Analysis of the Human AMELX Gene

Thais Gallo Nisenbaum[1,*]

[1] Biotechnology, Johns Hopkins University, Baltimore, MD, 21218, USA

* To whom correspondence should be addressed.

Name: Thais Gallo Nisenbaum
Email: tgallon1@jhu.edu

ABSTRACT

The *Homo sapiens* Amelogenin X-linked (AMELX) gene was analyzed for this assignment. Biological databases such as NCBI Entrez Gene, Ensembl, and UCSC Genome Browser were used to identify all known transcripts, coding exons, and genetic variations of this gene. Genetic variations, including single nucleotide polymorphisms and copy number variations, were further analyzed to determine their pathological and phenotypical effects. AMELX has been conserved across several species, and orthologs and paralogs were observed in Ensembl. Twelve orthologs were selected to be aligned with the *Homo sapiens* AMELX gene using MAFFT, and a phylogenetic tree was generated to demonstrate the molecular evolution and evolutionary divergence of the AMELX gene. Furthermore, an additional ortholog was identified in cyanobacteria using NCBI Blastn, emphasizing the importance of this gene and its function. Investigation of the AMELX gene in the NCBI Gene Expression Omnibus (GEO) database demonstrated the expression levels of this gene in response to vitamin D.

INTRODUCTION

Amelogenin X-linked (AMELX) is a protein-coding gene located in the forward strand of the X-chromosome from position 11,293,413 to 11,300,761. The AMELX gene is located in intron 1 of Rho GTPase Activating Protein 6 (ARHGAP6) (1). AMELX is flanked by Holocytochrome C Synthase (HCCS), located from position 11,111,332 to 11,123,086, and microRNA 548ax (MIR548AX), located from position 11,318,614 to 11,318,686 in the complement strand. AMELX encodes for the cell-adhesion protein amelogenin, which is responsible for the biomineralization of tooth enamel. Amelogenins are secreted by ameloblasts and are localized in the extracellular matrix (1). These proteins are conserved across species, which demonstrates their important function.

A PubMed search of the *Homo sapiens* AMELX gene yields 206 results consisting of various conditions impacting teeth, including Amelogenesis imperfecta, dental fluorosis, and dental caries. The gene ID for AMELX in the NCBI Entrez Gene database is 265. A PubMed search specifically for this ID yields 63 results, which include research on several mutations of the gene and its link to several pathogenic conditions, including severe diseases such as odontogenic tumors. Overall, there is great interest in the pathological manifestation of AMELX mutations and how they can be used as potential biomarkers.

MATERIAL AND METHODS

I. National Center for Biotechnology Information (NCBI) Databases

a. NCBI Entrez Gene:

NCBI Entrez Gene was used to determine known transcripts of AMELX according to the transcript table for the *Homo sapiens* AMELX entry, gene ID 265 (2). The gene entry and transcripts can be found in the data availability sections 1.1 and 1.3, respectively.

b. NCBI BLASTN:

An ortholog of the *Homo sapiens* AMELX gene, not included in the Ensembl database, was identified using NCBI nucleotide BLAST (3). The mRNA sequence of transcript variant 1 of the *Homo sapiens* AMELX gene, with the accession ID NM_001142.2, was used as the query. The results were limited to the nucleotide collection (nt) database and included only prokaryotes (prokaryotes (taxid:2157), prokaryotes (taxid:2)). The default scoring parameters were used, including a match score of 2, a mismatch score of -3, and gap cost with existence of 5 and extension of 2. The closest ortholog was determined according to the lowest E-value. The transcript variant and the identified prokaryote ortholog can be found in the data availability sections 1.4 and 1.5, respectively.

c. NCBI Variation Viewer:

NCBI Variation Viewer was used to determine the number of pathogenic copy number alterations (CNVs) and SNPs in the AMELX gene and their associated phenotypes.

d. NCBI Gene Expression Omnibus (GEO):

NCBI GEO was used to identify an experiment involving the *Homo sapiens* AMELX gene with regard to levels of its expression in different conditions.

II. Ensembl Genome Browser:

Ensembl genome browser release 108 was used to determine known transcripts, orthologs, and paralogs of AMELX according to the *Homo sapiens* AMELX entry with the accession ID ENSG00000125363 in the Ensembl database using the human genome assembly GRCh38.p13. The gene entry and transcripts can be found in the data availability section 2.1.

III. UCSC Genome Browser:

UCSC Genome Browser was used to determine known transcripts, the number of genetic variants, and the number of coding exons for the AMELX gene.

The known transcripts were determined according to the *Homo sapiens* AMELX gene located from position 11,293,413 to 11,300,761 in the X chromosome, using the Human Genome Browser with assembly ID hg38 initially released in December 2013 (Dec. 2013 (GRCh38/hg38)). Transcripts were noted according to the GENECODE V41 and the UCSC RefSeq Annotations track. The number of coding exons was noted for each transcript according to the GENCODE V41 annotations. Transcript information can be found in the data availability section 3.1.

The number of genetic variants was determined using the Table Browser tool set to the "Variation" group and dbSNP 153 track. The dbSNP is the single nucleotide polymorphism (SNP) consisting of 702 million genetic variants of around 50 base pairs in length. Genetic variations in this database include single-nucleotide variants (SNVs), insertions, and deletions. The results of both common and ClinVar SNPs were obtained (4). The number of coding exons was determined using the Table Browser tool set to the "Gene and Gene Predictions" group and the GENCODE V41 track, generating one browser extensible data (BED) record per coding exons.

Additionally, the GENSCAN track was used to compare the predicted to known gene models. GENSCAN uses a Generalized Hidden Markov Model (GHMM) algorithm to identify exon and intron structures to make gene predictions, taking gene density and unique CG regions into account (5).

IV. Galaxy

Galaxy is an analysis framework with the ability to integrate with several databases (6). Results generated utilizing the UCSC Genome Browser Table Browser tool were sent to and visualized in Galaxy.

V. MAFFT (Multiple Alignment using Fast Fourier Transform)

Twelve orthologs from Ensemble Genome Browser were selected and aligned with the *Homo sapiens* AMELX gene transcript variant 1. The FASTA sequences for the selected organisms were entered as the input in MAFFT, and the parameters were set to default values. Multiple sequence alignment in MAFFT includes progressive and iterative methods (7). A UPGMA phylogeny tree was generated based on the results from the multiple sequence alignment. The phylogenetic tree was visualized in *Phylo.io* 1.0.0, a web-based software used for viewing and scaling trees (8). The selected orthologs can be found in the data availability section 2.2.

RESULTS

I. AMELX transcripts according to three biological databases

a. *NCBI Entrez Gene:* According to NCBI Entrez Gene, AMELX has three known and one predicted transcript.

Transcript	Length (nt)	Protein	Length (aa)	Isoform
Transcript Variant 3 (NM_182680.1)	835	NP_872621.1	205	Amelogenin, X isoform 3
Transcript Variant 1 (NM_001142.2)	793	NP_001133.1	191	Amelogenin, X isoform 1 precursor
Transcript Variant 2 (NM_182681.1)	745	NP_872622.1	175	Amelogenin, X isoform 2
Predicted Transcript Variant X1 (XM_017029404.3)	1035	XP_016884893.2	344	Amelogenin, X isoform X1

Table 1

b. *Ensembl*: According to Ensembl, AMELX has three known transcripts.

Transcript	Length (nt)	Protein (UniProt)	Total Exons	Coding Exons	Length (aa)	Isoform
AMELX-202 (ENST00000380712.7)	835	Q99217-3	7	6	205	Isoform 3 of Amelogenin, X isoform
AMELX-203 (ENST00000380714.7)	793	Q99217	6	5	191	Amelogenin, X isoform
AMELX-201 (ENST00000348912.4)	745	Q99217-2	5	4	175	Isoform 2 of Amelogenin, X isoform

Table 2

c. *UCSC Genome Browser*: According to the UCSC Genome Browser, using the GENCODE V41 and UCSC RefSeq annotation trackers, AMELX has three known transcripts.

Transcript	Length (nt)	Transcript Position	Coding Region	Length (aa)	Isoform
Transcript Variant 3 (RefSeq NM_182680) (ENST00000380712.7)	835	hg38 chrX: 11,293,413-11,300,761 (+ strand) Total Exons: 7	hg38 chrX: 11,294,789-11,300,612 Coding Exons: 6	205	Amelogenin, X isoform 3
Transcript Variant 1 (RefSeq NM_001142) (ENST00000380714.7)	793	hg38 chrX: 11,293,413-11,300,761 (+ strand) Total Exons: 6	hg38 chrX: 11,294,789-11,300,612 Coding Exons: 5	191	Amelogenin, X isoform 1 precursor
Transcript Variant 2 (RefSeq NM_182681) (ENST00000348912.4)	745	hg38 chrX: 11,293,413-11,300,753 (+ strand) Total Exons: 5	hg38 chrX: 11,294,789-11,300,612 Coding Exons: 4	175	Amelogenin, X isoform 2

Table 3

II. Genetic variations and coding regions

According to the dbSNP153 track, there are 19 common SNPs and 15 ClinVar SNPs in the AMELX gene. Common SNPs include variants with a minor allele frequency of at least 1% in phase 3 of the 1000 Genomes dataset. ClinVar SNPs include variants that are mentioned in ClinVar (4). The output from the GENCODE V41 track resulted in 63 lines consisting of each gene transcript. Duplicate results were removed, as some exons are located in multiple transcripts and were listed more than once. The resulting table yielded 22 unique values, with six exons located in the positive strand and 16 located in the negative strand.

The AMELX gene contains two non-synonymous SNPs: dbSNP rs749213434 A/C missense variant and dbSNP rs2106416 C/A/T missense variant. rs749213434 is a common SNP, and rs2106416 is both a common and ClinVar SNP. Both of these variations are located in coding exons.

III. Orthologs and paralogs of the human AMELX gene

According to the Ensembl database, the *Homo sapiens* AMELX gene contains 108 orthologs and 1 paralog. Orthologs include primates, rodents, placental mammals, birds, and reptiles. A multiple sequence alignment was performed, resulting in 290 conserved sites across the sequences. A UPGMA phylogeny tree was generated for twelve selected orthologs, demonstrating the evolutionary path of the AMELX gene.

As depicted by the tree, the black snub-nosed monkey and the chimpanzee are the closest related to the human AMELX gene, while the Australian saltwater crocodile is the most distant.

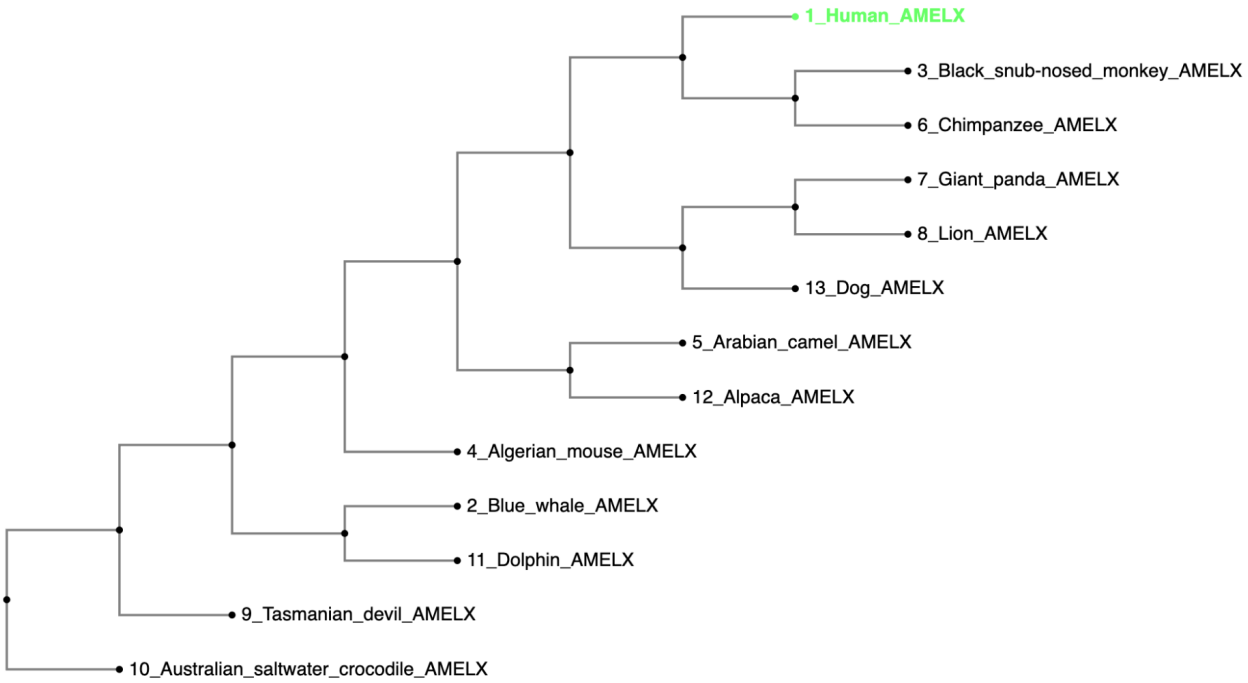

Figure 1

A BLASTN search limited to prokaryotes identified additional orthologs of the *Homo sapiens* AMELX gene. There were a total of 11 hits; the sequence closest related to the query was *Nostoc sphaeroides* CCNUC1 chromosome Gxm1, a strain of cyanobacteria, with an E-value of 2e-06 and 74% identity.

The paralog includes the amelogenin Y-linked (AMELY) Homo *sapiens* gene, with frogs and toads as the most recent common ancestor. The AMELY gene is located on the reserve strand of chromosome Y at positions 6,865,918 to 6,911,752. The AMELY gene codes for the amelogenin protein but only produces a small amount. Nearly all of the amelogenin in the body is made from the AMELX gene. The AMELY gene is not required for the biomineralization of tooth enamel (9).

IV. GENSCAN gene prediction tool and known gene models

The GENSCAN algorithm predicted a total of 4 exons, while the known gene models include 6 exons, 5 of which are coding. Only one of the exons predicted by GENSCAN (exon 3) matched the exact position of an exon on the known gene model (exon 4). The predicted mRNA and peptide sequences are compared to the known *Homo sapiens* amelogenin X-linked (AMELX) transcript variant 1 in the table below. Nucleotides and amino acid residues are colored green representing identical matches.

GENSCAN	**AMELX Known Gene Model**
Predicted mRNA sequence **999 bp** ATGATTCCAGTCCTTTATAAGGAAGATACTTATCAGAGGTCATT TGCACAGAAGTCCACAGGAGCAGTTGAGAGCAATGCTGAATT ATCCATCTTAGCTATTGAAAAGTACCCTTCCTATGGTTACGAGC CCATGGGTGGATGGCTGCACCACCAAATCATCCCCGTGCTGT CCCAACAGCACCCCCCGACTCACACCCTGCAGCCTCATCAC CACATCCCAGTGGTGCCAGCTCAGCAGCCCGTGATCCCCCA GCAACCAATGATGCCCGTTCCTGGCCAACACTCCATGACTCC AATCCAACACCACCAGCCAAACCTCCCTCCGCCCGCCCAGC AGCCCTACCAGCCCCAGCCTGTTCAGCCACAGCCTCACCAG CCCATGCAGCCCCAGCCACCTGTGCACCCCATGCAGCCCCT GCCGCCACAGCCACCTCTGCCTCCGATGTTCCCCATGCAGC CCCTGCCTCCCATGCTTCCTGATCTGACTCTGGAAGCTTGGC CATCAACAGACAAGACCAAGCGGGAGGAAGTGATGCTGCTCT GGGATTATTACCCCATGGAATCTTTCACGTCTTCCTGGCTTGG GTCCTACCTTGCTTCTAGCAGTAGGCACGAAACACAGAGCCT GAAATACACACACCGTATCTTCTTACCTCTCCAGGGTTCTTAC ACCTCAGCAATAGCAGCAGCTCCAGGCAGTAGCGAGCCCCC ATACTCCCCACGCAGAGCCCCTGTGCAGTGTGGGAACAGAG AAGGAGACAGGAGGCAGGGAGGGACACTGCTCAAAGTGCTA GAGGAGCCGGCCAGGCAGTCTCCCTGTCTGGAATGCACAAC TAGAGGATCTCGGAGGGAGAGAGCCTTGGGTCACTGTGATC CTCTGAGTCAGACACCCTGGGGAGCTCTTAGTTCTGAAAGTC CTCCCCAGGCACAGAAGGCTGAAAGCCCCCAAGCACCCACT AGGAAGAAAAGTGTAAAAGTTAAAAAGTTAAAAAGTTGA	**Homo sapiens amelogenin X-linked (AMELX), transcript variant 1, mRNA** **793 bp** AAAGGATCAAGCATCCCTGAGTTTCAAACAGAAACTTGCACTG AATACATTCAAAGAACCATCAAGAAATGGGGACCTGGATTTTAT TTGCCTGCCTCCTGGGAGCAGCTTTTGCCATGCCTCTACCACC TCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTG CTTACCCCTTTGAAGTGGTACCAGAGCATAAGGCCACCGTACC CTTCCTATGGTTACGAGCCCATGGGTGGATGGCTGCACCACCA AATCATCCCCGTGCTGTCCCAACAGCACCCCCCGACTCACAC CCTGCAGCCTCATCACCACATCCCAGTGGTGCCAGCTCAGCA GCCCGTGATCCCCCAGCAACCAATGATGCCCGTTCCTGGCCA ACACTCCATGACTCCAATCCAACACCACCAGCCAAACCTCCCT CCGCCCGCCCAGCAGCCCTACCAGCCCCAGCCTGTTCAGCC ACAGCCTCACCAGCCCATGCAGCCCCAGCCACCTGTGCACCC CATGCAGCCCCTGCCGCCACAGCCACCTCTGCCTCCGATGTT CCCCATGCAGCCCCTGCCTCCCATGCTTCCTGATCTGACTCTG GAAGCTTGGCCATCAACAGACAAGACCAAGCGGGAGGAAGTG GATTAAAAGATCAGAAGATGAGAGGGGAATGAATACTTCAGAT GCTTTCAGGAGTGACACAAGAACACAATGATTTTTGCTTATAAT CACTTTACTTAGCAAATTCTGTAACTAAAAAAGTACCATTAGCAG ACAATAAAATGCATTAAAAATCA
Predicted protein **332 aa** MIPVLYKEDTYQRSFAQKSTGAVESNAELSILAIEKYPSYGYEPM GGWLHHQIIPVLSQQHPPTHTLQPHHHIPVVPAQQPVIPQQPMM PVPGQHSMTPIQHHQPNLPPPAQQPYQPQPVQPQPHQPMQPQ PPVHPMQPLPPQPPLPPMFPMQPLPPMLPDLTLEAWPSTDKTK REEVMLLWDYYPMESFTSSWLGSYLASSSRHETQSLKYTHRIFL PLQGSYTSAIAAAPGSSEPPYSPRRAPVQCGNREGDRRQGGTL LKVLEEPARQSPCLECTTRGSRRERALGHCDPLSQTPWGALSS ESPPQAQKAESPQAPTRKKSVKVKKLKS	**Amelogenin, X isoform isoform 1 precursor [Homo sapiens]** **191 aa** MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLKWYQSI RPPYPSYGYEPMGGWLHHQIIPVLSQQHPPTHTLQPHHHIPVVP AQQPVIPQQPMMPVPGQHSMTPIQHHQPNLPPPAQQPYQPQPV QPQPHQPMQPQPPVHPMQPLPPQPPLPPMFPMQPLPPMLPDLT LEAWPSTDKTKREEVD

Table 4

Overall, GENSCAN predicted longer but similar mRNA and peptide sequences. The predicted mRNA sequence and transcript variant 1 have 427 identical base pairs. The predicted peptide sequence and isoform 1 have 142 identical amino acid residues.

V. Pathogenic variations of the human AMELX gene

According to the Variation Viewer in NCBI, there are 454 copy number variations (CNVs) in the *Homo sapiens* AMELX gene, out of which 264 are pathogenic. Phenotypes from pathogenic CNVs and SNPs include Amelogenesis imperfecta, type 1E.

Amelogenesis imperfecta (AI) refers to the underdevelopment of tooth enamel, resulting in small, discolored, and pitted teeth. Mutations in the AMELX gene resulting in this condition follow an X-linked dominant pattern of inheritance and account for 5 percent of AI cases. This condition is more severe in males than females, as males would only have one copy of the X chromosome (10).

VI. Effects of vitamin D on AMELX expression

A search of the *Homo sapiens* AMELX gene yields in NCBI GEO profiles yields 1732 results. GEO Profiles is a database consisting of several experiments that measure gene expression under various conditions (11). The selected experiment measures the effects of vitamin D on gene expression by intestinal epithelial cell line of human colorectal adenocarcinoma cells (Caco-2 cells). The experiment included 10 samples, out of which 5 were controls, belonging to the GSE444 series and the [HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array platform, and was published on October 26th, 2004. The researchers incubated experimental Caco-2 cells with 10^{-7} mol/l of 1,25-dihydroxyvitamin D_3 for 24 hours, and gene expression was measured with microarray analysis (12). According to the GEO profile graph, referenced in the data availability section 1.6, AMELX gene expression appears to be higher in response to the addition of vitamin D.

DISCUSSION

The analysis of the *Homo sapiens* AMELX gene demonstrated the molecular evolution and divergence of this gene, highlighting its conservation across several species, including a prokaryote, which was not included in Ensembl. All known transcripts, coding exons, and genetic variation were observed in biological databases, and further analysis demonstrated the pathological effects of mutations occurring in this gene. The most common disease associated with mutations in the AMELX gene is Amelogenesis imperfecta, type 1E. The Amelogenin gene is present in both sex chromosomes, and AMELX and AMELY are paralogues of one another. However, only the AMELX gene is required for the biomineralization of tooth enamel, as it is responsible for making nearly all amelogenin proteins in the body. This explains why Amelogenesis imperfecta is usually more severe in males than females.

The *Homo sapiens* amelogenin X-linked (AMELX) transcript variant 1 was compared with mRNA and peptide sequences predicted by GENSCAN. Predicted sequences were similar but longer than the known gene model. As referenced in section I, the transcript variant predicted by NCBI

(XP_016884893.2, Amelogenin, X isoform X1) is also longer than the known transcripts, with an mRNA sequence of 1035 nucleotides and a peptide sequence of 344 amino acid residues. Interestingly, the c-terminus of isoform X1 is identical to the c-terminus of the protein sequence predicted by GENSCAN. According to an experiment in the NCBI GEO database, AMELX expressions are higher in response to vitamin D when compared to controls. This experiment is relevant to this gene because Vitamin D plays a major role in calcium and phosphate regulation, which are the components of enamel. Overall, this gene is highly conserved and plays a critical role among several species.

DATA AVAILABILITY

1. National Center for Biotechnology Information: https://www.ncbi.nlm.nih.gov/
    - 1.1. AMELX amelogenin X-linked [Homo sapiens (human)] from the NCBI Gene Database (AMELX)
    - 1.2. Flanking Genes: HCCS holocytochrome c synthase [Homo sapiens (human)] (HCCS), MIR548AX microRNA 548ax [Homo sapiens (human)] (MIR548AX)
    - 1.3. Homo sapiens chromosome X, GRCh38.p14 Primary Assembly (NC_000023.11)
    - 1.4. NCBI known transcripts: Homo sapiens amelogenin X-linked (AMELX), transcript variant 3, mRNA (NM_182680.1), Homo sapiens amelogenin X-linked (AMELX), transcript variant 1, mRNA (NM_001142.2), Homo sapiens amelogenin X-linked (AMELX), transcript variant 2, mRNA (NM_182681.1), Predicted transcript: PREDICTED: Homo sapiens amelogenin X-linked (AMELX), transcript variant X1, mRNA (XM_017029404.3)
    - 1.5. Prokaryote Ortholog: *Nostoc sphaeroides CCNUC1* chromosome Gxm1, complete sequence (CP045226.1)
    - 1.6. Vitamin D effect on intestinal epithelial cells GEO Profile graph (GDS1847 / 34588_i_at)

2. Ensembl Genome Browser 108: https://useast.ensembl.org/index.html
    - 2.1. Ensembl transcripts: ENST00000380714.7 AMELX-203 (ENST00000380714.7), ENST00000380712.7 AMELX-202 (ENST00000380712.7), ENST00000348912.4 AMELX-201 (ENST00000348912.4)
    - 2.2. Selected orthologs: Algerian mouse AMELX gene (MGP_SPRETEiJ_G0034752), Alpaca AMELX gene (ENSVPAG00000008563), Arabian camel AMELX gene (ENSCDRG00005008628), Australian saltwater crocodile AMELX gene (ENSCPRG00005006432), Black snub-nosed monkey AMELX gene (ENSRBIG00000040414), Blue whale AMELX gene (ENSBMSG00010016602), Chimpanzee AMELX gene (ENSPTRG00000021659), Dog AMELX gene (ENSCAFG00845026844), Dolphin AMELX gene (ENSTTRG00000005725), Giant panda AMELX gene (ENSAMEG00000010171), Lion AMELX gene (ENSPLOG00000020208), Tasmanian devil AMELX gene (ENSSHAG00000009398).
    - 2.3. Ensembl paralog: Human AMELY gene (ENSG00000099721)

3. UCSC Genome Browser: https://genome.ucsc.edu/
    - 3.1. UCSC Genome Browser transcripts: *GENCODE V41:* Human Gene AMELX (ENST00000348912.4) from GENCODE V41 (ENST00000348912.4), Human Gene AMELX (ENST00000380712.7) from GENCODE V41 (ENST00000380712.7), Human Gene AMELX (ENST00000380714.7) from GENCODE V41 (ENST00000380714.7) *UCSC RefSeq Annotations:* RefSeq Gene AMELX, transcript

variant 1 (Transcript variant 1), RefSeq Gene AMELX, transcript variant 2 (Transcript variant 2), RefSeq Gene AMELX, transcript variant 3 (Transcript variant 3)

4. NCBI BLASTN: https://blast.ncbi.nlm.nih.gov/Blast.cgi
5. Galaxy: https://usegalaxy.org/
6. MAFFT: https://mafft.cbrc.jp/alignment/server/
7. *Phylo.io*: http://phylo.io

ACKNOWLEDGEMENT

I would like to acknowledge Dr. Jonathon Bennett for his time and effort in responding to my questions.

FUNDING

No funding to report.

CONFLICT OF INTEREST

No conflicts of interest to report.

REFERENCES

1. Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: {300391}. [Updated: January 14th, 2022]. Retrieved from https://omim.org/entry/300391
2. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. Nucleic acids research, 50(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
4. University of California, Sanata Cruz (UCSC). (2018, April 19). Short Genetic Variants from dbSNP release 153. Retrieved from https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg38&g=dbSnp153Composite
5. Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. JMB, 268(1), 78–94. https://doi.org/10.1006/jmbi.1997.0951)
6. Blankenberg, D., Coraor, N., Von Kuster, G., Taylor, J., Nekrutenko, A., & Galaxy Team (2011). Integrating diverse databases into an unified analysis framework: a Galaxy approach. Database : the journal of biological databases and curation, 2011, bar011. https://doi.org/10.1093/database/bar011
7. Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics, 20(4), 1160–1166. https://doi.org/10.1093/bib/bbx108
8. Robinson, O., Dylus, D., & Dessimoz, C. (2016). Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. Molecular biology and evolution, 33(8), 2163–2166. https://doi.org/10.1093/molbev/msw080
9. U.S. National Library of Medicine. (n.d.). *Amelx Gene: Medlineplus genetics*. MedlinePlus. [Updated: May 1, 2015]. Retrieved from https://medlineplus.gov/genetics/gene/amelx
10. U.S. National Library of Medicine. (n.d.). *Amelogenesis imperfecta: Medlineplus genetics*. MedlinePlus. [Updated: May 1, 2015]. Retrieved from https://medlineplus.gov/genetics/condition/amelogenesis-imperfecta/

11. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic acids research, 41(Database issue), D991–D995. https://doi.org/10.1093/nar/gks1193
12. Wood, R. J., Tchack, L., Angelo, G., Pratt, R. E., & Sonna, L. A. (2004). DNA microarray analysis of vitamin D-induced gene expression in a human colon carcinoma cell line. Physiological genomics, 17(2), 122–129. https://doi.org/10.1152/physiolgenomics.00002.2003

TABLE AND FIGURES LEGENDS

Table 1. NCBI Entrez Gene AMELX Transcripts

Information for known AMELX transcripts according to the NCBI Entrez Gene database.

Table 2. Ensembl AMELX Transcripts

Information for known AMELX transcripts according to the Ensembl Genome Browser.

Table 3. UCSC Genome Browser AMELX Transcripts

Information for known AMELX transcripts according to the UCSC Genome Browser.

Table 4. Comparison of GENSCAN predicted mRNA and peptide sequences with known gene models

The sequences predicted by GENSCAN were compared with *Homo sapiens* amelogenin X-linked (AMELX) transcript variant 1. Identical nucleotides and amino acids are colored green.

Figure 1. UPGMA Phylogeny Tree

UPMA Phylogeny tree generated in MAFFT and visualized with Phylo.io 1.0.0 following multiple sequence alignment. A total of 13 sequences were analyzed, including the human AMEX transcript 1 and 12 selected orthologs obtained from the Ensembl database. The dataset contained 741 total sites, 309 gap-free sites, and 290 conserved sites.