Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant Emory University and its agents the non-exclusive license to archive, make accessible, and display my dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select access restriction as part of the online submission of this dissertation. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use future works (such as articles and books) all or in part of this dissertation.

_____ _____

Alaine Case Keebaugh Date

Evolution of Purine Metabolism

By

Alaine C. Keebaugh
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Program in Population Biology, Ecology, and Evolution

_____
James W. Thomas, Ph.D
Adviser

_____
David Ledbetter, Ph.D
Committee Member

_____
Soojin Yi, Ph.D
Committee Member

_____
Todd Schlenke, Ph.D
Committee Member

_____
Michael Zwick, Ph.D
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D
Dean of the Graduate School

_____
Date

Evolution of Purine Metabolism

By

Alaine C. Keebaugh
B.S. Biology, University of Georgia, 2001

Advisor: James W. Thomas, Ph.D

An abstract of
A dissertation submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Program in Population Biology, Ecology, and Evolution
2009

Abstract
Evolution of Purine Metabolism
Alaine C. Keebaugh

Gene duplication has long been accepted as playing a major role in evolution. Recently, it has been hypothesized that gene loss might also play an important role in evolution. While this hypothesis has not systematically been tested, it is becoming clear that gene loss has played a role in recent human evolution. One pathway that has been repeatedly modified by both gene duplication and gene loss is purine metabolism. The goal of this dissertation is to use purine metabolism as a model system to explore the biological significance of gene duplication and gene loss with respect to evolution and disease. Through characterization of the purine catabolic pathway in mammals we identified a novel genetic difference between placental mammals and marsupials/monotremes. We hypothesize that this genetic difference distinguishes purine catabolism between these lineages. By performing a more detailed characterization of the purine catabolic pathway in vertebrates we discovered an unexpected fate following loss of UOX in the bird/reptile lineage. The patterns of gene retention and gene loss allowed us to hypothesize a new model of gene loss by which a single gene loss event can have long-term, indirect consequences on the evolutionary potential of many genes in the genome. Mutations in HPRT1, a salvage enzyme of purine metabolism, cause Lesch-Nyhan disease in humans; however, the clinical phenotype has not been successfully recapitulated in a mouse model. In this work, we reconstruct the evolutionary history of the HPRT-gene family and hypothesize that the lack of a disease phenotype in mouse is due to the absence of an HPRT1 paralog, PRTFDC1, in this species. To this end, we have developed a more humanized mouse model carrying the human PRTFDC1 locus. In

this dissertation we provide empirical evidence to support the theory of adaptive gene loss and highlight the importance of understanding the history of gene duplication and gene loss when modeling human disease.

Evolution of Purine Metabolism

By

Alaine C. Keebaugh
B.S. Biology, University of Georgia, 2001

Advisor: James W. Thomas, Ph.D

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Program in Population Biology, Ecology, and Evolution
2009

Acknowledgements

I would like to extend my deepest gratitude to my advisor, James W. Thomas, for his unwavering mentorship, support and technical assistance. I would like to thank my other committee members, David Ledbetter, Soojin Yi, Todd Schlenke, and Michael Zwick for their help and advice. I thank Bob Sullivan for his frequent help with computer programming as well as all the members, past and current, of the Thomas lab for their help, guidance, support and friendship. I would also like to thank the many others who were instrumental in helping me to complete this body of work.

My biggest thank you is to my family. Shane and Foster gave constant love and support. My parents, sister, and extended family were always enthusiastic about my research and nurtured my interest in science over many years.

Table of Contents

1 INTRODUCTION **1**

1.1 Gene duplication **4**

1.1.1 Mechanisms of gene duplication. 4

1.1.2 Physical properties of gene duplication. 6

1.1.3 Classical models of adaptive gene duplication. 8

1.2 Gene Loss **12**

1.2.1 Mechanisms of gene loss. 12

1.2.2 Classic models of adaptive gene loss. 14

1.2.3 Adaptive gene loss. 16

1.3 Purine Metabolism. **24**

1.3.1 Gene duplication. 24

1.3.2 Gene Loss. 28

1.3.3 Human disease. 29

1.3.4 Modeling disease in mouse. 32

1.3.5 Objectives of this dissertation. 34

2 THE GENOMES OF THE SOUTH AMERICAN OPOSSUM (MONODELPHIS DOMESTICA) AND PLATYPUS (ORNITHORHYNCHUS ANATINUS) ENCODE A MORE COMPLETE PURINE CATABOLIC PATHWAY THAN PLACENTAL MAMMALS **35**

2.1 INTRODUCTION **36**

2.2 MATERIALS AND METHODS **41**

2.2.1 Identification of orthologs encoding the purine catabolic pathway 41

2.2.2 Classification of orthologs 42

2.2.3 Sequence Evolution 42

2.3 RESULTS 43

2.3.1 The opossum and platypus genomes encode a more complete set of genes for purine catabolism than do placental mammals 43

2.3.2 Loss of ALLN is specific to placental mammals 47

2.4 Discussion 47

3 EVOLUTION OF THE PURINE CATABOLIC PATHWAY: AN EXPANDED VIEW OF THE EVOLUTIONARY SIGNIFICANCE OF GENE LOSS 51

3.1 Introduction 52

3.2 MATERIALS AND METHODS 57

3.2.1 Identification of orthologs encoding the purine catabolic pathway 57

3.2.2 Classification of orthologs 58

3.2.3 Sequence Evolution 58

3.2.4 Dating 60

3.2.5 RNA isolation and Sequencing 60

3.3 Results 62

3.3.1 Predicted gene content of the purine catabolic pathway in hominoids, birds and reptiles 62

3.3.2 Timing of the gene inactivation events in the hominoid lineage 69

3.3.3 Molecular characterization of the bird and reptile UOX loci 70

3.4 Discussion 72

4 GENE DUPLICATION AND INACTIVATION IN THE HPRT-GENE FAMILY 78

4.1 INTRODUCTION 79

4.2 METHODS 82

4.2.1 Targeted mapping and sequencing 82

4.2.2 Database searches for additional HPRT-gene family members 83

4.2.3 Genomic sequence annotation and analysis 83

4.2.4 Protein domain analysis 84

2.2.5 Phylogenetic analysis 85

4.2.6 Sequencing the mouse PRTFDC1 exons 86

4.2.7 Molecular dating of the mouse PRTFDC1 inactivation event 87

4.3 RESULTS 88

4.3.1 Compilation of the HPRT-gene family data set 88

4.3.2 Phylogeny of the HPRT-gene family 89

4.3.3 Genomic structure of the HPRT-gene family 92

4.3.4 Predicted enzymatic activity 96

4.3.5 Evolutionary rates in the HPRT-gene family 97

4.3.6 Inactivation of PRTFDC1 in mouse 99

4.4 DISCUSSION 101

5 INITIAL CHARACTERIZATION OF A NOVEL MOUSE MODEL FOR LESCH-NYHAN DISEASE 107

5.1 Introduction 108

5.2 Methods 113

5.2.1 Identification and characterization of a human PRTFDC1 BAC clone. 113

5.2.2 Generation of the $Tg^{Prtfdc1+/0}$ transgenic mouse. 113

5.2.3 Breeding & genotyping strategy. 114

5.2.4 Gross phenotyping. 115

5.2.5 Tissue collection. 115

5.2.6 Protein analysis. 116

5.2.7 Expression studies. 116

5.2.8 Statistical analysis. 118

5.3 Results **118**

5.3.1 Relative Transcript Abundance in Human. 118

5.3.2 Generation of the PRTFDC1 transgenic mouse. 119

5.3.3 Gross phenotyping. 120

5.3.4 Protein analysis. 121

5.3.5 Relative Transcript Abundance in Transgenic Mice. 123

5.4 Discussion **125**

6 DISCUSSION **130**

7 REFERENCES **137**

8 APPENDIX **156**

8.1 Supplementary Table 1 **156**

8.2 Supplementary Table 2 **158**

8.3 Supplementary Table 3 **160**

8.4 Supplementary Figure 4 **161**

Index of Figures

1.1 Classic theories of gene duplication 12

1.2 Classical model of gene loss 14

1.3 Synfunctionalization as a mechanism of gene loss 15

1.4 Purine metabolism in vertebrates 25

2.1 Purine catabolism 38

2.2 Gene content of the purine catabolic pathway in opossum and platypus 46

3.1 Gene content of the purine catabolic pathway in vertebrates 67

3.2 Intron/exon structure of UOX 72

3.3 An expanded model of gene loss 76

4.1 Phylogeny of the HPRT-gene family 91

4.2 Genomic structure and sequence composition of the HPRT-gene family loci 95

5.1 Catalytic role of *HPRT1* in purine salvage 108

5.2 Transcript profiles of *PRTFDC1* and *HPRT1* in human tissue 119

5.3 Chromosomal location and genotyping assay for human *PRTFDC1* BAC clone 120

5.4 HPRT1 and PRTFDC1 protein expression in line 9 123

5.5 Transcript profiles of *Prtfdc1* and *Hprt1* in mouse tissues 124

S.1 Protein phylogeny of the HPRT-gene family 161

Index of Tables

1.2 Human genetic diseases associated with purine metabolism 30

2.1 Estimation of Ka/Ks of the genes in the purine catabolic pathway 44

3.1 Estimation of Ka/Ks of coding sequence 64

3.2 Inactivating mutations in the hominoid HIU hydrolase genes 65

4.1 Inactivating mutations in the mouse *PRTFDC1* gene 101

5.1 *PRTFDC1* transmission frequencies in *Hprt1*-deficient males 121

S.1 Sources of HPRT-gene family genomic and cDNA sequences 187

S.2 Gene and genome sizes 189

S.3 Amino acid substitutions in HPRT critical sites 192

1 Introduction

1. Introduction

"From the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of higher animals directly follows. There is grandeur in this view of life, with its several powers, having been breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

Charles

Darwin

On the Origin of Species (1859)

The beauty of organismal complexity has long been appreciated. In the origin of species, Darwin sought to shed light on this complexity by (i) demonstrating the wide kinship of life (i.e. descent with modification) (Hillier et al.) by invoking natural selection as the mechanism responsible for those modifications (Darwin 1859). While Darwin could show that variation existed, and that natural selection could act on such variation, the origin and nature of variation remained unknown. A mechanism for the origin of novel structures was missing and thus many

doubted that natural selection could build new structures from existing variation. The canonical work on the subject is Ohno's *Evolution by Gene Duplication* where he proposes that the genetic redundancy provided by gene duplication is crucial to functional innovation. Ohno's ideas expand upon Darwin's principle of natural selection in that gene duplication is one type of genetic variation that natural selection can act on. When Ohno first proposed his ideas on the significance of gene duplication it was not technically possible to test many of them rigorously. Today, gene duplication is widely accepted as the primary mechanism for the acquisition of new functions as well as the ultimate material for evolutionary innovation. In contrast to this adaptive view of gene duplication, gene loss is generally associated with deleterious consequences. Genomics is revealing that genomes are more flexible and variable than expected and that evolution can work both ways, taking advantage of both gene duplication and gene loss. While the relative contribution of gene loss to organismal complexity is currently unknown, it is becoming increasingly clear that gene loss has played a role in adaptive evolution among contemporary human populations. With the advancement of genomic resources the connection between gene loss and organismal complexity is beginning to emerge.

1.1 Gene duplication

With the accumulation of complete genomic sequences over the last decade, it has become realized that significant portions of eukaryotic genomes consist of duplicated gene loci, that most gene families are large and have complex evolutionary histories, and that these are best explained by the widespread occurrence of whole genome duplications or tandem gene duplications (Aburomia, Khaner, and Sidow 2003). And, as predicted by Ohno over thirty years ago gene duplication and subsequent divergence of these duplicated genes is central to functional innovation.

1.1.1 Mechanisms of gene duplication.

Ohno suggested that whole genome duplications were likely more important than regional duplication in shaping the genome of early vertebrates (Ohno 1970). This assumption was based on isozyme complexity and genome size. While these measures were relatively inaccurate, more contemporary molecular studies have yielded convincing evidence favoring two rounds (1R/2R) of large-scale duplication events in early vertebrates. Specifically, genes in contemporary vertebrate species can be traced back to the base of the vertebrate phylogeny with a clear pattern of tetra-paralogy emerging (Blomme et al. 2006). In addition,

comparative genomic studies have found that many genes seem to have two copies in ray-finned fishes but only one cognate copy in other vertebrates. Most convincingly, Jaillon et al (2004) compared the distribution of genes between the human and *Tetraodon*, observing many cases where human syntenies were found in duplicate on two different *Tetraodon* chromosomes (Jaillon et al. 2004).

Many have speculated that whole genome duplications (WGD's) allowed for the diversification of vertebrates (Aburomia, Khaner, and Sidow 2003). Similarly, a fish specific WGD's may have contributed to the contemporary diversity of ray-finned fishes (Van de Peer 2004). In addition to the importance of WGD in vertebrate diversification, WGD's are associated with speciation in fly, worm and plants (Greenberg et al. 2003; De Bodt, Maere, and Van de Peer 2005; Scannell et al. 2006). Whole genome duplication is clearly a fundamental aspect of adaptive genome evolution.

In addition to these WGD's, smaller scale duplications of segments of chromosome (segmental duplications) or the transcribed portions of gene by retrotransposition are continuously occurring in the genome. The initial draft sequence of the human genome revealed an unexpectedly large number of segmental duplications as compared to other species such as worm and fly (Bailey et al. 2002). And, nearly half of the human genome, ~42%, is composed of

retrotransposed genes (Lander et al. 2001a). Microbial, fly and worm genomes have significantly fewer pseudogenes than human (Balasubramanian et al. 2009).

Gene duplication can also lead to copy number variation (CNV) between individuals, which can include whole genes and clusters of genes. Recent data from humans has shown that gene dosage differences caused by CNV's may underlie many aspects of human phenotypic variation and provide the raw material for gene family expansion and contraction (Perry et al. 2007; Xue et al. 2008).

1.1.2 Physical properties of gene duplication.

The presence of duplicate genes leads to unique physical properties that are subject to homology-driven evolutionary processes that facilitate subsequent duplication, deletion and inversion events (non-allelic homologous recombination (NAHR) such as reciprocal unequal crossing-over). Chromosomal rearrangements mediated by NAHR between duplications typically occur between segmental duplications with greater than 95% identity and are more than 10 kb in length. A second homology driven mechanism, gene conversion, is the process by which one copy exerts its sequence onto the other copy. Inverted repeats (Stefansson et al.) make up one class of segmental duplications whose homology

can be maintained through gene conversion. Studies have shown that IRs can contribute to gene regulation by formation of secondary structures that can modify chromatin structure (Warburton et al. 2004). IRs can also contribute to function beyond that of the gene in which they encode (Hanke, Hambor, and Kavathas 1995).

Site-specific recombination, like NAHR, can lead to inversion events and has been shown to occur at high frequencies in bacteria. This type of recombination is generally 'regulated' by physical properties and can provide a simple ON ↔ OFF switch for genes located within or adjacent to the invertible region (Abraham et al. 1985). One of the best-characterized examples of variation achieved by site-specific DNA inversion are those of type 1 fimbriae from *Escherichia coli*, where phase variation is mediated by the inversion of a 314 bp chromosomal region. In the 'ON' orientation, the promoter is in the correct position to initiate transcription and in the 'OFF' position transcription is inhibited (Abraham et al. 1985). It is generally believed that the most likely role for phase variation is to provide bacterium with a strategy for adapting to changing or multiple environmental conditions (Saunders 1994).

In some cases NAHR mediated rearrangements are highly disadvantageous and cause disease. Many genomic illnesses due to gene duplication are known while

only a handful of non-pathogenic and even less adaptive gene duplications have been identified. One obvious question is why are these physical properties maintained if they cause disease? An interesting example that might shed light on this enigma involves a 900-kb inversion polymorphism of chromosome 17q21.31 that is associated with segmental duplications (Stefansson et al. 2005). This particular inversion represents two distinct haplotypes that have been diverging for approximately 3 million years. One haplotype is found at a frequency of 20% in Europeans and indicates a history of positive selection such that carrier females have higher fertility rates than non-carriers. In contrast to its beneficial effect, microdeletions caused by NAHR between segmental duplications within this polymorphism have been identified in ~1% of European individuals with mental retardation (Sharp et al. 2006). To date, this is the only example where segmental duplication architecture predisposes a species to both positive and negative selective pressures. It is thus likely that the adaptive benefits of gene duplication counter balance the deleterious consequences, allowing for the creation of novel gene families and genetic and phenotypic variation.

1.1.3 Classical models of adaptive gene duplication.

The idea that gene duplication plays a fundamental role in the origin of phenotypic diversity has motivated numerous proposals to explain how a new

gene can emerge from its predecessor, and evolve a novel function. One of the earliest reports of gene duplication was in the BAR 'gene' of the fruit fly *Drosophila melanogaster* (Bridges 1936). As early as the 1930's geneticists, such as Haldane, speculated on the importance of gene duplication in evolution but its magnitude was not fully appreciated until the middle of the century.

In his seminal book, *Evolution by Gene Duplication*, Ohno introduced what is considered the classical models of gene duplication (Figure 1.1). Ohno proposed that the most likely fate of a duplicate gene is gene loss but in some cases a duplicate gene can be retained in the genome due to beneficial effects of gene dosage or neofunctionalization.

Ohno proposes that a duplicate gene can be retained due to the beneficial effects of increased gene dosage (producing more of a given protein). Under this model, the duplicates undergo concerted evolution and remain ‘identical’ to each other. Alternatively, because gene duplication creates a redundant locus that, being unencumbered by a functional role, is free to accumulate formerly forbidden mutations and eventually acquire a new function (Ohno 1970). This model is called neofunctionalization and was identified by Ohno as the primary mechanism for the preservation of duplicate genes. Interestingly, empirical data suggest that a

much greater proportion of gene duplicates are preserved than is predicted by the classical model (Lynch and Conery 2003).

Currently the mechanism of retention is an issue of debate in the literature. One key question is the relative importance of neofunctionalization and positive selection versus that of purely neutral processes. Hughes and Lynch (1994) proposed that preservation of a duplicate gene could be due to the accumulation of degenerative mutations causing complementary loss of function in the members of the pair. Thus, prior to gene duplication a single gene performs two functions and following duplication the ancestral function is partitioned between the duplicate genes so that each member of the pair performs a distinct function (Figure 1.1). Subsequently, an alternative subfunctionalization model was proposed, the duplication-degeneration-complementation (DDC) model. In this process, the preservation of duplicate genes is achieved through partitioning of the ancestral expression pattern rather than biochemical functions (Force et al. 1999). Both models of subfunctionalization are limited because they do not account for the acquisition of a novel function and they require the existence of a bifunctional ancestral state (Figure 1.1).

The debate over the relative importance of neofunctionalization and positive selection versus that of subfunctionalization and neutral processes has lead to the

emergence of hybrid views, such as neosubfunctionalization (He and Zhang 2005). Under neosubfunctionalization, rapid subfunctionalization occurs immediately following gene duplication and is accompanied by prolonged and substantial neofunctionalization. Like the model of subfunctionalization, this view assumes that the ancestral state had more than one function. This view highlights that while gene dosage, neofunctionalization and subfunctionalization are presented as mutually exclusive fates for duplicate gene pairs, they are not. Following retention, duplicate genes continue to evolve, and gene dosage and/or subfunctionalization could contribute to novelty simply by enabling duplicate genes to survive for long periods, increasing the chances of neofunctionalization.

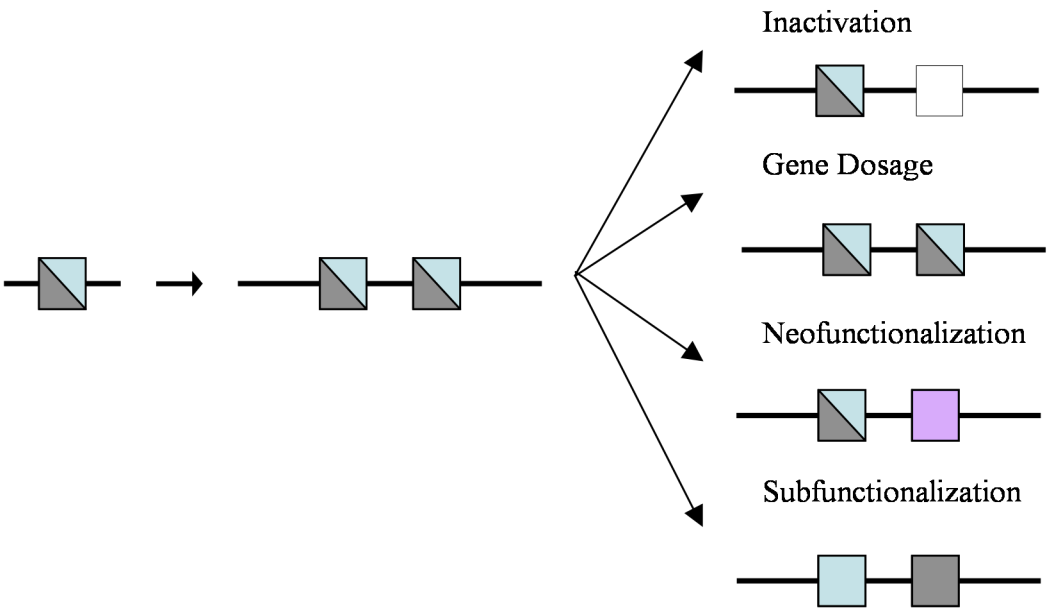


Figure 1.1 Classic theories of gene duplication.
Cartoon depicting the fate of a duplicate gene: inactivation, dosage compensation, neofunctionalization, and subfunctionalization, respectively. Colored blocks represent functional gene/s or regulatory domain/s. White boxes represent inactivated gene/s.

1.2 Gene Loss

1.2.1 Mechanisms of gene loss.

Loss of gene function can be achieved through loss of promoter activity, protein-truncating mutations, and by more subtle loss of function mutations. Given the type of mutation incurred and the time since the mutation originated, it can be difficult to identify the initial loss of function mutation. This is particularly true for regulatory regions because they are diverse and evolve rapidly. In contrast,

protein-coding mutations are more easily identified because of their high conservation and our greater understanding of these regions. Once a gene acquires an inactivating mutation, it will begin to evolve neutrally and genetically deteriorate. Thus, though there are many examples of gene loss, which mutation caused the initial gene inactivation event is often not known.

There are also circumstances where many genes might become inactivated due to unique genomic physical properties. Genome rearrangements, for example, are known to have occurred during eutherian evolution (Moses et al. 2003) and can result in gene deletion (Fitzgerald and Bateman 2004). For example, in a study by Fitzgerald and Bateman four human genes were identified that correspond to evolutionary breakpoints in rodents and ortholgos in mouse and rat were not found, supporting 'evolutionary breakpoint-associated gene deletion' as a mechanism of gene loss. Recently, sequencing of gibbon breakpoints revealed three genes disrupted by rearrangement that showed signatures of selection consistent with pseudogenization (Girirajan et al. 2009). Under this model gene loss can potentially disrupt developmental and biochemical processes or contributes to morphological and behavioral specialization and so fixation can be due to either neutral or beneficial gene loss.

Large-scale gene loss can be mediated by lack of recombination, which can lead to genetic degeneration and gene loss. The most classic example of genetic degeneration can be found in the sex chromosomes where the Y chromosome (or W chromosome) cannot recombine and has subsequently experienced substantial degeneration (Charlesworth and Charlesworth 2000). Gene loss by this mechanism is not restricted to the sex chromosomes, but has also occurred in the t-complex on mouse chromosome 17 as well (Ardlie 1998).

1.2.2 Classic models of adaptive gene loss.

In the "less is more hypothesis" Maynard Olson (1999) proposes that loss of gene function may play an important role in populations undergoing a change in the environment and consequently a change in selective pressure. Olson's idea that gene loss may be an adaptive mechanism of evolutionary change is counterintuitive because better fitness is generally associated with the acquisition of novel functions not broken ones. Nonetheless, there are several examples from the human population where loss of function improves fitness in particular environments (e.g. malaria and hemoglobinopathies).

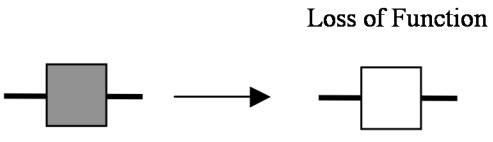


Figure 1.2 Classical model of gene loss. Loss of gene function occurs through the accumulation of a deleterious mutation/s. Clolored blocks represent protein funciton/s or expression pattern/s. Empty boxes represent loss of protein function or expression pattern.

In contrast to gene duplication, there has been little theoretical discussion on adaptive gene loss. Under the classical model, gene loss occurs when a functional gene acquires an inactivating mutation and becomes fixed in the population (Figure 1.2). Fixation of a null allele within a population can be due to random genetic drift, hitchhiking, or, as Olson suggests, natural selection. When gene loss is adaptive, the loss of function mutation is expected to occur frequently and spread rapidly through small populations.

A second model of gene loss, called synfunctionalization, was recently proposed (Figure 1.3). Under this model, loss of function in one paralog is preceded by the acquisition of the function in a different paralog. In contrast to the classical model of gene loss, synfunctionalization allows for a gene to be lost without its function being lost (Gitelman 2007); however, it requires that a paralog be present. Evidence for this model lies in the observation that within gene families among vertebrates, one function is executed in different species by different orthologs (Nguyen et al. 1998; Scholpp and Brand 2001; Locascio et al. 2002). Gene loss through synfunctionalization is likely not due to adaptive evolution because gene function is maintained.

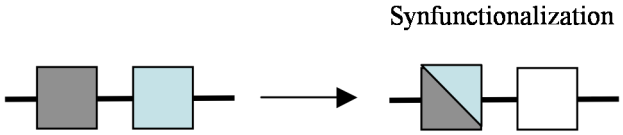


Figure 1.3 Synfunctionalization as a mechanism for gene loss. Under the model of synfunctionalization, gene loss can occur in a paralog when a different paralog acquires its function (Gittleman, 2007). Colored blocks represent protein function/s or expression pattern/s. Empty boxes represent loss of protein function or expression pattern

1.2.3 Adaptive gene loss.

Gene loss is an evolutionary phenomenon whereby a gene losses its function. Once a null allele arises in a population, it is generally considered to be detrimental to an organism and selectively disadvantageous. However, if a non-essential gene is inactivated or environmental conditions change there may be no selective disadvantage and a gene loss can be selectively neutral (Tanaka, Tateno, and Gojobori 2005). Fixation of such a null allele could be due to genetic drift, even when the allele is slightly deleterious. In the 'less-is-more' hypotheses, Olson proposes that some gene loss events may be selectively advantageous in populations undergoing a change in environment (Olson, 1999). Adaptive gene loss has been hypothesized in a number of species and has been associated with changes in diet (Li et al. 2005), morphology (Protas et al. 2006), and pathogen resistance (Haller, Acklin, and Staeheli 1987). Gene loss has also been implicated

in driving the anatomical and physiological differences between monotremes and other vertebrates (Brawand, Wahli, and Kaessmann 2008; Ordonez et al. 2008).

The idea that gene loss can be adaptive has become particularly interesting with respect to human evolution. Rapid population growth coupled with changes during the late Pleistocene and Holocene created new opportunities for adaptation. During the past 10,000 years human populations experienced rapid skeletal evolution and the appearance of many new genetic responses to diet and disease (Harper and Armelagos 2005). There are a number of human specific gene losses hypothesized to be associated with diet, morphological evolution, reproduction, and disease (Table 1.1).

Table 1.1 Hypothesized adaptive gene loss within the human population.

Gene Loss	Potential Biological Significance	Evidence for Adaptation[1,2,3,4,5]	Reference
UOX	Diet, disease, cognition	1,2,3,4,5	(Haldane 1955; Orowan 1955; Ames et al. 1981; Wu et al. 1989; Wu et al. 1992; Watanabe et al. 2002; Johnson et al. 2005)
AMY1	Diet	1,4	(Perry et al. 2007)
UGT2B17	Diet, metabolism	1	Xue et al 2008 p. 337
ACTN3	Metabolism	3,5	(North et al. 1999; Niemi et al. 2005; MacArthur et al. 2007; Yang et al. 2007)
CoxII	Metabolism, morphology	1,4	(Goldberg et al. 2003)
MYH16	Morphology	4	(Stedman et al. 2004)
YhHaA	Morphology	1	(Langbein et al. 2001; Winter et al. 2001)
ADAM1	Reproduction	n/a	(Jury et al. 1997)
SERPINA2	Reproduction, disease	1,2	(Seixas et al. 2007)
ELN	Disease	n/a	(Szabo et al. 1999)
CCR5Δ32	Infectious disease	1,4	(Galvani et al. 2005; Novembre et al. 2005)
G6PD	Infectious disease	1,4	(Ruwende et al. 1995; Tishkoff et al. 2001)
GPR33	Infectious disease	2,3,4	(Rompler et al. 2005)
FYO	Infectious disease	1,4	(Hamblin et al. 2002)
Caspase-12	Infectious disease	1,4	(Wang et al. 2006; Xue et al. 2006)
CMAH	Infectious disease	1,4,6	(Chou et al. 1998; Hayakawa et al. 2001; Varki 2001; Chou et al. 2002)
EMR4	Infectious disease	n/a	(Hamann et al. 2003; Kwakkenbos et al. 2004)
TCR ß	Infectious disease	n/a	(Meyer-Olson et al. 2003)
BASE	Infectious disease	n/a	(Hahn 2005)
MBL1	Infectious disease	2,5	(Sastry et al. 1995; Seyfarth et al. 2005)
GSTA4	Infectious disease		(Morel et al. 2002)
NRADD	Infectious disease	n/a	(Wang et al. 2003)
Nephrocan	Infectious disease	n/a	(Mochida et al. 2006; Matsushima et al. 2007)
CFTR ΔF508	Infectious disease	1,4	(Schroeder et al. 1995)
GGTA1	Infectious disease	1	Koike 2002
ABO	Infectious disease	1	(Yamamoto et al. 1990)
RT6	Infectious disease	n/a	(Haag et al. 1994)
APOBEC3H	Infectious disease	2	OhAinle 2008
Caspase-15	Infectious disease	2	Ekhart et al. 2006
FLJ33674	?	n/a	(Hahn 2005)
WBSCR27	?	n/a	(Hahn 2005)
DNAJB3	?	n/a	(Hahn 2005)
T2R62P	?	n/a	(Hahn 2005)
ZCCHC13	?	n/a	(Hahn 2005)
SULT1D1	?	n/a	(Meinl et al. 2001)
5-HT(5B)	?	n/a	(Grailhe et al. 2001)
GULOP	?	n/a	(Nishikimi et al. 1994; Ohta et al. 1999)
Centrin4p	?	n/a	(Gavet et al. 2003)
Acyl3	?	n/a	(Zhu et al. 2007)
Tex21	?	n/a	(Zhu et al. 2007)
Gucy2d	?	n/a	(Zhu et al. 2007)
Gm766	?	n/a	(Zhu et al. 2007)
ELA1	?	n/a	(Rose et al. 1997)
HPP	?	n/a	(McEvoy et al. 1988)
FMO2	?	n/a	(Dolphin et al. 1998)

Numbers indicate evidence supporting adaptive hypotheses:[1]Population genetic signature indicative of a selective sweep;
[2]Multiple independent gene loss events within or between species; [3]Loss of long-established genes (>50MYA);
[4]Inactivation coincides with environmental change; [5]In vivo studies support adaptive hypotheses.

Dietary changes. It has been proposed that specific loss of function mutations have provided a survival advantage to early primates by helping maintain blood pressure during periods of dietary change. Loss of UOX has occurred twice, once in the gibbon and once in the great apes. It has been proposed that loss of UOX was advantageous due to maintenance of blood pressure during the Paleolithic Period when sodium content was very low (Watanabe et al. 2002). Similarly, it is hypothesized that loss of $_{L}$-gulonolactone oxidase (GLO) occurred during the Eocene, a time of high vitamin C intake, and was partially advantageous because, in contrast to its synthesis, dietary intake of vitamin C does not generate oxidants. Subsequently, when climatic conditions changed, resulting in famine or starvation, lack of GLO resulted in oxidative stress and maintenance of high blood pressure (Johnson et al. 2008). In modern human populations, characterized by "western" diets and decreased physical activity, these once advantageous mutations are now associated with contemporary diseases such as hypertension and gout.

Morphological impacts. Human specific gene loss events can explain some of the morphological differences between humans and other great apes. Loss of a myosin heavy chain null allele (MYH16) is an example of gene loss that might be responsible for the marked size reduction in hominin masticatory muscles.

Population genetic, phylogenetic and morphological data suggests that this null allele has been under strong positive selection in the recent past and is hypothesized to have allowed for brain size expansion in contemporary humans (Stedman et al. 2004). Interestingly this gene is still transcribed, likely due to the fact that it is a recent inactivation event and is still in the degradation process. In another example, human type I hair keratin was lost in human but remains functional in chimpanzee and gorilla (Winter et al. 2001). This loss may represent differences in body hair patterns between humans and their great ape relatives.

Immune response. Human specific inactivation is frequent among genes that function in immune response. Within the human population many well known examples include null alleles that are segregating at high frequencies in distinct geographical areas (e.g. the chemoattractant receptors CCR5 and HIV1 infection, Duffy blood group locus FY, malaria and capsase12 that provide resistance to disease. Given the high disease prevalence associated with these regions, these inter-population differences in the frequency of segregating null alleles suggest population differences in encounters with certain pathogens. In these examples it is not clear if null alleles are less fit than the functional counterpart when the pathogens are rare or absent. These classic examples are clearly adaptive but they are not likely to become fixed.

There are several examples where inactive alleles are fixed in the human population (Table 1.1). GPR33, a gene that arose 120~190 million yrs ago, is present in eutherians and marsupials. This gene has been maintained in various mammalian species for a long time, however, multiple independent losses are observed in both the rodent (n=5) and primate (n=4) lineages, indicating that selection has acted on this locus multiple independent times (Rompler et al. 2005). It is speculated that these groups might be exposed to similar pathogen pressures. Unlike the Caspase12 null allele, population genetic analyses were unable to detect signatures of selection. These examples suggest a shift in pathogen infection and susceptibility in the human lineage since divergence from its great ape ancestor.

The rapid growth of human populations over the past millennia has also been implicated in driving pathogen virulence, with respect to human health, through gene loss. The bacterium responsible for whooping cough recently underwent dramatic gene loss, removing its ability to infect hosts other than humans but making it more efficient at infecting humans (Parkhill et al. 2003). Thus we see that gene loss in response to host-pathogen interactions is not unique to humans.

Unknown biological significance. There are also null alleles in the human population that have signatures of positive selection but no physiological benefit has been documented by the loss of function. For example, the ACTN3 null allele is segregating at a frequency of 10-15% in different human populations. Homozygotes for the null allele are overrepresented in endurance runner athletes and mice deficient for ACTN3 can run 33% further than wildtype mice (MacArthur et al. 2007). While population genetic analysis indicates selective sweeps for the null allele, the direct physiological benefit to humans remains unclear. In another example, the SERPINA2 null allele is segregating at very high frequencies in certain populations and exhibits population genetic signatures indicative of a selective sweep (Seixas et al. 2007). While expression patterns indicate this gene may have been involved in reproduction, loss of SERPINA2 may represent differences in host-pathogen interactions with respect to the reproductive tract. Again, the direct physiological benefit remains unclear.

While traditional population genetic parameters can be useful for detecting the molecular signatures left in the aftermath of a selective sweep, these methods are limited due to rapid decay of the genetic sequence as well as the limited availability of genomic resources. In addition to population genetics, other indicators can be useful for invoking adaptive hypotheses. For example, theoretical work predicts that duplicate genes have a half-life of 5MY (Lynch and

Conery 2003). Thus, loss of "old" genes, >50 MYO, is more likely to be due to selection than chance (Zhu et al. 2007). Furthermore, multiple independent losses are less likely to be due to chance than single events alone. Phylogenetic dating can be useful for dating the emergence of a null allele and can be indicative of selection when dating coincides with environmental, morphological and/or dietary changes. Additionally, *in vivo* experiments can be useful for modeling the functional effects of gene loss.

Finally, because gene inactivation is the most likely fate for duplicate genes, distinguishing between adaptive gene loss and gene loss due to redundancy is hard when the gene is a member of a large gene family. Within the human population alone, Menashe et al (2003) found 178 different combinations of functional and nonfunctional odorant receptors at 51 loci (Menashe et al. 2002). It is unclear if the intra-population and inter-population variation differences are shaped by random inactivating mutations and genetic drift or if they are in part due to differential natural selection in different geographic regions and environments. While table 1.1 lists many of the potentially adaptive human specific gene loss events, it is not a comprehensive list of all gene loss events as it excludes those associated with large gene families such as the olfaction-related genes.

1.3 Purine Metabolism.

Purine metabolism is an essential biological process that synthesizes and breaks down the purines adenine and guanine, which are precursors of DNA and RNA and frequently participate as coenzymes, second messengers, cell signaling molecules, and are critical for energy metabolism (Figure 1.4). Purine metabolism is a highly conserved pathway found among all domains of life, and is perhaps the oldest biochemical pathway on the planet (Caetano-Anolles, Kim, and Mittenthal 2007). However, despite its conservation and antiquity, purine metabolism has been repeatedly modified by gene duplication and gene loss throughout vertebrate evolution. Both gene duplication and gene loss have played important roles in the adaptive evolution of purine metabolism among vertebrates, leading to functional innovation and diversification between species.

1.3.1 Gene duplication.

Ancient duplication events are hypothesized to have played a large role in the establishment and diversification of purine metabolism (Horowitz 1945). Contemporary gene duplication has continued to shape the purine metabolic pathway between vertebrates, resulting in the preservation of duplicate genes through (i) gene dosage, (Hillier et al.) neofunctionalization, and (iii) subfunctionalization.

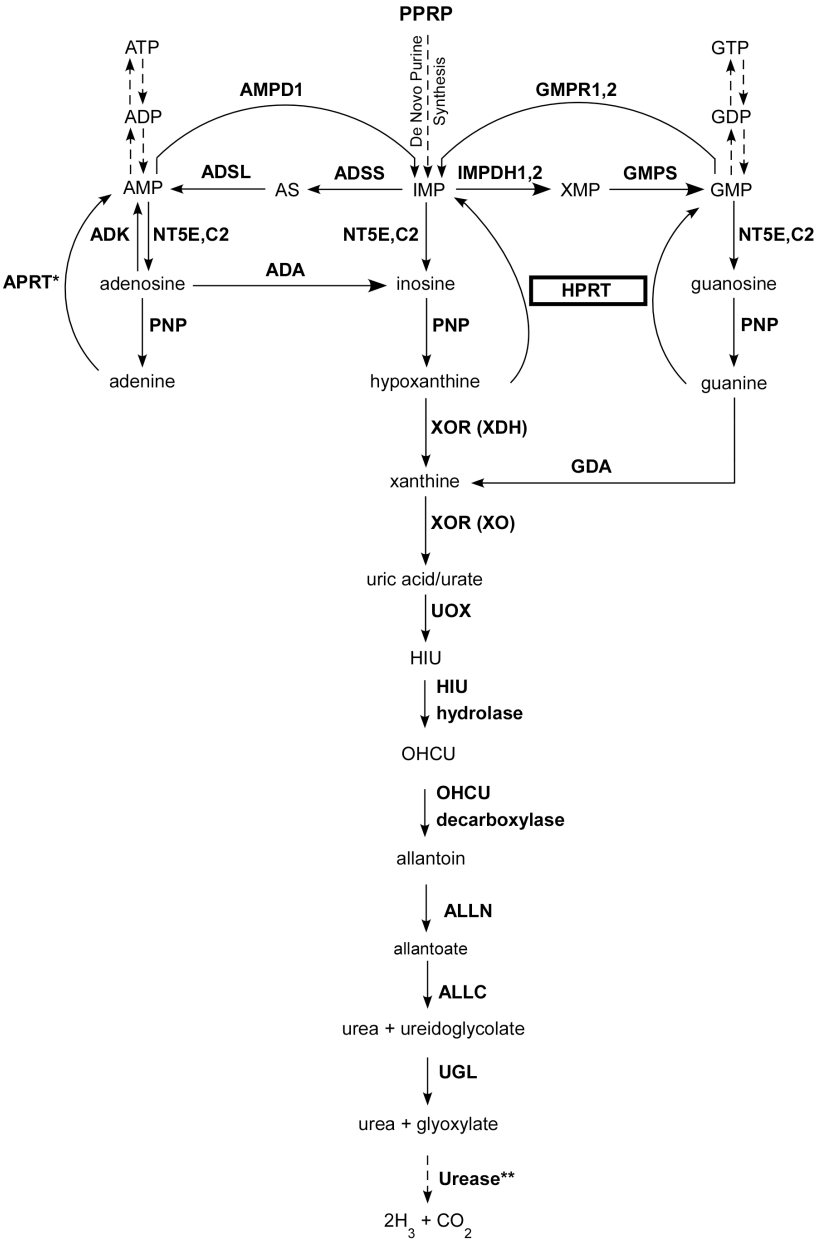


Figure 1.4 Purine metabolism in vertebrates.
Solid arrows indicate reactions in the pathway of specific interest to this review. Dashed lines indicate other portions of the pathway. Urease (**) activity is present in the gut of some fish.

Gene Dosage. The beneficial effects of gene dosage have led to the retention of duplicate genes within the purine metabolic pathway. For example, two distinct forms of human GMP reductase have been identified, GMPR1 and GMPR2. Both are able to reduce GMP (Zhang et al. 2003) and have overlapping expression patterns (Deng et al. 2002). Thus, it seems reasonable to assume that these duplicate gene pairs have been preserved by selection for increased gene dosage, as predicted by Ohno.

Neofunctionalization. Gene duplication has also led to the acquisition of novel functions not associated with purine metabolism. For example, following duplication of HIU hydrolase in the vertebrate ancestor, fish, amphibians, and mammals, retained the ancestral enzymatic activity associated with purine metabolism while its paralog, TTR, participates in hormone and vitamin transport (Zanotti et al. 2006). This example represents a dramatic functional shift between two closely related paralogous proteins and highlights the role of gene duplication in the acquisition of novel functions within this pathway.

Subfunctionalization. Purifying selection has also preserved duplicate genes within this pathway by partitioning the ancestral function among daughter copies. For example, in the duplicate gene trio, AMPD1, AMPD2, and AMPD3, all have

the same enzymatic function; however, these genes have evolved complementary patterns of expression (Morton et al. 1989) where AMPD1 encodes isoform M (muscle) and AMPD2 encodes isoform L (liver) and AMPD3 encodes the erythrocytic form. Because preservation of AMPD duplicates is achieved through partitioning of the ancestral expression pattern rather than biochemical functions, this example is consistent with the duplication-degeneration-complementation (DDC) model.

Finally, in some cases, retention of duplicate genes can be more complex. For example, adenosine deaminase, ADA, has undergone several gene duplication events leading to various gene families including ADA-like and adenosine deaminase-related growth factors, ADGF. Activity of these paralogous sequences is complex and involves a delicate balance of temporal and spatial expression with only some overlap of enzymatic activity (Maier, Galellis, and McDermid 2005). In this case, there is no simple model accounting for the preservation of duplicated genes. One possibility is that subfunctionalization and gene dosage initially allowed for preservation of duplicate copies and was followed by neofunctionalization.

1.3.2 Gene Loss.

The end products of purine catabolism vary among vertebrates and have been the subject of comparative biochemical studies since the early 20th century (for example see Hunter et al. 1914b). It is well established that differential gene loss has led to differences in the end products of purine metabolism (Figure 1.5). In hominoids, birds, and reptiles, the end product of purine metabolism is uric acid and is due to the loss of UOX. Loss of this gene has occurred at least three times during vertebrate history (Wu et al. 1992; Urich 1994a) and the beneficial effects of this loss have been of long standing interest. Gene loss of ALLN in other mammals has led to their end product being allantoin. Allantoate is the final product in some teleost fish and is due to loss of ALLC while other teleost fish and amphibians have retained functional ALLC and UGL genes so that they complete the pathway and have an end product of urea (Hayashi, Fujiwara, and Noguchi 2000b).

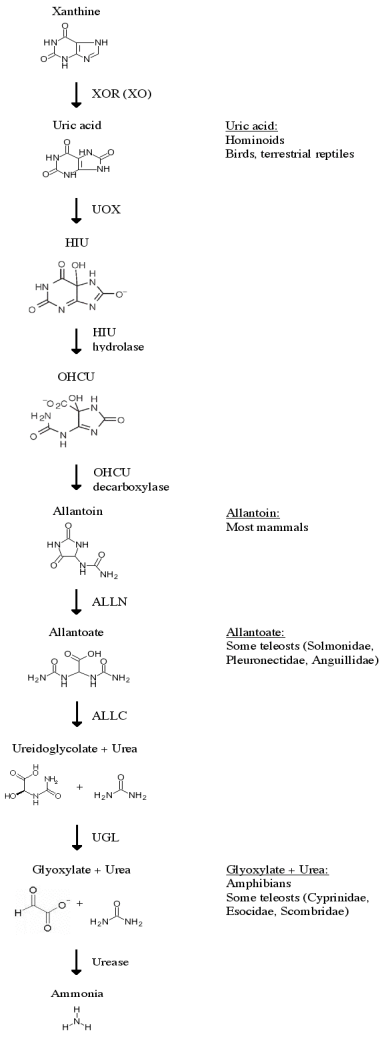


Figure 1.5 Purine catabolism.
The enzymes and catabolites of the terminal portion of the purine catabolic pathway are shown on the left. The end products of this pathway for different groups of vertebrates are shown on the right,(Urich 1994; Hayashi et al. 2000)

1.3.3 Human disease.

To date, at least eight human genetic diseases are associated with mutations in the purine metabolic pathway (Table 1.2). The clinical manifestations of these

diseases are highly variable and include a range of phenotypes not restricted to any specific aspect of human biology. For example, mutations to ADA or PNP lead to severe combined immune deficiency (Boss and Seegmiller 1982), while loss of IMPDH1 results in a dominant form of retintitis pigmentosa (Bowne et al. 2002). In the case of HPRT, the clinical phenotype can be as mild as gout to as severe as mental retardation and compulsive, self-injurious behavior (Jinnah and Friedann 2001). The variety of phenotypes resulting from defects in this pathway underscores the importance of purine metabolism in a wide spectrum of biological processes.

Table 1.2. Human genetic diseases associated with purine metabolism.

Gene	Disease	Knockout Mouse	Literature Cited
ADA	ADA deficiency	Lethal	(Boss and Seegmiller 1982)
ADSL	ADSL deficiency	N/A	(Kmoch et al. 2000)
AMPD1	AMPD deficiency	N/A	(Sabina et al. 1984; Morisaki et al. 1992)
APRT	APRT deficiency	Limited	(Boss and Seegmiller 1982)
HPRT	Lesch-Nyhan disease	Limited	(Jinnah and Friedann 2001)
IMPDH1*	Retinitis Pigmentosa	Limited	(Bowne et al. 2002)
PNP	PNP deficiency	N/A	(Boss and Seegmiller 1982)
XOR	Xanthinuria	Limited	(Boss and Seegmiller 1982; Ichida et al. 1997)

*Associated with missense mutations, no known nonsense mutations.

The biological significance of purine metabolism extends beyond the synthesis, salvage, and degradation of purines. Many genes involved in purine metabolism have been shown to have roles in a host of biological processes (Sollinger et al.

1992; Eugui and Allison 1993; Vorbach, Scriven, and Capecchi 2002; Akum et al. 2004; van der Knaap et al. 2005; Cheung et al. 2007). Thus, some of the clinical phenotypes may not be directly related to the known enzymatic activities, but rather due to the defects in other functions these genes have acquired in the recent evolution of vertebrates.

In addition to null alleles that cause disease, there are also non-pathogenic null alleles segregating in the human population. AMPD1 and AMPD3 null alleles are segregating at frequencies up to 12% in the human population (Ogasawara et al. 1987; Zydowo, Purzycka-Preis, and Ogasawara 1989; Morisaki et al. 1992). There is evidence that patients heterozygous for a null AMPD1 allele have better clinical outcomes following congestive heart failure and coronary artery disease (Loh et al. 1999; Anderson et al. 2000); however, this hypothesis has not been rigorously tested and a recent study did not find any relationship between presence of the null allele and better clinical outcomes (Agewall and Norman 2006). Nonetheless, the presence of null alleles at such high frequencies in the human population is intriguing with respect to ongoing modification/adaptation of this pathway.

1.3.4 Modeling disease in mouse.

Animal models can be an invaluable resource for developing a deeper understanding of, and treatments for, human genetic diseases. However, knockout mice do not always recapitulate the human phenotype and this seems to be especially true for diseases of purine metabolism (Table 1.2). For example, the human form of retinitis pigmentosa (RP) caused by mutations in the IMPDH1 gene is a severe autosomal dominant degenerative retinopathy; in contrast, IMPDH1 null mice display a slowly progressive form of retinal degeneration (Aherne et al. 2004). It is likely that protein mis-folding and aggregation, rather than reduced IMPDH1 enzyme activity, is the cause of the severe phenotype in the human form.

Some of the phenotypes are not directly related to the known enzymatic activity of the genes. For example, mouse XOR knockout models do not recapitulate the human disease phenotype but rather have a phenotype associated with the mammary gland of nursing females (Vorbach, Scriven, and Capecchi 2002). In another example, HPRT1-deficient humans have a suite of severe phenotypes including neurological and behavioral abnormalities, while HPRT1 knockout mice do not manifest any of the clinical phenotypes of Lesch-Nyhan Disease (Hooper et al. 1987; Kuehn et al. 1987a). The basis for the difference in loss of function phenotypes between species is not known but is hypothesized to be

related to gene content differences, genetic background and/or fundamental biological differences between species and supports rapid evolution of this pathway.

Clearly, gene content differences play an important role when studying human disease in model systems such as the mouse. For example, loss of UOX in humans is potentially adaptive whereas loss of UOX in mouse is lethal (Wu et al. 1994). In contrast, loss of HPRT1 in humans is lethal, whereas loss of HPRT1 in the mouse has no phenotype in a laboratory setting (Wu and Melton 1993b). These gene content differences may act as enhancers and/or suppressors of a disease phenotype. For example, when a double mutant has a less severe phenotype than the single mutant alone, the second mutation is said to be a suppressor. Interestingly, HPRT1/APRT double mutants are less severely affected than mice deficient for APRT alone (Engle et al. 1996). Thus, HPRT1 activity is required for full manifestation of the APRT mutant phenotype. Exploring the gene content differences between species within the purine metabolic pathway may reveal insight into other enhancer/suppressors that may be influencing the disparity in disease phenotypes between species and may have implications for modeling diseases of purine metabolism in model organisms.

1.3.5 Objectives of this dissertation.

The goal of the research presented in this dissertation is to characterize the specific aspects of gene duplication and gene loss in shaping interspecies differences with respect to the purine metabolic pathway. In chapters 2 and 3 we describe the history and pattern of gene loss in the terminal portion of the purine metabolic pathway. Chapter 2 provides novel insights into the importance of gene loss during mammalian evolution while chapter 3 combines theoretical discussion and experimental evidence to broaden our understanding on the long-term consequences of gene loss to adaptive evolution. In chapter 4 we show that gene duplication events have resulted in expansion of the vertebrate HPRT-gene family while subsequent differential gene loss has lead to interspecies differences. In chapter 5 we experimentally evaluate the role of a gene loss event to act as a suppressor of Lesch-Nyhan Disease.

2 The genomes of the South American opossum (Monodelphis domestica) and platypus (Ornithorhynchus anatinus) encode a more complete purine catabolic pathway than placental mammals [1]

[1]This chapter has been accepted for publication: Keebaugh, Alaine C., and James W. Thomas. 2009. The genomes of the South American opossum (*Monodelphis domestica*) and platypus (*Ornithorhynchus anatinus*) encode a more complete purine catabolic pathway than placental mammals. Comparative biochemistry and physiology. In press.

2.1 INTRODUCTION

Purine metabolism is an essential biochemical pathway that is conserved across a wide-range of phyla and is considered a likely candidate for the most ancient metabolic pathway on the planet (Caetano-Anolles, Kim, and Mittenthal 2007). The end products of purine catabolism, however, vary among vertebrates and have been the subject of comparative biochemical studies since the early 20th century (Hunter and Givens 1914; Hunter, Givens, and Guion 1914). Those studies as well as subsequent reports have led to the widely accepted view of the end products of purine catabolism in vertebrates as summarized in Fig. 2.1

The terminal portion of purine catabolism begins with the degradation of hypoxanthine to uric acid by xanthine dehyrogenase and xanthine oxidoreductase (XDH, EC 1.1.1.204; XOR, EC 1.2.3.2), which are encoded for by a single gene present in all vertebrates. Until recently, the next step in this pathway was reported to be the degradation of uric acid to allantoin by urate oxidase (UOX, EC 1.7.3.3), which is absent in hominoids, birds and terrestrial reptiles. However, Ramazzina *et al.*, (2006) recently identified two enzymes, HIU hydrolase and OHCU decarboxylase, that catalyze intermediate steps in the conversion of uric acid to allantoin, which is the end product of this pathway in most mammals. Allantoinase (ALLN, EC 3.5.2.5) then degrades allantoin to allantoate, which is

the final product of this pathway in some teleost fish. In other teleost fish and amphibians, allantoicase (ALLC, EC 3.5.3.4) catalyzes the hydrolysis of allantoate to ureidoglycolate and urea, which is followed by degradation of ureidoglycolate to glyoxylate and urea by ureidoglycolate lyase (UGL, EC 4.3.2.3)(Hayashi, Fujiwara, and Noguchi 2000a). Finally, urease activity (EC 3.5.1.5), whose presence has been detected in the gut of some fish but is encoded within bacteria living in the host and not the vertebrate genome, can generate the most terminal products of the pathway, ammonia and carbon dioxide (Urich 1994).

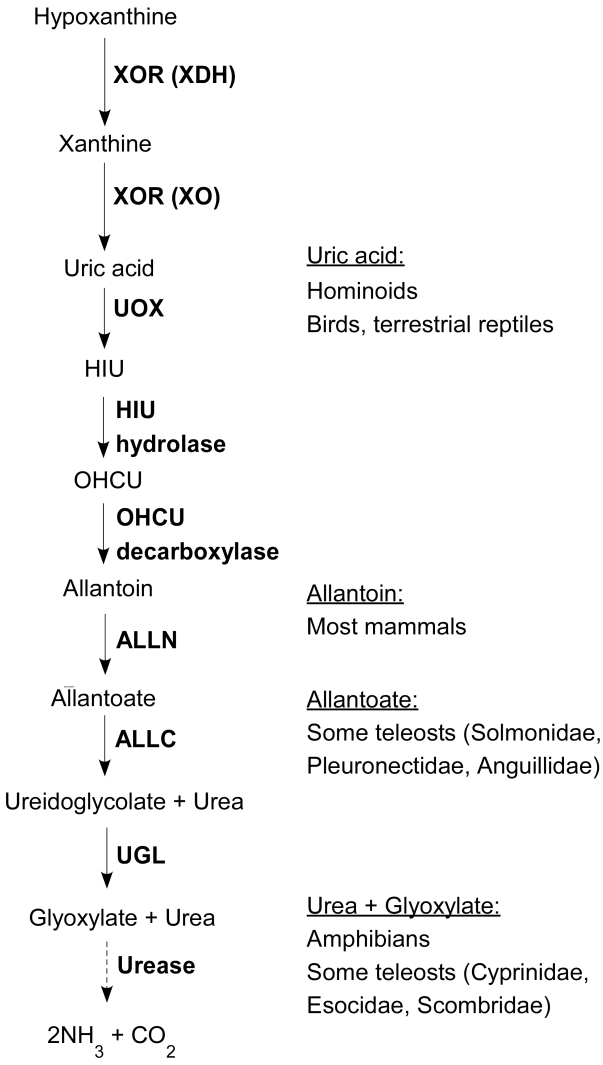


Figure 2.1 Purine catabolism.
The enzymes and catabolites of the terminal portion of the purine catabolic pathway are shown on the left. The end products of this pathway for different groups of vertebrates are shown on the right (Urich 1994a; Hayashi, Fujiwara, and Noguchi 2000a).

To date, comparative studies of the end products of purine catabolism in mammals have relied on the relative ratios of uric acid and allantoin in urine as a

standard by which to compare this pathway between species. In placental mammals, most species primarily excrete allantoin and a small amount of uric acid (Hunter and Givens 1914; Hunter, Givens, and Guion 1914), whereas hominoids primarily excrete uric acid and only trace amounts of allantoin (Wiechowski 1909a; Wiechowski 1912; Wells and Caldwell 1914a). More limited information is available for the other mammals, the marsupials and monotremes. In the case of marsupials, the enzymatic activity of UOX has been directly assayed and detected in the opossum (Caldwell and Wells 1914). With the exception of human and chimpanzee (Hunter, Givens, and Guion 1914), this species was shown to excrete more uric acid relative to allantoin than all placental mammals. In the monotremes, both the platypus and echidna have been shown to excrete uric acid and allantoin, but the total amount of these metabolites per a defined volume of urine was much smaller and included high proportions of uric acid than was observed in the placental mammal included in the same study (Denton, Reich, and Hird 1963). Purine catabolism therefore clearly proceeds beyond uric acid in both marsupials and monotremes, and the presence of significant amounts of allantoin indicates that the pathway extends at least as far in these species as in most placental mammals. Moreover, the distinct relative ratios of uric acid to allantoin in marsupials and monotremes compared to most placental mammals could be an indication that both uric acid and allantoin are intermediate metabolites in non-eutherian mammals. Thus, while purine catabolic

metabolites downstream of allantoin have not been directly assayed in marsupials and monotremes, these results are suggestive that the purine catabolic pathway in these species proceeds farther than it does in placental mammals.

Variation in the end product of purine catabolism among vertebrates is the result of truncation of the pathway by complete or partial inactivation of the genes encoding the purine catabolic enzymes (Wu et al. 1989; Wu et al. 1992; Andersen et al. 2006); Keebaugh and Thomas, 2009). Once a mutation truncates the purine catabolic pathway, the genes downstream of the mutated enzyme are presumably no longer needed and will be lost over time. Though this model is not always correct, (Fujiwara and Noguchi 1995; Vigetti et al. 2000; Vigetti, Monetti, and Bernardini 2001), a comparative genomic-based study focused strictly on the presence or absence of genes can be a good predictor of the functional content of the purine catabolic pathway (Wong and Wolfe 2005; Ramazzina et al. 2006b). Here we report the results of a comparative genomics study designed to compare the gene content in the terminal portion of the purine catabolic pathway in a marsupial, the South American opossum (*Monodelphis domestica*) and a monotreme, the platypus (*Ornithorhynchus anatinus*), versus a diverse set of other vertebrates.

2.2 MATERIALS AND METHODS

2.2.1 Identification of orthologs encoding the purine catabolic pathway

A three-tiered strategy was used to identify the platypus (*Ornithorhynchus anatinus*) and opossum (*Monodelphis domestica*) orthologs of the six known genes in the terminal portion of the purine catabolic pathway, as well as orthologs from all 17 species for XDH. First, pre-computed orthologs and the accompanying nucleotide and protein alignments were extracted from the TreeFam database (http://www.treefam.org/) (Li et al. 2006) for XDH (TF507706), UOX (TF323438), HIU hydrolase (TF300210), ALLN (TF300759) and ALLC (TF324677). Second, in cases where orthologs were missing (OHCU decarboxylase) or incomplete in TreeFam, Ensembl (http://www.ensembl.org) (Hubbard et al. 2007) annotation and ortholog designations were used to identify the genes of interest. Finally, homology searches followed by manual gene annotation were used to scan for orthologs absent in the combined TreeFam/Ensembl ortholog sets. The retrieval of orthologs from the 15 other vertebrates, excluding XDH, was done with a similar strategy and is described elsewhere (Keebaugh and Thomas 2009). The complete sets of orthologs were aligned with CLUSTALX (Jeanmougin et al. 1998a) and resulting alignments manually edited using the TreeFam alignments as a guide. The sequence

alignments are available from the authors upon request. Note that the specific genome assemblies analyzed here for platypus and opossum were ornAna1 and monDom4, respectively.

2.2.2 Classification of orthologs

The orthologs in each species were classified as a gene, a pseudogene, or absent. Orthologs were classified as a gene if they had an open reading frame, were evolving under purifying selection, and had a conserved intron/exon structure. Orthologs were classified as a pseudogene if the protein coding sequence was truncated by early nonsense mutations and evolving neutrally. Orthologs were classified as absent from a species only if homology searches of the assembled genome (and when available EST databases) and direct analysis of the predicted syntenic genomic location failed to detect the gene of interest.

2.2.3 Sequence Evolution

To evaluate the rate of evolution for the purine catabolic genes, a maximum likelihood method was implemented by the "codeml" program in PAML (Yang 2003) to estimate the ratio of nonsynonymous (K_a) to synonymous (K_s) substitutions, i.e. K_a/K_s. The tree phylogeny used in the following analyses is: (((((((Marmoset,(Rhesus,(Chimp,Human))),(Rat,Mouse)),(Dog,Cow)),Opossum),

Platypus),(Chicken,Anolis)),Xenopus,(((Fugu,Tetraodon),Stickleback),Zebrafish)); For each gene, K_a/K_s rate estimates were first compared between the marsupial and monotreme lineages using a likelihood ratio test (LRT) to determine if the estimated K_a/K_s values were significantly different ($p < 0.01$). Rate estimates for the marsupial and monotreme, either combined ($p > 0.01$) or individually ($p < 0.01$), were then compared to the K_a/K_s estimates for the genes known to have retained their function as purine catabolic enzymes in fish, amphibian, bird & reptile, and placental mammals to detect significant differences with other groups of vertebrates.

2.3 RESULTS

2.3.1 The opossum and platypus genomes encode a more complete set of genes for purine catabolism than do placental mammals

The end products of purine catabolism in placental mammals are well established but have yet to be determined in marsupials and monotremes. To genetically characterize the terminal portion of the purine catabolic pathway in marsupials and monotremes we determined the presence or absence of the known genes in this pathway in the platypus (monotreme) and opossum (marsupial) genomes. In particular, the platypus and opossum genomes were searched for the presence of

XOR/XDH, UOX, HIU hydrolase, OHCU decarboxylase, ALLN and ALLC as the basis for comparison to a diverse sampling of 15 other vertebrates (4 fish: stickleback, fugu, tetraodon and zebrafish, 1 amphibian: frog, 1 reptile: anolis lizard, 1 bird: chicken, and 8 placental mammals including 2 hominoids: human and chimpanzee and 6 other placental mammals: mouse, rat, dog, cow, rhesus monkey and marmoset)(Keebaugh and Thomas 2009). (Note that the identity of the gene encoding the UGL locus has yet to be identified in vertebrates (Takada and Noguchi 1986), and therefore could not be included in this study.)

Table 2.1 Estimation of Ka/Ks of the genes in the purine catabolic pathway.

	XDH		UOX		HIU hydrolase		OHCU decarboxylase		ALLN		ALLC	
	Syn	Non	Syn	Non	Syn	Non	Syn	Non	Syn	Non	Syn	Non
Species	656	1954	215	625	78	240	93	290	137	373	252	686
Marsupial	0.0865[a]		0.0797		0.0633		0.4613[a]		0.0957		0.3236	
Monotreme	0.1894[a]		0.0786		0.1455		0.0612[a]		0.0599		0.1401	
Fish & Amphibian	0.1164[c]		0.0350		0.2296		0.1296		0.0883		0.1102[b,c]	
Placental mammals	0.1492[b]		0.1101		0.1396		0.1916		n/a		n/a	

Ka/Ks estimates of the genes in the purine catabolic pathway were calculated for all lineages except the hominoids, and birds and reptiles. Note that though ALLC is present in the genomes of placental mammals it was not included in this table because it is no longer required to function as a purine catabolic enzyme.
[a]Indicates a significant difference ($p<0.01$) in estimated rates between marsupial and monotreme lineages.
[b]Indicates a significant difference ($p<0.01$) in estimated rates compared to marsupial.
[c]Indicates a significant difference ($p<0.01$) in estimated rates compared to monotreme.
'Syn' refers to the number of synonymous sites; 'Non' refers to the number of nonsynonymous sites

As expected from previous biochemical studies summarized in Fig. 2.1 and as reported elsewhere (Keebaugh and Thomas 2009), all the known purine catabolic genes are present in the amphibian and fish genomes (Fig. 2.2). Similarly, we also detected the full complement of purine catabolic genes, including *ALLN* which was absent in placental mammals, in both the opossum and platypus genomes (Fig. 2.2). (Note that the absence of *ALLN* in the genomes of placental mammals is consistent with allantoin being the end product of purine catabolism in those species, excluding the hominoids). Moreover, estimates of the rates at which the marsupial and monotreme purine catabolic genes have evolved were consistent with each encoding a functional protein (Table 2.1). In particular, the rate at which a protein has evolved can be inferred by calculating the ratio of the rates of nonsynonymous (K_a) and synonymous (K_s) nucleotide substitutions, i.e., K_a/K_s. In the case of the opossum and platypus, the genes in the purine catabolic pathway had K_a/K_s values between 0.07-0.47, consistent with all six genes evolving under purifying selection common for most vertebrate genes (Consortium 2001) (Table 2.1). While statistically significant differences in the estimated K_a/K_s values were detected between the opossum, platypus and other vertebrate lineages, no obvious pattern emerged that suggested the opossum and platypus genes were evolving under different selective constraints than the orthologs that are known to encode functional purine catabolic enzymes in other vertebrates (Table 2.1). The presence of all the genes encoding the purine catabolic pathway in the opossum and

platypus genomes and their pattern of evolution therefore suggests that marsupials and monotremes have the potential to encode a more complete purine catabolic pathway than placental mammals.

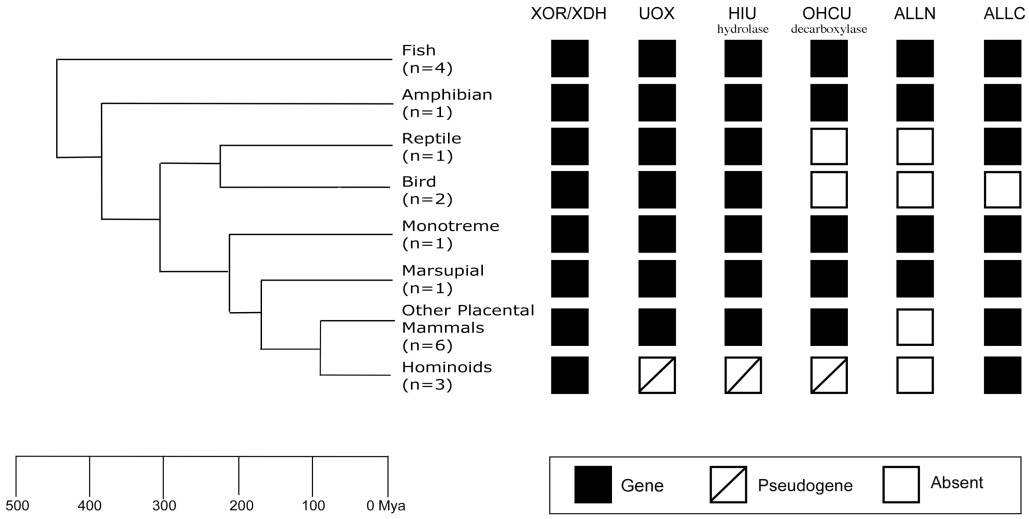


Figure 2.2 Gene content of the purine catabolic pathway in opossum and platypus.
The tree on the left represents the accepted vertebrate phylogeny and divergence dates (millions of years ago, Mya) (Kumar and Hedges 1998b). The species included in each category are fish (stickleback, fugu and tetraodon and zebrafish), an amphibian (frog), a reptile (anolis lizard),a bird (chicken), a monotreme (platypus), a marsupial (opossum), and placental mammals, including the hominoids (human and chimpanzee) and other placental mammals (mouse, rat, dog, cow, rhesus monkey and marmoset). The gene content of the purine catabolic pathway for each group of species is summarized on the right. Orthologs were classified as a gene if they had an open reading frame, were evolving under purifying selection, and had a conserved intron/exon structure, as pseudogenes if the protein coding sequence was truncated by early nonsense mutations and evolving neutrally, and absent when no gene or gene fragment was identified.

2.3.2 Loss of ALLN is specific to placental mammals

The presence of *ALLN* in opossum and platypus genomes and its absence in the genomes of placental mammals suggests that *ALLN* was most likely present in the eutherian lineage prior to the radiation of placental mammals ~100 million years ago (Murphy et al. 2001). We therefore scrutinized the predicted syntenic location of *ALLN* between the flanking *KIF1A* and *AGXT* loci in the genomes of the placental mammals to see if we could detect any remnants of this gene. However, we were unable to detect *ALLN*, or clear evidence for any other gene, in this region of eutherian genomes, including the finished genomes of human and mouse. Homology searches using the ALLN protein sequence from other species were also unable to detect this gene elsewhere in the genome of placental mammals. Thus, these results suggest that not only was the purine catabolic function of ALLN lost in placental mammals, but the gene encoding this enzyme has also been deleted or degenerated beyond recognition in the eutherian lineage.

2.4 Discussion

Variation in the end product of purine catabolism among vertebrates has been of interest to biologists for nearly a century (Caldwell and Wells 1914; Hunter and Givens 1914; Hunter, Givens, and Guion 1914; Keilin 1959). In textbooks and

reviews, mammals are grouped into two classes based on their end products of purine catabolism; the hominoids, whose end product is uric acid, and other mammals, whose end product is allantoin (Fig. 2.1). This dichotomy among mammals can be explained by the loss of ALLN activity in the common ancestor of all mammals, and subsequent recent inactivation(s) of UOX in hominoids (Wu et al. 1992; Oda et al. 2002). The results of this study provide an updated and expanded view of the sequence of genetic modifications that have shaped the evolution of this pathway in mammals.

The presence of the *ALLN* in opossum and platypus and absence in placental mammals strongly suggests that loss of *ALLN* is specific to placental mammals. The most parsimonious time frame for the loss of *ALLN* from the eutherian genome therefore pre-dates the most recent common ancestor of placental mammals ~100 million years ago (Murphy et al. 2001) and post-dates the split of the eutherian and marsupial lineages ~173 million years ago (Kumar and Hedges 1998b). It is possible that this genetic difference between placental mammals and marsupials and monotremes may not reflect a true biochemical difference between these groups of species. However, the presence of a full complement of the known genes in the terminal portion of the purine catabolic pathway in both the opossum and platypus demonstrates that these species have, at a minimum, retained the potential to encode a more complete pathway than is found in other

mammals. As such, we hypothesize that marsupials and monotremes have end products of purine catabolism comparable to that of fish and/or amphibians, i.e. glyoxylate and urea (Fig. 2.1), and thus represent a third class of mammals with respect to the end products of purine metabolism. We further propose that the distinct ratios and amounts of uric acid and allantoin previously observed in the urine of the Virginia opossum (*Didelphis virginiana*) (Caldwell and Wells 1914), platypus and echidna (Denton, Reich, and Hird 1963) compared to placental mammals is a reflection of this predicted difference in end products of purine catabolism. Direct biochemical studies will ultimately be needed to test these hypotheses and determine the specific end product of purine metabolism in marsupials and monotremes.

Purine metabolism has been repeatedly modified via differential gene loss throughout vertebrate evolution, and gene loss is increasingly being recognized as an important force in evolution (Olson 1999). Recently, clear associations have been made between gene loss and fundamental biological differences between placental mammals and monotremes (Brawand, Wahli, and Kaessmann 2008; Ordonez et al. 2008). Given that the loss of ALLN is specific to placental mammals, it is tempting to speculate on possible adaptive advantages the loss of this gene may have provided to our eutherian ancestors. Prior hypotheses on the adaptive advantages of modifications in nitrogen metabolism, including the urea

cycle and the end products of purine catabolism, have focused on diet, water conservation and embryonic environment (Packard 1966; Campbell, Vorhaben, and Smith 1987; Mommsen and Walsh 1989). Since the eutherian embryonic environment, which includes full-term development *in utero* with a chorioallantoic placenta, is distinct from that of both the marsupials and monotremes, one possibility is that loss of ALLN is somehow related to this phenotypic innovation that is common to all placental mammals. One possible experimental approach to assess the biological significance of truncation in placental mammals would be to genetically alter the mouse genome in order to reconstitute the functional purine catabolic pathway inferred to have been present in the common ancestor of all mammals. In summary, the results of our modern comparative genomic survey have yielded new insights into the evolution of the purine catabolic pathway in mammals and provide a rationale to revisit and expand the comparative biochemical studies of purine catabolism in marsupials and monotremes.

3 Evolution of the purine catabolic pathway: an expanded view of the evolutionary significance of gene loss[1]

[1]This chapter is in preparation for publication

3.1 Introduction

The evolutionary potential of gene loss has been recognized for some time (Ohno 1970). More recently it has been hypothesized that gene loss may be a common molecular mechanism in adaptive evolution (Ohno 1970; Olson 1999). Because null alleles are typically considered to be detrimental, the hypothesis that gene loss can be beneficial is counterintuitive to most experimental biologists. However a number of null alleles have been shown to be advantageous in certain environments (Olson 1999). Genome-wide scans for gene loss, along with targeted studies of specific genes, have demonstrated that numerous recent gene inactivation events have occurred during the course of human evolution (Wu et al. 1992; Varki 2001; Stedman et al. 2004; Wang, Grus, and Zhang 2006; Zhu et al. 2007). Null alleles for a number of genes are also segregating in the human population at high frequencies, some of which show signatures of positive selection or have been associated with defined beneficial fitness effects (Novembre, Galvani, and Slatkin 2005; Xue et al. 2006; Perry et al. 2007; Seixas et al. 2007; Xue et al. 2008). Recent genome sequencing revealed that a single human genome harbored 175 protein-truncating sequence variants (Bentley et al. 2008), each of which is a candidate for gene inactivating mutations. While most null alleles undoubtedly are detrimental or will have no impact on the fitness of an

individual, it is clear that the process of gene loss certainly has had the opportunity to influence the evolution of humans.

A classic example of a potential adaptive gene inactivation in the human lineage is the loss of urate oxidase (UOX, EC 1.7.3.3) activity. UOX is a peroxisomal enzyme in the purine catabolic pathway that functions primarily in the liver and converts uric acid to 5-hydroxyisourate (Hayashi, Fujiwara, and Noguchi 2000b; Ramazzina et al. 2006a). Loss of UOX activity was demonstrated in humans and great apes early in the 20th century by biochemical studies that revealed uric acid as the end product of purine metabolism in hominoids (Wiechowski 1909b; Wells and Caldwell 1914b). It has been speculated that the loss of UOX activity might have conferred an adaptive advantage, resulting in increased serum uric acid levels. Due to uric acid's structural similarity to caffeine, increased serum uric acid levels have been proposed to have led to increased cognitive abilities in hominoids due to its potential stimulatory effects (Haldane 1955; Orowan 1955). A second hypothesis is that because uric acid is an effective antioxidant**,** increased serum uric acid levels could result in a longer life span and reduce the incidence of age related cancer (Ames et al. 1981). More recently, a third hypothesis has been put forward, in which the loss of UOX provided a mechanism to maintain optimal blood pressure in ancestral hominoids that subsisted on a low-sodium, low-purine diet (Watanabe et al. 2002). It should be noted that UOX activity was

also lost, independently, in the bird and reptile lineage (Keilin 1959; Urich 1994a). The biological significance of UOX inactivation in these species has been linked to water conservation and/or embryonic environment (Campbell, Vorhaben, and Smith 1987).

Molecular studies focused on identifying the molecular basis of UOX inactivation in hominoids uncovered specific inactivating nonsense and/or frameshift mutations within this gene in all extant species included in this family of primates (Wu et al. 1989; Yeldandi et al. 1991). However, no inactivating mutations were found to be shared amongst all hominoids (Wu et al. 1989; Wu et al. 1992; Oda et al. 2002). This finding and subsequent molecular dating are consistent with the conclusion that UOX was inactivated twice in the hominoid lineage, once in the common ancestor of humans and the great apes, and independently in the gibbon lineage, with both inactivation events occurring ~10-15 million years ago (mya) (Oda et al. 2002).

In birds and reptiles, UOX was presumably inactivated one time in the common ancestor of those species, >200 million years ago, however the underlying molecular basis for that inactivation has not yet been elucidated. In particular, though it is well established that UOX activity is absent in the adult liver of birds and reptiles (Remy, Richert, and Westerfeld 1951; Varela-Echavarria, Montes de

Oca-Luna, and Barrera-Saldana 1988), there have been only cursory and somewhat contradictory studies of the UOX protein and gene sequence in these species. In particular, there is evidence for UOX transcripts and protein expression in the chicken embryo (Ito et al. 1989) as well as UOX activity (Keilin 1959). One interpretation of these results is that, unlike hominoids, which do not express the UOX protein (Varela-Echavarria, Montes de Oca-Luna, and Barrera-Saldana 1988), birds and reptiles may in fact express a functional UOX protein but only during a specific stage of embryonic development. Alternatively it may be that the UOX protein is more broadly expressed in birds and reptiles but has acquired subtle mutations that have specifically eliminated its purine catabolic activity (Andersen et al. 2006).

In addition to the lack of basic knowledge regarding the evolutionary fate of UOX in birds and reptiles, the fate of the purine catabolic genes downstream of UOX has not been rigorously evaluated in either lineage. It has generally been assumed that once UOX activity was lost in hominoids, and birds and reptiles, enzymes downstream of UOX in the purine catabolic pathway would be rendered dispensable, thereby resulting in their eventual inactivation (Keilin 1959). Under this model, the genes downstream of UOX in the hominoid and bird and reptile lineages would be expected to share a common fate-- loss of gene function. A cursory description of the two immediate downstream enzymes, HIU hydrolase

and OHCU decarboxylase, in human and chicken supported this view (Ramazzina et al. 2006a). In contrast, another downstream gene, ALLC, has been retained, and enzymatic activity has been detected in placental mammals (Vigetti et al. 2002; Vigetti et al. 2003). This finding indicates that gene loss may not be the only possible fate of these downstream genes. In light of these results, we decided to use the purine metabolic pathway as a model system in which to study the effect of gene loss upon the evolution of genes downstream in an important biological process.

Here we report the results of our study, the purposes of which were two-fold: to elucidate the evolutionary fate of (i) the UOX gene in birds and reptiles and (Hillier et al.) the downstream genes in the purine catabolic pathway in both hominoids, and birds and reptiles. Our results demonstrate that gene loss is not the exclusive fate of the downstream genes and show that a single gene loss event can have long-reaching effects and long-term benefits, providing an expanded view of the adaptive potential of gene loss.

3.2 MATERIALS AND METHODS

3.2.1 Identification of orthologs encoding the purine catabolic pathway

A three-tiered strategy was used to identify the orthologs of the five known genes in the terminal portion of the purine catabolic pathway. First, pre-computed orthologs and the accompanying nucleotide and protein alignments were extracted from the TreeFam database (http://www.treefam.org/) (Li et al. 2006) for 12 species [fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*), stickleback (*Gasterosteus aculeatu*), frog (*Xenopus tropicalis*), chicken (*Gallus gallus*), cow (*Bos taurus*), dog (*Canis familiaris*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), rhesus monkey (*Macaca mulatta*), chimpanzee (*Pan troglodytes*) and human (*Homo sapiens*)] and six genes/proteins [XDH (TF507706), UOX (TF323438), OHCU (323276), HIUase (TF300210), ALLN (TF300759) and ALLC (TF324677)]. Second, in cases where a set of orthologs were incomplete in TreeFam, the most recent Ensembl (http://www.ensembl.org) (Hubbard et al. 2007) annotation and ortholog designations were used to compile a complete set of orthologs for those species. Finally, homology searches followed by manual gene annotation were used to scan for orthologs absent in the combined TreeFam/Ensembl ortholog sets, as well as in the assembled genomes of the anolis lizard (*Anolis carolinesis*, AnoCar1.0), accessed via the UCSC genome browser (http://www.ucsc.genome.edu), zebra finch (*Taeniopygia*

guttata) unassembled whole-genome reads shotgun reads (http://www.ncbi.nlm.nih.gov), marmoset (*Callithrix jacchus*, Callithrix jacchus 2.0.1) and orangutan (*Pongo pygmaeus abelii*), which were downloaded from the Washington University Genome Center (http://genome.wustl.edu/pub/). The complete sets of orthologs were aligned with CLUSTALX (Jeanmougin et al. 1998a) and resulting alignments manually edited using the TreeFam alignments as a guide. The sequence alignments are available from the authors upon request.

3.2.2 Classification of orthologs

The orthologs in each species were classified as a gene, a pseudogene, or absent. An ortholog was classified as a gene if it had an open reading frame, was evolving under purifying selection, and had a conserved intron/exon structure. An ortholog was classified as a pseudogene if the protein coding sequence was truncated by early nonsense mutations and evolving neutrally. Orthologs were classified as absent from a species only if homology searches of the assembled genome (and when available EST databases) and direct analysis of the predicted syntenic genomic location failed to detect the gene of interest.

3.2.3 Sequence Evolution

To evaluate the rate of evolution in the purine catabolic genes, a maximum

likelihood method was implemented using the "codeml" program in PAML (Yang 2003) to estimate the ratio of nonsynonymous (K_a) to synonymous (K_s) substitutions, i. e. K_a/K_s. The following phylogeny was used in all PAML analyses: (((((((((Marmoset,(Rhesus,(Gibbon,(Orangutan,(Gorilla,(Chimp,Human)))))),(Rat,Mouse)),(Dog,Cow)),Opossum),Platypus),((Chicken,Zebrafinch),Anolis)),Xenopus),(((Fugu,Tetraodon),Stickleback),Zebrafish)),Fly,Worm). For each gene set, the hypothesis that all hominoid and all bird and reptile branches were evolving under the same constraint was tested (PAML free model versus a model where the branches are fixed). When this hypothesis was rejected the offending branches were excluded. For the UOX, HIUase, OHCU, and ALLC genes, K_a/K_s rate estimates for the hominoid lineage were compared to the K_a/K_s rate estimates for the other vertebrates. Next, K_a/K_s rate estimates for the bird and reptile lineage were compared to the K_a/K_s rate estimates for the other vertebrates. A likelihood ratio test (LRT) was used to determine if the estimated K_a/K_s values were significantly different ($p < 0.01$). Rate estimates within the hominoid and bird and reptile lineages were then compared to a model where the K_a/K_s estimate was set to one and a likelihood ratio test (LRT) was used to determine if the estimated K_a/K_s values were significantly different from one ($p < 0.01$).

3.2.4 Dating

A slightly modified method from Chou et al (2002) was used to date inactivation of the UOX, HIUase and OHCU hominoid pseudogenes. This model assumes that the null allele evolved under selective pressure similar to that of a closely related functional ortholog prior to inactivation and then neutrally following inactivation. We estimated the dates of inactivation independently for the hominid lineage (including human, chimp, orangutan, gorilla (DDBJ AB07432)), and gibbon (DDBJ AB074394) lineage. Each inactivation estimate was standardized using the Catarrhine primate divergence dates from Raaum et al., 2004, by transforming the raw ratio t*/t (t*=estimated inactivation date, t=divergence date), making the dating estimates comparable across lineages.

3.2.5 RNA isolation and Sequencing

Total RNA was isolated from ~100 mg of frozen zebra finch liver by homogenization with TRIzol (Invitrogen) and first strand cDNA synthesis was performed with the SuperScript First Strand Synthesis kit (Invitrogen). The region spanning the predicted exon 2 to the 3'UTR was amplified by RT-PCR using primers 5'-TGA TAC CTA CGG ACA CCA TAA AG-3' and 5'-TGA CAT TTC CTC TCA TCT CTC ATC-3'. To amplify the 5' end of the gene 5' RACE was carried out using the GeneRacer Core Kit (Invitrogen). The Gene Racer 5'

primer and UOX specific reverse primer 5'-GAT GTG GGG AAC GCC ATC CTC ATG-3' were used in the initial RT-PCR. Cycling conditions were: initial denaturation of one cycle at 94C for 2 minutes, 5 cycles at 94C for 30s followed by 72C for 1 minute, 5 cycles at 94C for 30s followed by 70C for 1 minute, 25 cycles with a denaturation temperature of 94C for 30s, an annealing temp of 65C for 30s and an elongation temp for 72C for 1 minute. The final extension was at 72C for 10 minutes. 1/5 (5μl) of the initial PCR reaction was used as the template for nested PCR using the GeneRacer 5' nested primer and the UOX specific nested reverse primer 5'-GAC GTA CGC CAC TTG GCA AAA TGT TGT C-3'. The same cycle parameters were used as for the initial RT-PCR. In both cases, a single PCR product was amplified and treated with Exonuclease I and Shrimp Alkaline Phosphatase. The amplified PCR product was sequenced using the PCR primers 5'-TGA TAC CTA CGG ACA CCA TAA AG-3'/5'-TGA CAT TTC CTC TCA TCT CTC ATC-3' and in the case of the 5' RACE product, primers 5'-ACA CTG ACA TGG ACT GAA GGA-3' and 5'-TGG TGT CCG TAC GTA TCA CAA-3'. Sequences for zebra finch UOX were assembled using the phred software (Ewing et al. 1998) and have been deposited in GenBank (EU068002 and EU068003).

3.3 Results

3.3.1 Predicted gene content of the purine catabolic pathway in hominoids, birds and reptiles

In order to determine the evolutionary fate of UOX and four downstream genes (HIU hydrolase, OHCU decarboxylase, ALLN and ALLC) in hominoids and birds/reptiles, genomic resources were used to determine the functional status of each gene in a diverse sampling of vertebrates. In particular, genomic sequences and annotation from 19 vertebrates, including 5 hominoids (human, chimpanzee, gorilla, orangutan and gibbon), 2 birds (chicken and zebra finch), 1 lizard (anolis), as well as species in which the pathway is truncated at ALLN (6 other placental mammals: dog, cow, rat, mouse, marmoset, and rhesus monkey) and species in which the purine catabolic pathway is not truncated downstream of ALLC (4 fish: stickleback, fugu, tetraodon and zebrafish, and 1 amphibian: *Xenopus tropicalis*), were searched to first determine if the purine catabolic genes were present or absent in each species, and second, whether the genes that were found to be present were likely functional or pseudogenes (see Methods). A summary of the results of these analyses is presented in Fig. 3.1. In the succeeding paragraphs we describe the specific findings for each gene.

UOX

UOX enzymatic activity is absent in hominoids, and birds and reptiles, but present in all other vertebrates (Urich 1994a). We therefore expected to detect functional copies of UOX in all sampled species except the hominoids and birds/reptiles. Indeed, the fish, amphibian, and other placental mammal UOX orthologs were all classified as functional genes (Fig 3.1). Moreover, as has been previously reported, the hominoid UOX orthologs all contained 1 or more inactivating mutations (Wu et al. 1992; Oda et al. 2002). They were found to be evolving at a rate not significantly different than the Ka/Ks ratio of ~1 expected for an inactivated gene (Ka/Ks = 0.87, $p = 0.71$, Table 3.1) and were therefore classified as pseudogenes (Fig 3.1). In contrast to the hominoids, no obvious inactivating mutations were detected in the bird and reptile UOX genes. Furthermore, the Ka/Ks ratio for the bird and reptile UOX lineages, though significantly higher than for the control set of vertebrates, was significantly less than 1 (Ka/Ks = 0.26, $p < 0.01$, Table 3.1). Thus, while UOX activity as a purine catabolic enzyme has been lost in birds and reptiles, these results suggest that the UOX locus in birds and reptiles continues to encode a functional protein that is evolving under purifying selection.

Table 3.1 Estimation of Ka/Ks of coding sequence

Gene	Other Vertebrates Ka/Ks	Hominoid				Bird & Reptile			
		Non	Syn	Ka/Ks	Ka/Ks ≠1 (p>0.01)	Non	Syn	Ka/Ks	Ka/Ks ≠1 (p>0.01)
UOX	0.0807	579	189	0.8754*	p=0.7083	602	193	0.2629*	p<0.01
HIU hydorlase	0.1180	133	38	0.8656#,*	p=0.7290#	148	44	0.2383	p<0.01
OHCU decarboxylase	0.1606	318	105	0.6042	p=0.1949	n/a	n/a	n/a	n/a
ALLC	0.1384	616	224	0.4844*	p=0.0001	639	225	0.2274	p<0.01

* Indicates a Ka/Ks significantly different (p<0.01) from that of the other vertebrates containing a functional purine catabolic gene.
# Indicates exclusion of orangutan lineage.

HIU hydrolase

HIU hydrolase was classified as a functional gene in all the fish, amphibian, and other placental mammals (Fig 3.1 and Table 3.1). In the hominoids, where presumably the purine catabolic function of HIU hydrolase is no longer required, a mixed pattern of functional and pseudogenized genes was observed (Fig 3.1). Specifically, in orangutan, no inactivating mutations were observed (Table 3.2). Furthermore, the Ka/Ks ratio for the orangutan lineage was significantly different from the rest of the hominoids, being significantly less than 1 (Ka/Ks=0.001, p=0.01), suggesting that orangutan HIU hydrolase may still be functional. However, based on the presence of inactivating mutations in human, chimpanzee and gibbon (Table 3.2) as well as Ka/Ks ratios that were not significantly different than 1 in all hominoid lineages except orangutan (Ka/Ks = 0.87, p = 0.73, Table

3.1), HIU hydrolase was classified as a pseudogene in the other four hominoids (Fig. 3.1 and Table 3.2). Among the birds and reptiles, HIU hydrolase was classified as a functional gene with a Ka/Ks ratio of 0.24, which was significantly different than 1 ($p < 0.01$, Table 3.1), was evolving at the same rate as the control set of vertebrate orthologs. Thus, for the most part HIU hydrolase appears to be genetically deteriorating in the hominoids while being actively retained in the bird and reptile genomes.

Table 3.2 Inactivating mutations in the hominoid HIU hydrolase genes

Species	Mutation				
	1A>G	2T>C	163_179del[b]	234_235del[c]	315_331del[d]
Human	-	+	+	-	+
Chimp	+	+	-	-	-
Orangutan	-	-	-	-	-
Gorilla	-	-	-	-	-
Gibbon	-	+	-	+	-

(+) Mutation present; (-) Mutation absent

[a] Mutations based on coordinates from mouse HIU hydrolase coding sequence (GenBank Accession No. NM_029821).

[b] Fourteen basepair frameshift deletion leads to an early stop at codon 67.

[c] Leads to early stop at codon 107

[d] Fourteen basepair frameshift deletion.

OHCU decarboxylase

As was observed for UOX and HIU hydrolase, functional copies of OHCU decarboxylase were inferred to be present in the sampled fish, amphibian and other placental mammalian genomes (Fig. 3.1). Within the hominoids, a single inactivating mutation, a 1-bp deletion at position 244 (relative to the mouse

protein coding sequence in GenBank Ac.# NM_001039678) that leads to a premature stop codon, was detected in the gibbon OHCU decarboxylase. In addition, all the hominoid orthologs were found to be evolving at a rate not significantly different than 1 (Ka/Ks = 0.6, p = 0.19, Table 3.1). As such, OHCU decarboxylase was classified as a pseudogene in all hominoids (Fig. 3.1). We were unable to detect OHCU decarboxylase in the bird and reptile genome assemblies and so classified the gene as absent in those species (Fig. 3.1). It is possible that the absence of this gene in the birds and reptiles could be due to extensive genetic degeneration or deletion, or it could simply be due to gaps in the current genome assemblies for these species. Nevertheless, there was no evidence for the presence of a functional OHCU decarboxylase gene in the genomes of the hominoids or in birds and reptile.

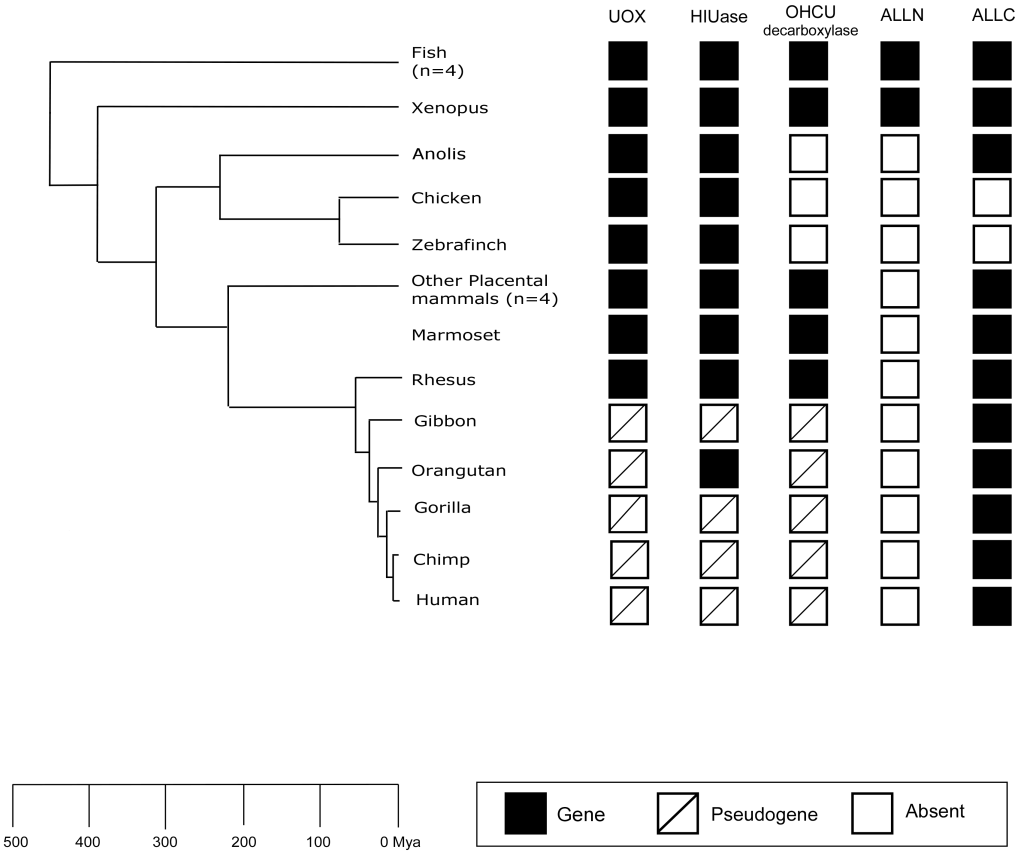


Figure 3.1 Gene content of the purine catabolic pathway in vertebrates. The tree on the left represents the accepted vertebrate phylogeny with divergence dates (millions of years ago, Mya) (Kumar and Hedges 1998b). The species included in each category are fish (stickleback, fugu and tetraodon and zebrafish), an amphibian (frog), a reptile (anolis lizard), two birds (chicken and zebra finch), and placental mammals, including the hominoids (gibbon, gorilla, orangutan, human, and chimpanzee) and other placental mammals (mouse, rat, dog, cow, rhesus monkey and marmoset). The gene content of the purine catabolic pathway for each group of species is summarized on the right. Genes were classified as functional if they had an open reading frame, were evolving under purifying selection, and had a conserved intron/exon structure. Genes were classified as pseudogenes if the protein coding sequence was truncated by early nonsense mutations. When no ortholog could be identified the gene was classified as absent.

ALLN

Functional ALLN genes were inferred to be present in the fish and amphibian lineages (Fig. 3.1). In contrast, ALLN was not detected in the genomes of the hominoids or other placental mammals (Fig. 3.1). As ALLN activity was presumably lost in the common ancestor of placental mammals > 100 million years ago, the lack of a detectable ALLN locus in these genomes likely reflects the degeneration and/or complete deletion of the inactivated ALLN locus in this lineage's distant past (Keebaugh and Thomas, 2009). Similarly, ALLN was not detected in the bird and reptile genomes. However, as with the lack of a detectable OHCU decarboxlase locus in the birds and reptile, this could reflect elimination of this locus from the genome or missing data in the current genome assemblies for these species.

ALLC

As expected, functional ALLC genes were inferred to be present in the fish and amphibian genomes (Fig. 3.1). In addition, functional ALLC genes were found in all placental mammals, including the hominoids. Thus, as has been reported previously in a smaller subset of mammals (Vigetti et al. 2003), though ALLN, the immediate upstream enzyme in the purine catabolic pathway, has been absent from the genome of placental mammals for >100 million years, ALLC has been

retained in the genome despite being despensible as a purine catabolic enzyme. Lastly, a functional ALLC gene was inferred to be present in the anolis lizard, but was absent from the two bird genomes. This variability within the birds and reptile as to the status of ALLC could reflect a true difference between these species, or once again could be the result of missing data in the bird genome assemblies.

3.3.2 Timing of the gene inactivation events in the hominoid lineage

Loss of UOX activity has been inferred to have occurred twice in the hominoid lineage, once in the gibbon lineage ~10 mya and once in the lineage leading to humans and the great apes ~15 mya (Oda et al. 2002). Assuming that the loss of UOX was the evolutionary force that precipitated the inactivation of HIU hydrolase and OHCU decarboxlase, we predicted that those genes would have begun to evolve neutrally at a time point that post-dates the inactivation of UOX. To test this hypothesis, we re-estimated the timing of the UOX inactivation events using the method described in Chou et al. 2002 and calibrated the date according to accepted divergence between hominoids and Old World monkeys (23 mya) (Raaum et al. 2005). Similar to what was reported previously, we estimated the inactivation of UOX in the gibbon lineage to have occurred ~13.1 mya and in the human and great ape lineage ~12.9 mya. Because of the small coding region of

HIU hydrolase, only a very limited number of synonymous and nonsynonymous sites were available for analysis, and thus we were unable to generate a similar estimate for the inactivation of this gene. However, the lack of an inactivating mutation common to all hominoids (Table 3.2), and the observation that the orangutan HIU hydrolase may still be evolving under purifying selection, suggests that the inactivation of HIU hydrolase in the other hominoids likely post-dates the UOX inactivations. In the case of OHCU decarboxylase we were able to generate independent estimates for the inactivation of this gene in each lineage, which ranged from ~4.8-12.6 million years ago. These estimates are consistent with the hypothesis that the degeneration of this gene began after the loss of UOX. Thus, the qualitative and quantitative estimates of the inactivation of HIU hydrolase and OHCU decarboxyase support the prediction that the loss of UOX likely triggered the genetic deterioration of the immediate downstream genes in the hominoid purine catabolic pathway.

3.3.3 Molecular characterization of the bird and reptile UOX loci

Different conclusions regarding the existence of a UOX gene in bird genomes have been reported in the literature (Przylecki and Rogalski 1927; Remy, Richert, and Westerfeld 1951; Varela-Echavarria, Montes de Oca-Luna, and Barrera-Saldana 1988; Ito et al. 1989). Our initial characterization of the loci in chicken,

zebra finch and anolis lizard were consistent with the presence of a functional gene being encoded by the UOX locus (Fig. 3.1 and Table 3.2). In addition, ESTs mapping to UOX exons 2-8 were detected in chicken, further supporting the conclusion that this locus encodes a functional gene (Fig. 3.2A). To infer the 5' end of the UOX gene in birds and confirm the transcription of this gene in another species, we performed 5' RACE and RT-PCR on cDNA generated from adult zebra finch liver total RNA. Sequencing of the resulting amplicons confirmed the transcription of exons 2-7 and detected a novel first coding exon that is also conserved in chicken. In chicken, this novel exon one encodes a predicted start codon and 3 additional amino acids (Fig. 3.2A). In particular, unlike the first exon of other UOX orthologs that are adjacent to exon 1 of DNASEB2, the first exon in zebra finch mapped upstream of the entire DNASEB2 locus (compare Fig. 3.2A and 3.2B). Finally, in the process of sequencing the zebra finch UOX partial cDNAs we identified 5 single-nucleotide polymorphisms in the protein coding portion of this gene. As would be expected for a functional gene evolving under purifying selection, the majority of these polymorphisms were synonymous (n = 4) while just 1 nonsynonymous polymorphism was observed. Thus, our analyses of the UOX transcripts in zebra finch were consistent with a functional protein being encoded by this locus, albeit one that contains a novel, protein coding first exon.

Figure 3.2a

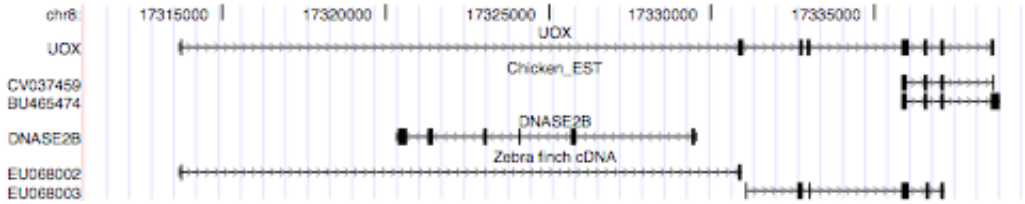


Figure 3.2b

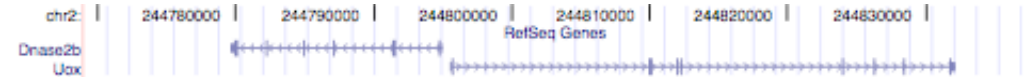


Figure 3.2 Intron/exon structure of UOX
(a) Intron/exon structure of bird UOX. Chicken ESTs indicated the potential for transcription of the bird UOX pseudogene. The 5' prime end and exon 1 of bird UOX were identified via sequencing (EU068002) and found to be located upstream of another gene, DNASE2B. (b) Intron/exon structure of mammalian UOX. In the rat (shown) and other mammals, the UOX 5 prime region and exon 1 do not span the region containing Dnase2b (Lander et al.).

3.4 Discussion

It is well established that UOX activity was lost twice independently in the hominoid lineage and, most likely, once in the common ancestor of birds and

reptiles (Urich 1994a; Oda et al. 2002). Here we have shown that, in contrast to the hominoid UOX genes, which have accumulated obvious inactivating mutations and are evolving neutrally as is expected for non-functional pseudogenes, the bird and reptile genomes contain a UOX locus associated with all the signatures of an active gene. Though a previous study had reported evidence for the transcription of UOX and expression of the UOX protein in the chicken embryo (Przylecki and Rogalski 1927; Ito et al. 1989), it has been widely accepted that UOX activity was lost in the bird and reptile lineage > 200 million years ago (Urich 1994a). In light of the detection of cDNA sequences corresponding to the UOX locus in adult chicken and zebra finch livers reported here but lack of UOX enzymatic activity in this tissue reported previously (Varela-Echavarria, Montes de Oca-Luna, and Barrera-Saldana 1988) we propose a scenario similar to that of Anderson et al., 2006. The most likely explanation is that the bird and reptile UOX gene acquired one or more mutations that led to the loss of the enzymatic activity associated with purine catabolism but escaped the fate of becoming a pseudogene. Because there is one report of UOX activity in chicken embryos (Przylecki and Rogalski 1927), we can not rule out that the mutations accounting for the loss of enzymatic activity may be further restricted to the protein isoforms expressed in the adults. Future experiments that assay recombinant bird and reptile UOX proteins will be needed to distinguish between these hypotheses.

Similar to the differential fate of the UOX locus in hominoids versus birds and reptiles, the inferred evolutionary fate of the genes downstream of UOX often varied between, and sometimes within, these clades. Overall, the results suggest that the HIU hydrolase and OHCU decarboxylase genes in hominoids are in the early stages of inactivation, likely beginning to genetically deteriorate after the loss of UOX activity. One exception to this finding was the HIU hydrolase gene in orangutan, which was classified as being functional. However, this result may simply reflect a very recent shift from gene to pseudogene in this lineage that cannot be detected by our methods. Within the birds and reptiles, no pseudogenes were detected in the purine catabolic pathway: each gene was classified as either functional or absent. Due to the incomplete nature of the genome assemblies of the birds and reptiles, the absence of any particular gene could reflect a true deletion from the genome as might be expected for an old pseudogene or could simply be an artifact of an incomplete genome assembly. However, because > 200 million years have elapsed since UOX activity was lost in birds and reptiles, we can confidently conclude that the classification of HIU hydrolase as functional in the birds and reptiles and the classification of ALLC as a functional gene in the reptile indicates that these two genes have been actively maintained in the genome, despite elimination of the requirement for these genes to function as enzymes in the purine metabolic pathway. Similarly, ALLC has been actively

maintained in the mammalian genome despite the inactivation of the upstream enzyme, ALLN (Vigetti et al. 2002; Vigetti et al. 2003).

An obvious question arises from the observation that UOX and its downstream genes were retained in birds and reptiles following an inactivation event truncating the purine catabolic pathway. Why have genes that are no longer required to function as purine catabolic enzymes been retained in the genome? Analogous to empirical data demonstrating that inactivation is the most likely fate for a duplicate gene due to loss of functional constraint (Lynch and Conery 2003), the most likely fate of a gene that has been freed of functional constraint by an inactivating mutation (either within the gene itself, or by a proximal mutation inactivating a pathway member upstream such as the UOX inactivation) is also inactivation (Tanaka, Tateno, and Gojobori 2005). Some duplicate genes, however, are retained in the genome due to the acquisition of a new function (neofunctionalization) (Ohno 1970; Proulx and Phillips 2006), the partitioning of the ancestral functions (subfunctionalization) (Hughes 1994; Force et al. 1999) or both (He and Zhang 2005). Similarly, it is possible that once a gene acquires an inactivating mutation or is rendered obsolete by a proximal mutation in a pathway, it can acquire previously ‘forbidden’ mutations that could result in the acquisition of a new beneficial function, thereby resulting in its retention in the genome. Alternatively, in cases where a gene has more than one function, a

mutation may abrogate one but not all the associated functions. Thus, a partially inactivated gene could be retained in the genome because it still encodes another function(s). Likewise, genes that are rendered obsolete by a proximal mutation in a pathway may simply be retained because they have one or more other functions (Tanaka, Tateno, and Gojobori 2005). Furthermore, the reduction in functional constraint on such genes could foster the acquisition of secondary mutations that optimzie the remaining function(s) of the gene. We therefore propose that the bird and reptile UOX locus and the genes downstream of pathway truncating mutations have been retained in the genome because they had a pre-existing important function outside of purine metabolism, or, because after becoming obsolete as purine catabolic enzymes they have acquired a novel function.

In light of these findings we propose a new model where, in addition to the direct benefits associated with inactivation (Olson 1999), a single gene loss may indirectly allow downstream genes to evolve via optimization of a secondary ancestral function, acquisition of a novel function, or both (Figure 3). Additionally, it is possible that gene inactivation removes the functional constraint imposed by direct protein-protein interactions. These proteins are then free to accumulate formerly 'forbidden' mutations in the same manner as the downstream genes. In summary, our results suggest that the evolutionary significance of gene loss can extend beyond the inactivated gene itself by altering

the evolutionary potential of many genes in the genome. How often this potential byproduct of gene loss contributes to evolutionary process is unknown and will require further investigation.

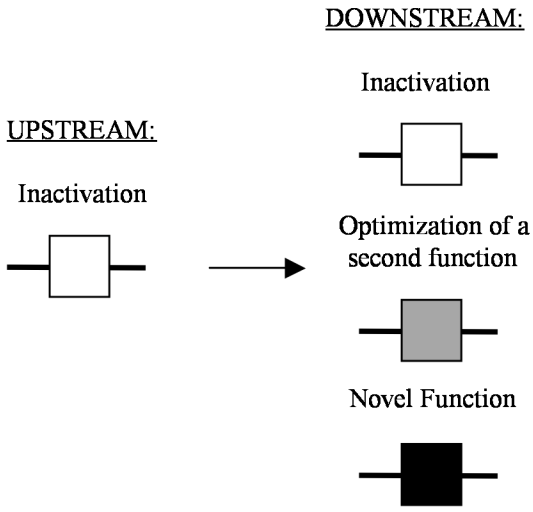


Figure 3.3 An expanded model of gene loss.
A single gene loss event removes the functional constraint imposed on the downstream genes, allowing them to be substrates of adaptive evolution. These genes can then become inactivated or evolve by optimizing a pre-existing second function, acquiring a novel function, or both. Colored blocks represent protein function/s or expression pattern/s. Empty boxes represent loss of protein function or expression pattern.

4 Gene duplication and inactivation in the HPRT-gene family[1]

[1]This chapter has been published: Keebaugh, Alaine C., Robert T. Sullivan, NISC Comparative Sequencing Program, and James W. Thomas. 2007. Gene duplication and inactivation of the vertebrate HPRT-gene family. Genomics 89(1): 134-42.

4.1 INTRODUCTION

Gene duplication is thought to play a critical role in adaptive evolution (Ohno 1970). Duplicate genes can improve the fitness of an organism by an increase in gene dosage, creating new genes with new functions by neofunctionalization (Ohno 1970), improving upon the function of a single ancestral gene by subfunctionalization (Hughes 1994), or by a combination of one or more of these mechanisms (He and Zhang 2005). Conversely, gene inactivation is also thought to be important for adaptive evolution (Ohno 1970; Olson 1999). For example, loss of function alleles can confer resistance to infectious disease in humans and mice (reviewed in (Olson 1999; Guenet and Bonhomme 2003)).

Among vertebrates, gene inactivation appears to be a recurring theme in the purine metabolic pathway. Though purine metabolism is an ancient and fundamental biological process, differential loss of enzymes in this pathway have led to variation in the end product of purine metabolism within vertebrates (Urich 1994b), and hypotheses have been put forward as to the adaptive advantage these gene loss events might confer. For example, because hominoids, birds and terrestrial reptiles lack a functional urate oxidase gene, the end product of purine metabolism in these species is uric acid. In the case of hominoids, it has been argued that since uric acid is an effective antioxidant, high levels of uric acid in

plasma may be a contributing factor to an increase in human lifespan and decrease in the incidence of age-related cancer (Ames et al. 1981). Due to the similarity in chemical structure to stimulants such as caffeine, it has also been debated whether an increase in uric acid plasma levels might be important for the increased cognitive abilities of hominoids (Haldane 1955; Orowan 1955). Though loss of function mutations in the purine metabolic pathway cause several human diseases (Boss and Seegmiller 1982), null alleles of two genes in this pathway, adenosine monophophosphate deaminase 1 and 3 (*AMPD1* and *AMPD3*), are segregating at frequencies up to 19% in human populations (Ogasawara et al. 1987; Morisaki et al. 1992). Strikingly, heterozygotes for the *AMPD1* null allele have been reported to be associated with better clinical outcomes after congestive heart failure (Loh et al. 1999) and are less likely to die of cardiovascular failure once diagnosed with coronary artery disease than are control patients (Anderson et al. 2000). Thus, not only has loss of gene function modified the purine metabolic pathway in the distant past, but it also may be a mechanism by which contemporary human populations are adaptively evolving. Here we have focused on elucidating the history of gene duplication and gene loss of the purine salvage enzyme, hypoxanthine phosphoribosyltransferase 1 (HPRT1, EC 2.4.2.8), in vertebrates.

HPRT1 is widely conserved across eukaryotes and prokaryotes and has a central role in the generation of purine nucleotides through the purine salvage pathway.

HPRT1 is a phosphoribosyl transferase domain (pfam00156) containing protein that catalyzes the conversion of hypoxanthine to inosine monophosphate (IMP) and guanine to guanosine monophosphate (GMP) via the transfer of the 5-phosphoribosyl group from 5-phosphoribosyl 1-pyrophosphate (PRPP). *HPRT1* is an X-linked gene in placental mammals and marsupials (Graves et al. 1979), whereas in other vertebrates, such as chicken, it is located on an autosome (Fukagawa et al. 1999). In humans, HPRT1-deficiency results in one of three disorders, the most severe of which is Lesch-Nyhan disease (LND, OMIM 300322). LND patients exhibit hyperuricemia, neurological dysfunction and self-injurious behavior, and although there is a good correlation between the level of HPRT1 activity and the severity of the disease, the etiology of the neurological phenotypes has yet to be established (Jinnah and Friedmann 2001b). In contrast, HPRT1-deficiency in the mouse does not lead to the neurological dysfunction or self-injurious behavior seen in humans (Hooper et al. 1987; Kuehn et al. 1987a).

To date, studies of the *HPRT1* duplication events in vertebrates have primarily focused on mammalian processed pseudogenes, which are likely the result of relatively recent retrotranspositions of the *HPRT1* messenger RNA. Though in at least one instance an *HPRT1* gene duplicate derived from a retrotransposition event was shown to encode a protein (Noyce, Conaty, and Piper 1997), until recently no other functional paralogs of *HPRT1* had been described (Nicklas

2005). Using a bioinformatic approach, Nicklas (Nicklas 2005) found an uncharacterized gene on human chromosome 10 with 68% amino acid identity to HPRT1 and a similar intron-exon structure, called phosphoribosyl transferase domain containing protein 1 (*PRTFDC1*), which is likely an ancient paralog of *HPRT1*. Here we report the systematic characterization of the vertebrate HPRT-gene family, including the gene duplications and gene inactivation events that have led to the extant variation in HPRT-gene family content across vertebrates, and discuss the potential importance of this variation for the development of a genetic model of LND.

4.2 METHODS

4.2.1 Targeted mapping and sequencing

A BAC-based targeted mapping and sequencing approach was used to sequence chromosomal segments containing HPRT-gene family members. Briefly, genomic libraries were screened as described in (Kellner et al. 2005) to identify bacterial artificial chromosome (BAC) clones containing the *HPRT1* or *PRTFDC1* loci. Selected clones were shotgun sequenced and then refined by comparative-grade finishing efforts to produce high-quality sequence in ordered and oriented contigs (Blakesley et al. 2004).

4.2.2 Database searches for additional HPRT-gene family members

Additional HPRT-gene family genomic loci were identified in public genomic databases, NCBI (http://www.ncbi.nlm.nih.gov) and the UCSC Genome Browser (http://www.genome.ucsc.edu), by sequence similarity to known HPRT1 and PRTFDC1 proteins. Full-length cDNAs were also identified by sequence similarity searches to the NCBI and TIGR (http://www.tigr.org/tdb/tgi/) gene and EST databases. The source of each gene sequence is listed in Supplementary Table 4.1.

4.2.3 Genomic sequence annotation and analysis

The intron-exon structure of each gene was inferred by cDNA-genomic (Spidey, (Wheelan, Church, and Ostell 2001)) and/or genomic-genomic (BLASTZ, (Schwartz et al. 2003)) alignments. For HPRT-gene family members where no cDNAs were available, interspecies alignments were used to infer the predicted open reading frame, protein and gene structure. The repetitive element content of each locus was determined with RepeatMasker (http://www.repeatmasker.org), and supplemented with species-specific repeat libraries for armadillo, bat, elephant, opossum, platypus and rabbit from

(ftp://ftp.ebi.ac.uk/pub/databases/ensembl/encode/repeat_libraries/). The repeat masking process was performed on species lacking a repetitive element database using the repetitive element library from the most closely related species available.

The annotated size of the individual gene loci was defined as the distance between the start and stop codons. In cases where the first exon was missing from the genomic assembly, interspecies alignments were used to identify the most 5' intron 1 sequence available and subsequently used to estimate the locus size. To measure the gene-genome size correlations, genome size estimates were based on whole-genome sequence assemblies, the Animal Genome Size Database (http://www.genomesize.com), and the literature (Redi et al. 2005) (see Supplementary Table 4.2).

4.2.4 Protein domain analysis

The domain content of each predicted protein was inferred by comparison to the cdd.v2.06 protein domain database with the NCBI conserved domain search tool (Marchler-Bauer and Bryant 2004).

2.2.5 Phylogenetic analysis

The inferred amino acid sequences were aligned with ClustalX (Jeanmougin et al. 1998b) and the resulting alignments were manually edited and used to estimate the maximum parsimony, maximum likelihood and neighbor joining phylogenies with Phylip (version 3.64) under the default parameters. For maximum likelihood and neighbor joining analyses the amino acid substitution matrix of D.T. Jones, W.K. Taylor, and J.M. Thornton was used (Jones, Taylor, and Thornton 1992). Bootstrap support was established from 1000 pseudo-replicate data sets. The protein alignments were converted to nucleotide alignments with CodonAlign (http://www.sinauer.com/hall/). Heuristic tree searching was preformed on the nucleotide alignments under maximum parsimony, maximum likelihood and Bayesian analysis. Maximum parsimony was implemented in Phylip (version 3.64). For maximum likelihood analysis, the program TREE PUZZLE (version 5.2) (Schmidt et al. 2002) and the tree searching algorithm, quartet puzzling, was used to estimate the maximum likelihood tree using the GTR+Γ+I model of substitution . For both maximum parsimony and maximum likelihood, bootstrap support was established from searches on 1000 pseudo-replicate data sets. Tree searching was also implemented in MrBayes (version 3.1) (Ronquist and Huelsenbeck 2003). Model parameters were estimated under the GTR+Γ+I model and the default priors used. Four chains were run simultaneously for 3,000,000 generations, with trees sampled every 1000 generations. After discarding the first

1500 sampled trees ("burn-in"), 1500 replicates were used to obtain clade credibility. For each method, MP, ML and Baysian only nodes with greater than 50% bootstrap support were resolved. A majority rules consensus tree was constructed from the three methods using the "consense" program in Phylip. Branch lengths were then computed in the program "baseml" in PAML (version 3.14) (Yang 1997). For all tree building methods gaps were excluded from the analysis and the *Ciona intestinalis* HPRT gene was used as the outgroup. Finally, a maximum-likelihood method implemented by the "codeml" program in PAML was used to estimate the rates of evolution at nonsynonymous (K_a) and synonymous (K_s) sites in the coding sequences. Likelihood ratio tests were used to compare models in which the number of K_a/ K_s rates varied within the tree.

4.2.6 Sequencing the mouse PRTFDC1 *exons*

Genomic DNA from wild-derived inbred mouse strains was purchased from The Jackson Laboratory and used as templates to individually PCR amplify the *PRTFDC1* exons. PCR amplicons were then isolated, sequenced and compared to the C57BL/6J genomic sequence. The panel of genomic DNA included: *M. m. domesticus*, WSB/EiJ; *M. m. musculus*, CZECH1/EiJ; *M. m. molossinus*, MOLC/RkJ; *M. m. castaneous*, CAST/EiJ; *M. spretus*, SPRET/EiJ; *M. spicilegus*,

PANCEVO/EiJ; *Mus caroli*/EiJ; and *Mus pahari*/EiJ. The sequences of the primers used to amplify each mouse *PRTFDC1* exon are available upon request.

4.2.7 Molecular dating of the mouse PRTFDC1 *inactivation event*

The method described in (Chou et al. 2002) was used to date the inactivation of the mouse *PRTFDC1* ortholog and assumes a model in which prior to inactivation the null allele was evolving under similar selective pressure as closely related functional orthologs, and then neutrally after inactivation. In the case of the inactive mouse *PRTFDC1* allele, the mouse lineage K_a would be expected to equal $f_N k(t - t_1) + k\, t_1$, where $f_N = K_a/K_s$ (in the rat + rodent lineages), k = neutral mutation rate [the average dS of the mouse + rodent and rat + rodent lineages / t = the time since the MRCA of mouse and rat (the average dS of the mouse and rat lineages / k)], and t_1 = the time since *PRTFDC1* inactivation. To derive the values necessary to solve for t_1, we used codeml from the PAML software package to analyze alignments of the coding region excluding the positions corresponding to the 5-bp frameshift insertion and nonsense codon in the mouse *PRTFDC1* gene between mouse, rat, human and opossum. Using the estimate of 88 MYA as the time since the MRCA of rodents and human (Springer et al. 2003), we solved the equation for $t_1 = [((\text{mouse lineage } K_a) / k) - f_N t] / (1 - f_N)$ with the values t_1 =

$[((0.0674) / (((0.3665 + 0.4773) / 2) / 88 \times 10^{6}) - ((0.0723) (((0.0570 + 0.1679) / 2) / 4.794 \times 10^{-9}))] / (1 - 0.0723) = 13.3$ MYA.

4.3 RESULTS

4.3.1 Compilation of the HPRT-gene family data set

Initial database searches comparing the human HPRT1 protein to a variety of vertebrate genomes revealed the presence of up to four HPRT homologs. As the first step toward resolving the evolutionary relationships and history of the vertebrate HPRT-gene family, we used targeted mapping and sequencing, and database searches to systematically assemble a large collection of these genes from a diverse panel of vertebrates (see Supplementary Table 4.1). In particular, a BAC-based targeted mapping and sequencing approach was used to identify 20 HPRT-gene family members from 15 species. The sequences of 33 additional HPRT-gene family members were also identified from publicly available genomic and/or EST and cDNA databases. In total, 53 predicted or known cDNA sequences from 28 species and 44 genomic sequences from 25 species were collected for characterization of the vertebrate HPRT-gene family.

4.3.2 Phylogeny of the HPRT-gene family

Initial sequence comparisons and gene distributions among the collection of HPRT-family members was consistent with the expected widespread distribution of *HPRT1*, as well as *PRTFDC1*, suggesting an ancient origin for both genes. However, no more than one HPRT-gene family member could be detected in the genomes of three invertebrates, *C. intestinalis*, *C. elegans*, and *D. melanogaster*, further suggesting that *PRTFDC1* might have originated as the result of a duplication event involving the *HPRT1* locus just prior to the most recent common ancestor (MRCA) of vertebrates. In addition, while in general two functional HPRT-gene family members were found in tetrapod genomes, up to four HPRT-gene family members were detected in fish, indicative of additional gene duplications or losses.

To resolve the evolutionary relationships of the HPRT homologs and estimate the timing of gene duplications or gene losses events in vertebrates, we first constructed a consensus tree using the amino acid sequences from our collection of functional HPRT-gene family members (Supplementary Fig. 4.1). Based on this protein consensus tree, the genes within the HPRT-gene family can be categorized into three groups: (1) *HPRT1*, which contains previously characterized *HPRT1* genes as well as newly identified genes from all of the vertebrate lineages sampled; (2) *PRTFDC1*, which also contains genes from all

the vertebrate lineages sampled and up to two genes (*PRTFDC1a* and *PRTFDC1b*) from some species of fish, including fugu, tetraodon and trout; and (3) *HPRT1L*, which is exclusively comprised of genes from fish.

Though the analysis of the amino acid sequences showed strong support for three distinct groups of genes within the vertebrate HPRT-gene family, the branching order within the groups was not well resolved. To better elucidate the evolutionary relationships between the genes within each group, we also constructed a concensus tree from the nucleotide sequence of the protein coding regions of each functional gene (Fig 4.1). As was observed in the protein tree, the phylogeny based on the nucleotide sequences strongly supported the *HPRT1*, *PRTFDC1* and *HPRT1L* gene groups. In addition, the branching order within each group and clustering of fish, amphibian, bird and alligator, and mammalian genes were consistent with the expected species phylogeny (Kumar and Hedges 1998a). Thus, the phylogeny of the HPRT-gene family indicates that gene duplication(s) prior to the MRCA of vertebrates and within the fish lineage are the primary basis for the expansion of the HPRT-gene family into three distinct groups and current taxonomic distribution. The inferred timing of duplications within the HPRT-gene family therefore coincide with timepoints hypothesized to be associated with large-scale duplications in the vertebrate lineage (Vandepoele et al. 2004).

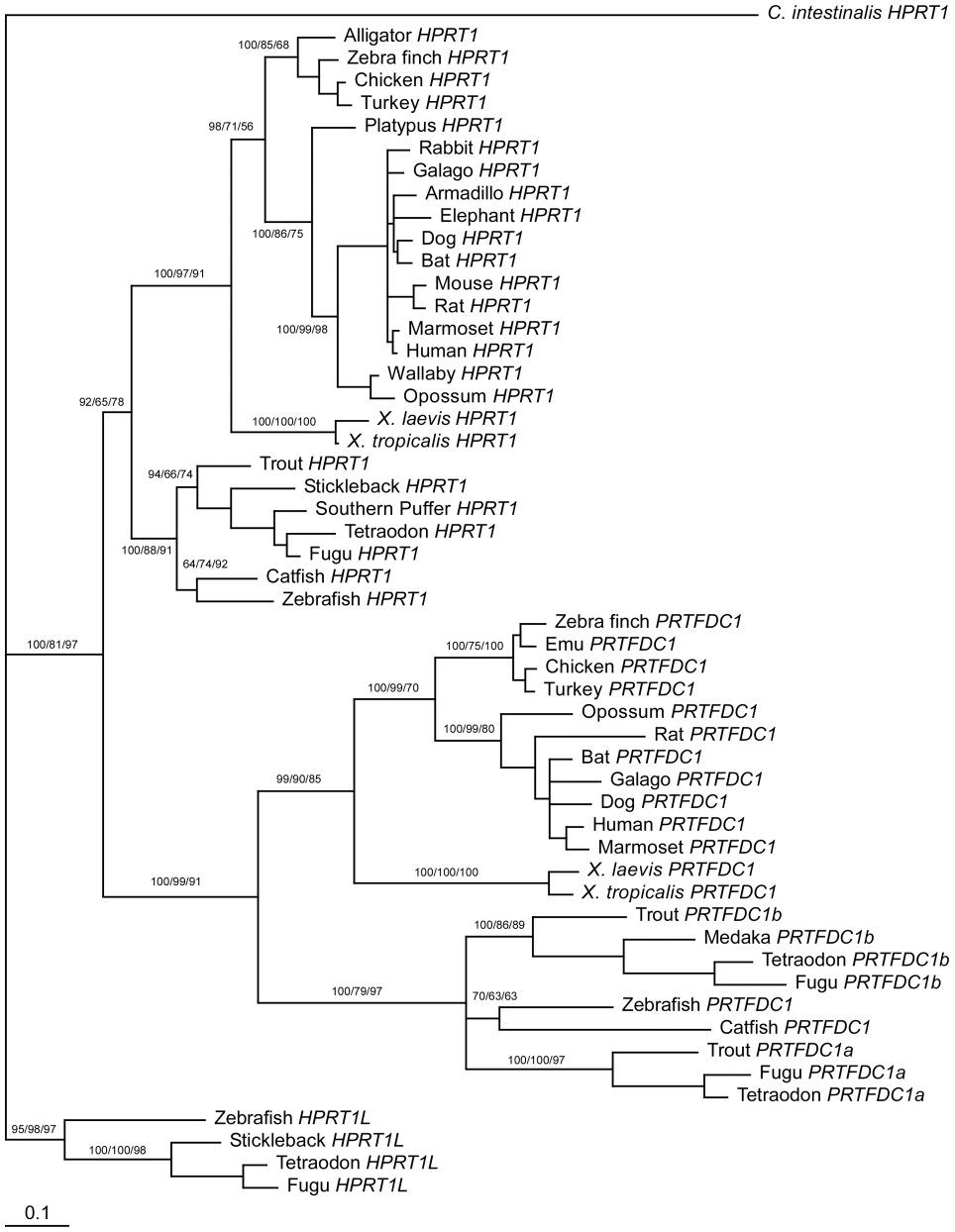


Figure 4.1 Phylogeny of the HPRT-gene family.
The nucleotide sequence of the protein coding portion of each HPRT-gene family member was used to build a consensus tree using Bayesian, maximum parsimony (MP) and maximum likelihood (ML) methods. Major nodes are labeled with the percent bootstrap support for the Bayesian, MP and ML methods, respectively. The branch lengths were estimated by ML and represent substitutions per site.

4.3.3 Genomic structure of the HPRT-gene family

Vertebrate orthologs display a high degree of conservation of intron-exon structure (Thomas et al. 2003). In contrast, the size and repetitive element content of orthologous introns tend to be strongly correlated with the genome size and repetitive element content of the individual species, and thus can be quite divergent (Vinogradov 1999; Thomas et al. 2003). In order to characterize and compare the HPRT-gene family at the genomic level, we inferred the intron-exon structure and intronic sequence composition of 44 HPRT-family members from 25 species (see Methods).

In support of their common evolutionary history, the intron-exon structures of the genes within the HPRT-gene family were found to be conserved across the entire family. Specifically, the genomic organization of each HPRT-gene family member consists of nine coding exons, of which only exons 1, 3 and 4 were observed to vary in size (Fig. 4a). However, this conserved gene structure was limited to vertebrates. For example, the HPRT homolog in *C. intestinalis* is an intronless gene, while the *C. elegans* homolog contains three introns. Thus, the conserved intron-exon structure within this gene family likely represents an ancestral genomic organization that predates the MRCA of vertebrates.

In contrast, the intron size and composition was found to be highly divergent between the HPRT-gene family members, and there was a clear correlation between genome size and the total size of the *HPRT1* (r^2=0.65), *PRTFDC1* (r^2=0.86) and *HPRT1L* (r^2=0.99) loci, with the genes from the 'large' mammalian genomes spanning the greatest chromosomal intervals and the orthologs from the 'small' puffer fish genomes the least (Fig. 4b and Supplementary Table 4.2). Similarly, the observed repetitive element content of the *HPRT1*, *PRTFDC1*, and *HPRT1L* loci were also consistent with available genome-wide averages and variation across species (Lander et al. 2001b; Aparicio et al. 2002; Waterston et al. 2002; Gibbs et al. 2004; Hillier et al. 2004; Jaillon et al. 2004) (Fig. 4b).

Interestingly, there was an unusual pattern of relative intron size between the *HPRT1* and *PRTFDC1* intraspecies paralogs. Within the mammals and birds, the *PRTFDC1* loci were 2.1- to 3.5-fold larger than the *HPRT1* loci. However, with the exception of zebrafish which had a *PRTFDC1*/*HPRT1* size ratio of 2.8, the other fish and amphibian paralogous gene pair size ratios were between 1.1-1.2 (Fig. 4b). While conservation of local variation in the rate of DNA gain and loss at the *PRTFDC1* and *HPRT1* loci could account for this difference in amniotes, surprisingly the size differential was not evenly distributed across the locus. Instead the size difference mapped primarily to a single intron, intron 3, which was 3.6- to 11.2-fold larger in *PRTFDC1* compared to *HPRT1*. The length

variation between the paralogous third intron in *PRTFDC1* and *HPRT1* accounted for 44-91% of the overall size differences, which in all cases was significantly higher than expected (p-values < 1e-25, df = 1, χ^2 test) if all the paralagous introns had expanded/contracted proportionally. Thus, while the evolution of the vertebrate HPRT-gene family introns was strongly correlated with genome-wide properties, within the amniotes, intron 3 was associated with an unusual pattern of evolution, perhaps consistent with a signature of differential selection influencing intron size within that clade.

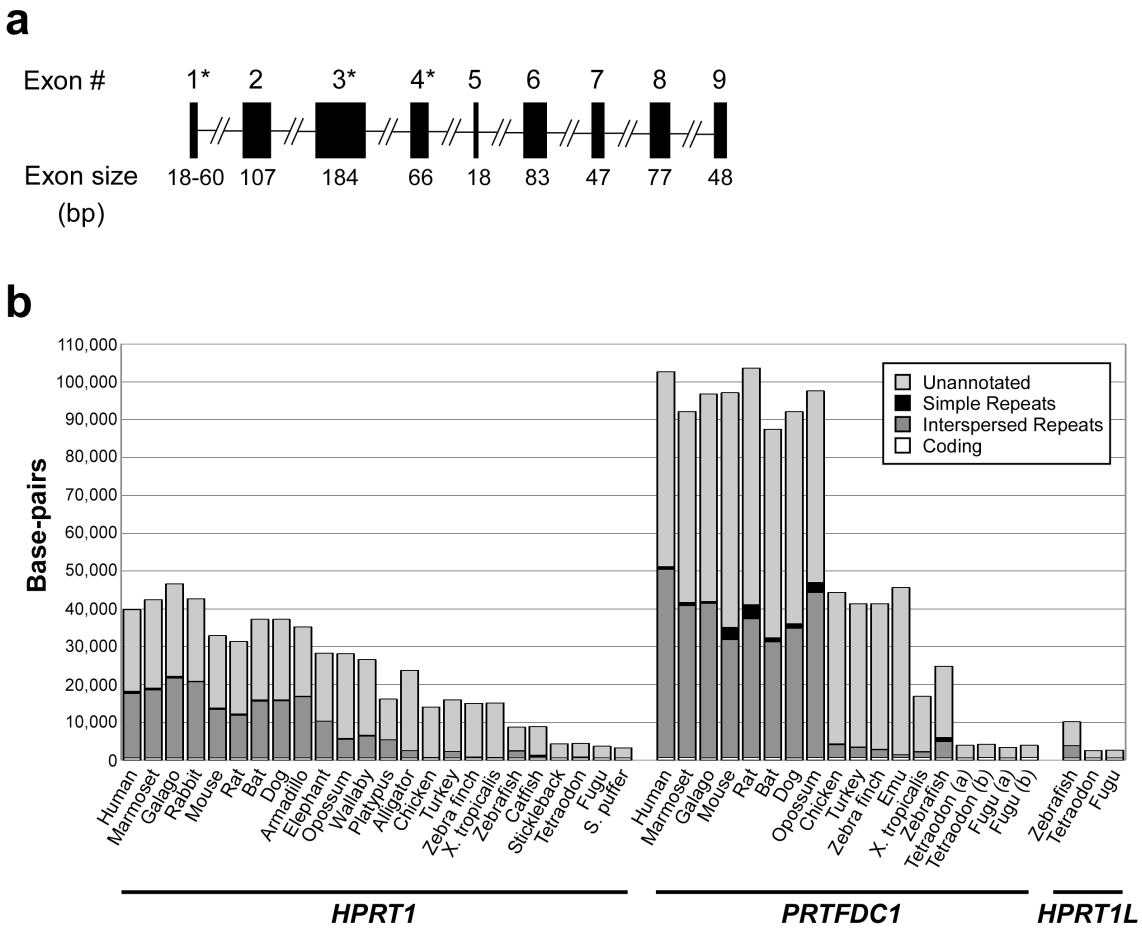


Figure 4.2 Genomic structure and sequence composition content of the HPRT-gene family loci.
(a) The consensus intron-exon structure of the vertebrate HPRT-gene family consists of nine-coding exons. Variation in exon length (*) was noted for exon 1 (18, 21, 27, 30, 33, 45, 48, 51 or 60-bp), exon 3 (184 or 181-bp in the fugu and tetraodon *PRTFDC1a* genes) and exon 4 (66 or 63-bp in the fugu and tetraodon *HPRT1L* genes). (b) Genomic sequence composition of the *HPRT1*, *PRTFDC1* and *HPRT1L* loci.

4.3.4 Predicted enzymatic activity

The only function associated with the HPRT-gene family is the ancient enzymatic role of HPRT1 as a phosphoribosyltransferase in the purine salvage pathway. Analysis of the domain content within the HPRT-gene family predicted that all of the family members contain a phosphoribosyl transferase domain (pfam00156). In addition, HPRT1, PRTFDC1 and HPRT1L proteins overall are quite similar, with the average percent identity between intraspecies HPRT1-PRTFDC1, HPRT1-HPRT1L and PRTFDC1-HPRT1L paralogs equal to 65%, 75% and 56%, respectively. We therefore sought to ascertain whether or not each gene family member had likely retained or lost the known enzymatic function of HPRT1. Specifically, we hypothesized that the genes within the *HPRT1* group would have retained the ancestral enzyme activity, while the others might have divergent function(s). To test this hypothesis, we compared the amino acid sequence of each gene at eleven conserved residues that flank or are near the active site and are presumed to be especially critical for HPRT function (Craig and Eakin 2000). As expected, the *HPRT1* orthologs with experimentally validated HPRT activity, such as those from human and chicken, had no amino acid substitutions at the eleven conserved positions. Moreover, the eleven critical amino acids were completely conserved in all the *HPRT1* genes sampled, consistent with the hypothesis that those proteins have retained the ancestral function in the purine salvage pathway. In contrast, none of the PRTFDC1 or HPRT1L proteins showed

conservation at all eleven critical sites, suggesting that each of those proteins had lost the ancestral HPRT enzymatic activity (Supplementary Table 4.3). Thus, the predicted enzymatic function, or lack thereof, was consistent with our phylogenetic classification of the HPRT-gene family.

4.3.5 Evolutionary rates in the HPRT-gene family

Because the proteins encoded by the *PRTFDC1* and *HPRT1L* genes are predicted to have lost their ancestral HPRT enzymatic activity, it is likely that these duplicate genes have been retained in the genome by virtue of evolving a new function, and thus that distinct selective forces have acted on the *HPRT1*, *PRTFDC1* and *HPRT1L* genes. The rate at which a gene has evolved can be inferred by calculating the ratio of nonsynonymous (K_a) to synonymous (K_s) substitutions, i.e. K_a/K_s. By applying a maximum-likelihood method to the nucleotide-based phylogeny shown in Figure 4.1 (see Methods), we estimated the K_a/K_s rate ratio for each of the three groups of genes and found that the *HPRT1* genes had the slowest rate of evolution, $K_a/K_s = 0.024$, *HPRT1L* an intermediate rate, $K_a/K_s = 0.053$, while *PRTFDC1* displayed the highest rate of evolution, $K_a/K_s = 0.098$. A comparison between a model in which each of the three groups had a distinct K_a/K_s value versus a model in which all three groups had the same K_a/K_s value indicated that the observed differences in the rate of evolution

between the three groups was significant ($2\Delta l = 125.83$, df = 2 p-value = 4.74×10^{-28}). Further tests directly comparing the K_a/K_s rates between all possible pairs of the three gene groups also supported the hypothesis that there was significant variation in the rate of evolution between these groups (p-values < 0.003, data not shown). Thus, while all three gene groups were found to be evolving under strong purifying selection, each group was associated with a distinct rate of evolution.

Having discovered evolutionary rate differences between the three groups of genes, we sought to determine if significant changes in the rate of evolution could be detected within the large sets of *HPRT1* and *PRTFDC1* genes. A systematic comparison of the rates of evolution between genes from the major taxonomic clades, i.e. tetrapods versus fish, and then a series of pairwise comparisons within tetrapods between placental mammals (eutherians), marsupials, birds and amphibians did not reveal any significant variation in the rate of evolution within the set of *PRTFDC1* genes. However, analogous comparisons with the *HPRT1* genes were able to partition this group by taxonomy and evolutionary rate into two distinct clusters; the eutherians, $K_a/K_s = 0.052$, and the remaining tetrapods and all fish, $K_a/K_s = 0.015$ ($2\Delta l = 37.68$, df = 1, p-value = 8.34×10^{-10}). Thus, while these results suggest that the rate of evolution of the *PRTFDC1* genes has been fairly uniform, the rate variation detected among the *HPRT1* genes is a

potential signal of a recent relaxation of selection, or positive selection, on *HPRT1* in the eutherian lineage.

4.3.6 Inactivation of PRTFDC1 *in mouse*

A mouse *PRTFDC1* locus was readily identified on mouse chromosome 2 at the predicted position within a large block of conserved synteny with the human, rat, dog and chicken genomes (Waterston et al. 2002; Gibbs et al. 2004; Hillier et al. 2004; Lindblad-Toh et al. 2005). However, closer examination of the coding sequence revealed that the mouse *PRTFDC1* gene was inactive. Specifically, the predicted mouse *PRTFDC1* open reading frame inferred from interspecies alignments to the assembled C57BL/6J genome was disrupted by three independent mutations: a missense mutation in the first exon that eliminates the start codon, a frameshift mutation in exon 2 caused by a 5-bp insertion that leads to an early stop codon and truncation of three-quarters of the protein, and a nonsense mutation in exon 8 that eliminates the final eleven amino acids. Furthermore, while interspecies genomic and cDNA alignments were consistent with a conserved intron/exon structure, including consensus splice acceptor and donor sites, no mouse ESTs or cDNAs representing a *PRTFDC1* mRNA were identified by thorough searches of public databases. In contrast, 147 and 38 ESTs

or cDNAs have been assigned to the human and rat *PRTFDC1* UniGene (Schuler 1997) clusters, respectively.

Because *PRTFDC1* is predicted to be functional in rat and all other mammals surveyed, we hypothesized that the inactive allele observed in the inbred C57BL/6J strain had arisen in the mouse lineage since the MRCA of mice and rats. In order to more precisely time the inactivation event, we sought to determine if *PRTFDC1* was also inactive in other mouse species by PCR amplifying and sequencing the coding exons of the *PRTFDC1* gene from a diverse panel of wild-derived mice (see Methods). At least one inactivating mutation in the *PRTFDC1* gene was observed in all eight strains surveyed (Table 4.1), suggesting that the inactive *PRTFDC1* allele predates the MRCA of all the species of *Mus* represented in our sample. We also used the method described in (Chou et al. 2002) to date the emergence of the inactive *PRTFDC1* allele (see Methods). In this case, based on an empirically determined divergence time between mouse and rat of 23.5 million years, we estimate that the inactivation of the mouse *PRTFDC1* allele occurred 13.3 MYA. Thus, the most parsimonious explanation for these results is that the *PRTFDC1* gene was inactivated in the mouse lineage after the divergence from the MRCA of mouse and rat. The inactive allele then became fixed in an ancestral population prior to the

divergence of the *Mus* lineages present in all laboratory and wild-derived strains of mice.

Table 4.1 Inactivating mutations in the mouse *PRTFDC1* gene.

	Mutation[a]		
Mouse Strain	1A>G	131_132insACCAC	619C>T
C57BL/6J	+	+	+
WSB/EiJ	n.a.	-	+
CZECH1/EiJ	+	+	+
MOLC/RkJ	+	+	+
CAST/EiJ	+	+	+
SPRET/EiJ	+	+	-
PANCEVO/EiJ	+	n.a.	-
Mus caroli/EiJ	n.a.	+[b]	+
Mus pahari/EiJ	n.a.	-	+

(+) indicates the mutation was present, (-) that the mutation was absent, and (n.a) that sequence was not available.

[a] The position of the mutations are based on coordinates from the human *PRTFDC1* coding sequence (GenBank Ac# NM_020200) and the inferred mutation by comparison to the predicted rat coding sequence (GenBank Ac# XM_214518).

[b] The mutation in *Mus caroli*/EiJ is 131_132insACCCCAC.

4.4 DISCUSSION

As a result of a combination of important properties including enzymatic function in the purine salvage pathway, the development and subsequent widespread use of *HPRT1* as a somatic cell genetic marker, and clinical significance of HPRT1-deficiency in human disease, *HPRT1* is perhaps one of the best characterized

genes in the human genome. Here we have reported for the first time a systematic study of the evolution of the *HPRT1* locus and HPRT-gene family in vertebrates.

Phylogenetic analysis indicated that the vertebrate HPRT-gene family consists of three groups of genes: *HPRT1*, which encodes proteins with predicted or known HPRT activity; and *PRTFDC1* and *HPRT1L*, which arose as the result of the duplication of the *HPRT1* locus. In particular, we favor a model in which a gene duplication of the *HPRT1* locus prior to the MRCA of vertebrates some ~450 MYA gave rise to *PRTFDC1*. Subsequent duplication of the *PRTFDC1* locus in the fish lineage followed by the differential retention or loss of the gene duplicates then led to the observation that some fish genomes have two *PRTFDC1* genes, while others have only one. Both of these interpretations are compatible with the whole-genome and/or other large-scale duplication events hypothesized to have occurred prior to the MRCA of vertebrates and subsequently in a common ancestor of just the ray-finned fish (Vandepoele et al. 2004). In contrast, though the *HPRT1L* was only found in fish, the position of the *HPRT1L* orthologs within the tree suggests an origin prior to the divergence of fish and tetrapods. Thus, *HPRT1L* may have arisen as the result of a duplication event prior to the MRCA of vertebrates, but was subsequently lost in the tetrapod lineage. Alternatively, *HPRT1L* may have arisen more recently by duplication in the fish lineage, but the tree-building methods we used here did not faithfully reconstruct the true history

of this gene. Nonetheless, our findings indicate that three ancient duplication events resulted in the expansion of the vertebrate HPRT-gene family into three groups: *HPRT1*, *PRTFDC1*, and *HPRT1L*.

All members of the vertebrate HPRT-gene family share a common intron-exon structure, suggesting that the extant genomic organization of these genes has been conserved over a cumulative evolutionary timescale of several billion years. In contrast, the intron length and composition of orthologs within each of the three gene clades were highly divergent and representative of genome-wide differences between the species. Intraspecies comparisons of paralogous *HPRT1* and *PRTFDC1* loci revealed a clear difference in the size of the introns of these genes that was primarily restricted to intron 3 in mammals and birds. Though the length of paralogous introns is not tightly correlated (Yandell et al. 2006), the general conservation of the length variation of *HPRT1*/*PRTFDC1* intron 3 across mammals and birds is perhaps suggestive of selection acting at the level of genomic size (Vinogradov 2004).

The function(s) of the *PRTFDC1* and *HPRT1L* genes are unknown, but like HPRT1, both the PRTFDC1 and HPRT1L proteins contain a phosphoribosyl transferase domain. However, comparison of the amino acid composition at conserved sites critical for HPRT enzymatic activity predicted that all PRTFDC1

and HPRT1L proteins have lost this ancestral function. In addition, the ability to select for the presence or absence of HPRT activity, for example in human and chicken cell lines which also contain the *PRTFDC1* gene, provides indirect experimental evidence that the function of PRTFDC1 is not redundant with HPRT1 in this regard. Therefore, *PRTFDC1* and *HPRT1L* have likely been retained in the genome by virtue of acquiring a new and yet to be determined function as predicted by the neofunctionalization model (Ohno 1970).

Similar to most protein coding genes, the members of the HPRT-gene family were found to be evolving under strong purifying selection ($K_a/K_s < 0.1$) (Gibbs et al. 2004). Significant variation in the rate of protein evolution was observed between the three groups of genes as well as among the *HPRT1* genes. In particular, the eutherian *HPRT1* genes were found to be evolving at an elevated rate relative to other tetrapods and fish, and likely also to the ancestral rate of evolution. One potential cause of the observed elevated rate is the position of *HPRT1* on the X chromosome in eutherians compared to its autosomal linkage in most other vertebrate genomes. However, though not statistically significant, it should be noted that the marsupial *HPRT1* genes, which are also X-linked (Graves et al. 1979), were evolving at a slower rate than their eutherian orthologs ($K_a/K_s = 0.02$ versus $K_a/K_s = 0.052$, $2\Delta l = 1.52$, df = 1, p-value = 0.08). Thus,

other factors are also likely involved in this observed difference in evolutionary rate of eutherian *HPRT1* orthologs.

Finally, we detected a recent gene inactivation event specific to the *Mus* lineage in *PRTFDC1*. Without any knowledge regarding the function of this gene, hypotheses regarding the basis of the fixation of the *PRTFDC1* null allele in mice are limited to general mechanisms, including genetic drift, hitchhiking, or positive selection. Regardless of the means by which *PRTFDC1* function was lost in the mouse lineage, the inactivation and fixation of this null allele is strikingly reminiscent of the presumably adaptive gene inactivations that have occurred in the vertebrate purine metabolic pathway. While the potential selective advantage the loss of *PRTFDC1* function did or did not confer to an ancestral population of mice can not be directly tested, we propose that the lack of PRTFDC1 activity in the mouse is a strong genetic candidate for the phenotypic difference between HPRT-deficient humans and mice. In particular, we hypothesize that the presence of PRTFDC1 is required for the manifestation of the severe neurological phenotypes observed in HPRT-deficient humans, which are absent in HPRT-deficient mice. In other words, the progression of the neurological symptoms is dependent on the mis-regulation of PRTFDC1 either by the absence of direct protein-protein interactions with HPRT1, or indirectly, for example via the increased levels of PRPP in HPRT-deficient cells which have been shown to up-

regulate the activity of other phosphoribosyl transferase domain containing proteins (Jinnah and Friedmann 2001b). Our model therefore has two testable hypotheses: (1) that HPRT-deficiency in any other mammal will likely result in a neurological phenotype; and (2) that the reintroduction of an active copy of *PRTFDC1* into HPRT-deficient mice will lead to a neurological phenotype, and thus a genetic model for LND.

In conclusion, a comparative genomics approach was used to determine a detailed evolutionary history of the HPRT-gene family in vertebrates. The series of 'evolutionary alleles' (Xue and Noll 1996) for the *HPRT1* and *PRTFDC1* genes reported here in the form of genomic clones and sequences from a diverse set of vertebrates will provide invaluable resources for future experimental studies designed to dissect and compare the function of these genes, particularly in light of their evolutionary history and role in human disease.

5 Initial characterization of a novel mouse model for Lesch-Nyhan Disease[1]

[1]A portion of this chapter is in preparation for publication.

5.1 Introduction

The enzyme, hypoxanthine phosphoribosyltransferase, *HPRT1*, has a central role in the recycling of purine nucleotides through the purine salvage pathway (Figure 5.1). Mutations to *HPRT1* result in Lesch-Nyhan Disease (LND, OMIM 300322), an X-linked disorder with an incidence of 3/million persons. Patients with LND exhibit severe developmental and neurological abnormalities such as testicular atrophy, hypotonia, dysarthria, mental retardation, and compulsive aggressive self-injurious behavior (Jinnah and Friedmann 2001a). Furthermore, these patients exhibit a 70-90% loss of dopamine in the basal ganglia (Lloyd et al. 1981). The mechanism by which *HPRT1* deficiency leads to the neurobehavioral abnormalities and the dopamine loss seen in LND patients is not well understood.

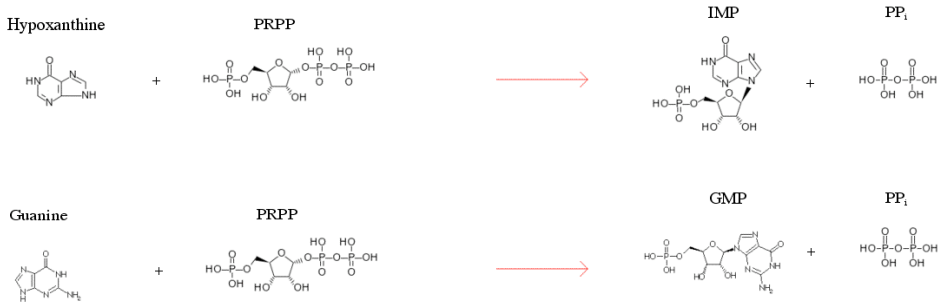


Figure 5.1 Catalytic role of *HPRT1* in purine salvage.
HPRT1 catalyzes conversion of hypoxanthine to inosine monophosphate (*IMP*) and guanine to guanosine monophosphate (*GMP*) via transfer of the 5-phosphoribosyl group (PPi) from 5-phosphoribosyl 1-pyrophosphate (*PRPP*).

In order to gain a better understanding of the neurological consequences of *HPRT1*-deficiency, animal models have been developed through both pharmacological (Lloyd et al. 1981; Minana and Grisolia 1986; Nehlig, Daval, and Debry 1992) and gene manipulation methods (Hooper et al. 1987) (Kuehn et al. 1987b). Of the pharmacological models developed, some do not directly block HPRT1 activity but rather illicit self-injurious behavior by means of neonatal dopamine loss in rats. (Breese et al. 1984). While some of the pharmacological models do display an "enhanced behavioral sensitivity" to amphetamines, none address the full spectrum of LND phenotypes. In the mouse genetic models of LND, subtle differences in brain dopamine levels were observed and they did recapitulate the biochemical hallmarks of human patients (Finger et al. 1988; Dunnett et al. 1989; Jinnah, Gage, and Friedmann 1991; Jinnah, Langlais, and Friedmann 1992; Jinnah et al. 1994; Jinnah et al. 1999); however, they do not display any of the neurobehavioral disorders or testicular atrophy seen in human patients (Finger et al. 1988; Jinnah, Gage, and Friedmann 1991; Jinnah, Langlais, and Friedmann 1992). Thus, the current animal models for LND remain incomplete and the etiology of *HPRT1*-deficiency in the nervous system and testes remains largely unknown. To date there is no treatment for the neurological aspects of LND.

Since the initial *Hprt1*-gene knockout models were developed, several groups have tried to make better, more "humanized" models for LND by accounting for genetic differences in purine metabolism between humans and mice. Unlike mice and other mammals, humans lack an enzyme downstream of *HPRT1* in the purine metabolic pathway called urate oxidase (*UOX*) (Keilin 1959; Wu et al. 1992). *UOX* is an interesting candidate

because loss of *UOX* exacerbates the accumulation of uric acid. It was hypothesized that these increased levels of uric acid were associated with the neurobehavioral phenotypes (Henderson 1968). This isn't likely the case. Treatment of LND patients with uric acid degrading drugs has no influence on the development of neurobehavioral abnormalities (Marks et al. 1968). Additionally, it has been reported that *Hprt1/Uox* double mutants do not develop neurological or behavioral phenotypes (Jinnah and Friedmann 2001a). Another genetic candidate tested was *APRT*, another salvage enzyme of purine metabolism. It was hypothesized that, compared to humans, mice were more reliant on *Aprt* than *Hprt1* for purine salvage. However, *Hprt1/Aprt* double mutants do not represent a better model than the initial *Hprt1* knockout (Wu and Melton 1993a; Engle et al. 1996).

UOX and *APRT* are clearly not the only genetic differences between mice and man. However, prior to the availability of whole genome sequences, insufficient genetic information was available to identify and characterize genetic differences between vertebrates on a large scale. Recently, an *HPRT1* paralog in humans was briefly described, *PRTFDC1* (Nicklas 2006). Subsequently, *PRTFDC1* was further characterized among vertebrates through a large-scale comparative genomics study of the *HPRT1* gene family (Keebaugh, Sullivan, and Thomas 2007), where *PRTFDC1* was shown to have emerged from a duplication event of *HPRT1* prior to the diversification of vertebrates. Interestingly, *PRTFDC1* was found to be functional in all vertebrates surveyed except the mouse. Keebaugh et. al. hypothesized *PRTFDC1* as a candidate

genetic difference accounting for the disparity in phenotypes between *Hprt1*-deficient mice and man (Keebaugh, Sullivan, and Thomas 2007).

Currently, the function of *PRTFDC1* is unknown and the only function associated with *HPRT1* is its ancient enzymatic role in purine metabolism. It is not likely that *PRTFDC1* has redundant enzymatic activity with *HPRT1* because loss of *HPRT1* function can be selected for in human cell lines, which contain a functional *PRTFDC1* locus. *PRTFDC1* is a member of the phosphoribosyl transferase domain containing family and thus is predicted to bind *PRPP*. In *HPRT1*-deficient cells *PRPP* is elevated and leads to the up-regulation of all other characterized proteins in this gene family (n=8). Homology modeling of *PRTFDC1* from the *HPRT1* crystal structure also supports the prediction that *PRTFDC1* binds *PRPP* and identified a fixed difference between vertebrate *HPRT1* and *PRTFDC1*, D137G, which is hypothesized to result in *PRTFDC1* binding a different base than does *HPRT1* (guanine and hypoxanthine; see Figure 5.1). *PRTFDC1* has also been identified as a potential tumor suppressor gene (Suzuki et al. 2007; Cai 2007). It is likely that *PRTFDC1* has retained the phosphoribosyl transferase domain containing activity and the ability to bind *PRPP* but has acquired a novel function that is coupled to purine metabolism via an interaction with *PRPP*.

Until recently *HPRT1* was not known to physically interact with any other protein. Homology modeling predicts that *PRTFDC1* and *HPRT1* proteins can directly interact through the formation of heterotetramers and yeast 2-hybird studies have detected an interaction between *PRTFDC1* and *HPRT1* (Rual et al. 2005). Additionally, both genes

are ubiquitously expressed in human, indicating that while *PRTFDC1* and *HPRT1* genes likely have divergent functions, they do have broad and overlapping expression patterns. Thus, *PRTFDC1* and *HPRT1* may be able to regulate one another through direct protein-protein interactions or indirect regulatory interactions. Furthermore, lack of *Prtfdc1* in mouse represents an excellent candidate gene that contributes to the lack of LND like phenotypes in *Hprt1*-deficient mice. Thus, we further hypothesize that the presence of *PRTFDC1* in humans enhances or is required for the phenotypes associated with the loss of *HPRT1* and LND, whereas the absence of *PRTFDC1* in mice acts to suppress those phenotypes.

To directly test the hypothesis that *PRTFDC1* is an enhancer/suppressor of LND we have generated *Hprt1*-deficient mice carrying the human *PRTFDC1* locus. This more humanized model of LND will allow us to characterize the expression profile and interactions of *PRTFDC1* relative to *HPRT1* and LND. As a means towards this end, we have characterized various aspects of this transgenic mouse model including (i) viability and fertility phenotypes, (Hillier et al.) protein expression, (iii) transcript profile, and (iv) potential molecular interactions between *Hprt1* and $Tg^{Prtfdc1+/0}$, establishing a basic body of knowledge for the *PRTFDC1* gene and its potential interactions with *HPRT1* and LND. While there is no evidence for indirect regulatory interactions between *Hprt1* and $Tg^{Prtfdc1+/0}$, there is evidence for a direct interaction between these two proteins. The relationship between this physical interaction and LND is not yet clear; however, behavioral testing is currently underway.

5.2 Methods

5.2.1 Identification and characterization of a human PRTFDC1 BAC clone.

The human BAC clone, RP11-129O7, was identified based on gene location in the human genome and validated by restriction-enzyme fingerprint analysis and PCR. The BAC restriction-enzyme fingerprints matched the virtual DNA digests of the corresponding sequence from human chromosome 10. PCR amplification of sequences from the 5' and 3' flanks, and introns 1 and 7 of the *PRTFDC1* gene were also consistent with the expected composition of the human *PRTFDC1* BAC. PCR amplification of sequences included regions from the 5' (F: 5'-TCTGTGTGCAGTGTTACCTGTG-3'; R: 5'-CAAACCAACTGAAAGAGGTTCC) and 3' (F: 5'-ACAGATGGTTGGTTCACTTTCC-3'; R: 5'-GAGGCAGGTTTCATAGGTGAAC-3') flanks as well as within introns 1 (F: 5'-TCTCCTGGGATCAGAGTAGACC-3'; R: 5'-ACATCAAGTCCTAAGCGAGAGG-3') and 7 (F:5'-CTGAAGTCTGCCTATGCAAGTG-3'; R:5'-TGTTCTCTAGGGCTTCCTCATC-3') of the PRTFDC1 gene. Cycle conditions were 94°C for 5 minutes followed by 35 cycles of 94°C for 30s, 58°C for 30s, and 72°C for 1 minute, followed by a final 7-minute extension at 72°C.

5.2.2 Generation of the $Tg^{Prtfdc1+/0}$ transgenic mouse.

The BAC RP11-129O07 was linearized with PI-*Sce*1, gel isolated, and agarase treated. The Emory Mouse Transgenic Core (http://corelabs.emory.edu/home.cfm#tmc) performed pronuclear microinjections of our BAC-transgene into fertilized FVB eggs. To

genotype the mice according to the presence/absence of the *PRTFDC1* transgene, crude DNA preps from tail clips were taken at day 14 and the PCR primers from the initial validation step were used to screen for transgenic founders.

5.2.3 Breeding & genotyping strategy.

In order to recapitulate the human disease genotype, an active *PRTFDC1* gene must be introduced into the *Hprt1*-deficient mouse genome. The current *Hprt1*-deficient mouse is in the C57BL/6J background; thus, our breeding strategy was to cross $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ mice in the FVB background to $Prtfdc1^{-/0}$ $Hprt1^{+/0}$ mice in the C57BL/6J background. Furthermore, in order to minimize the effect that transgene integration site might have on any observed phenotypes, two independent transgenic lines were generated, L9 and L13. Once in the C57BL/6J background (F5 generation), males from L9 and L13 were bred with $Prtfdc1^{-/0}$ $Hprt1^{-/-}$ females, $Hprt1^{b\text{-}m3}$ (Jackson Laboratories)(Hooper et al. 1987), to produce $Tg^{Prtfdc1+/0}$ $Hprt^{-/0}$ and $Prtfdc1^{-/0}$ $Hprt1^{-/0}$ (Table 5.1). In addition, $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ males were bred to $Prtfdc1^{-/0}$ $Hprt1^{+/+}$ females, $Hprt1^{b\text{-}m3}$ (Jackson Laboratories), to produce $Tg^{Prtfdc1+/0}Hprt1^{+/0}$ and $Prtfdc1^{-/0}$ $Hprt1^{+/0}$ males. The $Tg^{Prtfdc1+/0}Hprt1^{+/0}$, $Prtfdc1^{-/0}$ $Hprt1^{+/0}$, and $Prtfdc1^{-/0}$ $Hprt1^{-/0}$ provide important controls in the phenotype analysis. The *PRTFDC1* transgene was followed by PCR (see validation step) and the *Hprt1* allele was genotyped following a modified protocol as described in (McEwan and Melton 2003). Briefly, a multiplex PCR flanking the *Hprt1* breakpoint was used to amplify both the deleted *Hprt1* gene and the wildtype *Hprt1* gene. In total, we characterized the phenotypes of six genotypes [1: $Prtfdc^{-/0}$ $Hprt1^{+/0}$, 2: $Prtfdc^{-/0}$ $Hprt1^{-/0}$, 3-4: $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ (i.e. both transgenic lines), and 5-6: $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ (i.e.

both transgenic lines)] from four different crosses to assess the effect of *PRTFDC1* expression on Hprt1-deficient mice.

Table 5.1

5.2.4 Gross phenotyping.

Litters were closely monitored on the day of birth and thereafter until weaning to measure the number of pups that **1)** die before weaning, **2)** fail to feed, as indicated by a lack of milk in the stomach, **3)** are clearly runted, and **4)** do not walk or have difficulty moving. Dead pups were tail clipped and genotyped. Genotype frequencies were calculated under a chi-square distribution. $Tg^{Prtfdc1+/0}Hprt1^{-/0}$ males from both transgenic lines (> 8 weeks of age; n=3/line) were mated with wild-type C57BL/6J females to determine if male $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ mice were fertile.

5.2.5 Tissue collection.

Following euthanization by CO_2 inhalation, brain, testis and liver were collected at p0, p28 and 7 week-old male mice representing six genotypes [1: $Prtfdc^{-/0}$ $Hprt1^{+/0}$, 2: $Prtfdc^{-/0}$ $Hprt1^{-/0}$, 3-4: $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$, and 5-6: $Prtfdc1^{+/0}$ $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$]. Tissues were immediately frozen on dry ice and stored at -80°C until RNA and protein extraction.

5.2.6 Protein analysis.

Ice-cold RIPA buffer was added to frozen tissue at 3mL per gram. Tissues were homogenized and incubated on ice for 30 minutes. Lysate was centrifuged at approximately 12,000 rpm for 30 minutes at 4°C. Lysate was collected and protein concentration determined by Bradford assay. Protein (40ug) was loaded into a 12% polyacrylamide gel and separated by SDS-polyacrylamide electrophoresis. Protein was transferred to a nitrocellulose membrane (New England Biolabs, Inc.). Membranes were blocked in 5% non-fat dry milk and 0.2% Tween20. Membranes were incubated with primary antibody for 1 hour at room temperature (Prtfdc1: Proteintech Group; Hprt1 & β-actin: Santa Cruz Biotechnology). Antibody binding was visualized using HRP-linked IgG secondary antibody (Jackson ImmunoResearch Laboratories, Inc.). Immunoreactive bands were visualized via chemilumescence (New England Biolabs, Inc.) on film (Hyperfilm-ECL Amersham, Arlington Heights, IL). Films were analyzed using the Kodak Imaging Software (version 4.0).

5.2.7 Expression studies.

Total mouse RNA from brain, was prepared using the RNeasy Lipid tissue Mini Kit (Quiagen, Hilden, Germany) according to the manufacture's recommendations for whole brain. Briefly, RNeasy Lipid Tissue Kits integrate phenol/guanidine-based sample lysis, designed to facilitate lysis of fatty tissues and inhibit RNases, and silicon membrane purification of total RNA. Total mouse RNA from testis and liver was prepared from Trizol (Invitrogen) according to manufacture's recommendations. cDNA was synthesized from 5ug of total mouse RNA or human RNA (Clontech Laboratories, Inc.,

Mountain View, CA and US Biologicals, Swampscott, MA) using Superscript II reverse transcriptase (Invitrogen) primed with oligo(dT) according to manufacture's recommendations. Two cDNA's were synthesized for each RNA extraction and four replicate quantitative reverse-transcription PCRs (qRT-PCR) were completed for each cDNA. Relative transcript abundance was measured using qRT-PCR on a Roche LightCycler V3 for three mouse genes (*Actb*: NM_007393.1 , *RPol2a*: NM_009089.2, and *Hprt1*: NM_013556.2) and four human genes (*ACTB*: NM_001101.3, *RPOL2a*: NM_000937.2, *HPRT1*: NM_000194.2 and *PRTFDC1*: NM_020200.5). Gene specific primers were purchased from SABiosciences (http://www.sabiosciences.com/) and validated for amplification of only a single product of the correct size and high PCR efficiency. SyberGreen (Invitrogen) reactions were preformed in 20ul volumes according to manufacture's recommendations. Cycle conditions were 95°C for 10 minutes followed by 45 cycles of 95°C for 15s, 60°C for 60s and a final 40s extension at 72°C. A dissociation curve was run to ensure that only a single product was amplified. To control for efficiencies of RNA extraction and cDNA synthesis, we regressed the inverse of each critical threshold value ($1/C_T$) from the qRT-PCR against the corresponding '$1/ C_T$ ' measure for two internal control genes, *Actb* and *RPol2a*. Relative expression was measured as the fold change in expression, $2^{(-(\Delta\Delta CT))}$. RT-PCR was preformed on cDNA's from both lines at 7 weeks to detect residual expression of *PRTFDC1* and *Hprt1*. Primers used for amplification of PRTFDC1 were F: 5'-GTCTTCCCTTCCCGCGTTCC-3' and R: 5'-GCTTCCCAAGTACAGGCGTAAA-3'. Primers used for *Hprt1* amplification were F: 5'-GCTGGATTACATCAAAGCACTG-3' and R: 5'-CAAGGGCATATCCTACAACAAAC-3'. Cycle conditions were 94°C for 5 minutes

followed by 35 cycles of 94°C for 30s, 58°C for 30s, and 72°C for 1 minute, followed by a final 7-minute extension at 72°C.

5.2.8 Statistical analysis.

Analysis of variance (ANOVA) using regression residuals of $1/C_T$ was preformed to test for genotype effects with sequential Bonferroni corrections (Rice 1989). Using raw C_T values provided similar results. Differences in expression levels were considered significant when the p-value was less than 0.05, relative to both control genes, *Actb* and *RPol2a*.

5.3 Results

5.3.1 Relative Transcript Abundance in Human.

Preliminary analyses of public expression databases indicate that *PRTFDC1* is broadly expressed in humans (http://www.ncbi.nlm.nih.gov/geo/; data not shown). To refine the expression profile of *PRTFDC1* in different tissues of relevance to LND, we characterized the normal expression profile of *PRTFDC1* relative to *HPRT1* in a broad panel of human tissue with an emphasis on brain and testis (Figure 5.2). We found *PRTFDC1* to be expressed higher than *HPRT1* in all brain tissues, including fetal whole brain. While previous studies have yielded mixed reports as to the levels and pattern of HPRT expression in the brain (ref), we found that *HPRT1* transcript abundance (normalized to *ACTB* and *RPOL2a*) in the substantia nigra and cortex was significantly different than the putamen, caudate nucleus, cerebellum and whole brain. Similarly,

PRTFDC1 transcript abundance (normalized to *ACTB* and *RPOL2a*) in the cortex was significantly different from that of the putamen, caudate nucleus, cerebellum, substantia nigra and whole brain. These differences, while significant, were subtle. In contrast to brain, *HPRT1* was expressed higher than *PRTFDC1* in testis and no difference in expression was seen in adult and fetal liver. While the patterns of *PRTFDC1* and *HPRT1* expression varied between tissues and at different time points in development, the results are consistent with a ubiquitous pattern of expression.

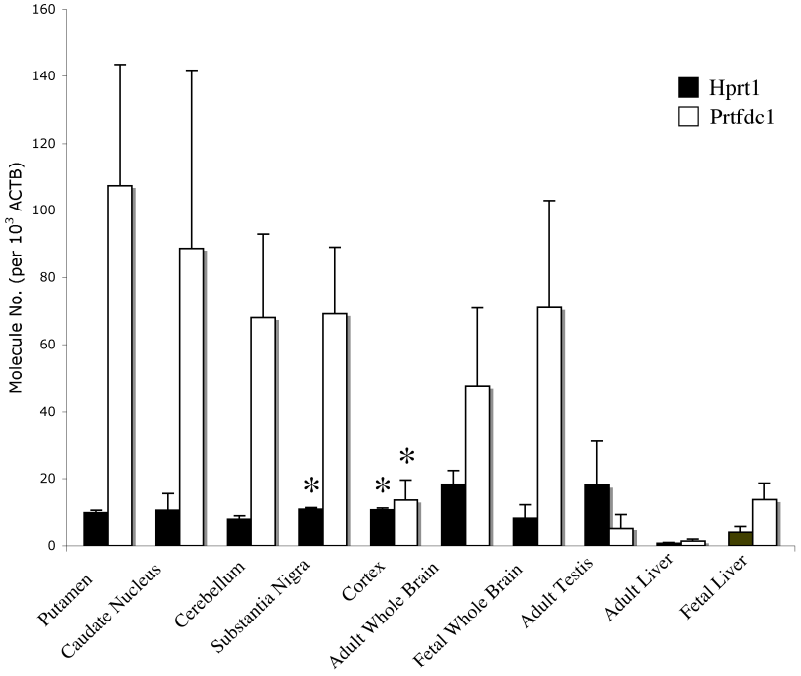


Figure 5.2 Transcript profiles of *PRTFDC1* and *HPRT1* in human tissue.
Relative transcript abundance of *HPRT1* (black) and *PRTFDC1* (white) is shown per 10,000 *ACTB* molecules. * Indicates a significant difference ($p<0.05$) in expression between human brain sub-regions, including whole brain.

5.3.2 Generation of the PRTFDC1 transgenic mouse.

In order to generate mice expressing *PRTFDC1,* a BAC clone containing the human locus (Fig 5.3) was identified and used to generate transgenic mice by pronuclear

injection. Four transgenic founders were identified, and confirmed to express human *PRTFDC1* transcripts by RT-PCR. In addition, initial qRT-PCR studies indicated that all four lines expressed *PRTFDC1* at similar levels. Furthermore, we observed no skew in *PRTFDC1* transmission (n≥99/line), where we expected *PRTFDC1* to be transmitted to 50% of the offspring (n_{total}=417 individuals genotyped at postnatal day 14). No gross phenotypes were observed in transgenic carriers. As there were no observable differences between the transgenic lines, we arbitrarily picked two lines for detailed characterization.

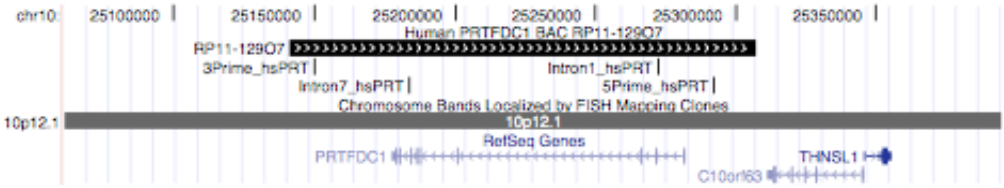


Figure 5.3 Chromosomal location and genotyping assay for human *PRTFDC1* BAC clone.
The position of the sequenced human BAC clone (RP11-29O7 (GenBank Acc. AL512598) is shown relative to human chromosome 10 (NCBI Build #17) and the *PRTFDC1* locus. This BAC contains significant amounts of sequence 5' and 3' of *PRTFDC1* but does not contain any other genes. The arrows indicate the direction of gene transcription. Tall and short rectangles indicate coding exons and UTRs, respectively.

5.3.3 Gross phenotyping.

Because LND is an X-linked recessive disorder that primarily affects males, we have focused our phenotype analysis on the male mice. Given that LND patients display developmental motor delay within the first year of life, never learn to walk, and have low

birth weights and a slow growth curve (Jinnah and Friedmann 2001a), we expect those phenotypes to manifest post-natally as a general failure to thrive. To assess this phenotype in $Tg^{Prtfdc1+/0}$ $Hprt^{-/0}$ offspring, $Tg^{Prtfdc1+/0}$ $Hprt^{+/0}$ males were mated with $Prtfdc1^{-/0}$ $Hprt^{-/-}$ females (Table 5.1). Transmission of the *PRTFDC1* transgene was normal for lines 9 and 13 in *Hprt1*-deficient males (p=0.32 and 0.78, respectively; Table 5.1). Furthermore, we did not see any evidence of pups that (1) died before weaning, (2) failed to feed, (3) were clearly runted, and (4) did not walk or had difficulty moving. Thus, $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ mice were no different with respect to their ability to thrive as compared to their $Prtfdc1^{-/0}$ $Hprt1^{-/0}$ littermates.

Table 5.1 *PRTFDC1* transmission frequencies in *Hprt1*-deficient males.

	$Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ x $Prtfdc1^{-/0}$ $Hprt1^{-/-}$	
	$Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$	$Prtfdc1^{-/0}$ $Hprt1^{-/0}$
Line 9	81	94
Line 13	103	107

In addition to basic viability, because male LND patients undergo testicular atrophy and fail to advance to puberty, we set up breeding tests for n=3 $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ males from each transgenic line to determine if they were fertile. All of the males tested sired offspring; thus, $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ males are both viable and fertile.

5.3.4 Protein analysis.

Protein expression was analyzed in brain, testis and liver via western bolt (Figure 5.2). As expected, we were able to detect a band for PRTFDC1 (26 kDa) using a PRTFDC1

polyclonal antibody in genotypes Tg$^{Prtfdc1+/0}$ *Hprt1*$^{+/0}$ and Tg$^{Prtfdc1+/0}$ *Hprt1*$^{-/0}$ but we did not detect a band for genotype *Prtfdc1*$^{-/0}$ *Hprt1*$^{+/0}$. Antibody for PRTFDC1 cross-reacted with HPRT1 (23.4 kDa), which was detected in genotypes *Prtfdc1*$^{-/0}$ *Hprt1*$^{+/0}$ and Tg$^{Prtfdc1+/0}$ *Hprt1*$^{+/0}$; HPRT1 protein was not detected in genotype Tg$^{Prtfdc1+/0}$ *Hprt1*$^{-/0}$. The same pattern of HPRT1 protein presence/absence was detected using an HPRT1 monoclonal antibody (data not shown). Unexpectedly, comparison of the relative levels of PRTFDC1 between the genotypes Tg$^{Prtfdc1+/0}$ *Hprt1*$^{+/0}$ and Tg$^{Prtfdc1+/0}$ *Hprt1*$^{-/0}$ indicated that PRTFDC1 levels were lower in the absence of HPRT1 versus when HPRT1 was present. Specifically, a 4-fold decrease in PRTFDC1 expression was observed in the absence of HPRT1 in brain for transgenic line 9 (Figure 5.4) and 2-fold decrease was observed in line 13 (data not shown). Similarly, a 1.5 fold decrease was observed in testis in both lines (data not shown). There was no difference in PRTFDC1 expression between genotypes in liver (data not shown).

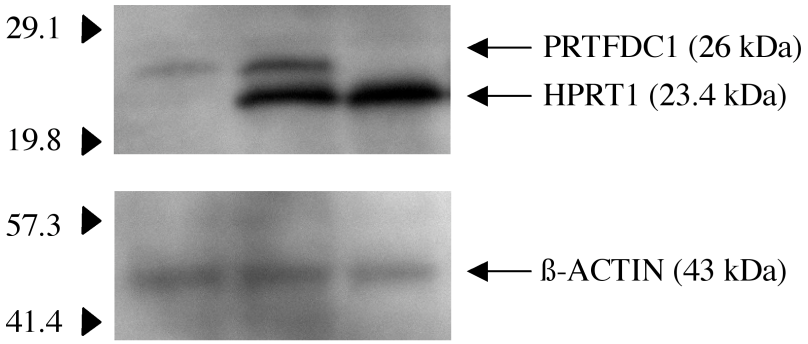


Figure 5.4 HPRT1 and PRTFDC1 protein expression in line 9.
Western blots from line 9, 7 week brains were prepared and probed with anit-PRTFDC1 as described. Lane 1, $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$; lane 2, $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$; lane 3, $Prtfdc1^{-/0}Hprt1^{+/0}$. The molecular weight (kDa) of the marker is indicated to the left and protein reactivity to the right. β-ACTIN (43 kDa) was detected in all genotypes and used as a control to quantify PRTFDC1 and HPRT1 expression.

5.3.5 Relative Transcript Abundance in Transgenic Mice.

Given the difference in PRTFDC1 protein levels in the presence and absence of HPRT1, we were interested in looking at transcript levels of both *Hprt1* and *Prtfdc1*. In both transgenic lines, we measured relative transcript levels of *Hprt1* and *Prtfdc1* across age-matched brain, testis and liver at pre- (p0), and post- (p28 and 7 weeks) sexual maturity (Figure 5.5). We compared *Prtfdc1* expression in the presence/absence of *Hprt1* ($Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ *vs.* $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$) and *Hprt1* expression in the presence/absence of *Prtfdc1* ($Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ *vs.* $Prtfdc1^{-/0}$ $Hprt1^{+/0}$). In total, 70 pair-wise comparisons between genotypes in various tissues and developmental stages were preformed, including tissues where the change in PRTFDC1 protein expression was observed (see Figure 5.5). Four significant differences ($p<0.05$) were detected: *Prtfdc1* expression in

L13 7 week and p28 brain, *Prtfdc1* expression in L9 p28 testis, and *Hprt1* expression in L13 p28 testis. These significant differences were not replicated across lines. Thus, no reproducible differences were observed in the level of *Hprt1* or *Prtfdc1* transcripts.

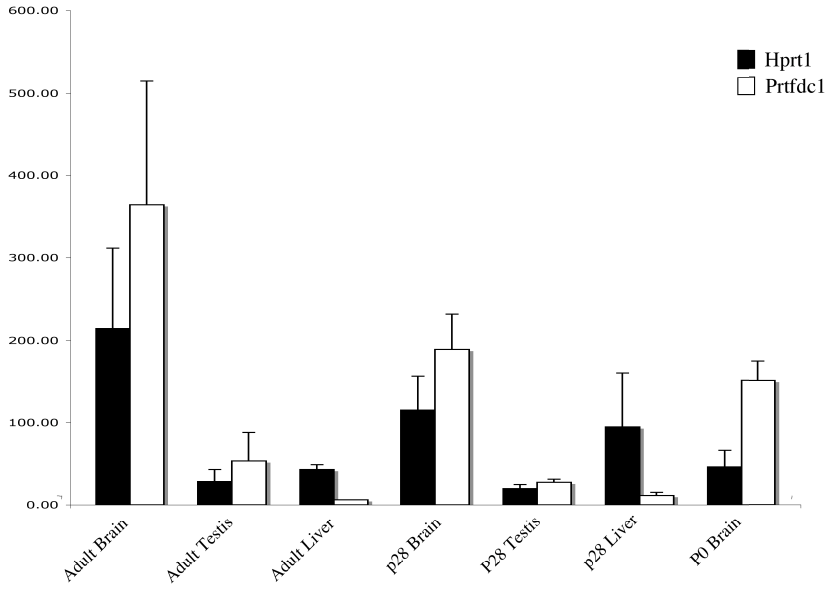


Figure 5.5 Transcript profiles of *Prtfdc1* and *Hprt1* in mouse tissues.
Relative transcript abundance of *Prtfdc1* (white) and *Hprt1* (black) per 10,000 *Actb* molecules in the $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$ mouse. No differences were observed between genotypes or transgenic lines.

While it is generally difficult to compare transcript levels across species, an important aspect of modeling human biology in mouse is determining if the model recapitulates the human expression patterns. In brain, the primary tissue affected by LND, no obvious differences in *Prtfdc1* and *Hprt1* expression patterns were observed between the transgenic mouse ($Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$) and human.

5.4 Discussion

Despite decades of research, the pathophysiology of LND remains unknown. Animal models are currently available, however each is limited in its utility for studying the neurological aspects of this disease. A null *Hprt1* allele was previously created at the *HPRT1* locus in mouse (Hooper et al. 1987). Given that mice do not have a functional *Prtfdc1* allele, *Hprt1*-deficient mice do not have the same genotype as *HPRT1*-deficient humans. To test the hypothesis that the absence of *PRTFDC1* in mouse acts to suppress the phenotypes associated with LND, we have generated a more humanized *Hprt1*-deficient mouse carrying the human *PRTFDC1* allele.

Hprt1-deficient male mice carrying the human *PRTFDC1* allele do not display any obvious phenotypic abnormalities with respect to viability and fertility when compared to littermates ($Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ *vs.* $Tg^{Prtfdc1+/0}$ $Hprt1^{+/0}$, $Prtfdc1^{-/0}$ $Hprt1^{-/0}$, *and* $Prtfdc1^{-/0}$ $Hprt1^{+/0}$). Currently, behavioral studies are underway to detect any neurological phenotype associated with this new model. Specifically, three key aspects associated with the neurobehavioral abnormalities of LND are being assessed. The altered dopamine system of the basal ganglia is being assessed via administration of an amphetamine that promotes dopamine release. Second, compulsive self-injurious behavior is being measured by stereotypy. And, finally, aggression is being evaluated by resident-intruder aggression tests. This battery of behavioral experiments will test whether the presence of *PRTFDC1* modifies the neurobehavioral phenotype in *Hprt1*-deficient mice. One limitation of the gross phenotyping and behavioral experiments is

that they have only been carried out in males heterozygous for the presence of *Prtfdc1*. There could be a dosage dependent effect. Increasing the dosage (i.e. $Tg^{Prtfdc1+/+}$) could affect the phenotype. Furthermore, our analysis does not provide information on how *Prtfdc1* expression affects the female phenotype. Future experiments could include males homozygous for the presence of *Prtfdc1* as well as the full complement of females.

PRTFDC1 was expressed in all tissue types analyzed, as seen in human. It is unclear if the levels of expression are comparable across species. Furthermore, the genotype (i.e. the presence and absence of either *Prtfdc1*and *Hprt1*) did not have any significant effect on transcript levels of *Prtfdc1* or *Hprt1*. A key assumption in studying mRNA expression is that it is informative in the prediction of protein expression (Chen et al. 2002; Greenbaum et al. 2003). Transcript levels, however, did not accurately reflect the amount of protein in all tissues. Specifically, PRTFDC1 protein expression decreased in the absence of HPRT1 in brain and testis, the two tissues affected by LND.
These results indicate that transcriptional regulation is not the mechanism accounting for the observed differences in PRTFDC1 expression but rather suggest that translational processes control the expression of PRTFDC1.

While the mechanism by which HPRT1 influences PRTFDC1 mean protein levels remains to be elucidated, different biological mRNA and protein degradation rates might affect mRNA and protein correlations (Guo et al. 2008). Conceivably, direct protein-protein interactions between HPRT1 and PRTFDC1 may act to stabilize PRTFDC1 by increasing its half-life. Formation of an HPRT1-PRTFDC1 heterotetramer, as supported

by homology modeling, is a potential mechanism for this stability. It is possible that the loss of an interaction with a functional HPRT1 protein leads to the mis-regulation of PRTFDC1 and manifestation of the clinically relevant phenotypes associated with LND. Recently, there have been numerous studies exploring the relationship between mRNA and protein expression. The results from studies have been largely inconsistent in both yeast (Futcher et al. 1999; Gygi et al. 1999; Greenbaum et al. 2003) and human (Anderson and Seilhamer 1997; Chen et al. 2002; Lichtinghagen et al. 2002; Guo et al. 2008), suggesting that gene expression at the transcriptional and translational levels are more likely to be discordant for genes involved in development, regulation, and disease. Of outstanding interest is to determine if the same pattern of PRTFDC1 expression is observed in human brain and testis. One potential avenue for progress lies in comparing protein expression from wildtype human cell lines and matched HPRT1-deficient patient cell lines.

The utility of this new animal model for studying the etiology of LND is promising. In addition to the negative correlation seen in PRTFDC1 protein expression in the presence and absence of HPRT1, preliminary behavioral studies in line 9 have identified an increased trend toward aggression and stereotypy in $Tg^{Prtfdc1+/0}$ $Hprt1^{-/0}$ mice. This behavioral phenotype is more representative of the neurobehavioral phenotype seen in patients with LND as compared to the other mouse genetic animal models of LND developed to date. Thus, this more humanized model potentially represents a new and improved model for studying the neurobehavioral syndrome of LND.

Previous studies in *Hprt1*-deficient mice provide indirect evidence for an underlying abnormality of catecholaminergic function in the brains of *Hprt1*-deficient mice, which is analogous to the abnormalities of dopamine function seen in patients with LND (Lloyd et al. 1981). The idea is that the neurobehavioral phenotype seen in LND patients may be the result of over stimulation of supersensitive dopaminergic receptors in the basal ganglia (Breese et al. 1984; Casas-Bruge et al. 1985; Goldstein et al. 1986). LND patients show a 70-90% reduction (Lloyd et al. 1981), while HPRT1-deficient mice show only a 19-45% reduction in basal ganglia dopamine (Dunnett et al. 1989; Jinnah, Langlais, and Friedmann 1992; Jinnah et al. 1999). Changes in dopamine receptor sensitivity do not occur unless tissue dopamine levels are reduced by more than 80% (Zigmond et al. 1990) and this fact may account for the absence of dopamine receptor sensitivity (and neurobehavioral dysfunction) in *Hprt1*-deficient mice as compared to LND patients. One possibility is that *PRTFDC1* acts as a modifier in the threshold effect seen with dopaminergic loss and neurobehavioral dysfunction. Future experiments could address this hypothesis by measuring dopamine levels in the basal ganglia to determine if the presence of *Prtfdc1* could further reduce basal ganglia dopamine in *Hprt1*-deficient mice. In anticipation of this study we have collected tissues from the basal ganglia (nucleus accumbems and caudoputamen) and olfactory bulb for each of the six genotypes (n=8/genotype).

The two hypotheses discussed, (i) mis-regulation of PRTFDC1 via post-translational modification and (Hillier et al.) dopamine hypersensitivity, are not mutually exclusive and one can easily imagine a situation where *HPRT1*-deficiency leads to post-

translational modification of PRTFDC1 which in turn alters the dopamine system of the basal ganglia. If basal ganglia dopaminergic neuronal loss is key to the development of this pathway, PRTFDC1 may be acting as a modifier in the threshold effect seen with dopaminergic loss and neurobehavioral dysfunction. While we still have a long way to go in characterizing the usefulness of this model and the etiology of LND, therapeutics targeting down-regulation of PRTFDC1 might prove be clinically efficacious, if shown to be safe in humans.

Gene loss has been hypothesized to play a role in adaptive evolution (Olson 1999). While the significance, if any, of the loss of *PRTFDC1* in the mouse is unknown the difference in gene content it caused between humans and mice has the potential to provide a striking example of how gene loss can act as a disease modifier. Furthermore, this hypothesis highlights the importance of considering gene content differences when modeling human disease in other species.

6 Discussion

6 Discussion

Independent gene inactivation events have truncated the purine catabolic pathway and as a consequence the end product of purine catabolism varies amongst vertebrates. Mammals have traditionally been grouped into two classes based on their end product of purine catabolism: most mammals, whose end product is allantoin due to the loss of ALLN and the hominoids, whose end product is uric acid due to the loss of UOX (Keilin 1959; Urich 1994a). Information regarding the end product of this pathway in marsupials and monotremes, however, is limited and it is possible that purine catabolism proceeds further in these mammals than it does in the placental mammals. In order to infer the likely functional content of the purine catabolic pathway in the marsupials and monotremes, we compared the gene content in the terminal portion of the purine metabolic pathway in a marsupial and a monotreme to a diverse set of vertebrates in chapter 2. We found that unlike other mammals, the marsupial and monotreme encode the full complement of purine catabolic genes. This analysis identified that loss of ALLN is specific to the placental mammals and represents a previously undetected genetic difference distinguishing marsupials/monotremes from placental mammals. These results suggested that marsupials and monotremes have end products of purine catabolism comparable to that of fish and amphibians and thus represent a third class of mammals. Biochemical studies that measure and compare the levels of intermediate metabolites in the purine catabolic pathway in urine from marsupials and monotremes will ultimately be required to test this hypothesis. The presence of glyoxylate and urea in the urine would confirm that marsupials and monotremes are distinct from the other mammals. Loss of

ALLN in placental mammals is intriguing; while fixation could have been due to drift, it might also have led to a phenotypic innovation distinct to placental mammals, such as the placenta. One potential way to test the biological significance of this loss would be to alter the mouse genome to include the full complement of genes assumed to have been present in the mammalian ancestor.

Chapter 2 highlights the significance of gene loss in shaping interspecies differences in purine catabolism and reveals novel insights into the importance of gene loss during mammalian evolution. Under the "less is more hypothesis" gene loss is proposed as playing a role in adaptive evolution by conferring an immediate and direct benefit under certain environmental conditions. However, the potential consequence of gene loss on the fate of other genes, connected by function, is poorly understood. In chapter 3 we used the purine metabolic pathway as a model system in which to explore this question and present a new model of adaptive gene loss. Under this model, a single gene loss event removes the functional constraint imposed on the downstream genes and affords them the opportunity to optimize a secondary ancestral function, acquire a novel function, or both. These results broaden the view of adaptive gene loss, demonstrating that a single gene loss event can have long-term, indirect consequences on the adaptive potential of many genes in a pathway like purine catabolism. While these results show that gene loss, like gene duplication, can provide the raw material for evolutionary innovation, they do not provide a quantitative measure of the contribution of this process to evolution. Like purine metabolism, the VB6 pathway has experienced multiple, independent gene loss events (Tanaka, Tateno, and Gojobori 2005). It would be interesting to determine the

frequency at which downstream genes have been retained following gene loss in various metabolic pathways like purine and VB6 metabolism.

Gene duplication has also led to interspecies differences with respect to the salvage portion of purine metabolism. Phylogenetic analysis in chapter 4revealed that ancient vertebrate-specific gene duplication events resulted in expansion of the vertebrate HPRT-gene family. Specifically, our findings indicated that duplication of the HPRT1 locus gave rise to PRTFDC1 about 450MYA and subsequent duplication events specific to the ray-finned fish likely gave rise to HPRT1L and a second copy of PRTFDC1. Both of these interpretations are compatible with the whole-genome and/or other large-scale duplication events hypothesized to have occurred prior to the MRCA of vertebrates and then subsequently in the MRCA of the ray-finned fish (Vandepoele et al. 2004). One confounding factor to this interpretation was the position of HPRT1L in the tree, which suggested an origin prior to the divergence of fish and tetrapods. Thus, an alternative interpretation is that HPRT1L emerged from a duplication event of HPRT1 prior to the diversification of vertebrates and was subsequently lost in the tetrapod lineage. Sequence comparisons indicated that both PRTFDC1 and HPRT1L lost their ancient enzymatic function associated with purine metabolism and thus were retained due to the acquisition of a novel function/s as predicted by the neofunctionalization model.

Among mammals, we identified an inactivation event of PRTFDC1 specific to the mouse lineage. In light of its conservation in all other vertebrate lineages surveyed, the adaptive significance for the likely fixation of this null allele is intriguing. Fixation of the null

PRTFDC1 allele in the *Mus* lineage was demonstrated by sequencing of nine wild derived inbreed strains. Because inbreed strains are not representative of the genetic variation seen in the natural population, it is possible that the null allele is not fixed; however, this is highly unlikely given the broad geographic range of the species sampled and level of divergence. Future studies could sample more individuals from wild populations to provide further evidence for this fixation event. Moreover, this inactivation event represents a new genetic difference between humans and mice and we propose PRTFDC1 as a candidate genetic difference contributing to the phenotypic disparity seen between HPRT-deficient mice and LND patients.

While HPRT1-deficiency in humans leads to Lesch-Nyhan Disease, HPRT1-deficient mice do not develop the neurobehavioral abnormalities seen in LND patients. It has long been hypothesized that genetic differences between humans and mice account for this phenotypic disparity. We developed a genetic mouse model to experimentally evaluate the role of PRTFDC1 as a candidate genetic difference in contributing to this difference in phenotype in chapter 5. At a gross phenotypic level, the mouse model did not recapitulate the human phenotype. At the molecular level, the results provide evidence for a direct interaction between HPRT1 and PRTFDC1 in the brain and testis, the two tissues affected by the disease. It will be of interest to determine if these protein expression patterns also occur in human tissues. The relationship between PRTFDC1 and the neurobehavioral abnormalities associated with LND, however, has not yet been fully evaluated in this transgenic model. Behavioral studies are in progress to determine

if the presence of PRTFDC1, in both the presence and absence of HPRT1, has a behavioral phenotype.

In addition to the neurobehavioral phenotype associated with LND, abnormalities of dopamine function have also been observed in LND patients (Lloyd et al. 1981). It is hypothesized that over stimulation of supersensitive dopaminergic receptors in the basal ganglia result in the neurobehavioral phenotype of LND patients (Breese et al. 1984). While HPRT1-deficient mice have reduced basal ganglia dopamine, their reduction is not as severe as that seen in human patients and thus they do not experience dopamine receptor sensitivity (Lloyd et al. 1981). It is possible that PRTFDC1 acts as a modifier in the threshold effect seen with dopaminergic loss and neurobehavioral dysfunction. While the dopamine system of the basal ganglia in our transgenic model is being assessed via behavioral studies, testing of such a 'threshold effect' is beyond the scope of this dissertation. Future studies, however, could experimentally measure the effect of PRTFDC1 on dopamine levels, in both the presence and absence of HPRT1, to determine if the presence of PRTFDC1 further reduces basal ganglia dopamine in mouse.

Loss of PRTFDC1 in mouse has the potential to provide a striking example of how considering gene content differences between humans and mice resulting from gene loss can be critical when modeling human disease in other species. There is a precedent for creating an improved mouse model of human disease in cases where the gene of interest has recently been inactivated. Like PRTFDC1, cholesteryl ester transfer protein (CETP) was recently inactivated in the rodent lineage (Ha and Barter 1982). To test whether or

not CETP was associated with atherosclerosis, transgenic mice expressing human CETP were developed (Lie et al. 2001). In this case the presence of CETP was shown to make the mice more susceptible to atherosclerosis and to be a better model of human disease (Lie et al. 2001). As previously mentioned, other genetic differences between human and mouse have been considered with respect to HPRT1-defcieint mice, for example loss of UOX in human. If PRTFDC1 is shown to modify susceptibility to LND it may be useful to generate HPRT1/UOX double knockout mice carrying a functional PRTFDC1 locus. The extent to which a "more humanized" model will be informative is unclear and this may represent an even better model for understanding the etiology of LND.

In conclusion, the data presented in this dissertation establishes an evolutionary picture of how gene duplication and gene loss have shaped the purine metabolic pathway and the potential importance of those events with respect to modeling human disease. In the broader context, the results reported contribute to an expanded model of adaptive gene loss, and directly assess the significance of one particular gene loss event with respect to modeling a human disease.

7 References

Abraham, J. M., C. S. Freitag, J. R. Clements, and B. I. Eisenstein. 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of Escherichia coli. Proc Natl Acad Sci U S A **82**:5724-5727.

Aburomia, R., O. Khaner, and A. Sidow. 2003. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. J Struct Funct Genomics **3**:45-52.

Agewall, S., and B. Norman. 2006. Association between AMPD1 gene polymorphism and coagulation factors in patients with coronary heart disease. Pathophysiol Haemost Thromb **35**:440-444.

Aherne, A., A. Kennan, P. F. Kenna, N. McNally, D. G. Lloyd, I. L. Alberts, A. S. Kiang, M. M. Humphries, C. Ayuso, P. C. Engel, J. J. Gu, B. S. Mitchell, G. J. Farrar, and P. Humphries. 2004. On the molecular pathology of neurodegeneration in IMPDH1-based retinitis pigmentosa. Hum Mol Genet **13**:641-650.

Akum, B. F., M. Chen, S. I. Gunderson, G. M. Riefler, M. M. Scerri-Hansen, and B. L. Firestein. 2004. Cypin regulates dendrite patterning in hippocampal neurons by promoting microtubule assembly. Nat Neurosci **7**:145-152.

Ames, B. N., R. Cathcart, E. Schwiers, and P. Hochstein. 1981. Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. Proc Natl Acad Sci U S A **78**:6858-6862.

Andersen, O., T. S. Aas, S. Skugor, H. Takle, S. van Nes, B. Grisdale-Helland, S. J. Helland, and B. F. Terjesen. 2006. Purine-induced expression of urate oxidase and enzyme activity in Atlantic salmon (Salmo salar). Cloning of urate oxidase liver cDNA from three teleost species and the African lungfish Protopterus annectens. Febs J **273**:2839-2850.

Anderson, J. L., J. Habashi, J. F. Carlquist, J. B. Muhlestein, B. D. Horne, T. L. Bair, R. R. Pearson, and N. Hart. 2000. A common variant of the AMPD1 gene predicts improved cardiovascular survival in patients with coronary artery disease. J Am Coll Cardiol **36**:1248-1252.

Anderson, L., and J. Seilhamer. 1997. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis **18**:533-537.

Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J. M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M. D. Gelpke, J. Roach, T. Oh, I. Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S. F. Smith, M. S. Clark, Y. J. Edwards, N. Doggett, A. Zharkikh, S. V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y. H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, and S. Brenner. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science **297**:1301-1310.

Ardlie, K. G. 1998. Putting the brake on drive: meiotic drive of t haplotypes in natural populations of mice. Trends Genet **14**:189-193.

Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler. 2002. Recent segmental duplications in the human genome. Science **297**:1003-1007.

Balasubramanian, S., D. Zheng, Y. J. Liu, G. Fang, A. Frankish, N. Carriero, R. Robilotto, P. Cayting, and M. Gerstein. 2009. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. Genome Biol **10**:R2.

Bentley, D. R.S. BalasubramanianH. P. SwerdlowG. P. SmithJ. MiltonC. G. BrownK. P. HallD. J. EversC. L. BarnesH. R. BignellJ. M. BoutellJ. BryantR. J. CarterR. Keira CheethamA. J. CoxD. J. EllisM. R. FlatbushN. A. GormleyS. J. HumphrayL. J. IrvingM. S. KarbelashviliS. M. KirkH. LiX. LiuK. S. MaisingerL. J. MurrayB. ObradovicT. OstM. L. ParkinsonM. R. PrattI. M. RasolonjatovoM. T. ReedR. RigattiC. RodighieroM. T. RossA. SabotS. V. SankarA. ScallyG. P. SchrothM. E. SmithV. P. SmithA. SpiridouP. E. TorranceS. S. TzonevE. H. VermaasK. WalterX. WuL. ZhangM. D. AlamC. AnastasiI. C. AnieboD. M. BaileyI. R. BancarzS. BanerjeeS. G. BarbourP. A. BaybayanV. A. BenoitK. F. BensonC. BevisP. J. BlackA. BoodhunJ. S. BrennanJ. A. BridghamR. C. BrownA. A. BrownD. H. BuermannA. A. BunduJ. C. BurrowsN. P. CarterN. CastilloE. C. M. ChiaraS. ChangR. Neil CooleyN. R. CrakeO. O. DadaK. D. DiakoumakosB. Dominguez-FernandezD. J. EarnshawU. C. EgbujorD. W. ElmoreS. S. EtchinM. R. EwanM. FedurcoL. J. FraserK. V. Fuentes FajardoW. Scott FureyD. GeorgeK. J. GietzenC. P. GoddardG. S. GoldaP. A. GranieriD. E. GreenD. L. GustafsonN. F. HansenK. HarnishC. D. HaudenschildN. I. HeyerM. M. HimsJ. T. HoA. M. HorganK. HoschlerS. HurwitzD. V. IvanovM. Q. JohnsonT. JamesT. A. Huw JonesG. D. KangT. H. KerelskaA. D. KerseyI. KhrebtukovaA. P. KindwallZ. KingsburyP. I. Kokko-GonzalesA. KumarM. A. LaurentC. T. LawleyS. E. LeeX. LeeA. K. LiaoJ. A. LochM. LokS. LuoR. M. MammenJ. W. MartinP. G. McCauleyP. McNittP. MehtaK. W. MoonJ. W. MullensT. NewingtonZ. NingB. Ling NgS. M. NovoM. J. O'NeillM. A. OsborneA. OsnowskiO. OstadanL. L. ParaschosL. PickeringA. C. PikeA. C. PikeD. Chris PinkardD. P. PliskinJ. PodhaskyV. J. QuijanoC. RaczyV. H. RaeS. R. RawlingsA. Chiva RodriguezP. M. RoeJ. RogersM. C. Rogert BacigalupoN. RomanovA. RomieuR. K. RothN. J. RourkeS. T. RuedigerE. RusmanR. M. Sanches-KuiperM. R. SchenkerJ. M. SeoaneR. J. ShawM. K. ShiverS. W. ShortN. L. SiztoJ. P. SluisM. A. SmithJ. Ernest Sohna SohnaE. J. SpenceK. StevensN. SuttonL. SzajkowskiC. L. TregidgoG. TurcattiS. VandevondeleY. VerhovskyS. M. VirkS. WakelinG. C. WalcottJ. WangG. J. WorsleyJ. YanL. YauM. ZuerleinJ. RogersJ. C. MullikinM. E. HurlesN. J. McCookeJ. S. WestF. L. OaksP. L. LundbergD. KlenermanR. Durbin, and A. J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456**:53-59.

Blakesley, R. W., N. F. Hansen, J. C. Mullikin, P. J. Thomas, J. C. McDowell, B. Maskeri, A. C. Young, B. Benjamin, S. Y. Brooks, B. I. Coleman, J. Gupta, S. L. Ho, E. M. Karlins, Q. L. Maduro, S. Stantripop, C. Tsurgeon, J. L. Vogt, M. A. Walker, C. A. Masiello, X. Guan, G. G. Bouffard, and E. D. Green. 2004. An intermediate grade of finished genomic sequence suitable for comparative analyses. Genome Res **14**:2235-2244.

Blomme, T., K. Vandepoele, S. De Bodt, C. Simillion, S. Maere, and Y. Van de Peer. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol **7**:R43.

Boss, G. R., and J. E. Seegmiller. 1982. Genetic defects in human purine and pyrimidine metabolism. Annu Rev Genet **16**:297-328.

Bowne, S. J., L. S. Sullivan, S. H. Blanton, C. L. Cepko, S. Blackshaw, D. G. Birch, D. Hughbanks-Wheaton, J. R. Heckenlively, and S. P. Daiger. 2002. Mutations in the inosine monophosphate dehydrogenase 1 gene (IMPDH1) cause the RP10 form of autosomal dominant retinitis pigmentosa. Hum Mol Genet **11**:559-568.

Brawand, D., W. Wahli, and H. Kaessmann. 2008. Loss of egg yolk genes in mammals and the origin of lactation and placentation. PLoS Biol **6**:e63.

Breese, G. R., A. A. Baumeister, T. J. McCown, S. G. Emerick, G. D. Frye, K. Crotty, and R. A. Mueller. 1984. Behavioral differences between neonatal and adult 6-hydroxydopamine-treated rats to dopamine agonists: relevance to neurological symptoms in clinical syndromes with reduced brain dopamine. J Pharmacol Exp Ther **231**:343-354.

Bridges, C. B. 1936. The Bar "Gene" a Duplication. Science **83**:210-211.

Caetano-Anolles, G., H. S. Kim, and J. E. Mittenthal. 2007. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc Natl Acad Sci U S A **104**:9358-9363.

Caldwell, G. T., and H. G. Wells. 1914. The purine enzymes of the opossum *(Didelphus virginiana)*. J Biol Chem **19**:279-283.

Campbell, J. W., J. E. Vorhaben, and D. D. Smith, Jr. 1987. Uricoteley:its nature and origin during the evolution of tetrapod vertebrates. J Exp Zool **243**:349-363.

Casas-Bruge, M., C. Almenar, I. M. Grau, J. Jane, M. Herrera-Marschitz, and U. Ungerstedt. 1985. Dopaminergic receptor supersensitivity in self-mutilatory behaviour of Lesch-Nyhan disease. Lancet **1**:991-992.

Charlesworth, B., and D. Charlesworth. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci **355**:1563-1572.

Chen, G., T. G. Gharib, C. C. Huang, J. M. Taylor, D. E. Misek, S. L. Kardia, T. J. Giordano, M. D. Iannettoni, M. B. Orringer, S. M. Hanash, and D. G. Beer. 2002. Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics **1**:304-313.

Cheung, K. J., I. Tzameli, P. Pissios, I. Rovira, O. Gavrilova, T. Ohtsubo, Z. Chen, T. Finkel, J. S. Flier, and J. M. Friedman. 2007. Xanthine oxidoreductase is a regulator of adipogenesis and PPARgamma activity. Cell Metab **5**:115-128.

Chou, H. H., T. Hayakawa, S. Diaz, M. Krings, E. Indriati, M. Leakey, S. Paabo, Y. Satta, N. Takahata, and A. Varki. 2002. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc Natl Acad Sci U S A **99**:11736-11741.

Consortium, I. H. G. S. 2001. Initial sequencing and analysis of the human genome. Nature **409**:860-921.

Craig, S. P., 3rd, and A. E. Eakin. 2000. Purine phosphoribosyltransferases. J Biol Chem **275**:20231-20234.

Darwin, C. 1859. On the Origins of Species. A. L. Burt Company, 52-58 Duane Street, New York.

De Bodt, S., S. Maere, and Y. Van de Peer. 2005. Genome duplication and the origin of angiosperms. Trends Ecol Evol **20**:591-597.

Deng, Y., Z. Wang, K. Ying, S. Gu, C. Ji, Y. Huang, X. Gu, Y. Wang, Y. Xu, Y. Li, Y. Xie, and Y. Mao. 2002. NADPH-dependent GMP reductase isoenzyme of human (GMPR2). Expression, purification, and kinetic properties. Int J Biochem Cell Biol **34**:1035-1050.

Denton, D. A., M. Reich, and F. J. R. Hird. 1963. Ureotelism of Echidna and Platypus. Science **139**:1.

Dunnett, S. B., D. J. Sirinathsinghji, R. Heavens, D. C. Rogers, and M. R. Kuehn. 1989. Monoamine deficiency in a transgenic (Hprt-) mouse model of Lesch-Nyhan syndrome. Brain Res **501**:401-406.

Engle, S. J., D. E. Womer, P. M. Davies, G. Boivin, A. Sahota, H. A. Simmonds, P. J. Stambrook, and J. A. Tischfield. 1996. HPRT-APRT-deficient mice are not a model for lesch-nyhan syndrome. Hum Mol Genet **5**:1607-1610.

Eugui, E. M., and A. C. Allison. 1993. Immunosuppressive activity of mycophenolate mofetil. Ann N Y Acad Sci **685**:309-329.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8**:175-185.

Finger, S., R. P. Heavens, D. J. Sirinathsinghji, M. R. Kuehn, and S. B. Dunnett. 1988. Behavioral and neurochemical evaluation of a transgenic mouse model of Lesch-Nyhan syndrome. J Neurol Sci **86**:203-213.

Fitzgerald, J., and J. F. Bateman. 2004. Why mice have lost genes for COL21A1, STK17A, GPR145 and AHRI: evidence for gene deletion at evolutionary breakpoints in the rodent lineage. Trends Genet **20**:408-412.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**:1531-1545.

Fujiwara, S., and T. Noguchi. 1995. Degradation of purines: only ureidoglycollate lyase out of four allantoin-degrading enzymes is present in mammals. J Biol Chem:315-318.

Fukagawa, T., N. Hayward, J. Yang, C. Azzalin, D. Griffin, A. F. Stewart, and W. Brown. 1999. The chicken HPRT gene: a counter selectable marker for the DT40 cell line. Nucleic Acids Res **27**:1966-1969.

Futcher, B., G. I. Latter, P. Monardo, C. S. McLaughlin, and J. I. Garrels. 1999. A sampling of the yeast proteome. Mol Cell Biol **19**:7357-7368.

Gibbs, R. A.G. M. WeinstockM. L. MetzkerD. M. MuznyE. J. SodergrenS. SchererG. ScottD. SteffenK. C. WorleyP. E. BurchG. OkwuonuS. HinesL. LewisC. DeRamoO. DelgadoS. Dugan-RochaG. MinerM. MorganA. HawesR. GillCeleraR. A. HoltM. D. AdamsP. G. AmanatidesH. Baden-TillsonM. BarnsteadS. ChinC. A. EvansS. FerrieraC. FoslerA. GlodekZ. GuD. JenningsC. L. KraftT. NguyenC. M. PfannkochC. SitterG. G. SuttonJ. C. VenterT. WoodageD. SmithH. M. LeeE. GustafsonP. CahillA. KanaL. Doucette-StammK. WeinstockK. FechtelR. B. WeissD. M. DunnE. D. GreenR. W. BlakesleyG. G. BouffardP. J. De JongK. OsoegawaB. ZhuM. MarraJ. ScheinI. BosdetC. FjellS. JonesM. KrzywinskiC. MathewsonA. SiddiquiN. WyeJ. McPhersonS. ZhaoC. M. FraserJ. ShettyS. ShatsmanK. GeerY. ChenS. AbramzonW. C. NiermanP. H. HavlakR. ChenK. J. DurbinA. EganY. RenX. Z. SongB. LiY. LiuX. QinS. CawleyA. J. CooneyL. M. D'SouzaK. MartinJ. Q. WuM. L. Gonzalez-GarayA. R. JacksonK. J.

KalafusM. P. McLeodA. MilosavljevicD. VirkA. VolkovD. A. WheelerZ. ZhangJ. A. BaileyE. E. EichlerE. TuzunE. BirneyE. MonginA. Ureta-VidalC. WoodwarkE. ZdobnovP. BorkM. SuyamaD. TorrentsM. AlexanderssonB. J. TraskJ. M. YoungH. HuangH. WangH. XingS. DanielsD. GietzenJ. SchmidtK. StevensU. VittJ. WingroveF. CamaraM. Mar AlbaJ. F. AbrilR. GuigoA. SmitI. DubchakE. M. RubinO. CouronneA. PoliakovN. HubnerD. GantenC. GoeseleO. HummelT. KreitlerY. A. LeeJ. MontiH. SchulzH. ZimdahlH. HimmelbauerH. LehrachH. J. JacobS. BrombergJ. Gullings-HandleyM. I. Jensen-SeamanA. E. KwitekJ. LazarD. PaskoP. J. TonellatoS. TwiggerC. P. PontingJ. M. DuarteS. RiceL. GoodstadtS. A. BeatsonR. D. EmesE. E. WinterC. WebberP. BrandtG. NyakaturaM. AdetobiF. ChiaromonteL. ElnitskiP. EswaraR. C. HardisonM. HouD. KolbeK. MakovaW. MillerA. NekrutenkoC. RiemerS. SchwartzJ. TaylorS. YangY. ZhangK. LindpaintnerT. D. AndrewsM. CaccamoM. ClampL. ClarkeV. CurwenR. DurbinE. EyrasS. M. SearleG. M. CooperS. BatzoglouM. BrudnoA. SidowE. A. StoneB. A. PayseurG. BourqueC. Lopez-OtinX. S. PuenteK. ChakrabartiS. ChatterjiC. DeweyL. PachterN. BrayV. B. YapA. CaspiG. TeslerP. A. PevznerD. HausslerK. M. RoskinR. BaertschH. ClawsonT. S. FureyA. S. HinrichsD. KarolchikW. J. KentK. R. RosenbloomH. TrumbowerM. WeirauchD. N. CooperP. D. StensonB. MaM. BrentM. ArumugamD. ShteynbergR. R. CopleyM. S. TaylorH. RiethmanU. MudunuriJ. PetersonM. GuyerA. FelsenfeldS. OldS. Mockrin, and F. Collins. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature **428**:493-521.

Girirajan, S., L. Chen, T. Graves, T. Marques-Bonet, M. Ventura, C. Fronick, L. Fulton, M. Rocchi, R. S. Fulton, R. K. Wilson, E. R. Mardis, and E. E. Eichler. 2009. Sequencing human-gibbon breakpoints of synteny reveals mosaic new insertions at rearrangement sites. Genome Res.

Gitelman, I. 2007. Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. Mol Biol Evol **24**:1912-1925.

Goldstein, M., S. Kuga, N. Kusano, E. Meller, J. Dancis, and R. Schwarcz. 1986. Dopamine agonist induced self-mutilative biting behavior in monkeys with unilateral ventromedial tegmental lesions of the brainstem: possible pharmacological model for Lesch-Nyhan syndrome. Brain Res **367**:114-120.

Graves, J. A., G. K. Chew, D. W. Cooper, and P. G. Johnston. 1979. Marsupial--mouse cell hybrids containing fragments of the marsupial X chromosome. Somatic Cell Genet **5**:481-489.

Greenbaum, D., C. Colangelo, K. Williams, and M. Gerstein. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol **4**:117.

Greenberg, A. J., J. R. Moran, J. A. Coyne, and C. I. Wu. 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. Science **302**:1754-1757.

Guenet, J. L., and F. Bonhomme. 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet **19**:24-31.

Guo, Y., P. Xiao, S. Lei, F. Deng, G. G. Xiao, Y. Liu, X. Chen, L. Li, S. Wu, Y. Chen, H. Jiang, L. Tan, J. Xie, X. Zhu, S. Liang, and H. Deng. 2008. How is mRNA

expression predictive for protein expression? A correlation study on human circulating monocytes. Acta Biochim Biophys Sin (Shanghai) **40**:426-436.
Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol **19**:1720-1730.
Ha, Y. C., and P. J. Barter. 1982. Differences in plasma cholesteryl ester transfer activity in sixteen vertebrate species. Comp Biochem Physiol B **71**:265-269.
Haldane, J. B. 1955. Origin of man. Nature **176**:169-170.
Haller, O., M. Acklin, and P. Staeheli. 1987. Influenza virus resistance of wild mice: wild-type and mutant Mx alleles occur at comparable frequencies. J Interferon Res **7**:647-656.
Hanke, J. H., J. E. Hambor, and P. Kavathas. 1995. Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. J Mol Biol **246**:63-73.
Harper, K. N., and G. J. Armelagos. 2005. Genomics at the origins of Agriculture, Part One. Evolutionary Anthropology **14**:68-77.
Hayashi, S., S. Fujiwara, and T. Noguchi. 2000a. Evolution of urate-degrading enzymes in animal peroxisomes. Cell Biochem Biophys **32**:123-129.
Hayashi, S., S. Fujiwara, and T. Noguchi. 2000b. Evolution of urate-degrading enzymes in animal peroxisomes. Cell Biochem Biophys **32 Spring**:123-129.
He, X., and J. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics **169**:1157-1164.
Henderson, J. F. 1968. Possible functions of hypoxanthine-guanine phosphoribosyltransferase and their relation to the biochemical pathology of the Lesch-Nyhan syndrome. Fed Proc **27**:1075-1077.
Hillier, L. W.W. MillerE. BirneyW. WarrenR. C. HardisonC. P. PontingP. BorkD. W. BurtM. A. GroenenM. E. DelanyJ. B. DodgsonA. T. ChinwallaP. F. CliftenS. W. CliftonK. D. DelehauntyC. FronickR. S. FultonT. A. GravesC. KremitzkiD. LaymanV. MagriniJ. D. McPhersonT. L. MinerP. MinxW. E. NashM. N. NhanJ. O. NelsonL. G. OddyC. S. PohlJ. Randall-MaherS. M. SmithJ. W. WallisS. P. YangM. N. RomanovC. M. RondelliB. PatonJ. SmithD. MorriceL. DanielsH. G. TempestL. RobertsonJ. S. MasabandaD. K. GriffinA. VignalV. FillonL. JacobbsonS. KerjeL. AnderssonR. P. CrooijmansJ. AertsJ. J. van der PoelH. EllegrenR. B. CaldwellS. J. HubbardD. V. GrafhamA. M. KierzekS. R. McLarenI. M. OvertonH. ArakawaK. J. BeattieY. BezzubovP. E. BoardmanJ. K. BonfieldM. D. CroningR. M. DaviesM. D. FrancisS. J. HumphrayC. E. ScottR. G. TaylorC. TickleW. R. BrownJ. RogersJ. M. BuersteddeS. A. WilsonL. StubbsI. OvcharenkoL. GordonS. LucasM. M. MillerH. InokoT. ShiinaJ. KaufmanJ. SalomonsenK. SkjoedtG. K. WongJ. WangB. LiuJ. YuH. YangM. NefedovM. KoriabineP. J. DejongL. GoodstadtC. WebberN. J. DickensI. LetunicM. SuyamaD. TorrentsC. von MeringE. M. ZdobnovK. MakovaA. NekrutenkoL. ElnitskiP. EswaraD. C. KingS. YangS. TyekuchevaA. RadakrishnanR. S. HarrisF. ChiaromonteJ. TaylorJ. HeM. RijnkelsS. Griffiths-JonesA. Ureta-VidalM. M. HoffmanJ. SeverinS. M. SearleA. S. LawD. SpeedD. WaddingtonZ. ChengE. TuzunE. EichlerZ. BaoP. FlicekD. D. ShteynbergM. R. BrentJ. M. ByeE. J. HuckleS. ChatterjiC. DeweyL. PachterA. KouranovZ. MourelatosA. G.

HatzigeorgiouA. H. PatersonR. IvarieM. BrandstromE. AxelssonN. BackstromS. BerlinM. T. WebsterO. PourquieA. ReymondC. UclaS. E. AntonarakisM. LongJ. J. EmersonE. BetranI. DupanloupH. KaessmannA. S. HinrichsG. BejeranoT. S. FureyR. A. HarteB. RaneyA. SiepelW. J. KentD. HausslerE. EyrasR. CasteloJ. F. AbrilS. CastellanoF. CamaraG. ParraR. GuigoG. BourqueG. TeslerP. A. PevznerA. SmitL. A. FultonE. R. Mardis, and R. K. Wilson. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature **432**:695-716.

Hooper, M., K. Hardy, A. Handyside, S. Hunter, and M. Monk. 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature **326**:292-295.

Horowitz, N. H. 1945. On the Evolution of Biochemical Syntheses. Proc Natl Acad Sci U S A **31**:153-157.

Hubbard, T. J., B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. 2007. Ensembl 2007. Nucleic Acids Res **35**:D610-617.

Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci **256**:119-124.

Hunter, A., and M. H. Givens. 1914. Studies in the comparative biochemsitry of purine metabolism. II. The excretion of purine catabolites in the urine of ungulates. J Biol Chem:403-416.

Hunter, A., M. H. Givens, and C. M. Guion. 1914. Studies in the biochemistry of purine metabolism. I. The excretion of purine catabolites in the urine of marsupials, rodents and carnivora. J Biol Chem:387-401.

Ichida, K., Y. Amaya, N. Kamatani, T. Nishino, T. Hosoya, and O. Sakai. 1997. Identification of two mutations in human xanthine dehydrogenase gene responsible for classical type I xanthinuria. J Clin Invest **99**:2391-2397.

Ito, M., M. Nakamura, H. Ogawa, and Y. Takagi. 1989. Sequence analysis of rat liver uricase-cDNA and the possible presence of the homologous cDNA sequences in chicken embryo. Adv Exp Med Biol **253A**:507-510.

Jaillon, O., J. M. Aury, F. Brunet, J. L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biemont, Z. Skalli, L. Cattolico, J. Poulain, V. De Berardinis, C. Cruaud, S. Duprat, P. Brottier, J. P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K. J. McKernan, P. McEwan, S. Bosak, M. Kellis, J. N. Volff, R. Guigo, M. C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quetier, W. Saurin, C.

Scarpelli, P. Wincker, E. S. Lander, J. Weissenbach, and H. Roest Crollius. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature **431**:946-957.

Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998a. Multiple sequence alignment with Clustal X. Trends Biochem Sci **23**:403-405.

Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998b. Multiple sequence alignment with Clustal X. Trends Biochem.Sci. **23**:403-405.

Jinnah, H. A., and T. Friedann. 2001. Lesch-Nyhan Disease and Its Variants in The Metabolic & Molecular Bases of Inherited Disease. McGraw-Hill: New York.

Jinnah, H. A., and T. Friedmann. 2001a. Lesch-Nyhan Disease and its variants, in The Metabolic & Molecular Basis of Inherited Disease. McGraw Hill, New York.

Jinnah, H. A., and T. Friedmann. 2001b. Lesch-Nyhan Disease and Its Variants. Pp. 2537-2570 *in* C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds. The Metabolic & Molecular Bases of Inherited Disease. McGraw-Hill, New York.

Jinnah, H. A., F. H. Gage, and T. Friedmann. 1991. Amphetamine-induced behavioral phenotype in a hypoxanthine-guanine phosphoribosyltransferase-deficient mouse model of Lesch-Nyhan syndrome. Behav Neurosci **105**:1004-1012.

Jinnah, H. A., M. D. Jones, B. E. Wojcik, J. D. Rothstein, E. J. Hess, T. Friedmann, and G. R. Breese. 1999. Influence of age and strain on striatal dopamine loss in a genetic mouse model of Lesch-Nyhan disease. J Neurochem **72**:225-229.

Jinnah, H. A., P. J. Langlais, and T. Friedmann. 1992. Functional analysis of brain dopamine systems in a genetic mouse model of Lesch-Nyhan syndrome. J Pharmacol Exp Ther **263**:596-607.

Jinnah, H. A., B. E. Wojcik, M. Hunt, N. Narang, K. Y. Lee, M. Goldstein, J. K. Wamsley, P. J. Langlais, and T. Friedmann. 1994. Dopamine deficiency in a genetic mouse model of Lesch-Nyhan disease. J Neurosci **14**:1164-1175.

Johnson, R. J., E. A. Gaucher, Y. Y. Sautin, G. N. Henderson, A. J. Angerhofer, and S. A. Benner. 2008. The planetary biology of ascorbate and uric acid and their relationship with the epidemic of obesity and cardiovascular disease. Med Hypotheses **71**:22-31.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8**:275-282.

Keebaugh, A. C., R. T. Sullivan, and J. W. Thomas. 2007. Gene duplication and inactivation in the HPRT gene family. Genomics **89**:134-142.

Keebaugh, A. C., and J. W. Thomas. 2009. Evolution of the purine catabolic pathway: an expanded view of the evolutionary significance of gene loss. unpublished data.

Keilin, J. 1959. The biological significance of uric acid and guanine excretion. Biol. Rev. **34**:265-296.

Kellner, W. A., R. T. Sullivan, B. H. Carlson, and J. W. Thomas. 2005. Uprobe: A genome-wide universal probe resource for comparative physical mapping in vertebrates. Genome Res **15**:166-173.

Kmoch, S., H. Hartmannova, B. Stiburkova, J. Krijt, M. Zikanova, and I. Sebesta. 2000. Human adenylosuccinate lyase (ADSL), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for ADSL deficiency in six patients. Hum Mol Genet **9**:1501-1513.

Kuehn, M. R., A. Bradley, E. J. Robertson, and M. J. Evans. 1987a. A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature **326**:295-298.

Kuehn, M. R., A. Bradley, E. J. Robertson, and M. J. Evans. 1987b. A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature **326**:295-298.

Kumar, S., and S. B. Hedges. 1998a. A molecular timescale for vertebrate evolution. Nature **392**:917-920.

Kumar, S., and S. B. Hedges. 1998b. A molecular timescale for vertebrate evolution. Nature **392**:917-920.

Lander, E. S.L. M. LintonB. BirrenC. NusbaumM. C. ZodyJ. BaldwinK. DevonK. DewarM. DoyleW. FitzHughR. FunkeD. GageK. HarrisA. HeafordJ. HowlandL. KannJ. LehoczkyR. LeVineP. McEwanK. McKernanJ. MeldrimJ. P. MesirovC. MirandaW. MorrisJ. NaylorC. RaymondM. RosettiR. SantosA. SheridanC. SougnezN. Stange-ThomannN. StojanovicA. SubramanianD. WymanJ. RogersJ. SulstonR. AinscoughS. BeckD. BentleyJ. BurtonC. CleeN. CarterA. CoulsonR. DeadmanP. DeloukasA. DunhamI. DunhamR. DurbinL. FrenchD. GrafhamS. GregoryT. HubbardS. HumphrayA. HuntM. JonesC. LloydA. McMurrayL. MatthewsS. MercerS. MilneJ. C. MullikinA. MungallR. PlumbM. RossR. ShownkeenS. SimsR. H. WaterstonR. K. WilsonL. W. HillierJ. D. McPhersonM. A. MarraE. R. MardisL. A. FultonA. T. ChinwallaK. H. PepinW. R. GishS. L. ChissoeM. C. WendlK. D. DelehauntyT. L. MinerA. DelehauntyJ. B. KramerL. L. CookR. S. FultonD. L. JohnsonP. J. MinxS. W. CliftonT. HawkinsE. BranscombP. PredkiP. RichardsonS. WenningT. SlezakN. DoggettJ. F. ChengA. OlsenS. LucasC. ElkinE. UberbacherM. FrazierR. A. GibbsD. M. MuznyS. E. SchererJ. B. BouckE. J. SodergrenK. C. WorleyC. M. RivesJ. H. GorrellM. L. MetzkerS. L. NaylorR. S. KucherlapatiD. L. NelsonG. M. WeinstockY. SakakiA. FujiyamaM. HattoriT. YadaA. ToyodaT. ItohC. KawagoeH. WatanabeY. TotokiT. TaylorJ. WeissenbachR. HeiligW. SaurinF. ArtiguenaveP. BrottierT. BrulsE. PelletierC. RobertP. WinckerD. R. SmithL. Doucette-StammM. RubenfieldK. WeinstockH. M. LeeJ. DuboisA. RosenthalM. PlatzerG. NyakaturaS. TaudienA. RumpH. YangJ. YuJ. WangG. HuangJ. GuL. HoodL. RowenA. MadanS. QinR. W. DavisN. A. FederspielA. P. AbolaM. J. ProctorR. M. MyersJ. SchmutzM. DicksonJ. GrimwoodD. R. CoxM. V. OlsonR. KaulC. RaymondN. ShimizuK. KawasakiS. MinoshimaG. A. EvansM. AthanasiouR. SchultzB. A. RoeF. ChenH. PanJ. RamserH. LehrachR. ReinhardtW. R. McCombieM. de la BastideN. DedhiaH. BlockerK. HornischerG. NordsiekR. AgarwalaL. AravindJ. A. BaileyA. BatemanS. BatzoglouE. BirneyP. BorkD. G. BrownC. B. BurgeL. CeruttiH. C. ChenD. ChurchM. ClampR. R. CopleyT. DoerksS. R. EddyE. E. EichlerT. S. FureyJ. GalaganJ. G. GilbertC. HarmonY. HayashizakiD. HausslerH. HermjakobK. HokampW. JangL. S. JohnsonT. A. JonesS. KasifA. KaspryzkS. KennedyW. J. KentP. KittsE. V. KooninI. KorfD. KulpD. LancetT. M. LoweA. McLysaghtT. MikkelsenJ. V. MoranN. MulderV. J. PollaraC. P. PontingG. SchulerJ. SchultzG. SlaterA. F. SmitE. StupkaJ. SzustakowskiD. Thierry-MiegJ. Thierry-MiegL. WagnerJ. WallisR. WheelerA. WilliamsY. I. WolfK. H. WolfeS. P. YangR. F. YehF. CollinsM. S. GuyerJ.

PetersonA. FelsenfeldK. A. WetterstrandA. PatrinosM. J. MorganP. de JongJ. J. CataneseK. OsoegawaH. ShizuyaS. Choi, and Y. J. Chen. 2001a. Initial sequencing and analysis of the human genome. Nature **409**:860-921.

Lander, E. S.L. M. LintonB. BirrenC. NusbaumM. C. ZodyJ. BaldwinK. DevonK. DewarM. DoyleW. FitzHughR. FunkeD. GageK. HarrisA. HeafordJ. HowlandL. KannJ. LehoczkyR. LeVineP. McEwanK. McKernanJ. MeldrimJ. P. MesirovC. MirandaW. MorrisJ. NaylorC. RaymondM. RosettiR. SantosA. SheridanC. SougnezN. Stange-ThomannN. StojanovicA. SubramanianD. WymanJ. RogersJ. SulstonR. AinscoughS. BeckD. BentleyJ. BurtonC. CleeN. CarterA. CoulsonR. DeadmanP. DeloukasA. DunhamI. DunhamR. DurbinL. FrenchD. GrafhamS. GregoryT. HubbardS. HumphrayA. HuntM. JonesC. LloydA. McMurrayL. MatthewsS. MercerS. MilneJ. C. MullikinA. MungallR. PlumbM. RossR. ShownkeenS. SimsR. H. WaterstonR. K. WilsonL. W. HillierJ. D. McPhersonM. A. MarraE. R. MardisL. A. FultonA. T. ChinwallaK. H. PepinW. R. GishS. L. ChissoeM. C. WendlK. D. DelehauntyT. L. MinerA. DelehauntyJ. B. KramerL. L. CookR. S. FultonD. L. JohnsonP. J. MinxS. W. CliftonT. HawkinsE. BranscombP. PredkiP. RichardsonS. WenningT. SlezakN. DoggettJ. F. ChengA. OlsenS. LucasC. ElkinE. UberbacherM. FrazierR. A. GibbsD. M. MuznyS. E. SchererJ. B. BouckE. J. SodergrenK. C. WorleyC. M. RivesJ. H. GorrellM. L. MetzkerS. L. NaylorR. S. KucherlapatiD. L. NelsonG. M. WeinstockY. SakakiA. FujiyamaM. HattoriT. YadaA. ToyodaT. ItohC. KawagoeH. WatanabeY. TotokiT. TaylorJ. WeissenbachR. HeiligW. SaurinF. ArtiguenaveP. BrottierT. BrulsE. PelletierC. RobertP. WinckerD. R. SmithL. Doucette-StammM. RubenfieldK. WeinstockH. M. LeeJ. DuboisA. RosenthalM. PlatzerG. NyakaturaS. TaudienA. RumpH. YangJ. YuJ. WangG. HuangJ. GuL. HoodL. RowenA. MadanS. QinR. W. DavisN. A. FederspielA. P. AbolaM. J. ProctorR. M. MyersJ. SchmutzM. DicksonJ. GrimwoodD. R. CoxM. V. OlsonR. KaulN. ShimizuK. KawasakiS. MinoshimaG. A. EvansM. AthanasiouR. SchultzB. A. RoeF. ChenH. PanJ. RamserH. LehrachR. ReinhardtW. R. McCombieM. de la BastideN. DedhiaH. BlockerK. HornischerG. NordsiekR. AgarwalaL. AravindJ. A. BaileyA. BatemanS. BatzoglouE. BirneyP. BorkD. G. BrownC. B. BurgeL. CeruttiH. C. ChenD. ChurchM. ClampR. R. CopleyT. DoerksS. R. EddyE. E. EichlerT. S. FureyJ. GalaganJ. G. GilbertC. HarmonY. HayashizakiD. HausslerH. HermjakobK. HokampW. JangL. S. JohnsonT. A. JonesS. KasifA. KaspryzkS. KennedyW. J. KentP. KittsE. V. KooninI. KorfD. KulpD. LancetT. M. LoweA. McLysaghtT. MikkelsenJ. V. MoranN. MulderV. J. PollaraC. P. PontingG. SchulerJ. SchultzG. SlaterA. F. SmitE. StupkaJ. SzustakowskiD. Thierry-MiegJ. Thierry-MiegL. WagnerJ. WallisR. WheelerA. WilliamsY. I. WolfK. H. WolfeS. P. YangR. F. YehF. CollinsM. S. GuyerJ. PetersonA. FelsenfeldK. A. WetterstrandA. PatrinosM. J. MorganJ. SzustakowkiP. de JongJ. J. CataneseK. OsoegawaH. ShizuyaS. Choi, and Y. J. Chen. 2001b. Initial sequencing and analysis of the human genome. Nature **409**:860-921.

Li, H., A. Coghlan, J. Ruan, L. J. Coin, J. K. Heriche, L. Osmotherly, R. Li, T. Liu, Z. Zhang, L. Bolund, G. K. Wong, W. Zheng, P. Dehal, J. Wang, and R. Durbin. 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res **34**:D572-580.

Li, X., W. Li, H. Wang, J. Cao, K. Maehashi, L. Huang, A. A. Bachmanov, D. R. Reed, V. Legrand-Defretin, G. K. Beauchamp, and J. G. Brand. 2005. Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. PLoS Genet **1**:27-35.

Lichtinghagen, R., P. B. Musholt, M. Lein, A. Romer, B. Rudolph, G. Kristiansen, S. Hauptmann, D. Schnorr, S. A. Loening, and K. Jung. 2002. Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue. Eur Urol **42**:398-406.

Lie, J., R. de Crom, M. Jauhiainen, T. van Gent, R. van Haperen, L. Scheek, H. Jansen, C. Ehnholm, and A. van Tol. 2001. Evaluation of phospholipid transfer protein and cholesteryl ester transfer protein as contributors to the generation of pre beta-high-density lipoproteins. Biochem J **360**:379-385.

Lindblad-Toh, K.C. M. WadeT. S. MikkelsenE. K. KarlssonD. B. JaffeM. KamalM. ClampJ. L. ChangE. J. Kulbokas, 3rdM. C. ZodyE. MauceliX. XieM. BreenR. K. WayneE. A. OstranderC. P. PontingF. GalibertD. R. SmithP. J. DeJongE. KirknessP. AlvarezT. BiagiW. BrockmanJ. ButlerC. W. ChinA. CookJ. CuffM. J. DalyD. DeCaprioS. GnerreM. GrabherrM. KellisM. KleberC. BardelebenL. GoodstadtA. HegerC. HitteL. KimK. P. KoepfliH. G. ParkerJ. P. PollingerS. M. SearleN. B. SutterR. ThomasC. WebberJ. BaldwinA. AbebeA. AbouelleilL. AftuckM. Ait-ZahraT. AldredgeN. AllenP. AnS. AndersonC. AntoineH. ArachchiA. AslamL. AyotteP. BachantsangA. BarryT. BayulM. BenamaraA. BerlinD. BessetteB. BlitshteynT. BloomJ. BlyeL. BoguslavskiyC. BonnetB. BoukhgalterA. BrownP. CahillN. CalixteJ. CamarataY. CheshatsangJ. ChuM. CitroenA. CollymoreP. CookeT. DawoeR. DazaK. DecktorS. DeGrayN. DhargayK. DooleyP. DorjeK. DorjeeL. DorrisN. DuffeyA. DupesO. EgbiremolenR. ElongJ. FalkA. FarinaS. FaroD. FergusonP. FerreiraS. FisherM. FitzGeraldK. FoleyC. FoleyA. FrankeD. FriedrichD. GageM. GarberG. GearinG. GiannoukosT. GoodeA. GoyetteJ. GrahamE. GrandboisK. GyaltsenN. HafezD. HagopianB. HagosJ. HallC. HealyR. HegartyT. HonanA. HornN. HoudeL. HughesL. HunnicuttM. HusbyB. JesterC. JonesA. KamatB. KangaC. KellsD. KhazanovichA. C. KieuP. KisnerM. KumarK. LanceT. LandersM. LaraW. LeeJ. P. LegerN. LennonL. LeuperS. LeVineJ. LiuX. LiuY. LokyitsangT. LokyitsangA. LuiJ. MacdonaldJ. MajorR. MarabellaK. MaruC. MatthewsS. McDonoughT. MehtaJ. MeldrimA. MelnikovL. MeneusA. MihalevT. MihovaK. MillerR. MittelmanV. MlengaL. MulrainG. MunsonA. NavidiJ. NaylorT. NguyenN. NguyenC. NguyenR. NicolN. NorbuC. NorbuN. NovodT. NyimaP. OlandtB. O'NeillK. O'NeillS. OsmanL. OyonoC. PattiD. PerrinP. PhunkhangF. PierreM. PriestA. RachupkaS. RaghuramanR. RameauV. RayC. RaymondF. RegeC. RiseJ. RogersP. RogovJ. SahalieS. SettipalliT. SharpeT. SheaM. SheehanN. SherpaJ. ShiD. ShihJ. SloanC. SmithT. SparrowJ. StalkerN. Stange-ThomannS. StavropoulosC. StoneS. StoneS. SykesP. TchuingaP. TenzingS. TesfayeD. ThoulutsangY. ThoulutsangK. TophamI. ToppingT. TsamlaH. VassilievV. VenkataramanA. VoT. WangchukT. WangdiM. WeiandJ. WilkinsonA. WilsonS. YadavS. YangX. YangG. YoungQ. YuJ. ZainounL. ZembekA. Zimmer, and E. S.

Lander. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438**:803-819.

Lloyd, K. G., O. Hornykiewicz, L. Davidson, K. Shannak, I. Farley, M. Goldstein, M. Shibuya, W. N. Kelley, and I. H. Fox. 1981. Biochemical evidence of dysfunction of brain neurotransmitters in the Lesch-Nyhan syndrome. N Engl J Med **305**:1106-1111.

Locascio, A., M. Manzanares, M. J. Blanco, and M. A. Nieto. 2002. Modularity and reshuffling of Snail and Slug expression during vertebrate evolution. Proc Natl Acad Sci U S A **99**:16841-16846.

Loh, E., T. R. Rebbeck, P. D. Mahoney, D. DeNofrio, J. L. Swain, and E. W. Holmes. 1999. Common variant in AMPD1 gene predicts improved clinical outcome in patients with heart failure. Circulation **99**:1422-1425.

Lynch, M., and J. S. Conery. 2003. The evolutionary demography of duplicate genes. J Struct Funct Genomics **3**:35-44.

MacArthur, D. G., J. T. Seto, J. M. Raftery, K. G. Quinlan, G. A. Huttley, J. W. Hook, F. A. Lemckert, A. J. Kee, M. R. Edwards, Y. Berman, E. C. Hardeman, P. W. Gunning, S. Easteal, N. Yang, and K. N. North. 2007. Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. Nat Genet **39**:1261-1265.

Maier, S. A., J. R. Galellis, and H. E. McDermid. 2005. Phylogenetic analysis reveals a novel protein family closely related to adenosine deaminase. J Mol Evol **61**:776-794.

Marchler-Bauer, A., and S. H. Bryant. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res **32**:W327-331.

Marks, J. F., J. Baum, D. K. Keele, J. L. Kay, and A. MacFarlen. 1968. Lesch-Nyhan syndrome treated from the early neonatal period. Pediatrics **42**:357-359.

McEwan, C., and D. W. Melton. 2003. A simple genotyping assay for the Hprt null allele in mice produced from the HM-1 and E14TG2a mouse embryonic stem cell lines. Transgenic Res **12**:519-520.

Menashe, I., O. Man, D. Lancet, and Y. Gilad. 2002. Population differences in haplotype structure within a human olfactory receptor gene cluster. Hum Mol Genet **11**:1381-1390.

Minana, M. D., and S. Grisolia. 1986. Caffeine ingestion by rats increases noradrenaline turnover and results in self-biting. J Neurochem **47**:728-732.

Mommsen, T. P., and P. J. Walsh. 1989. Evolution of urea synthesis in vertebrates: the piscine connection. Science **243**:72-75.

Morisaki, T., M. Gross, H. Morisaki, D. Pongratz, N. Zollner, and E. W. Holmes. 1992. Molecular basis of AMP deaminase deficiency in skeletal muscle. Proc Natl Acad Sci U S A **89**:6457-6461.

Morton, C. C., R. L. Eddy, T. B. Shows, P. R. H. Clark, R. L. Sabina, and E. W. Holmes. 1989. Human AMP deaminase-1 gene (AMPD1) is mapped to chromosome 1. Cytogenet. Cell cenet **51**:1048-1049.

Moses, A. M., D. Y. Chiang, M. Kellis, E. S. Lander, and M. B. Eisen. 2003. Position specific variation in the rate of evolution in transcription factor binding sites. BMC Evol Biol **3**:19.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science **294**:2348-2351.

Nehlig, A., J. L. Daval, and G. Debry. 1992. Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. Brain Res Brain Res Rev **17**:139-170.

Nguyen, V. H., B. Schmid, J. Trout, S. A. Connors, M. Ekker, and M. C. Mullins. 1998. Ventral and lateral regions of the zebrafish gastrula, including the neural crest progenitors, are established by a bmp2b/swirl pathway of genes. Dev Biol **199**:93-110.

Nicklas, J. A. 2006. Pseudogenes of the human HPRT1 gene. Environ Mol Mutagen **47**:212-218.

Nicklas, J. A. 2005. Pseudogenes of the human HPRT1 gene. Environ Mol Mutagen **47**:212-218.

Novembre, J., A. P. Galvani, and M. Slatkin. 2005. The geographic spread of the CCR5 Delta32 HIV-resistance allele. PLoS Biol **3**:e339.

Noyce, L., J. Conaty, and A. A. Piper. 1997. Identification of a novel tissue-specific processed HPRT gene and comparison with X-linked gene transcription in the Australian marsupial Macropus robustus. Gene **186**:87-95.

Oda, M., Y. Satta, O. Takenaka, and N. Takahata. 2002. Loss of urate oxidase activity in hominoids and its evolutionary implications. Mol Biol Evol **19**:640-653.

Ogasawara, N., H. Goto, Y. Yamada, I. Nishigaki, T. Itoh, I. Hasegawa, and K. S. Park. 1987. Deficiency of AMP deaminase in erythrocytes. Hum Genet **75**:15-18.

Ohno, S. 1970. Evolution by Gene Duplication. Springer-Verlag, Berlin.

Olson, M. V. 1999. When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet **64**:18-23.

Ordonez, G. R., L. W. Hillier, W. C. Warren, F. Grutzner, C. Lopez-Otin, and X. S. Puente. 2008. Loss of genes implicated in gastric function during platypus evolution. Genome Biol **9**:R81.

Orowan, E. 1955. The origin of man. Nature **175**:683-684.

Packard, G. C. 1966. The Influence of Ambient Temperature and Aridity on Modes of Reproduction and Excretion of Amniote Vertebrates. The American Naturalist **100**:667-682.

Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet **35**:32-40.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, and A. C. Stone. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet **39**:1256-1260.

Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J. Tabin. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet **38**:107-111.

Proulx, S. R., and P. C. Phillips. 2006. Allelic divergence precedes and promotes gene duplication. Evolution Int J Org Evolution **60**:881-892.

Przylecki, S. J., and L. Rogalski. 1927. La Loi Biogenetique et les fonctions des organismes vivants. Physiologie **28**:423-430.

Raaum, R. L., K. N. Sterner, C. M. Noviello, C. B. Stewart, and T. R. Disotell. 2005. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. J Hum Evol **48**:237-257.

Ramazzina, I., C. Folli, A. Secchi, R. Berni, and R. Percudani. 2006a. Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. Nat Chem Biol **2**:144-148.

Ramazzina, I., C. Folli, A. Secchi, R. Berni, and R. Percudani. 2006b. Completing the uric acid degradation pathway throught phylogenetic comparison of whole genomes. Nat Chem Biol **2**:144-148.

Redi, C. A., H. Zacharias, S. Merani, M. Oliveira-Miranda, M. Aguilera, M. Zuccotti, S. Garagna, and E. Capanna. 2005. Genome sizes in afrotheria, xenarthra, euarchontoglires, and laurasiatheria. J Hered **96**:485-493.

Remy, C., D. A. Richert, and W. W. Westerfeld. 1951. The determination of xanthine dehydrogenase in chicken tissues. J Biol Chem **193**:649-657.

Rompler, H., A. Schulz, C. Pitra, G. Coop, M. Przeworski, S. Paabo, and T. Schoneberg. 2005. The rise and fall of the chemoattractant receptor GPR33. J Biol Chem **280**:31068-31075.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572-1574.

Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal. 2005. Towards a proteome-scale map of the human protein-protein interaction network. Nature **437**:1173-1178.

Sabina, R. L., J. L. Swain, C. W. Olanow, W. G. Bradley, W. N. Fishbein, S. DiMauro, and E. W. Holmes. 1984. Myoadenylate deaminase deficiency. Functional and metabolic abnormalities associated with disruption of the purine nucleotide cycle. J Clin Invest **73**:720-730.

Saunders, N. J. 1994. Why monitor peak vancomycin concentrations? Lancet **344**:1748-1750.

Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature **440**:341-345.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502-504.

Scholpp, S., and M. Brand. 2001. Morpholino-induced knockdown of zebrafish engrailed genes eng2 and eng3 reveals redundant and unique functions in midbrain--hindbrain boundary development. Genesis **30**:129-133.

Schuler, G. D. 1997. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J.Mol.Med. **75**:694-698.

Schwartz, S., W. J. Kent, A. Smit, Z. Zhang, R. Baertsch, R. C. Hardison, D. Haussler, and W. Miller. 2003. Human-mouse alignments with BLASTZ. Genome Res **13**:103-107.

Seixas, S., G. Suriano, F. Carvalho, R. Seruca, J. Rocha, and A. Di Rienzo. 2007. Sequence diversity at the proximal 14q32.1 SERPIN subcluster: evidence for natural selection favoring the pseudogenization of SERPINA2. Mol Biol Evol **24**:587-598.

Sharp, A. J., S. Hansen, R. R. Selzer, Z. Cheng, R. Regan, J. A. Hurst, H. Stewart, S. M. Price, E. Blair, R. C. Hennekam, C. A. Fitzpatrick, R. Segraves, T. A. Richmond, C. Guiver, D. G. Albertson, D. Pinkel, P. S. Eis, S. Schwartz, S. J. Knight, and E. E. Eichler. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet **38**:1038-1042.

Sollinger, H. W., F. O. Belzer, M. H. Deierhoi, A. G. Diethelm, T. A. Gonwa, R. S. Kauffman, G. B. Klintmalm, S. V. McDiarmid, J. Roberts, J. T. Rosenthal, and et al. 1992. RS-61443 (mycophenolate mofetil). A multicenter study for refractory kidney transplant rejection. Ann Surg **216**:513-518; discussion 518-519.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc.Natl.Acad.Sci.U.S.A **100**:1056-1061.

Stedman, H. H., B. W. Kozyak, A. Nelson, D. M. Thesier, L. T. Su, D. W. Low, C. R. Bridges, J. B. Shrager, N. Minugh-Purvis, and M. A. Mitchell. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature **428**:415-418.

Stefansson, H., A. Helgason, G. Thorleifsson, V. Steinthorsdottir, G. Masson, J. Barnard, A. Baker, A. Jonasdottir, A. Ingason, V. G. Gudnadottir, N. Desnica, A. Hicks, A. Gylfason, D. F. Gudbjartsson, G. M. Jonsdottir, J. Sainz, K. Agnarsson, B. Birgisdottir, S. Ghosh, A. Olafsdottir, J. B. Cazier, K. Kristjansson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, A. Kong, and K. Stefansson. 2005. A common inversion under selection in Europeans. Nat Genet **37**:129-137.

Takada, Y., and T. Noguchi. 1986. Ureidoglycollate lyase, a new metalloenzyme of peroxisomal urate degradation in marine fish liver. J Biol Chem **235**:391-397.

Tanaka, T., Y. Tateno, and T. Gojobori. 2005. Evolution of vitamin B6 (pyridoxine) metabolism by gain and loss of genes. Mol Biol Evol **22**:243-250.

Thomas, J. W., J. W. Touchman, R. W. Blakesley, G. G. Bouffard, S. M. Beckstrom-Sternberg, E. H. Margulies, M. Blanchette, A. C. Siepel, P. J. Thomas, J. C.

McDowell, B. Maskeri, N. F. Hansen, M. S. Schwartz, R. J. Weber, W. J. Kent, D. Karolchik, T. C. Bruen, R. Bevan, D. J. Cutler, S. Schwartz, L. Elnitski, J. R. Idol, A. B. Prasad, S. Q. Lee-Lin, V. V. Maduro, T. J. Summers, M. E. Portnoy, N. L. Dietrich, N. Akhter, K. Ayele, B. Benjamin, K. Cariaga, C. P. Brinkley, S. Y. Brooks, S. Granite, X. Guan, J. Gupta, P. Haghighi, S. L. Ho, M. C. Huang, E. Karlins, P. L. Laric, R. Legaspi, M. J. Lim, Q. L. Maduro, C. A. Masiello, S. D. Mastrian, J. C. McCloskey, R. Pearson, S. Stantripop, E. E. Tiongson, J. T. Tran, C. Tsurgeon, J. L. Vogt, M. A. Walker, K. D. Wetherby, L. S. Wiggins, A. C. Young, L. H. Zhang, K. Osoegawa, B. Zhu, B. Zhao, C. L. Shu, P. J. de Jong, C. E. Lawrence, A. F. Smit, A. Chakravarti, D. Haussler, P. Green, W. Miller, and E. D. Green. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. Nature **424**:788-793.

Urich, K. 1994a. Comparative Animal Biochemistry. Springer-Verlag, Berlin.

Urich, K. 1994b. Comparative Animal Biochemistry. Springer-Verlag, Berlin.

Van de Peer, Y. 2004. Computational approaches to unveiling ancient genome duplications. Nat Rev Genet **5**:752-763.

van der Knaap, J. A., B. R. Kumar, Y. M. Moshkin, K. Langenberg, J. Krijgsveld, A. J. Heck, F. Karch, and C. P. Verrijzer. 2005. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Mol Cell **17**:695-707.

Vandepoele, K., W. De Vos, J. S. Taylor, A. Meyer, and Y. Van de Peer. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A **101**:1638-1643.

Varela-Echavarria, A., R. Montes de Oca-Luna, and H. A. Barrera-Saldana. 1988. Uricase protein sequences: conserved during vertebrate evolution but absent in humans. Faseb J **2**:3092-3096.

Varki, A. 2001. Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol **Suppl 33**:54-69.

Vigetti, D., G. Binelli, C. Monetti, M. Prati, G. Bernardini, and R. Gornati. 2003. Selective pressure on the allantoicase gene during vertebrate evolution. J Mol Evol **57**:650-658.

Vigetti, D., C. Monetti, F. Acquati, R. Taramelli, and G. Bernardini. 2000. Human allantoicase gene: cDNA cloning, genomic organization and chromosome localization. Gene **256**:253-260.

Vigetti, D., C. Monetti, and G. Bernardini. 2001. Molecular cloning of mouse allantoicase cDNA. Biochim Biophys Acta **1519**:117-121.

Vigetti, D., L. Pollegioni, C. Monetti, M. Prati, G. Bernardini, and R. Gornati. 2002. Property comparison of recombinant amphibian and mammalian allantoicases. FEBS Lett **512**:323-328.

Vinogradov, A. E. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? Trends Genet **20**:248-253.

Vinogradov, A. E. 1999. Intron-genome size relationship on a large evolutionary scale. J Mol Evol **49**:376-384.

Vorbach, C., A. Scriven, and M. R. Capecchi. 2002. The housekeeping gene xanthine oxidoreductase is necessary for milk fat droplet enveloping and secretion: gene sharing in the lactating mammary gland. Genes Dev **16**:3223-3235.

Wang, X., W. E. Grus, and J. Zhang. 2006. Gene losses during human origins. PLoS Biol **4**:e52.

Warburton, P. E., J. Giordano, F. Cheung, Y. Gelfand, and G. Benson. 2004. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res **14**:1861-1869.

Watanabe, S., D. H. Kang, L. Feng, T. Nakagawa, J. Kanellis, H. Lan, M. Mazzali, and R. J. Johnson. 2002. Uric acid, hominoid evolution, and the pathogenesis of salt-sensitivity. Hypertension **40**:355-360.

Waterston, R. H.K. Lindblad-TohE. BirneyJ. RogersJ. F. AbrilP. AgarwalR. AgarwalaR. AinscoughM. AlexanderssonP. AnS. E. AntonarakisJ. AttwoodR. BaertschJ. BaileyK. BarlowS. BeckE. BerryB. BirrenT. BloomP. BorkM. BotcherbyN. BrayM. R. BrentD. G. BrownS. D. BrownC. BultJ. BurtonJ. ButlerR. D. CampbellP. CarninciS. CawleyF. ChiaromonteA. T. ChinwallaD. M. ChurchM. ClampC. CleeF. S. CollinsL. L. CookR. R. CopleyA. CoulsonO. CouronneJ. CuffV. CurwenT. CuttsM. DalyR. DavidJ. DaviesK. D. DelehauntyJ. DeriE. T. DermitzakisC. DeweyN. J. DickensM. DiekhansS. DodgeI. DubchakD. M. DunnS. R. EddyL. ElnitskiR. D. EmesP. EswaraE. EyrasA. FelsenfeldG. A. FewellP. FlicekK. FoleyW. N. FrankelL. A. FultonR. S. FultonT. S. FureyD. GageR. A. GibbsG. GlusmanS. GnerreN. GoldmanL. GoodstadtD. GrafhamT. A. GravesE. D. GreenS. GregoryR. GuigoM. GuyerR. C. HardisonD. HausslerY. HayashizakiL. W. HillierA. HinrichsW. HlavinaT. HolzerF. HsuA. HuaT. HubbardA. HuntI. JacksonD. B. JaffeL. S. JohnsonM. JonesT. A. JonesA. JoyM. KamalE. K. KarlssonD. KarolchikA. KasprzykJ. KawaiE. KeiblerC. KellsW. J. KentA. KirbyD. L. KolbeI. KorfR. S. KucherlapatiE. J. KulbokasD. KulpT. LandersJ. P. LegerS. LeonardI. LetunicR. LevineJ. LiM. LiC. LloydS. LucasB. MaD. R. MaglottE. R. MardisL. MatthewsE. MauceliJ. H. MayerM. McCarthyW. R. McCombieS. McLarenK. McLayJ. D. McPhersonJ. MeldrimB. MeredithJ. P. MesirovW. MillerT. L. MinerE. MonginK. T. MontgomeryM. MorganR. MottJ. C. MullikinD. M. MuznyW. E. NashJ. O. NelsonM. N. NhanR. NicolZ. NingC. NusbaumM. J. O'ConnorY. OkazakiK. OliverE. Overton-LartyL. PachterG. ParraK. H. PepinJ. PetersonP. PevznerR. PlumbC. S. PohlA. PoliakovT. C. PonceC. P. PontingS. PotterM. QuailA. ReymondB. A. RoeK. M. RoskinE. M. RubinA. G. RustR. SantosV. SapojnikovB. SchultzJ. SchultzM. S. SchwartzS. SchwartzC. ScottS. SeamanS. SearleT. SharpeA. SheridanR. ShownkeenS. SimsJ. B. SingerG. SlaterA. SmitD. R. SmithB. SpencerA. StabenauN. Stange-ThomannC. SugnetM. SuyamaG. TeslerJ. ThompsonD. TorrentsE. TrevaskisJ. TrompC. UclaA. Ureta-VidalJ. P. VinsonA. C. Von NiederhausernC. M. WadeM. WallR. J. WeberR. B. WeissM. C. WendlA. P. WestK. WetterstrandR. WheelerS. WhelanJ. WierzbowskiD. WilleyS. WilliamsR. K. WilsonE. WinterK. C. WorleyD. WymanS. YangS. P. YangE. M. ZdobnovM. C. Zody, and E. S. Lander. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420**:520-562.

Wells, H. G., and G. T. Caldwell. 1914a. The purine enzymes of the orang-utan (Simia satyrus) and chimpanzee (Anthropopithecus troglodytes). J Biol Chem **18**:157-165.

Wells, H. G., and G. T. Caldwell. 1914b. The purine enzymes of the orang-utan (Simia satyrus) and chimpanzee (Anthropopithecus troglodytes). J Biol Chem **18**:157-165.

Wheelan, S. J., D. M. Church, and J. M. Ostell. 2001. Spidey: a tool for mRNA-to-genomic alignments. Genome Res. **11**:1952-1957.

Wiechowski, W. 1909a. Das Vorhandensein von Allantoin im normalen Menschenharn und seine Bedeutung fur Beurteilung des menschlichen Harnsaurestoffwechsels. Biochemische Zeitschrift **19**:368-383.

Wiechowski, W. 1912. Ein Beitrag zur Kenntnis des Purinstoffwechsels der Affen. Prag. med. Wochenschr **37**:275.

Wiechowski, W. 1909b. Das Vorhandensein van Allantoin im normalen Menschenharn und seine Bedeutung fur die Beurteilung des menschlichen Harnsaurestoffwechsels. Biochemische Zeitschrift **19**:368-383.

Winter, H., L. Langbein, M. Krawczak, D. N. Cooper, L. F. Jave-Suarez, M. A. Rogers, S. Praetzel, P. J. Heidt, and J. Schweizer. 2001. Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence. Hum Genet **108**:37-42.

Wong, S., and K. H. Wolfe. 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat Genet **37**:777-782.

Wu, C.-L., and D. W. Melton. 1993a. Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. Nat Genet **3**:235-240.

Wu, C. L., and D. W. Melton. 1993b. Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. Nat Genet **3**:235-240.

Wu, X., M. Wakamiya, S. Vaishnav, R. Geske, C. Montgomery, Jr., P. Jones, A. Bradley, and C. T. Caskey. 1994. Hyperuricemia and urate nephropathy in urate oxidase-deficient mice. Proc Natl Acad Sci U S A **91**:742-746.

Wu, X. W., C. C. Lee, D. M. Muzny, and C. T. Caskey. 1989. Urate oxidase: primary structure and evolutionary implications. Proc Natl Acad Sci U S A **86**:9412-9416.

Wu, X. W., D. M. Muzny, C. C. Lee, and C. T. Caskey. 1992. Two independent mutational events in the loss of urate oxidase during hominoid evolution. J Mol Evol **34**:78-84.

Xue, L., and M. Noll. 1996. The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. EMBO J. **15**:3722-3731.

Xue, Y., A. Daly, B. Yngvadottir, M. Liu, G. Coop, Y. Kim, P. Sabeti, Y. Chen, J. Stalker, E. Huckle, J. Burton, S. Leonard, J. Rogers, and C. Tyler-Smith. 2006. Spread of an inactive form of caspase-12 in humans is due to recent positive selection. Am J Hum Genet **78**:659-670.

Xue, Y., D. Sun, A. Daly, F. Yang, X. Zhou, M. Zhao, N. Huang, T. Zerjal, C. Lee, N. P. Carter, M. E. Hurles, and C. Tyler-Smith. 2008. Adaptive evolution of UGT2B17 copy-number variation. Am J Hum Genet **83**:337-346.

Yandell, M., C. J. Mungall, C. Smith, S. Prochnik, J. Kaminker, G. Hartzell, S. Lewis, and G. M. Rubin. 2006. Large-Scale Trends in the Evolution of Gene Structures within 11 Animal Genomes. PLoS Comput Biol **2**:e15.

Yang, Z. 2003. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555-556.

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

Yeldandi, A. V., V. Yeldandi, S. Kumar, C. V. Murthy, X. D. Wang, K. Alvares, M. S. Rao, and J. K. Reddy. 1991. Molecular evolution of the urate oxidase-encoding gene in hominoid primates: nonsense mutations. Gene **109**:281-284.

Zanotti, G., L. Cendron, I. Ramazzina, C. Folli, R. Percudani, and R. Berni. 2006. Structure of zebra fish HIUase: insights into evolution of an enzyme to a hormone transporter. J Mol Biol **363**:1-9.

Zhang, J., W. Zhang, D. Zou, G. Chen, T. Wan, M. Zhang, and X. Cao. 2003. Cloning and functional characterization of GMPR2, a novel human guanosine monophosphate reductase, which promotes the monocytic differentiation of HL-60 leukemia cells. J Cancer Res Clin Oncol **129**:76-83.

Zhu, J., J. Z. Sanborn, M. Diekhans, C. B. Lowe, T. H. Pringle, and D. Haussler. 2007. Comparative Genomics Search for Losses of Long-Established Genes on the Human Lineage. PLoS Comput Biol **3**:e247.

Zigmond, M. J., E. D. Abercrombie, T. W. Berger, A. A. Grace, and E. M. Stricker. 1990. Compensations after lesions of central dopaminergic neurons: some clinical and basic implications. Trends Neurosci **13**:290-296.

Zydowo, M. M., J. Purzycka-Preis, and N. Ogasawara. 1989. Deficiency of AMP deaminase in human erythrocytes. Adv Exp Med Biol **253A**:31-34.

8 Appendix

8.1 Supplementary Table 1

Supplementary Table 1. Sources of HPRT-gene family genomic and cDNA sequences.

Species	Gene	Genomic	cDNA
Human (*Homo sapiens*)	*HPRT1*	WGA, hg17	NM_000194
Human (*Homo sapiens*)	*PRTFDC1*	WGA, hg17	NM_020200
Marmoset (*Callithrix jacchus*)	*HPRT1*	**AC148367, AC148366**	Predicted
Marmoset (*Callithrix jacchus*)	*PRTFDC1*	**AC155198**	Predicted
Galago (*Otolemur garnettii*)	*HPRT1*	**AC148730**	Predicted
Galago (*Otolemur garnettii*)	*PRTFDC1*	**AC154057**[a]	Predicted[a]
Rabbit (*Oryctolagus cuniculus*)	*HPRT1*	**AC148374**	Predicted
Rat (*Rattus norvegicus*)	*HPRT1*	WGA, rn3	XM_217584
Rat (*Rattus norvegicus*)	*PRTFDC1*	WGA, rn3	XM_214518
Mouse (*Mus musculus*)	*HPRT1*	WGA, mm7	NM_013556
Mouse (*Mus musculus*)	*PRTFDC1*	WGA, mm7	Predicted[d]
Bat (*Rhinolophus ferrumequinum*)	*HPRT1*	**AC148443**	Predicted
Bat (*Rhinolophus ferrumequinum*)	*PRTFDC1*	**AC154166**	Predicted
Dog (*Canis familiaris*)	*HPRT1*	WGA, canFam2	Predicted
Dog (*Canis familiaris*)	*PRTFDC1*	WGA, canFam2[a]	Predicted[a]
Armadillo (*Dasypus novemcinctus*)	*HPRT1*	**AC151455**	Predicted
Elephant (*Loxodonta africana*)	*HPRT1*	**AC154077**	Predicted
Wallaby (*Macropus eugenii*)	*HPRT1*	**AC154070**	Predicted
Opossum (*Monodelphis domesticata*)	*HPRT1*	WGA, monDom1	Predicted
Opossum (*Monodelphis domesticata*)	*PRTFDC1*	WGA, monDom1	Predicted
Platypus (*Ornithorhynchus anatinus*)	*HPRT1*	**AC148426**	Predicted
Chicken (*Gallus gallus*)	*HPRT1*	WGA, galGal2	AJ132697
Chicken (*Gallus gallus*)	*PRTFDC1*	WGA, galGal2	Predicted
Turkey (*Meleagris gallopavo*)	*HPRT1*	**AC148422**	Predicted
Turkey (*Meleagris gallopavo*)	*PRTFDC1*	**AC155203**[b]	Predicted[b]
Zebra finch (*Taeniopygia guttata*)	*HPRT1*	**AC148572**	Predicted
Zebra finch (*Taeniopygia guttata*)	*PRTFDC1*	**AC151212**	Predicted[e]
Emu (*Dromaius novaehallandae*)	*PRTFDC1*	**AC155221**	Predicted
Alligator (*Alligator mississippiensis*)	*HPRT1*	**AC149026**[a]	Predicted[a]
Xenopus tropicalis	*HPRT1*	WGA, xenTro1	CR760889
Xenopus tropicalis	*PRTFDC1*	WGA, xenTro1	BX715188
Xenopus laevis	*HPRT1*	n.a.	CK798518
Xenopus laevis	*PRTFDC1*	n.a.	BC077960
Zebrafish (*Danio rerio*)	*HPRT1*	WGA, danRer3	BC046003
Zebrafish (*Danio rerio*)	*PRTFDC1*	WGA, danRer3[b]	BC047173
Zebrafish (*Danio rerio*)	*HPRT1L*	WGA, danRer3	BC071336
Catfish (*Ictalurus punctatus*)	*HPRT1*	**AC154053**	Predicted
Catfish (*Ictalurus punctatus*)	*PRTFDC1*	n.a.	CB940771
Rainbow trout (*Oncorhynchus mykiss*)	*HPRT1*	n.a.	BX076136
Rainbow trout (*Oncorhynchus mykiss*)	*PRTFDC1a*	n.a.	BX084317, BX084316
Rainbow trout (*Oncorhynchus mykiss*)	*PRTFDC1b*	n.a.	BX875181
Medaka (*Oryzias latipes*)	*PRTFDC1b*	n.a.	TC37910[f]

Stickleback (*Gasterosteus aculeatus*)	*HPRT1*	**AC148363**	Predicted
Stickleback (*Gasterosteus aculeatus*)	*HPRT1L*	n.a.	CD507371
Tetraodon nigroviridis	*HPRT1*	WGA, tetNig1	CR641420
Tetraodon nigroviridis	*PRTFDC1a*	WGA, tetNig1	Predicted
Tetraodon nigroviridis	*PRTFDC1b*	WGA, tetNig1	CR701223
Tetraodon nigroviridis	*HPRT1L*	WGA, tetNig1	Predicted
Fugu (*Takifugu rubripes*)	*HPRT1*	WGA, fr1	Predicted
Fugu (*Takifugu rubripes*)	*PRTFDC1a*	WGA, fr1	Predicted
Fugu (*Takifugu rubripes*)	*PRTFDC1b*	WGA, fr1[c]	Predicted
Fugu (*Takifugu rubripes*)	*HPRT1L*	WGA, fr1	Predicted
Southern puffer (*Spheroides nephelus*)	*HPRT1*	**AC153754**	Predicted
Ciona (*Ciona intestinalis*)	*HPRT1*	WGA, ci1	Predicted

The GenBank accession numbers for genomic sequences generated by the authors are indicated in **bold-type**. Other genomic sequences were obtained from whole-genome assemblies (WGA, version), or additional GenBank records. Where applicable, the GenBank accession number is also listed for each cDNA. n.a. = not available.

[a] Missing exon 1.
[b] Missing exon 5.
[c] The fugu whole-genome shotgun reads ti437971708 and ti436367218 were used to extend the fr1 assembly.
[d] Inactive.
[e] Low quality genomic sequence manually examined (ti 650756902) to resolve coding region disparity.
[f] TIGR Gene Indices record.

8.2 Supplementary Table 2

Supplementary Table 2. Gene and genome sizes.

Species	Gene	Gene size (bp)	Genome size (x 10^9 bp)
Human	*hsHPRT1*	39766	2.9 [a]
Marmoset	*cjHPRT1*	42471	3.4 [b]
Galago	*ogHPRT1*	46606	n.a.
Rabbit	*ocHPRT1*	42688	3.1 [b]
Mouse	*mmHPRT1*	32891	2.6 [a]
Rat	*rnHPRT1*	31395	2.7 [a]
Bat	*rfHPRT1*	37205	2.2 [b]
Dog	*cfHPRT1*	37197	2.4 [a]
Armadillo	*dnHPRT1*	35233	5.3 [c]
Elephant	*laHPRT1*	28343	n.a.
Opossum	*mdHPRT1*	28133	3.5 [a]
Wallaby	*meHPRT1*	26603	n.a.
Platypus	*oaHPRT1*	16113	3.0 [b]
Alligator	*amHPRT1*	23729	2.4 [b]
Chicken	*ggHPRT1*	14029	1.1 [a]
Turkey	*mgHPRT1*	15984	1.5 [b]
Zebra finch	*tgHPRT1*	15021	1.2 [b]
X. tropicalis	*xtHPRT1*	15095	1.6 [a]
Zebrafish	*drHPRT1*	8702	1.6 [a]
Catfish	*ipHPRT1*	8892	1.0 [b]
Stickleback	*gaHPRT1*	4301	0.6 [b]
Tetraodon	*tnHPRT1*	4403	0.3 [a]
Fugu	*trHPRT1*	3666	0.4 [a]
S. puffer	*snHPRT1*	3234	0.5 [b]
Human	*hsPRTFDC1*	102732	2.9 [a]
Marmoset	*cjPRTFDC1*	92140	3.4 [b]
Galago	*ogPRTFDC1*	96771	n.a.
Mouse	*mmPRTFDC1*	97147	2.6 [a]
Rat	*rnPRTFDC1*	103706	2.7 [a]
Bat	*rfPRTFDC1*	87477	2.2 [b]
Dog	*cfPRTFDC1*	92128	2.4 [a]
Opossum	*mdPRTFDC1*	97617	3.5 [a]
Chicken	*ggPRTFDC1*	44357	1.1 [a]
Turkey	*mgPRTFDC1*	41395	1.5 [b]
Zebra finch	*tgPRTFDC1*	41385	1.2 [b]
Emu	*dnoPRTFDC1*	45673	1.6 [b]
X. tropicalis	*xtPRTFDC1*	16875	1.6 [a]

Zebrafish	*drPRTFDC1*	24789	1.6 [a]
Tetraodon	*tnPRTFDC1a*	3920	0.3 [a]
Tetraodon	*tnPRTFDC1b*	4184	0.3 [a]
Fugu	*trPRTFDC1a*	3380	0.4 [a]
Fugu	*trPRTFDC1b*	3912	0.4 [a]
Zebrafish	*drHPRT1L*	10122	1.6 [a]
Tetraodon	*tnHPRT1L*	2554	0.3 [a]
Fugu	*trHPRT1L*	2666	0.4 [a]

n.a. = not available.
[a] Based on whole-genome sequence.
[b] Average of the genome size estimates from the Animal Genome Size Database (http://www.genomesize.com).
[c] Redi et al, J Heredity, 2005 96:485-493.

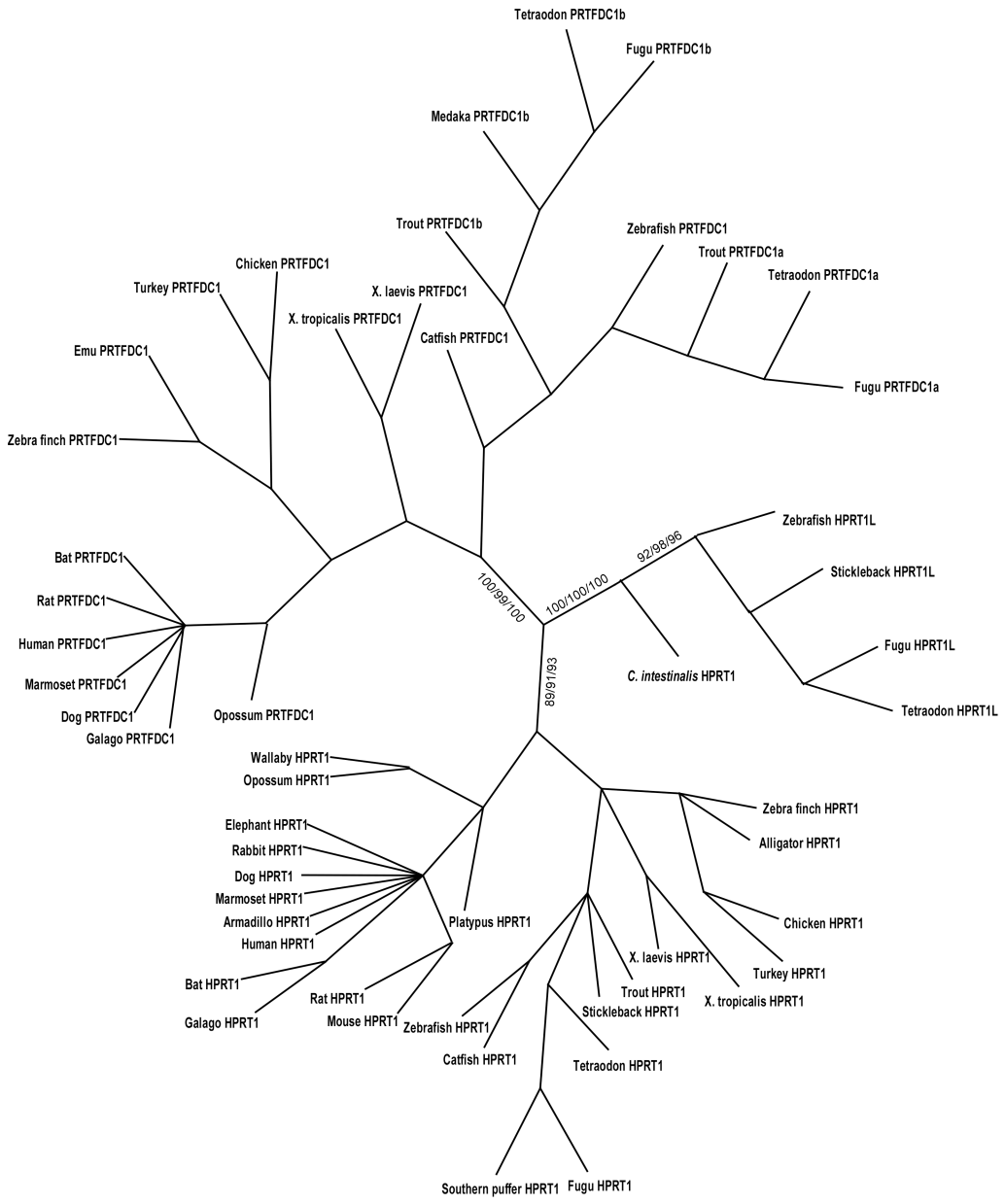
8.3 Supplementary Table 3

Supplementary Table 3 Amino acid substitutions in the HPRT critical sites

	Critical sites										
Gene	Leu-67	Gly-69	Ser-103	Tyr-104	Glu-133	Asp-134	Asp-137	Lys-165	Gly-189	Asp-193	Arg-199
hsPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
cjPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
ogPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
rnPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
rfPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
cfPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
mdPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
ggPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
mgPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
tgPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
dnoPRTFDC1	—	—	—	—	—	—	**Gly**[a]	—	—	—	—
xtPRTFDC1	—	—	**Cys**[a]	—	—	—	**Asn**	—	—	—	—
xlPRTFDC1	—	—	**Cys**[a]	—	—	—	**Asn**	—	—	—	—
drPRTFDC1	—	—	—	—	—	**Ala**[a]	—	—	—	—	—
ipPRTFDC1	—	—	—	—	—	**Ala**[a]	—	—	—	—	—
omPRTFDC1a	—	—	—	—	—	**Ala**[a]	—	—	—	—	—
omPRTFDC1b	—	**Ala**	—	—	—	**Ala**[a]	—	—	—	—	—
olPRTFDC1b	—	—	—	—	—	**Gly**[a]	**Gly**[a]	—	—	—	—
tnPRTFDC1a	—	—	—	—	—	**Ala**[a]	—	—	—	—	—
tnPRTFDC1b	—	—	—	—	—	**Ala**[a]	**Gly**[a]	—	—	—	—
trPRTFDC1a	—	—	—	—	—	**Ala**[a]	—	—	—	—	—
trPRTFDC1b	—	—	—	—	—	**Ala**[a]	**Gly**[a]	—	—	—	—
drHPRT1L	—	—	—	—	—	—	**Glu**	—	—	—	—
tnHPRT1L	—	—	—	—	—	—	**Glu**	—	—	—	—
trHPRT1L	—	—	—	—	—	—	**Glu**	—	—	—	—
gaHPRT1L	—	—	—	—	—	—	**Glu**	—	—	—	—

'—' indicates no amino acid substitution.
[a] Radical substitution.

8.4 Supplementary Figure 4



Supplementary Figure 1. Protein phylogeny of the HPRT-gene family.
Alignments of the amino acid sequences of the functional HPRT-gene family members was used to build maximum likelihood, neighbor joining and maximum parsimony trees. The consensus tree from of those three methods along with bootstrap support values from each method at the major branch points separating the three groups of genes is shown here.