GBE

Saliva Protein Genes in Humans were Shaped During Primate Evolution

Petar Pajic [1,2], Luane Landau [1], Omer Gokcumen [1,*], Stefan Ruhl [3,*]

[1]Department of Biological Sciences, University at Buffalo, the State University of New York, Buffalo, NY 14260, USA

[2]Department of Chemistry, Yale University, New Haven, CT 06511, USA

[3]Department of Oral Biology, School of Dental Medicine, University at Buffalo, the State University of New York, Buffalo, NY 14214, USA

*Corresponding authors: E-mails: omergokc@buffalo.edu; shruhl@buffalo.edu

Accepted: August 20, 2025

Abstract

Genes within the secretory calcium-binding phosphoprotein locus diversified along with the formation of a calcified skeleton in vertebrates, the emergence of tooth enamel in fish, and the introduction of lactation in mammals, at each stage marking major transitions in life history. The secretory calcium-binding phosphoprotein (SCPP) locus also harbors genes expressed primarily and abundantly in the saliva of humans. Here, we explored the phylogeny and evolution of the saliva-related SCPP genes by harnessing available genomic and transcriptomic resources. We observe extensive diversification of SCPP genes within mammals, driven by gene duplications and losses, with the most pronounced changes occurring in the SCPP genes that are expressed in salivary glands. When comparing rodent and human SCPP genes, we concluded that regulatory shifts and gene turnover events likely facilitated the accelerated gain of salivary gland expression. In primate genomes, we found more recent duplication events that affected genes coding for proteins secreted in saliva. Several saliva-related SCPP genes in the primate lineage show signatures of positive selection, while the other genes in the SCPP locus remain conserved. Our results position saliva-related SCPP genes as highly malleable to evolutionary innovation. Variations shaped by dietary and pathogenic pressures likely influenced the functional properties of saliva proteins, impacting metabolic and immune-related traits in oral health among primates, including humans.

Key words: gene duplication, gene turnover, tooth enamel, milk caseins, saliva proteins.

Significance

Saliva plays an essential role in digestion, maintenance of tooth mineralization, immune defense, and oral health, yet little is known about how salivary proteins have evolved in response to environmental pressures. By analyzing the evolution of genes that encode key saliva proteins, we found that these genes have undergone frequent duplications, losses, and regulatory changes, which became particularly evident in the primate lineage. Our findings suggest that rapid evolution in saliva proteins occurred in primates. This work highlights the unique evolutionary flexibility of saliva-related genes and provides new insights into how adaptations to diet and disease may have influenced primate biology, including humans.

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

Introduction

In our previous work, we have shown that the genes encoding saliva mucin-7 (*MUC7*) and saliva amylase (*AMY1*) (Xu et al. 2016; Pajic et al. 2019; Pajic et al. 2022) show variation among different mammalian species, and hypothesized that saliva protein genes, in general, are more subject to selective evolutionary pressure due to differences in species-specific diets and habitats. The secretory calcium-binding phosphoprotein (SCPP) gene locus offers an excellent model for investigating this hypothesis, as it harbors genes expressed in saliva, which are neighbored by genes exerting fundamentally different functions such as tooth mineralization (Kawasaki et al. 2005) and calcium homeostasis in milk (Kawasaki et al. 2011).

The SCPP gene family has been of interest for its regulation of calcium-phosphate concentration in tissues as diverse as bone, dentin, enamel, mammary glands, and salivary glands (Kawasaki et al. 2009). The SCPP gene repertoire, comprising more than 20 genes in humans, was proposed to have originated from a single ancestral gene, *SPARCL1*, involved in the mineralization of skeletal tissue in bony fish (Kawasaki et al. 2007). SCPP genes evolved as calcium-binding proteins (Kawasaki et al. 2004). They diversified in function through major evolutionary transitions, including the formation of a calcified skeleton (Kawasaki et al. 2004), the emergence of teeth (Kawasaki and Weiss 2008), and the introduction of lactation in mammals (Kawasaki et al. 2011). Many members of this gene family are relevant to oral biology as they encode not only proteins that form tooth enamel but also others that protect and remineralize the enamel of erupted teeth and rank among the most abundant proteins in human saliva (Saitou et al. 2020).

The SCPP genes can be divided into two subfamilies that form two distinct gene clusters on chromosome 4. One cluster encodes the D/E/S-rich proteins and the other encodes the P/Q-rich proteins (Kawasaki et al. 2007). The D/E/S-rich SCPP genes code for five acidic proteins (DSPP, DMP1, IBSP, MEPE, and SPP1), which have functions in bone and dentin mineralization (George and Veis 2008). The P/Q-rich locus comprises 16 syntenically clustered genes along with one distal enamel-related gene, *AMEL*, located on the sex chromosomes. The P/Q-rich SCPP genes function in the mineralization of the enamel matrix and sustain calcium-binding in milk (Kawasaki and Weiss 2008; Kawasaki et al. 2011). Notably, 5 out of the 17 P/Q-rich genes in this locus are highly expressed in salivary glands (Saitou et al. 2020) and secreted into saliva (Lau et al. 2021) (Table 1).

Here, we set out to test the hypothesis that the saliva-related genes in this cluster evolved faster than their neighboring enamel-related or milk-related genes. To achieve this goal, we conducted a bioinformatic examination of the locus across vertebrate genomes representing major evolutionary lineages, including fishes, amphibians, reptiles, birds, and mammals. Given that the SCPP genes code for some of the most abundant saliva proteins in primates, including humans (Saitou et al. 2020; Thamadilok et al. 2020), the insights gained from this study will ultimately enhance our understanding of how mammalian, and especially human saliva proteins, co-evolved in their functions with species-specific habitat-driven diet preferences and environmental pathogen challenges.

Results

SCPP Genes With Functions in Saliva Are Derived From Conserved Enamel-related Genes

The SCPP gene locus has been pointed out as an example of rapid gene turnover (Ponting 2017). To find out how different functional categories of SCPP genes are affected by this gene turnover, we analyzed the genetic variation in the SCPP locus in 61 vertebrate species comprising mammals ($n = 48$), birds ($n = 4$), reptiles ($n = 4$), amphibians ($n = 1$), lobe-finned fishes ($n = 1$), ray-finned fishes ($n = 3$), and cartilaginous fishes ($n = 1$) (Fig. 1). Our study builds upon pioneering work by Kawasaki and coworkers, which linked the variation in this locus to calcification of skeleton and teeth, as well as calcium homeostasis in milk (Kawasaki and Weiss 2008; Kawasaki 2009; Kawasaki et al. 2011). The availability of a greater number of well-annotated, long-read sequenced vertebrate genomes allowed us to probe for the presence and absence of SCPP genes in a wide selection of vertebrate species. To investigate the evolution of the saliva-related SCPP genes within the context of the evolution of the entire locus, we used a three-pronged approach: (i) searching for sequence similarity, (ii) comparing gene order, and (iii) analyzing individual gene annotations in multiple databases (see Methods).

We defined four functional categories (enamel-related, milk-related, broad, and saliva-related) based on expression trends of SCPP genes as previously defined (Kawasaki and Weiss 2008) (Table 1, Fig. 1). Our analysis supports earlier studies on the evolution of the SCPP locus in vertebrates and mammals (Kawasaki 2009; Kawasaki et al. 2011). We found that the presence of enamel-related genes is highly conserved across vertebrates, echoing previous work (Kawasaki and Amemiya 2014). The milk-related beta-casein (CSN2) and kappa-casein (CSN3) genes are also conserved across all surveyed mammalian species, as suggested previously (Zhou et al. 2021) (Fig. 1). It has also been shown that *CSN1S2A* and *CSN1S2B* are pseudogenized in humans (Kawasaki et al. 2011). We were able to show that these pseudogenization events happened first in Old World monkeys (*CSN1S2B*) and subsequently in

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

Table 1 SCPP gene names, reported functions, and tissue expression trends in humans.

Gene	Function	Expression
CSN1S1: Casein Alpha S1	Calcium binding, nutrition source	Lactating breast (milk)
CSN2: Casein Beta	Calcium binding, nutrition source	Lactating breast (milk)
STATH: Statherin	Calcium binding, antimicrobial	Salivary glands (saliva)
HTN3: Histatin 3	Calcium binding, antimicrobial	Salivary glands (saliva)
HTN1: Histatin 1	Calcium binding, antimicrobial	Salivary glands (saliva)
CSN1S2: Casein Alpha S1	Pseudogene	Unknown
PRR27: Proline-Rich Protein 27	Unknown	Broad
ODAM: Odontogenic, Ameloblast Associated	Enamel mineralization	Broad
FDCSP: Follicular Dendritic Cell Secreted Protein	Immune regulation	Broad (salivary glands)
CSN3: Casein Kappa	Calcium binding, nutrition source	Lactating breast (milk)
CABS1: Calcium Binding Protein, Spermatid *Associated 1*	Calcium binding, sperm flagella	Broad
SMR3A: Submaxillary Gland Androgen Regulated Protein 3A	Regulation of sensory perception of pain	Salivary glands (saliva)
SMR3B: Submaxillary Gland Androgen Regulated Protein 3B	Regulation of sensory perception of pain	Salivary glands (saliva)
PROL1 Basic Proline-Rich Lacrimal Protein	Regulation of sensory perception of pain	Lacrimal and salivary glands (saliva)
MUC7: Mucin 7	Microbial binding, regulation of physical properties of saliva	Salivary glands (saliva)
AMTN: Amelotin	Enamel mineralization	Ameloblasts
AMBN: Ameloblastin	Enamel mineralization	Ameloblasts
ENAM: Enamelin	Enamel mineralization	Ameloblasts
AMEL: Amelogenin (x-linked)	Enamel mineralization	Ameloblasts
SCPPPQ1:SCPP Proline-Glutamine Rich 1	Odontogenesis	Ameloblasts
SPARCL1: SPARC Like 1	Calcium/metal-binding	Broad
DSPP: Dentin Sialophosphoprotein	Calcium binding, dentinogenesis	Odontoblasts
DMP1: Dentin Matrix Acidic Phosphoprotein 1	Dentinogenesis	Osteoblasts and odontoblasts
IBSP: Integrin Binding Sialoprotein	Bone mineralization	Osteoblasts
MEPE: Matrix Extracellular Phosphoglycoprotein	Bone mineralization, phosphate regulation	Osteoblasts
SPP1: Secreted Phosphoprotein 1	Bone mineralization, signaling	Broad (osteoblasts)

Genes are listed in the table following their syntenic locations in the SCPP gene cluster. Functions are derived from the Human Protein Atlas and Gene Cards.

hominins, predating the human-Neanderthal ancestor (*CSN1S2A*) (Appendix Fig. S1). We also note the presence of a conserved gene, *CABS1*, within the SCPP locus across mammals, which to our knowledge has not been previously described as belonging to the SCPP family. Although this gene is nested within the SCPP gene family, its origin remains unclear, and further investigation is needed to elucidate its evolutionary history. Among all the SCPP genes, we found that the saliva-related genes (Table 1) were the most dynamic in their presence or absence among different species (Fig. 1). Thus, in our subsequent analysis, we focused on resolving the evolution of the saliva-related subgroup of SCPP genes.

Expression Patterns of SCPP Genes in Human and Mouse Salivary Glands Differ Substantially

The mouse is commonly used as an outgroup to investigate primate-specific evolution. It also currently is the only species besides humans for which a complete set of expression data from all three major salivary gland tissues is available (Gao et al. 2018). The null hypothesis, based on previous work (Chan et al. 2009; Brawand et al. 2011), is that for the majority of human genes, mouse orthologs should exist, exhibiting similar expression patterns. Contrary to these expectations, out of the 17 functionally annotated genes in the human SCPP locus, only 12 are found in the mouse genome. More specifically, the human saliva-related SCPP genes *HTN1*, *HTN3*, *STATH*, and *MUC7* do not have orthologs in the mouse, while the mouse contains the genes *Smr2, Gm7048, Gm7709, Gm7714, Gm7721* in its SCPP locus that do not have orthologs in humans (Fig. 2). We note that the 4 human SCPP genes, which do not have mouse orthologs, are among the top 20 genes expressed in salivary glands (Saitou et al. 2020), underscoring the distinctiveness of the mouse and human saliva protein repertoires.

Analyzing RNA-sequencing data, we investigated which of the 12 orthologous SCPP genes are expressed in salivary gland tissue of mice (Gao et al. 2018) and humans ((Saitou et al. 2020), Appendix Fig. S2). We found that none of these are expressed in the salivary glands

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

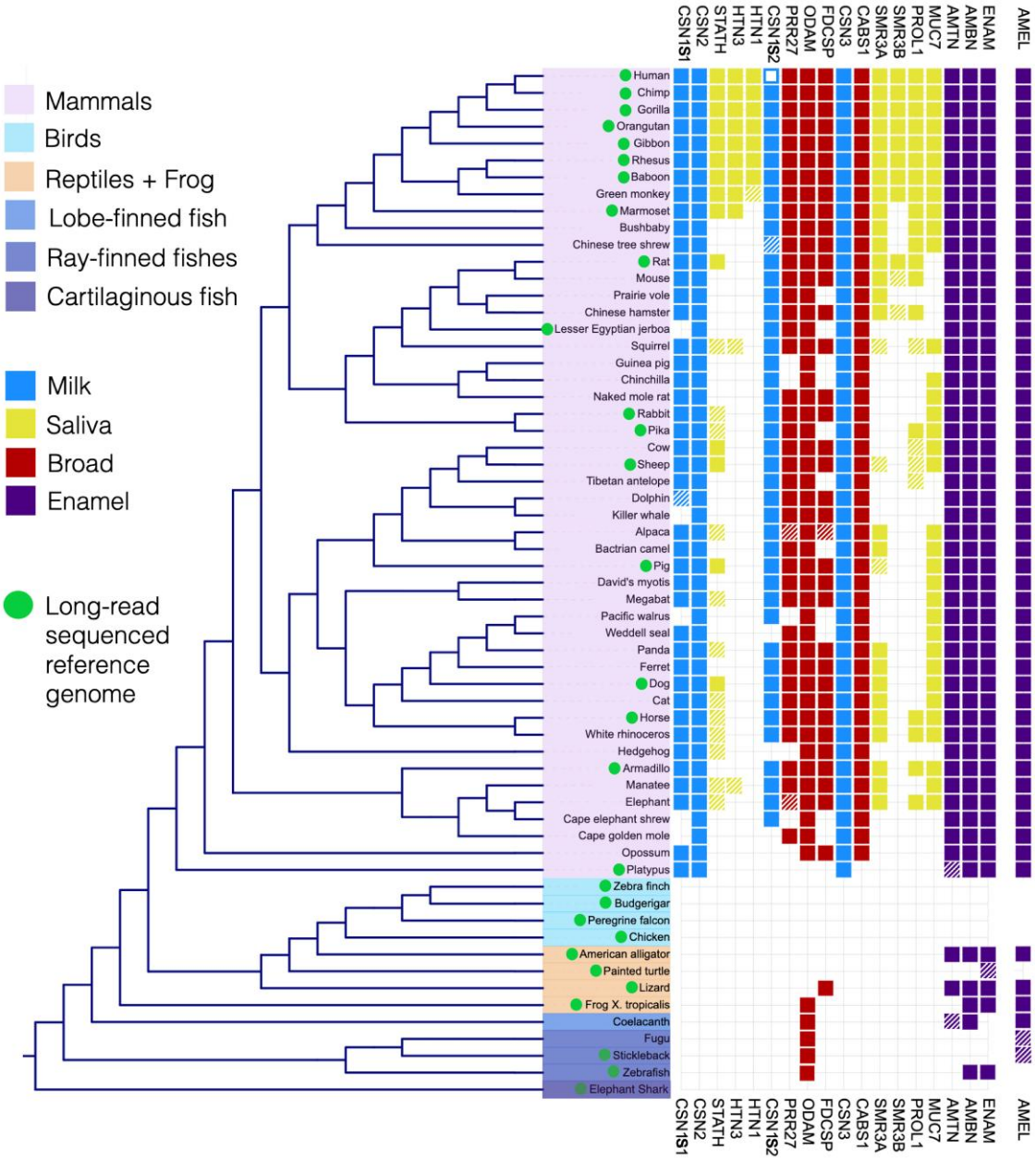

Fig. 1. P/Q-rich SCPP genes in vertebrates. The presence and absence of orthologs to the P/Q-rich SCPP genes found in humans are indicated across 61 vertebrate species. Genes are colored by expression categories: purple, enamel; yellow, salivary glands (saliva); blue, mammary glands (milk); red, broad expression (broad). Hatched boxes indicate genes found in only one of the two gene annotation databases used (see Methods). The databases, genome assemblies examined, and accession numbers of genes can be found in Table S1.

of both species. Of the 12 orthologs, only 3, namely *PRR27*, *FDCSP*, and *ODAM,* show expression in human salivary glands. Only two genes, namely *Smr3a* and *Prol1,* show expression in mouse salivary glands (Fig. 2). The case of *PROL1*/*Prol1* provides clues for the unexpected expression differences. As described earlier (Pajic et al. 2022), despite being evolutionary orthologs at the genetic level, the protein products of human *PROL1,* PROL1, and mouse *Prol1*, MUC10, are not functional orthologs. Human PROL1 is secreted primarily in the lacrimal glands, while mouse MUC10 is secreted in the salivary glands. Moreover, the protein product of mouse *Prol1* gained a mucin function in the rodent lineage, which is why it was annotated as MUC10, not as PROL1. In that same context, it is also noteworthy that the human SCPP gene product MUC7 is lost in the mouse lineage and has been functionally replaced by MUC10, which exhibits similar expression quantities and protein features.

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

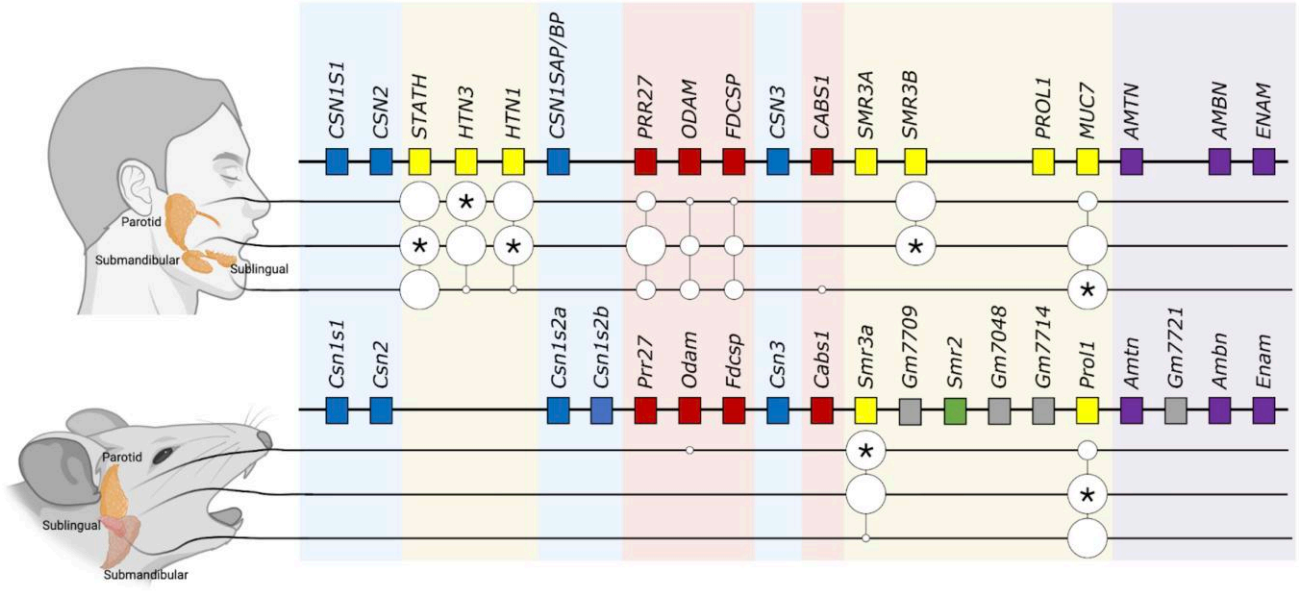

Fig. 2. Comparison of SCPP gene expression in the three major salivary gland types between human and mouse. SCPP genes are colored based on expression categories: purple, enamel; yellow, salivary glands (saliva); blue, mammary glands (milk); red, broad expression (broad). The size of the bubbles indicates the expression level categorized as low, medium, or high. For genes with high levels of expression, the glands with the highest expression are indicated by stars. Figure created in BioRender.

In sum, our results show that a combination of gene turnover events (loss and emergence of genes) and putative changes in regulatory sequences led to almost absolute difference in salivary gland expression trends between mouse and human SCPP genes.

Saliva-related SCPP Genes in Primates Evolved Faster Than Milk- and Enamel-related Ones

Given the diversity of diets consumed by different mammalian species, we hypothesized that the saliva-expressed SCPP genes evolved faster than their milk- and enamel-related counterparts. We investigated whether there is selection acting on the saliva-related genes within the primate lineage at the sequence level. We applied PhyloP (Pollard et al. 2010) (see Methods), which measures the evolutionary conservation of individual nucleotide positions across multiple species. Higher scores indicate slower evolution (higher sequence conservation), and lower scores indicate faster evolution (lower sequence conservation) at that particular position in our phylogeny of SCPP genes. We found that enamel-related and milk-related genes, along with the broadly-expressed genes, show significantly more sequence-level conservation than the saliva-related SCPP genes ($P < 2.2 \times 10^{-16}$, Fig. 3a & 3b). Because some saliva-related genes exhibit less gene presence than other SCPP genes across species (Fig. 1), our analysis compares per-base conservation scores from aligned regions only, allowing direct comparison across gene categories despite variation in gene presence. Our observation is consistent with a scenario where enamel-related SCPP genes have evolved under purifying selection, as previously suggested (Alazem and Abramyan 2019), while saliva-related genes have experienced bouts of rapid lineage-specific selection.

Further support for our hypothesis comes from the above-mentioned observation that almost half of the saliva-related SCPP genes in primates do not have orthologs in nonprimate genomes, whereas most of the other genes in this locus do (Fig. 1). To visualize the rate of gene turnover among mammals, we show the SCPP gene locus across nine species for which high-quality genome assemblies and annotations were available (Fig. 3c). These species exemplify major mammalian phylogenetic groups with unique habitats, ecological contexts, and different diets. We found multiple lineage-specific genes in these mammals, including so-called "orphan" mucins, which we previously showed recurrently arose from saliva-related precursor genes that code for proteins rich in proline (Pajic et al. 2022; Pajic 2025). Another gene called *histatherin* (*HSTN*), which we found in the cow SCPP gene cluster (Fig. 3c), was previously hypothesized to be a chimera, resulting from a gene fusion of statherin and histatin (de Sousa-Pereira et al. 2013). A more parsimonious explanation for its presence, given its syntenic location (Fig. 3c) and sequence similarity to the annotated *STATH* in cow (Appendix Fig. S3), is that it arose through lineage-specific duplication within the Artiodactyla lineage. Overall, our results indicate that saliva-related P/Q-rich SCPP genes are

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

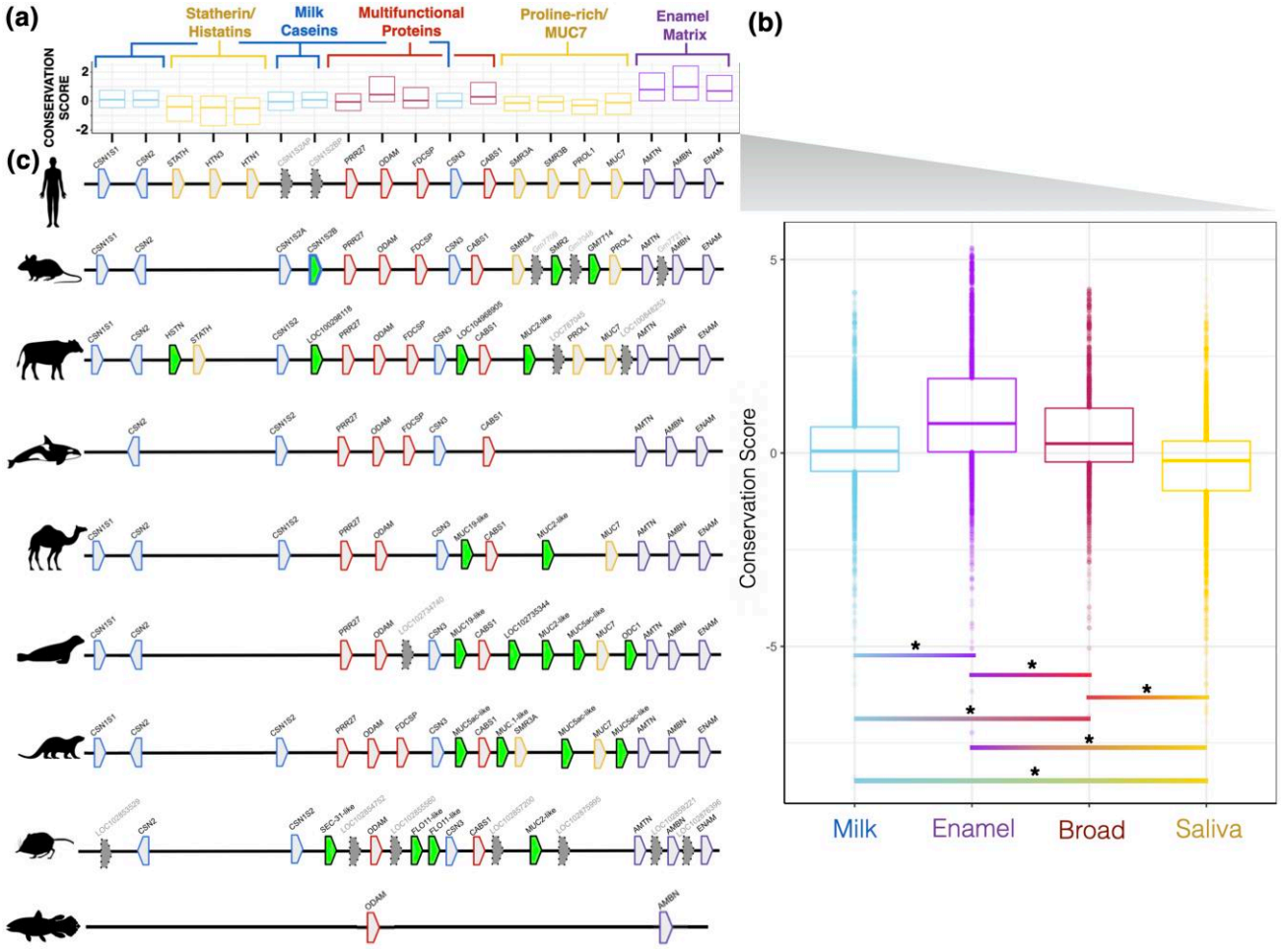

Fig. 3. Gene conservation and turnover within the SCPP locus. a) Box plots represent conservation scores calculated for different SCPP genes across 100 vertebrate species (phyloP100way) for each SCPP gene on a per-base-pair basis. Colors denote the functional category: purple, enamel; yellow, salivary glands (saliva); blue, mammary glands (milk); red, broad expression (broad). b) Box plot representing conservation scores calculated per base pair for the P/Q-rich SCPP genes and merged by functional category (enamel; $n = 3$, milk; $n = 5$, broad; $n = 4$, saliva; $n = 7$). Stars (*) indicate significance scores $P < 2.2 \times 10^{-16}$. c) The SCPP loci for nine selected species are shown to illustrate the turnover of genes within the locus. Genes are labeled according to their functional categories (same as in panel a). Lineage-specific genes are colored in green. Pseudogenes are colored in dark grey with dotted outlines.

more often affected by gene turnover events than P/Q-rich milk-related genes in the same locus.

Based on our gene conservation analysis, we further hypothesized that lineage-specific positive selective forces affected saliva-related SCPP genes more than the milk- and enamel-related genes in the locus. To test this hypothesis at the nucleotide level, we searched for lineage-specific selection acting on human SCPP genes and their orthologs in nonhuman primates. We focused our analysis on primates because the majority of saliva-related human SCPP genes do not have orthologs in nonprimate mammals (Fig. 1). We constructed multiple-sequence alignments involving the primate species analyzed in this study. Using BUSTED, a test for episodic positive selection affecting select branches in a phylogenetic tree (Murrell et al. 2015), we found evidence of lineage-specific selection for three saliva-related SCPP genes (*STATH*, *SMR3A*, and *MUC7)*. We also found evidence of lineage-specific selection for two other SCPP genes (*FDCSP* and *CSN1S1*) (Fig. 4a). Using dN/dS (nonsynonymous vs. synonymous substitution) pairwise comparisons, we found that saliva-related SCPP genes, when grouped together, show higher levels of amino-acid changing mutations than nonsaliva-related genes, indicating positive selection (Fig. 4b). We identified *HTN1*, *STATH*, and *HTN3* as exhibiting the highest proportion of nonsynonymous mutations among all SCPP genes. Collectively, these results support the hypothesis that positive selection led to the accelerated evolutionary diversification of some saliva-related SCPP genes in primates.

Diverse Origins of Saliva-related SCPP Genes in Primates

Given that the most abundantly expressed human saliva-related SCPP genes originated in the primate lineage,

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

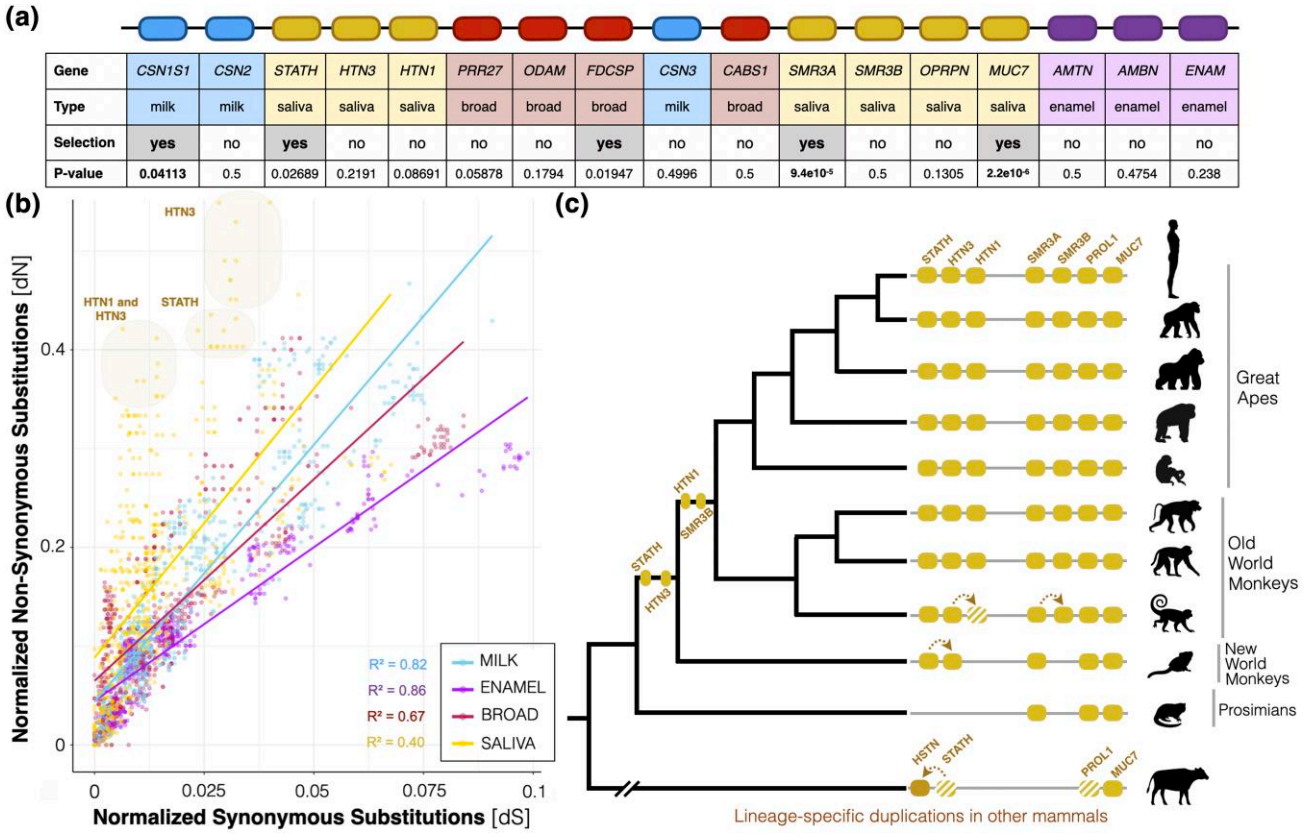

Fig. 4. Saliva-related SCPP genes in primates evolved rapidly and exhibit evidence for positive selection. a) Table summarizing the results of the selection analysis using BUSTED. Colors denote the category of function for each gene. b) Scatterplot of nonsynonymous versus synonymous nucleotide substitutions in enamel-related genes (*ENAM*, *AMBN*, *AMTN*; purple), milk-related casein genes (*CSN1S1*, *CSN2*, *CSN3*; light blue), broadly-expressed genes (*PRR27*, *CABS1*, *ODAM*, *FDCSP;* red), and saliva-related genes (*MUC7*, *SMR3A&B*, *HTN1&3*, *STATH;* yellow). dN and dS values were normalized by total protein length and total transcript length, respectively. Linear regression lines are shown for each gene type with each functional category. Shaded areas highlight genes showing the highest ratio of dN/dS. c) The repertoire of saliva-related SCPP genes in a phylogeny of primate species (from top to bottom: human, chimpanzee, gorilla, orangutan, gibbon, baboon, rhesus macaque, marmoset, and galago). The cow was included for the occurrence of *HSTN*, which resembles histatins but may have originated independently. Hatched boxes indicate database discrepancies in the presence of genes. The emergence of novel genes is indicated at the stem of branches. Arrows show the most likely duplication pathways based on phylogenetic analysis.

their evolutionary origins and histories may provide clues to saliva biology and function in the context of environmental and dietary pressures. We first investigated *SMR3A* and *SMR3B*, along with *PROL1* (also named *OPRPN*), which comprise a group of saliva-related SCPP genes that code for proteins rich in proline. We previously showed that these proteins can provide a foundation for the evolution of mucin function (Pajic et al. 2022). Our analysis of nucleotide sequence homology supports earlier hypotheses (Dickinson and Thiesse 1996) that *SMR3A* and *SMR3B* originated as duplicates of *PROL1*. Examining gene presence and absence (Fig. 1) and conducting phylogenetic analysis (Appendix Fig. S4), we found that *SMR3A* duplicated first from *PROL1*, followed by *SMR3B* duplicating from *SMR3A* (Fig. 4c). We found *SMR3*-like sequences present without apparent relation to phylogeny in 19 of 36 nonprimate placental mammals, approximately half (53%) of the species analyzed. In the primate lineage, however, *SMR3A* and *SMR3B* have become consistently present in catarrhini after the split of Old World and New World monkeys (Fig. 1).

Furthermore, we investigated the origin and evolution of *STATH*, which has been shown to maintain calcium in saliva at a supersaturated state (Hay et al. 1984; Xiao et al. 2015), a function vital for enamel remineralization. *STATH* duplicated from a casein gene, *CSN1S2* (Kawasaki and Weiss 2003; Sire et al. 2007). The presence of *STATH* in the marmoset genome suggests that the gene originated before the divergence of Old World and New World monkeys (Fig. 4c, Appendix Fig. S1). Our analysis identified *STATH*-like sequences scattered across mammals (Fig. 1, Appendix Fig. S3). Such sporadic occurrences of genes can be explained under two possible scenarios: (i) *STATH* has deep ancestral origins, dating back to the common ancestor of placental mammals, with frequent lineage-specific losses or (ii) statherin-like genes arose through

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

lineage-specific duplications in an independent pattern. Resolution between these two scenarios is challenging because the short and similar sequences of the *STATH* genes do not provide the necessary statistical power to resolve phylogenetic connections (Appendix Fig. S4).

STATH has been suggested to have given rise to the histatin genes (*HTN1* and *HTN3*) (Sabatini et al. 1993). Histatin proteins did not retain the calcium-binding function of STATH but were reported to contain zinc- and copper-binding motifs (Puri et al. 2015). Here, we found that *HTN3* is the phylogenetically older version of the histatins since it duplicated from *STATH* as early as in the ancestor of Old and New World monkeys Appendix Fig. S4. A consecutive duplication event in the ancestor of Old World monkeys and apes then gave rise to *HTN1* (Fig. 4c). Similar to *STATH*, we found isolated occurrences of *HTN*-like sequences in squirrel and manatee genomes, likely unrelated to the primate histatins (Fig. 1, Appendix Fig. S4).

Discussion

Gene evolution in the SCPP cluster occurred through three major evolutionary transitions: (i) the formation of a calcified skeleton (Kawasaki et al. 2004), (ii) the emergence of teeth (Kawasaki and Weiss 2008), and (iii) the introduction of lactation in mammals (Kawasaki et al. 2011). Focusing on the P/Q-rich SCPP gene locus, where enamel- and milk-related genes are lined up in the vicinity of saliva-related genes, we found that the saliva-specific SCPP genes exhibit signals of accelerated evolution in the primate lineage, both through gene duplications and amino-acid substitutions. Specifically, several of these genes (*HTN1*, *HTN3*, *STATH*, *SMR3A*, *MUC7*) show (i) lower phyloP conservation scores, (ii) elevated dN/dS ratios, and (iii) evidence of lineage-specific gene duplication in primates. The combined evidence from amino-acid changes and gene turnover patterns highlights the dynamic evolutionary history of saliva-related SCPP genes in primates.

Saliva-related SCPP gene expansions are likely a result of gene duplications from other SCPP gene family members. We propose that the expansion and diversification of saliva-related SCPP genes within the primate lineage may have been driven by species-specific selective pressures. Indeed, dietary change has been discussed as a major adaptive pressure in primates (Tishkoff et al. 2007; Veilleux et al. 2023; Bolognini et al. 2024; Yılmaz et al. 2024). Several lines of evidence support the hypothesis that saliva-related SCPP genes have evolved to accommodate the diverse dietary needs of primates, shaping taste perception (Azen et al. 1990; Melis et al. 2017) and the physical properties of saliva (Pushpass et al. 2019; Thamadilok et al. 2020). In parallel, numerous saliva-related SCPP genes have been identified as key players in binding and interacting with bacteria and other pathogens in saliva (Walz et al. 2009; Heo et al. 2013; Frenkel and Ribbeck 2015; Puri et al. 2015; Thamadilok et al. 2016; Barnard et al. 2020). Within this context, it is conceivable that not only dietary pressures but also an ongoing evolutionary competition between pathogens and saliva-related SCPP proteins partially underlie the observed SCPP genetic variation among primates. While our study focused on primates, we analyzed a broad set of mammalian genomes to establish the ancestral context of the SCPP locus. Although we did not examine potential saliva-related SCPP expansions in nonprimate mammals, similar ecological and pathogenic pressures that shaped the evolution of saliva-related SCPPs in primates may have also influenced SCPP gene repertoires in other lineages, warranting further investigation.

Recognizing that dietary and pathogenic pressures rapidly shaped the variation in saliva-related SCPP genes in primates and, ultimately, humans, it becomes of significant interest to investigate how variations in these genes and recent evolutionary trends have occurred among different human populations with diverse diets and pathogenic exposures. One example is the recent evolution of saliva amylase (Pajic et al. 2019; Bolognini et al. 2024; Yılmaz et al. 2024), which accompanied starch-rich diets and may impact metabolic and microbiome variation. Similarly, variations in the saliva-related SCPP genes *SMR3B* and *SMR3A* were recently shown to significantly affect oral microbiome composition (Kamitaki et al. 2025), with potential downstream effects on disease susceptibility. Characterizing both the genetic variation and posttranslational modifications, particularly glycosylation, of these proteins may offer new avenues for precision therapeutic delivery in disease contexts (Yang et al. 2025).

Beyond starch consumption, broader dietary shifts, including those enabled by the controlled use of fire, likely shaped the human saliva proteome (Thamadilok et al. 2020). In parallel, human-specific genomic changes, particularly those affecting the glycan decoration of saliva mucins and other glycoproteins, may have given rise to a distinct molecular profile of human saliva, with downstream effects on oral microbial colonization (O'Bleness et al. 2012; Cross and Ruhl 2018). A key example in this context is the loss-of-function mutation in the *CMAH* gene, which abolished endogenous synthesis of N-glycolylneuraminic acid (Neu5Gc) and resulted in a monopoly of N-acetylneuraminic acid (Neu5Ac) on human glycoproteins (Tangvoranuntakul et al. 2003; Varki 2022). The dominance of Neu5Ac must have affected the posttranslational properties of SCPPs, likely lending human saliva a unique biochemical signature compared to that of other mammals (Barnard et al. 2020). In this context, our findings here set the stage for transformative research utilizing health-related genomic datasets to identify links between saliva-related SCPP genetic variation and human oral and systemic health or disease.

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

Methods

Gene Tree and BLAST Parameters

Species radiation and divergence parameters used for Fig. 1 were downloaded from the UCSC Genome Browser: http://hgdownload.cse.ucsc.edu/goldenpath/hg19/multiz100way/hg19.100way.nh. NCBI BLAST was used to determine the presence or absence of genes in both long and short read based reference genome assemblies of species. We used these BLAST results along with manual alignments and synteny checks to identify true orthologs. See Supplementary Methods for details. The databases used, genome assemblies examined, and all accession numbers of genes can be found in Table S1.

Gene Expression Analysis

RNA-sequencing data used to construct Fig. 2 are taken from Saitou et al. (2020) for human salivary gland expression, from *Gao* et al. (Gao et al. 2018) for mouse salivary gland expression, and from GTEx (GTEx Consortium 2013) for expression data from multiple tissues. Gene expression data can be found in **Research Data 2**. For further details, see Supplementary Methods.

Conservation Score

Site-by-site, pairwise, conservation scores were downloaded from the UCSC Genome Browser using the 100 Vertebrates Basewise Conservation by PhyloP (phyloP100way) for each SCPP gene (Fig. 3a & 3b) (Pollard et al. 2010). PhyloP scores measure evolutionary conservation at individual nucleotide sites across species. For interpretation of the scores, they are compared to scores expected under neutral drift. Positive scores indicate conservation. Negative scores indicate accelerated evolution. Output scores can be found in **Research Data 1**. Note that not all SCPP genes are present or alignable across the full set of 100 vertebrate species due to lineage-specific gene gains or losses. PhyloP scores are calculated only for aligned nucleotide positions, and missing data in some species can affect the depth of the multiple alignment, particularly for saliva-related genes that have less gene presence across taxa compared to other SCPP gene categories. To mitigate this, our analyses were conducted on a per-base level, using only the available aligned positions per gene. The distributions of per-base PhyloP scores were then compared across functional gene categories using nonparametric Wilcoxon rank-sum tests, enabling robust comparison despite differences in gene length or alignment completeness. This approach emphasizes local sequence constraint rather than gene presence/absence alone.

Selection Analysis

Positive selection analysis was performed using BUSTED (Murrell et al. 2015). Using MEGA X (Stecher et al. 2020), we calculated global dN/dS (the ratio of substitution rates for nonsynonymous and synonymous SNPs). For further details, see the Supplementary Methods.

Phylogenetic Trees and Duplication History

The maximum likelihood trees shown in Appendix Fig. S4 were generated using saliva-related SCPP peptide and coding nucleotide sequences obtained from the Ensembl database for various mammalian species. Sequences were aligned using Clustal Omega in Seaview. We included gaps in our calculations. Maximum likelihood trees were constructed using IQtree 2 (Minh et al. 2020) with 100 bootstrap replicates (Pajic et al. 2016). All sequence alignments can be found in Additional File S1. For further details, see the Supplementary Methods and Appendix Fig. S4 legend.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Vincent Lynch for technical help, discussions, suggestions, and feedback on the manuscript. We thank Benjamin Stein for additional bioinformatic validation of the presence and absence of genes in species' genomes.

Author Contributions

P.P., S.R., and O.G. contributed to conception, design, data acquisition, and interpretation, drafted and critically revised the manuscript. L.L. contributed to data acquisition and interpretation, and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

Funding

This study was funded by NSF grant no. 2049947 (to O.G.) and in part by the National Institute of Dental and Craniofacial Research (NIDCR) grants R01 DE019807 and R21 DE025826 (to S.R.), and National Cancer Institute (NCI) grant U01 CA221244 (to S.R.).

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

All data generated and sequences analyzed can be found in the supplementary files.

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

Literature Cited

Alazem O, Abramyan J. Reptile enamel matrix proteins: selection, divergence, and functional constraint. J Exp Zool B Mol Dev Evol. 2019:332:136–148. https://doi.org/10.1002/jez.b.22857.

Azen EA, Hellekant G, Sabatini LM, Warner TF. mRNAs for PRPs, statherin, and histatins in von Ebner's gland tissues. J Dent Res. 1990:69:1724–1730. https://doi.org/10.1177/00220345900690110401.

Barnard KN, et al. Modified sialic acids on mucus and erythrocytes inhibit influenza A virus hemagglutinin and neuraminidase functions. J Virol. 2020:94:e01567-19. https://doi.org/10.1128/JVI.01567-19.

Bolognini D, et al. Recurrent evolution and selection shape structural diversity at the amylase locus. Nature. 2024:634:617–625. https://doi.org/10.1038/s41586-024-07911-1.

Brawand D, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011:478:343–348. https://doi.org/10.1038/nature10532.

Chan ET, et al. Conservation of core gene expression in vertebrate tissues. J Biol. 2009:8:1–17. https://doi.org/10.1186/jbiol130.

Cross BW, Ruhl S. Glycan recognition at the saliva–oral microbiome interface. Cell Immunol. 2018:333:19–33. https://doi.org/10.1016/j.cellimm.2018.08.008

de Sousa-Pereira P, et al. An evolutionary perspective of mammal salivary peptide families: cystatins, histatins, statherin and PRPs. Arch Oral Biol. 2013:58:451–458. https://doi.org/10.1016/j.archoralbio.2012.12.011.

Dickinson DP, Thiesse M. cDNA cloning of an abundant human lacrimal gland mRNA encoding a novel tear protein. Curr Eye Res. 1996:15:377–386. https://doi.org/10.3109/02713689608995828.

Frenkel ES, Ribbeck K. Salivary mucins in host defense and disease prevention. J Oral Microbiol. 2015:7:29759. https://doi.org/10.3402/jom.v7.29759.

Gao X, Oei MS, Ovitt CE, Sincan M, Melvin JE. Transcriptional profiling reveals gland-specific differential expression in the three major salivary glands of the adult mouse. Physiol Genomics. 2018:50:263–271. https://doi.org/10.1152/physiolgenomics.00124.2017.

George A, Veis A. Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. Chem Rev. 2008:108:4670–4693. https://doi.org/10.1021/cr0782729.

GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet. 2013:45:580–585. https://doi.org/10.1038/ng.2653.

Hay DI, Smith DJ, Schluckebier SK, Moreno EC. Relationship between concentration of human salivary statherin and inhibition of calcium phosphate precipitation in stimulated human parotid saliva. J Dent Res. 1984:63:857–863. https://doi.org/10.1177/00220345840630060901.

Heo SM, et al. Host defense proteins derived from human saliva bind to *Staphylococcus aureus*. Infect Immun. 2013:81:1364–1437. https://doi.org/10.1128/iai.00825-12.

Kamitaki N, et al. 2025 Apr 1. Human and bacterial genetic variation shape oral microbiomes and health [preprint]. medRxiv. 25324952. https://doi.org/10.1101/2025.03.31.25324952

Kawasaki K. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. Dev Genes Evol. 2009:219:147–157. https://doi.org/10.1007/s00427-009-0276-x.

Kawasaki K, Amemiya CT. SCPP genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. J Exp Zool. 2014:322:390–402. https://doi.org/10.1002/jez.b.22546.

Kawasaki K, Buchanan AV, Weiss KM. Gene duplication and the evolution of vertebrate skeletal mineralization. Cells Tissues Organs. 2007:186:7–24. https://doi.org/10.1159/000102678.

Kawasaki K, Buchanan AV, Weiss KM. Biomineralization in humans: making the hard choices in life. Annu Rev Genet. 2009:43:119–142. https://doi.org/10.1146/annurev-genet-102108-134242.

Kawasaki K, Lafont A-G, Sire J-Y. The evolution of milk casein genes from tooth genes before the origin of mammals. Mol Biol Evol. 2011:28:2053–2061. https://doi.org/10.1093/molbev/msr020.

Kawasaki K, Suzuki T, Weiss KM. Genetic basis for the evolution of vertebrate mineralized tissue. Proc Natl Acad Sci U S A. 2004:101:11356–11361. https://doi.org/10.1073/pnas.0404279101.

Kawasaki K, Suzuki T, Weiss KM. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. Proc Natl Acad Sci U S A. 2005:102:18063–18068. https://doi.org/10.1073/pnas.0509263102.

Kawasaki K, Weiss KM. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. Proc Natl Acad Sci U S A. 2003:100:4060–4065. https://doi.org/10.1073/pnas.0638023100.

Kawasaki K, Weiss KM. SCPP gene evolution and the dental mineralization continuum. J Dent Res. 2008:87:520–531. https://doi.org/10.1177/154405910808700608.

Lau WW, Hardt M, Zhang YH, Freire M, Ruhl S. The human salivary proteome wiki: a community-driven research platform. J Dent Res. 2021:100:1510–1519. https://doi.org/10.1177/00220345211014432.

Melis M, et al. Sensory perception of and salivary protein response to astringency as a function of the 6-n-propylthioural (PROP) bitter-taste phenotype. Physiol Behav. 2017:173:163–173. https://doi.org/10.1016/j.physbeh.2017.01.031.

Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:37:1530–3064. https://doi.org/10.1093/molbev/msaa015.

Murrell B, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015:32:1365–1371. https://doi.org/10.1093/molbev/msv035.

O'Bleness M, Searles VB, Varki A, Gagneux P, Sikela JM. Evolution of genetic and genomic features unique to the human lineage. Nat Rev Genet. 2012:13:853–866. https://doi.org/10.1038/nrg3336.

Pajic P, et al. Independent amylase gene copy number bursts correlate with dietary preferences in mammals. Elife. 2019:8:e44628. https://doi.org/10.7554/eLife.44628.

Pajic P, et al. A mechanism of gene evolution generating mucin function. Sci Adv. 2022:8:eabm8757. https://doi.org/10.1126/sciadv.abm8757.

Pajic P, Lin Y-L, Xu D, Gokcumen O. The psoriasis-associated deletion of late cornified envelope genes LCE3B and LCE3C has been maintained under balancing selection since human denisovan divergence. BMC Evol Biol. 2016:16:265. https://doi.org/10.1186/s12862-016-0842-6.

Pajic PM. Mucins provide insights into mechanisms of functional evolution. 2025. https://search.proquest.com/openview/1ffb7e89c47fde5d95fda45b10e26bbd/1?pq-origsite=gscholar&cbl=18750&diss=y

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010:20:110–121. https://doi.org/10.1101/gr.097857.109.

Ponting CP. Biological function in the twilight zone of sequence conservation. BMC Biol. 2017:15:71–NaN. https://doi.org/10.1186/s12915-017-0411-5.

Puri S, Li R, Ruszaj D, Tati S, Edgerton M. Iron binding modulates candidacidal properties of salivary histatin 5. J Dent Res. 2015:94:201–208. https://doi.org/10.1177/0022034514556709.

Pushpass R-AG, Daly B, Kelly C, Proctor G, Carpenter GH. Altered salivary flow, protein composition, and rheology following taste and

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026

TRP stimulation in older adults. Front Physiol. 2019:10:652. https://doi.org/10.3389/fphys.2019.00652.
Sabatini LM, Ota T, Azen EA. Nucleotide sequence analysis of the human salivary protein genes HIS1 and HIS2, and evolution of the STATH/HIS gene family. Mol Biol Evol. 1993:10:497–511. https://doi.org/10.1093/oxfordjournals.molbev.a040022.
Saitou M, et al. Functional specialization of human salivary glands and origins of proteins intrinsic to human Saliva. Cell Rep. 2020:33:108402. https://doi.org/10.1016/j.celrep.2020.108402.
Sire J-Y, Davit-Béal T, Delgado S, Gu X. The origin and evolution of enamel mineralization genes. Cells Tissues Organs. 2007:186:25–48. https://doi.org/10.1159/000102679.
Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evol. 2020:37:1237–1239. https://doi.org/10.1093/molbev/msz312.
Tangvoranuntakul P, et al. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. Proc Natl Acad Sci U S A. 2003:100:12045–12095. https://doi.org/10.1073/pnas.2131556100.
Thamadilok S, et al. Human and nonhuman primate lineage-specific footprints in the salivary proteome. Mol Biol Evol. 2020:37:395–405. https://doi.org/10.1093/molbev/msz223.
Thamadilok S, Roche-Håkansson H, Håkansson AP, Ruhl S. Absence of capsule reveals glycan-mediated binding and recognition of salivary mucin MUC7 by *Streptococcus pneumoniae*. Mol Oral Microbiol. 2016:31:175–263. https://doi.org/10.1111/omi.12113.
Tishkoff SA, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet. 2007:39:31–40. https://doi.org/10.1038/ng1946.
Varki A. Evolution of glycan diversity. In: Varki A, et al., editors. Essentials of glycobiology. 4th ed. 2022. Cold Spring Harbor Laboratory Press; p. 265–278. Chapter 20.
Veilleux CC, et al. Human subsistence and signatures of selection on chemosensory genes. Commun Biol. 2023:6:683. https://doi.org/10.1038/s42003-023-05047-y.
Walz A, et al. Identification of glycoprotein receptors within the human salivary proteome for the lectin-like BabA and SabA adhesins of *Helicobacter pylori* by fluorescence-based 2-D bacterial overlay. Proteomics. 2009:9:1582–1674. https://doi.org/10.1002/pmic.200700808.
Xiao Y, et al. Hydroxyapatite growth inhibition effect of pellicle statherin peptides. J Dent Res. 2015:94:1106–1112. https://doi.org/10.1177/0022034515586769.
Xu D, et al. Recent evolution of the salivary mucin MUC7. Sci Rep. 2016:6:31791. https://doi.org/10.1038/srep31791.
Yang Q, et al. Tuning the tropism and infectivity of SARS-CoV-2 virus-like particles for mRNA delivery. Nucleic Acids Res. 2025:53:gkaf133. https://doi.org/10.1093/nar/gkaf133.
Yılmaz F, et al. Reconstruction of the human amylase locus reveals ancient duplications seeding modern-day variation. Science. 2024:386:eadn0609. https://doi.org/10.1126/science.adn0609.
Zhou Y, et al. Platypus and echidna genomes reveal mammalian biology and evolution. Nature. 2021:592:756–762. https://doi.org/10.1038/s41586-020-03039-0.

Associate editor: Mashaal Sohail

Downloaded from https://academic.oup.com/gbe/article/17/9/evaf165/8239972 by University at Buffalo user on 03 April 2026