Submitted: 02/12/2024
Accepted: 02/17/2025
Published: 04/10/2026

Research Article

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Keywords

MAMLD1, positive selection, pleiotropy, primate evolution, hypospadias

Email Correspondence

yunhua.ren@mail.utoronto.ca

https://doi.org/10.26443/msurj.v21i1.227

Ren, Yunhua[1]

Positive Selection on the Primate MAMLD1 Gene Suggests Adaptive Evolution in a Pleiotropic Regulatory Factor

Abstract

Gene regulatory factors (GRFs) are essential regulators of human phenotypic diversity. Among these, MAMLD1 acts as a transcriptional co-activator critical for male genital development. Mutations in *MAMLD1* are clinically associated with X-linked hypospadias type 2, a condition that alters the placement of the male urethral opening and potentially impacts reproductive fitness. In evolutionary biology, natural selection is often detected by comparing the rates of non-synonymous (*dN*) and synonymous (*dS*) substitutions ($\omega = dN/dS$). Positive selection ($\omega > 1$) represents the rapid fixation of beneficial mutations that enhance an organism's fitness, distinguishing it from purifying selection ($\omega < 1$), which removes deleterious mutations, and neutral evolution ($\omega = 1$), which involves random genetic drift. Building upon broader genomic surveys, this study provides a high-resolution analysis of *MAMLD1* using the PAML v4.9 suite across 20 primate species. By employing site models to assess selection across the primate phylogeny, we identified significant signals of positive selection. Findings suggest that *MAMLD1* has undergone adaptive evolution, likely driven by sexual selection and its role in reproductive isolation. This research underscores the gene's multifaceted role in shaping human-specific developmental phenotypes and evolutionary strategies.

Introduction

Sequence differences in gene coding regions and variance in gene regulatory systems may both contribute to phenotypic differences between humans and other species [1]. Variations in gene regulatory mechanisms can arise from changes in the DNA sequence of a regulatory region, which may impact gene expression, or changes in sequences of so-called gene regulatory factors (GRFs), which may affect the expression of their target genes [2]. GRFs play a variety of roles in gene regulation, including regulating the timing and tissue-specificity of a gene's expression. In this study, I focused on MAMLD1, a transcriptional co-activator expressed during male urogenital development and implicated in disorders of sex development. I selected MAMLD1 because (i) it is a pleiotropic regulatory factor with multiple downstream targets, and (ii) prior comparative genomic work has suggested that some GRFs show signatures of adaptive evolution in the human lineage [3]. I therefore tested the specific hypothesis that a subset of codons in MAMLD1 experienced episodic positive selection across the primate phylogeny, potentially reflecting selection on reproductive or developmental phenotypes. Hypospadias type 2, a disorder of sex development associated with MAMLD1 mutations, is characterized by abnormal placement of the male urethral opening on the underside of the penis [4]. This condition can lead to challenges in the normal function of male genitalia, including issues with urination and, in severe cases, reproductive difficulties [5]. Given its pleiotropic properties, regulatory control over multiple genes, and essential role in normal male genital development, I hypothesized that MAMLD1 is subject to evolutionary constraint, particularly within its functional domains. These constraints are likely to preserve the integrity of crucial sequences, maintaining conserved functions across species due to the gene's essential roles in development and physiological processes [1]. It has also been proposed that, in contrast to its target genes, gene regulatory systems evolve under less selective restrictions than their target genes [1].

The significant differences between humans and great apes, as well as between great apes and other primates provide a framework for investigating the genetic mechanisms underlying interspecific variation [6]. Recent studies have shown an average genetic difference of 1.2% between humans and chimpanzees, with non-coding regions showing slightly higher genetic differences than coding regions [7]. The difference is between 3-4% when non-alignable regions of the genome, such as insertions, deletions, and rearrangements, are taken into consideration. The average genetic difference between humans and the rhesus macaque, a primate species that is more evolutionarily distant from us, was determined to be significantly greater at 6.46%, or up to 9.24% when minor insertions and deletions are taken into account [8]. Some of the sequence changes might result from neutral evolution, which involves random genetic mutations that do not confer any selective advantage or disadvantage. Other changes might be caused by ongoing adaptive interactions between genomes and the environment, leading to positive selection [9]. Under positive selection, genetic variants that increase an organism's fitness or reproductive success increase in frequency over time. Depending on the dating technique, the human lineage diverged from its close relatives, chimpanzees, and bonobos, about 5.5–11.5 million years ago (Ma) [10]. The resulting human phenotype has historically been attributed to continual adaptations to niches and local environments [11]. The discovery of genes that have evolved by positive selection can shed light on how animals adapt to their environment and provide answers to some crucial biological problems, such as how a particular phenotype developed [8]. Across primates, many coding genes are dominated by purifying selection, although the identity and strength of positive selection can differ among lineages depending on ecological, demographic, and reproductive factors. Purifying selection is the evolutionary process that eliminates deleterious alleles from a population, thereby preserving the fitness of the species by preventing harmful genetic variations from becoming widespread. It has long been hypothesized that identifying genes that have undergone positive selection along the human lineage could contribute to uncover biologically significant genetic changes that uniquely identify humans [12]. As a result, researchers have used various methods to search the human genome for signs of positive selection in particular genes. Notably, Jovanovic et al. (2021) re-

ported evidence of positive selection in MAMLD1 using expanded genomic sampling and complementary methods, identifying candidate codons such as 726 and 728. Following the findings of Jovanovic et al. (2021), the goal of the present study is to independently assess how animal model choice affects inference of the previously identified positively selected sites using a streamlined PAML analysis**.** Shared environmental challenges, known as selective pressures, can lead to similar evolutionary responses across a group of species [13], resulting in uniform selection patterns, or traits that are consistently favored by natural selection across these species. This phenomenon suggests that certain traits may emerge not by coincidence but as adaptive responses to similar environmental demands. However, positively selected mutations are seldom found at polymorphic sites [3]. Through the process of positive selection, beneficial mutations are typically fixed rapidly within a population. This rapid fixation means that advantageous genetic variations tend to become common or even universal among individuals of a species, as they confer an increased chance of survival and reproduction. The exclusion of probable false positive candidates discovered at the species level is thus made possible by examining the polymorphism of positively selected codons at the population level. The main objective of this paper is to examine the MAMLD1 gene as a potential candidate that may exhibit signs of positive selection, with a focus on how selective pressures on MAMLD1 are distributed across primate species, supporting the gene's functional importance in human evolution. The ability to detect positive selection within a gene is significantly influenced by the number of available sequences for analysis. A larger dataset of sequences, including those from humans and their close relatives, enhances the statistical power of evolutionary models to identify subtle changes indicative of positive selection. Consequently, this study utilizes 20 sequences, encompassing both human and closely related species, to ensure a robust analysis capable of detecting signs of positive selection on the MAMLD1 gene.

Importantly, MAMLD1 is located on the X chromosome, which can experience evolutionary dynamics that differ from autosomes due to hemizygosity in males, sex-biased effective population size, and potentially greater exposure of recessive alleles to selection. In addition, human evolutionary history includes demographic bottlenecks and genetic drift, which can reduce genetic diversity and sometimes mimic or obscure signals of selection in sequence-based tests. Because codon models infer selection from patterns of fixed substitutions along branches, demographic history is not modeled explicitly here; therefore, lineage-level evidence should be interpreted cautiously and ideally complemented by population-level tests of fixation and by comparison to control loci.

Table 1. Summary of sequence data used. Species name, corresponding MAMLD1 accession number on NCBI and abbreviation. The first 10 sequences are the original dataset provided by Prof. Irwin; the last 10 sequences are displayed in search order.

Species name	Accession number	Abbreviation
Homo sapiens	ENST00000370401	Human
Pan troglodytes	ENSPTRT00000107245	T_glodytes
Nomascus leucogenys	ENSNLET00000017270	L_cogenys
Macaca mulatta	ENSMMUT00000074913	Mulatta
Microcebus murinus	ENSMICT00000017610	Murinus
Otolemur garnettii	ENSOGAT00000013472	Garnettii
Cebus imitator	ENSCCAT00000018615	Imitator
Rhinopithecus roxellana	ENSRROT00000058882	Roxellana
Aotus nancymaae	ENSANAT00000034312	Nancymaae
Pongo abelii	ENSPPYT00000024273	Abelii
Pan paniscus	XM_008968068	Paniscus
Gorilla gorilla gorilla	XM_004065005	Gorilla
Papio anubis	XM_031661028	Anubis
Trachypithecus francoisi	XM_045384134	Francoisi
Cercocebus atys	XM_012081351	Atys_dna
Macaca thibetana thibetana	XM_050775309	Thibetana
Hylobates moloch	XM_032757793	Moloch
Macaca nemestrina	XM_011765830	N_estrina
Macaca fascicularis	XM_045384134	F_icular
Rhinopithecus bieti	XM_017865534	Bieti

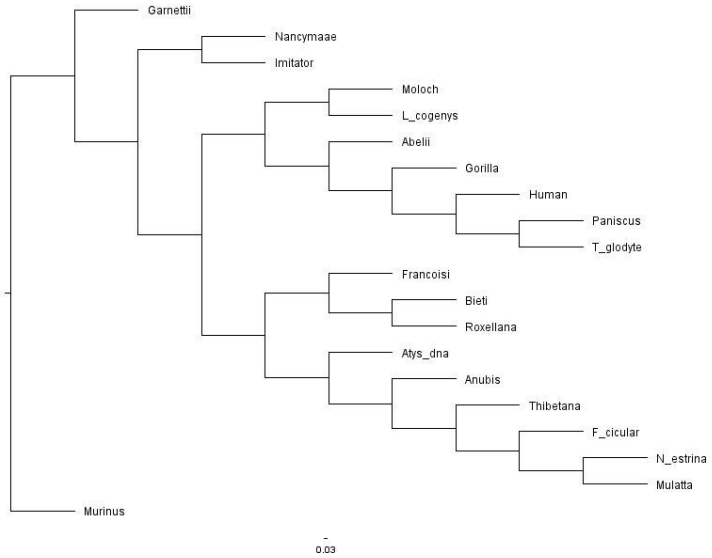

Figure 1. . Primate species tree used for codon-model analyses. Species are labeled using common names (human, chimpanzee, bonobo, gorilla, orangutan, baboon, rhesus macaque, etc.) to improve readability. *Microcebus murinus* is used to root the primate tree

Materials and Methods

Sequence Dataset and phylogenetic analysis

A total of 20 sequences were used for this analysis, one human sequence and 19 ortholog sequences from closely related primates. First ten sequences in Table 1 were downloaded from Ensembl and the last ten sequences were from the NBCI nBLAT website. Alignment was performed using MUSCLE (codon-aware) in MEGA, and the phylogeny was inferred using PhyML; FigTree was used only for tree visualization. The TN93 + G model of evolution was chosen as it was the best fit model according to AICc values. This analysis uses one curated reference coding sequence per species (i.e., an orthologous representative) to infer long-term substitution patterns along branches of the species tree. In codon-based phylogenetic models, branch lengths summarize substitutions accumulated since divergence and therefore reflect lineage-level evolutionary history rather than within-species polymorphism. Consequently, these models can detect recurrent selection operating over millions of years, but they cannot determine whether candidate sites are fixed or polymorphic within modern populations. We therefore interpret lineage-level signals using log-likelihood ratio tests (LRTs) based on ω as evidence of historical selection pressures and treat population-level verification as an important limitation and future direction. In general, the above displayed tree (Figure 1) has a similar topology to a recently published species tree [14].

Candidate sequences were obtained as annotated one-to-one orthologs from Ensembl when available. Additional primate sequences were retrieved from NCBI and verified by reciprocal best-hit BLAST against the human MAMLD1 coding sequence to reduce the risk of paralog inclusion.

Site Model test for Positive Selection

Estimating the ratio of non-synonymous to synonymous substitutions (ω) between the coding regions of homologs is a common method for measuring the selective pressure acting on protein-coding genes. The empirical p-values were calculated under the assumption that the log-ratio tests (LRT) had a $\chi 2$ distribution. The Bayes Empirical Bayes (BEB) approach, implemented in PAML v4.9, was used to estimate the posterior probabilities that individual codon sites belong to site classes with dN/dS (ω) > 1. This method identifies codons with strong statistical support for positive

Table 2. Summary table of PAML (Phylogenetic Analysis by Maximum Likelihood) analysis, showcasing the results of running several evolutionary models on the MAMLD1 gene sequences.

Model	np	lnL	k					Null	LRT	df	P
M0	39	-3678.103535	1.15655					n/a			
M1a	40	-35016.051963	0.8736	**p:**	0.62208	0.37792		M0	3564.103	1	0.0000
				w:	0.05164	1.00000					
M2a	42	-33345.590264	1.16499	**p:**	0.16244	0.71091	0.12665	M1a	3340.923	2	0.0000
				w:	0.19078	1.00000	95.07202				
M2a_rel	42	-33345.590623	1.16499	**p:**	0.16245	0.71090	0.12665	M1a	3340.923	2	0.0000
				w:	0.19078	1.00000	95.07109				
M3	43	-33345.355812	1.16018	**p:**	0.15537	0.71707	0.12756	M0	6905.495	4	0.0000
				w:	0.17866	0.95563	91.89691				
M7	40	-35127.479377	0.85969	**p:**	p=0.31920	q=0.39792		n/a			
M8	42	-33331.356206	1.13260	**p:**	p=0.85848	q=0.34956		M7	3592.246	2	0.0000
				w:	79.05894						

selection across the primate phylogeny. Subsequent investigations should concentrate on codons identified as being under positive selection.

M0 (one-ratio): assumes a single rate of evolution across all branches. M1a (nearly neutral): contrasting neutral versus selection sites. M2a (positive selection): allows for sites under positive selection. M2a_rel: a refined version of M2a offering more detailed insights into selection. M3 (discrete): permits variable selection across sites. M7 (beta): modeling variation under a beta distribution without positive selection. M8 (beta&ω): introducing positive selection on top of the beta distribution. Each model plays a crucial role in identifying different aspects of evolutionary pressures acting on the gene, with particular emphasis on pinpointing specific codons where positive selection may have occurred, offering insights into the adaptive evolution of the MAMLD1 gene in humans and their close relatives.

Results

Positive Selection Across the Primate MAMLD1 Gene

Testing for positive selection pressure was detected using codon models of evolution implemented in PAML (Yang 2007). These sites are critical positions within the gene's coding region where amino acid substitutions occur. The ω (omega) parameters, which represent the ratio of non-synonymous (*dN*) to synonymous (*dS*) substitution rates, vary among these sites. A ratio greater than 1 (ω > 1) at any site suggests that the site is under positive selection, indicating that amino acid changes at these positions are favored by natural selection because they likely confer some adaptive advantage. Our analysis revealed that under models M2a, M3, M7, and M8, several sites across the MAMLD1 gene exhibit ω parameters indicative of positive selection pressure, suggesting these coding sites have undergone adaptive changes that could be functionally relevant to organismal fitness. These adaptive changes in the protein indicate positive selection.

In Table 2, column 'p' represents the estimated proportion of sites belonging to a specific ω category, while column 'w' represents the estimated ω (*dN/dS*) value for that category. The p-value (Column M) represents the statistical significance of the Likelihood Ratio Test (LRT). Specifically, model M8 identified a class of sites with a highly elevated ω = 79.06, and model M2a identified a class of sites with ω = 95.07, indicating strong positive selection at these positions.

Discussion

The codon-model analyses are consistent with episodic positive selection acting on a subset of MAMLD1 codons. These methods primarily detect recurrent selection that leaves multiple fixed nonsynonymous substitutions. They may miss adaptation driven by a single major change, and they cannot on their own distinguish fixation from standing polymorphism. Biologically, MAMLD1 is implicated in male urogenital development, so one plausible selective context is sexual selection or other reproductive selection pressures acting on fertility-related phenotypes. However, linking sequence evolution in a single locus to speciation requires additional evidence such as demonstrated reproductive barriers, functional effects of the substituted residues, or consistent divergence across populations. Thus we treat speciation as a hypothesis that invites further research in this area.

Although MAMLD1 has been identified as a GRF under positive selection, no interactions between MAMLD1 and other GRFs have been reported to our knowledge. This lack of reported interactions highlights the unique evolutionary pathway of MAMLD1 and underscores the importance of further research into its specific role and mechanisms within the regulatory networks influencing human development and evolution [14]. It is possible to hypothesize that this is because the set of genes it controls are separate or functioning via different pathways, and the epistasis between them cannot be identified using the data currently available. The independenceof MAMLD1's effects on different traits and its potential role in complex phenotypes can be accounted if potential pleiotropic effect is considered. Because MAMLD1 normally regulates the expression of several genes, it can have physiological and morphological implications across scales, affecting cells, tissues or at the level of whole organisms [15]. These implications could be independently adaptive; this means that different mutations or variations within this gene can lead to separate adaptive advantages. It has been shown that minor mutations, even within a single gene, may provide a quick road to phenotypic adaption [16]. The different positively selected sites (PSSs) that were found within MAMLD1 could add to the pleiotropic effect and connect with distinct phenotypes. This has been observed before for certain other genes, where various polymorphism sites have varied trait relationships [4]. Another reasonable explanation for the lack of gene interactions is that the selective pressures on MAMLD1 were acting at a distinct time points after the split of the human lineage. For MAMLD1, this could mean that the gene has undergone several rounds of optimization, each enhancing different aspects of its regulatory functions. Early in human evolution, for instance, a selective event might have favored mutations that improved reproductive success, while much later, another event might have selected for variations that enhanced the gene's role in development or disease resistance.

The evolutionary history of MAMLD1, marked by these separate yet cumulative adaptations, underscores its potential significance in understanding the genetic foundations of human development, health, and disease.

A key limitation of the current design is the absence of comparator loci that would contextualize model behavior such as an X-linked gene with no prior expectation of positive selection as a negative control, and an autosomal gene with reported positive selection as a positive control. Includ-

ing such controls would help evaluate whether the magnitude of likelihood improvements observed for MAMLD1 is exceptional or typical under the same alignment and model settings. In addition, outgroup choice can influence branch length estimates; here I used *Microcebus murinus* as an outgroup within primates to root the tree while avoiding the extreme divergence and life-history differences that would be introduced by using rodents. Future work should implement explicit control-gene analyses and sensitivity checks for outgroup selection.

Acknowledgments

This research was conducted as part of an undergraduate course HMB460H1: Molecular Evolution and Genomics at University of Toronto St. George Campus. I would like to thank Prof. David M. Irwin and Prof. Belinda Chang for their guidance, support, and valuable insights.

References

1. Anderson, J., Vilgalys, T. & Tung, J. Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. *Curr. Opin. Genet. Dev.* **62,** 16–22 (2020).
2. Siepel, A. & Arbiza, L. Cis-regulatory elements and human evolution. *Curr. Opin. Genet. Dev.* **29,** 81–89 (2014).
3. Jovanovic, V. et al. Positive selection in gene regulatory factors suggests adaptive pleiotropic changes during human evolution. *Front. Genet.* **12** (2021).
4. Chen, Y. Mutational study of the MAMLD1-gene in hypospadias. *Eur. J. Med. Genet.* **53,** 122–126 (2010).
5. Fukami, M. et al. CXorf6 is a causative gene for hypospadias. *Nat. Genet.* **38,** 1369–1371 (2006).
6. Nickel, G., Tefft, D., Goglin, K. & Adams, M. An empirical test for branch-specific positive selection. *Genetics* **179,** 2183–2193 (2008).
7. Kronenberg, Z. et al. High-resolution comparative analysis of great ape genomes. *Science* **360** (2018).
8. Su, Z. et al. Species specific exome probes reveal new insights in positively selected genes in nonhuman primates. *Sci. Rep.* **6** (2016).
9. Varki, A., Geschwind, D. & Eichler, E. Human uniqueness: genome interactions with environment, behaviour and culture. *Nat. Rev. Genet.* **9,** 749–763 (2008).
10. Jeong, C. & Rienzo, A. Adaptations to local environments in modern human populations. *Curr. Opin. Genet. Dev.* **29,** 1–8 (2014).
11. Clark, A. et al. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302,** 1960–1963 (2003).
12. Shultz, A. & Sackton, T. Immune genes are hotspots of shared positive selection across birds and mammals. *eLife* **8** (2019).
13. Slodkowicz, G. & Goldman, N. Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proc. Natl. Acad. Sci. U.S.A.* **117,** 5977–5986 (2020).
14. Yang, Z. & Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19,** 908–917 (2002).
15. Linnen, C. et al. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339,** 1312–1316 (2013).
16. Mackay, T., Stone, E. & Ayroles, J. The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10,** 565–577 (2009).