

L-gulonolactone Oxidase (GULO) Evolution: A Bioinformatic Exploration of L-gulonolactone Oxidase (GULO) protein in mammals

Hameed Sanusi

Department of Biology, Georgia State University

Advisor: Dr. Eric Gaucher

Committee: Dr. Eric Gaucher, Dr. Jessica Carter

April 19th, 2024

Table of Contents

List of Figures and Tables.....	4
Acknowledgments.....	5
Introduction.....	6
1.1 Importance of Vitamin C	6
1.2 Vitamin C Synthesis and L-gulonolactone Oxidase (GULO)	7
1.3 Evolutionary Loss of GULO in Mammals.....	9
1.4 Bioinformatics analysis of GULO mutation	11
1.5 Research Objectives	13
1.6 Hypothesis	13
Materials and Methods.....	14
2.1 GULO Protein Sequence Collection	14
2.2 Multiple Sequence Alignment	14
2.3 Phylogenetic Analysis of GULO sequences	15
2.4 Reconstruction of Ancestral Sequences	15
Results and Discussions	14
3.1 Phylogenetic Analysis	1
3.2 Reconstruction of Ancestral Sequences with Amino Acid Replacement	19
Discussion and Conclusion	21

References

List of Figures and Tables

Figure 1. A pathway showing the biosynthesis of ascorbic acid (Vitamin C)	8
Table 1: Parameters used for the phylogenetic analysis using PAML.....	13
Figure 2: A collapsed phylogram showing the evolutionary relationship of several mammalian species GULO protein sequences.....	15
Figure 3: A section of the phylogram showing the node label for the mammalian species belonging to the <i>Strepsirrhini</i> order.....	16
Table 2: Summary of amino acid changes along node 194 to 195 with posterior probabilities.....	17
Supplemental Table 1: Accession number for 168 GULO protein sequences collected with accession number.....	20
Supplemental Figure 1: MSA Alignment of 126 GULO protein Sequence.....	28
Supplemental Figure 2: Phylogenetic Tree of all 126 GULO sequence.....	28

Acknowledgments

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

In the name of God, the most gracious, the most merciful.

First and foremost, I express my deepest gratitude to Dr. Eric Gaucher, who has been a terrific mentor throughout this project. I walked into your doors ignorant of how to approach scientific problems and I can say that I am more confident about finding a place to advance the frontiers of science. His guidance and the materials he provided ensured the smooth completion of this work.

To Dr. Jessica Carter, you have been an amazing mentor, and I am truly grateful for your help in helping me find a place in science.

At the later stages of the project, I am grateful for the input of Nuria Ramirez and Allison Lipstein for their help in trying to provide context to some of the results I was presenting at our meetings. To our wonderful lab manager Lily Tran, I am grateful for your help in getting situated into the lab.

To my parents, Mr and Mrs. Sanusi, you have done your best in raising me, I cannot thank you enough for your constant prayers and for ensuring that I get the best education.

To my uncle, Mukaram Enifeni, thank you for being a shoulder of support and an enabler of dreams.

To my dear siblings, especially, my sister Azeezat, thank you for listening to my rants and the continuous motivation to never give up.

To my beloved, Zainab, love gives purpose, a reason to fight, to stand tall, and to never give up. I cannot appreciate enough, your presence in my life.

I am truly grateful to my friends, Moshood, Funmi B, Oreoluwa, and Abdul Hameed, your emotional support throughout this period of carrying out this project is invaluable.

1.0 Introduction

1.1 Importance of Vitamin C

Vitamins are defined as organic compounds that are essential for growth and development and are usually required in trace quantities in diets. As an organic compound, vitamins are distinct from the major macromolecules – fats, carbohydrates, and proteins (Combs and McClung, 2019). It occurs naturally in components of food, typically present in minute quantities, yet crucial for normal physiological function, including maintenance, growth, and development. Notably, its deficiency can lead to a specific syndrome when the organism cannot synthesize it in sufficient amounts to meet its needs or uptake it from diets (Combs and McClung, 2019).

Vitamins can be sourced from plants, dietary supplements, and animal food products. This inability to synthesize in sufficient quantities, or at all, necessitates a dietary intake (Drouin *et al.*, 2011). This highlights the organism's dependence on specific enzymes for vitamin biosynthesis and the potential consequences of their absence or limited activity. Vitamin C (ascorbic acid) is one example.

Ascorbic acid possesses various biological functions, many of which involve its redox properties (Combs and McClung, 2019), positioning it as a key player in cellular antioxidant defense (Combs and McClung, 2019). Additionally, maintaining adequate ascorbic acid levels in the body is essential for collagen, catecholamine, and carnitine biosynthesis (Michels, 2012). Ascorbic acid is the primary water-soluble antioxidant in plasma and tissues which allows it to participate in the contribution of redox recycling of α -tocopherol, enhances the bioavailability of non-heme iron, and maintains optimal oxidation states of enzyme-bound metals for proper

enzymatic function (Combs and McClung, 2019). As a result, compromises in these processes are known contributors to the onset of Vitamin C deficiency.

Higher primates such as tarsier, orangutans, apes belonging to the *Haplorhini* order, and humans, as well as guinea pigs, are among the mammalian species that are unable to synthesize Vitamin C and consequently rely on their diets to obtain this vitamin to prevent scurvy – a pathophysiological onset because of Vitamin C deficiency (Nishikimi *et al.*, 1994).

1.2 Vitamin C Synthesis and L-gulonolactone Oxidase (GULO)

The synthesis of Vitamin C in mammals from glucose occurs via the glucuronic acid pathway (Combs and McClung, 2019). In mammals, the liver is responsible for the synthesis of Vitamin C while the kidney is responsible for its synthesis in fishes, amphibians, older bird orders, and reptiles. Interestingly, there have been changes in the organ responsible for the synthesis of Vitamin C twice during evolution: once from the kidney to the liver in mammals and in a similar manner in birds (Drouin *et al.*, 2011). For egg-laying mammals, ascorbic acid synthesis of Vitamin C occurs only in their kidneys, and many marsupials can use both their liver and kidneys for Vitamin C synthesis (Combs and McClung, 2019). This switch across organs has been explained to be because of selective pressures to balance biochemical homeostasis under stressful conditions (Drouin *et al.*, 2011).

In the pathway shown in Figure 1 below, GULO is the protein that catalyzes the last step in the biosynthesis of Vitamin C. A mutation or disruption of the expression of GULO results in a deficiency of Vitamin C production meaning that the conversion of L-gulonolactone to L-ascorbic acid will not happen.

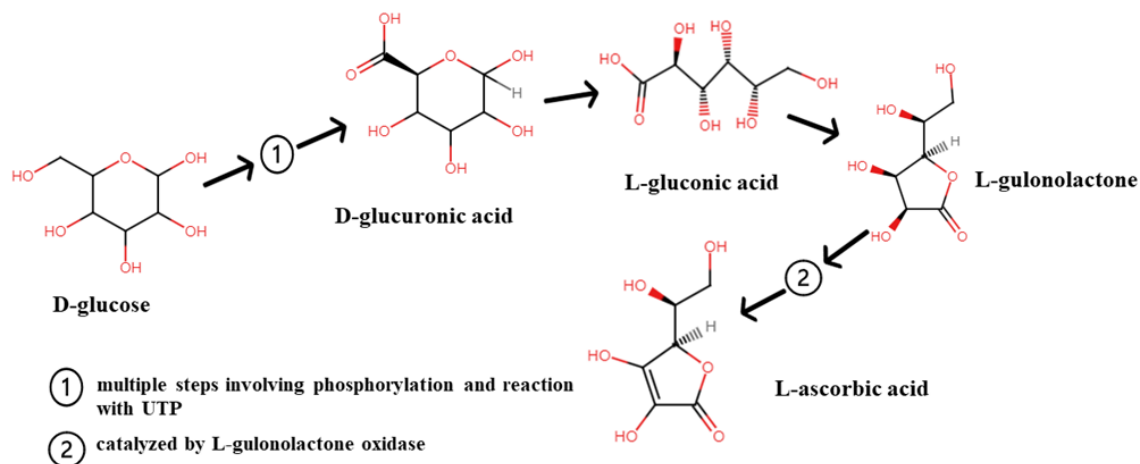


Figure 1: A pathway showing the biosynthesis of ascorbic acid (Vitamin C).

1.3 Evolutionary Loss of GULO in Mammals

Despite the biological significance of Vitamin C, mammalian species such as primates, some bat species, and guinea pigs are unable to produce it and this is due to a loss of a functional GULO gene that has become pseudogene (Henriques *et al.*, 2019). Pseudogenes are described as DNA segments that contain a high degree of sequence similarity to functional genes but are non-functional (Zhang *et al.*, 2019). Through sequence comparison of lineages that have a function and non-functional GULO gene, the inactivation dates of GULO have been placed at 14 and 61 MYA respectively for guinea pig and primate genes (Lachapelle & Drouin, 2011). Pseudogenes can contain disruptive effects, and the pseudogenization of a gene that was once functional is likely caused by a single mutational event. Such events of loss and regaining of function in GULO have been reported in some species of bats and fishes where GULO genes that were once

thought to be nonfunctional have regained function (Cui *et al.*, 2011; Cho *et al.*, 2007). These events can introduce premature stop codons, disrupt splice junctions, cause frameshifts in the coding sequence, or impair the functionality of transcriptional regulatory elements (Zhang *et al.*, 2019).

Pseudogenes that are particularly valuable for functional studies are unitary pseudogenes of which GULO is an example. These are unprocessed pseudogenes lacking functional counterparts within the same genome that arise from disruptive mutations in functional genes, rendering them incapable of being transcribed or translated successfully (Zhang *et al.*, 2019).

1.4 Bioinformatics analysis of GULO evolution

Bioinformatics analysis describes the process of using computational technology to retrieve, process, analyze, and predict biological information. Usually, the first step involved is collecting sequences of interest from biological data banks depending on the macromolecule of interest.

Once sequences are collected, comparative analysis of sequences from species of interest is done through multiple sequence alignment (Ramsden, 2023). Basic Local Alignment Search Tool (BLAST) is one such tool that checks for similar regions between biological sequences by comparing nucleotide or protein sequences to sequences in the database and calculates the statistical significance (Altschul *et al.*, 1997). The comparison of sequences can allow for the inference of biologically significant relationships as well as identifying gene and protein families (Altschul *et al.*, 1990).

Through phylogenetic analysis, computational and mathematical approaches are used to analyze biological datasets using maximum likelihood. This analysis generates trees based on likelihood scores that can give evolutionary insight into the divergence and relatedness of species. Additionally, through the reconstruction of ancestral sequences, amino acid replacements that are highly probable can give insight into the loss of function or otherwise in a protein of interest. By subjecting the GULO protein sequences across several mammalian species to these bioinformatics processes, gaining insight into the possible loss of function of GULO which results in the inability of Vitamin C synthesis in some higher mammals can be identified.

1.5 Research Objectives and Hypothesis

While a good number of mammalian species are currently able to produce Vitamin C due to a functional GULO gene, some mammalian species possess GULOP – a pseudogene. Reconstruction of ancestral sequences can provide evolutionary insights into the mutations and gene conservations that are present in these genes across the several species analyzed. This study utilized ancestral sequence reconstruction and phylogenetic analysis to investigate the evolutionary divergence of the Gulonolactone Oxidase (GULO) gene into pseudogenes (GULOP) in specific mammalian lineages. In particular, this research focused strongly on identifying predicted key mutations of GULO in the *Strepsirrhini* sub-order of primates following ancestral sequence reconstruction. This is because the primates are divided into two suborders: *Strepsirrhini* and *Haplorhini* and they are known to have evolved from a common ancestor. The *Haplorhini* orders containing humans and apes are known to possess pseudogenes and are unable to synthesize Vitamin C.

Amino acid replacements, particularly those affecting amino acid residues critical for enzyme function, that happened in the internal branch before the pseudogene evolved may have played a significant role in the subsequent loss of functional GULO enzyme activity in the *Haplorhini* order.

2.0 Materials and Methods

The bioinformatic analysis of GULO protein sequences was carried out by first collecting sequences from the National Center for Biotechnology Information (NCBI) and querying for the sequences using the Basic Local Alignment Search Tool (BLAST). Following the sequences collection, multiple sequence alignment was carried out on the European Bioinformatics Institute job dispatcher tool using the Clustal Omega algorithm, and the result of the alignment was visualized using Jalview.

After the sequence alignment, the sequences were then passed to Randomized Axelerated Maximum Likelihood (RAXML) to generate a phylogenetic tree. Mesquite was used to readjust the tree branches to conform to a Zoonomia project tree that used whole genome alignment of 240 phylogenetically diverse species of eutherian mammals. Finally, ancestral sequences were then reconstructed using Phylogenetic Analysis by Maximum Likelihood (PAML) to provide evolutionary insight into the divergence of the current mammalian sequences from early mammalian sequences.

2.1 Sequence Collection

Initially, GULO protein sequences were searched for by entering the word “L-gulonolactone Oxidase” on the NCBI protein database. The results were then filtered to include

only sequences for mammals. However, this search was not comprehensive enough as the number of mammalian species with the GULO protein sequence was limited.

A [BLAST](#) (Altschul *et al.*, 1997) protein search was then conducted using the *Mus musculus* protein sequence with accession number NP_848862 as the query search. The non-redundant protein sequences (nr) were selected as the choice database on BLAST, blastp (protein-protein BLAST) as the algorithm and the maximum target sequences were set to 5000.

2.2 Multiple Sequence Alignment

Over 165 mammalian GULO sequences were collected and 3 non-mammalian species belonging to two birds and one fish (supplemental table 1). All 168 sequences were entered into the EBI Clustal Omega job dispatcher (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) tool (Madeira, 2022) for multiple sequence alignment (MSA) with the default settings and the fasta sequence generated were visualized using the Jalview version 2 software (Waterhouse, 2009).

Gaps were removed from the alignment using the Jalview edit tool and visualized as well (supplemental figure 1). Finally, trimming was done manually to get rid of poorly aligned regions from the multiple sequence alignment and even some mammalian species sequences were completely removed. Finally, an MSA fasta file containing 123 mammalian sequences and 3 non-mammalian sequences ([supplemental data](#)) was then used for phylogenetic analysis.

2.3 Phylogenetic Analyses

All 126 sequences were entered into RAxML GUI 2.0 (Edler *et al.*, 2021) for Mac which uses the RAxML tool version 8.0 (Stamatakis, 2014) and RAxML-NG v1.2.1 (Kozlov *et al.*,

2019) for phylogenetic analysis. A model test based on ModelTest-NG (Darriba *et al.*, 2020) was initially run and the parameters (JTT+I+G) for the substitution models with the highest score and weight were selected. Analysis was carried out with the ML + Transfer + Bootstrap Expectation + Consensus based on the RaxML-NG tool (Kozlov *et al.*, 2019). *Acipenser trasnmontanus* (White sturgeon fish) with accession number ABO15549 was selected as the outgroup. Gamma (mean) was selected for rate heterogeneity, 100 for replicates, and the +I (ML estimate) for the proportion of invariant sites.

Following the phylogenetic analysis with RAxML, the best tree generated (figure 2 and supplemental figure 2) visualized with FigTree v1.4.4 was then redrawn to model the Zoonomia project phylogenetic tree of mammalian species (Zoonomia consortium, 2020) using Mesquite 3.81.

2.4 Reconstruction of Ancestral Sequences

The tree output Newick format from Mesquite and all 126 MSA sequences were entered into PAML version 4.9j (Yang, 1997) as tree file and sequence file respectively following the CODEML program specification and parameters in Table 1 below.

Table 1: Parameters used for the phylogenetic analysis using PAML.

Parameter	Value
Outfile	mlc
aaRatefile	wag.dat

Noisy	9
Model	3
fix_alpha, clock, RateAncestor	1
verbose, seqtype,	2
runmode, getSE, aaDist, NSsites, Mgene, alpha, Malpha, getSE, cleandata, fix_blength, method	0
Small_Diff	.5e-6

3.0 Result

3.1 Phylogenetic analysis

Following the analysis by RAxML, the tree in Figure 2 was redrawn using Mesquite to follow the Zoonomia project phylogenetic tree of mammalian species.

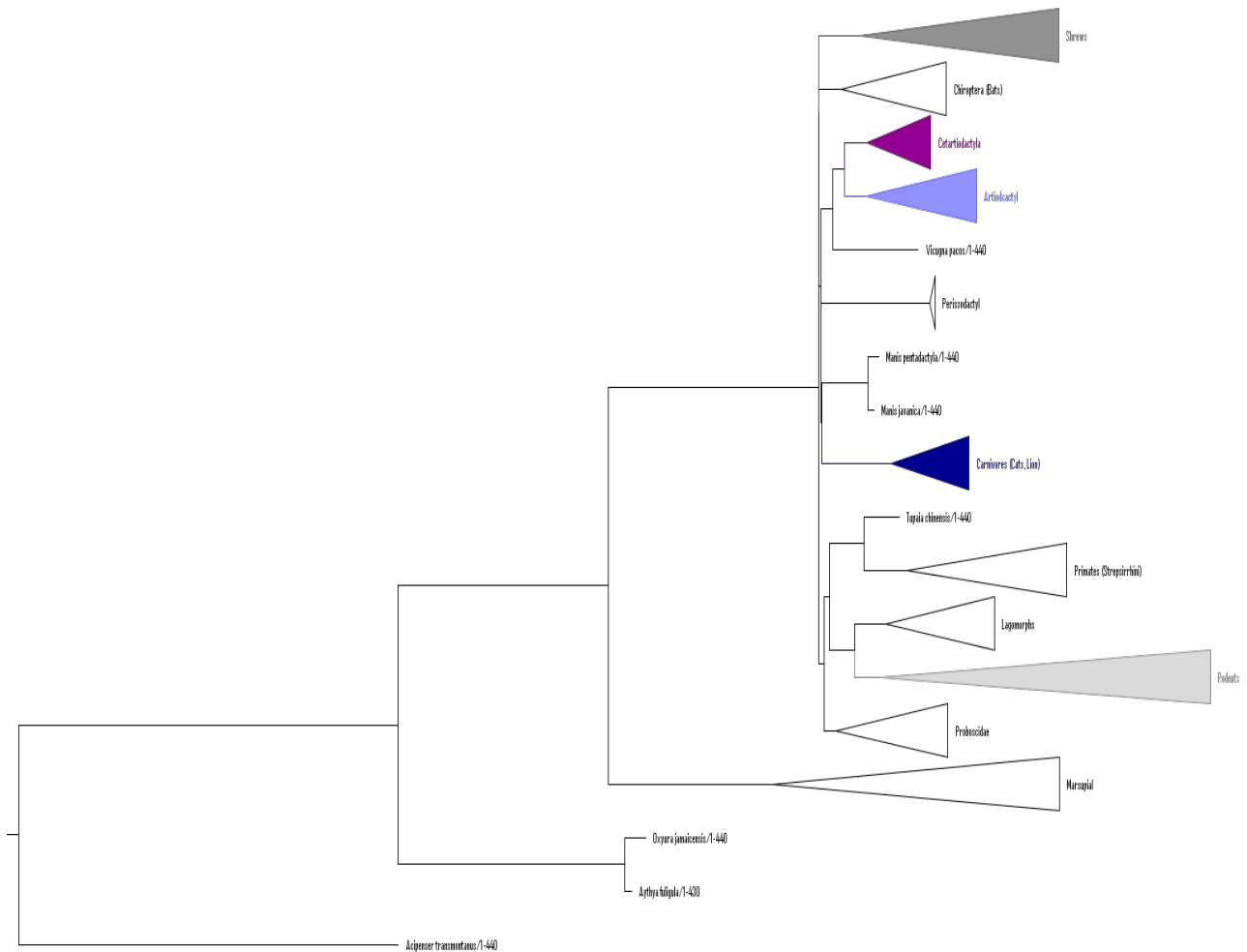


Figure 2: A collapsed phylogram showing the evolutionary relationship of several mammalian species GULO protein sequences. The tree was prepared using Figtree v1.4.4

3.2 Reconstruction of Ancestral Sequences

After the sequences were run through PAML, node 194 shown in Figure three below (figure not drawn to scale) containing mammalian species belonging to the order *Strepsirrhini* is of interest.

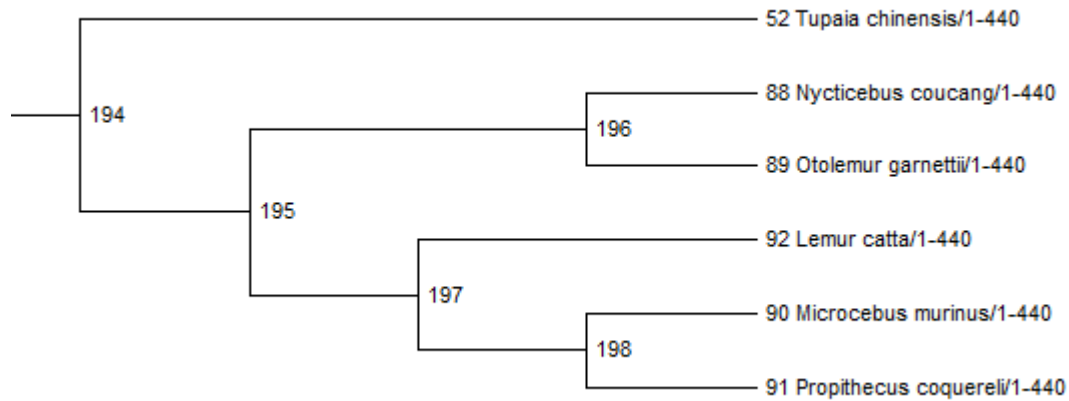


Figure 3: A section of the phylogram showing the node label for the mammalian species belonging to the *Strepsirrhini* order. The tree was prepared using Figtree v1.4.4

Reconstruction of ancestral sequences using PAML showed that moving from node 194 to 195, the ancestral amino acid Q(Glutamine) at position 11 has a high probability (0.995) of being replaced by E (Glutamic acid) with a probability of 0.993. Similarly, at position 119, the ancestral amino acid A (Alanine) is predicted to be replaced by G (Glycine) with a high probability of 0.998. In total, moving from node 194 to node 195, a total of 10 amino acid replacements are predicted and their posterior probabilities are shown in table 2 below.

Table 2: Summary of amino acid changes along nodes 194 to 195 with posterior probabilities

Amino Acid Position	Amino Acid at node 194	Posterior probability of being replaced	Amino Acid at node 195	Posterior Probability of being replaced
11	Glutamine (Q)	0.995	Glutamic Acid (E)	0.993
119	Alanine (A)	0.998	Glycine (G)	0.998
236	Asparagine (N)	0.996	Serine (S)	0.995
251	Isoleucine (I)	0.991	Valine (V)	0.947
283	Threonine (T)	0.997	Asparagine (N)	0.354
286	Lysine (K)	0.991	Arginine (R)	0.995
288	Asparagine (N)	0.996	Histidine (H)	0.996
290	Asparagine (N)	0.997	Aspartic Acid (D)	0.534
337	Tyrosine (Y)	0.990	Phenylalanine (F)	0.991

433	Alanine (A)	0.991	Serine (S)	0.988
-----	-------------	-------	------------	-------

Lastly, the ancestral sequence in supplemental data ([supplemental data](#)) along node 195 was also predicted using PAML.

4.0 Discussion and Conclusion

In Figure 2, the phylogenetic tree reveals the relationship between several groups of mammalian species. The tree starts with placing *Acipenser transmontanus*, a species of fish as the outgroup supporting the theory that life evolved first on water. Also, the relationship between species belonging to the order lagomorphs (*Ochotona pirnceps*) and rodents (*Mus musculus* and *Rattus Novergicus*) is well established showing descent from a common ancestor. Importantly, the population of rodent species is worthy of mention. Perhaps, the representative population in rodents will indicate the success of GULO expression as studies (Ohta & Nishikimi, 1999; Nishikimi *et al.*, 1994) have described the presence of their twelve exons that result in a functional GULO as a basis for comparing the GULO in species such as human that have only six out of the eleven conserved exons responsible for GULO expression.

In Table 2, the amino acid replacements that have occurred can serve as a starting point to questioning the effect of those changes on the GULO expression since the node is the site of divergence between the *Haplorhini* and *Strepsirrhini* species. Amino acid change can have effects on the structure and function of the protein. This is because, changes in amino acid properties such as the hydrophobicity and charges can affect protein function, interactivity with other molecules, or enzymatic activity.

While this study does not proceed to provide a structure for the protein that can be predicted from the generated ancestral sequence, it serves as a basis to further question whether the pseudogenization of the GULO gene is a result of these amino acid replacements.

Supplemental Data

Supplemental Table 1: Accession number for 168 GULO protein sequences collected with accession number.

S/N	Organism	Common Name	Accession Number
1	Rattus_norvegicus	Brown rat	P10867
2	Rattus rattus	Black rat	XP_032773268
3	Arvicanthis niloticus	African grass rat	XP_034353883
4	Mastomys coucha	Multimammate mouse	XP_031217601
5	Grammomys surdaster	African climbing mouse	XP_028614624
6	Mus pahari	Asian house mouse	XP_021059328
7	Mus caroli	Ryukyu Islands spiny rat	XP_021036878
8	Mus musculus	House mouse	NP_848862
9	Peromyscus californicus insignis	Inyo deer mouse	XP_052587792
10	Psammomys obesus	Fat sand rat	XP_055465793
11	Meriones unguiculatus	Mongolian gerbil	XP_021510212
12	Onychomys torridus	Northern grasshopper mouse	XP_036055953
13	Peromyscus maniculatus bairdii	Baird's deer mouse	XP_006988833
14	Peromyscus leucopus	White-footed mouse	XP_028745864
15	Apodemus sylvaticus	Long-tailed field mouse	XP_052046219

16	Acomys_russatus	Greater Egyptian spiny mouse	XP_051016690
17	Nannospalax galili	Galilee mole rat	XP_017652394
18	Phodopus roborovskii	Roborovski hamster	XP_051051450
19	Cricetulus griseus	Chinese hamster	EGW05786
20	Marmota marmota marmota	Alpine marmot	XP_015354934
21	Mesocricetus auratus	Golden hamster	XP_012972737
22	Sciurus carolinensis	Eastern gray squirrel	MBZ3869398
23	Marmota monax	Woodchuck	XP_046298902
24	Marmota flaviventris	Yellow-bellied marmot	XP_027782986
25	Ictidomys tridecemlineatus	Thirteen-lined ground squirrel	XP_040148376
26	Urocitellus parryii	Arctic ground squirrel	XP_026254654
27	Loxodonta africana	African elephant	XP_003412410
28	Elephas maximus indicus	Indian elephant	XP_049721450
29	Tupaia chinensis	Chinese tree shrew	ELW62543
30	Bubalus bubalis	Water buffalo	XP_006057367
31	Trichechus manatus latirostris	Florida manatee	XP_012412689
32	Bos taurus	Cow (domesticated)	NP_001029215
33	Cervus hanglu yarkandensis	Yarkand deer	KAF4019432

34	<i>Orycteropus afer afer</i>	Aardvark	XP_042636607
35	<i>Cervus canadensis</i>	Elk (Wapiti)	XP_043309734
36	<i>Propithecus coquereli</i>	Coquerel's sifaka	XP_012518430
37	<i>Diceros bicornis minor</i>	Western black rhinoceros	KAF5923530
38	<i>Oryctolagus cuniculus</i>	European rabbit	XP_008247496
39	<i>Muntiacus muntjak</i>	Indian muntjac	KAB0351009
40	<i>Meles meles</i>	European badger	XP_045853715
41	<i>Chinchilla lanigera</i>	Chinchilla	XP_005373805
42	<i>Lutra lutra</i>	Eurasian otter	XP_047573464
43	<i>Balaenoptera musculus</i>	Blue whale	XP_036710021
44	<i>Moschus berezovskii</i>	Siberian musk deer	XP_055283247
45	<i>Lontra canadensis</i>	River otter	XP_032711554
46	<i>Heterocephalus glaber</i>	Naked mole-rat	XP_004848449
47	<i>Vicugna pacos</i>	Alpaca	XP_006203265
48	<i>Ailuropoda melanoleuca</i>	Giant panda	XP_011216970
49	<i>Equus quagga</i>	Plains zebra	XP_046511783
50	<i>Castor canadensis</i>	Beaver	XP_020028781
51	<i>Ochotona curzoniae</i>	Plateau pika	XP_040822282
52	<i>Balaenoptera acutorostrata</i> <i>scammoni</i>	Gray whale	XP_007192798
53	<i>Manis pentadactyla</i>	Pangolin	XP_036745019
54	<i>Lemur catta</i>	Ring-tailed lemur	XP_045391779
55	<i>Equus caballus</i>	Horse (domesticated)	XP_005607755

56	Jaculus jaculus	Greater jerboa	XP_012805930
57	Sus scrofa	Wild boar	NP_001123420
58	Mustela erminea	Ermine (short-tailed weasel)	XP_032188235
59	Cervus elaphus	Red deer	XP_043726483
60	Camelus bactrianus	Camelus bactrianus: Bactrian camel	XP_010946865
61	Suricata suricatta	Meerkat	XP_029794159
62	Arvicola amphibius	Water vole	XP_038165953
63	Equus asinus	Donkey (domesticated)	XP_044622796
64	Vulpes lagopus	Arctic fox	XP_041624173
65	Hyaena hyaena	Striped hyena	XP_039073292
66	Ursus arctos	Brown bear	XP_044247110
67	Rangifer tarandus platyrhincus	Svalbard reindeer	CAI9171074
68	Octodon degus	Octodon degus: Degu	XP_012370088
69	Odocoileus virginianus texanus	Texas white-tailed deer	XP_020754301
70	Lagenorhynchus obliquidens	Pacific white-sided dolphin	XP_026943748
71	Monodon monoceros	Narwhal	XP_029059207
72	Manis javanica	Sunda pangolin	XP_036880511
73	Enhydra lutris kenyon	Sea otter	XP_022372216
74	Callorhinus ursinus	Northern fur seal	XP_025729573

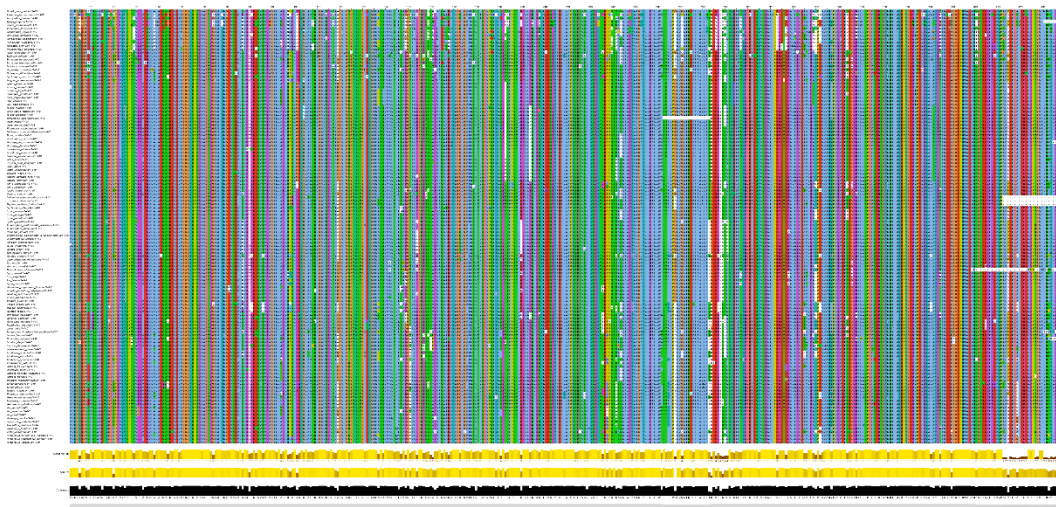
75	<i>Crocota crocuta</i>	Spotted hyena	KAF0878269
76	<i>Canis lupus familiaris</i>	Dog (domesticated)	XP_005635730
77	<i>Odobenus rosmarus divergens</i>	Pacific walrus	XP_004411725
78	<i>Neogale vison</i>	<i>Neogale vison</i> : American mink	XP_044080929
79	<i>Delphinapterus leucas</i>	Beluga whale	XP_022450701
80	<i>Orcinus orca</i>	Killer whale	XP_004270744
81	<i>Vulpes vulpes</i>	Red fox	XP_025863664
82	<i>Zalophus californianus</i>	California sea lion	XP_027453215
83	<i>Microtus oregoni</i>	Creeping vole	XP_041530035
84	<i>Globicephala melas</i>	Short-finned pilot whale	XP_030734958
85	<i>Sousa chinensis</i>	Indo-Pacific humpback dolphin	TEA24377
86	<i>Neophocaena asiaeorientalis asiaeorientalis</i>	Yangtze finless porpoise	XP_024623502
87	<i>Ursus americanus</i>	Black bear	XP_045630038
88	<i>Phoca vitulina</i>	Harbor seal	XP_032245123
89	<i>Choloepus didactylus</i>	<i>Choloepus didactylus</i> : Two-toed sloth	XP_037669235
90	<i>Halichoerus grypus</i>	Atlantic grey seal	XP_035960322
91	<i>Mirounga leonina</i>	Southern elephant seal	XP_034878500

92	<i>Dipodomys spectabilis</i>	Banner-tailed kangaroo rat	XP_042553762
93	<i>Mustela putorius furo</i>	Ferret (domesticated)	XP_004775164
94	<i>Microcebus murinus</i>	Grey mouse lemur	XP_012630826
95	<i>Camelus ferus</i>	Wild Bactrian camel	XP_032327007
96	<i>Dasypus novemcinctus</i>	Nine-banded armadillo	XP_004454215
97	<i>Phocoena sinus</i>	Gulf of California porpoise	XP_032491018
98	<i>Chrysochloris asiatica</i>	Golden mole	XP_006864497
99	<i>Mirounga angustirostris</i>	Northern elephant seal	XP_045735505
100	<i>Ochotona princeps</i>	Pika	XP_012786868
101	<i>Dipodomys ordii</i>	Ord's kangaroo rat	<u>XP_012879644</u>
102	<i>Fukomys damarensis</i>	Damaraland mole-rat	XP_010621731
103	<i>Neomonachus schauinslandi</i>	Hawaiian monk seal	XP_044768742
104	<i>Felis catus</i>	Domestic cat	XP_019684152
105	<i>Lynx canadensis</i>	Canada lynx	XP_030168519
106	<i>Microtus fortis</i>	Singing vole	XP_050008092
107	<i>Puma yagouaroundi</i>	Yaguarundi	XP_040319797
108	<i>Nycticebus coucang</i>	Slow loris	XP_053434594
109	<i>Panthera tigris</i>	Tiger	XP_015399705
110	<i>Eumetopias jubatus</i>	Steller sea lion	XP_027970024
111	<i>Rhinolophus sinicus</i>	Chinese horseshoe bat	XP_019588647
112	<i>Erinaceus europaeus</i>	European hedgehog	XP_016041119
113	<i>Rousettus aegyptiacus</i>	Egyptian fruit bat	KAF6401440

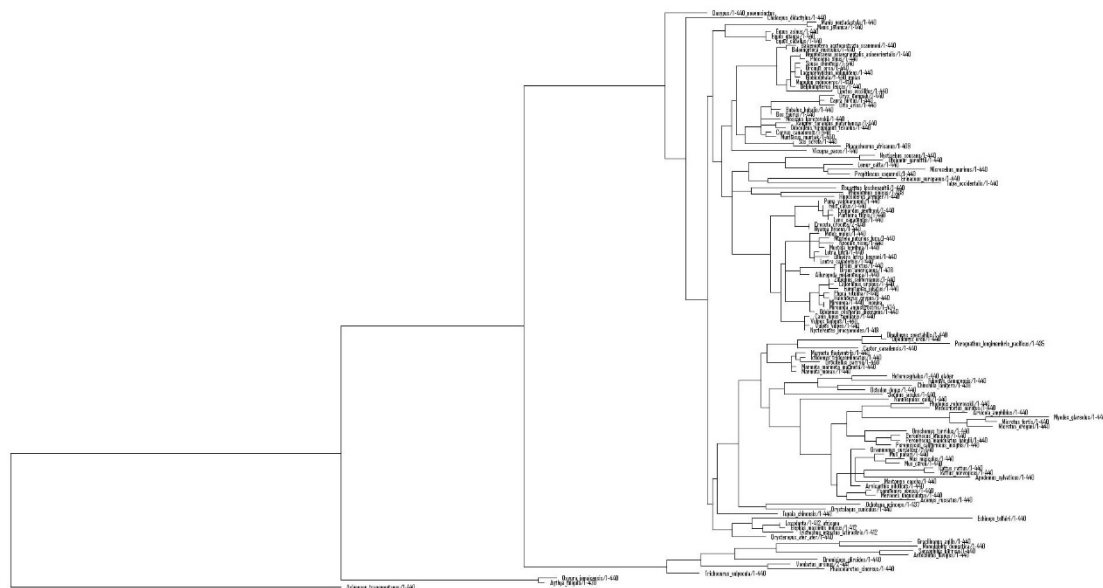
114	<i>Leopardus geoffroyi</i>	Geoffroy's cat	XP_045317203
115	<i>Panthera uncia</i>	Snow leopard	XP_049486030
116	<i>Rousettus leschenaultii</i>	Leschenault's rousette	ADP88813
117	<i>Panthera leo</i>	Lion	XP_042792602
118	<i>Hipposideros armiger</i>	Round-eared horseshoe bat	ADP88814
119	<i>Perognathus longimembris</i> <i>pacificus</i>	Pacific long-eared pocket mouse	XP_048186939
120	<i>Otolemur garnettii</i>	Allen's bushbaby	XP_012663175
121	<i>Lynx rufus</i>	Bobcat	XP_046948263
122	<i>Oryx dammah</i>	Scimitar-horned oryx	XP_040092092
123	<i>Capra hircus</i>	Goat (domesticated)	XP_005684101
124	<i>Ovis aries</i>	Sheep (domesticated)	XP_004004204
125	<i>Lipotes vexillifer</i>	Baiji (extinct)	XP_007466224
126	<i>Myodes glareolus</i>	Bank vole	XP_048277885
127	<i>Echinops telfairi</i>	Lesser hedgehog tenrec	XP_045142718
128	<i>Talpa occidentalis</i>	Broad-footed mole	XP_037382605
129	<i>Budorcas taxicolor</i>	Takin	XP_052500500
130	<i>Ceratotherium simum</i> <i>simum</i>	White rhinoceros	XP_014645409
131	<i>Ursus maritimus</i>	Polar bear	XP_040481655
132	<i>Sus scrofa</i>	Wild boar	XP_020927730
133	<i>Nyctereutes procyonoides</i>	Raccoon dog	XP_055156208
134	<i>Prionailurus viverrinus</i>	Fishing cat	XP_047713650

135	Rhinolophus ferrumequinum	Greater horseshoe bat	XP_032990522
136	Camelus_dromedarius	Dromedary camel	KAB1256122
137	Microtus ochrogaster	Prairie vole	XP_005355607
138	Bubalus carabanensis	Water buffalo (swamp buffalo)	XP_055435716
139	Bos mutus	Yak	MXQ93553
140	Trichosurus vulpecula	Common brushtail possum	XP_036608784
141	Suncus etruscus	Etruscan shrew	XP_049628589
142	Vombatus ursinus	Wombat	XP_027706054
143	Sorex araneus	Common shrew	XP_054999526
144	Phascolarctos cinereus	Koala	XP_020839637
145	Sarcophilus harrisii	Tasmanian devil	XP_023351196
146	Dromiciops gliroides	Monito del monte	XP_043846010
147	Antechinus flavipes	Yellow-footed antechinus	XP_051832348
148	Puma concolor	Cougar (mountain lion)	XP_025778041
149	Phacochoerus africanus	Common warthog	XP_047617902
150	Gracilinanus agilis	Agile antechinus	XP_044516966
151	Monodelphis domestica	White-eared opossum	XP_007475977
152	Bos indicus	Zebu (humped cattle)	XP_019821156
153	Tachyglossus aculeatus	Short-beaked echidna	XP_038627035
154	Lynx pardinus	Spanish lynx	VFV44087
155	Eschrichtius robustus	Gray whale	KAJ8798511

156	<i>Equus przewalskii</i>	Przewalski's horse	XP_008517522
157	<i>Balaenoptera physalus</i>	Fin whale	KAB0399106
158	<i>Condylura cristata</i>	Star-nosed mole	XP_004682583
159	<i>Bison bison bison</i>	American bison (buffalo)	XP_010838441
160	<i>Cavia porcellus</i>	Guinea pig (domesticated)	XP_012998768
161	<i>Neotoma lepida</i>	Desert woodrat	OBS73910
162	<i>Pteropus giganteus</i>	Indian flying fox	XP_039731702
163	<i>Physeter catodon</i>	Sperm whale	XP_028349288
164	<i>Leptonychotes weddellii</i>	Weddell seal	XP_006740879
165	<i>Galemys pyrenaicus</i>	Pyrenean desman	KAG8506324
166	<i>Acipenser Transmontanus</i>	White sturgeon	ABO15549
167	<i>Aythya fuligula</i>	Pochard	XP_032040416
168	<i>Oxyura jamaicensis</i>	Ruddy duck	XP_035178535



Supplemental Figure 1: MSA Alignment of 126 GULO protein sequence. Alignment was visualized using Jalview.



Supplemental Figure 2: Phylogenetic Tree of all 126 GULO sequence. Alignment was visualized using Figtree.

References

- 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.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389-3402.
- Andrew Rambaut. FigTree Tee Figure Drawing Tool. Institute of Evolutionary Biology, University of Edinburgh. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>
- Cho, Y. S., Douglas, S. E., Gallant, J. W., Kim, K. Y., Kim, D. S., & Nam, Y. K. (2007). Isolation and characterization of cDNA sequences of L-gulonono-gamma-lactone oxidase, a key enzyme for biosynthesis of ascorbic acid, from extant primitive fish groups. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 147(2), 178-190.
- Combs Jr, G. F., & McClung, J. P. (2016). *The vitamins: fundamental aspects in nutrition and health*. Academic press.
- Cui, J., Pan, Y. H., Zhang, Y., Jones, G., & Zhang, S. (2011). Progressive pseudogenization: Vitamin C synthesis and its loss in bats. *Molecular biology and evolution*, 28(2), 1025-1031.

- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular biology and evolution*, 37(1), 291-294.
- Drouin, G., Godin, J. R., & Pagé, B. (2011). The genetics of Vitamin C loss in vertebrates. *Current genomics*, 12(5), 371-378.
- Edler, D., Klein, J., Antonelli, A., & Silvestro, D. (2021). raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution*, 12(2), 373-377.
- Henriques, S. F., Duque, P., López-Fernández, H., Vázquez, N., Fdez-Riverola, F., Reboiro-Jato, M., ... & Vieira, J. (2019). Multiple independent L-gulonolactone oxidase (GULO) gene losses and Vitamin C synthesis reacquisition events in non-Deuterostomian animal species. *BMC Evolutionary Biology*, 19, 1-12.
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453-4455.
- Lachapelle, M. Y., & Drouin, G. (2011). Inactivation dates of the human and guinea pig Vitamin C genes. *Genetica*, 139, 199-207.
- Maddison, W. P., & Maddison, D.R. (2023). Mesquite): a modular system for evolutionary analysis. Version 3.81. <http://www.mesquiteproject.org>.

- Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., ... & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research*, 50(W1), W276-W279.
- Nishikimi, M., Fukuyama, R., Minoshima, S., Shimizu, N., & Yagi, K. (1994). Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *Journal of biological chemistry*, 269(18), 13685-13688.
- Ohta, Y., & Nishikimi, M. (1999). Random nucleotide substitutions in primate nonfunctional gene for L-gulono- γ -lactone oxidase, the missing enzyme in L-ascorbic acid biosynthesis. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1472(1-2), 408-411.
- Ramsden, J. (2023). *Bioinformatics: an introduction*. Springer Nature.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Stipanuk, M. H., & Caudill, M. A. (2012). *Biochemical, Physiological, and Molecular Aspects of Human Nutrition-E-Book: Biochemical, Physiological, and Molecular Aspects of Human Nutrition-E-Book*. Elsevier health sciences.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood.

Computer applications in the biosciences, 13(5), 555-556.

Zoonomia Consortium (2020). A comparative genomics multitool for scientific discovery and

conservation. *Nature* **587**, 240–245. <https://doi.org/10.1038/s41586-020-2876-6>