sharpe 2001b).

In the second model, wavelike patterns of activation and inhibition arise through the *dynamic interaction* of diffusible morphogens (e.g. signaling factor molecules). In the dynamic interaction model the process of quantitative interactions, rather than the specific genes present, specify the developing pattern. The dynamic interaction model gets its inspiration from the reaction-diffusion processes that Alan Turing (Turing 1952) described for the action of diffusing chemicals. In the most basic case you have two types of molecules, an *activator* which activates its own activity and that of the second molecule, an *inhibitor*. The interaction of these molecules
as they diffuse across space can produce regular wave-like patterns of activation and inhibition. The peaks of activation are thought to cross a threshold that initiates a downstream cascade of activity producing a trait such as a tooth or tooth cusp. An important aspect of these models is that slight changes in the parameters such as the diffusion rate, initial morphogen levels, or the space in which the reaction takes place can result in major changes in the outcome of the pattern. Although, these models are inspired by Turing processes it is important to point out that it is not necessary and probably unrealistic to literally apply the mathematical models to biological processes. The important aspect is the idea of a quantitative process that can produce an emergent pattern.

Reaction-diffusion-type models have been used to describe the patterning of a number of traits including animal pelage patterns (Meinhardt 1996), mollusk shell patterns (Meinhardt et al. 1995), fish striping (Kondo and Asai 1995; Kondo 2002), feather patterning (Jung et al. 1998; Jung et al. 1999), leaf shape and vascularization (Dengler and Kang 2001), and teeth. Jukka Jernvall’s group in Helsinki has produced copious examples, from fitting variation in cusp number in seal teeth to a reaction-diffusion-like model (Jernvall 2000), to actually simulating the formation of teeth in silico (Salazar-Ciudad and Jernvall 2002). Importantly, by slightly changing the parameters of their simulation model it can shift from producing mouse-like teeth with parallel cusps to the closely related vole-like teeth with offset cusps. They have since updated this model to take into consideration different ways variation can arise during development and how these changes will affect the eventual morphological pattern (Salazar-Ciudad et al. 2003; Salazar-Ciudad and Jernvall 2004). Most recently, they reported an activation/inhibition model for relative molar tooth size (Kavanagh et al. 2007).
TEETH AS A MODEL FOR DEVELOPMENTAL PATTERNING PROCESSES

Teeth have been extensively studied at both the gross morphological level and at the molecular level; every nook and cranny can be named by specialists and we have a good general understanding of the molecular processes involved in tooth development. Because of this the mammalian dentition is an excellent model to study the processes involved in producing variation in serially homologous traits. Mammalian teeth exemplify duplication and specialization coming in a wide array of shapes, sizes and numbers from the simplified teeth of the orca to our own fairly generalized dentition to extremely specialized structures like the incisor of the narwhal (figure 1.3A-C). All of this variety, in tooth number, shape, and arrangement is derived from the hypothesized ancestral placental mammalian dentition with a dental formula of 3 incisors, 1 canine, 4 premolars, and 3 molars (figure 1.3D).

Figure 1.3: Mammalian dental diversity: Orca (A), Human (B), Narwhal (C) show range of mammalian tooth form from simple to specialized derived from the hypothesized ancestral placental mammal dentition (D). Sources: D: redrawn from (Conroy 1990))
Variation in tooth patterning occurs at two levels. Teeth vary in both the number and organization of teeth along the tooth row and in the structure and arrangement of cusps within individual teeth (figure 1.4). Through decades of research biologists and dental anthropologists have documented innumerable examples of variants, from missing or extra teeth to slight cuspal variants (e.g. Gruneberg 1965; e.g. Colyer et al. 1988). Although by no means unique in this regard, teeth are therefore an excellent model for studying developmental patterning processes for serially homologous traits, but beyond this, teeth have played an essential role in the development of biological anthropology.

Figure 1.4: Variation in tooth row (A) and cusp (B) pattern in mammalian dentitions. Variation in the wave-like pattern of the dentition is apparent at both the level of the entire tooth row and cusp pattern of individual teeth. Sources: A: modified from (Hillson 1986), B: redrawn from (Thenius 1989).
The mammalian dentition is a historically important model in biology in general, but it has played a particularly key role in the history of biological anthropology. Before the advent of specialized imaging technology teeth were the only skeletal element available for direct study in living organisms. This is especially useful for studying humans who were, much to the chagrin of many an early anthropologist, not generally shot in the field and added to museum collections. But there are several more fundamental reasons that teeth have played such a prominent role in anthropology. We see similar variation in tooth number, size and morphology looking among primate species (figure 1.5) as we saw among mammals in general and this variation can be used to reveal aspects of animal behavior, relationships, and growth and development.

Figure 1.5: **Variation in tooth pattern among primate.** Like mammalian dental diversity the teeth of different primate species vary in number, size, and cusp arrangement. Modified from (Schwartz 1995; Swindler 2002).

Teeth have a direct functional, adaptive relationship to dietary specialization and by extension behavior (Kay 1975). Among primates there is a close correlation between diet and
group size and social organization. By analogy anthropologists can also use the morphology and microwear patterns of the teeth of extant primate species, where the diet is known, to infer diet in extinct species. Additionally, one of the major constituents of teeth is enamel. Consisting of calcium hydroxyapatite, enamel is the most densely mineralized animal tissue (Nanci 2003); it is resistant to taphonomic processes and therefore preserves preferentially in the fossil record (Wood 1981). Because of this set of facts, anthropologists frequently use comparisons of tooth morphology in establishing taxonomic relationships among living and extinct primates and human ancestors (Wood 1981; Beard et al. 1996; Flynn et al. 1999; Chaimanee et al. 2000; Dean 2000; Seiffert et al. 2003). Furthermore, the structure of the growing enamel leaves traces of daily growth intervals, longer period increments such as Striae of Retzius, and major life events (e.g. birth, disease) used in studies of growth and life history. And finally, forensic anthropologists and bioarchaeologists use teeth in the identification of and age estimation in skeletal remains.

For nearly a century dental anthropologists have studied and catalogued tooth variation among human populations (e.g. Campbell 1938; Moorrees 1951; Pedersen 1949; Dahlberg 1971; Brothwell 1963) and among fossil human ancestors (e.g. Robinson 1956; Weidenreich 1937). This variation can be quantitative, such as crown size or enamel thickness, or qualitative, including such traits as Carabelli’s trait, shovel-shaped incisors and the Y5/+4 molar groove pattern. A main goal of this work was to identify tooth characters that could be used to differentiate among major continental populations, and often this research was racially motivated. However, no tooth traits have been identified that perfectly predict continental origin. There are frequency differences among populations; Carabelli’s trait is more frequent in Europeans, shovel-shaped incisors in Asians and Native Americans, and like most traits this
variation is geographically patterned, but clear distinctions between ‘racial’ groups do not exist (Scott and Turner 1997).

In paleoanthropology it has been estimated that teeth make up between 70 and 75 percent of the total skeletal sample of hominid remains (Wood 1981). Inference regarding hominid relationships at sites such as Omo, Ethiopia, and Laetoli, Tanzania, are based almost exclusively on teeth (Wood 1981). To classify hominid fossil remains and establish taxonomic relationships among hominids, paleoanthropologists use dental characters such as tooth size, tooth shape, cusp arrangement and enamel thickness (Wood 1981; Dean 2000). Taxonomic decisions have been based on less than 2mm difference in tooth length (Wood 1981). Recent studies of dental variation in baboons (Hlusko 2004; Hlusko et al. 2004) and mice (Kangas et al. 2004) have demonstrated that many dental characters are nonindependent, questioning the validity of using individual traits for making phylogenetic inference. Therefore, inference based on fossil tooth morphology would be greatly enhanced by an understanding of the genetic and developmental processes involved in producing the kind of morphological variation found in the fossil record and the natural variation found among modern humans and primates (Weiss 1993; Dean 2000).

The tooth characters used by anthropologists are the direct result of patterning processes that take place during development. To date, genetic studies of variation in these characters have been restricted to attempts to characterize the inheritance of these traits within the mendelian model (Scott and Turner 1997). Although numerous defects of dental development (e.g. ectodermal dysplasia, amelogenesis imperfecta, hypodontia) have been at least partially characterized at the molecular level, none of the examples of normal tooth variation studied by dental anthropologists have succumbed to genetic analysis beyond hypothetical models of inheritance and heritability scores. In this way, the state of affairs in the genetics of normal
variation in teeth is similar to many other traits. Teeth are complex traits and studies of their genetic basis suffer all the problems described previously in studying complex traits in general. But because of the vast amount of work describing normal tooth variation and the genetic models of dental development that have emerged over the last several decades the dentition is one of a few traits where normal variation may be amenable to study. However, uncovering the processes that produce dental variation requires the manipulation of embryonic development and is for obvious reasons impossible in humans and at the very least undesirable in non-human primates. Thus, a good model organism is needed and the best available is the mouse.

THE ANTHROPOLOGY OF MICE

The use of model organisms in anthropology is not unprecedented; in fact with a growing understanding of the importance of development in evolution their use seems to be on the rise. Important anthropological work was done using non-primate animal models in the middle of the 20th century; Sherwood Washburn, for example, studied skull development using rats, rabbits and pigs and Albert A. Dahlberg experimented on tooth development in dogs. Later, because individual traits were thought to have direct relationships to individual genetic changes, traits were tacitly considered genetically independent and the relevance of non-primate models became unclear and the enthusiasm for using model systems faded. But, the use of animal models in anthropology has recently resurfaced in studies of limb development (Chiu and Hamrick 2002; Cohn et al. 2002; Lovejoy 2003), Down syndrome (Richtsmeier et al. 2000), cranial evolution and many others.
Most of what we know about tooth development comes from studies in the mouse. The mouse has a long history as a model organism. Mice are small, relatively inexpensive to keep and easy to care for. Generation time is short for a mammal and mouse husbandry is simple; because mice form a mucus vaginal plug following conception, timed embryonic stages can be obtained. Well-established lab protocols for manipulation of mouse embryos, hundreds of inbred strains, and vast genetic and genomic resources (including whole genome sequences) make the mouse the best available model system for the study of complex traits. There are limitations to using the mouse as a model for mammalian or primate dental development and evolution. The mouse has a highly derived dentition consisting of only one incisor and three molars separated by a toothless diastema in each tooth quadrant (figure 1.6), but we have several reasons to think that the mouse is a useful model for understanding dental patterning in general and in the primate dentition more specifically.

Figure 1.6: Upper dentition of the mouse: Although the dentition of the mouse is highly derived among mammals consisting of a single incisor and three molars separated by a toothless diastema, conserved genes and genetic pathways make it a valid model for studying primate dental variation.
Recent work in evolutionary-developmental biology has shown clearly why information gained in studying the mouse is directly relevant to questions about human or primate dental patterning. First, the genes that cause many human dental anomalies have been found to produce similar phenotypes when mutated in mouse (Table 1.1) (many reviewed in Thyagarajan *et al.* 2003). For example, the human ectodermal dysplasias result in lost or malformed teeth (among other symptoms) and are caused by mutations in the gene *edal* (ectodysplasin 1); this same gene is responsible for the *tabby* phenotype in mouse (Tucker *et al.* 2000; Mikkola and Thesleff 2003; Tucker *et al.* 2004). The teeth of Tabby mice are malformed including occasional missing molars and low, indistinct tooth cusps. Second, one of the great breakthroughs of the recent study of evolution and development is the recognition of the extent to which patterning genes and signaling pathways are conserved through evolution. For example, the ‘homeobox code’ for basic anterior-posterior patterning is highly conserved between human and flies (Carroll *et al.* 2001). Many of the genes and genetic pathways thus far discovered to be involved in tooth development are similarly conserved. The details vary and it is that variation that is likely to explain the nature and pattern of evolution among mammals, but the mouse is a valid model for dental development among mammals. The details of how the system has been adjusted in other species, including primates, can then be documented.

<table>
<thead>
<tr>
<th>Human Disorder</th>
<th>Gene</th>
<th>Mouse phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ectodermal Dysplasia</td>
<td>Eda1</td>
<td>Tabby (cusp patterning defects, missing teeth)</td>
</tr>
<tr>
<td>Oligodontia</td>
<td>Pax9</td>
<td>Tooth arrest at bud stage</td>
</tr>
<tr>
<td>Hypodontia, oligodontia</td>
<td>Msx1</td>
<td>Tooth arrest at bud stage</td>
</tr>
<tr>
<td>Amelogenesis Imperfecta</td>
<td>AmelX</td>
<td>Enamel hypoplasia</td>
</tr>
<tr>
<td>Dentinogenesis Imperfecta</td>
<td>Dspp</td>
<td>Defective dentin</td>
</tr>
</tbody>
</table>

*Table 1.1: Human dental anomalies and mouse mutants (Source: OMIM)*
THE DEVELOPMENT AND PATTERNING OF THE MAMMALIAN DENTITION

Teeth are ectodermally derived organs that arise in the maxillary and mandibular processes of the first branchial arch and fronto-nasal process (Peters and Balling 1999; Nanci 2003). Their development is similar to many other epithelial appendages (e.g. hair, feathers, and mammary glands) and involves a series of interactions between the oral epithelium and neural crest-derived ectomesenchyme. Through tissue recombination experiments (e.g. Kollar and Baird 1970; Kollar and Fisher 1980; Lemus 1995) and molecular genetic studies (reviewed in Peters and Balling 1999; Thesleff and Mikkola 2002; Tucker and Sharpe 2004) much of the process has been worked out. The first sign of tooth development is a thickening of the oral epithelium in the regions where teeth will develop and initially, control of development resides in the epithelium. Signaling molecules from the epithelium initiate a cascade of events that transfers control of tooth development to the mesenchyme, which in turn signals back to the epithelium which begins to invaginate forming the tooth bud. As morphogenesis continues a signaling center comprised of non-dividing epithelial cells known as the enamel knot forms which is believed to control continued epithelial growth and folding. Later, the enamel knot disappears through apoptosis and secondary enamel knots form at the future cusp tips. Eventually, epithelial cells differentiate into ameloblasts and secrete the proteins for enamel mineralization and mesenchymal cells form odontoblasts and secrete dentin matrix proteins. Once the tooth crown mineralizes it erupts into the oral cavity with further remodeling. The process is nested, using and reusing the same set of developmental genes to first establish the number and position of teeth along the jaw and then the cusp patterns within those teeth (figure 1.7).
Figure 1.7: Development of the dentition. Illustrations of tooth development showing many of the genes involved and their interactions between the epithelium and mesenchyme (A), and the hierarchical determination of tooth identity and form. Sources: A: redrawn from (Jernvall and Thesleff 2000), B: from (Weiss and Buchanan 2004) redrawn from (Jernvall and Thesleff 2000).

As previously described there are two basic models for the patterning processes involved in tooth development. The odontogenic homeobox code and the dynamic interaction model of tooth development. Although experimental results to date have not excluded either model and it is possible that both types of patterning are active at different stages of development, there seems
to be a growing body of evidence in favor of the dynamic interaction model. Unfortunately, most of the evidence comes from knock-outs and transgenic experiments that have produced major, pathologic-scale phenotype effects such as supernumerary or lost teeth, or extremely malformed cusps. Such studies demonstrate that a gene is involved in dental development and can be used to identify interactions among genes and pathways but they are extremely difficult to interpret in an evolutionary sense. Although, tooth loss and gain play a major role in the evolution of mammalian dentitions, most dental evolution occurs through minor natural variation. For example, old world monkeys, apes, and human all share the 2:1:2:3 dental formula, but there have been major changes in tooth size and shape among these species.

**SUMMARY**

Continued progress in understanding the evolution and development of complex traits will rely on facing the complexity of the genotype-phenotype relationship head on. The most promising approaches will choose traits, such as the dentition, that have a history of study in terms of natural variation, embryology, and molecular genetics in a model system, like the mouse, with sufficient genetic and experimental resources. Unfortunately there is likely to be a trade-off between studying pathologic-scale phenotypic effects which are easy to identify and relatively simple to interpret but tell us little about evolution and studying natural variation which will produce subtle, complex results that are difficult to interpret, but may have substantial explanatory power.
This thesis addresses the question of how the relationship between genotype and phenotype influences the evolution and development of traits and their variation, and attempts to uncover the genetic and developmental processes that produce normal variation in traits. Chapters 2, 3, and 4 contain the main substantive portions of this work based upon the background information provided in this introduction. The final 5th chapter presents the conclusions and future directions of this work. Chapter 2 is an analytical discussion of the many processes that affect the relationship between genes and traits which aims to temper the widespread practice in genetics of treating this relationship as one-to-one. Chapters 3 and 4 present two experimental approaches to understanding normal variation in traits using the mammalian dentition as a model system. The first approach uses forward genetics to attempt to uncover genetic variation associated with slight cuspal variation between the teeth of two inbred strains of mouse by genetic mapping in a set of recombinant inbred mice. The second approach uses reverse genetics to investigate the role of altering the expression pattern of the gene Bmp4 transgenically.
REFERENCES


Jung, H. S., L. Wolpert and D. Davidson (1999). "The formation of the feather pattern in chick skin after a proportion of cells have been killed by X-irradiation." Int J Dev Biol 43(2): 117-23.


INTRODUCTION

Mutations in the gene that codes for Phenylalanine Hydroxylase (PAH) can result in the recessive human disease Phenylketonuria (PKU, OMIM *261600). PAH converts the amino acid phenylalanine, obtained primarily through diet, into tyrosine. If, as in PKU, PAH does not function properly the resulting build up of excess phenylalanine can lead to severe mental retardation, and a suite of other symptoms. Many mutations are known and most individuals with PKU are heterozygotes, having a different mutation on each allele. Because nearly all cases of PKU are caused by mutations in PAH it has long been considered a classic example of a ‘simple’ genetic (or mendelian) disorder (but see Scriver and Waters 1999). However, early detection (infants are generally screened) and a diet low in phenylalanine, while reportedly unpleasant, can prevent or at least greatly reduce the severity of most symptoms. In other words,
even a PKU genotype does not lead to a PKU phenotype when the *environment*, in this case diet, is appropriately altered.

This classic example illustrates one of the most basic, and often overlooked, problems in biology: what factors determine the relationship between *genotype*, an organism’s or specific trait’s genetic makeup, and *phenotype*, the morphological, biochemical, or behavioral character of that organism or trait, including its entire developmental path (Lewontin 1992; Weiss 2003). Historically, a one-to-one relationship has dominated views of the genotype-phenotype

---

**Figure 2.1.** Old (A) and new (B) views of the relationship between genotype and phenotype.
relationship. Thus, much work has been devoted to the study of genes and traits, but little in between. But even the famous ‘one gene – one enzyme/protein’ hypothesis holds true in only the simplest cases. Alternative splicing, the formation of protein multimers, epigenetics, epistasis, and a host of other complicating factors, including the environment, all act to complicate this relationship (figure 2.1), or even the very definition of a ‘gene’.

It is crucial to understand that natural selection acts only indirectly on genes (Weiss and Buchanan 2003). There has been long-standing debate about the ‘unit of selection’ (e.g. Lloyd 1992). While many have postulated the gene as the locus of selection, it is the collection of phenotypes that make up entire organisms that confront selection, not just their genomes. This filtering through phenotypes, which are the result of the interaction between genotype, developmental process, and environment, can lead to variation in the genotype-phenotype relationship. From the point of view of natural selection, as long as an acceptable trait is achieved, the underlying genetic mechanism can vary. Hemoglobin mutations in malarial areas of the world provide a familiar example in which the phenotype of relative malarial resistance can be obtained through several different genetic mechanisms. In fact, in most complex traits in which multiple genes contribute to the phenotype (i.e. stature, blood pressure) there will almost certainly be different genotypes which produce equivalent phenotypes. Where this phenogenetic equivalence exists there can also be phenogenetic drift for basically the same reasons that we all accept genetic drift (see below; Weiss and Fullerton 2000). We can therefore expect, and selection experiments (Rutledge et al. 1974; Mackay 1995, 2001) and simulations (Yedid and Bell 2002) demonstrate, that there will be a many-to-many relationship between genotype and phenotype. Factors other than the genes themselves can be nearly as important, if not equally so. Can PKU really be considered solely, or even mainly, a genetic disorder if it can be treated
environmentally? More importantly, can characterizing diseases or traits as genetically determined influence the way we view them, preventing us from seeing environmental factors or solutions (Moore 2002)?

Ask a geneticist (or any biologist for that matter) about the role of phenogenetics, the relationship between genotype and phenotype, in producing variation during development and over evolutionary time and you’re sure to get the ‘right’ answer. Almost no one in the field maintains the position that the genome is solely responsible for every aspect and nuance of the phenotype of an individual. Nearly all acknowledge the role of epigenetic factors, the environment, gene-environment interactions and developmental process in the formation of phenotypes. Unfortunately, the extent to which this information is applied and the relative importance it is given is belied by the numerous reports of ‘genes for’ diseases or traits that appear on a regular basis in the news and scientific literature (e.g. Kaplan and Pigliucci 2001). These findings while not trivial often only apply to genes with strong effect discovered in isolated families with a particularly severe form of the disease or trait. The deterministic interpretations that these results are often given, the sensationalism with which they are presented, and attempts to generalize the results to all instances of the trait not only overstate the facts but also have the potential to mislead.

The importance of phenogenetic relationships has taken a backseat in biology since the modern synthesis of the 1930s and 40s and the resulting exclusion of development from studies of evolution, although scientists have argued for much longer than that over the relative importance of genes and traits in evolution (Gilbert et al. 1996; Gilbert 1998, 2003). The successes of population genetics, the theoretical underpinnings of the modern synthesis, further promoted a ‘gene’s eye’ view of evolution. Additionally, experimental methods (using inbred
plants, flies, or mice reared in homogeneous environments) attempt to remove nuisance factors that may fog the gene to phenotype correlation cementing the belief that phenotypes can be computed directly from genotypes, but does this accurately model reality (Schlichting and Pigliucci 1998; Lewontin 2000; Weiss and Fullerton 2000)? Even these methods cannot really isolate the effects of genes from the environment, they only demonstrate the effect of the genes in that one specific environment (Sultan 2003) and we know from numerous examples that genotypes do not always respond linearly to changing environments. This does not minimize the contributions of population genetics or the advances in genetics that have come since the modern synthesis, many of which have been truly groundbreaking. We owe our current advanced state of understanding in genetics to the 20th century’s somewhat deterministic view of nature.

Similarly, the use of inbred model organisms is indispensable in many areas of research. But, if the goal of science is to apply lessons learned from the laboratory to nature, the underlying phenogenetic complexities that have thus far been left largely unexplored must become the focus of study and results from experiments undertaken in these most extreme of unnatural conditions must be interpreted with care. Looking at evolution with a ‘trait’s eye’ view is an important first step. Here we present some sobering facts (and only a sampling at that) with the hopes of tempering, not discouraging research into the genotype-phenotype relationship.

MECHANISM VERSUS VARIATION

There are two perspectives with which geneticists have traditionally viewed the relationship between genotype and phenotype (Weiss 2003). The first view focuses on a causal
relationship between genes and traits, and can be characterized as mechanistic. Biologists who take this stance concentrate on the decipherment of the cellular and molecular processes that read and interpret a given DNA sequence for the production of proteins and therefore it is often presumed phenotypes. Traditionally, the mechanistic view of the phenogenetic relationship invokes an image of genetic determinism and such oft (mis)used analogies as the ‘genetic blueprint’, DNA as the ‘book of life’ or phenotypes ‘computed’ from genotypes (Kay 2000). But as will be shown, there can be a great deal of variation and complexity in the genotype-phenotype relationship even when mechanism is the focus of study. Alternative splicing and other mRNA editing, RNA interference, the formation of protein multimers, gene regulation and a host of other factors all act to complicate the relationship between genes and phenotypes. This relationship is rarely, if ever, one-to-one.

Genes, in the mechanical or physiological sense, are involved in the production of all traits but how or whether genes, and their variation, determine traits and the variation in those traits among individuals or populations is a separate question. The many steps between DNA sequence and trait (e.g. regulatory control of expression, mRNA processing, translation, interaction with other genes or proteins and the environment, and chance) add variation to the genotype-phenotype relationship. This variation is the focus of study of the second phenogenetic perspective. No two individuals share the same environment, development – the time and space over which genotype and environment interact to create phenotype – or for that matter the same genes (even monozygotic twins accumulate different somatic mutations from the time their individual cells begin to divide). Thus, the specific interactions of genes, development and environment that produce the phenotypes of an individual are unique and can introduce a great deal of variation. The existence of this variation in the genotype-phenotype relationship is
not surprising when one views evolution as acting at the level of the phenotype. Comparing the results of gene knockout experiments between strains of inbred animals (e.g. LeCouter et al. 1998) and the relationship between plants grown in various environments as famously shown by Clausen, Keck and Hiesey in the 1940s (for review see Nunez-Farfán and Schlichting 2001) provide beautiful examples. Strictly mendelian traits or diseases are rare, and when they are closely studied almost never turn out to be ‘simple’ (Scriver and Waters 1999). Most biologists know about variation in the expressivity and penetrance of diseases, but few have focused on the evolutionary implications of these and other similar factors, especially in non-disease related traits. Non-genetic inheritance (i.e. language, learned behaviors and often important environmental factors such as diet) also plays an important role in the understanding of variation due to phenogenetic interactions, and not just among humans.

This variation has a huge impact on our ability to map genotype to phenotype, or vice versa. For instance, discovering the ‘gene for’ a particular trait in one individual or family may or may not tell us how that trait develops in others, a fact borne out in the difficulty that is faced in repeating such studies (Long and Langley 1999; Goring et al. 2001; Ioannidis et al. 2001). In this chapter we will attempt to highlight the importance of phenogenetics in modern biology by exploring the complexity and variation that arises in the relationship between genes and phenotypes whether one focuses on mechanism or variation. We do not suggest that the majority of scientists are unaware of the importance of the phenogenetic relationship, but that despite their awareness business continues as usual and the primacy of genes in evolution and determining phenotypes is taken for granted.
FROM GENOTYPE TO PHENOTYPE: MECHANISM

We now know that the ‘one gene – one enzyme’ hypothesis first suggested by Beadle and Tatum (1941) through their work with the bread mold *Neurospora* greatly over simplifies reality. Many genes play multiple functional roles (pleiotropy) within and among taxa, and because selection works at the phenotypic level, the opposite, that multiple different genes can fulfill equivalent functions, is also true. The updated ‘one gene – one polypeptide’ comes closer to the truth by eliminating the idea of ‘one function’, but we will see several examples where even this may not be accurate. Understanding the many-to-many relationships that result from this complexity requires an appreciation of the mechanisms that relate gene structure to protein function. This entails more than transcription and translation as these topics are often taught in introductory biology courses. The gene-to-protein fidelity that is implied in this necessarily simplistic view is misleading, leaving out the importance of such factors as mRNA editing by splicing or ADAR (see below), somatic recombination, and homo- and hetero-multimers to protein function. Additionally, questions of gene regulation, where and when a particular gene will be expressed and produce a particular protein, are left unasked. Mechanisms that influence the gene to phenotype relationship can be classified into two general categories: those that act prior to and influence transcription itself, and those that work during and after transcription at the RNA or protein level.

A quick digression concerning DNA sequence: arbitrary and saturated

It is noteworthy to mention that our understanding of these processes do not come from DNA sequence alone. In fact, the sequence of As, Ts, Cs, and Gs that make up a genome would
be completely uninterpretable without the vast amount of experimentation that has demonstrated the function of specific DNA elements. There is nothing that we know, for instance, biochemically about the sequence of nucleotides ATG that suggests that it should indicate the start of translation or that it should code for the amino acid methionine, but it has been experimentally shown to do so. In other words, the genetic code is functionally arbitrary. This is also true of the genetic signaling and gene regulation that control cellular differentiation during development (Weiss 2002). Common signaling pathways (e.g. Wnt, notch/delta, TGF-β) and transcription factors while often named for the context in which they were discovered are used generically in many different situations during development and in many different tissues. Gene prediction algorithms that appear to find genes from sequence alone must use experimentally derived information such as the genetic code to look for open reading frames, splice donor and acceptor sites to find exon/intron borders, and regulatory elements like the TATA and CAAT boxes to locate the beginning of genes. An additional layer of complexity is added by the fact that the genome is saturated with such sites. In the extreme case, the two nucleotide canonical splice donor and acceptor sites (GT and AG respectively) would be expected to occur every 16 base pairs (bp) by chance alone (figure 2.2); in a genome with some 3 billion bp (approximately the size of the human genome) that would be more than 180 million sites each! Obviously just locating these sites alone tells us little about their potential function; the sequence context in which they are found must contain as much information as the sites themselves and must be taken into consideration as well. This saturation and functional arbitrariness makes understanding the mechanisms that lead from gene to functional protein even more difficult.
Pre-transcriptional mechanisms

The most universal of the pre-transcriptional mechanisms is gene regulation. Basically every cell in every tissue of an organism has an identical set of genetic information (ignoring mutations and some exceptions noted below); what distinguishes these cells and tissue types is the complement of genes active during their development (Freiman and Tjian 2003). Gene regulation is a hierarchical process that depends on the developmental history of a cell and its interaction with neighboring cells. The mechanism of regulation for any given gene requires several factors. The main components of this process are the actual machinery of transcription, usually RNA polymerase II; promoters, which facilitate the activity of the polymerase; cis-acting (on the same chromosome) enhancer, repressor and insulator modules (i.e. transcription factor binding sites); and the trans-acting transcription factors themselves, often used in concert, which bind to the DNA and influence the ability of RNA polymerase to do its job (Blackwood and
Kadonaga 1998). Other aspects involve sequence variation related to chromosomal packaging and manipulation in the cell.

The complexity of this system of regulation is best illustrated by the function of the *cis*-regulatory elements. Although regulatory modules are usually found within a few kilobase (kb) pairs 5’ of a gene, they are also found 3’, in introns, and even 100s of kb from the gene they regulate (Lettice et al. 2002; Lettice et al. 2003; Nobrega et al. 2003). These modules, of which there may be many for a particular gene used in different spatial or temporal contexts, contain the binding sites for several transcription factors that may work cooperatively or antagonistically to regulate the expression of a gene. The complement and concentration of transcription factors (TF) present in a cell and the number, location and specificity of binding sites in a gene’s enhancer modules establish its expression domain. But all of these factors are statistical rather than deterministic. For example, TFs recognize and bind to specific DNA sequences, but these recognition sequences can vary. This variation may influence the affinity of TF binding and therefore the expression levels of a gene (Dermitzakis and Clark 2002). Because of their short length (as few as 5 – 10 nucleotides) binding sites evolve quite differently than genes themselves. New genes generally arise through the duplication and divergence of existing genes; new TF binding sites however, can easily appear through mutation in raw sequence (Stone and Wray 2001). Binding sites can also mutate out of existence in the same manner and this volatility has an important role in phenotypic evolution (see below). In addition, binding sites of TFs may overlap and consequently compete with one another. In this case the relative concentration of TFs and the location and number of enhancer and repressor elements present will influence expression (Blackwood and Kadonaga 1998), which may also be affected by the way different enhancers are arranged in regulatory ‘cassettes’. 
There is also an element of chance involved in gene regulation. Individual TFs produced in the cytoplasm of a cell must find their way into the nucleus and locate their specific binding sites in a veritable sea of nucleic acids and proteins. By chance this doesn’t necessarily happen in the same way in every cell of a given tissue, even though we may say that the tissue is expressing the gene. What seems to be important is that ‘enough’ cells (whatever that means) in the tissue behave appropriately (Weiss and Buchanan 2004). Binding of TF to DNA alone is often not enough to drive expression, also necessary are the protein-protein interaction among several TFs and possibly other proteins. An example is found in Tricho-dento-osseous disease (TDO, OMIM #190320; Hart et al. 1997). The disease has been mapped to and mutations found in the homeobox containing transcription factor $Dlx3$. The mutations however do not affect the DNA binding domain of the gene, but instead truncate the gene removing a protein interaction domain. The truncated gene can still bind to DNA, but without binding to an as yet unidentified protein partner its function is impaired (Price et al. 1998; Dodds et al. 2003).

Also important in gene regulation are epigenetic factors, such as DNA methylation or histone acetylation, affecting the structure rather than the sequence of DNA. For example high levels of DNA methylation which generally affects the cytosine in CpG dinucleotides can make certain regions of the genome inaccessible to the transcriptional machinery. Interestingly, these non-sequence factors can be inherited both cell-to-cell within an individual’s lifetime and across generations. A well known example is genomic imprinting, in which methylation shuts off the allele of a gene an individual receives from one parent (either maternal or paternal). As many as 70 imprinted genes have been found in mammals, but estimates of the total number are much higher (Murphy and Jirtle 2003). A different epigenetic mechanism is used in X-inactivation to compensate for the increased gene dosage in XX females compared to XY males. Here, RNA
from the gene *Xist* physically covers all but the small pseudoautosomal region of one randomly chosen copy of the X in each cell (Plath et al. 2002). This RNA coat prevents genes on the inactive X from being expressed.

Phenotypic variation can also arise within an individual’s lifetime through somatic mutations that occur during each mitosis and are inherited in descendant mitotic lineages thereafter (as imprinting can be). These can, but need not be, pathological leading for example to cancers. In organisms without sequestered germlines like sponges that reproduce by sloughing off groups of cells, offspring genomes can be different from parental even without the usual mechanisms employed in sexual reproduction. In this regard, not all the cells in a sponge need even have an identical genome. In flowering plants, meristems develop individually so that even within a single plant different flowers may accumulate different mutations and thus have different genomes. If these mutations are beneficial and lead that meristem to survive or reproduce more effectively the ‘acquired character’ can be passed to the next generation and could eventually replace the original in a form of ‘Lamarckism’ that is also completely consistent with our current ideas of genetics and natural selection.

In other systems, several alternative genes, often but not always directly related as members of gene families, are used for a given function, but only one or a subset of these genes is used in a given context. The mammalian adaptive immune system provides an example of this kind of pre-transcriptional mechanism that doesn’t fit with the traditional view of the gene to protein relationship. Our immune system must produce a practically limitless amount of variation in order to cope with an unpredictable array of invaders (Weiss and Buchanan 2004). It does this in many ways, but of particular interest is the *somatic recombination* of the immunoglobulin genes in the production of antibodies. Antibodies are multimers composed of
two heavy and two light immunoglobulin chains. Each of these chains is comprised by several subunits: constant (C), joining (J) and variable (V) regions in the light chain, heavy chains have these plus an additional diversity (D) region. The genome contains multiple copies of each of these subunits, as many as 51 for the heavy chain V region (Nossal 2003). During lymphocyte development individual copies of each region are chosen, apparently at random, to be transcribed. Recombination brings the chosen regions together into a single transcription unit and the unused copies are removed from the genome of the cell. This recombination can produce an enormous variety of immunoglobulin chains, but additional variation is added by diversity in the junctions between subunits by including one or several additional nucleotides often shifting the downstream reading frame and the somatic hypermutability of these regions. A similar mechanism is used in the production of T-cell receptors, another component of the immune system.

This cellular gene choice, expression of a single gene, and often only one allele of that gene, to the exclusion of other closely related genes occurs in several other contexts. For example olfactory neurons usually only express one of the more than 1000 possible odorant receptor genes (Serizawa et al. 2003; Mombaerts 2004), cone cells express either a red or green opsin gene, but not both (Smallwood et al. 2002). In addition many other genes are expressed monoallelically, either through imprinting or other mechanisms (Knight 2004).

**Post-transcriptional mechanisms**

Getting transcribed at the right time, place and level is only the first step in the transition from gene to protein. Post-transcriptional mechanisms that influence the genotype-phenotype
relationship include, but are not limited to, alternative splicing, RNA interference, and the joining of individual polypeptides into functional proteins.

Most eukaryotic genes are organized as a series of exons (protein coding sequence) interrupted by intervening, non-coding sequence known as introns. Following transcription, splice donor and acceptor sites (mentioned above) signal cellular machinery to remove the introns forming the mature messenger RNA (mRNA). For many genes this splicing can be done in several different ways, removing or adding coding sequence, leading to altered protein structure and probably altered function. In humans, the prevalence of alternative splicing has been estimated to be as high as nearly 60% of known genes (Lander et al. 2001), but it is difficult to estimate how many of these splice variants, usually identified through ESTs (expressed sequence tags), are functional and how many are merely accidents of cellular processes (Sorek et al. 2004). For the case of alternative splicing we can update the classic phrase to ‘one gene – many polypeptides’.

Another form of mRNA editing known as ADAR (Adenosine Deaminase Acting on RNA) actively changes specific A nucleotides in mRNA to I (inosine), which acts as G during translation (Reenan 2001). At least in some cell types and for selected genes, this can result in amino acid changes. For example, the glutamate receptor gene GluR-B is edited nearly 100 percent of the time in neurons changing a glutamine codon to arginine. Reduction in the efficiency of this editing has been associated with human diseases such as ALS (Kawahara et al. 2004) and malignant gliomas (Maas et al. 2001). What’s curious, of course, is why such an elaborate process for editing would evolve rather than just changing the DNA sequence. However, whatever the reason, this system is deeply conserved and has neural function in
species from Drosophila to mammals. Presumably it allows spatial and temporal specificity of editing, so that in certain contexts we might see ‘one gene – a different polypeptide’.

Anti-sense RNA (RNA with sequence complimentary to a functional gene) has long been suspected of playing a role the in the post-transcriptional regulation of gene function (for review see Wagner and Flardh 2002), and the recent explosion of interest in RNA interference (RNAi) seems to validate that suspicion. In fact RNA is involved in many other cellular processes, including pre-transcriptional gene silencing, RNA editing (both splicing and ADAR), translation, and possibly even DNA editing through reverse transcription (Herbert 2004). Mattick (2003, 2004) suggests that as much as 98% of transcription in humans is non-protein coding (including introns). This means either that our cells produce a large amount of useless RNA or that this non-coding RNA plays an important functional role. Recent work has demonstrated the existence of RNAi in species ranging from bacteria to mammals, suggesting its early evolution and nearly universal application (Wagner and Flardh 2002; Cerutti 2003). The mechanism of RNAi is different in different taxa but the implication is the same: in some cells at some times genes get transcribed at levels inappropriate for that cell’s function. Although nary a day goes by without some new report of the action of RNAi we are still too early in our understanding of this process to know for sure the size of its role. But, if it is as prevalent as so far seems to be the case this means that knowing when and how much of a particular gene is being expressed in a cell may not mean knowing its importance. One gene – no protein?

Another important general factor to consider in understanding the genotype to phenotype mechanism is that genes only code for strings of amino acids; functional proteins typically are modified and/or require specific proteins to protect them, fold them into functional three-dimensional shape, or transport them in the cell. Many functional proteins are complex forms
consisting of multiple copies of the same (homo-multimers) or different (hetero-multimers) polypeptides (that is, coded by different genes). A functional hemoglobin molecule for example consists of two alpha cluster and two beta cluster chains and different members of the alpha and beta clusters are used at different points during development. Many genes – one protein.

In this mechanistic sense, because all traits rely on proteins, which with the above caveats are coded by the genome, all traits are genetic. But as we have tried to show even when one concentrates on this view there is no simple gene to protein relationship, and as we will show below, producing proteins is only a first step in the series of interactions that lead to phenotype. All of the factors we have seen so far work at the level of the cell, but in multicellular organisms these complications must be coordinated and choreographed as cells differentiate and like cells congregate into tissues. Amazingly, it almost always works. Below we will discuss the further complication of understanding how variation in the genotype leads to variation in phenotype.

FROM GENOTYPE TO PHENOTYPE: VARIATION

Thus far, we have focused on the mechanisms relating genes to phenotypes, but a host of other factors, genetic, developmental, environmental, and stochastic influence the timing, spatial organization and precision of these mechanisms, and therefore the genotype-phenotype relationship. In certain contexts these factors can lead to additional phenotypic variation from the same genes and in others to stable phenotypes and cryptic genetic variation. Also, there is a commonly held view that evolution takes place by incremental change in end-product genes, those with structural or enzymatic function for example. But these genes make up only a small
percentage of most genomes. The greater portion of genes is involved in gene regulation, inter- and intracellular signaling, and protein processing. These genes are used generically and interact during development so that end-product or structural genes can be deployed at the right time and place to do their usually narrow task. Change in these regulatory or processing genes, their interactions with other genes or proteins and the environment, or even more likely changes in the regulatory elements (e.g. TF binding sites) of genes, at any stage during development can alter phenotypes and may be more important in phenotypic evolution than changes in the structural genes themselves (e.g. Carroll et al. 2001). It cannot be stressed too greatly that genes only code for polypeptides which are several steps removed from what we usually call a trait. In this section we will discuss some of the ways that traits develop and through that how they can evolve. Important throughout this section is the labile relationship between genotype and phenotype. If we have learned anything in the century and a half since Darwin it’s that the only criterion of evolution is that if it works, it works. A prescriptive view of the phenogenetic relationship may find that the exceptions outnumber the rules.

A lexicographer’s nightmare: Canalization, Robustness, Plasticity, Polyphenism . . .

There is an unfortunately long list of sometimes ungainly terms that pop up in discussions of the relationship between genotype and phenotype. Their meanings overlap, but there are nuances and not everyone uses them in exactly the same way. Many of these concepts are discussed in more detail elsewhere in this volume (see e.g. Gibson, Stearns), so here we will try to reduce them to generalities and give a brief description of the impact that they have on our understanding of the relationship between genes and traits. All have to do with the interaction of the genotype with the environment (here broadly defined to include everything from genetic
background to external temperature) to produce phenotype. They can essentially be boiled down to two phenomena both of which undermine the traditional one-to-one view in phenogenetics:

(1) a given phenotype produced reliably despite slight variation in either the underlying genes or the environment (e.g., canalization, robustness, developmental stability, cryptic variation, buffering), or (2) different phenotypes produced by the same genotype in different environments (e.g., phenotypic plasticity, norms of reaction, polyphenism).

Waddington and Schmalhausen, two of the few biologists actually interested in connecting development and evolution during the height of influence of the modern synthesis, introduced the first phenomenon, what Waddington (1942) termed canalization. The impetus came from the fact that ‘wild type’ organisms appeared to be much less variable than ‘mutants’. Waddington suggested that this reduced variation resulted from selection; he argued that it would be advantageous for organisms to produce some optimal norm even in the face of minor genetic or environmental perturbation. This makes intuitive sense, but we also tend to think that canalization reduces evolvability, a potential liability. Recent work has shown that while this may be true in the short term, the hidden variation that results can have a large long term evolutionary impact. The evolution of discreet traits has been a problem in evolutionary theory since its inception, especially because these traits often have a complex genetic architecture so that single mutations are unlikely to produce the necessary change (but see the next section for a situation in which this could occur). Goldschmidt’s notion of ‘hopeful monsters’ has been universally panned but several lines of recent evidence suggest that saltational change may have played a role in evolution (Dietrich 2003). Probably most important for reviving at least some acceptability for the notion of sudden change was the discovery of the genetic basis of change in the numbers or arrangement of segmental or meristic traits. The classic instance was the finding
that Hox gene mutation could produce extra wings or leg-antennal substitutions in flies, examples of traits considered long ago in this context by William Bateson (Bateson 1894). Change in segment number can be brought about by mutation in a single gene, though the change is often not completely viable or competitively ‘fit’, it shows that small genetic change can lead to organized morphological change, and some of this resembles evolutionary changes in body plan (Carroll et al. 2001).

A basic idea is that the majority of changes necessary for a discrete shift in phenotype could be phenotypically silent. But at some point one more change in this newly created genetic background could generate a qualitatively different phenotype. Lauter and Doebley (2002) suggest that this could be the case in teosinte, the wild ancestor of maize, which harbors genetic variation for invariant traits that distinguish it from maize. This variation manifests by changing the genetic background through maize-teosinte hybrids (Lauter and Doebley 2002). This phenomenon has also recently been modeled using the secondary structure of RNA (the use of the term plastogenetic congruence in this work may take the jargon award!; Ancel and Fontana 2000; Fontana 2002).

One buffering mechanism that has received a lot of recent attention is the heat shock proteins (HSPs). HSPs act as protein chaperons in both normal conditions and under stress ensuring proper protein folding or degrading damaged proteins (Rutherford 2000, 2003). Experimental alteration of HSP expression level has demonstrated its role both by ameliorating the harmful effects of mutations when over-expressed (Fares et al. 2002) and by revealing hidden variation when reduced (Rutherford and Lindquist 1998; Queitsch et al. 2002).

Recently, Siegal and Bergman (2002) have suggested that canalization could evolve even without selection; it could simply be an inevitable result of the interconnectedness of complex
developmental systems. Regardless of how it has come about (for a recent discussion see de Visser et al. 2003) one result is that a certain amount of genetic variation has no impact on phenotype under so-called ‘normal’ conditions. To come full circle, Waddington, in a set of famous experiments (Waddington 1953, 1956) also demonstrated that if hidden genetic variation could be induced into expression by a large enough developmental shock it could be selected for and eventually expressed in the absence of the shock, a process he called genetic assimilation.

Since the end of the 19th century we have known examples of the opposite phenomenon, different phenotypes from the same genotype (Gilbert 2003). Despite this, it is often thought that even if we can’t predict genotype from phenotype, the converse should be relatively easy. And while it may sometimes be true, for many complex traits the importance of non-genetic factors such as the environment makes even this task daunting. This is relevant to ideas about the potential genetic specificity of natural selection, which also has to screen genetic effects indirectly via phenotypes. One early and famous experimental example mentioned above involved growing plants at different elevations, but there are also many natural examples: seasonal morphs in butterflies and caterpillars, predator induced polyphenism in some fish and other organisms, and environmental sex determination to name a few. Wing polyphenism in ants (Abouheif and Wray 2002) provides a particularly instructive example because it also demonstrates the instability of genotype-phenotype relationships. Environmental cues determine whether ants develop into winged queens or wingless non-reproductive castes. All extant ant species share this trait so it is believed to have evolved only once. However, the genetic mechanism with which wing development is interrupted varies both between and within ant species. Abouheif and Wray (2002) show that wing development is disrupted differently in the
two wingless castes (soldiers and workers) of the ant species *Pheidole morrisi* and even between the fore and hind wings of one caste.

Both of the phenomena discussed above have been known for the better part of a century. While interest grows, the importance of environmental contributions to phenotypes is still underplayed by a majority of biologists. Also, these are exactly the kinds of things that are minimized by our experimental methods. There has however been a call (Gilbert 2001; Sultan 2003) to shift biology out of the laboratory and back into the field. This ecological-developmental biology, as it is called, will be important if we are to understand evolution and development as they happen in nature rather than in the controlled conditions of the laboratory.

**Developmental process: patterning repeated traits**

Repeated structures abound in nature in both plants and animals. Look in the mirror; your hair, your teeth, your fingers and toes, your ribs and vertebrae all represent series of repeated elements often with some variation on the theme. Stripes on fish, pelage patterns in animals, segments in insects, and branch and leaf patterns in plants all provide additional examples. Is there a gene for each tooth, stripe or leaf? Almost certainly not. The interaction of many genes and other factors determine the development of these complex traits. Two complimentary models of developmental processes have been proposed to explain how these repeated, serially homologous traits develop and there is mounting experimental and computer simulated evidence to suggest that these processes are widespread. The first model relies on gradients of diffusing morphogens (e.g. signaling, transcription, or growth factors) that establish heterogeneous gene expression across an embryo that hierarchically sequester regions with different developmental fates. The identity of these regions may then be determined by a
combinatorial code such as the famous Hox code used in antero-posterior (AP) axis
determination in both invertebrates and vertebrates. The second model suggests that wave-like
patterns, such as those found in teeth and pelage patterns for example, can emerge through the
dynamic interaction of diffusing genetic or chemical factors. Again, combinatorial gene
expression is involved, this time involving different combinations in areas of structure growth
compared to surrounding inhibition zones.

Complex developmental and regulatory networks underlie most of the traits patterned in
this manner, and many others. Often these networks are used in several different contexts during
development in a single species (e.g. Hox genes also pattern limbs) so that presumably a given
network evolved in a single context and was subsequently co-opted as novel morphologies
evolved (Carroll et al. 2001). Amazingly, many of these networks are highly conserved across
deep phylogenetic distance, often having similar roles in vertebrates and invertebrates. In plants
similar patterning process are used but with a different set of genes, for example MADS box
rather than homeobox, even though plants do have homeobox genes. Meyerowitz (2002)
therefore suggests that the similar processes evolved independently in plants and animals
demonstrating how fundamental these processes are to development and evolution. While the
level of this conservation is sometimes surprisingly high, it fits with our evolutionary view of life
and the idea of common ancestry. This characteristic also forms the basis of our ability to use
experimental model organisms. In many respects a human is genetically a mouse, or even a fly.
But model organisms do not always behave as we expect; for example, knockout experiments of
what seem like important genes in mice often produce an unexpected or ‘no phenotype’
(Thyagarajan et al. 2003), and up to 50% of yeast genes when inactivated have been
characterized as nonessential (Thatcher et al. 1998). Of course it is unknown and possibly
unknowable in these cases whether changing the environment would induce effects (Tautz 2000).

Because the first model is familiar to most and reviews can be easily found (e.g. Carroll et al. 2001), especially concerning axis determination in the fruit fly, we will focus attention on the second. But one particularly interesting aspect of axis development in flies for our understanding of the genotype-phenotype relationship is that the initial conditions for segmentation are set before the activation of the zygotic genome. So-called ‘maternal effect’ genes form gradients of high to low concentration from both ends (e.g. bicoid – anterior; nanos and caudal – posterior) of the embryo. These gradients establish the heterogeneous environment that allows the developmental cascade of gap, pair-rule, segment polarity and finally Hox genes to sequester and successively define segments and their identities. Thus, the maternal genome plays an important role in determining the phenotypes of its offspring. This type of pattern formation has withstood the test of extensive experimentation and is universally accepted.

Less well known is the second model mentioned above which draws its inspiration from the chemical reaction-diffusion processes first described by Turing (1952). A basic reaction-diffusion process in essence entails the diffusion of two substances in an initially uniform field. One substance, called the *activator*, induces its own activity and that of the second substance, the *inhibitor* which reduces the activity of the activator. In such situations periodic wavelike patterns of activator level often emerge (figure 2.3). The amount of each substance, the level of their effect on one another, and their respective rates of diffusion determine the dynamics of the system. One can easily see how such a process could be at the root of periodically patterned morphological traits. For example, peaks of activator level could cross a threshold initiating a developmental cascade that eventually produces a morphological unit (e.g. tooth, stripe), while
valleys correspond to the spaces between units. Importantly the specific genes used in these processes are not as important as the dynamic properties of the system. In fact the same genes (often FGFs or BMPs) could be used generically in many patterned traits; the process specifies the pattern, not the trait. Computer simulations have shown that reaction-diffusion type processes can accurately model tooth patterns (Salazar-Ciudad and Jernvall 2002; Streelman et al. 2003), stripes on fish (Kondo 2002), mollusk shell pigmentation patterns (Meinhardt et al. 1995), feather patterns (Jung et al. 1998), vascular patterns in leaves (Dengler and Kang 2001) and many others.

Figure 2.3. **A basic reaction-diffusion model** shown at different sequential times creating a repetitive pattern. The horizontal axis represents a layer of otherwise similar cells. Cells produce an activator (A) that catalyzes its own continued production. As A diffuses across the tissue, cells are stimulated to produce A and a rapidly diffusing inhibitor (I). Where the concentration of A relative to I exceeds a threshold level gene switching occurs leading to activity peaks where units, like teeth or stripes, form.

One of the most interesting aspects of dynamic patterning processes is that slight tweaking of the parameters (e.g. diffusion rates, the space in which the reaction takes place) genetically or environmentally can lead to large qualitative changes in the pattern produced. This could possibly explain evolutionary changes in repeated discrete traits because the genes responsible for the actual construction of the trait would not be affected. Changes that lead to
additional peaks of activator could increase the number of units of a repeated trait because the activator simply recruits the existing developmental program. Jernvall (2000) has shown how variation in cusp number in the teeth of seals fits with just these types of changes. Similarly, so-called atavistic traits like horse’s toes could be the result of relatively simple changes in patterning dynamics, rather than requiring back mutations to undo millions of years of evolution (e.g. Weiss and Sholtis 2003).

These two models are almost certainly not mutually exclusive and combinations probably are important in the development of many traits. In teeth where the patterning mechanism is still not completely understood, plausible arguments have been made for both types of patterning processes (Thomas et al. 1997; Weiss et al. 1998). What is important for our understanding of phenogenetics is that in both of these processes many of the genes important in the evolution and development of repeated traits determine the pattern rather than the trait.

**Gene regulation and the evolution of phenotypes**

We described the mechanism of gene regulation above and discussed how easily new binding sites for transcription factors can evolve. These characteristics of the regulation of gene expression are extremely important in understanding the evolution of phenotypes. Because of the closeness of human and chimpanzee proteins (>98% identity, and we now know that our genomes have similar levels of identity) King and Wilson (1975) suggested early that our morphological differences may be explained by differential gene regulation, rather than changes in end-product genes. In 1975 this was a mere guess, but as our understanding of the processes of gene regulation has grown and examples have been identified, thanks largely to microarray technology, it has proven to be truly prophetic.
Differential gene expression has been demonstrated both within and among several species. Fay and Wu (1999) describe differential gene expression in nine yeast strains correlated with resistance to copper sulfate, an antimicrobial. Oleksiak and others (Oleksiak et al. 2002) show variation in gene expression in fish species of the genus *Fundulus*. Enard and colleagues (Enard et al. 2002) compared gene expression levels among several primate species, including humans, and found differences. This result received a lot of attention because expression differences were greater in human brain tissue than in liver or blood leukocytes when compared to other primates. But similar to the case of ‘genes for’ traits an ‘expression difference for’ interpretation of the evolution of humans is probably far too simplistic. Many other examples can be found (Wray 2003 reviews several) suggesting that differential gene regulation must play an important evolutionary role.

To complicate things further, we also see conserved expression patterns despite wholesale changes in regulatory machinery. For example, both the *cis*-regulatory elements and the transcription factors that control expression of the gene *Endo16* have changed in two species of sea urchin without changing its expression pattern (Romano and Wray 2003). Similarly, α- and β-globins which shared a common ancestor approximately 450 million years ago continue to be expressed coordinately with almost no conserved regulatory elements (Hardison 1998).

**Phenogenetic drift: the role of chance in the evolution of genotype-phenotype relationships**

At the crux of all the above mentioned examples of problems in phenogenetics is the labile relationship between genes and traits. Genetic drift is usually applied to DNA sequences to model the effects of finite population size on the change of allele frequencies over time, when the alleles in question do not affect darwinian fitness. The safest way to do this is to restrict
attention to areas of the genomes in question that are least likely to be affected by selection. This naturally leads to a focus on intergenic or intronic sequence, or third (‘wobble’) positions in codons and the like.

It is often tacitly assumed that drift does not apply to variation that has function or phenotype, but that is mistaken (a fact well-known in population genetics). The key fact would have to be that phenotypic variation associated with the genetic variation does not affect fitness. This is selectively neutral variation and might include such traits as stature, leaf shape, number of caudal vertebrae, details of facial morphology, and the like. Even these kinds of apparent details are of course speculative examples that might not even be affected by genetic variation, and it must be said that some biologists tend to insist that all phenotypic (or, indeed, genetic) variation is likely to affect fitness in some way. Allelic variation associated with such phenotypes can change stochastically as genetic drift, and there can be associated phenotypic drift in traits the alleles affect.

A different kind of stochastic change in phenogenetic relationships can take place even when there is natural selection (Weiss and Fullerton 2000; Weiss 2003; Weiss and Buchanan 2003). The reason is that selection acts on phenotypes, not genotypes, so that alternative genotypes associated with the same phenotype will experience genetic drift relative to each other. This can be referred to as phenogenetic drift and has two major applications. At any given time, what appears to be the same phenotype (or, equivalently if it can be known, the same at least with respect to fitness) can be caused by different genotypes. Mapping of quantitative traits, studies of disease or mutational effects and so on have identified this kind of allelic and locus heterogeneity. A practical implication is that even supposing the risky proposition to be true that the phenotype can be accurately predicted from knowledge of the genotype, the reverse
may not apply. The genotype may not necessarily be reliably predicted from knowledge of the phenotype. This kind of genotypic equivalence has been well-known since the earliest models of the genetic basis of quantitative or complex traits.

Another effect of phenogenetic drift can be experienced at the evolutionary macro-scale. Over time, selection can favor a trait, such as the expression of a given gene in a given place in an embryo or a particular morphology, while its underlying genetic basis changes. Comparison among taxa that presumably share the trait since they diverged from a common ancestor can reveal that its underlying basis differs (e.g. Abouheif and Wray 2002; Romano and Wray 2003).

SUMMARY

The relationship between genotype and phenotype is complex whether one focuses on the mechanisms that relate gene structure to protein function or on variation. Mechanisms such as gene regulation, alternative splicing, ADAR, somatic recombination, and the formation of protein multimers lead to a many-to-many relationship between genes and proteins. The environment can alter phenotype without changing genotype. Genetic mechanisms can lead to stable phenotypes despite changing genes or environment. Developmental processes often produce patterns rather than traits. And chance further complicates the situation. Much of modern biology is predicated on the assumption that by knowing the genotype of an individual we can predict the phenotype, or even vice versa, but the examples we show above demonstrate that in many situations this may not be possible. As our knowledge grows these complicating factors become harder to ignore and biologists need to take them into consideration whether they
are trying to decipher the genetic basis of disease or the evolution of species. This may seem a depressing fact (especially to the pharmaceutical industry) but what it in fact means is that we are getting closer in our understanding of genetics, development and evolution to how they have and continue to work in nature. As in any maturing science, the simple models that formed the foundation of our current knowledge must be reevaluated as new evidence is revealed. We must start by acknowledging that evolution works at the phenotypic level, not directly on genes, and that the resulting relationship between genotype and phenotype is labile and complex.
REFERENCES:


MAPPING NATURAL DENTAL VARIATION IN RECOMBINANT
INBRED MICE

INTRODUCTION

Variation in teeth among animal species was described as early as by Aristotle (Schour 1929) and dentists and surgeons such as John Hunter studied the human dentition extensively in the 18th century. However, Hunter devoted only a few short passages in his *The Natural History of the Human Teeth* (Hunter 1778) to the “Formation of Teeth in the Foetus” stating that:

The depression, or first rudiments of the *Alveoli* observable in a Foetus of three or four months, is filled with four or five little pulpy substances, which are not very distinct at this age. About the fifth month both the processes themselves and the pulpy substances become more distinct; the anterior of which are the most complete (p. 77).

Dental development was not described rigorously until histological studies in the late 19th century (Schour 1929). These studies established the basic tissues involved and stages of tooth
development: beginning with the initial thickening of the dental lamina, followed by bud, cap, and bell stages, and finally eruption of the tooth into the oral cavity (figure 3.1). Classic studies in the 1970s and 80s utilizing the tools of developmental biology, including tissue recombination, characterized how the two main tissues (oral epithelium and neural crest-derived ectomesenchyme) interacted with control of odontogenesis switching between the tissues as development progressed (Kollar and Baird 1970; Kollar and Fisher 1980; Lemus et al. 1986; Kollar and Mina 1991; Lemus 1995). Later, molecular and developmental genetic approaches such as tooth culture, in vitro manipulation of protein level using protein soaked acrylic beads, gene knockouts, and expression studies using in situ hybridization identified many of the genes involved and the interactions among them at various stages of tooth development (Vainio et al. 1993; Jernvall et al. 1994; Sharpe 1995; Thesleff et al. 1995; Chen et al. 1996; Aberg et al. 1997; Neubuser et al. 1997; Davideau et al. 1999). The expression patterns of nearly 300 genes have been characterized in developing teeth (http://bite-it.helsinki.fi).

Figure 3.1: Stages of tooth development: Tooth development has been studied since Aristotle but the basic stages of tooth development: thickening of oral epithelium, bud, cap, bell stages, and finally eruption, were not described until histological studies of the late 19th century. Oral epithelium shown in light pink, ectomesenchyme in purple (modified from Tucker and Sharpe 2004).
This work over the course of more than a century established the basic stereotypic model of tooth development as was described in chapter 1. Scientists have been developing models for the patterning processes responsible for dental development (e.g. the field model of Butler (Butler 1939) and the clone model of Osborn (Osborn 1978)) since the early part of the 20th century, but within the past decade studies have focused on building more specific gene-based models of the patterning processes and how variation in these processes affect tooth morphology and evolution. The two main models currently discussed in the literature are the “odontogenic homeobox code” (McCollum and Sharpe 2001a; McCollum and Sharpe 2001b) and the dynamic interaction models inspired by Turing’s reaction-diffusion processes (e.g. Weiss et al. 1998) described in chapter 1. Although it is possible that both models account for some aspect of dental patterning and they are not necessarily mutually exclusive, the weight of recent evidence seems to support a reaction-diffusion type model (Salazar-Ciudad and Jernvall 2002; Kassai et al. 2005; Cai et al. 2007; Kavanagh et al. 2007).

As a first step in building an understanding of dental patterning and evolution we need to identify the genes and the types of genetic changes involved in tooth variation. There are two main approaches for identifying a relationship between genetic and phenotypic variation. In the first, referred to as forward genetics, a phenotype is identified and geneticists work to identify the gene or genes involved through linkage and association mapping studies or positional cloning. The basic idea is that the traits will be associated, or inherited together with genetic markers that are physically linked to the causal allele on a chromosome. Once a statistical association is established between the trait and genetic markers the chromosomal region can be searched for candidate genes and causal variation. The second approach, reverse genetics targets a mutation in a particular gene (e.g. by knockout or transgenic over-expression) and identifies
and characterizes the resultant phenotype. The use of both approaches has successfully identified genes involved in tooth development.

The vast majority of work to date identifying genes involved in dental variation has relied on natural dental anomalies or induced mutations all of which have tended to involve pathologic scale phenotypic effects. This work is based in the traditional mendelian model and although it has added much to our knowledge of the developmental genetic processes involved in tooth development, because the methods used tend to produce frankly pathologic results it is difficult to interpret the findings in an evolutionary framework. Many of the genes thought to be most important in patterning the dentition are highly pleiotropic, involved in the development of many organs and often early axial patterning, and therefore knockout mice are frequently fatal before any useful information concerning tooth patterning can be gained (many reviewed in Thyagarajan et al. 2003). The majority of transgenic experiments investigating tooth development to date have used the keratin-14 (K14) enhancer to drive transgenic expression (e.g. Mustonen et al. 2003; Tucker et al. 2004; Wang et al. 2004; Plikus et al. 2005). K14 is a highly expressed gene (Coulombe et al. 1989) and transgenes driven by the K14 enhancer tend to produce major patterning disruptions of tooth development including loss or gain of teeth. Because it is expressed in all ectodermally derived epithelial tissues (Coulombe et al. 1989; Vassar et al. 1989) transgenes made with the K14 driver also affect more than just teeth. There are generally concomitant phenotypic effects to skin, hair, vibrissae, limbs, mammary glands, and external genitalia (e.g. Mustonen et al. 2003; Plikus et al. 2004); it is therefore also difficult to know if any tooth effects are direct or syndromic.

Although tooth gain and loss are prominent aspects of mammalian dental evolution at the level of the genus and above and in mutant mice tooth gain is often suggested to be indicative of
a reawakening of evolutionarily lost teeth (Peterkova et al. 2005) (rudiments of teeth that have been observed but eventually regress in the developing diastema have also been suggested to be remnant or vestigial lost teeth (Peterka et al. 2000; Viriot et al. 2000; Peterkova et al. 2002; Viriot et al. 2002)), most of tooth evolution is probably driven by minor natural dental variants within species. Among the nearly 150 catarrhine primate species (old world monkeys, apes and humans) for example there is no variation in tooth number, all share the 2:1:2:3 dental formula, but there is great variation in the size, shape and cusp patterns of teeth among these primate species (Swindler 2002).

Physical anthropologists have documented numerous examples of normal variation within and among human populations, including Carabelli’s Trait, shovel-shaped incisors, and the Y5/+4 molar cusp patterns (figure 3.2 A&B), but beyond showing that these traits are heritable we know essentially nothing about their genetic underpinnings. Normal variation in the teeth of non-human primates has also been documented (e.g. Gingerich and Schoeninger 1979; Colyer et al. 1988; Uchida 1998), although to a much lesser extent. To date there has been only one attempt in primates to uncover the genetic mechanisms involved in producing this variation. Leslea Hlusko for her dissertation at Penn State identified a minor cuspal variant (figure 3.2C) among a captive pedigree of baboons housed at the Southwest Foundation for Biomedical Research Primate Research Center in San Antonio, Texas. Using a quantitative genetic approach she showed that the variation is heritable and has done preliminary linkage mapping studies (Hlusko 2000). But getting beyond descriptive quantitative genetics measures and statistically suggestive associations between the trait and candidate chromosomal regions to actually identify the genetic variation responsible has thus far proved intractable. This is at least partially due to the limitations of baboons as model organisms. Genetic and genomic resources are not yet fully
developed in non-human primates such as baboons and dissecting the genetic architecture of these traits requires the ability to genetically manipulate embryonic development, something that is currently impossible and likely undesirable in baboons. But the problem goes deeper and relates to the basic issues of dissecting natural variation in complex traits.

**Figure 3.2: Normal variation in teeth:** Four levels of expression of shovel-shaped incisors (A) and Carabelli’s trait (B) and variation in expression of cusplet in baboons at SFBR (C). Although extensively catalogued, we still know almost nothing about the genetic and developmental origins of normal variation in teeth among humans and other primates. Source: A and B: (Scott and Turner 1997), C: modified from (Hlusko 2002a).

Despite the obvious evolutionary importance of understanding natural variation in teeth (and other traits) little experimental work has been done to try to explain the developmental and genetic origins of natural variation. This dearth of research on normal variation may seem surprising, but given the history of genetics using the methodologies of the mendelian model,
normal variation has been particularly hard to decipher. The subtle genetic changes that are likely to be responsible for normal variation are difficult to detect; generating enough statistical power using standard methods for example requires huge sample sizes. As was described in chapter 1 and 2, one of the major stumbling blocks in dissecting the genetics of complex traits is heterogeneity of causation, what on the surface appears to be the same trait may have different genetic or environmental causation. To work, genetic mapping methods must assume that the members of a sample with a trait of interest have that trait for the same genetic reasons. But for complex traits this assumption is often false and may be one of the major reasons for the lack of success when attempting to genetically map normal variation in complex traits in genetically heterogeneous or outbred populations like humans.

Many of the problems associated with studying normal variation in teeth can be, at least partially, ameliorated with the use of an appropriate animal model system and the best available is the mouse. Experimental methodologies for manipulating the mouse embryo are well established, there are vast genetic and genomic resources available, including whole genome sequences for several strains of mouse and thousands of genotyped polymorphic genetic markers, and there are hundreds of lines of inbred mice easily available through biological storehouses such as the Jackson Laboratories in Bar Harbor, Maine (www.jax.org). Mice also fulfill the usual requirements for model organisms of being small, easy to maintain in minimal space for minimal expense, having relatively short generation time (at least for mammals), and easy husbandry, including the ability to monitor and time conception.

Recent experimental work and comparative genomics has also demonstrated that the majority of genes and developmental pathways involved in tooth development are highly conserved between humans and mice, and often to much more distantly related organisms (e.g.
invertebrates). We also know that the genes responsible for several dental anomalies in humans produce similar phenotypes when mutated, either experimentally or naturally, in the mouse. For example, teeth in the classic mouse mutant Tabby are small, have low, disorganized cusps and are sometimes missing (Gruneberg 1965; Kristenova-Cermakova et al. 2002). The Tabby mutation has been traced to the gene \emph{ectodysplasin-1 (Eda1)}, a gene that is associated with ectodermal dysplasia in humans (Laurikkala et al. 2001; Mikkola and Thesleff 2003). None of this is surprising given the known evolutionary relationship between mice and other mammals, like us, and the fact that teeth have been a consistent feature of vertebrates since before the split between cartilaginous and bony fishes around 450 million years ago (Reif 1982; Kawasaki and Weiss 2006). One shortcoming of mice as a model for mammalian, and more specifically primate or human dental development and patterning is that mice have a highly derived dental formula consisting of a single incisor and three molars in each tooth quadrant. As with any model making analogies to other systems or organisms must be done cautiously, but to understand the evolution of traits of interest to anthropologists in humans and their relatives we must understand their development and the mouse is an appropriate model in which to do so.

The use of inbred strains of mice and the crosses derived from them allows for the control of many of the problems associated with genetic mapping in outbred populations. Inbred mice are created by brother-sister mating for at least 20 generations after which point mice are homozygous at approximately 98.7% of genetic loci (Lyon and Searle 1990; Silver 1995). Most commonly used inbred strains have been inbred in excess of 60 generations making them effectively homozygous for the entire genome (Silver 1995). The use of crosses among inbred mice in mapping studies greatly reduces the amount of genetic variation that must be screened and traits identified among the offspring are likely to share a single genetic cause.
One particularly useful resource in this regard is recombinant inbred (RI) sets of mice. Recombinant inbred mice are created by inbreeding the hybrid offspring of two inbred progenitor strains of mice through brother-sister mating (figure 3.3) (Taylor 1978; Broman 2005). The process is repeated many times producing a large number of RI lines each with an independently derived approximately 50% of the genomes of the two progenitor strains (Taylor 1978). Traits that differ among the progenitor strains can be scored in the RI lines and mapped using standard procedures. An additional advantage of mapping using RI sets is that no large scale breeding program is necessary and the available strains have been genotyped at literally thousands of genome-spanning polymorphic genetic markers (Wellcome-CTC Mouse Strain SNP Genotype Set http://www.well.ox.ac.uk/mouse/INBREDS).

Figure 3.3: Breeding scheme for creating recombinant inbred (RI) mice by brother-sister mating (redrawn from Broman 2005). By inbreeding the hybrid offspring of two progenitor strains of inbred mice sets of RI mice are created each with a unique combination of progenitor strain genomes. Because of their genomic structure these RI lines are useful tools for mapping complex traits.
For mapping with RI sets the main work lies in identifying and scoring a mappable trait that varies among the progenitor strains. This is not a trivial step. Choosing an appropriate trait is the key to successful mapping, but there are no standard set of a priori rules for doing so. The trait must be genetic and simple enough that a small number of alleles with large enough effect can be identified given the sample size and power of the mapping algorithms. As a start, it is useful to go back to Mendel and identify traits that vary discretely between progenitor strains of mouse. Mapping studies of normal variation in teeth to date have been limited to quantitative measures of tooth size and shape (Workman et al. 2002; Shimizu et al. 2004). These studies have identified up to 18 quantitative trait loci (QTLs) associated with their measures, but have not been able to go beyond this to identify candidate genes or causal variation.

The goal of this work is to identify tooth variation among the progenitor strains of RI sets and use that variation to identify candidate genes and genetic variation involved in normal tooth variation by linkage mapping. We sought to identify traits that varied discretely between strains of mice in the hope that they would be simple enough to map. Discrete natural variation in teeth among inbred strains of mice has been known since the 1960s (Gruneberg 1965) but to date there has been no attempt to discover its genetic or developmental architecture. This type of variation also mimics the kinds of variants (e.g. Carabelli’s trait) that dental anthropologists have classically characterized among human populations (Brothwell 1963; Dahlberg 1971; Scott and Turner 1997), Hlusko identified in baboons (Hlusko 2002b; Hlusko 2002a) and has been observed in the hominid fossil record (e.g. Robinson 1956; Van Reenen and Reid 1995). Understanding the genetic processes and types of genetic change that produce this variation among members of a species should therefore greatly enhance our ability to interpret similar variation found among humans, other primates, and their fossil ancestors.
MATERIALS AND METHODS

Mice

Use of mice was approved by the Pennsylvania State University Institutional Animal Care and Use Committee (IACUC #27388).

Inbred progenitor mice: For the initial screening for tooth phenotypes we obtained 2 mice each from 14 inbred strains that are the progenitors of 16 RI sets (table 3.1). An additional 4 mice each of the A/j (A) and C57BL/6j (B) mouse strains were then obtained from Jackson Labs to ensure there was no environmental or stochastic variation in the trait.

Table 3.1: Progenitor Inbred Strains

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbr</th>
<th>Stock #</th>
<th>Source*</th>
<th>Recombinant Inbred Sets (# lines)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/J</td>
<td>A</td>
<td>000646</td>
<td>Jax</td>
<td>AXB/BXA(25), SMXA(26)</td>
</tr>
<tr>
<td>AKR/J</td>
<td>AK</td>
<td>000648</td>
<td>Jax</td>
<td>AKXD(20), AKXL(15)</td>
</tr>
<tr>
<td>Balb/cJ</td>
<td>C</td>
<td>000651</td>
<td>Jax</td>
<td>CXB(7), CXJ(7), CBXC(9)</td>
</tr>
<tr>
<td>C3H/HeJ</td>
<td>H</td>
<td>000659</td>
<td>Jax</td>
<td>BXH(13)</td>
</tr>
<tr>
<td>C57BL/6J</td>
<td>B</td>
<td>000664</td>
<td>Jax</td>
<td>AXB/BXA(25), BXD(57), BXH(13), BXJ(1), CXB(7), LXB(1)</td>
</tr>
<tr>
<td>C57L/J</td>
<td>L</td>
<td>000688</td>
<td>Jax</td>
<td>AKXL(15), LXB(1), SWXL(7)</td>
</tr>
<tr>
<td>CBA/J</td>
<td>CB</td>
<td>000656</td>
<td>Jax</td>
<td>CBXC(9)</td>
</tr>
<tr>
<td>DBA/2J</td>
<td>D</td>
<td>000671</td>
<td>Jax</td>
<td>AKXD(20), BXD(57)</td>
</tr>
<tr>
<td>ISS</td>
<td>S</td>
<td></td>
<td>U Col</td>
<td>LXS(77)</td>
</tr>
<tr>
<td>ILS</td>
<td>L</td>
<td></td>
<td>U Col</td>
<td>LXS(77)</td>
</tr>
<tr>
<td>NZB/BINJ</td>
<td>N</td>
<td>000684</td>
<td>Jax</td>
<td>NXSM(17)</td>
</tr>
<tr>
<td>SJL/J</td>
<td>J</td>
<td>000686</td>
<td>Jax</td>
<td>BXL(1), CXJ(7), SWXJ(14)</td>
</tr>
<tr>
<td>SM/J</td>
<td>SM</td>
<td>000687</td>
<td>Jax</td>
<td>NXSM(17), SMXA(26)</td>
</tr>
<tr>
<td>SWR/J</td>
<td>SW</td>
<td>000689</td>
<td>Jax</td>
<td>SWXJ(14), SWXL(7)</td>
</tr>
</tbody>
</table>

*Jax: Jackson Laboratories, Bar Harbor, ME (www.jax.org)
*U Col: Institute for Behavior Genetics, University of Colorado at Boulder (Williams et al. 2004)
AXB/BXA recombinant inbred mice: These mice were created by inbreeding the hybrid offspring of the A/J (A) and C57BL/6J (B) strains of mouse. This process was repeated to create a set of 25 independent recombinant inbred lines (table 3.2). We obtained breeding sets (2 female and 1 male mouse) for 9 of the 25 AXB/BXA RI lines and bred them to confirm inheritance of the trait (marked with * in table). For the remaining 16 AXB/BXA strains 2 heads (1 female and 1 male) were obtained from Jackson Labs and scored for the trait. Strain distribution patterns (SDPs) of marker genotypes for the 25 AXB/BXA lines were provided by Professor Robert W. Williams, Department of Anatomy and Neurobiology, University of Tennessee Health Science Center (Williams et al. 2001).

<table>
<thead>
<tr>
<th>Name</th>
<th>Stock #</th>
<th>Name</th>
<th>Stock #</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXB1*</td>
<td>001673</td>
<td>BXA1*</td>
<td>001692</td>
</tr>
<tr>
<td>AXB2</td>
<td>001674</td>
<td>BXA2</td>
<td>001693</td>
</tr>
<tr>
<td>AXB4</td>
<td>001676</td>
<td>BXA4</td>
<td>001694</td>
</tr>
<tr>
<td>AXB5</td>
<td>001677</td>
<td>BXA7</td>
<td>001696</td>
</tr>
<tr>
<td>AXB6</td>
<td>001678</td>
<td>BXA11*</td>
<td>001699</td>
</tr>
<tr>
<td>AXB8</td>
<td>001679</td>
<td>BXA12*</td>
<td>001700</td>
</tr>
<tr>
<td>AXB10*</td>
<td>001681</td>
<td>BXA13</td>
<td>001701</td>
</tr>
<tr>
<td>AXB12*</td>
<td>001683</td>
<td>BXA14</td>
<td>001702</td>
</tr>
<tr>
<td>AXB15*</td>
<td>001685</td>
<td>BXA16*</td>
<td>001703</td>
</tr>
<tr>
<td>AXB18*</td>
<td>001686</td>
<td>BXA17</td>
<td>001697</td>
</tr>
<tr>
<td>AXB23</td>
<td>001690</td>
<td>BXA24</td>
<td>001710</td>
</tr>
<tr>
<td>AXB24</td>
<td>001691</td>
<td>BXA25</td>
<td>001711</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BXA26</td>
<td>001999</td>
</tr>
</tbody>
</table>

Chromosome substitution strains: A complete set of chromosome substitution strains was made by backcrossing the hybrid offspring A/J and C57BL/6J mice into the C57BL/6J strain using marker assisted selection to replace a single C57BL/6J chromosome with the homologous chromosome from A/J (Nadeau et al. 2000; Singer et al. 2004). 1 female and 1 male mouse
from chromosome substitution strains for the A/J chromosomes 11, 13, and 19 in a C57BL/6J background were purchased from Jackson Labs: C57BL/6J-Chr11^{A/J}/NaJ (stock #004389), C57BL/6J-Chr13^{A/J}/NaJ (#004391), and C57BL/6J-Chr19^{A/J}/NaJ (#004397). Their teeth were analyzed for the presence of the tooth trait.

Figure 3.4: Photographs depicting the dentition of one C57BL/6J mouse. Sets of photos like this were compared side-by-side to identify qualitative tooth variation between progenitor strains of recombinant inbred sets.

Skull preparation and phenotyping

To examine the teeth, skulls of young adult mice (4 – 6 weeks) were macerated by removing the skin and boiling for one hour in ~100ml distilled water. Muscle tissue and brain were removed using a blunt-end forceps followed by digestion in 1.0% trypsin (Fisher) in a 30%
saturated sodium borate solution for 1 - 2 hours. Any remaining soft tissue was removed by rinsing with distilled water. Skulls were then rinsed in ethanol and allowed to dry.

Photographs of the molar tooth rows (upper right, upper left, lower right, lower left) were taken from lingual, occlusal, and buccal perspectives at 20X magnification under a dissecting microscope using a CCD camera. Complete sets of photographs for the progenitor strains (e.g. figure 3.4) for each RI set were compared visually to identify minor cuspal variants.

Genetic mapping

Interval mapping was carried out with WebQTL (http://www.genenetwork.org) using 2,445 genome spanning markers. The trait was treated as present/absent with the 25 AXB/BXA lines scored as either 1 or 0, if their teeth resembled A/J or C57BL/6J mice, respectively. WebQTL computes a likelihood of the odds (LOD) ratio measuring the association between trait values and marker alleles at regular intervals across the genome. Empirical suggestive (LOD = 2.24) and significant (LOD = 3.65) thresholds were determined by permutation tests (Churchill and Doerge 1994). Chromosomal regions spanning a one LOD drop-off from mapping peaks (Lander and Botstein 1989) were analyzed for causal variation.

Sequence analysis

C57BL/6J (Waterston et al. 2002) and A/J (Kerlavage et al. 2002; Mural et al. 2002) mouse genomes have both been sequenced. Datasets containing all single nucleotide polymorphisms (SNPs) between A/J and C57BL/6J mice for mouse chromosomes 11, 13, and 19 were kindly provided by Dr. Tony Cox of the Wellcome Trust Sanger Center, Hinxton, UK and analyzed using the “custom track” feature at the UCSC genome browser
All analysis used the July 2007, Build 37 (mm9) assembly of the mouse genome. SNPs provided by Dr. Cox were converted to Build 37 using the ‘lift-over’ tool in Galaxy (http://main.g2.bx.psu.edu/).

Genes with amino acid changing (non-synonymous) SNPs between A/J and C57BL/6J mice in the one LOD intervals were identified using the WebQTL SNP Browser (http://www.genenetwork.org/webqtl/snpBrowser.py). Five kilobase pairs (kb) of sequence flanking genes was examined for SNPs that could affect gene expression using the ESPERR 7X regulatory Potential track (King et al. 2005; Taylor et al. 2006) on the UCSC browser. SNPs with high regulatory potential (>0.05) in predicted transcription factor binding sites were identified using TFSEARCH (http://mbs.cbrc.jp/htbin/nph-tfsearch) (Yutaka Akiyama: "TFSEARCH: Searching Transcription Factor Binding Sites", http://www.rwcp.or.jp/papia/) which identifies transcription factor binding motifs in sequence using the TRANSFAC database (http://www.gene-regulation.com) (Wingender et al. 1996; Wingender et al. 2000).

Candidate SNPs in Glis3 and upstream of epiprofin/Sp6 were sequenced in all AXB/BXA lines strain following amplification from genomic DNA isolated from tail clips with the following primers: 5'-AGCAACCAGGGCCCACCTCCA-3' and 5'-AGAGGGCAGGTCCAAGGCACTG-3' for Glis3-SNP, 5'-TAGCCCTGTGGAGACATCGTGAG-3' and 5'-GTCCGCTACCTAAGCTGTCCTGC-3' for Epiprofin-SNP1, 5'-TAATAATGCCCCACTTGTAGC-3' and 5'-GTCCGCTACCTAAGCTGTCCTGC-3' for Epiprofin-SNP2, and 5'-GAGTATTATCACCAGGGCATTTTTTGGAGC-3' and 5'-CCCCTGTGGAGACATCGTGAG-3' for Epiprofin-SNP3.
Identification of candidate genes

Candidate genes in the one LOD intervals of mapping peaks were identified by extensive literature search including the Gene Expression in Tooth Database (http://bite-it.helsinki.fi), EST databases, and PubMed literature searches. Genes were considered as potential candidates if they had known expression in tooth, were included in tooth, craniofacial, mandible or maxilla EST libraries, or were related through gene family membership to genes that were. This list of genes was further narrowed by direct sequence comparison between A/J and C57BL/6J mice, such that genes with non-synonymous SNPs between the two mouse strains and genes with SNPs in regions of high regulatory potential that overlapped with predicted transcription factor binding sites within five kilobase pairs (kb) up and down stream of the gene were selected for further analysis.

In situ hybridization

Mouse embryos were collected at various developmental stages (E12.5, E14.5, E16.5, and E18.5); day of discovery of vaginal plug was considered E0.5. Heads were dissected in phosphate buffered saline (PBS, Sigma), fixed in 4% paraformaldehyde (Sigma) in PBS at 4° overnight, dehydrated to 100% methanol, and stored at -20° C until use. For sectioning, heads were embedded in Paraplast plus (SPI Supplies, West Chester, PA), and sectioned at 6-8 µm.

PCR primers were designed for Glis3 and epiprofin/Sp6 using PrimerSelect (DNAStrar, Inc). In situ probes were amplified from a cDNA library made by reverse transcription PCR from E12.5 embryonic jaw using the following primers: 5’-
AGCTCGGGCGTCTCCAACTGAG-3’, 5’-AAGGGATGGGGACCTGGGAGTG-3’ for epiprofin/Sp6 and 5’-GCAAGAGTCTCTGGTGTTCCA-3’ and 5’-
CTGCAGTCTATCCAACGGCA-3’ for Glis3. Epiprofin/Sp6 and Glis3 fragments were cloned using the StrataClone PCR Cloning Kit (Stratagene) into the pSC-A vector and confirmed by sequencing. Plasmids were digested with HindIII for antisense and NotI for sense probes and digoxigenin labeled (Dig-UTP) RNA probes were synthesized using T7 (antisense) and T3 (sense) RNA polymerases.

Prior to hybridization sections were deparaffinized with ProPar Clearant (Anatech, Inc, Battle Creek, MI), rehydrated in an ethanol series, fixed in 4% paraformaldehyde, permeabilized with Proteinase K (12µg/ml) for 15 minutes and fixed again. Sections were then dehydrated to 100% methanol and allowed to air dry. Hybridization took place overnight at 60° in a humid chamber using a hybridization buffer consisting of 50% formamide, 2XSSC, 10% dextran sulfate, 0.02% SDS, and 100 µg/ml yeast tRNA (Sigma) with probe concentration of 0.5 µg/ml. Non-hybridized probes were washed away in 50% formamide/2XSSC and maleic acid buffer with tween (MABT). Sections were treated with anti-digoxigenin antibody conjugated with alkaline phosphatase (Roche) in 1% blocking solution (Roche) overnight and signal visualized with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (NBT-BCIP, Promega).

**GenePaint**

Expression patterns for all genes in the one LOD intervals for the three mapping peaks on mouse chromosomes 11, 13, and 19 were searched in the GenePaint (http://www.genepaint.org) database. The GenePaint database has whole body section in situ hybridization results showing mRNA expression for approximately 17,000 genes, mostly at embryonic day (E) 14.5 in the C57BL/6J strain. When a gene from a mapping interval was found in GenePaint, sections were screened for expression in developing molars and incisors.
RESULTS

Trait discovery, selection and phenotyping

To identify variation in cusp pattern between the teeth of progenitor strains of recombinant inbred (RI) sets of mice photographs of the progenitor strains for 16 RI sets were compared visually side-by-side. We identified minor cuspal variation between the progenitor strains of the AXB/BXA, CXJ, and SMXA RI sets. Two minor dental variants (missing fifth cusp on lower second molar (figure 3.5A) and variation in groove separating two lingual cusps on upper second molar (figure 3.5B)) were observed between the teeth of Balb/cJ and SJL/J inbred mice which are the progenitors of the CXJ RI set. A small cusplet on the distal end of the upper first molar (figure 3.5C) was present in SM/J mice and absent in A/J mice the progenitors of the SMXA RI set. And, finally, the A/J and C57BL/6J mice, progenitors of the AXB/BXA RI set differed in the presence of a groove separating the two lingual cusps (L1 and L2) of the upper second molar (figure 3.5D) similar to the trait identified between Balb/cJ and SJL/J.

We chose to focus further study on the trait identified between A/J and C57BL/6J mice (figure 3.5D) because of the size and availability of the RI set. Of the RI progenitors for which traits were discovered the AXB/BXA set was large comprising 25 independent RI lines. In addition, colonies of live mice from the AXB/BXA set are maintained at Jackson Labs and were therefore easily available. The CXJ RI set has only 7 lines that are cryopreserved at Jackson Labs; cost of reviving these mice was prohibitive and with only 7 lines it is unlikely that mapping would have produced reliable results. The SMXA RI set comprises 26 lines maintained at Hamamatsu University School of Medicine, Institute for Experimental Animals in Japan
(Nishimura et al. 1995) but after repeated attempts to contact the source were left unanswered, we were unable to acquire these mice.

Figure 3.5: **Four tooth variants identified between the progenitor strains of recombinant inbred sets.** (A) A/J and C57BL/6J progenitors of AXB/BXA RI set. (B and C) Balb/cJ and SJL/J progenitors of CXJ RI set. (D) A/J and SM/J progenitors of SMXA RI set.

**Trait description:** The tooth variation identified between the A/J and C57BL/6J mice involves the two lingual cusps of the upper second molar. In C57BL/6J mice the two cusps (L1 and L2) are separated by a distinct groove. In A/J mice, this groove is greatly reduced or absent and the two cusps form a continuous ridge. Tooth cusp pattern is controlled by transient signaling.
centers in the dental epithelium that secrete molecules from numerous signaling factor families (e.g. Shh, Bmp, Fgf, Wnt, Eda) (Jernvall et al. 1994; Vahtokari et al. 1996; Tucker et al. 2000; Thesleff et al. 2001). Signal from the enamel knot has been suggested to control spacing of cusps in manner similar to feather patterning and has been linked to variation in cusp number (Jernvall and Jung 2000). The variation observed between A/J and C57BL/6J mice fits with a model of reduced spacing between secondary enamel knots or an increase in size of the enamel knots leading to fusion of the two lingual cusps, and probably occurs late in development.

Although there is variation in the expression of the trait, the two strains of mice had clearly non-overlapping ranges of variability and thus we treated the difference as present or absent. Three-dimensional reconstructions of micro-CT scans of the molar tooth rows for A/J and C57BL/6J mice (figure 3.6) show that the teeth of these mice vary at more than just the lingual cusps of the upper second molars. A/J molars appear broader with more rounded cusps and the lingual cusps of the first molar are less distinctly separated similar to the trait identified in the second molar. The trait used in this study may be a proxy for a more widespread difference in the teeth.

Figure 3.6: **3D reconstruction of micro-CT scans of upper molars of C57BL/6J and A/J mice** show extensive variation between the teeth of these strains of mice although only the trait on the upper second molar was used for mapping. Images courtesy Dr. Timothy Ryan, Penn State Center for Quantitative Imaging.
**AXB/BXA Recombinant Inbred Mice:** The AXB/BXA RI set was developed by reciprocal crosses of A/J and C57BL/6J mice (Nesbitt and Skamene 1984) to study behavior, resistance to infection and tumorigenesis. 25 AXB and 25 BXA strains were originally developed of which 21 have gone extinct. Of the 29 remaining strains several lines (AXB13/AXB14, AXB18/AXB19/AXB20, BXA8/BXA17) were determined to be genetically contaminated and therefore non-independent and only one member of each of these groups should be used for mapping leaving 25 independent AXB/BXA RI lines (Sampson *et al.* 1998).

<table>
<thead>
<tr>
<th>Name</th>
<th>Phenotype*</th>
<th>Name</th>
<th>Phenotype*</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXB1</td>
<td>B</td>
<td>BXA1</td>
<td>B</td>
</tr>
<tr>
<td>AXB2</td>
<td>B</td>
<td>BXA2</td>
<td>B</td>
</tr>
<tr>
<td>AXB4</td>
<td>B</td>
<td>BXA4</td>
<td>B</td>
</tr>
<tr>
<td>AXB5</td>
<td>B</td>
<td>BXA7</td>
<td>A</td>
</tr>
<tr>
<td>AXB6</td>
<td>A</td>
<td>BXA11</td>
<td>B</td>
</tr>
<tr>
<td>AXB8</td>
<td>B</td>
<td>BXA12</td>
<td>B</td>
</tr>
<tr>
<td>AXB10</td>
<td>B</td>
<td>BXA13</td>
<td>B</td>
</tr>
<tr>
<td>AXB12</td>
<td>B</td>
<td>BXA14</td>
<td>B</td>
</tr>
<tr>
<td>AXB15</td>
<td>B</td>
<td>BXA16</td>
<td>A</td>
</tr>
<tr>
<td>AXB18</td>
<td>B</td>
<td>BXA17</td>
<td>B</td>
</tr>
<tr>
<td>AXB23</td>
<td>A</td>
<td>BXA24</td>
<td>B</td>
</tr>
<tr>
<td>AXB24</td>
<td>B</td>
<td>BXA25</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BXA26</td>
<td>B</td>
</tr>
</tbody>
</table>

* A = A/J phenotype, B = C57BL/6J phenotype

We first obtained breeding sets (2 female and 1 male) for 9 AXB/BXA lines: AXB1, 10, 12, 15, 18 and BXA1, 11, 12, 16. These mice were bred and offspring and parents were scored for the tooth trait to ensure inheritance of the phenotype. In all cases offspring clearly resembled parents and each RI line resembled either the A/J or C57BL/6J progenitor strain. We then
obtained 1 male and 1 female mouse from the remaining 16 lines and scored them for the trait. Once again, all RI lines clearly resembled one or the other progenitor strain. Tooth phenotypes for the 25 RI lines are shown in table 3.3. Of the 25 AXB/BXA lines four (AXB6, AXB23, BXA7, and BXA16) resembled the A/J tooth trait, the remaining 21 lines all resembled C57BL/6J.

**Genetic mapping of AXB/BXA tooth trait**

Traits scores of 0 (C57BL/6J-like) and 1 (A/J-like) were used to genetically map the trait using the interval mapping function at WebQTL (http://www.gnenetwork.org) (Wang *et al.* 2003). Interval mapping tests for statistical association between trait values and marker genotypes across the genome. Strain distribution patterns of marker genotypes for 2445 microsatellite and single nucleotide polymorphism (SNP) markers polymorphic between A/J and C57BL/6J mice were used. The results of the full-genome scan are shown in figure 3.7. The analysis yielded suggestive evidence of association of the trait and 3 chromosomal regions found on mouse chromosomes 11, 13, and 19 (figure 3.8).

![Figure 3.7: Full genome interval mapping results for tooth trait in AXB/BXA recombinant inbred mice. Peaks on chromosomes 11, 13, and 19 reach the suggestive significance threshold (LOD = 2.24) and were explored for potential candidate variation involved in producing normal tooth variation.](image-url)
Figure 3.8: **Detail of mapping results for chromosomes 11, 13, and 19.** Red boxes on X-axis represent the ~1 LOD drop-off intervals from mapping peaks that were explored for potential candidate variation.
Significance of association is measured by LOD score (logarithm of the odds) and thresholds were determined by permutation test (Churchill and Doerge 1994). The suggestive cutoff (LOD = 2.24) is set such that you expect one false positive per genome scan (Lander and Kruglyak 1995). No regions reached the significant cutoff (LOD = 3.65). As in any mapping project sample size is the key to generating the statistical power to detect association. Although the RI breeding scheme results in numerous recombination events increasing the effective genetic length up to four times and RI sets are much more efficient for mapping than backcrosses (Lander and Botstein 1989) with only 25 lines suggestive results are probably the best we could hope for.

We therefore identified candidate regions for further study using a one LOD drop-off (red boxes in figure 3.8) from the three mapping peaks that reached the suggestive cutoff. The use of the one LOD interval is used by convention and has been suggested to approximate a 95% confidence interval for the location of the associated genetic change (Lander and Botstein 1989).

**Analysis of candidate chromosomal regions**

The one LOD intervals are quite large, up to 20 megabase pairs (Mb) on chromosome 13 and contain up to 351 genes (chromosome 11) (summarized in table 3.4).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Peak LOD score</th>
<th>~1 LOD interval (Mb)</th>
<th>Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2.24</td>
<td>92-110</td>
<td>351</td>
</tr>
<tr>
<td>13</td>
<td>2.24</td>
<td>30-50</td>
<td>113</td>
</tr>
<tr>
<td>19</td>
<td>3.01</td>
<td>26-32</td>
<td>42</td>
</tr>
</tbody>
</table>

*Table 3.4: Summary of Genes in 1 LOD interval of mapping peaks.*
Chromosome 11: There are two nearby peaks on chromosome 11 that reach the suggestive LOD cutoff of 2.24, they are located at markers rs6357939 (96.9 Mb) and rs6407687 (105.1 Mb). Because of the proximity of these two peaks the one LOD intervals overlap and they were treated as a single region. The one LOD interval spans 18 Mb from 92 Mb to 110 Mb and contains 351 genes. An extensive literature search turned up evidence of a possible role in tooth development for 17 of these genes (table 3.5).

Chromosome 13: The region on chromosome 13 was the largest spanning 20 Mb from 30 Mb to 50 Mb with its peak reaching the suggestive level of cutoff at marker rs13481799 (46.6 Mb). The one LOD interval contained 113 genes of which 7 were found in the tooth development literature (table 3.5).

Chromosome 19: The highest peak was found at D19Mit47 (27.6 Mb) of chromosome 19. The one LOD interval spanned 6 Mb and contained 42 genes, 8 of these had evidence for a role in tooth development (table 3.5).

Coding variation in candidate genes

Analysis of the A/J and C57BL/6J genomes for variation in coding sequence that led to amino acid substitutions using the GeneNetwork SNP Browser revealed 88 total genes with non-synonymous SNPs in the mapping regions (table 3.6). 81 of these are located on chromosome 11, 5 on chromosome 13, and only 2 on chromosome 19. Searching the literature turned up evidence for three of these genes in tooth development (Mrpl10, Itga3, and Glis3).
Table 3.5: Genes in 1 LOD intervals with link to tooth development

<table>
<thead>
<tr>
<th>Chr</th>
<th>Accession #</th>
<th>Gene</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>NM_007689</td>
<td>Chad</td>
<td>Bone/cartilage extracellular matrix</td>
</tr>
<tr>
<td>11</td>
<td>NM_007742</td>
<td>Col1a1</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>11</td>
<td>NM_013565</td>
<td>Itga3</td>
<td>Family relation to integrins</td>
</tr>
<tr>
<td>11</td>
<td>NM_010055</td>
<td>Dlx3</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>11</td>
<td>NM_007867</td>
<td>Dlx4</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>11</td>
<td>NM_030220</td>
<td>Sp2</td>
<td>Family relation to Sp6</td>
</tr>
<tr>
<td>11</td>
<td>NM_031183</td>
<td>Epiprofin/Sp6</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>11</td>
<td>NM_026154</td>
<td>Mrpl10</td>
<td>EST: Soares_NKWMD_mandible</td>
</tr>
<tr>
<td>11</td>
<td>NM_008379</td>
<td>Kpnb1</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>11</td>
<td>NM_008942</td>
<td>Npepps</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>11</td>
<td>NM_025927</td>
<td>Mrpl45</td>
<td>EST: E8.5 Mouse Craniofacial</td>
</tr>
<tr>
<td>11</td>
<td>NM_009545</td>
<td>Pcgf2</td>
<td>EST: Soares_NKWMD_mandible</td>
</tr>
<tr>
<td>11</td>
<td>NM_011971</td>
<td>Psmb3</td>
<td>EST: Soares_NKWMD_mandible</td>
</tr>
<tr>
<td>11</td>
<td>NM_022891</td>
<td>Rpl23</td>
<td>EST: Soares_NKWMD_mandible</td>
</tr>
<tr>
<td>11</td>
<td>NM_011719</td>
<td>Wnt9b</td>
<td>Family relation to Wnt genes</td>
</tr>
<tr>
<td>11</td>
<td>NM_009521</td>
<td>Wnt3</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>11</td>
<td>NM_016780</td>
<td>Itgb3</td>
<td>Family relation to integrins</td>
</tr>
<tr>
<td>13</td>
<td>NM_007556</td>
<td>Bmp6</td>
<td>Family relation to Bmp genes</td>
</tr>
<tr>
<td>13</td>
<td>NM_172143</td>
<td>Ofcc1</td>
<td>Orofacial cleft candidate</td>
</tr>
<tr>
<td>13</td>
<td>NM_010104</td>
<td>Edn1</td>
<td>(Neuhaus and Byers 2007)</td>
</tr>
<tr>
<td>13</td>
<td>NM_007526</td>
<td>Barx1</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>13</td>
<td>NM_025711</td>
<td>Aspn</td>
<td>(Henry et al. 2001)</td>
</tr>
<tr>
<td>13</td>
<td>NM_012050</td>
<td>Omd</td>
<td>Regulates Bmp2</td>
</tr>
<tr>
<td>13</td>
<td>NM_008760</td>
<td>Ogn</td>
<td>Bone/cartilage formation</td>
</tr>
<tr>
<td>19</td>
<td>NM_013703</td>
<td>Vldlr</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>19</td>
<td>NM_177474</td>
<td>D19Bwg1357e</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>19</td>
<td>NM_175459</td>
<td>Glis3</td>
<td>Family relation to Gli genes</td>
</tr>
<tr>
<td>19</td>
<td>NM_025950</td>
<td>Cdc37l1</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>19</td>
<td>NM_021525</td>
<td>Rcl1</td>
<td>EST: Soares_NKWMD_mandible</td>
</tr>
<tr>
<td>19</td>
<td>NM_011272</td>
<td>Rln1</td>
<td>bite-it.helsinki.fi</td>
</tr>
<tr>
<td>19</td>
<td>NM_001081213</td>
<td>Ermp1</td>
<td>EST: NIH_MGC_190 (pooled molar)</td>
</tr>
<tr>
<td>19</td>
<td>NM_010051</td>
<td>Dkk1</td>
<td>(Fjeld et al. 2005)</td>
</tr>
</tbody>
</table>
Mrpl10: The mitochondrial ribosomal protein 10 (*Mrpl10*) located on chromosome 11 was found as an expressed sequence tag (EST) isolated from mouse embryonic mandible (Soares_NKWMD_mandible). It is a structural component of the ribosome, with no known role in tooth development and is unlikely to be involved in dental patterning.

**Table 3.6: Genes with non-synonymous SNPs in 1 LOD intervals**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Gene</th>
<th>Chr</th>
<th>Gene</th>
<th>Chr</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>LOC629636</td>
<td>11</td>
<td>Lrrc46</td>
<td>11</td>
<td>1700039D13Rik</td>
</tr>
<tr>
<td>11</td>
<td>Nme2</td>
<td>11</td>
<td>Mrpl10*</td>
<td>11</td>
<td>4933439F11Rik;170003</td>
</tr>
<tr>
<td>11</td>
<td>LOC629647</td>
<td>11</td>
<td>Osbp17</td>
<td>11</td>
<td>Gfap</td>
</tr>
<tr>
<td>11</td>
<td>Spag9</td>
<td>11</td>
<td>5330439C02Rik</td>
<td>11</td>
<td>3000004C01Rik</td>
</tr>
<tr>
<td>11</td>
<td>Tob1</td>
<td>11</td>
<td>LOC665512</td>
<td>11</td>
<td>Hexim2</td>
</tr>
<tr>
<td>11</td>
<td>Wfikkn2</td>
<td>11</td>
<td>Ccddc49</td>
<td>11</td>
<td>Map3k14</td>
</tr>
<tr>
<td>11</td>
<td>Abcc3</td>
<td>11</td>
<td>Plxdc1</td>
<td>11</td>
<td>LOC237954</td>
</tr>
<tr>
<td>11</td>
<td>Cacna1g</td>
<td>11</td>
<td>Stard3</td>
<td>11</td>
<td>Gm884</td>
</tr>
<tr>
<td>11</td>
<td>Spata20</td>
<td>11</td>
<td>Krt1-c29</td>
<td>11</td>
<td>LOC629960</td>
</tr>
<tr>
<td>11</td>
<td>Epn3</td>
<td>11</td>
<td>4732494G06Rik</td>
<td>11</td>
<td>LOC666749</td>
</tr>
<tr>
<td>11</td>
<td>Mycbpap</td>
<td>11</td>
<td>Ka36</td>
<td>11</td>
<td>LOC632867</td>
</tr>
<tr>
<td>11</td>
<td>Rsad1</td>
<td>11</td>
<td>1110054P19Rik</td>
<td>11</td>
<td>LOC217246</td>
</tr>
<tr>
<td>11</td>
<td>BC018371</td>
<td>11</td>
<td>LOC666021</td>
<td>11</td>
<td>Cd79b</td>
</tr>
<tr>
<td>11</td>
<td>Eme1;Lrrc59</td>
<td>11</td>
<td>4733401H21Rik</td>
<td>11</td>
<td>Scn4a</td>
</tr>
<tr>
<td>11</td>
<td>RP23-290B5.6</td>
<td>11</td>
<td>LOC432602</td>
<td>11</td>
<td>Polg2</td>
</tr>
<tr>
<td>11</td>
<td>LOC544807</td>
<td>11</td>
<td>Krt1-4</td>
<td>11</td>
<td>Ccddc45</td>
</tr>
<tr>
<td>11</td>
<td>LOC665174</td>
<td>11</td>
<td>Krt1-1</td>
<td>11</td>
<td>Falz</td>
</tr>
<tr>
<td>11</td>
<td>Sgca</td>
<td>11</td>
<td>Krt1-5</td>
<td>11</td>
<td>Ccddc46</td>
</tr>
<tr>
<td>11</td>
<td>Samd14</td>
<td>11</td>
<td>Krt1-13</td>
<td>11</td>
<td>Rgs9</td>
</tr>
<tr>
<td>11</td>
<td>Itga3*</td>
<td>11</td>
<td>Krt1-16</td>
<td>11</td>
<td>9930022D16Rik</td>
</tr>
<tr>
<td>11</td>
<td>LOC544809</td>
<td>11</td>
<td>1110036003Rik</td>
<td>11</td>
<td>Abca6</td>
</tr>
<tr>
<td>11</td>
<td>Phospho1;LOC665273</td>
<td>11</td>
<td>A930006D11Rik</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Atp5g1</td>
<td>11</td>
<td>D11Lgp2e</td>
<td>13</td>
<td>Serpinb9c</td>
</tr>
<tr>
<td>11</td>
<td>Ttl6</td>
<td>11</td>
<td>Naglu</td>
<td>13</td>
<td>Serpinb9e</td>
</tr>
<tr>
<td>11</td>
<td>LOC432593</td>
<td>11</td>
<td>Aoc2</td>
<td>13</td>
<td>Cage1</td>
</tr>
<tr>
<td>11</td>
<td>Hoxb7</td>
<td>11</td>
<td>Rundc1</td>
<td>13</td>
<td>Dsp</td>
</tr>
<tr>
<td>11</td>
<td>LOC544810</td>
<td>11</td>
<td>Brca1</td>
<td>13</td>
<td>Kif13a</td>
</tr>
<tr>
<td>11</td>
<td>Scap1</td>
<td>11</td>
<td>Sost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Snx11</td>
<td>11</td>
<td>Grn</td>
<td>19</td>
<td>Glis3*</td>
</tr>
<tr>
<td>11</td>
<td>Scrn2</td>
<td>11</td>
<td>Itga2b</td>
<td>19</td>
<td>LOC434460</td>
</tr>
</tbody>
</table>

* genes with link to tooth development
**Itga3**: Unfortunately, due to the shifting nature of sequence information as new builds of the mouse genome are released two missense mutations in *Itga3*, integrin alpha 3, did not appear until the final revision of this report. *Itga3*, located on chromosome 11, was chosen as a potential candidate gene because of its family relationship to other integrin genes known to be expressed in teeth (e.g. *Itga4, Itga6, Itgb1, Itgb4, and Itgb5*). Integrins are cell surface receptors involved in cell-cell interactions and have been suggested to be important in epithelial-mesenchymal interactions during tooth development (Salmivirta et al. 1996). *Itga3* was recently shown by immunostaining to be expressed in the dental epithelium at E13.5 (Fukumoto et al. 2006). The first missense mutation is a G to A transition at amino acid (aa) position 511 in exon 11 resulting in a change from arginine (CGG) to glutamine (CAG). Arginine is a hydrophilic, basic amino acid and glutamine is hydrophilic, polar (non-charged), thus a shift from arginine to glutamine is likely to have an effect on protein structure. The second is a transition from C to T at aa position 895 in exon 20 resulting in a shift from alanine (GCC) to valine (GTC), a conservative change between two hydrophobic, non-polar amino acids.

**Glis3**: *Glis3* (Gli-similar 3), chromosome 19, was identified through its relationship to the Gli (Glioma-associated oncogene) family of genes which are Krüppel-like transcription factors involved in sonic hedgehog (Shh) signaling in the enamel knot (Hardcastle et al. 1998; Gritli-Linde et al. 2002; Cobourne et al. 2004). We chose to focus further study on *Glis3* because of this relationship and the hypothesized relationship between enamel knot signaling and cusp spacing. There are two SNPs in the coding region of *Glis3* between A/J and C57BL/6J mice (figure 3.9). One was a silent T to C mutation in the third position of an alanine codon (GCT → GCC) in exon 4. The second was a G to A transition at aa position 379 in exon 8 which led to a
glycine (GGG) to glutamic acid (GAG) amino acid change. Glycine is a small, non-polar amino acid and glutamic acid is hydrophilic acidic and is therefore a major non-conservative change likely to influence protein structure.

Figure 3.9: UCSC genome browser shot of Glis3 region on chromosome 19. Red track shows all SNPs between A/J and C57BL/6L genomes. Two SNPs (indicated by arrows) overlapped with the Glis3 coding region: a silent mutation (T to C) in exon 4 and a missense mutation (G to A) in exon 8 that led to a Glycine to Glutamic Acid amino acid shift.

Glis3 is highly expressed in osteoblasts where it promotes differentiation through regulation of Fgf18 (Beak et al. 2007) and mutations in Glis3 have been linked to neonatal diabetes (Senee et al. 2006). Glis3 was known through whole-mount in situ hybridization to be expressed in the developing facial prominences (Kim et al. 2003), but not in tooth specifically. We used in situ hybridization to analyze Glis3 expression in developing tooth at stages E12.5, E14.5, E16.5 and E18.5. Weak expression in the oral and dental epithelium was detected at all stages and no differences in expression were noted between A/J and C57BL/6J mice. At E12.5 Glis3 expression was most prominent in the epithelium on the lingual side near the base of the tooth bud (figure 3.10). At E14.5 and E16.5 expression was extremely weak but detectable in epithelial cells and maintained a pattern of slightly stronger expression on the lingual aspect of the tooth germ. At E18.5 tooth expression is again found in the epithelium, primarily in preameloblast cells, but expression is also seen in mesenchymal cells directly underlying the epithelium (figure 3.10). Expression was also seen in brain, vibrissae and in the mesenchymal
cells of the lining of the oropharynx (not shown). The expression pattern, although weak, is consistent with a role in production of the tooth trait, and Glis3 is therefore considered a good candidate for functional evaluation for a role in tooth development and patterning.

Figure 3.10: **In situ hybridization showing Glis3 expression.** Glis3 expression in the epithelium of developing tooth shown at stages E12.5 (frontal section) and E18.5 (sagittal section) indicated by arrowheads. DE, dental epithelium; DM, dental mesenchyme; OE, oral epithelium; PA, preameloblasts.

**Narrowing the chromosomal regions**

At this point in the project a useful new resource was developed and became available. A complete set of chromosome substitution strains were developed using the C57BL/6J and A/J strains of inbred mice (Nadeau *et al.* 2000; Singer *et al.* 2004). This set of mouse strains each with a single A/J chromosome in an otherwise C57BL/6J genetic background was created by backcrossing the F1 hybrid of A/J and C57BL/6J mice into the C57BL/6J strain with marker assisted selection for each entire chromosome. We obtained one female and one male mouse of the three lines with A/J chromosome 11, 13, and 19 (our three mapping peaks) and checked their
teeth. They all resembled the C57BL/6J tooth trait which led us to conclude that no single region is sufficient to produce the trait. We thus adopted a multi-gene model for the trait that involved interaction among the three regions. We used this information to narrow the candidate chromosomal regions by looking for a pattern of markers on the three chromosomes that are uniquely shared by RI strains that have the A/J phenotype.

![Figure 3.11](image)

**Figure 3.11: Strain distribution patterns of AXB/BXA RI mice showing shared chromosomal regions.** RI lines with A/J tooth trait are outlined in red, BXA11, outlined in blue, shared the pattern of regions but had the C57BL/6J tooth phenotype. Marker genotypes B (C57BL/6J) are shaded grey; A (A/J) are white.

The 4 RI strains with the A/J phenotype (AXB6, AXB23, BXA7 and BXA16) shared a pattern of smaller chromosomal regions with one other RI strains (BXA11) that had the C57BL/6J tooth phenotype (figure 3.11). These narrowed chromosomal intervals are summarized in table 3.7. The shared regions corresponded perfectly to the peaks of the LOD scores on each chromosome, including two regions on chromosome 11. The inclusion of BXA11 in this group could be explained if there was a double recombination event that included the causal variant between two typed markers. Although this seemed unlikely given the density of the markers used we checked for this by sequencing four candidate SNPs in the regions (the non-synonymous SNP in *Glis3* and 3 potential regulatory SNPs upstream of *epiprofin/Sp6*, discussed below) in all RI lines. No double recombinants were identified and the shared pattern in BXA11 remains unexplained. It is possible that an unidentified factor outside of the mapping
regions influencing the trait explains the BXA11 exception, but there is no direct evidence for this. Despite this exception, the narrowed regions were used for further analysis.

### Table 3.7: Narrowed Chromosomal Intervals

<table>
<thead>
<tr>
<th>Chr</th>
<th>~1 LOD interval (Mb)</th>
<th># Genes</th>
<th>Tooth Genes</th>
<th>Narrowed Interval (Mb)</th>
<th># Genes</th>
<th>Tooth Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>92-110</td>
<td>351</td>
<td>17</td>
<td>97.7-97.8</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>105-106</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>30-50</td>
<td>113</td>
<td>7</td>
<td>46-47</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>26-32</td>
<td>42</td>
<td>8</td>
<td>27-29</td>
<td>22</td>
<td>5</td>
</tr>
</tbody>
</table>

**Regulatory variation in candidate genes**

We did not want to limit our search for causal variation to just protein coding regions, as changes in gene regulatory elements that control when and where genes are expressed are just as likely to produce the type of variation we see (e.g. Wray 2007). Identifying the regulatory elements of genes bioinformatically is a major challenge because transcription factor binding sites are short sequences (6-10 base pairs) that tolerate variation and thus traditional homology searches are ineffective. The approach taken here for identifying potential regulatory variation uses the best currently available tools, but it should be noted that potential regulatory elements would require experimental confirmation.

For genes in the narrowed chromosomal regions that had some existing evidence of relation to tooth development we looked at sequence 5 kilobase pairs (kb) up and downstream of the gene and in introns for SNPs in regions of high regulatory potential (>0.05) based on the 7x regulatory potential track (ESPERR 7xRP) on the UCSC genome browser. These regions are selected not only for high conservation in seven species sequence alignments but also for
similarity to sequence motifs in a training set of known regulatory elements (King et al. 2005; Taylor et al. 2006). Although many distal regulatory elements have been identified (e.g. Lettice et al. 2002; Nobrega et al. 2003), the majority of regulatory elements are expected to be within 5 kb of genes (Birney et al. 2007). Sequence containing SNPs with high regulatory potential were then analyzed using TFSEARCH, a program that identifies potential transcription factor binding sites using the TRANSFAC database. Genes with SNPs with high regulatory potential and in predicted transcription factor binding sites were considered candidates for a role in producing the tooth phenotype.

This analysis was originally done before the regulatory potential track was available on the mouse genome. Coordinates for sequence with high regulatory potential were converted from the human genome to the mouse genome. From this analysis one gene, epiprofin/Sp6 had three SNPs in the 5 kb upstream of the gene that had both high regulatory potential and were located in predicted transcription factor binding sites. The experimental follow-up of this gene is described below. However, since the original analysis the regulatory potential track has become available on newer versions of the mouse genome and the data was reanalyzed greatly increasing the number of SNPs with high regulatory potential. Table 3.8 summarizes the number of SNPs found in this new analysis upstream (5′), in introns, and downstream (3′) for each of the 14 ‘tooth’ genes in the narrowed chromosomal regions. Table 3.8 also shows the number of these SNPs that are located in predicted transcription factor binding sites.

The new analysis has identified evidence for potential regulatory variation in several genes that were not in the original analysis. SNPs in regions of high regulatory potential and in predicted transcription factor binding sites were found in sequence flanking Sp2, epiprofin/Sp6, Mrpl10, Npepps, Pcgf2, Psmb3, and Glis3. The expression pattern of Glis3 between the A/J and
C57BL/6J progenitor strains was examined at several developmental stages as described above and no difference in expression was observed. The remaining genes with newly identified regulatory variation could be examined in future work.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Gene</th>
<th>5'</th>
<th>Intronic</th>
<th>3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Sp2</td>
<td>16 (3)</td>
<td>14 (2)</td>
<td>7 (1)</td>
</tr>
<tr>
<td>11</td>
<td>Sp6</td>
<td>13 (5)</td>
<td>7 (1)</td>
<td>6 (0)</td>
</tr>
<tr>
<td>11</td>
<td>Mrpl10</td>
<td>19 (4)</td>
<td>7 (1)</td>
<td>18 (4)</td>
</tr>
<tr>
<td>11</td>
<td>Kpnb1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Npepps</td>
<td>0</td>
<td>12 (1)</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Mrpl45</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>11</td>
<td>Pcgf2</td>
<td>1 (0)</td>
<td>5 (2)</td>
<td>2 (0)</td>
</tr>
<tr>
<td>11</td>
<td>Psmb3</td>
<td>3 (1)</td>
<td>0</td>
<td>1 (0)</td>
</tr>
<tr>
<td>11</td>
<td>Rpl23</td>
<td>0</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>19</td>
<td>Glis3</td>
<td>2 (1)</td>
<td>166 (25)</td>
<td>3 (0)</td>
</tr>
<tr>
<td>19</td>
<td>Cdc37l1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>Rcl1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>Rln1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>Ermp1</td>
<td>0</td>
<td>1 (0)</td>
<td>0</td>
</tr>
</tbody>
</table>

* number of SNPs in predicted transcription factor binding sites using TFSEARCH in parentheses

**Epiprofin/Sp6:** Results of the initial screening for regulatory variation led us to explore the expression patterns of one of these genes, *epiprofin/Sp6*. *Epiprofin/Sp6* is a Krüppel-like transcription factor with three zinc-finger domains (Scohy et al. 2000). It is a member of the Specificity Protein/Krüppel-like factor (SP/KLF) gene family which consists of 25 related transcription factors that play multiple developmental roles (Suske et al. 2005). *Epiprofin/Sp6* is expressed continually in the epithelium of developing teeth from E11.5 to E17.5 and was also
detected in limb buds and hair follicles (Nakamura et al. 2004). It has been shown to be a positive regulator of cell growth and proliferation (Nakamura et al. 2004).

![Figure 3.12: Epiprofin/Sp6 knockout results. Histological sections showing tooth bud at various developmental stages in mice heterozygous (top) and homozygous (bottom) for epiprofin/Sp6 knockout. Continuous budding of epithelium (panel f and g) leads to multiple supernumerary teeth in adult (radiographs of six month old mouse). Modified from (Nakamura et al. 2008).](image)

 Recently, two groups (Hertveldt et al. 2008; Nakamura et al. 2008) have created *epiprofin/Sp6* knockout mice which produce major patterning disruptions of tooth development as well as phenotypic effects to skin and limbs. Knockout mice have multiple supernumerary teeth (hyperdontia) with poorly developed cusps that appear flattened (figure 3.12). The teeth appear yellow and chalky and have severely reduced or completely missing enamel and dentin. Nakamura et al. (Nakamura et al. 2008) suggest that the knockout phenotype is caused by reduced cell proliferation and failure of dental epithelial cells to differentiate; they have shown that *epiprofin/Sp6* promotes epithelial differentiation in culture. Undifferentiated epithelial cells continually produce tooth buds (figure 3.12) leading to supernumerary teeth and fail to form
ameloblasts that secrete the proteins necessary for enamel mineralization. Both groups also note that the enamel knots of developing teeth are disorganized and fail to undergo apoptosis.

Figure 3.13: **UCSC Browser shot of epiprofin/Sp6 including 5 kb up and down stream.** Light blue track shows regulatory potential. Track indicated by arrow shows SNPs in regions of regulator potential greater than 0.05.

<table>
<thead>
<tr>
<th>SNP position</th>
<th>Location</th>
<th>Predicted Binding Site*</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>96825480</td>
<td>5'</td>
<td>GGGTCTTTCC</td>
<td>NF-Kappaβ</td>
</tr>
<tr>
<td>96827373</td>
<td>5'</td>
<td>ACGATGGC</td>
<td>GATA-2</td>
</tr>
<tr>
<td>96827988</td>
<td>5'</td>
<td>GGAGATGGAG</td>
<td>GATA-1</td>
</tr>
<tr>
<td>96828247</td>
<td>5'</td>
<td>CAGAAAG</td>
<td>cdx-1</td>
</tr>
<tr>
<td>96829565</td>
<td>5'</td>
<td>GCCGATAGGA</td>
<td>GATA-1</td>
</tr>
<tr>
<td>96835685</td>
<td>intron 1</td>
<td>GCCCGCGCCCC</td>
<td>Sp1</td>
</tr>
</tbody>
</table>

*position of SNP underlined

Screening of non-coding sequence including five kilobase pairs flanking *epiprofin/Sp6* revealed 26 SNPs that fell in regions of high regulatory potential (figure 3.13). Five SNPs upstream of the gene and one in the first intron were also found in predicted transcription factor binding sites (table 3.9). Expression of *epiprofin/Sp6* was therefore examined in A/J and C57BL/6J mice through *in situ* hybridization to look for expression differences between the two
strains. Strong expression was detected in the epithelium at all developmental stages examined (figure 3.14) consistent with previous reports of epiprofin/Sp6 expression (Nakamura et al. 2004), but no differences were noted between the two strains.

Figure 3.14: In situ hybridization showing expression of epiprofin/Sp6 at stages E12.5, E14.5, E16.5 (frontal sections), and E18.5 (sagittal section) in A/J and C57BL/6J mice. DE, dental epithelium; DM dental mesenchyme.
Detecting quantitative changes in gene expression is difficult using *in situ* hybridization and it is possible that slight changes in expression of *epiprofin/Sp6* undetectable by *in situ* could be involved in production of the trait. Methods to quantify gene expression differences in teeth are not yet available; current methods (e.g. microarrays, quantitative RT-PCR) require large amounts of tissue or amplification of RNA that could bias results. Thus, although no differential expression was found in our study, because of its known role in tooth development and the knockout phenotype, *epiprofin/Sp6* cannot be completely dismissed as having a role in the tooth variation observed between A/J and C57BL/6J mice. Quantitative gene expression techniques under development that utilize new high-throughput sequencing technology to sequence the entire population of mRNA in cells may allow more precise comparisons in the near future, at which point the expression of epiprofin/Sp6 could be reevaluated.

**GenePaint: an unbiased approach to identifying candidate genes?**

The identification of candidate genes presented thus far is limited by the availability of information relating specific genes to tooth development in the literature, and this data is inherently biased. Genes usually get reported because of their link to some specific disease or phenotype and are treated (and often named) as if they were ‘for’ that phenotype. Any other possible roles for the gene are ignored. A newly available resource, GenePaint (http://www.genepaint.org)(Visel *et al.* 2004), makes it possible to reduce this bias. The GenePaint database contains whole-body expression data by section *in situ* hybridization in E14.5 mouse for nearly 17,000 genes, and new genes are continuously added. This sort of general expression data has been available for certain tissues in the form of microarray expression databases, but this is the first in which high-resolution images of sections through
entire embryos can be searched for gene expression. GenePaint is therefore a useful tool to identify potential candidate genes from mapping studies, although it is still limited by the fact that primarily only one developmental stage is represented.

To try to identify candidate genes for the tooth variation between A/J and C57BL/6J mice with the least possible bias all genes in the one LOD intervals of our mapping peaks were searched in the GenePaint database. The results are summarized in table 3.10. Out of 506 total genes 352 were found in the GenePaint database and more than half of these (187) were expressed in teeth. Expression does not necessarily equal function and many of these genes are likely to be involved in generic cellular processes (‘housekeeping’) that have no effect on tooth patterning, but for most genes we have no way to determine this a priori.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Peak LOD</th>
<th>~1 LOD Interval (Mb)</th>
<th>Total Genes</th>
<th>Genes in tooth literature</th>
<th>Genes in GenePaint</th>
<th>GenePaint Genes expressed in teeth</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2.24</td>
<td>92-110</td>
<td>351</td>
<td>17</td>
<td>242</td>
<td>136</td>
</tr>
<tr>
<td>13</td>
<td>2.24</td>
<td>30-50</td>
<td>113</td>
<td>7</td>
<td>84</td>
<td>43</td>
</tr>
<tr>
<td>19</td>
<td>3.01</td>
<td>26-32</td>
<td>42</td>
<td>8</td>
<td>26</td>
<td>18</td>
</tr>
</tbody>
</table>

For many of the genes discovered to be expressed in teeth there would be no way prior to the availability of GenePaint to identify them as potential candidates for tooth patterning. For example, figure 3.15A shows a section from GenePaint stained for expression of the Huntington-associated protein 1 (Hap1) gene. Hap1 was discovered because of its ability to bind the huntingtin protein in a yeast two-hybrid assay and its role in neurodevelopment has been studied extensively (e.g. Walling et al. 1998; Fujinaga et al. 2004; Sheng et al. 2008). In GenePaint we
see that it is expressed in developing brain and nerve tissue as we might expect given its association with Huntington disease, but a closer look (inset) reveals very specific and strong expression in the epithelium of developing molar. Figure 3.15B shows another example, the testes expressed gene 2 (Tex2), expressed in both molars and incisors.

Figure 3.15: **Tooth expression in GenePaint.** Genes such as the Huntington associated protein 1 (Hap1) (A, inset shows detail of developing molar) and the Testes expressed gene 2 (Tex2) (B, insets show incisor and molar expression) could not be identified as potential candidates for tooth variation using tradition literature searches.

Teeth are complex traits but they are small and derived from a limited number of tissues, thus seeing the number of genes expressed in teeth is quite surprising. Other traits are likely to share this pattern reminiscent of Sewell Wright’s notion of universal pleiotropy. Analysis of candidate chromosomal regions with resources like GenePaint should alter irrevocably the interpretation of mapping results; candidate genes can no longer be cherry-picked from the literature. Mapping regions generally span megabases and with the known levels of SNP variation even among two inbred strains of mice (1000 – 2000/Mb) the number of potential causal variants is daunting. In the current case we can still identify all the genes expressed in teeth with coding variation between the two strains of mouse but the increased from three to 21.
Additionally, we have no *a priori* reason to eliminate any of the 187 genes identified as expressed in teeth from an analysis for regulatory variation.

**SUMMARY**

Normal variation in traits is the raw material for evolution, but traditional genetic approaches have relied on mutations of large, pathologic-scale effect in order to discover a link between genes and traits. Minor, natural variants in traits like teeth result from subtle and often complex genetic and developmental causation which are difficult to parse, but given the obvious evolutionary importance of understanding normal variation in traits methods must be developed to identify the genes and types of genetic changes responsible. This study presents one such method that relies on the identification of variation between the progenitor strains of recombinant inbred mouse sets.

We identified slight variation in cusp pattern between the A/J and C57BL/6J strains of mouse which are progenitors of the AXB/BXA recombinant inbred set. Association mapping revealed three chromosomal regions associated with the trait variation on mouse chromosomes 11, 13, and 19. We used direct sequence comparison to identify three genes with known expression in teeth (*Mrpl10, Itga3*, and *Glis3*) in these regions with non-synonymous SNPs between the two strains. *Itga3* was identified too late to evaluate its expression but *Glis3* was shown to be expressed in the epithelium of developing tooth. Although *Mrpl10* is unlikely to be involved in dental patterning, *Itga3* and *Glis3* are good candidate genes to be evaluated.
functionally by either replacing the A/J gene sequence with the homologous sequence from C57BL/6J or vice versa or by transgenic over-expression and knockout.

Potential regulatory variation was identified for seven genes with known tooth expression (Sp2, epiprofin/Sp6, Mrpl10, Npepps, Pcgf2, Psmb3, and Glis3). Expression studies by in situ hybridization of epiprofin/Sp6 and Glis3 did not reveal any difference in expression of these genes between the two mouse strains but it is possible that subtle expression differences not detectable by in situ could be involved in producing the trait variation. Expression of the remaining five candidate genes could be examined in future work.

Analysis of the genes in the candidate chromosomal regions with GenePaint revealed an additional layer of complexity for identifying candidate genes from mapping results. Over half of the genes in the regions were observed to be expressed in teeth. This result is likely to be generalizable to other traits and the chromosomal regions associated with them through mapping studies. Identifying candidate genes solely through previous reports in the literature will potentially miss hundreds candidate genes. Dealing with the glut of genes that must be analyzed experimentally will rival the existing problems of identifying causal variants in mapping studies.

The main issues with any mapping study, and the current work is no exception, are strength of statistical association between the trait and candidate chromosomal regions and the size of those regions. Both of these issues are largely a factor of sample size. Future work for this project includes increasing the sample size with a set of recombinant congenic strains (RCS) of mice developed from the A/J and C57BL/6J mice (Fortin et al. 2001). RCS mice are made through two rounds of backcrosses followed by inbreeding. The resulting strains of mice will have a random approximately 12.5% of the genome of one strain in the background of the
second strain (Demant and Hart 1986). 37 RCS strains were developed (Fortin et al. 2001) and are maintained at Emerillon Therapeutics in Montreal, Canada.
REFERENCES


Chapter 4

TRANSGENIC ALTERATION OF *BMP4* EXPRESSION IN THE DEVELOPING DENTITION OF THE MOUSE

INTRODUCTION

The mammalian dentition consists of an array of repeated units (teeth) specialized into four tooth types (incisors, canines, premolars and molars). Teeth develop through reciprocal interactions between the oral epithelial and neural-crest derived ectomesenchyme in the first branchial arch and fronto-nasal process (Peters and Balling 1999; Pispa and Thesleff 2003). Tissue recombination experiments have shown that the initial inductive signal resides in the overlying epithelium and later the odontogenic potential shifts to the mesenchyme (Kollar and Baird 1970; Kollar and Fisher 1980; Lemus 1995). Although the molecular details of dental patterning are still being worked out the dentition can be thought of as a set of nested waves at both the morphological level and in the expression of genes that establish its pattern. The first
wave establishes the number and location of teeth along the jaw, and subsequent waves produce
tooth identity and cusp patterns within teeth.

Numerous experimental approaches (e.g. Jernvall and Jung 2000; Kassai et al. 2005;
Plikus et al. 2005; Cai et al. 2007; Kavanagh et al. 2007) and computer simulations of
developing teeth (Salazar-Ciudad and Jernvall 2002; Salazar-Ciudad et al. 2003; Salazar-Ciudad
and Jernvall 2004) have shown that the dynamic interaction model of dental patterning inspired
by Turing’s reaction-diffusion processes (Weiss et al. 1998) fits best with the type of variation
observed in teeth. Additionally, observations in baboons (Hlusko 2004; Hlusko et al. 2004) and
experiments in mice (Kangas et al. 2004) have demonstrated clearly that dental characters are
correlated between upper/lower and right/left teeth. This non-independence of tooth traits
supports the dynamic interaction model where the dental pattern derives from a quantitative
process rather than individual genes for each tooth trait.

The basic idea of the dynamic interaction model is that morphogens diffusing across
space interact to produce a periodic wave-like pattern of activation and inhibition. Peaks of
activation are thought to cross a threshold initiating a downstream developmental cascade of
events that produces teeth and the cusps within teeth. Valleys of activator level correspond to
the spaces between units. The diffusing morphogens of development are extracellular signaling
factors such as members of the bone morphogenetic protein (BMP), fibroblast growth factor
(FGF) or tumor necrosis factor (TNF) families. Cells detect the level of these ligands through
cell surface receptors, which in turn activate a series of secondary messenger molecules that
ultimately influence gene expression by activating transcription factors, genes that control the
expression of other genes within the cell. The changes in gene expression that result lead to cell
differentiation, the expression of enamel and dentin matrix proteins, mineralization, and finally,
eruption of the tooth into the oral cavity. During dental development the expression patterns of many genes resemble a wave generating process (Thesleff and Sharpe 1997; Keranen et al. 1998; Keranen et al. 1999; Peters and Balling 1999; Thesleff et al. 2001; Thesleff and Mikkola 2002) and are reminiscent of expression patterns in evolutionary related traits like feather patterning where reaction-diffusion like processes have been clearly demonstrated (figure 4.1) (Jung et al. 1998; Jung et al. 1999; Jernvall and Jung 2000).

![Figure 4.1: Similarities between feather and tooth patterning. Tooth development including gene expression patterns resembles feather patterning which has been shown experimentally to fit a reaction-diffusion process (Jernvall and Jung 2000).](image)

Several factors contribute to the pattern that results from a reaction-diffusion-like process including the space in which the reaction takes place, the characteristics of the interactions among signaling molecules, diffusion rates, and the amount of signaling molecules present at any given time. Slight changes in any one of these factors can lead to dramatic change in the resulting pattern. Using computer simulation, Salazar-Ciudad and colleagues have shown how
this can influence tooth development and evolution. By altering the parameters of their model they were able to change a simulated mouse-like dental pattern to that of the closely related vole (figure 4.2) (Salazar-Ciudad and Jernvall 2002). To experimentally test the dynamic interaction model of dental patterning in vivo the easiest of these parameters to manipulate is the amount of signaling molecules present. There are established experimental protocols for eliminating a signaling molecule altogether (null mutations or gene knockouts) and increasing their expression transgenically. More subtle down-regulation of genes using RNA interference or morpholinos is available in many model systems, but not yet standard in mammals.

![Image of predicted and observed dental patterns](image)

**Figure 4.2:** Computer simulation of tooth development mimicking mouse (left) and vole (right) dental patterns. Simulations of the dynamic interaction model of tooth development can recreate natural dental patterns (Salazar-Ciudad and Jernvall 2002).

To date, the approaches taken that test the model have all involved pathologic-scale natural or experimentally induced mutations (many examples in Thyagarajan et al. 2003; Fleischmannova et al. 2008) and little is known about the genetic and developmental processes that produce natural variation in teeth. While these studies have successfully identified numerous genes involved in dental patterning and many of the interactions among them; the
large-scale phenotypic effects are difficult to interpret in evolutionary terms. Many of the genes known to be involved in tooth development are highly pleiotropic and thus null mutations tend to be embryonic lethal and halt tooth development well before any useful information regarding their role in patterning can be gained (Thyagarajan et al. 2003). Most transgenic approaches have utilized the enhancer from keratin-14 (K14) a gene that is highly expressed in all ectodermally derived epithelial tissue (Coulombe et al. 1989; Vassar et al. 1989). Transgenes driven by the K14 enhancer thus tend to produce major patterning disruption including loss of teeth or supernumerary teeth, as well as phenotypic effects to skin, hair, limbs, and external genitalia (Mustonen et al. 2003; Tucker et al. 2004; Wang et al. 2004a; Plikus et al. 2005).

Because teeth are serially homologous structures they do vary meristically among species, so tooth gain and loss has played an important role in the evolution of the dentition, but the majority of tooth evolution, especially among closely related species likely involves small-scale natural variation in teeth among members of a population.

The goal of this project is to use existing experimental techniques in a novel way in an attempt to produce variation in teeth that mimics natural variation. By using a tissue specific enhancer from a transcription factor gene naturally involved in tooth development we hoped to transgenically alter gene expression subtly. We therefore utilized a portion of the enhancer of the distal-less homeobox 2 (Dlx2) gene, a transcription factor naturally expressed during tooth development, that drives expression specifically in the oral and dental epithelium (Thomas et al. 2000). Importantly, this enhancer did not drive expression in the brain where Dlx2 is naturally expressed. We used this enhancer to drive over-expression of the bone morphogenetic protein 4 (Bmp4) gene which is known to be involved in tooth patterning at multiple developmental stages. The aim was to produce subtle variation in teeth by slightly altering the expression level of this
important morphogen and thus test in as natural a sense as possible the dynamic interaction model of dental patterning.

**Transgenic alteration of dental patterning**

Since beginning this project several papers have appeared that use a similar approach. Many of these have produced dramatic alteration of dental patterning. Several recent studies (Mustonen et al. 2003; Tucker et al. 2004; Wang et al. 2004a; Wang et al. 2004b; Plikus et al. 2005) have shown that quantitative changes in gene expression level can affect tooth patterning as predicted by the dynamic interaction model. *Ectodysplasin (Eda)* and its receptor *Edar* are members of the TNF family of signaling molecules implicated in the formation of the enamel knots (Tucker et al. 2000). Transgenic over-expression of these genes led to major disruption of dental patterning and a suite of other phenotypic effects to ectodermal organs. Mice over-expressing *Eda* using the K14 enhancer had disrupted enamel formation and a supernumerary fourth molar (figure 4.3A) although cusp pattern seemed to be unaffected (Mustonen et al. 2003). Over-expression of *Edar* also occasionally led to extra molars, but the most prominent effect was the production of additional cusps (figure 4.3B). Based on the results of these two studies, Tucker et al. (Tucker et al. 2004) suggest that the natural function of EDA signaling may be to pattern the number of teeth and cusps within teeth by influencing enamel knot size and spacing.

The transgenic over-expression of several genes encoding natural antagonists of BMP signaling using the K14 enhancer have also produced major disruptions of dental patterning. Two of these, *follistatin* (Tucker et al. 2004; Wang et al. 2004a; Wang et al. 2004b) and *ectodin* (Kassai et al. 2005) have known roles in dental patterning and are discussed below. A third
BMP antagonist, *noggin*, although not directly implicated in normal tooth development disrupts dental patterning when over-expressed in the dental and oral epithelium using the K14 enhancer (Plikus et al. 2005). *K14-Noggin* mice lack all mandibular molars and maxillary third molars. Mutant molars are small, have shallow disorganized cusps, and lack enamel (figure 4.3C). In addition to tooth effects *K14-Noggin* mice have defects to hair, limbs, genitalia, and skin (Plikus et al. 2004).

Figure 4.3: **Transgenic alteration of dental patterning.** (A) Over-expression of *ectodysplasin* (*Eda*) results in supernumerary fourth molar (Mustonen et al. 2003). (B) Histological sections of K14-Edar (top) and wild-type mice showing additional cusps (marked by *) in Edar mutant (Tucker et al. 2004). (C) Over-expression of BMP antagonist noggin leads to reduced molar size, reduced cusps, and missing third molars (Plikus et al. 2005).
The result of down-regulating BMP signaling transgenically seems to show that the level of BMP signal is vital to normal dental patterning. However, use of the K14 enhancer produces such high levels of over-expression that the resulting phenotypes are pathologic and therefore not the kind of variation important in evolution. Thus, by over-expressing Bmp4 using the Dlx2 epithelial enhancer we hoped to produce reciprocal but smaller-scale effects on dental patterning. It should also be noted that many BMP antagonists, including follistatin, ectodin, and noggin have binding affinity for multiple BMPs (Balemans and Van Hul 2002; Canalis et al. 2003). Over-expression of these antagonists affects more than just BMP4 signaling in developing tooth where many BMPs play key roles in patterning.

The Dlx2 epithelial enhancer

The vertebrate distal-less homeobox (Dlx) genes are transcription factors related to the Drosophila distal-less (Dll) gene (Stock et al. 1996). Dll is necessary for the formation of the distal ends of limbs, mouthparts and antennae in flies (Cohen et al. 1989). In mammals there are six Dlx genes found in pairs on the same chromosomes as three Hox clusters. The mammalian Dlx genes are expressed developmentally in numerous tissues including branchial arches, limb buds, forebrain, eyes, ears and teeth (Stock et al. 1996).

The expression pattern of Dlx2 in tooth development is well characterized (figure 4.4). Expression is first seen at the dental lamina stage (~E11.5) where it is expressed in both the oral and dental epithelium and underlying mesenchyme. Expression continues throughout tooth development becoming restricted to the dental epithelium, including the enamel knot, and mesenchymal cells of the dental papilla (Thomas et al. 1995; Zhao et al. 2000b). Dlx2 is found
on mouse chromosome 2 linked to \textit{Dlx1} near the HoxD cluster (McGuinness \textit{et al.} 1996). While single knockouts of \textit{Dlx1} and \textit{Dlx2} have no overt affect on tooth development, double knockouts (Dlx1\textsuperscript{-/-}, Dlx2\textsuperscript{-/-}) results in loss of maxillary molars suggesting functional redundancy among the Dlx genes in tooth development (Qiu \textit{et al.} 1997).

![Figure 4.4: Endogenous \textit{Dlx2} expression in developing tooth.](http://bite-it.helsinki.fi)  

The epithelial and mesenchymal expression domains of \textit{Dlx2} are believed to be regulated independently (Thomas \textit{et al.} 2000). Using a reporter construct containing a 4.7 kilobase pair (kb) fragment of genomic DNA upstream of the \textit{Dlx2} gene linked to the \textit{LacZ} gene, Thomas and colleagues recapitulated the epithelial expression domain of \textit{Dlx2} (figure 4.5). \textit{LacZ} expression was observed in the epithelium of the first branchial arch as well as the apical ectodermal ridge (AER) of developing limb buds. We obtained this reporter construct from Professor John Rubenstein at the Nina Ireland Laboratory of Developmental Neurobiology, University of California, San Francisco for use in our study.
Figure 4.5: **Expression of Dlx2-LacZ reporter construct during development.** (A) Diagram of reporter construct. (B-E) LacZ expression at E10.5. (F-I) Endogenous Dlx2 expression. (Thomas *et al.* 2000)

**Bmp4 in tooth development**

Signaling molecules in the bone morphogenetic protein (BMP) family are members of the transforming growth factor β (TGFβ) superfamily of extracellular ligands and play multiple roles during development. Signal from BMPs is transduced through membrane-bound serine/threonine kinase receptors (BMP receptors type I and type II). Upon binding of the BMP ligand to its receptor, associated transcription factor proteins of the Smad family are phosphorylated and transported to the cell nucleus where they influence the expression of the target genes of Bmp signaling (Chen *et al.* 2004). More than 30 BMPs have been identified (Laurikkala *et al.* 2003) and they are divided into subfamilies based on phylogenetic analysis.
Bmp4 (along with Bmp2) is a member of the decapentaplegic (dpp) subfamily because of its relationship to the Drosophila dpp gene (Nie et al. 2006).

The role of Bmp4 in development has been extensively studied. It is expressed prior to gastrulation during embryogenesis and plays a role in dorsal/ventral patterning, neural patterning, skeletal and limb development, the hematopoietic system, and the development of teeth. Because of its early expression and pleiotropic role in so many fundamental developmental processes, knocking out Bmp4 results in early embryonic lethality prior to embryonic day (E) 9.5 in the mouse (Zhao 2003).

![Figure 4.6: Endogenous expression of Bmp4 during tooth development.](http://bite-it.helsinki.fi)

During tooth development Bmp4 is expressed in a dynamic pattern shifting between the epithelium and mesenchyme (figure 4.6). It plays an important role as an induction signal between these two tissues and its expression has been associated with the shift of odontogenic potential between epithelium and mesenchyme (Vainio et al. 1993; Aberg et al. 1997). Prior to any morphological signs of tooth development and at the epithelial thickening stage (E10 – 11)
of tooth development, \(Bmp4\) expression is detected in the oral epithelium. BMP4 signal from the epithelium is sufficient to induce its own expression and that of many other genes in the mesenchyme (Vainio et al. 1993) and by mid-tooth development (bud stage, E12 - 13) \(Bmp4\) is expressed exclusively in the condensing cells of the dental mesenchyme where it is involved in the induction of the enamel knots, transient signaling centers in the epithelium (Jernvall and Thesleff 2000). Subsequently, at cap stage (E14 – 15) \(Bmp4\) expression continues in the mesenchyme but is also detected in the enamel knots themselves which determine the final cusp pattern of the tooth (Jernvall et al. 1998). By bell stage (E16 – 19) \(Bmp4\) expression is again confined to the dental mesenchyme in cells of the dental papilla. Finally, Bmp4 is expressed in both the ameloblast and odontoblast cell layers (Aberg et al. 1997).

![Figure 4.7: Schematic showing relationship between interaction of Bmp4 and Fgf8 in the epithelium at ~E10.5 and location of future teeth. Region of oral epithelium that expression Fgf8, but not Bmp4 induces expression of downstream genes such as Pax9 in the dental mesenchyme initiating the developmental cascade that will eventually produce teeth. Modified from (Peters and Balling 1999).]
A number of experimental studies have shown the critical role that \textit{Bmp4} plays in dental patterning. Initially, the interaction of \textit{Bmp4} and \textit{Fgf8} in the epithelium establishes the location of future teeth along the jaw (figure 4.7). Prior to any morphological signs of tooth development \textit{Fgf8} is expressed across the entire oral epithelium. Subsequently \textit{Bmp4} is expressed in specific regions of the epithelium where it inhibits the activity of \textit{Fgf8}. In regions where it is still active, \textit{Fgf8} initiates tooth development by inducing the expression of several transcription factors, including \textit{Pax9}, in the underlying mesenchyme. Thus, at this stage \textit{Bmp4} acts to repress tooth development where it is expressed (Neubuser \textit{et al.} 1997; Peters and Balling 1999). In this way \textit{Bmp4} seems to act as an inhibitor in the classical reaction-diffusion sense. Later in development BMP4 signal from the mesenchyme has been shown to induce expression of \textit{p21} and \textit{Msx2}, genes important in the establishment of in the enamel knot (Jernvall \textit{et al.} 1998). Expression of \textit{Bmp4} in the enamel knot itself is believed to be important in inducing apoptosis for the eventual removal of the enamel knot (Jernvall \textit{et al.} 1998).

**Downstream targets of BMP4 signaling**

BMP signaling has been shown experimentally to influence the expression of several genes in tooth development largely through \textit{in vitro} tissue culture using BMP soaked acrylic beads.

\textit{Pax9}: The \textit{paired-box-containing transcription factor 9 (Pax9)} is required for tooth development to progress beyond the bud stage and is an early marker for regions of the jaws in which teeth will develop. Its expression is restricted to the dental mesenchyme throughout development. At early stages \textit{Pax9} is induced by FGF8 signal from the epithelium and
antagonized by BMP4. *Bmp4* thus acts as a negative regulator of *Pax9* expression in the mesenchyme (Neubuser *et al.* 1997).

**Msx1**: The *muscle segment homeobox 1 (Msx1)* gene is a homeobox containing transcription factor that is involved in inductive signaling in tooth development by inducing the expression of *Bmp4*, *Lef1*, and *syndecan1* (Chen *et al.* 1996). Its expression is restricted to the mesenchyme throughout tooth development and can be induced in isolated dental mesenchyme by BMP4 soaked beads (Vainio *et al.* 1993). The teeth of Msx1 knockouts fail to develop past the bud stage (Satokata and Maas 1994).

**Msx2**: Closely related to *Msx1*, *Msx2* has a more dynamic expression pattern during tooth development (Jowett *et al.* 1993). Early in development its expression is restricted to the dental mesenchyme and this expression can be induced in isolated mesenchyme by BMP4 (Vainio *et al.* 1993). At the bud and cap stages *Msx2* expression shifts to the epithelium where it is expressed in the enamel knot. This epithelial expression pattern can also be induced by BMP4 in form of protein soaked beads (Jernvall *et al.* 1998). Knocking out *Msx2* in the mouse results in malformed, brittle teeth with almost no enamel; tooth development progresses normally until the bell stage where defects are seen in the enamel organ (Satokata *et al.* 2000).

**Shh**: In tooth development the *Sonic hedgehog (Shh)* gene plays in important role in cell proliferation and cytodifferentiation, and tooth morphogenesis (Dassule *et al.* 2000; Cobourne *et al.* 2001; Gritli-Linde *et al.* 2002). It is expressed exclusively in the epithelium and is first detected prior to tooth initiation. At bud and cap stages *Shh* expression is restricted to the tip of
the invaginating tooth bud and the enamel knot. Later, Shh expression is found across the inner enamel epithelium and in differentiating ameloblasts (Dassule et al. 2000). BMP4 protein applied to the dental epithelium of the Msx1 knockout mouse partially rescues Shh expression in the epithelium, suggesting that BMP4 from the mesenchyme is necessary to maintain Shh expression (Zhang et al. 2002).

p21: The cyclin-dependent kinase inhibitor, p21, is an early marker of the enamel knot. It has been associated with terminal differentiation in muscle cells and stopping cell proliferation in the apical ectodermal ridge (AER) of limb buds. It is has been suggested to be involved in cell differentiation and apoptosis in the enamel knot. In cultured dental epithelial cells p21 expression is induced by beads soaked in BMP4 (Jernvall et al. 1998).

BMP antagonists and tooth development

BMP signaling during development is controlled by extracellular molecules that bind to the BMP ligand preventing it from associating with it cell-surface receptors (Balemans and Van Hul 2002; Canalis et al. 2003; Yanagita 2005). These BMP antagonists help to maintain the precise levels of active BMPs that are required for certain biological processes. For example, Liu and others (Liu et al. 2005) have identified Bmp4 dose-dependent regulation of genes in the developing mandible and in Xenopus and Drosophila different cell types are generated along a BMP activity gradient (Balemans and Van Hul 2002). Many BMP antagonists (e.g. noggin, bambi, and ectodin) are downstream targets of BMP signaling suggesting a complex autoregulatory feedback loop in the control of BMP signal strength (Laurikkala et al. 2003). The roles of two BMP antagonists (follistatin and ectodin) have been specifically studied in teeth.
Figure 4.8: **Teeth of follistatin mutants** (A) Upper and Lower molars of newborn follistatin knockout mouse showing shallow, disorganized cusps. (B) Upper and lower molars of K14-Follistatin transgenic mouse with missing third molar and cusp dysmorphology. Modified from (Wang et al. 2004a).

**Follistatin:** *Follistatin* (*Fst*) was initially identified by its ability to repress follicle-stimulating hormone from the pituitary (Balemans and Van Hul 2002). It binds to activin proteins (TGFβ superfamily members) as well as members of the BMP family (BMP2, BMP4 and BMP7) (Canalis et al. 2003). *Fst* knockout mice die at birth and have abnormal tooth development including absent or severely delayed incisors (Matzuk et al. 1995). Its expression during tooth development has been detected at cap and bell stages where it expressed in the oral epithelium, inner and outer enamel epithelium, and enamel knots (Heikinheimo et al. 1997; Wang et al. 2004a; Wang et al. 2004b). Close examination of the teeth of *Fst* knockout mice and mice over-expressing *Fst* under control of the K14 enhancer has demonstrated its role in ameloblast differentiation in incisors and molar cusp patterning (Wang et al. 2004a; Wang et al. 2004b). Ameloblast differentiation is induced by BMPs and *Fst* prevents ameloblast differentiation by antagonizing the BMP signal. In the incisors of *Fst* knockout mice ectopic ameloblasts form on the lingual surface of the tooth, whereas with transgenic over-expression of *Fst* ameloblasts fail to differentiate on the labial tooth surface. Molars in *Fst* knockout mice have disorganized, shallow cusps and appear to have an abnormal number of large enamel knots (figure 4.8A). *Fst*
over-expression leads to complete agenesis of the third molars and first and second molars that have no clear cusps and prematurely worn enamel (figure 4.8B).

Figure 4.9: Range of phenotypic effects in teeth of ectodin knockout mice (Kassai et al. 2005). Ectodin is a natural inhibitor of BMP signaling during tooth development, knocking it out produces a range of phenotypic effects including supernumerary molars.

_Ectodin:_ Ectodin (Ecdn, also known as uterine sensitization-associated gene 1, USAG1) is a recently identified antagonist of BMP signaling that is highly expressed in teeth and other ectodermal organs and the kidney (Laurikkala et al. 2003; Yanagita et al. 2004). It is expressed in both the epithelium and mesenchyme during tooth development, but expression is absent from the tip of the tooth bud including the enamel knots (Laurikkala et al. 2003). _Ecdn_ expression is a downstream target of BMP signaling and its expression is positively regulated by BMPs (Laurikkala et al. 2003). Other signals from the enamel knot (FGFs, SHH) down regulate _Ecdn_ giving its expression the distinctive negative image of the enamel knots (Kassai et al. 2005). Knocking out the _Ecdn_ gene produces major disruptions of tooth patterning (figure 4.9). First and second molars are often fused and an ectopic molar forms anterior to the first molar. The enamel knots are enlarged and anterior-posterior cusps are fused (Kassai et al. 2005).
MATERIALS AND METHODS

Mice

Use of mice was approved by the Pennsylvania State University Institutional Animal Care and Use Committee (IACUC #27388).

Reporter mice

We obtained a reporter construct from Professor John Rubenstein at UCSF consisting of 4.7 kb of mouse genomic DNA cloned from the 5’ region of the transcription factor Dlx2 including the transcription start site linked to the LacZ gene and SV40 poly(A) sequence (Thomas et al. 2000). This construct was purified and injected into the pronuclei of fertilized eggs of FVB mice (Taconic). Surviving eggs were transferred to the oviducts of pseudopregnant CD1 (Charles River) mice. Embryos were collected at embryonic day (E) 10.5, E12.5, and E14.5. The day of the appearance of the vaginal plug was considered embryonic day 0.5. Embryos were then stained for β-galactosidase activity with X-gal. For X-gal staining, embryos were fixed in 0.25% glutaraldehyde (Sigma) for 30 min, washed in phosphate buffered saline (PBS, Sigma), then placed in a staining cocktail containing 10ml staining solution (1.8mM spermidine, 2mM MgCl2, 0.02% NP40, 0.01% Sodium Deoxycholate, in 1XPBS), 120µl potassium ferro/ferri cyanide, and 50µl X-gal (100mg/ml in N, N-dimethylformamide). Tissue was allowed to stain for 2-3 hours then staining was stopped by washing in PBS and fixing in 10% formalin in PBS overnight. Tissue was washed in PBS and stored in 70% ethanol.
Transgenic construct

We assembled the Dlx2-Bmp4 transgene by modifying the reporter construct (figure 4.10). The transgene is composed of three portions: a 4.7 kb genomic DNA fragment cloned from the 5’ region of the mouse Dlx2 gene, cDNA sequence from the human Bmp4 gene, and the SV40 poly(A) sequence. Using the human Bmp4 cDNA allowed differentiation between endogenous and transgene expression by designing specific PCR primers, described below. The human Bmp4 is amplified from a cDNA clone (Accession # BC020546, www.openbiosystems.com) and cloned using the TOPO TA Cloning Kit (Invitrogen) into the pcrII-TOPO vector. The 5’-primer (GAACGCGTAGCCCGGCGGAAGCTAGGA) is located in the 5’-UTR 160 bp upstream of the translation initiation site (ATG), and the 3’-primer (TGCTAGAGTGGGGTGAAGATGGGAACGTGTG) is 110 bp downstream of the stop codon within the 3’-UTR. The addition of appropriate restriction enzyme recognition sites (MluI and XbaI) at the end of the PCR primers (underlined in the primer sequences) facilitates the unidirectional cloning. The SV40 poly(A) sequence is used because of its strong polyadenylation signal, necessary to terminate transcription.

We first subcloned the reporter construct into a pBluescript SK+ plasmid vector with its BamHI, XbaI, and NotI restriction enzyme sites inactivated. An MluI site was then inserted just upstream of the Dlx2 transcription start site by site directed mutagenesis (ACCCGT -> ACGCGT). The LacZ gene was excised by restriction digestion with MluI and XbaI and the construct containing the Dlx2 enhancer and SV40 poly(A) sequence purified. The MluI-XbaI fragment of Bmp4 could then be ligated into the transgenic construct. All amplified products are confirmed by sequencing to avoid clones with mutations accumulated during the PCR reaction. Our transgenic construct has been specifically designed to allow easy exchange of the transgene
by excision of the \textit{MluI-XbaI} fragment by restriction enzyme digestion and ligation to genes which have been amplified by PCR with the appropriate restriction sites included in the PCR primers.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.10.png}
\caption{\textbf{Constructing the Dlx2-Bmp4 transgene.} Schematic showing modifications of reporter construct to make Dlx2-Bmp4 transgene.}
\end{figure}

\textbf{Generation and genotyping of transgenic mice}

Transgenic mice were generated in the laboratory of Professor Cooduvalli Shashikant in the Department of Dairy and Animal Science at the Pennsylvania State University. The transgenic construct was excised by restriction enzyme digestion (\textit{EcoRI}, figure 4.11), purified...
from the vector sequence by centrifugation through a 10-40% linear sucrose gradient and injected into the pronuclei of one-cell mouse embryo using standard methods (Shashikant et al. 1995; Nagy et al. 2003). Briefly, 15 female egg donor mice of the FVB strain (Taconics) are superovulated with FSH (Follicle-stimulating hormone: 10 units/30 gm body wt) on day one, and then with LH (Luteinizing hormone: 10 units) on day three. These mice are then mated with stud males (FVB) approximately 18 hrs prior to the experiment. At the same time, 20 female foster mice (CD1, Charles River) are mated with vasectomised males to obtain 5-6 pseudopregnant foster mice. On the day of the experiment, the donor females are euthanized, ovariectomized and the fertilized one-cell embryos are collected for the purpose of microinjection of purified DNA fragments in the pronuclei. Surviving one-cell embryos are then implanted into the oviduct of an anesthetized foster animal. This procedure is done by making an incision into the peritoneal cavity and isolating one of the ovaries. The ovary, still attached to the uterine horn, is drawn outside through the incision. The infundibulum is located and the fertilized eggs are drawn into a micropipet, which is inserted into the infundibulum. The eggs are then physically blown into the oviduct, the ovary is returned to the peritoneal cavity and the incisions are sutured.

![Diagram](image)

**Figure 4.11:** **Detail of transgenic construct showing restriction sites** (not to scale).

The founder generation mice are screened for those that carry the transgene by Southern hybridization and PCR on genomic DNA isolated from tail biopsies using the Gentra Puregene
Kit (Qiagen). The following primers were used for PCR: 5’-GCCCGCAGCCTAGCAAGAG-3’ and 5’-GGCAGTCCCCATGGCAGTAG-3’. Founder mice positive for incorporation of the transgene were mated with wild-type FVB mice and the offspring used to check teeth for phenotype and confirm embryonic expression of the transgene. In subsequent generations mice carrying the transgene were identified by PCR.

**Adult teeth**

To examine the teeth, skulls of young adult *Dlx2-Bmp4* mice (4 – 6 weeks) and their non-transgenic littermates were macerated by removing the skin and boiling for one hour in ~100ml distilled water. Muscle tissue and brain were removed using a blunt-end forceps followed by digestion in 1.0% trypsin (Fisher) in a 30% saturated sodium borate solution for 1 - 2 hours. Any remaining soft tissue was removed by rinsing with distilled water. Skulls were then rinsed in ethanol and allowed to dry. Photographs of the molar tooth rows of transgenic and non-transgenic mice taken at 20X magnification in a dissecting microscope were compared directly.

**Transgene expression**

Embryos from crosses between *Dlx2-Bmp4* and wild-type FVB mice were collected at various developmental stages. Embryos were dissected into head, body/tail, and limb sections which were used as follows. Heads were set aside for histology or *in situ* hybridization (see following sections), genomic DNA isolated from the body/tail segment was used to identify embryos carrying the transgene by PCR (as described above), and limbs were placed in RNAlater (Ambion) at -20° to stabilize RNA for extraction and cDNA synthesis. For RNA extraction limb tissue was disrupted and homogenized for 20 – 40 seconds with a Tissue Tearor
(BioSpec Products Inc., Bartlesville, OK). RNA was then extracted using the RNeasy Mini Kit (Qiagen) and first strand cDNA synthesized with SuperScript III reverse transcriptase (Invitrogen) primed with an oligo(dT)$_{17}$ adapter. Transgene expression was confirmed by PCR from first strand cDNA isolated from limb buds using the following primers specific to the human $Bmp4$ cDNA: 5’- GACTGCCCTTTCCACTGGCTGAC-3’ and 5’- TTCCACCACCTGCTCCATTCA-3’.

**Histology**

Dlx-$Bmp4$ embryos and their non-transgenic litter mates collected at embryonic day (E) 13.5 and E17.5; day of discovery of vaginal plug considered E0.5. Heads were dissected in phosphate buffered saline (PBS, Sigma) and fixed in 4% paraformaldehyde (Sigma) overnight, dehydrated to 100% methanol and stored at -20° until use. Heads or dissected jaws were then embedded in Paraplast plus (SPI Supplies, West Chester, PA) and sectioned at 6-7 µm. Tissue sections are stained with Alcian Blue (Sigma), Mayer’s Hematoxylin (Sigma), and Papanicolaou stain (Gill’s modified EA formula, Ricca Chemical, Arlington, TX).

**In situ hybridization**

Mouse embryos were collected at various developmental stages (E11.5, E13.5, E16.5, and E18.5); day of discovery of vaginal plug was considered E0.5. Heads were dissected in phosphate buffered saline (PBS, Sigma), fixed in 4% paraformaldehyde (Sigma) in PBS at 4° overnight, dehydrated to 100% methanol, and stored at -20° C until use. For sectioning, heads were embedded in Paraplast plus (SPI Supplies, West Chester, PA), and sectioned at 6-8 µm.
PCR primers were designed for Bmp4, Msx1, Msx2, Shh, p21, Pax9, Ecdn, and Fst using PrimerSelect (DNAStar, Inc). In situ probes were amplified from cDNA libraries made by reverse transcription PCR from E12.5 or E18.5 embryonic jaw using the following primers: 5’-GCCGCAGCCTAGCAAGAG-3’ and 5’-GGCAGTCCCCATGGCAGTAG-3’ for Bmp4; 5’-GCTGGAGAAGCTAGATGGGC-3’ and 5’-CGGAAGCAGCTGATGGAGTCTC-3’ for Msx1; 5’-ACTGCAAGAGCGGAAGACTG-3’ and 5’-CAGGTCTTAGTGCCTTCACCT-3’ for Msx2; 5’-CCACGCGAGATATGAAGGGAAG-3’ and 5’-GCACGCGGTGGCAAGAG-3’ for Shh; 5’-GAGCAGTTGCGCCGTGATTG-3’ and 5’-GAGGCTAAGGCGCAAGATGG-3’ for p21; 5’-CGCTGCCCTACAACCACATTTA-3’ and 5’-TTGCACGTTGCAAGATGG-3’ for Pax9; 5’-TCTCTCATTCCCTGCTTGC-3’ and 5’-CTTGCTGGCTTGGCTC-3’ for Ecdn; 5’-CTCCGCCAAGCAAAAGAC-3’ and 5’-ACTCGCTGGCGTATGG-3’ for Fst. Gene fragments were cloned using either the TOPO TA Cloning Kit (Invitrogen) into the pcrII-TOPO vector (Bmp4, Fgf8) or the StrataClone PCR Cloning Kit (Stratagene) into the pSC-A vector (Msx1, Msx2, Shh, p21, Pax9, Ecdn, Fst) and confirmed by sequencing. Plasmids were digested using appropriate restriction enzymes and antisense and sense digoxigenin labeled (Dig-UTP) RNA probes were synthesized using T3, Sp6 or T3 RNA polymerases.

Prior to hybridization sections were deparaffinized with ProPar Clearant (Anatech, Inc, Battle Creek, MI), rehydrated in an ethanol series, fixed in 4% paraformaldehyde, permeabilized with Proteinase K (12μg/ml) for 15 minutes and fixed again. Sections were then dehydrated to 100% methanol and allowed to air dry. Hybridization took place overnight at 60° in a humid chamber using a hybridization buffer consisting of 50% formamide, 2XSSC, 10% dextran sulfate, 0.02% SDS, and 100 μg/ml yeast tRNA (Sigma) with probe concentration of 0.5 μg/ml. Non-hybridized probes were washed away in 50% formamide/2XSSC and maleic acid buffer.
with tween (MABT). Sections were treated with anti-digoxigenin antibody conjugated with alkaline phosphatase (Roche) in 1% blocking solution (Roche) overnight and signal visualized with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (NBT-BCIP, Promega).

RESULTS

Confirmation of Dlx2 enhancer activity

To confirm the activity and expression domain of Dlx2 enhancer we injected the Dlx2-LacZ reporter construct into fertilized eggs of FVB mice and collected embryos at various developmental stages (E10.5, E12.5 and E14.5) to check for β-galactosidase activity. X-gal staining (figure 4.12) revealed enhancer activity in the maxillary and mandibular processes of the first branchial arch, fronto-nasal process, whisker follicles, and in the apical ectodermal ridge of the developing limbs as expected based on the published report (Thomas et al. 2000).

![Figure 4.12: X-gal staining showing β-galactosidase activity under control of the Dlx2 epithelial enhancer. Staining can be seen in first branchial arch, fronto-nasal process, whisker buds, and AER.](image)
Production, genotyping, and phenotyping of Dlx2-Bmp4 mice

We used a human Bmp4 cDNA to make the transgene in order to be able to distinguish between transgenic and endogenous Bmp4 by designing PCR primers specific to the human sequence. The nucleotide sequence of the human and mouse Bmp4 genes are 91% identical and the amino acid sequence of the encoded proteins share 96% identity. Despite the slight sequence difference the human Bmp4 gene has previously been used successfully as a transgene in the mouse (Bellusci et al. 1996; Zhang et al. 2000; Zhao et al. 2000a). The human Bmp4 cDNA was cloned into a vector downstream of 4.7 kb of genomic DNA derived from the upstream region of the transcription factor Dlx2 and upstream of the SV40 poly(A) sequence. The Dlx2-Bmp4 transgenic construct was excised and purified from vector sequence and used for microinjection into fertilized eggs of FVB mice. Incorporation of the transgene was confirmed in 11 founder animals by Southern hybridization and PCR (table 4.1).

<table>
<thead>
<tr>
<th>Ear Tag #</th>
<th>Southern</th>
<th>PCR</th>
<th>Expression</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>40</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>45</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.13: **Teeth of Dlx2-Bmp4 mice.** Comparison of teeth of Dlx2-Bmp4 transgenic mice with non-transgenic littermate revealed no overt phenotypic effect of the over-expression of Bmp4. M1, M2, M3, first, second, and third molars respectively.

Figure 4.14: **Histological analysis of teeth.** Sections of Dlx2-Bmp4 transgenic and non-transgenic littermates’ teeth at E13.5 and E17.5 stained with Alcian Blue, Hematoxylin and Papanicolaou Stain. DE, dental epithelium; DM, dental mesenchyme; OC, oral cavity.
Founder animals positive for transgene incorporation by Southern, PCR, or both were mated to wild-type FVB mice to establish transgenic lines and to check the offspring for phenotypes. Examination of the teeth of young adult mice of all 11 transgenic lines revealed no overt morphological phenotypic effect of the transgene (e.g. figure 4.13). Limbs, where the transgene is also expressed, also appeared normal. Histological sections at E13.5 and E17.5 were made to look for transient effects of the transgene during development (figure 4.14). Again, no phenotypic effect could be detected. Animals homozygous for the transgene also showed no overt phenotypic effect and all further analysis was performed on heterozygotes.

**Transgene expression**

Because none of the lines positive for transgene incorporation showed any overt phenotype we needed to establish that the transgene was being expressed. We focused on transgenic founders that were positive for the transgene by both Southern blot and PCR. We first checked for expression of the transgene by reverse transcriptase-polymerase chain reaction (RT-PCR) on RNA extracted from E13.5 limb buds. Of the three founder lines checked by RT-PCR, two were positive for transgene expression (table 4.1). Transgene expression in the developing teeth was confirmed in these lines by in situ hybridization.

Because of the sequence similarity between human and mouse Bmp4 mRNA (91%) the *in situ* probe cross-reacted between transgene and endogenous Bmp4. However we were able to distinguish transgene expression because of the non-overlapping expression domains of Bmp4 and Dlx2 during development. Bmp4 expression is restricted to the mesenchyme by bud stage of tooth development except for expression in the enamel knots and differentiating ameloblasts.
(figure 4.6), where as Dlx2 is expressed in the entire enamel organ throughout tooth development (figure 4.4).

Figure 4.15: Bmp4 expression. Expression of Bmp4 in Dlx2-Bmp4 transgenic mice and non-transgenic littermates at E13.5, E16.5, and E18.5. Endogenous expression of Bmp4 at these stages is restricted to the mesenchyme. Transgene expression can be seen in the epithelium at each stage (arrowheads in E18.5). Separation between epithelium and mesenchyme marked by dotted line. DE, dental epithelium; DM, dental mesenchyme.

We tested Bmp4 expression at E13.5, E16.5, and E18.5 in Dlx2-Bmp4 transgenic mice and their wild-type litter mates (figure 4.15). At E13.5 endogenous Bmp4 expression is seen in
the mesenchyme in both the *Dlx2-Bmp4* mice and their non-transgenic littermates, however clear expression is also seen in the epithelium of the *Dlx2-Bmp4* mice in a pattern that matches *Dlx2* epithelial expression at this stage. Epithelial expression continues at E16.5 and E18.5 confirming that the transgene is in fact expressed. We therefore needed to establish that the transgene was making a functional protein. To do this we looked at the expression of genes known to be downstream of *Bmp4* during tooth development including two antagonists of BMP signaling with known roles in tooth development (*follistatin* and *ectodin*).

**Expression of downstream targets of BMP4**

To establish that a functional transgenic BMP4 protein was being produced we looked at the expression patterns of several genes that have been experimentally shown to respond to BMP signaling in tooth development. We examined the expression pattern of five downstream targets of BMP4 (*Pax9, Msx1, Msx2, Shh,* and *p21*). If the expression pattern of any of these genes was altered by transgenic *Bmp4* expression it would show that the transgene was functional and the lack of phenotypic effect would need to be explained by other mechanisms. However, if no difference was found in the expression of these downstream targets of BMP4 it could mean either that no functional BMP4 protein is being produced by the transgene, or that the BMP4 signal is being regulated following translation of the protein. Because BMP signaling is known to be strictly controlled by several mechanisms including extra-cellular antagonists we also looked at the expression pattern of two BMP antagonists with known roles in tooth development (discussed in following section).
Figure 4.16: Expression of genes downstream of Bmp4. In situ hybridization results for genes downstream of Bmp4 during tooth development.

We looked at the expression of the five downstream targets of BMP4 at E13.5, E16.5, and E18.5 (figure 4.16). In all cases expression appeared the same between Dlx2-Bmp4 mice
and their non-transgenic littermates, no effect of the increased \textit{Bmp4} expression could be detected. Although this is not an exhaustive list of downstream targets of BMP signaling in tooth development and it is possible that the expression of genes not yet identified as targets of \textit{Bmp4} are affected by the transgene, it seems clear that if the transgenic \textit{Bmp4} is producing a functional protein its signal must be somehow restricted. Because of the known role of extracellular antagonists of BMP signaling in tooth development we looked at the expression of \textit{ectodin} and \textit{follistatin} as possible mechanisms for the lack of BMP4 activity.

**Expression of BMP antagonists**

Because we were unable to detect any effect of the transgenic over-expression of \textit{Bmp4} phenotypically or in the expression of downstream targets of BMP4 signaling we tested the expression of \textit{follistatin} (\textit{Fst}) and \textit{ectodin} (\textit{Ecdn}), two BMP antagonists with known roles in tooth development. Expression of many BMP antagonists, including \textit{Ecdn} (Laurikkala et al. 2003) but not \textit{Fst}, have been shown to be direct targets of BMPs controlling the level of BMP signaling in an autoregulatory feedback loop (Balemans and Van Hul 2002). If such a system regulates BMP signaling during tooth development we might expect to see an increase in \textit{Bmp4} expression directly compensated by a concurrent increase in the expression of BMP antagonists. We therefore checked expression of \textit{Fst} and \textit{Ecdn} at E13.5, E16.5, and E18.5 (figure 4.17) to look for differences between \textit{Dlx2-Bmp4} mice and their non-transgenic littermates. Although differences in expression were found, they did not seem to fit a simple feedback loop mechanism for controlling the level of BMP signaling.
**Follistatin**: At E13.5 *Fst* expression appears to be increased in the dental epithelium but not the oral epithelium (figure 4.17). Although this increase of expression is what you would expect if a simple feedback mechanism were controlling BMP signaling in dental development at later stages the picture is not as clear. At E16.5 *Fst* expression in the epithelium appeared to be the same between *Dlx2-Bmp4* mice and their non-transgenic littermates, however mesenchymal expression in the transgenic mice appeared to increase. This increase in mesenchymal expression could be a response to diffusing BMP4 signal from the epithelium. By E18.5 no difference in *Fst* expression could be seen.

![Figure 4.17: Expression of BMP antagonists.](image)

*Figure 4.17: Expression of BMP antagonists. Follistatin (Fst) expression (left) appears increased in dental epithelium at E13.5, but not in oral epithelium (arrowheads). At E16.5 Fst expression in mesenchyme appears to increase (arrowheads) and E18.5 no difference in expression was observed. Ectodin (Ecdn) expression (right) in the epithelium was reduced at all tested stages in the *Dlx2-Bmp4* transgenic mice (arrowheads at E16.5 show inner enamel epithelium). Dotted lines highlight junction of epithelium and mesenchyme.*
*Ectodin:* Surprisingly *Ecdn* expression was decreased in the epithelium at all tested stages (figure 4.17). At E13.5 wild-type *Ecdn* expression is found in the mesenchymal cells surrounding the tooth bud and in the epithelium. In *Dlx2-Bmp4* mice the epithelial signal is lost. At E16.5 expression in the inner enamel epithelium (arrowheads in figure) is lost in the transgenic mice and at E18.5 *Ecdn* expression in the preameloblasts is reduced. This is the opposite effect of what would be expected if *Ecdn* was compensating for increased *Bmp4* in a feedback mechanism and opposite of previous reports showing that *Ecdn* was positively regulated by BMP2 and BMP7 (Laurikkala *et al.* 2003).

**SUMMARY**

The patterning of the mammalian dentition relies on the interaction among several different families of signaling molecules, including BMPs, which control inductive relationships between the oral epithelium and neural crest-derived ectomesenchyme. The expression level of several of these signaling molecules has been shown experimentally to influence dental development fitting the dynamic interaction model of dental patterning. However, to date all attempts to manipulate gene expression of these molecules have resulted in pathologic scale phenotypic effects that are unlikely to be important evolutionarily. We attempted to manipulate the expression level and domain of *Bmp4* in tooth development using an epithelium specific enhancer from the transcription factor *Dlx2*. By using a natural tooth enhancer from a transcription factor gene we hoped to subtly change *Bmp4* expression to produce natural scale variation in the tooth patterning process and test the dynamic interaction model of tooth development in as natural a sense as possible.
The experiment worked in the sense that the transgene was incorporated into the genome and we were able to detect its expression by both RT-PCR on mRNA extracted from embryonic mouse jaw and *in situ* hybridization. We confirmed activity of the the encoded BMP4 protein by examining the expression patterns of genes downstream of BMP signaling in tooth development and antagonists of BMP. But, interestingly, increasing *Bmp4* expression in the epithelium during tooth development had no overt effect on tooth morphology. In fact, the only effect of the transgenic over-expression of *Bmp4* we were able to detect were subtle changes in the expression of the two BMP antagonists, *follistatin* and *ectodin*. Although many BMP antagonists have been found to be regulated by BMP signaling in an autoregulatory feedback system (Balemans and Van Hul 2002), the changes in expression observed do not fit with a simple feedback loop mechanism where increase in *Bmp4* expression would lead to increase in expression of the BMP antagonists.

Many members of the BMP family of proteins are expressed during tooth development and have been shown to have overlapping and sometimes redundant functions. It is therefore possible that treating these genes individually is not the proper approach to determining their role in normal tooth development where slight changes in expression of one member of the gene family may be compensated by changes in expression of the others. To test this, expression of all members of the BMP family would need to be tested quantitatively in the *Dlx2-Bmp4* transgenic animals. In fact, because *Ecdn* has been shown to be positively regulated by BMP2 and BMP7 (Laurikkala *et al.* 2003), the loss of *Ecdn* expression in the epithelium of transgenic animals could be explained by a decrease in expression of *Bmp2* and *Bmp7* in response to increased *Bmp4* expression. Again, quantitative analysis of expression of all members of the BMP family may be able to address this question.
Although this is the first attempt to describe the effect of increased Bmp4 expression in the dental epithelium in vivo, previous studies have reported a similar lack of response of tooth development to excess BMP4 applied in culture (Kassai et al. 2005). Over-expression of Bmp4 under control of the Msx1 enhancer that drives expression exclusively in the mesenchyme also produces no phenotype (Zhao et al. 2000a; Zhang et al. 2002). These results and those of this study seem to point to an extreme robustness of tooth development to over-expression of Bmp4 although the mechanism is by no means clear. There is clearly an effect of Bmp4 over-expression on the expression of BMP antagonists, Fst and Ecdn, but it does not fit with a simple autoregulatory feedback mechanism. It is also possible that the expression of BMP receptors plays a large role in regulating the level of BMP signaling. If the presence or amount of BMPRs is the limiting factor in BMP signaling altering the expression level of the ligand would not be expected to affect phenotype. This could be tested by creating transgenic animals that over-express BMPRs, which could then be crossed with mice over-expressing Bmp4.

Although we were unable to produce variation in the dental pattern by over-expressing Bmp4 under the control of the Dlx2 epithelial enhancer subtle changes in the expression of Fst and Ecdn suggest that there is a complex mechanism for buffering tooth development to increases in Bmp4 expression. Alteration of dental pattern during evolution by upregulating Bmp4 therefore seems unlikely. Conversely, down regulating BMP signaling in teeth by transgenic over-expression of BMP antagonists using the K14 enhancer has a dramatic effect on dental patterning (Wang et al. 2004a; Kassai et al. 2005; Plikus et al. 2005). As a further step to test the function of our Dlx2-Bmp4 transgene we have acquired K14-Noggin mice (Plikus et al. 2005). These mice were made in a C57BL/6J background and over-express the BMP antagonist Noggin producing major disruptions of tooth development and patterning. We have backcrossed
our Dlx2-Bmp4 mice into a C57BL/6J background and will cross them with the K14-Noggin mice to test for any compensatory effects. In a similar manner, Msnl-Bmp4 transgenic mice when crossed with Msnl knockout mice were able to rescue some of the Msnl knockout phenotype even thought they had no phenotype on their own (Zhang et al. 2000; Zhao et al. 2000a; Zhang et al. 2002).
REFERENCES


Jung, H. S., L. Wolpert and D. Davidson (1999). "The formation of the feather pattern in chick skin after a proportion of cells have been killed by X-irradiation." Int J Dev Biol 43(2): 117-23.


morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches." Dev Biol 185(2): 165-84.


Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

INTRODUCTION

The relationship between genes and traits is complex; development involves integrating inputs from genes and the environment, and chance to build a trait or organism. But deciphering this relationship is vital to building an understanding of complex traits and the processes involved in their evolution. By reintegrating development into the study of evolution evolutionary-developmental biology (evo-devo) has brought these questions back to the forefront of modern evolutionary biology and genetics. Beginning to solve complex problems often involves making simplifying assumptions or taking on only the simplest and rarest cases at the extremes of the distribution. The hope is that by starting simple generalities will be discovered that will help explain the complex. Out of necessity, this has been the strategy adopted by geneticists. For most of the past century geneticists have focused on trying to understand simple
mendelian traits or rare debilitating diseases and have produced vast progress in our understanding of genetic processes at both the molecular and population level.

But a danger arises out of the success of the method. One can become entrapped by the relative ease of success and forget that the answers come simply because we chose the easy questions. To date, work on the simple has not led to any great successes in the study of the complex. As geneticists have begun to try to tackle the more common, complex cases the hope of discovering generalities which, like the physical laws, would apply at all levels of complexity, seems less likely. Biology and evolution are fields defined by exceptions, not rules. In evolution every force involves a component of randomness and if there can be said to be a target, optimum fit to the environment, it is a moving one. There are no simple answers to the complex questions of genetics.

The situation just described provides the motivation for the two main goals of the work presented in this dissertation.

**Phenogenetics**

The first objective is to lay out the complexities of the genotype-phenotype relationship focusing on the ways in which this relationship influences variation in traits. Understanding how evolution and development produce phenotypic variation directly impacts the biomedical and pharmaceutical fields, and can also influence social policy. By explicating the molecular, cellular, environmental, and stochastic processes at work during development that produce variation in the genotype-phenotype relationship I hope to temper the widely held view of a simple, deterministic connection between genes and traits. Since the publication of *Phenogenetics: Genotypes, Phenotypes, and Variation* (Sholtis and Weiss 2005), presented here
as chapter 2, many of the topics discussed have seen advancement in understanding and new complexities in the genotype-phenotype relationship have come to light. These will be discussed in a following section.

**Normal variation and dental patterning**

The second main goal of this dissertation is to begin to more directly confront the complexity of the genotype-phenotype relationship by developing methods to study natural variation in traits using the patterning of the mammalian dentition as a model system. The strategy taken is to study natural variation in traits in *unnatural* organisms. Because natural variation in traits likely results from subtle and complex interplay between genes, environment, and chance it is still helpful, and necessary, to reduce the degrees of freedom by using inbred mice or trying to isolate the effects of manipulating a single gene. To begin to dissect the genetic and developmental processes that influence natural variation in teeth I used two approaches, one forward (RI mapping) and one reverse (*Dlx2-Bmp4* transgenics) genetics technique.

**Recombinant inbred mapping:** The first approach aimed to identify genes and developmental processes involved in the production of normal variation in the cusp pattern of teeth. Slight variation in cusp size, arrangement and number is known in humans and other primates and has been studied extensively but little or no progress has been made in understanding the specific genetic or developmental mechanisms that are responsible. This lack of progress is largely due to the difficulties of studying this type of variation in humans and other primates. Here, I identified slight variation in cusp pattern between two inbred strains of mice and was able to
obtain suggestive evidence for the role of three chromosomal regions in the production of the trait.

Because of the genetic and experimental resources available in the mouse I was able to identify all of the genes in these regions with genetic changes that led to amino acid substitutions in their encoded proteins. Two of these, Itga3 and Mrpl10, were known to be expressed during tooth development, although Mrpl10 seems an unlikely candidate because of its function as a structural component of the ribosome. A third gene, Glis3, was not known to be involved in tooth development itself, but is related to the Gli family of Krüppel-like transcription factors which are involved in Shh signaling in the enamel knots. Confirmation that Glis3 is indeed expressed in teeth suggests that it, as well as Igta3 may be good candidates for functional experimental follow-up.

I was also able to identify genes with variation in potential regulatory sites that could affect where and when the gene is expressed. One of these, epiprofin/Sp6, had five SNPs in sites of high regulatory potential found in predicted transcription factor binding sites. This gene is known from experimental knockout to be involved in cell differentiation in developing teeth and tooth patterning. Unfortunately I was unable to identify any variation in epiprofin/Sp6 expression between the inbred mice used. It is, however, possible that expression differences exist at developmental stages that were not checked or at levels below the ability of in situ hybridization to detect.

Finally, this mapping project led to the insight that new methods for identifying candidate genes from chromosomal regions identified by mapping are necessary to avoid the bias inherent in the biomedical literature. By using the GenePaint database, I found that more than half of the genes in the candidate chromosomal regions were expressed in teeth. Nearly all previous
mapping studies have relied on existing literature to identify candidate genes ignoring other genes in the regions, but with the availability of massive databases of expression data the problem of eliminating genes as candidates has become more difficult. Attempts to identify candidate genes will have to consider the fact that more than half of all genes are expressed in a tissue of interest and cannot be eliminated \textit{a priori} as candidates.

\textit{Over-expression of Bmp4}: The second approach to understanding normal variation in the patterning of the dentition used a transgenic strategy to over-express the signaling factor gene \textit{Bmp4} specifically in the oral and dental epithelium. Changes in the level of expression of signaling molecules during tooth development has been shown to affect tooth development fitting the dynamic interaction model of tooth patterning. However, most such work to date has utilized the K14 enhancer which produces high levels of over-expression resulting in major, pathology-scale disruptions of dental development. The goal here was to increase \textit{Bmp4} expression subtly hoping to mimic a natural change by using a natural tooth enhancer from the transcription factor \textit{Dlx2}.

Interestingly, although the experiment worked by all technical measures, no overt phenotype resulted. Thus, the experiment showed that tooth development is robust to slight levels of over-expression of \textit{Bmp4} in the epithelium. The mechanism for this buffering remains unclear. No effect of the transgene was detected in the expression of genes that have previously been shown to be downstream of \textit{Bmp4} during tooth development; however the transgene did affect the expression of two BMP antagonists with known roles in tooth development, \textit{follistatin} and \textit{ectodin}. The changes to \textit{follistatin} and \textit{ectodin} did not however fit a simple autoregulatory feedback mechanism to control BMP signaling as has been suggested in the past (Balemans and
Van Hul 2002). Overall, this experiment showed that development can be buffered in complex ways to over-expression of a signaling molecule like Bmp4, but deciphering the precise mechanism of this buffering will require additional experiments.

PHENOGENETICS: AN UPDATE

New and exciting insight into the relationship between genes and traits has been gained since the publication of my chapter on the role of phenogenetics in the production of variation. Renewed interest in the link between genes and traits pushed by the emerging field of evolutionary-developmental biology has spurred a rapid expansion of research into the phenogenetic relationship. As with any relatively new scientific endeavor more new questions are generated than answers provided and the full impact of many of these discoveries may not be known for some time. Observations of widespread gene expression reminiscent of Wright’s universal pleiotropy like the GenePaint results presented in this dissertation and massive transcription of non-coding DNA from novel transcription start sites (Birney et al. 2007) will continue to make identifying causation of variation in the genotype-phenotype relationship challenging. But it is clear that coming to grips with the complexity of the genotype-phenotype relationship is a prerequisite for understanding the origin and evolution of variation in traits.

Mechanisms influencing the genotype-phenotype relationship

One of the most important mechanisms influencing the relationship between genes and traits is the regulatory control of gene expression. As discussed in the original chapter, the
importance of differential use of genes, rather than protein change, in evolution has been suggested since the mid-70s (King and Wilson 1975). Only recently has technology and our understanding advanced to a point that allows systematic study of the mechanisms involved in gene regulation and it has become a major growth field of study.

Although bioinformatic discovery of gene regulatory elements is still extremely difficult, computation analysis of multispecies DNA alignments have led to the discovery of highly conserved non-coding regions of the genome that are believed to have a role in gene regulation (Dermitzakis et al. 2002; Bejerano et al. 2004). The function of most of these regions remains a mystery but many have now been shown to have enhancer activity in a mouse when linked to a reporter gene transgenically (Pennacchio et al. 2006). An interesting class of conserved non-coding regions termed human accelerated regions have been identified that seem to have undergone positive selection on the human lineage (Pollard et al. 2006a; Prabhakar et al. 2006). These have been suggested to be important in the evolution of human specific traits and one such element has been shown to produce a non-coding RNA that is expressed in the brain (Pollard et al. 2006b). The full significance of the human accelerated conserved non-coding sequences remains unknown but further exploration of these regions may be the most direct test of King and Wilson’s predictions.

The location of many of these conserved non-coding regions in vast gene deserts has led to a renewed interest in long-distance or trans-regulation of gene expression. Mutations in long-range enhancers up to a megabase away from the gene for Sox9 (Wunderle et al. 1998) and Shh (Lettice et al. 2002; Lettice et al. 2003) have been demonstrated to be associated with disease, but until recently the mechanism by which these distal enhancers operate was unknown. By crosslinking chromatin proteins with formaldehyde the 3C (capturing chromatin conformation)
method has shown that distal enhancers come into contact with local promoters by physical looping of chromosomes (Dekker et al. 2002). Once there they modify chromatin structure allowing access of the transcription machinery to the DNA (West and Fraser 2005). The upshot is that variation almost anywhere in the genome may be related to variation in traits and we can no longer focus just on genes and their immediate vicinity.

Another area discussed in the original chapter that has seen vast progress is the role of non-coding RNAs in chromatin structure. I discussed the role of the Xist gene in X-inactivation, but more examples have come to light. Dosage compensation of genes on the X chromosome of the fruit fly, Drosophila melanogaster, has also been shown to be controlled by non-coding RNAs. Male Drosophila have a single X chromosome. A complex of non-coding RNAs including RoX (RNA on the X) binds to the male X promoting histone acetylation and upregulating X-linked genes (Deng and Meller 2006b; Deng and Meller 2006a). In yeast (Schizosaccharomyces pombe), RNAi has been demonstrated to be involved in the formation and maintenance of heterochromatin, important in gene regulation (Folco et al. 2008; White and Allshire 2008).

Understanding the mechanisms that lead from genotype to phenotype requires knowing more than just gene sequence. Gene regulation involves interactions above the sequence level including epigenetic modification of DNA, non-coding RNAs, protein interactions among transcription factors, and chromatin structure and histone modification. To understanding how variation in traits is produced all of these factors need to be considered, but the phenogenetic relationship is not prescriptive and it is not sufficient to understand just these genetic and cellular mechanisms. There is variation in the genotype-phenotype relationship in which the same genes
lead to varying traits and the same traits result from varying genes that cannot be explained mechanistically.

**Variation in the genotype-phenotype relationship**

Variation in the genotype-phenotype relationship can arise through environmental influences or random processes. In the original chapter three main areas were discussed that lead to variation in the phenogenetic relationship: canalization and plasticity, developmental patterning processes, and phenogenetic drift. Interesting new examples and implications of each of these have been discussed in recent literature.

Canalization and plasticity are two sides of the same coin. Canalization refers to a situation where the production of a trait is robust to variation in genes or environment. Plasticity refers to a situation in which variation in a trait arises through environmental factors without variation in genes. These two concepts have been at the center of recent discussions about the importance of studying evolvability, the ability of an organism or species to produce phenotypic variation and adapt to changing environments (Kirschner and Gerhart 1998; Draghi and Wagner 2008). It has been suggested that evolvability evolves (Pigliucci 2008) and even that it is the result of natural selection (Earl and Deem 2004). Hendrikse and colleagues go so far as to say evolvability should be the main focus of evolutionary-developmental biology (Hendrikse *et al.* 2007).

In the original chapter the section that dealt with this topic was titled “A lexicographer’s nightmare” because of the already long list of terms used to refer to the two basic ideas just described. The addition of the term evolvability seems to add little conceptually to the discussion. The buildup of cryptic genetic variation beneath robust traits can be said to produce
evolvability, but this idea is hardly new (Ancel and Fontana 2000; Fontana 2002; Lauter and Doebley 2002). And the idea that phenotypically plastic traits could come under genetic control (the Baldwin effect or genetic assimilation) thus becoming heritable has been discussed since the turn of the last century (Waddington 1942; Simpson 1953; Waddington 1953). Elevating the study of evolvability to a primary goal of an entire field of biological study seems excessive; it is only one aspect of the complex set of interactions that make up the genotype-phenotype relationship.

Developmental patterning processes are the topic of much of this dissertation; they are important to understanding the genotype-phenotype relationship because the resulting trait and variation in that trait are emergent properties of interactions among many factors rather than the direct product of genetic variation. New experimental and computational results have added further support to the role of reaction-diffusion-like processes in the development and evolution of repeated traits. The computer simulations of Salazar-Ciudad and colleagues were mentioned in the original article and have been discussed elsewhere in this dissertation (Salazar-Ciudad and Jernvall 2002). Since original publication they have expanded their model to account for various ways in which variation can arise. They divide pattern formation into inductive and morphogenetic phases and show through simulations how variation in each affects form (Salazar-Ciudad et al. 2003; Salazar-Ciudad and Jernvall 2004). As another example, using a combination of computational modeling and experiment, Sick and colleagues have shown that a reaction-diffusion model also fits the spacing of hair follicles. WNT signaling molecules and their antagonist DKK interact quantitatively to produce the final pattern (Sick et al. 2006).

Finally, a likely new example of phenogenetic drift has been identified in the evolution of the mineralized tissues of teeth. Teeth have been a constant feature of vertebrates since before
the split of cartilaginous and bony fish around 450 million years ago (Reif 1982) but in modern fishes and tetrapods many of the genes involved in their mineralization appear to have evolved separately. Kawasaki et al. have identified a family of genes that code for proteins in milk, saliva, bone, dentin and enamel whose primary function is to bind calcium (Kawasaki and Weiss 2003). Because these proteins are secreted by cells and tend to act in the extracellular matrix they were named the secreted calcium-binding phosphoprotein (SCPP) family. The chemical properties that allow calcium binding appear to be more important than the amino acid sequence of the proteins and little sequence identity is found among SCPP family members. Surprisingly, the SCPP genes of modern fishes and tetrapods seem to be independently derived from a common ancestor (Kawasaki *et al.* 2004; Kawasaki *et al.* 2005). Thus Kawasaki and his co-workers have shown that as long as functional teeth are produced, the specific genes that control their mineralization have been allowed to vary during evolution (Kawasaki *et al.* 2005; Kawasaki and Weiss 2006).

**FUTURE DIRECTIONS**

Attempting to dissect the genetic and developmental process involved in the production of normal variation is, not surprisingly, difficult due to the complex, subtle nature of the variation involved. The results presented here are thus not as definitive as one would hope but there are a number of ways in which future work could expand upon and improve the work presented in this dissertation.
Recombinant inbred mapping

The main issue with any mapping study is producing enough statistical power to detect an association between the trait and marker alleles linked to the causal variation. Power in mapping schemes is most directly influenced by sample size. Increasing the sample size in the case of recombinant inbred (RI) mapping means finding or producing comparable inbred strains created from the same progenitor strains. In the RI mapping study presented here the AXB/BXA RI set comprises 25 independent lines; although this is a relatively large set we were only able to obtain suggestive statistical evidence. To expand the mapping panel a set of recombinant congenic strains (RCS) of mice have been developed from crosses of A/J and C57BL/6J mice (Fortin et al. 2001). This set of mice, designated as AcB/BcA, is comprised of 37 separate lines created by backcrossing the offspring of A/J and C57BL/6J mice to one of the progenitor strains twice and then inbreeding by brother-sister mating. Adding the RCS mice would more than double the number of strains in the mapping study.

The AcB/BcA mice are maintained by Emerillon Therapeutics Inc. a biotech company in Montreal, Quebec, Canada and we have tried to acquire them for three years. If or when we get these mice, adding them to the mapping panel could potentially increase the statistical evidence for the mapping peaks we have found and sharpen them, narrowing the one LOD intervals. Narrowing the chromosomal regions in which to search for candidate genes takes on new importance given the insights provided by the GenePaint analysis. If, as it seems, half or more of any random set of genes will be expressed in any given trait, narrowing the number of possible candidates by other means may become the most crucial stage in any mapping study. It is also possible that adding the RCS mice would wash out the signal found by mapping with RI lines alone. If the causation of the trait is more complex than hoped for it is possible that there are
multiple combinations of alleles even within the limited genomic variation found with the two inbred lines that could produce the trait.

In addition to trying to increase the size of mapping panel, several genes identified by the current study could be followed up to examine their role in the production of normal variation in tooth cusp patterns. Variation in the coding sequence of integrin alpha 3 (Itga3) was discovered too late for this gene to be examined in detail, but the known role of integrins in tooth development suggest that it would be a good candidate for further study. Glis3 also is a good candidate for functional analysis. One possibility in the case of these two genes would be to engineer mice carrying the A/J amino acid sequence in an otherwise C57BL/6J background by homologous recombination.

For epiprofin/Sp6 we identified potential regulatory variation but were unable to detect a difference in expression between A/J and C57BL/6J mice. As mentioned above, if a difference does exist it is possible that in situ hybridization is too blunt a tool to detect it. New technologies that quantify gene expression using high-throughput sequencing instead of hybridization (e.g. illumina digital gene expression, http://www.illumina.com) may be sensitive enough and a more accurate means to detect difference in expression between strains (Marioni et al. 2008). These new techniques are also able to use much smaller tissue samples than microarrays including cells collected by laser capture microdissection (Akiyama et al. 2008).

Finally, the list of potential candidates generated by the GenePaint analysis is enormous. 21 genes with amino acid substitutions between A/J and C57BL/6J mice were found to be expressed in teeth. All of these must be treated as candidates for having a causal role in the production of the tooth trait and could be followed up with expression and functional studies.
Additionally, all 187 genes in the three mapping regions found to be expressed in teeth could be analyzed for potential regulatory variation.

**Dlx2-Bmp4 transgenics**

Over-expression of Bmp4 in the epithelium of developing tooth did not produce the expected effects on tooth morphology. Although this lack of phenotypic effect precluded a direct test of the dynamic interaction model of dental patterning it did demonstrate that tooth development is buffered against the over-expression of Bmp4. The mechanism of this buffering remains unknown and in future work I would like to attempt to work out the molecular details of the system. Using what has been referred to as a ‘systems biology’ approach may be necessary because of the complex network of genes that must be considered in BMP signaling. Expression of Bmps, BMP receptors, BMP antagonists, SMAD transcription factors and other secondary messenger molecules, and finally the downstream targets of BMP would need to be analyzed. Additionally because BMPs signal across tissues it would be necessary to understand the molecular response of both the epithelium and the mesenchyme to excess Bmp4 in the epithelium. Again, the new techniques of quantitative gene expression analysis using high-throughput sequencing could be used (Mardis 2008; Marioni et al. 2008). Expression profiling in teeth using current microarray technology is extremely difficult because of the large amount of tissue needed to obtain the proper quantity of RNA. This requires pooling tissue dissected from a large number of teeth and many individuals which could introduce error into the analysis. New techniques require much less starting RNA (Meyer et al. 2008) and the use of laser capture microdissection would allow specific cells of the epithelium and mesenchyme to be collected and analyzed separately.
Additional future work includes analyzing the ability of over-expression of \textit{Bmp4} to compensate the effects of over-expressing the BMP antagonist \textit{noggin}. Cheng-Ming Chuong at the University of Southern California has kindly provided mice carrying the \textit{K14-Noggin} transgene (Plikus \textit{et al.} 2004; Plikus \textit{et al.} 2005) and we are in the process of crossing these to the \textit{Dlx2-Bmp4} mice.

\textbf{SUMMARY}

Unfortunately, acknowledging the complexities of the genotype-phenotype relationship does not allow one to overcome them. My own attempts to confront this complexity by studying normal variation in the patterning of the mammalian dentition have not produced definitive explanations of the processes that produce natural variation. Genetic mapping of normal variation in the teeth of recombinant inbred mice has identified candidate chromosomal regions on three chromosomes none of which is sufficient to produce the trait on their own. Although a few candidate gene have been identified and warrant further study, hundreds of genes in these regions are expressed in teeth. Over-expression of the signaling factor gene \textit{Bmp4} during tooth development produced no morphological phenotype suggesting that tooth development is robust to excess \textit{Bmp4} but based on the expression changes to BMP antagonists the mechanism for this buffering may be extremely complex.

Incomplete explanations and negative results are frustrating, but they are the reality of trying to deal with complexity. The importance of explaining complex common traits and the production of normal variation for biomedical purposes and our basic understanding of evolution
means that we cannot hide from this complexity or pretend it does not exist. Results like those presented here are necessary to define directions of future study by indentifying ways in which the complexities of the genotype-phenotype relationship manifest in particular traits, and to develop new methods to handle the complexity. There are no universal laws governing the evolution and development of complex traits each of which will have its own unique history, but strategies, like the multiple approaches taken here, can be developed that will begin to reveal these histories.
REFERENCES


VITA

Samuel Jacob Sholtis

Education
2008  PhD  Anthropology  The Pennsylvania State University
2003  MA  Anthropology  The Pennsylvania State University
2001  BA  Anthropology  The Pennsylvania State University

Professional Experience
2008  Research Assistant, Drs. Kenneth M. Weiss and Joan T. Richtsmeier, Cranial Evolution Project, Pennsylvania State University
2007  Instructor, Anthropology 021 – Introduction to Biological Anthropology, Pennsylvania State University (Summer)
       Teaching Assistant, Anthropology 021 – Introduction to Biological Anthropology, Pennsylvania State University
2006  Instructor, Anthropology 021 – Introduction to Biological Anthropology, Pennsylvania State University (Summer)
2005  Teaching Assistant, Anthropology 001 – Introduction to Anthropology, Pennsylvania State University
2004-2006  Teaching Assistant, Anthropology/Biology 460H – Honors Human Genetics, Pennsylvania State University (three semesters)

Publications
Buchanan, A. V., S. Sholtis, J. Richstmeier, K. M. Weiss (in press). What are genes ‘for’ or where are traits ‘from’? What is the question?. BioEssays.

Selected Honors and Awards
American Association of Anthropological Genetics Student Prize for presentation at American Association of Physical Anthropologists 2007 Annual Meeting
Weiss Graduate Scholars Fellowship, Pennsylvania State University, 2002-2006

Professional Memberships
American Association of Physical Anthropologists
American Association of Anthropological Genetics