Evaluation of Diversity and Selection Pressure on CEACAM-I Cellular Receptor in Human

SUPERIOR UNIVERSITY

Thesis Submitted to

The Superior University Lahore

In Partial Fulfillment of the

Requirement for the Degree of

Master of Philosophy in Biotechnology

By

KHADIJA TABASSUM

Roll No. SU92-MSBTW-F22-011

Session: 2022-2024

Faculty of Biological Sciences

Author's Declaration

I hereby state that my MS/M.Phil. thesis titled **"Evaluation of diversity and selection pressure on CEACAM-I cellular receptor in human"** is my work and has not been submitted previously by me for taking any degree from this University,

The Superior University, Lahore,

or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the university has the right to withdraw my MS/M.Phil. degree.

Name of Student: **Khadija Tabassum**

Date: _____

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled **"Evaluation of diversity and selection pressure on CEACAM-I cellular receptor in human"** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and University,

The Superior University, Lahore,

towards plagiarism. Therefore, I as author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as a reference is properly referred/cited. I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis, even after awarding of MS/M.Phil. degree, the University reserves the rights to withdraw/revoke my MS/M.Phil. degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted a plagiarized thesis.

Student/Author Signature: _____

Name: Khadija Tabassum

Research Completion Certificate

This is to certify that the thesis entitled **"Evaluation of diversity and selection pressure on CEACAM-I cellular receptor in human"** submitted by **"Khadija Tabassum "**has been accepted towards the partial fulfillment of the requirement for MS/M.Phil. **"MS Biotechnology"**. The quality of the work contained in this thesis is adequate for the award of degree.

Supervisor Name: Dr. Rabia Nawaz

Designation: Assistant Professor

Signature: _____

Certificate of Approval

This is to certify that the research work presented in this thesis, titled **"Evaluation of diversity and selection pressure on CEACAM-I cellular receptor in human"** was conducted by **"Khadija Tabassum"** under the supervision of **"Dr. Rabia Nawaz"**

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Faculty of Economics and Commerce, The Superior University, Lahore in partial fulfillment of the requirements for the degree of Master of Philosophy in the field of "**Molecular Biology"** in Faculty of Economics and Commerce at The Superior University, Lahore.

Student Name: Khadija Tabassum Signature:_____

Examination Committee:

a) External Examiner: Signature:

_____

b) Supervisor Name: Dr. Rabia Nawaz Signature:

_____

c) Name of Program Leader/HOD: Dr. Uqba Mehmood

Signature: _____

DEDICATION

I dedicate this work to my mother, who, despite her strong desire for education, was unable to complete middle school. She made the selfless choice to care for my grandmother during her illness, sacrificing her own studies.

Khadija Tabassum

ACKNOWLEDGMENTS

In the name of Allah, the most Gracious, the most Merciful.

I am extremely thankful to Almighty ‘Allah’ Who is the entire source of knowledge and wisdom endowed to mankind, for providing me with the acumen and vision to complete this endeavor.

I am grateful to my parents and siblings for always having my back.

I am thankful to my friends and research fellows for keep guiding me during all my research time.

I would regard the support of Management of Al Noor Diagnostic Centre, Lahore where I am doing Job. I appreciate the support of my work colleagues too.

At last, but not the least, I am humbly thankful to Dr Rabia Nawaz, my supervisor, for keep pushing me to complete the research work. She keeps guiding and counseling me at my hardest times. I am eternally grateful for all the effort and time she has invested in my academic journey.

Additionally, I would like to take this chance to acknowledge the Superior University for providing me a constructive environment for my research.

Finally, I would like to thank everyone who has directly or indirectly supported me throughout this process. Your kindness and encouragement have been greatly appreciated.

Khadija Tabassum

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Understanding the molecular diversity of viral receptors is pivotal in unraveling the mechanisms of host-pathogen interactions, especially in the context of zoonotic transmission. This thesis investigates the CEACAM1 receptor, a key molecule exploited by Mouse Hepatitis Virus (MHV), to evaluate how genetic variation and evolutionary pressures influence susceptibility to infection and potential cross-species viral transmission. By analyzing CEACAM1 sequences across various species, this study applies Shannon's Diversity Index to quantify receptor variability. Additionally, evolutionary analyses using MEME and aBSREL in the HyPhy framework were employed to detect site-specific selection pressures. The findings highlight both conserved and adaptive regions of CEACAM1, shedding light on the receptor's role in viral adaptability. These results not only deepen our understanding of host-virus co-evolution but also offer meaningful insight into the molecular underpinnings that may facilitate or hinder zoonotic spillovers.

CHAPTER 1

INTRODUCTION

The study of receptor diversity is essential for understanding the intricacies of viral infections, particularly zoonotic viruses that cross species barriers. Receptor diversity refers to the molecular variations in cell receptors, often resulting from genetic polymorphisms, mutations, alternative splicing, or post-translational modifications (1). These variations can alter how receptors interact with viral particles, which can either increase or decrease susceptibility to infection. Not only does receptor diversity influence the infection dynamics within a species, but it also plays a critical role in shaping interspecies viral transmission. Consequently, investigating receptor variation across species is essential for grasping the mechanisms behind viral adaptation, host specificity, and zoonotic potential (2).

Zoonotic viruses which are capable of infecting multiple species, pose a significant threat to global health due to their ability to evolve and adapt to new hosts. Many of these viruses, such as influenza A, SARS-CoV, and MERS-CoV, rely on specific host cell receptors to initiate infection. For instance, the ACE2 receptor is pivotal in SARS-CoV-2's entry into human cells, and its structural variability across species sheds light on potential cross-species transmission pathways(3,4). Research into these receptor-virus interactions helps elucidate why certain species are more susceptible to infection than others and offers clues about potential barriers to interspecies transmission. By focusing on the structural and functional aspects of receptors used by zoonotic viruses, scientists can better predict and mitigate the risk of spill overs, which can be instrumental in controlling emerging pandemics.

The relationship between viruses and host receptors is shaped by evolutionary pressures, driving adaptations on both sides of this biological "arms race." Viruses evolve to recognize conserved elements of receptors across various hosts, allowing them to spread among diverse species (5). Meanwhile, host organisms may undergo evolutionary changes in receptor structure, reducing the binding affinity for viral particles as a protective mechanism. This co-evolutionary struggle not only impacts viral fitness but also provides valuable insights into the pathways of zoonotic transmission (6). Studying receptor diversity, particularly in relation to viruses capable

of crossing species barriers, is therefore crucial for understanding and potentially interrupting these evolutionary adaptations.

Mouse Hepatitis Virus (MHV) serves as a model virus for examining the effects of receptor diversity on viral transmission (7). MHV infects mice through the CEACAM1 receptor, a highly variable molecule that illustrates the complexities of receptor-mediated viral entry (8). The diversity of CEACAM1 in mammals can demonstrate how genetic variability within a receptor can influence viral adaptation, pathogenesis, and spread within and beyond a host species.

This study examines the diversity and evolutionary dynamics of the CEACAM1 receptor, which is critical in understanding mechanisms underlying viral entry, adaptability, and cross-species transmission. By assessing the diversity of CEACAM1 across species using Shannon's Diversity Index, we aimed to characterize receptor variability that may influence susceptibility to viral infections. Furthermore, molecular evolutionary analyses using MEME and aBSREL in HyPhy provided insights into site-specific selection pressures, highlighting evolutionary adaptations that may affect viral binding and zoonotic potential. This approach helps elucidate both genetic diversity and evolutionary selection on receptors, offering implications for zoonotic transmission risks and potential therapeutic targets.

AIM AND OBJECTIVE

- To analyse CEACAM-1 receptor in humans and mice to study its evolutionary host-virus interactions
- To investigate specie-specific differences that are important in immune responses

CHAPTER 2

LITERATURE REVIEW

2.1. RECEPTOR DIVERSITY

Receptor diversity is defined as the variations in the molecular structure of a cell receptor. Cell receptors are mainly made up of proteins and glycoproteins, present on the surface of cells and bind with particular molecules such as hormones, neurotransmitters, or viruses. The causes of diversity are genetic changes, alternative splicing, post-translational modifications, and environmental factors (1).

Cell receptors are sensitive to some of the cell pathogens and viruses, which is why the study of receptor diversity is important. Different individuals and different species can have different variants of a receptor. These variations affect the binding and entrance of a virus into the cell (9). The study of these variations might help to understand cross-species transmission, which can help to provide information to public health strategies to prevent and control an epidemic.

2.2. INTER-TRANSMISSIBLE VIRUSES AND RECEPTOR DIVERSITY

Receptor diversity is important particularly in the perspective of viruses which can jump from one species to another, including human beings. For example, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East Respiratory syndrome coronavirus (MERS-CoV) were first originated in animals (3,4), but due to special interactions, they were able to infect humans. SARS-CoV 2 first appeared in China and spread internationally, while MERS-CoV was first identified in Saudi Arabia in 2012 (10,11).

2.2.1. Influenza A virus

Influenza A virus is basically a bird virus originating from wild waterfowl (12). It transmits into humans either directly from infected birds or through an intermediate host such as pigs, which can provide a habitat for these reassorted strains of the virus. This virus uses sialic acid receptors present on the epithelial cells of the respiratory pathway of humans for its entrance. The severity of infection caused by Influenza A virus can range from mild symptoms like the flu to pneumonia and severe

complications associated with acute respiratory distress syndrome (ARDS). Animal epidemics of seasonal flu cause millions of deaths; moreover, epidemics such as the Spanish flu in 1918 caused damage to millions of people (13,14).These kinds of epidemics also affect the economy of a country severely; it includes healthcare expenses, loss of yield, and economic disruption during the epidemic.

2.2.2. Coronaviruses

SARS-CoV, MERS-CoV, and SARS-CoV-2, which cause COVID-19, are types of coronaviruses and are considered to have zoonotic origin (15). SARS-CoV is believed to transmit from bats to humans through civet cats as intermediate hosts (16), while MERS-CoV transfers from bats to camels before infecting human beings. SARS-CoV-2 is believed to originate from bats before starting infection in humans, certainly using an intermediate host which is not identified yet. Both SARS-CoV and SARS-CoV-2 use the ACE2 receptor to enter human cells, while MERS-CoV uses the DPP4 receptor for its entrance (17,18). Damage caused by these viruses can range from mild respiratory syndromes to severe pneumonia and major organ damage. SARS-CoV-2 has caused millions of deaths globally, resulting in international lockdowns, overwhelmed healthcare systems, and economic losses in different fields and professions of life.

2.2.3. Human immunodeficiency virus (HIV)

Human immunodeficiency virus (HIV) evolved from simian immunodeficiency virus (SIV), which is found in primates. It is believed that the consumption of bushmeat is the cause of its infection in humans. HIV targets CD4 receptors present on T-helper cells, resulting in acquired immunodeficiency syndrome (AIDS), which severely affects the immune system (19). Lack of immunity leads to opportunistic infections and cancer. Since the beginning of the epidemic, HIV has caused 20 million deaths globally and 36.8 million people are currently living with HIV (20). Infected individuals bear the economic burden of lifelong antiretroviral therapy too. Severely affected regions put economic pressure on healthcare systems and cause production losses.

2.2.4. Ebola virus

The origin of the Ebola virus is considered bats, with its subsequent transmission into humans happening by direct contact with infected animals such as

bats and other primates (16). Ebola virus uses the NPC1 receptor to enter human cells (Wang et al., 2016). It causes severe haemorrhagic fever, whose death ratio is exceedingly high. For example, in the outbreak of 2014 to 2016 in West Africa, there were more than 11,000 casualties (21). The economic impact of the Ebola virus includes emergency response expenses, healthcare expenses, and disruption of social and economic activities in the affected regions.

2.2.5. Nipah virus

Nipah virus originated in fruit bats and moved into humans through contact with infected bats, contaminated food, or person-to-person contact (16). It uses the Ephrin B2 and Ephrin B3 receptors for entrance (22). Nipah virus can cause severe respiratory and neurological infections. The death ratio due to Nipah virus in different outbreaks was 40% to 70% (22). Economic impacts by Nipah virus infection include outbreak management expenses, healthcare expenses, and economic disruption in affected regions.

2.2.6. Hendra virus

Hendra virus is also a bat-borne virus that transfers into humans through infected horses, which work as intermediate hosts (16). This virus causes severe respiratory and neurological disorders, leading to a very high death ratio (23). Economic impacts by this virus include epidemic control measures, healthcare expenses, and economic losses in the livestock industry due to infected livestock.

These inter-transmissible viruses highlight the importance of receptor diversity as variations in human receptors can affect the ability of virus binding and transmission potential.

2.3. VIRAL STRATEGIES TO EXPLOIT CONSERVED/ HOMOLOGOUS RECEPTORS

Inter-transmissible viruses exploit the common receptor in different species due to the conserved structure of a receptor and its adaptability towards a virus. Many receptors are proteins whose sequences and structures are conserved in different species. They perform important biological functions and are not prone to large mutations. Especially their peculiar binding sites for virus binding are very conserved regions. Due to these conserved binding sites of a receptor, a virus that can bind to a

particular receptor of a species can also bind with the similar receptor belonging to another species. Viruses have the tendency to mutate rapidly (24). The mutations in the surface protein of a virus, such as in the spike protein of the coronavirus, can improve the ability of a virus to bind with the receptor of a new host. Additionally, the exchange of genetic material between different viruses in recombination can also enhance viral entry in different species (25). In an environment that harbours many host species, natural selection also facilitates those viral variants which can identify the conserved regions of receptors, hence increasing the host spectrum.

Another possibility that can decrease the distance between different species is the intermediate hosts. For example, in the case of the influenza virus, pigs can be affected by both human and bird variants, which can lead to a new variant potentially harmful to humans (26). Direct zoonotic spill-over happens when humans live close to infected animals. For example, SARS-CoV-2 transmission in humans from infected bats probably used an intermediate host species (27). Receptor diversity in species affects the adaptability and evolution of viruses, whereas selection pressure facilitates the adaptability of these viruses which can exploit these conserved structures of receptors. A special component of viral protein, such as the spike protein, interacts with the binding domains of the host receptor, and adaptation in those domains can enhance the binding ability between the virus and species. Moreover, similarities between host cell machinery of species, upon which viruses depend to enter into the host cell after binding, also facilitate the infection in species (28). The exploitation of similar receptors by viruses is a complex play of selection pressure, structure conservation, and viral adaptation. To predict and manage zoonotic spill-over, the understanding of these mechanisms, especially receptor diversity and selection pressure, is of immense importance.

2.4. SELECTION PRESSURE

Selection pressures are environmental factors that influence the survival and reproductive success of organisms, resulting in variations in specific traits within a population. This concept is fundamental to natural selection, where individuals with traits that are more suited to their environment have a greater likelihood of surviving and reproducing. The sources of selection pressure can differ, including factors like diseases, predation, competition, and changes in climate. It serves as a key driving force behind evolutionary changes, helping populations adapt more effectively to their

habitats. In terms of receptor diversity, selection pressure is crucial in the development of cell receptors, particularly regarding viral entry. The outcomes of this evolutionary struggle indicate that selection pressure impacts both sides (29,30).

The ability of viruses to bind with host receptors is essential for their transmission and replication. Mutations that enhance this binding capability are naturally selected, leading to the emergence of more virulent and infectious strains. For instance, both HIV and MERS-CoV specifically target receptors in humans. HIV attaches to the CCR5 receptor for transmission, while MERS-CoV interacts with the DPP4 receptor (31,32). Changes in these receptors can greatly affect how well the virus binds. A notable example is the CCR5-Δ32 mutation found in some individuals, which offers resistance to HIV infection (33). Likewise, the sialic acid receptor has various forms that the influenza A virus exploits for infection, with these variants showing different sensitivities during various stages of the viral lifecycle (34). Conversely, hosts experience selection pressure that leads to the development of receptor variations aimed at blocking viral entry or enhancing the immune system's ability to detect the virus. This pressure encourages the evolution of receptor variations that can either reduce viral binding or bolster the host's immune response (29). For example, certain mammals have evolved receptor variants that confer resistance to specific viral strains, such as some types of avian influenza (34).

This ongoing interaction between viral strains and host receptors highlights how selection pressure influences the evolution of both viruses and their hosts. Understanding these interactions is vital for predicting potential zoonotic viruses that may jump to humans and cause infections. It also plays a key role in developing countermeasures, such as vaccines and antiviral treatments, by underscoring the importance of considering receptor diversity and viral binding mechanisms in preventive and therapeutic strategies.

2.5. MOUSE HEPATITIS VIRUS (MHV)

Mouse hepatitis virus (MHV), also referred to as murine hepatitis virus, is a type of coronavirus that primarily affects mice (35). It is part of the *Coronaviridae* family and falls under the *Betacoronavirus* genus (36). This virus serves as a valuable model for studying the pathogenicity of other coronaviruses. It plays a role in veterinary medicine and is among the most researched coronaviruses, including those that affect

humans. MHV spreads extensively in laboratory mouse colonies, with its prevalence influenced by biosecurity measures and animal husbandry practices (7).

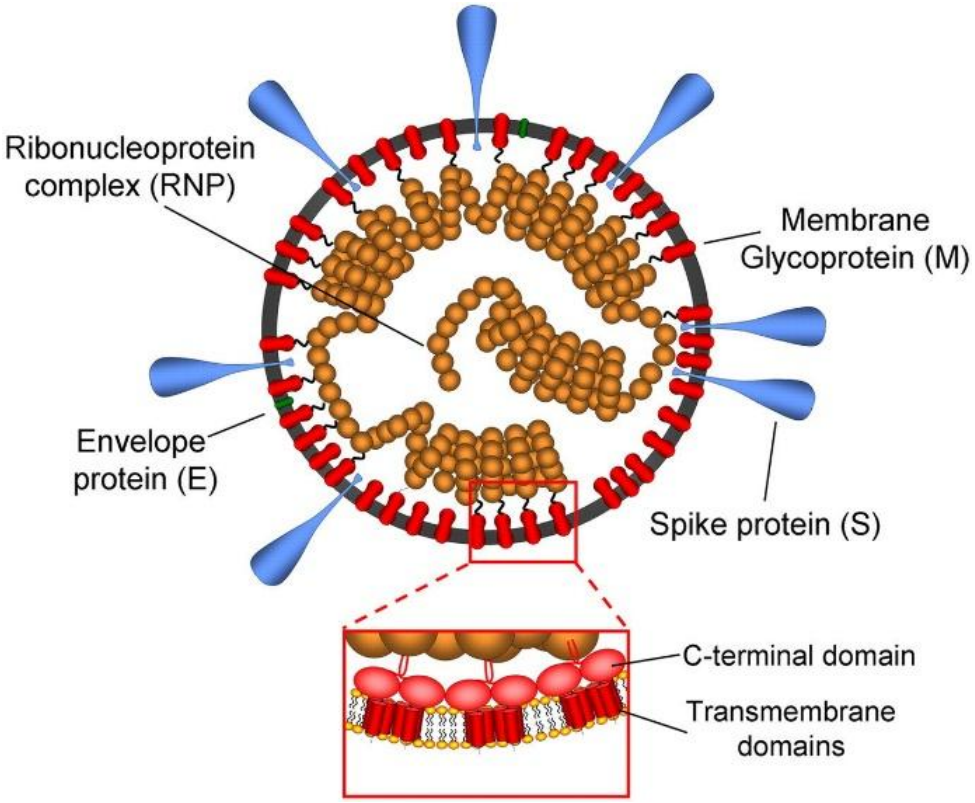

Figure 2.1. Coronavirus Structural Model from Tomogram Analysis. Inside the virus, disordered RNP segments alternate with coiled structures. M proteins exist in different forms; monomers, dimers, trimers, or tetramers; and are spread throughout the envelope, interacting locally. The C-terminal domains of these M proteins create an additional layer beneath the lipid membrane. The connections between the envelope and RNP highlight key anchor points essential for viral assembly and stability (7).

MHV infection can be acute or chronic, and its severity depends upon the viral strain and the immune system of the host (37). MHV can infect directly or indirectly by means of contaminated bedding, food, water, and other fomites. Nostrils, faecal, and oral routes are important means of viral transmission. MHV is highly contagious; once entered into the mouse colony, it spreads rapidly. Faeces, urine, and respiratory droplets released by infected mice facilitate the virus's spread to other mice. This virus can remain on surfaces for a long time, contributing to its persistence in the environment. To prevent and manage MHV outbreaks in laboratories, it is crucial to implement effective control measures, including maintaining a clean environment, quarantining newly infected animals, and conducting regular checks and balances (38).

In mice, MHV primarily targets the liver and central nervous system (CNS), but it can also infect the gastrointestinal and respiratory systems, along with other organs. The virus enters host cells by utilizing specific receptors, particularly the carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). This glycoprotein is found on various cell types, including hepatocytes, endothelial cells, and certain immune cells. CEACAM1 plays a crucial role in allowing the virus to enter the host cell, leading to the spread of the infection. The interaction between MHV and CEACAM1 is vital for the virus's pathogenicity and influences the duration and severity of the disease in infected mice (8). Gaining insights into the MHV receptor and its transmission mechanisms is essential not only for managing its effects but also for predicting potential outbreaks in other species, including humans.

2.6. CARCINO-EMBRYONIC ANTIGEN (CEA)

Carcino-embryonic antigen (CEA, CD66e) was initially identified as a tumour antigen (39). Currently, all glycoproteins associated with the immunoglobulin superfamily (IGSF) are categorized under the CEA family. These glycoproteins, which can either be anchored to the cell membrane or secreted, are found on epithelial cells, leukocytes, endothelial cells, and in the placenta. In humans, there are 29 genes and pseudogenes that fall into two subfamilies: the CEA-related cell adhesion molecule (CEACAM) subfamily and the pregnancy-specific glycoprotein (PSG) subfamily. These proteins share several common structural characteristics (40).

2.6.1. Carcino-embryonic antigen-related molecule 1 (CEACAM1)

CEACAM1, or carcino-embryonic antigen-related molecule 1, is the first member of the CEACAM family of glycosylated immunoglobulin (Ig) molecules. It is also known as cluster of differentiation 66a (CD66a) and biliary glycoprotein (BGP). Initially identified as biliary glycoprotein, it is found on the gallbladder and the surface of hepatocytes. Its expression begins during early embryonic development and is present on placental trophoblasts, pre-implanting embryos, and infiltrating leukocytes. While most species have a single CEACAM1 gene, mice possess two related genes known as CEACAM1 and CEACAM2 (41).

2.6.2. CEACAM1 Structure

CEACAM1 is a glycoprotein that belongs to the CEA family and is known for its complex and multifaceted structure. In mice, there are four isoforms of CEACAM1

produced through alternative mRNA splicing. These isoforms contain either two or four immunoglobulin (Ig)-like domains, referred to as D1 to D4. These domains are located on the cell surface and play a crucial role in maintaining the protein's structural integrity. The transmembrane segment of the protein helps it to embed in the cell membrane. A cytoplasmic tail, which can be either long or short, is also part of the protein structure. The long cytoplasmic tail is known for the ITIM-like motif (immune receptor tyrosine-based inhibition motif). It can undergo tyrosine phosphorylation, but the exact method of gene expression of the ectodomain by natural ligands and CEACAM1 signalling are not well known. The glycosylation of these domains, specifically broad N-linked glycosylation, makes the crystallization of CEACAM1 and similar proteins to investigate their in-depth structure (42,43).

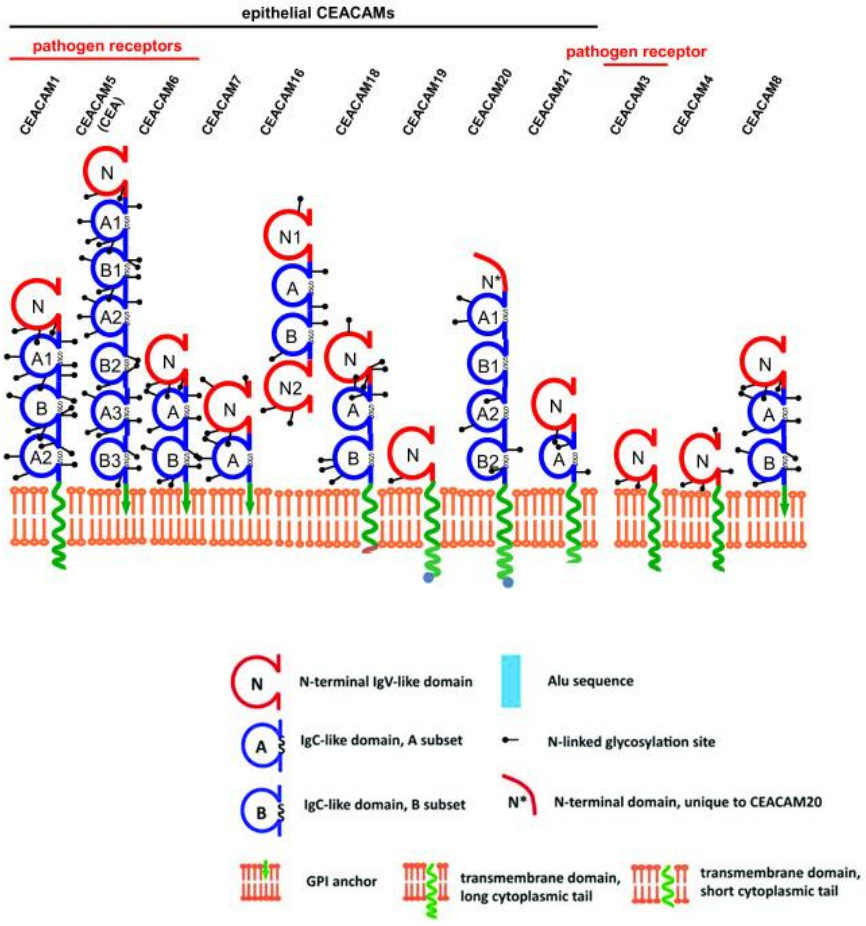

Figure 2.2: Overview of the structural variations in CEACAM1 splice variants and human CEACAM proteins. The 12 CEACAM genes found in humans, including CEACAM1, CEACAM3, CEACAM4, and CEACAM5 (CEA), encode functional proteins. While CEACAM1L, 3L, 4L, and CEACAM18-21 have transmembrane anchors with cytoplasmic tails, CEACAM16 is soluble. CEACAM5-8 exhibit GPI-linkage. The 12 isoforms of CEACAM1 are produced by alternative splicing; the most researched forms have three or four extracellular immunoglobulin-like domains and either a long (L) or short (S) cytoplasmic tail. Both membrane-bound and soluble versions are among the variations, some of which have unidentified functions (44).

The structure of CEACAM1 in humans closely resembles that in mice. It features several Ig-like domains that contribute to its role in various cell functions, a transmembrane segment that anchors it to the cell membrane, and a cytoplasmic tail that varies in length (44). Human CEACAM1 (hCEACAM1) is also glycosylated, which is crucial for its stability and interactions (45).The specific arrangement of its domains and glycosylation enables homophilic binding interactions, which are essential for its structural characterization within the cellular environment. This intricate structure, comprising its Ig-like domain and variable cytoplasmic tail, effectively positions it as a member of the CEA family of glycoproteins (43,46,47).

2.6.3. Functions of CEACAM1 in Mouse

Both CEACAM1 and CEACAM2 are proteins found in cells that enable homophilic cell adhesion without the need for calcium in mice. CEACAM1 functions as a co-inhibitory receptor in immune responses, insulin activity, and blood vessel formation (48–51). It suppresses cytokine production, T-cell growth, and TCR-mediated cytotoxicity (51). Additionally, it is involved in the responses of T-cells, NK cells, and neutrophils (50,51). CEACAM1 also inhibits NK-cell-mediated cytolysis in tumour cells and regulates ROS activity and IL1B production in neutrophils (50,52). In response to EGF, CEACAM1 down-regulates cell proliferation and inhibits migration and scattering. Regarding insulin management, CEACAM1 aids in insulin clearance and regulates lipogenesis in the liver (48). It also plays a role in the transport of bile acids and negatively impacts osteoclastogenesis (53).

CEACAM2, on the other hand, enhances the population of T-cells that control the production and release of IgA, which is associated with resistance to enteropathogens and commensal microbiota. Furthermore, in the context of microbial infection, CEACAM1 acts as a receptor for the spike glycoprotein (S1) of mouse coronavirus (MHV), facilitating viral entry into host cells (54). The distinct functions of CEACAM1 and CEACAM2 highlight their significance in regulating the immune system, cell signalling, and interactions with pathogens.

2.6.4. Functions of CEACAM1 in Humans

In humans, CEACAM1 works as a cell adhesion protein which mediates homophilic cell adhesion independent of calcium (55). In immune responses and insulin action, it acts as a co-inhibitory receptor. It works as an activator in the process

of angiogenesis (56). Its co-inhibitory function depends upon PTPNG1, which suppresses signal transduction of associated receptors by dephosphorylation of its downstream regulation (55). It plays an important role in autoimmunity and anti-tumour immunity by inhibiting the production of TCR-mediated cytotoxicity and cytokines in T-cells by interacting with HAVCR2 (55). In neutrophils, it down-regulates the production of IL1B, which lowers the disruption of lysosomes and ROS, and regulates inflammasome activity. CEACAM1 also manages insulin action by promoting insulin clearance and regulates lipogenesis in the liver through insulin signalling pathways. In addition, it plays a role in vascular permeability, angiogenesis, and down-regulates cell development in response to EGF by interacting with SHC1 and EGFR (57).

2.6.5. Functions of CEACAM1 in Rats

CEACAM1 has similar functions in rats as in mice. It helps in cell adhesion and plays a role in immune response, insulin action, and angiogenesis (58,59). CEACAM1 functions as a phosphorylation and PTPN6-dependent co-inhibitory receptor that influences the activity of T-cells, NK cells, and neutrophils. It aids in insulin clearance and lipogenesis in the liver, leading to the endocytosis and degradation of insulin after phosphorylation via INSR (59). Additionally, CEACAM1 inhibits cell migration and scattering through its interaction with FLNA, while also facilitating blood vessel remodelling and regulating vascular permeability through VEGFR2 signalling (60).

2.7. BINDING AND TRANSMISSION OF MHV

The interaction between MHV and CEACAM1 plays a crucial role in the process of viral infection. CEACAM1, a member of the immunoglobulin (Ig) superfamily, serves as a key receptor for MHV. Unlike some other viruses that rely on an endosomal pathway for entry, MHV utilizes a non-endosomal pathway. This pathway is made possible by the direct binding of MHV's spike protein (S) to the CEACAM1 receptor located on the cell surface (61,62). Initially, the attachment occurs at the N-terminal of the CEACAM1 domain, which contains essential residues such as Ser-32, Tyr-34, Val-39, Gln-44, Gln-89, and Ile-91 that are critical for viral binding (43,47,61,63). This binding not only secures the virus to the host cell but also triggers the fusogenic properties of the S protein, leading to a conformational change

that allows the viral envelope to fuse directly with the host cell's plasma membrane, facilitating the entry of viral RNA into the cytoplasm (61).

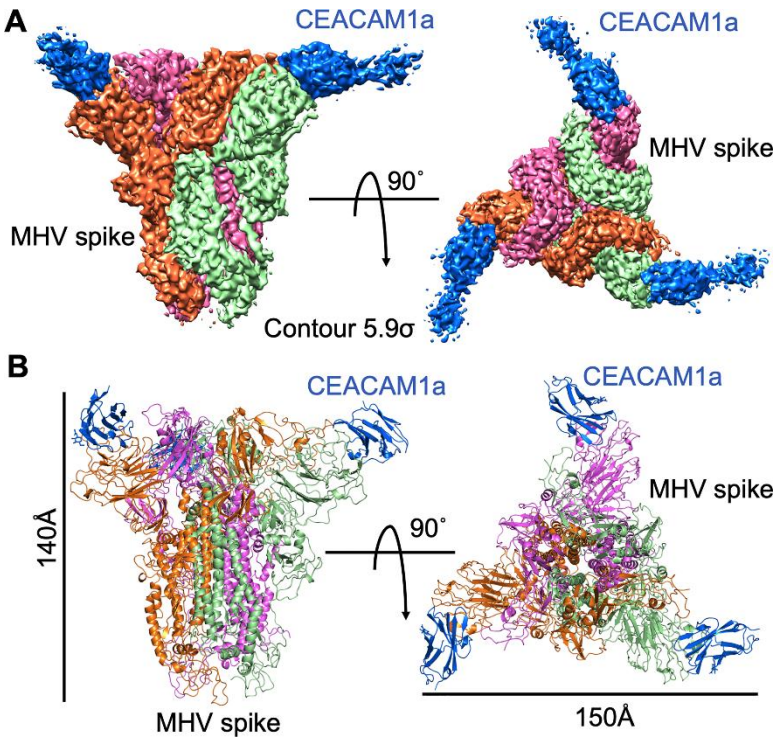

Figure 2.3: MHV Spike Protein/CEACAM1a Complex. (A) Cryo-EM map of the MHV spike protein bound to CEACAM1a, shown from the side and top views. The trimeric spike is in its pre-fusion state, with each subunit uniquely coloured, and CEACAM1a highlighted in blue. (B) Atomic structure of the same complex, matching the views and colours in panel (62).

After binding, the S protein undergoes important changes in its shape, transitioning from a fusion-negative to a fusion-positive condition. This change is necessary for membrane fusion, which allows the viral genome to enter the host cell and start infection. The ability of the N-domain of CEACAM1 to identify and bind with the S protein of MHV is crucial for this structural binding activity. Although the N-domain alone is not sufficient for complete activation of the receptor, other Ig-like domains and associated molecules are also of immense importance for viral entry and subsequent infection. Different types of mice have different sensitivities to MHV. This sensitivity depends on the presence of specific alleles of CEACAM1. The CEACAM1a isoform is found in most laboratory mouse types and is more sensitive to MHV, while the CEACAM1b isoform is found in resistant SJL mice. CEACAM1a shows relatively stronger binding and receptor activity than CEACAM1b (64). Investigation and understanding of the different shapes and special structural interactions between MHV and CEACAM1 are important for the strategic development of targeted cures against MHV infections and related diseases (61).

N-terminal domain or D1 domain of CEACAM1 receptor binds with MHV. Binding residues of D1 for MHV aligned with human CEACAM1, and bovine CEACAM1 are showing in the figure 2.4 and 2.5. Positions “29, 30, 34, 39, 41, 42, 47, 49, 51, 52, 54, 56, 57, 58, 59, 93, 94, and 96” are the contact positions of D1 for MHV binding with CEACAM1 receptor (65). D1 domain is the region of CEACAM1 receptor from amino acid position 35 to 142 of CEACAM1. Hence, the contact residues of CEACAM1 receptor according to the amino acid position in CEACAM1 amino acid sequence are “63, 64, 68, 73, 75, 76, 81, 83, 85, 86, 88, 90,91, 92, 93, 127, 128, 130”.

MHV	Y15	F19	R20	I22	Q23	L24	V25	N26	S27	G29	L89	L160	N172	L174
mCEACAM1a	I41	N94	Y34	G29	R47	M54	R47	F56	T57	Q59	I41	I41	V49	E93
	D42		T39	A30		F56	F56	T57	G58		D42		N51	
			R96	V49					Q59		Q59		S52	

Figure 2.4: Binding sites of MHV with D1 domain of CEACAM1 (65)**.**

```
                        21          31          41          51          91
                        |           |           |           |           |
  Murine-CEACAM1a       VHNLPLALGA FAWYKGNTTA IDKEIARFVP NSNMNFTGQA  TDENYRRTQA
  Murine-CEACAM1b       VHNLPLALGA FAWYKGNPVS TNAEIVHFVT GTNKTTTGPA  TDENFRRTEA
  Bovine-CEACAM1a       AHNVTKNPLG YAWYRGERVD NSQLIASYRV AT---TKGPA  TKDDLQTERQ
  Bovine-CEACAM1b       AQNVTKSPLG YSWYRGERVD NTQLIASYRV DTNATTKGPA  TKDDLQTERP
  Human-CEACEAM1        VHNLPQQLFG YSWYKGERVD GNRQIVGYAI GTQQATPGPA  IKSDLVNEEA
                        .:*:.    . ::**:*: .      *. :    :     * *   ..:    .
```

Figure 2.5: Alignment of D1 domain residues of CEACM1 amino acid residues of murine, bovine and human from position 21-60 and 91-100 of D1. Red and bold residues are the viral contact residues of D1. Positions 29, 30, 34, 39, 41, 42, 47, 49, 51, 52, 54, 56, 57, 58, 59, 93, 94, and 96 are the contact positions of D1 for MHV binding with CEACAM1 receptor(65).

CHAPTER 3

METHODOLOGY

The major steps of this study include the alignment of coding sequences of CEACAM1, their analysis for similarity to humans, calculations of diversity index, and analysing the amino acid residues of CEACAM1 under positive selection and selection pressure on CEACAM1 receptor of different species.

3.1. INSTALLATION OF ALL NECESSARY SOFTWARE AND TOOLS

3.1.1. ALIVIEW alignment tool

For alignment, “Aliview alignment tool” was deployed. To download and install Aliview software following steps were followed.

1. Google web browser was opened.
2. On the search bar “Aliview alignment tool” was searched.
3. The first result shown on the result bar was opened (http link: https://ormbunkar.se/aliview/).
4. Aliview version 1.x for windows was downloaded from download link.
5. “Aliview.exe” for Aliview tool was downloaded in download folder of window.
6. “Aliview.exe” was opened for installation.
7. Windows was allowed to install it by clicking on YES.
8. “Aliview alignment tool” was installed successfully.

3.1.2. Notepad ++

1. To download and install Notepad++, the Google browser was opened.
2. In the search bar, “Notepad++” was entered and searched.
3. The first site on the results page was opened (http link: https://notepad-plus-plus.org/downloads/).
4. The latest version of Notepad++ for 64-bit Windows (version: Notepad++ 64-bit x64 8.6.9) was selected to download.
5. A program file named “npp.8.6.9.Installer.x64.exe” appeared in the Downloads folder.

6. This file was opened, and Windows was allowed to install the app by clicking YES.
7. Notepad++ was successfully installed.

3.1.3. R and R-Studio

R-Studio is a powerful and user-friendly integrated development environment (IDE) that uses the R language for statistical analysis and visualization. R-Studio requires both R and R-Studio to be installed.

1. To download R and R-Studio, the Google browser was opened, and R and R-Studio were searched.
2. The first link on the results page (link https://posit.co/download/rstudio-desktop/) was opened.
3. On this page, there were links to both R and R-Studio: “Download and install R” and “Download R-Studio Desktop for Windows”.
4. The link “Download R-Studio Desktop for Windows” was opened to start downloading R-Studio, and “Download and install R” was opened in a new tab.
5. On the resulting page, another link labelled “Download R for Windows” was selected to start downloading R.
6. Both R (version 4.4.1) and R-Studio were downloaded into the Downloads folder.
7. The folder was opened, and the installer for R was run and allowed to install by clicking YES.
8. Then, the installer for R-Studio was opened and allowed to install by clicking YES.
9. Both R and R-Studio were successfully installed.

3.1.4. Ubuntu

1. To download Ubuntu, the Microsoft Store in Windows 11 was opened, and “Ubuntu” was searched in the search bar.
2. On the results page, Ubuntu 22.04.5 LTS was selected.
3. On the app detail page, the Get button was clicked to download Ubuntu.
4. Once the download was complete, the app was opened, and a login was registered by providing a username and password.

5. Ubuntu was installed successfully.

3.1.5. HyPhy

HyPhy (Hypothesis Testing Using Phylogenies) is a powerful tool and was used for evolutionary analyses, including MEME (Mixed Effects Model of Evolution) and aBSREL (Adaptive Branch-Site Random Effects Likelihood).

1. HyPhy was installed on Ubuntu using Linux commands.
2. Ubuntu was opened.
3. It opened the home directory by default which was "root@DESKTOP-78D5R1A:~$"
4. Below commands were then used for the installation of HyPhy in this directory:

```
sudo apt update (to update the package lists from the ubuntu
   software repositories)
sudo apt upgrade (to upgrade all installed packages)
mkdir hyphy-demo
git clone https://github.com/veg/hyphy.git hyphy-demo
ls
cd hyphy-demo
cmake ..
make -j4
sudo make install
hyphy
```

5. HyPhy version 2.5.62 was installed successfully

3.2. DATA RETRIEVAL

1. CEACAM1 coding sequences were downloaded from the NCBI Nucleotide database for all mammal species, including all available transcript variants.
2. Each sequence was further analysed for CEACAM1 viral binding sites, year of submission, and source of origin.
3. For species with more than one transcript variant, the latest uploaded variant containing viral binding regions and a valid source of origin was selected, providing one sequence per species. A total of 102 CEACAM1 sequences were selected for analysis.
4. After selecting one sequence for each species, an **Accession.txt** file was created, containing all accession numbers.

5. This file was uploaded to the NCBI Batch Entrez database. All retrieved coding sequences (CDS) were downloaded and saved as **CDS_CEACAM1_Sequences.fas**.
6. The same Accession.txt file was re-uploaded to the NCBI Batch Entrez database.
7. A summary file of all sequence entries (with organism names by default) was downloaded and saved as **CEACAM1_Species_Names.txt.**
8. The details in **CDS_CEACAM1_Sequences.fas** were edited to include species names from **CEACAM1_Species_Names.txt** and renamed **cds_names_accession_seq.fas** in FASTA format.
9. The order and family names for each specie were researched in the literature and added alongside the species names.
10. Accession IDs, along with species, order, and family names, are provided in Table 4.1.

3.3. ALIGNMENT

1. The Aliview alignment tool was opened
2. The file **CEACAM1_cds_seq.fas** was uploaded for alignment.
3. Sequences were aligned by codon using the MUSCLE alignment algorithm.
4. Large gaps were removed
5. Small gaps were edited as "N"s using the human mRNA-cds sequence as a guide sequence.
6. All stop codons were removed.
7. The DNA alignment was saved as **CEACAM1_ntd_Align.fas**.
8. Aligned DNA sequences were translated to amino acid sequences using the "Translate protein" function in Aliview and saved as **CEACAM1_AllRes_align.fas.**

3.4. INVESTIGATION OF IMPORTANT BINDING RESIDUES

3.4.1. Preparing files

1. To calculate the similarity of amino acid contact residues with human and mouse sequences and measure the diversity index, R 4.4.1 was used.
2. Excel comma-separated values (.csv) files of amino acid residues and contact residues were created using the following steps:

i. **CEACAM1_AllRes_align.fas** was opened and saved as **AllRes_edited.txt.**
ii. To open AllRes_edited.txt in Excel, some editing was required first.
iii. Each species name, along with its amino acid sequence, was edited to ensure each species and its sequence occupied a single row in the text file, allowing it to open properly in Excel.
iv. Excel was opened, and **AllRes_edited.txt** was imported as a space-delimited file.
v. The same steps were followed to create an Excel file of aligned nucleotide sequences from **CEACAM1_ntd_Align.fas**, saved as **CEACAM1_ntd.csv**.

3. Each row in Excel now contained a species name, Accession ID, and the aligned amino acid sequence.
4. The entire amino acid sequence was in a single cell in the third column.
5. The Character Splitter formula **{=MID ($C2, SEQUENCE (1, LEN($C2)), 1)}** was applied to separate each amino acid letter into individual cells.
6. This formula was entered in cell D2 and dragged down through rows D3, D4, D5, …, D103, so that each letter was in its own cell along the row.
7. Each amino acid residue column was numbered from 1 to 528, representing the 528 amino acid residues of the CEACAM1 receptor for each species.
8. It was confirmed that no "X" residues (indicating unknown amino acids) were present, as they could bias the results.
9. The file was saved as AllRes.csv, and given as “Appendix 8”.
10. A separate file was created containing only the aligned contact/binding residues from the AllRes.csv file.
11. Columns corresponding to positions 63, 64, 68, 73, 75, 76, 81, 83, 85, 86, 88, 90, 91, 92, 93, 127, 128, and 130 were copied into a new comma-separated values (.csv) file.
12. Order and family names were added by copying them from **cds_names_accesion_seq.fas** and pasting them alongside species names.
13. This file was saved as Cres.csv, and given as “Appendix 9”.
14. To calculate the most common amino acid at each position, the frequency mode calculator formula **{=MODE.MULT(F104:W104)}** was used**.**

15. This formula was entered in cell F104 to extract the most frequent amino acid for each position, and the file is given as “appendix 9”.
16. To calculate the similarity score, each amino acid was assigned a score based on its similarity to the reference sequence.
17. For similarity to humans, the human contact residue sequence served as the reference, while the mouse (*Mus musculus*) contact residue sequence was used to calculate similarity to mice.
18. Amino acids that matched the reference sequence were assigned a score of 1, while dissimilar amino acids were assigned a score of -1.
19. Similarity was assessed based on shared properties, including hydrogen bonding, charge (e.g., replacement of a negative with a positive amino acid), and Van Der Waals interactions (17).
20. Table 4.2 shows the amino acids grouped by similarity for score calculations.

3.4.2. Investigation of Similarity Score and Diversity Index

1. A folder named **CEACAM1_Reproduce** was created on the desktop, and the **CRes.csv** file was saved in this folder and renamed as **CEACAM1.csv**.
2. An R script was written in Notepad++ for calculating and analysing similarity scores and Shannon’s diversity index, saved as **diversity_script.R** in a new folder on the desktop named **Diversity_Reproduce**.
3. The R script is provided in the appendices as **Appendix 1: diversity_script.R**.
4. Running the commands as in **Appendix 1** produced two output files: **Shannons_Index.csv** and **CEACAM1_Likeliness_Score.csv**, representing the Shannon Diversity index and similarity scores, respectively.

3.5. MOLECULAR EVOLUTION ANALYSIS

To better understand the selection pressure on the CEACAM1 receptor, the analysis was conducted in three sub-sections:

1. Across all species.
2. Only within the order Primates.
3. For species outside the order Primates.

Both MEME and aBSREL analyses required a Maximum Clade Credibility (MCC) phylogenetic tree in NEWICK format and a relevant codon alignment file in FASTA format.

3.5.1. Maximum Clade Credibility (MCC) Trees

To accurately analyse evolutionary patterns and selection pressures on genes of interest, a time tree was employed. This tree provides a calibrated timeline for the evolutionary history of study organisms. By integrating this temporal information into HyPhy analyses, more precise estimates of the timing of key evolutionary events and rates of molecular evolution across lineages can be obtained.

MCC trees are crucial for evolutionary analysis as they provide insights into the timing of species divergence. These trees also offer the most credible branching patterns, enhancing the confidence in evolutionary inferences. the MCC trees were extracted from a large mammalian clade tree provided by Upham et al. (66).

3.5.1.1. Extracting and Pruning of MCC Trees

1. A new folder named **CEACAM1_SP** was created on the desktop.
2. The Mammalian MCC tree was downloaded from the supplementary data of the article **"Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation"** by Upham et al. (66). This was saved as **Upham_MCC_no_fossil.tre.txt** in the CEACAM1_SP folder.
3. The **CEACAM1_ntd.csv** codon alignment file, created in the alignment section, was copied into the CEACAM1_SP folder and renamed to **CEACAM1_Upham_Names.csv.**
4. R-Studio was opened, and the MCC trees were extracted using the R commands provided in the appendices as **Appendix 2: Pruned_MCC.R**.

5. The following MCC trees were extracted:
    i. **Upham_MCC_CEACAM1.newick**
    ii. **Upham_MCC_CEACAM1_Primates.newick**
    iii. **Upham_MCC_CEACAM1_Non-Primates.newick**
6. A new folder named **HyPhy_Reproduce** was created on the desktop.
7. The three extracted MCC phylogenetic trees were copied into this folder.
8. Out of the 102 species initially selected, only 91 species were accepted for the extraction of MCC trees.
9. Four species were excluded due to repetition, and seven species were not included in the parent **Upham_MCC_no_fossil.tre.txt tree.**

3.5.2. To ready Alignment files

1. The alignment file, "**CEACAM1_ntd_Align.fas",** from the alignment section was opened and saved as "**CEACAM1_Alignment_All.fas".**
2. The file "**CEACAM1_Alignment_All.fas**" was filtered for primate species only, and the resulting alignment file was saved in FASTA format as "**Primates_Alignment.fas".**
3. " **CEACAM1_Alignment_All.fas "** was then filtered to exclude non-primates, and the alignment of all remaining species was saved as " **Non-Primates_Alignment.fas** " in FASTA format.
4. All three alignment files were saved in a folder named "**HyPhy_Reproduce**" on the Desktop.

3.5.3. Mixed effects model of evolution (MEME) analysis

1. MEME analysis was conducted using HyPhy (Hypothesis Testing Using Phylogenies) version 2.5.62. The steps below outline the procedure.
2. The Ubuntu terminal was opened by typing "ubuntu" in the search bar of the Start menu.
3. The HyPhy directory was accessed using the following command:
    i. `cd hyphy-demo`
4. To start MEME analysis on the CEACAM1 receptor for all 91 species, the following Linux command was executed:
    i. `hyphy ~/hyphy-demo/res/TemplateBatchFiles/SelectionAnalyses/MEME.bf /mnt/c/users/dell/desktop/HyPhy_Reproduce/CEACAM1_Alignment_All.`

```
fas
/mnt/c/users/dell/desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_2024
-10-20.newick ENV=TOLERATE_NUMERICAL_ERRORS=1
```

5. This command produced an output file named "**CEACAM1_Alignment_All.fas.MEME.json",** which was renamed " **CEACAM1_MEME_All.json".**
6. To run MEME analysis on the CEACAM1 receptor for primates only, the following command was used:
   i.
```
hyphy ~/hyphy-
demo/res/TemplateBatchFiles/SelectionAnalyses/MEME.bf
/mnt/c/users/dell/desktop/HyPhy_Reproduce/Primates_Alignment_25-
01-25.fas
/mnt/c/users/dell/desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_Prim
ate_2025-01-25.newick ENV=TOLERATE_NUMERICAL_ERRORS=1
```

7. This produced an output file named " **Primates_Alignment.fas.MEME.json",** which was renamed " **Primates_MEME.json".**
8. To conduct MEME analysis on CEACAM1 for species other than primates, the following command was run:
   i.
```
hyphy ~/hyphy-
demo/res/TemplateBatchFiles/SelectionAnalyses/MEME.bf
/mnt/c/users/dell/desktop/HyPhy_Reproduce/Non-
primates_Alignment_25-01-25.fas
/mnt/c/users/dell/desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_Non-
Primate_2025-01-25.newick ENV=TOLERATE_NUMERICAL_ERRORS=1
```

9. This generated an output file named "**Non-primates_Alignment.fas.MEME.json**", which was renamed **"Non-primates_MEME.json".**
10. Each MEME analysis took approximately 4 to 6 hours, depending on the number of species included.

3.5.4. Adaptive branch site random effect likelihood (aBSREL) analysis

The aBSREL analysis was also conducted using HyPhy version 2.5.62. The steps below outline the procedure.

1. The Ubuntu terminal was opened by typing "Ubuntu" in the Start menu search bar.
2. The HyPhy directory was accessed with the command:
   i.
```
cd hyphy-demo
```

3. To start aBSREL analysis on the CEACAM1 receptor for all 91 species, the following command was executed:
    i. `hyphy ~/hyphy-demo/res/TemplateBatchFiles/SelectionAnalyses/aBSREL.bf /mnt/c/users/dell/desktop/HyPhy_Reproduced/HyPhy_Align_All_91_22-10-24.fas /mnt/c/users/dell/desktop/HyPhy_Reproduced/Upham_MCC_renamed_91spp_forCEACAM1_2024-10-20.newick ENV=TOLERATE_NUMERICAL_ERRORS=1`
4. This produced an output file named "**CEACAM1_Alignment_All.fas.aBSREL.json**", which was renamed "**CEACAM1_All_aBSREL.json**".
5. To conduct aBSREL analysis on the CEACAM1 receptor for order Primates only, the following command was used:
    i. `hyphy ~/hyphy-demo/res/TemplateBatchFiles/SelectionAnalyses/aBSREL.bf /mnt/c/users/dell/desktop/HyPhy_Reproduce/Primates_Alignment_25-01-25.fas /mnt/c/users/dell/desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_Primate_2025-01-25.newick ENV=TOLERATE_NUMERICAL_ERRORS=1`
6. This generated an output file named "**Primates_Alignment.fas.aBSREL.json",** which was renamed "**Primates_aBSREL.json**".
7. To run aBSREL analysis on CEACAM1 for species other than Primates, the following command was executed:
    i. `hyphy ~/hyphy-demo/res/TemplateBatchFiles/SelectionAnalyses/aBSREL.bf /mnt/c/users/dell/desktop/HyPhy_Reproduce/Non-primates_Alignment_25-01-25.fas /mnt/c/users/dell/desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_Non-Primate_2025-01-25.newick ENV=TOLERATE_NUMERICAL_ERRORS=1`
8. This generated an output file named **"Non-Primates_Alignment.fas.aBSREL.json",** which was renamed "**Non-Primates_aBSREL.json**".
9. Each aBSREL analysis took approximately 8 to 10 hours, depending on the number of species analysed.

3.6. EXTRACTION OF MEME AND ABSREL RESULTS FROM JSON (JAVASCRIPT OBJECT NOTATION) FILES TO CSV EXCEL FILES

HyPhy generated output files in JSON format. To make the data more accessible, the JSON files were converted into CSV format using R.

1. The R script, titled "**Appendix 3: Extraction_JSONS_Hyphy.R**", was used to extract data from all JSON files and save them as CSV files.
2. The resulting CSV files were saved with the following names:
    i. meme_91_summary_analysis.csv
    ii. meme_68_summary_analysis.csv
    iii. meme_23_summary_analysis.csv
    iv. whole_gene_absrel.csv
    v. non_primates_absrel.csv
    vi. primates_absrel.csv
3. MEME analysis was conducted at two significance levels: $p < 0.05$ and $p < 0.1$.
4. Results were then analysed to identify the following categories across the three MEME result files (for all species, primates only, and species other than primate mammals):
    i. Contact or binding residues under positive selection
    ii. Non-contact residues under positive selection
    iii. Contact residues not under positive selection
    iv. Non-contact residues not under positive selection
5. Fisher's exact tests, along with odds ratio calculations, were performed to evaluate significant associations between the following groups:
    i. For all species at $p < 0.05$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not under positive selection vs. non-contact residues not under positive selection.
    ii. For all species at $p < 0.1$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not under positive selection vs. non-contact residues not under positive selection.
    iii. For primates only at $p < 0.05$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not

under positive selection vs. non-contact residues not under positive selection.

iv. For primates only at $p < 0.1$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not under positive selection vs. non-contact residues not under positive selection.

v. For species other than primates at $p < 0.05$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not under positive selection vs. non-contact residues not under positive selection.

vi. For species other than primates at $p < 0.1$: Contact residues under positive selection vs. non-contact residues under positive selection, and contact residues not under positive selection vs. non-contact residues not under positive selection.

6. A results table for the MEME analysis was created and saved as Table S4.
7. All analyses were conducted using R version 4.4.1, with the relevant R script provided in the appendices under the title "**Appendix 4: MEME_Result_Analysis_Script.R**."
8. For the aBSREL analysis, a significance threshold of $p < 0.05$ was applied, without correction. Branches with $p < 0.05$ were identified as being under positive selection.
9. aBSREL analysis was conducted across all three files (whole gene, primates only, and species other than primates), focusing on both terminal and non-terminal branches.
10. Fisher's exact tests and odds ratios were calculated to examine associations between the following groups:

    i. The number of primate branches under selection vs. primate branches not under selection, compared to non-primate branches under selection vs. non-primate branches not under selection.

    ii. The number of terminal primate branches under selection vs. terminal primate branches not under selection, compared to terminal non-primate branches under selection vs. terminal non-primate branches not under selection.

11. The results of the aBSREL analysis were saved as a CSV file. All analyses were conducted in R version 4.4.1.
12. The R script for the aBSREL analysis is provided in the appendices as "**Appendix 5: aBSREL_Result_Analysis_Script.R**."

3.7. VARIABILITY AND COMPLETENESS ANALYSIS OF SAMPLE DATASET

The independent distribution of codon base pairs between primate and non-primate species was analysed using the Wilcoxon Rank-Sum Test. To assess the overall distribution shape across the entire dataset, a Shapiro-Wilk test was performed using R version 4.4.1. The R script used for these analyses is provided in the appendices under the title "**Appendix 6: Variable_sites_and_Completeness.R**."

3.8 FIGURES AND GRAPHS

1. To analyse the similarity score of binding residues in relation to human and mouse binding residues, a boxplot was created with overlaid jittered points (Figure 4.1) using R version 4.4.1.
2. The R script used to generate this figure is included in the appendices as "**Appendix 7: Figure.R**."
3. Another figure was created to illustrate selection pressure on branches of the phylogenetic tree.
4. For this figure, an online tool iTOL (Interactive Tree of Life) was used to generate a circular phylogenetic tree. Different branch colours were used to indicate selection pressure at $p < 0.05$ (link: https://itol.embl.de/).
5. The phylogenetic tree file **"Upham_MCC_CEACAM1_2024-10-20.newick"** was uploaded to "iTOL".
6. The circular view style was selected for the tree visualization.
7. In the final figure, branches under positive selection were marked with black tip labels, non-selected primate branches were marked with blue tip labels, and non-primate branches were marked with grey tip labels.
8. The final figure was downloaded as both a PDF and a PNG image file.

CHAPTER 4

RESULT

4.1 Retrieved Data

Coding sequences were retrieved from NCBI (http link: https://www.ncbi.nlm.nih.gov/nucleotide/) . Accession IDs, along with species, order, and family names, are provided in Table 4.1.

Table 4.1: Accession IDs, Specie name, order and family name of CEACAM1 coding sequences.

Sr.	**Accession IDs**	***Specie Names***	**Order**	**Family**
1	AF259566.1	*Cercopithecus aethiops*	Primates	Cercopithecidae
2	MF564057.1	*Equus caballus*	Perissodactyla	Equidae
3	MG874670.1	*Macaca mulatta*	Primates	Cercopithecidae
4	NM_001024912.3	*Homo sapiens*	Primates	Hominidae
5	NM_001033860.1	*Rattus norvegicus*	Rodentia	Muridae
6	NM_001039185.1	*Mus musculus*	Rodentia	Muridae
7	NM_001097557.1	*Canis lupus familiaris*	Carnivora	Canidae
8	NM_001132423.1	*Pongo abelii*	Primates	Hominidae
9	NM_001712.5	*Homo sapiens*	Primates	Hominidae
10	XM_002762196.3	*Callithrix jacchus*	Primates	Callitrichidae
11	XM_003915629.4	*Papio Anubis*	Primates	Cercopithecidae
12	XM_003943210.3	*Saimiri boliviensis*	Primates	Cebidae
13	XM_004015280.6	*Ovis aries*	Cetartiodactyla	Bovidae
14	XM_004670324.2	*Jaculus*	Rodentia	Dipodidae
15	XM_004780425.3	*Mustela putorius*	Carnivora	Mustelidae
16	XM_004873005.2	*Heterocephalus glaber*	Rodentia	Heterocephalidae

Sr.	Accession IDs	*Specie Names*	Order	Family
17	XM_005338246.4	*Ictidomys tridecemlineatus*	Rodentia	Sciuridae
18	XM_005412355.1	*Chinchilla lanigera*	Rodentia	Chinchillidae
19	XM_005589370.4	*Macaca fascicularis*	Primates	Cercopithecidae
20	XM_005596278.3	*Equus caballus*	Perissodactyla	Equidae
21	XM_005655889.2	*Sus scrofa*	Cetartiodactyla	Suidae
22	XM_006041224.4	*Bubalus bubalis*	Cetartiodactyla	Bovidae
23	XM_006871507.1	*Chrysochloris asiatica*	Afrotheria	Chrysochloridae
24	XM_007086642.2	*Panthera tigris*	Carnivora	Felidae
25	XM_007532949.3	*Erinaceus europaeus*	Eulipotyphla	Erinaceidae
26	XM_007997002.2	*Chlorocebus sabaeus*	Primates	Cercopithecidae
27	XM_008249692.3	*Oryctolagus cuniculus*	Lagomorpha	Leporidae
28	XM_008709687.2	*Ursus maritimus*	Carnivora	Ursidae
29	XM_008964144.4	*Pan paniscus*	Primates	Hominidae
30	XM_010358723.2	*Rhinopithecus roxellana*	Primates	Cercopithecidae
31	XM_010623328.3	*Fukomys damarensis*	Rodentia	Bathyergidae
32	XM_010842547.1	*Bison*	Cetartiodactyla	Bovidae
33	XM_011289901.4	*Felis catus*	Carnivora	Felidae
34	XM_011764531.2	*Macaca nemestrina*	Primates	Cercopithecidae
35	XM_011936539.1	*Colobus angolensis*	Primates	Cercopithecidae
36	XM_012086961.1	*Cercocebus atys*	Primates	Cercopithecidae
37	XM_012461510.3	*Aotus nancymaae*	Primates	Aotidae
38	XM_012653468.1	*Propithecus coquereli*	Primates	Indriidae

Sr.	Accession IDs	*Specie Names*	Order	Family
39	XM_012733660.1	*Condylura cristata*	Eulipotyphla	Talpidae
40	XM_015496801.2	*Marmota marmota*	Rodentia	Sciuridae
41	XM_015562203.1	*Myotis davidii*	Chiroptera	Vespertilionidae
42	XM_016129903.2	*Rousettus aegyptiacus*	Chiroptera	Pteropodidae
43	XM_017849842.1	*Rhinopithecus bieti*	Primates	Cercopithecidae
44	XM_019434422.2	*Panthera pardus*	Carnivora	Felidae
45	XM_020871207.1	*Odocoileus virginianus*	Cetartiodactyla	Cervidae
46	XM_023229215.2	*Piliocolobus tephrosceles*	Primates	Cercopithecidae
47	XM_023704520.1	*Octodon degus*	Rodentia	Octodontidae
48	XM_024132797.3	*Physeter catodon*	Cetartiodactyla	Physeteridae
49	XM_024577349.3	*Desmodus rotundus*	Chiroptera	Phyllostomidae
50	XM_024977969.2	*Bos taurus*	Cetartiodactyla	Bovidae
51	XM_025366225.1	*Theropithecus gelada*	Primates	Cercopithecidae
52	XM_026403696.1	*Urocitellus parryii*	Rodentia	Sciuridae
53	XM_026482217.4	*Ursus arctos*	Carnivora	Ursidae
54	XM_027858153.1	*Vombatus ursinus*	Diprotodontia	Vombatidae
55	XM_028530369.2	*Phyllostomus discolor*	Chiroptera	Phyllostomidae
56	XM_028860001.2	*Peromyscus leucopus*	Rodentia	Cricetidae
57	XM_029570052.1	*Nannospalax galili*	Rodentia	Spalacidae
58	XM_030297609.1	*Lynx canadensis*	Carnivora	Felidae
59	XM_030797105.1	*Nomascus leucogenys*	Primates	Hylobatidae
60	XM_031004080.3	*Gorilla gorilla*	Primates	Hominidae

Sr.	Accession IDs	*Specie Names*	Order	Family
61	XM_031681826.1	*Vicugna pacos*	Cetartiodactyla	Camelidae
62	XM_032172816.2	*Hylobates moloch*	Primates	Hylobatidae
63	XM_032894175.1	*Rattus rattus*	Rodentia	Muridae
64	XM_033197411.1	*Trachypithecus francoisi*	Primates	Cercopithecidae
65	XM_036168114.1	*Onychomys torridus*	Rodentia	Cricetidae
66	XM_036346012.1	*Myotis myotis*	Chiroptera	Vespertilionidae
67	XM_036833342.1	*Balaenoptera musculus*	Cetartiodactyla	Balaenopteridae
68	XM_036896477.2	*Manis pentadactyla*	Pholidota	Manidae
69	XM_037013906.1	*Manis javanica*	Pholidota	Manidae
70	XM_037516693.2	*Talpa occidentalis*	Eulipotyphla	Talpidae
71	XM_039912179.1	*Ornithorhynchus anatinus*	Monotremata	Ornithorhynchidae
72	XM_040634031.1	*Ursus maritimus*	Carnivora	Ursidae
73	XM_040734318.1	*Mesocricetus auratus*	Rodentia	Cricetidae
74	XM_041654886.1	*Microtus oregoni*	Rodentia	Cricetidae
75	XM_042267688.1	*Peromyscus maniculatus*	Rodentia	Cricetidae
76	XM_042668120.1	*Dipodomys spectabilis*	Rodentia	Heteromyidae
77	XM_045292270.1	*Echinops telfairi*	Afrosoricida	Tenrecidae
78	XM_048440439.1	*Myodes glareolus*	Rodentia	Cricetidae
79	XM_049621270.1	*Panthera uncia*	Carnivora	Felidae
80	XM_049703839.1	*Orcinus orca*	Cetartiodactyla	Delphinidae
81	XM_051207508.1	*Phodopus roborovskii*	Rodentia	Cricetidae

Sr.	Accession IDs	*Specie Names*	Order	Family
82	XM_052168799.1	*Apodemus sylvaticus*	Rodentia	Muridae
83	XM_052655777.1	*Budorcas taxicolor*	Cetartiodactyla	Bovidae
84	XM_053657140.1	*Artibeus jamaicensis*	Chiroptera	Phyllostomidae
85	XM_054463351.2	*Pongo pygmaeus*	Primates	Hominidae
86	XM_054567634.1	*Pteronotus mesoamericanus*	Chiroptera	Mormoopidae
87	XM_055145459.1	*Sorex Araneus*	Eulipotyphla	Soricidae
88	XM_055311974.1	*Nyctereutes procyonoides*	Carnivora	Canidae
89	XM_055412081.1	*Moschus berezovskii*	Cetartiodactyla	Moschidae
90	XM_055551361.1	*Bubalus kerabau*	Cetartiodactyla	Bovidae
91	XM_055620670.1	*Psammomys obesus*	Rodentia	Muridae
92	XM_056135625.1	*Sorex fumeus*	Eulipotyphla	Soricidae
93	XM_059253485.1	*Peromyscus eremicus*	Rodentia	Cricetidae
94	XM_062175861.1	*Lepus europaeus*	Lagomorpha	Leporidae
95	XM_063252825.1	*Cavia porcellus*	Rodentia	Caviidae
96	XM_063252826.1	*Cavia porcellus*	Rodentia	Caviidae
97	XM_063801767.1	*Pan troglodytes*	Primates	Hominidae
98	XM_064293748.1	*Loxodonta Africana*	Proboscidea	Elephantidae
99	XM_064489351.1	*Camelus dromedarius*	Cetartiodactyla	Camelidae
100	XM_064581140.1	*Mirounga angustirostris*	Carnivora	Phocidae
101	XM_066271982.1	*Saccopteryx bilineata*	Chiroptera	Emballonuridae
102	XM_066373583.1	*Saccopteryx leptura*	Chiroptera	Emballonuridae

4.2. ANALYSIS of BINDING RESIDUES

4.2.1. Diversity analysis of binding residues

Unique amino acid combination at contact sites of CEACAM1 along all mammal sequences are shown in table 4.2. Mammals are mostly diversified on contact position of CEACAM1. Out of 102 sequences there were 89 unique amino acid combinations. *Bubalus bubalus* and *Bubalus kerabau* (n=48), *Rhinopithecus bieti* and *Rhinopithecus roxellana* (n=50), *Pongo abelii* and *Pongo pygmaeus* (n=51), and *Peromyscus leucopus* and *Peromyscus maniculatus* (n=78) were having identical amino acid combinations on CEACAM1 binding sites (Table 4.2).

Table 4.2: Unique amino acid combination at contact sites of CEACAM1

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
MF564057.1	*Equus caballus*	AAYVHHSKGPETPGPTSQ	1
XM_005596278.3	*Equus caballus*	AAYVPHSVNTQTPGPANQ	2
XM_045292270.1	*Echinops telfairi*	AFYVKNTASSITRGPADV	3
XM_066373583.1	*Saccopteryx leptura*	AGYVSNSVATETPGPKNL	4
XM_055311974.1	*Nyctereutes procyonoides*	ASFVPALVETITPGPRKE	5
NM_001097557.1	*Canis lupus familiaris*	ASFVPTLVDTITPGPQKE	6
XM_042668120.1	*Dipodomys spectabilis*	ELYTESASNSQTPGPRDA	7
XM_011764531.2	*Macaca nemestrina*	FGYVAKAVGTQTPGPGDV	8
XM_033197411.1	*Trachypithecus francoisi*	FGYVAKAVRTQTPGPEDV	9
NM_001024912.3	*Homo sapiens*	FGYVGNGAGTQTPGPSDV	10

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
NM_001712.5	*Homo sapiens*	FGYVGNGAGTQTPGPSDV	10
XM_031004080.3	*Gorilla gorilla*	FGYVGNGVATQTPGPSDV	11
XM_063801767.1	*Pan troglodytes*	FGYVGNGVGTQTPGPSDV	12
NM_001039185.1	*Mus musculus*	GAYTIDRVNSMFTGQENR	13
XM_012733660.1	*Condylura cristata*	GIYVGESKGSAGTGPRDE	14
XM_054567634.1	*Pteronotus mesoamericanus*	GKFVTSSRDNETRGPKHS	15
XM_004780425.3	*Mustela putorius*	GRYTENSIDTKVIGPRTE	16
XM_024577349.3	*Desmodus rotundus*	GRYVSKSMETAVPGPKDS	17
XM_010623328.3	*Fukomys damarensis*	IAFVSNALADTTQGPPNL	18
XM_023704520.1	*Octodon degus*	IAFVSSTEDTRTPGRSNQ	19
XM_005338246.4	*Ictidomys tridecemlineatus*	IGFANRSMSTLTPGSANR	20
XM_030297609.1	*Lynx canadensis*	IGFIPHSADLETPGFRNQ	21
XM_011289901.4	*Felis catus*	IGFIPRSADSETLGFRNQ	22
XM_005412355.1	*Chinchilla lanigera*	IGFSSSAVSTVTPGPDNW	23
XM_026403696.1	*Urocitellus parryii*	IGFTDSSSLTNNTGPNDR	24
XM_064581140.1	*Mirounga angustirostris*	IGFVPRLADTVTPGPRDQ	25
XM_007532949.3	*Erinaceus europaeus*	IGFVSAGEATQTTGPSGQ	26
XM_012086961.1	*Cercocebus atys*	IGHVAKAVGTQTPGPEDV	27

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
XM_025366225.1	*Theropithecus gelada*	IGHVAKAVGTQTPGPGDV	28
MG874670.1	*Macaca mulatta*	IGHVAKAVRTQTPGPGDV	29
XM_005589370.4	*Macaca fascicularis*	IGYVAKAVESQTPGPGDV	30
XM_003915629.4	*Papio anubis*	IGYVAKAVGTQTPGPEDV	31
XM_032172816.2	*Hylobates moloch*	IGYVANAVENQSLGPLDV	32
XM_030797105.1	*Nomascus leucogenys*	IGYVGNAAGTQTPGPSDV	33
XM_008964144.4	*Pan paniscus*	IGYVGNAAQNQTRGPSDV	34
XM_066271982.1	*Saccopteryx bilineata*	IGYVSNSAATETPGPQNL	35
XM_049621270.1	*Panthera uncia*	LGFIPHSADSETPGFRNQ	36
XM_019434422.2	*Panthera pardus*	LGFIPRSADSETPGFRNQ	37
XM_007086642.2	*Panthera tigris*	LGFIPRSADSETPGLRNQ	38
XM_063252825.1	*Cavia porcellus*	LGFTSDRVNNIVTGKETD	39
XM_026482217.4	*Ursus arctos*	LGFVASSVDPETPGPKNQ	40
XM_031681826.1	*Vicugna pacos*	LGFVNSSVSSLIPGPDNS	41
XM_008709687.2	*Ursus maritimus*	LGFVTSSVDTVTPGPKNQ	42
XM_007997002.2	*Chlorocebus sabaeus*	LGHVAKAVGTQTPGPQDV	43
AF259566.1	*Cercopithecus aethiops*	LGHVANAVGTQTPGPQDV	44
XM_004015280.6	*Ovis aries*	LGHVNASRDNATNGSNDQ	45

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
XM_052655777.1	*Budorcas taxicolor*	LGHVNASRGNATNGPEDQ	46
XM_055412081.1	*Moschus berezovskii*	LGHVNSSRDTVTRGPVDQ	47
XM_006041224.4	*Bubalus bubalis*	LGHVNTSRATLTKGPDDQ	48
XM_055551361.1	*Bubalus kerabau*	LGHVNTSRATLTKGPDDQ	48
XM_023229215.2	*Piliocolobus tephrosceles*	LGYVAKAVETRTPGPEDV	49
XM_017849842.1	*Rhinopithecus bieti*	LGYVAKAVRTQTPGPEDV	50
XM_010358723.2	*Rhinopithecus roxellana*	LGYVAKAVRTQTPGPEDV	50
NM_001132423.1	*Pongo abelii*	LGYVANGVSDLTPGPSDV	51
XM_054463351.2	*Pongo pygmaeus*	LGYVANGVSDLTPGPSDV	51
XM_037013906.1	*Manis javanica*	LGYVFNSENKATPGIGSL	52
XM_036896477.2	*Manis pentadactyla*	LGYVFNSVNTATPGIGSL	53
XM_036833342.1	*Balaenoptera musculus*	LGYVKNSRDVTVTGPNDQ	54
XM_011936539.1	*Colobus angolensis*	LGYVNKAVRTQTPGPEDV	55
XM_049703839.1	*Orcinus orca*	LGYVNNSRDVTTTGLNDQ	56
XM_005655889.2	*Sus scrofa*	LGYVNNTRDTATQGPNDQ	57
XM_010842547.1	*Bison bison*	LGYVNSSRATLTKGPDDQ	58
XM_020871207.1	*Odocoileus virginianus*	LGYVNSSRDTVTTGSNDQ	59
XM_024977969.2	*Bos taurus*	LGYVNTSRDTATKGPDDQ	60

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
XM_012653468.1	*Propithecus coquereli*	LGYVSNAVSTVNTGPADQ	61
XM_024132797.3	*Physeter catodon*	LGYVSNSRDGAATGPNDQ	62
XM_037516693.2	*Talpa occidentalis*	LKYVKSSRDTNIPGPNNL	63
XM_036346012.1	*Myotis myotis*	LLYISNSVDTANFGPKSK	64
XM_028530369.2	*Phyllostomus discolor*	LTFASSAVQTKTPGPKDL	65
XM_002762196.3	*Callithrix jacchus*	NGYVSTGVASLTPGPADV	66
XM_029570052.1	*Nannospalax galili*	NIYTDDRLSTEVEGLSRN	67
XM_062175861.1	*Lepus europaeus*	PVYAPNTVSLNAPGPILV	68
XM_059253485.1	*Peromyscus eremicus*	QAYVDSRLSITKTGPEND	69
XM_064489351.1	*Camelus dromedarius*	QGFVNSSISSAVPGPDKI	70
XM_041654886.1	*Microtus oregoni*	QVHTNSRVSDVKTGPQGD	71
XM_051207508.1	*Phodopus roborovskii*	QVYGESRISTMTTGPIND	72
XM_040734318.1	*Mesocricetus auratus*	QVYLDSRISSTSTGPAND	73
NM_001033860.1	*Rattus norvegicus*	QVYLPDRISDMKTGPQQN	74
XM_032894175.1	*Rattus rattus*	QVYTLNRISDTKTGPQDN	75
XM_052168799.1	*Apodemus sylvaticus*	QVYTVNRVQTRTMGPPDN	76

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
XM_004873005.2	*Heterocephalus glaber*	RAFTEDTMDETLEGPRNL	77
XM_028860001.2	*Peromyscus leucopus*	RAYVDSRLSSTKTGPEND	78
XM_042267688.1	*Peromyscus maniculatus*	RAYVDSRLSSTKTGPEND	78
XM_015562203.1	*Myotis davidii*	RGYISESVDTTTLGPENQ	79
XM_003943210.3	*Saimiri boliviensis*	TGHVSTGVATLTRGPANV	80
XM_012461510.3	*Aotus nancymaae*	TGYVSTGVTTQIPGPADV	81
XM_063252826.1	*Cavia porcellus*	VGFTPNRVTNVVPGKETD	82
XM_053657140.1	*Artibeus jamaicensis*	VGYTNNSVQTKTPGSESQ	83
XM_015496801.2	*Marmota marmota*	VGYTQNSSLTSIQGPPNQ	84
XM_016129903.2	*Rousettus aegyptiacus*	VGYVSNSVENTVPGPKNL	85
XM_055145459.1	*Sorex araneus*	VRCISKTKTPNTNGPMGD	86
XM_006871507.1	*Chrysochloris asiatica*	YGYALSALENSALGVNDQ	87
XM_004670324.2	*Jaculus jaculus*	YGYVESRISTIVTGPEVQ	88
XM_064293748.1	*Loxodonta africana*	Residues with gaps or missing Sequence	89
XM_048440439.1	*Myodes glareolus*	Residues with gaps or missing Sequence	89
XM_036168114.1	*Onychomys torridus*	Residues with gaps or missing Sequence	89

Accession	*Specie*	Contact Residue Sequence	Unique Type (n)
XM_039912179.1	*Ornithorhynchus anatinus*	Residues with gaps or missing Sequence	89
XM_008249692.3	*Oryctolagus cuniculus*	Residues with gaps or missing Sequence	89
XM_055620670.1	*Psammomys obesus*	Residues with gaps or missing Sequence	89
XM_056135625.1	*Sorex fumeus*	Residues with gaps or missing Sequence	89
XM_040634031.1	*Ursus maritimus*	Residues with gaps or missing Sequence	89
XM_027858153.1	*Vombatus ursinus*	Residues with gaps or missing Sequence	89

Table 4.3: Number of each amino acid at contact position in all mammals. Highest number of each Amino acid is highlighted with Brown Colour.

Residue Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
63	6	0	0	1	6	5	0	18	0	33	0	2	1	8	4	0	2	5	0	2
64	9	0	0	0	1	64	0	2	2	2	0	0	0	0	3	2	1	7	0	0
68	0	1	0	0	23	0	12	0	0	0	0	0	0	0	0	0	0	0	0	57
73	4	0	0	0	0	1	0	8	0	2	0	0	0	0	0	1	13	64	0	0
75	16	0	6	5	2	7	1	1	3	2	0	16	12	1	0	18	2	1	0	0
76	4	0	5	2	0	0	4	0	14	0	0	30	0	0	5	22	7	0	0	0
81	22	0	0	0	0	10	0	0	0	3	0	0	0	0	14	38	6	0	0	0
83	12	0	0	3	0	0	0	7	3	6	3	0	0	0	14	3	0	42	0	0
85	10	0	24	7	0	13	0	0	0	2	0	6	0	4	5	19	3	0	0	0
86	0	0	6	1	0	1	0	1	1	2	0	9	3	0	0	15	52	2	0	0
88	11	0	0	11	0	0	0	5	3	9	3	4	0	23	3	2	11	8	0	0
90	3	0	0	0	1	1	0	4	6	1	0	3	0	0	0	2	63	9	0	0
91	0	0	0	2	1	0	0	1	4	4	1	3	50	3	5	0	19	0	0	0
92	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93	0	0	0	0	4	0	0	2	2	3	0	0	75	1	1	4	0	1	0	0
127	8	0	7	17	0	6	0	2	8	1	1	9	3	7	11	11	1	1	0	0

Residue Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
128	0	0	47	0	0	3	1	0	3	1	0	27	0	1	1	5	3	1	0	0
130	1	0	9	4	0	0	0	1	1	9	0	4	0	30	3	3	0	27	1	0

Table 4.4: Number of each amino acid at contact position in non-primate mammals. Highest number of each Amino acid is highlighted with Brown Colour.

Residue Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
63	6	0	0	1	0	5	0	10	0	24	0	1	1	8	4	0	0	5	0	2
64	9	0	0	0	1	38	0	2	2	2	0	0	0	0	3	2	1	7	0	0
68	0	1	0	0	23	0	6	0	0	0	0	0	0	0	0	0	0	0	0	37
73	4	0	0	0	0	1	0	8	0	2	0	0	0	0	0	1	13	38	0	0
75	1	0	6	5	2	1	1	1	3	2	0	15	12	1	0	14	2	1	0	0
76	4	0	5	2	0	0	4	0	2	0	0	19	0	0	5	22	4	0	0	0
81	5	0	0	0	0	1	0	0	0	3	0	0	0	0	14	38	6	0	0	0
83	8	0	0	3	0	0	0	7	3	6	3	0	0	0	14	3	0	20	0	0
85	7	0	24	4	0	3	0	0	0	2	0	6	0	3	0	16	2	0	0	0
86	0	0	4	1	0	1	0	1	1	2	0	7	3	0	0	13	32	2	0	0
88	11	0	0	11	0	0	0	5	3	5	3	4	0	3	2	2	11	7	0	0
90	3	0	0	0	1	1	0	3	6	1	0	2	0	0	0	1	40	9	0	0
91	0	0	0	2	1	0	0	1	4	3	1	3	28	3	3	0	18	0	0	0
92	0	0	0	0	0	67	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93	0	0	0	0	4	0	0	2	2	3	0	0	49	1	1	4	0	1	0	0
127	4	0	7	10	0	2	0	2	8	0	1	9	3	5	11	3	1	1	0	0
128	0	0	22	0	0	3	1	0	3	1	0	26	0	1	1	5	3	1	0	0
130	1	0	9	4	0	0	0	1	1	9	0	4	0	29	3	3	0	2	1	0

Table 4.5: Number of each amino acid at contact position in primates. Highest number of each Amino acid is highlighted with Brown Colour.

Residue Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
63	0	0	0	0	6	0	0	8	0	9	0	1	0	0	0	0	2	0	0	0
64	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	20
73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0
75	15	0	0	0	0	6	0	0	0	0	0	1	0	0	0	4	0	0	0	0
76	0	0	0	0	0	0	0	0	12	0	0	11	0	0	0	0	3	0	0	0
81	17	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0
85	3	0	0	3	0	10	0	0	0	0	0	0	0	1	5	3	1	0	0	0
86	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	2	20	0	0	0
88	0	0	0	0	0	0	0	0	0	4	0	0	0	20	1	0	0	1	0	0
90	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	23	0	0	0
91	0	0	0	0	0	0	0	0	0	1	0	0	22	0	2	0	1	0	0	0
92	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0
127	4	0	0	7	0	4	0	0	0	1	0	0	0	2	0	8	0	0	0	0
128	0	0	25	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Residue Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
130	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	25	0	0

4.2.2. Shannon's diversity index analysis

Shannon's diversity index accounts for both the evenness and richness of a species, making it more informative than a simple species count. It typically ranges from 0 to 4.5, with values closer to 4.5 indicating greater diversity, while a value of 0 suggests that only one species is present at that position. Shannon's diversity index was calculated for the contact residues of CEACAM1 using R version 4.4.1. Table 4.3 is showing number of each amino acid at contact position throughout all species in study. Table 4.4 presents the Shannon diversity index values, along with the evenness and unique residues observed at each position of CEACAM1 in all mammals under study. Table 4.5 presents the Shannon diversity index values, along with the evenness and unique residues observed at each position of CEACAM1 in all non-primate mammals under study. Table 4.6 presents the Shannon diversity index values, along with the evenness and unique residues observed at each position of CEACAM1 in all primates under study.

Table 4.6: Shannon's Diversity Index of Contact residues of CEACAM1 along with species richness and evenness in mammals.

Residue Position	Human Sequence	Shannon's Diversity Index	Unique_Species	Evenness
63	F	2.045002	13	0.797288
64	G	1.216407	10	0.528279
68	Y	0.958522	4	0.691428
73	V	1.058622	7	0.544024
75	G	2.26978	15	0.83816
76	N	1.853266	9	0.843457
81	G	1.519156	6	0.847857
83	A	1.722889	9	0.784121

Residue Position	Human Sequence	Shannon's Diversity Index	Unique_Species	Evenness
85	G	2.04621	10	0.888658
86	T	1.493036	11	0.622644
88	Q	2.247399	12	0.90442
90	T	1.252318	10	0.543875
91	P	1.536245	11	0.640664
92	G	0	1	NA
93	P	0.866251	9	0.394248
127	S	2.418193	15	0.892964
128	D	1.437124	11	0.599327
130	V	1.863185	12	0.749801

Table 4.7: Shannon's Diversity Index of Contact residues of CEACAM1 in non-primate mammals along with species richness and evenness.

Residue Position	Human Sequence	Shannon's Non-Primates	Species Non-Primates	Evenness Non-Primates
63	F	1.970199	11	0.821637
64	G	1.511175	10	0.656295
68	Y	0.973784	4	0.702437
73	V	1.292161	7	0.664039
75	G	2.210082	15	0.816116
76	N	1.824845	9	0.830523
81	G	1.260384	6	0.703434
83	A	1.950181	9	0.887566
85	G	1.817885	9	0.827355
86	T	1.675098	11	0.69857

Residue Position	Human Sequence	Shannon's Non-Primates	Species Non-Primates	Evenness Non-Primates
88	Q	2.308401	12	0.928969
90	T	1.427697	10	0.620041
91	P	1.735391	11	0.723714
92	G	0	1	NA
93	P	1.102335	9	0.501694
127	S	2.377958	14	0.901063
128	D	1.657705	11	0.691317
130	V	1.872305	12	0.753471

Table 4.8: Shannon's Diversity Index of contact residues of CEACAM1 in primates along with species richness and evenness.

Residue Position	Human Sequence	Shannon's Primates	Species Primates	Evenness Primates
63	F	1.390889	5	0.864208
64	G	0	1	NA
68	Y	0.540204	2	0.77935
73	V	0	1	NA
75	G	1.069001	4	0.771121
76	N	0.969959	3	0.882895
81	G	0.645033	2	0.930586
83	A	0.429323	2	0.619382
85	G	1.682691	7	0.864732
86	T	0.79373	4	0.572555
88	Q	0.740411	4	0.534094
90	T	0.48439	4	0.349414
91	P	0.58928	4	0.425076
92	G	0	1	NA
93	P	0	1	NA
127	S	1.614498	6	0.901069

Residue Position	Human Sequence	Shannon's Primates	Species Primates	Evenness Primates
128	D	0.163024	2	0.235193
130	V	0.163024	2	0.235193

4.2.3. Residue Similarity Analysis

1. To calculate the similarity score, each amino acid was assigned a score based on its similarity to a reference sequence.
2. For similarity to humans, the human contact residue sequence served as the reference, while the mouse (*Mus musculus*) contact residue sequence was used to calculate similarity to mice.
3. Amino acids that matched the reference sequence were assigned a score of 1, while dissimilar amino acids were assigned a score of -1.
4. Similarity was assessed based on shared properties, including hydrogen bonding, charge (e.g., replacement of a negative with a positive amino acid), and Van Der Waals interactions (17).

Table 4.9 shows the amino acids grouped by similarity for score calculation

Table 4.9: Classification of amino acids according to similar properties with reference sequence amino acid.

Position - CEACAM1	All_Common_AA	Mice Reference Sequence	Similar_ AA (Score: 1)	Neutral_AA (Score: 0)	Dissimilar_AA (Score: -1)	Human Sequence (Interest)	Similar_ AA (Score: 1)	Neutral_AA (Score: 0)	Dissimilar_AA (Score: -1)
63	63- T, Q, V, L, N, A, I, Y, G, E, F, R, P	G	G,A,V,L,I	T,Q,N	F,Y,P,E,R	F	F,Y,L,I	A,V,N,Q,T	R,P,E,G
64	64- G, V, S, I, R, L, F, A, T, K	A	A,G,V,L,I	S,T,F	K,R	G	G,A,V,L,I	S,T	R,K,F
68	68- Y, H, F, C	Y	Y,F	-	H,C	Y	Y,F	-	H,C
73	73- V, T, S, A, I, L, Q, G, P, E	T	A,I,T,S	V,G,L	P,E,Q	V	V,A,L,I,T	S,G	P,E,Q,
75	75- S, V, N, K, P, A, L, G, E, H, F, Q, D, I, T	I	G,I,A,L,V,F	P,S,T,N,Q	D,E,K,H	G	G,A,V,L,I	S,T,Q,N	D,F,P,E,K,H
76	76- T, N, S, A, D, K, E, H, R	D	D,E,N,S,T,H	A	K,R	N	D,N,T,H,E,S	A	R,K
81	81- G, R, S, L, A, T, K, H, Y	R	K,H,R	S,T,A,G	L,Y	G	A,S,L,G	T,H	Y,K,R
83	83- V, R, I, L, K, M, S, A, E, T	V	V,A,T,L.I.M	S	R,K,E	A	A,V,L,I	F,S,T	E,K,R
85	85- T, Q, D, A, G, S, N, E, R, L	N	N, Q, S, T,	G, A, L	R,D,E	G	G,A,L,I	T,Q,N,S	E,R,D
86	86- T, V, N, S, P, D, E, L, K, I, G	S	S,T,N,I	G,P	D,E,V,L ,K	T	I,T,S,N,	V,G,L	K,D,E,P

Position - CEACAM1	All_Common_AA	Mice Reference Sequence	Similar_ AA (Score: 1)	Neutral_AA (Score: 0)	Dissimilar_AA (Score: -1)	Human Sequence (Interest)	Similar_ AA (Score: 1)	Neutral_AA (Score: 0)	Dissimilar_AA (Score: -1)
88	88- Q, R, K, T, L, A, I, V, S, E, N, M, G	M	M,L,I,V	A,N,Q,T,S	E,K,R,G	Q	Q,N,T,S	A,M,G	E,K,I,L,V,R
90	90- I, T, V, A, G, L, S, K, F, N	F	F, L,I	G,A,S,T,N,V	K	T	A,I,T,S,N,	V,G,L	K,F
91	91- P, M, T, K, N, L, R, Q, E, I, F, D, S	T	I,T,S,N,	M,L	K,D,F,R,E,Q,P	P	P	Q,N,V,G,I,M,L,S	T,R,F,D,K
92	92- G	G	G	_	_	G	G	_	_
93	93- P, S, K, V, F, I, Q, L, R, A	Q	S,Q	I,K,P	A,L,F,V,R	P	S,P,A	Q,I,K,V,L	R,F
127	127 - A, P, E, N, D, Q, R, K, T, S, L, I, G, V, M,	E	D,E,	G,Q,N,S,T	L,I,M,V,A,R,K,P	S	S,T, Q, N,I, A,	G,P,L,M,V	D,E,R,K
128	128- D, S, K, T, N, G, V, L, R, H, Q	N	N, Q, S, T,	G,A,L,V,F	D,H,K,R	D	D, Q, S, T,	G,L,V	N,K,R,H
130	130- V, N, Q, I, E, D, W, S, A, L, R, K, T, Y	R	R,K,Q,N	S,T,A	D,E,I,W,V,Y	V	V,L,I,A,T	S,N,Q,W	R,K,E,D,Y

4.2.4. Residue Similarity Score to Human and Mice

Similarity of all 18 contact residues to human and mice was also calculated on R 4.4.1 and shown in table 4.10.

Table 4.10: Similarity scores to human and mice.

Sr.	Accession	*Specie names*	Human similarity score	Mice similarity score
1	XM_012461510.3	*Aotus nancymaae*	0.833333	0.222222
2	XM_052168799.1	*Apodemus sylvaticus*	0.5	0.444444
3	XM_053657140.1	*Artibeus jamaicensis*	0.555556	0.611111
4	XM_036833342.1	*Balaenoptera musculus*	0.388889	0.111111
5	XM_010842547.1	*Bison bison*	0.444444	0.333333
6	XM_024977969.2	*Bos taurus*	0.388889	0.222222
7	XM_006041224.4	*Bubalus bubalis*	0.333333	0.222222
8	XM_055551361.1	*Bubalus kerabau*	0.333333	0.222222
9	XM_052655777.1	*Budorcas taxicolor*	0.388889	0.222222
10	XM_002762196.3	*Callithrix jacchus*	0.777778	0.166667
11	XM_064489351.1	*Camelus dromedarius*	0.5	0.277778
12	NM_001097557.1	*Canis lupus familiaris*	0.333333	0.111111
13	XM_063252825.1	*Cavia porcellus*	0.111111	0.722222
14	XM_063252826.1	*Cavia porcellus*	0.222222	0.611111
15	XM_012086961.1	*Cercocebus atys*	0.666667	0.111111
16	AF259566.1	*Cercopithecus aethiops*	0.888889	0.166667
17	XM_005412355.1	*Chinchilla lanigera*	0.444444	0.555556
18	XM_007997002.2	*Chlorocebus sabaeus*	0.777778	0.055556
19	XM_006871507.1	*Chrysochloris asiatica*	0.722222	0.277778

Sr.	Accession	*Specie names*	Human similarity score	Mice similarity score
20	XM_011936539.1	*Colobus angolensis*	0.611111	0.111111
21	XM_012733660.1	*Condylura cristata*	0.222222	0.222222
22	XM_024577349.3	*Desmodus rotundus*	0.166667	-0.05556
23	XM_042668120.1	*Dipodomys spectabilis*	0.555556	0.111111
24	XM_045292270.1	*Echinops telfairi*	0.388889	0.166667
25	MF564057.1	*Equus caballus*	0.333333	0.166667
26	XM_005596278.3	*Equus caballus*	0.5	0.444444
27	XM_007532949.3	*Erinaceus europaeus*	0.555556	0.333333
28	XM_011289901.4	*Felis catus*	0.111111	0.222222
29	XM_010623328.3	*Fukomys damarensis*	0.611111	0.222222
30	XM_031004080.3	*Gorilla gorilla*	1	0.166667
31	XM_004873005.2	*Heterocephalus glaber*	0	0.111111
32	NM_001024912.3	*Homo sapiens*	1	0.166667
33	NM_001712.5	*Homo sapiens*	1	0.166667
34	XM_032172816.2	*Hylobates moloch*	0.777778	0.222222
35	XM_005338246.4	*Ictidomys tridecemlineatus*	0.388889	0.5
36	XM_004670324.2	*Jaculus jaculus*	0.222222	0.555556
37	XM_062175861.1	*Lepus europaeus*	0.555556	0.111111
38	XM_064293748.1	*Loxodonta africana*	0	0
39	XM_030297609.1	*Lynx canadensis*	0.111111	0.166667
40	XM_005589370.4	*Macaca fascicularis*	0.722222	0.111111
41	MG874670.1	*Macaca mulatta*	0.611111	0
42	XM_011764531.2	*Macaca nemestrina*	0.833333	0.055556

Sr.	Accession	*Specie names*	Human similarity score	Mice similarity score
43	XM_037013906.1	*Manis javanica*	0.444444	0.277778
44	XM_036896477.2	*Manis pentadactyla*	0.666667	0.5
45	XM_015496801.2	*Marmota marmota*	0.555556	0.444444
46	XM_040734318.1	*Mesocricetus auratus*	0.333333	0.388889
47	XM_041654886.1	*Microtus oregoni*	0	0.277778
48	XM_064581140.1	*Mirounga angustirostris*	0.388889	0.111111
49	XM_055412081.1	*Moschus berezovskii*	0.277778	0.055556
50	NM_001039185.1	*Mus musculus*	-0.05556	1
51	XM_004780425.3	*Mustela putorius*	0.111111	0.166667
52	XM_048440439.1	*Myodes glareolus*	0	0
53	XM_015562203.1	*Myotis davidii*	0.277778	0.444444
54	XM_036346012.1	*Myotis myotis*	0.444444	0.388889
55	XM_029570052.1	*Nannospalax galili*	0.111111	0.277778
56	XM_030797105.1	*Nomascus leucogenys*	1	0.277778
57	XM_055311974.1	*Nyctereutes procyonoides*	0.166667	0
58	XM_023704520.1	*Octodon degus*	0.277778	0.166667
59	XM_020871207.1	*Odocoileus virginianus*	0.444444	0.388889
60	XM_036168114.1	*Onychomys torridus*	0	0
61	XM_049703839.1	*Orcinus orca*	0.444444	0.111111
62	XM_039912179.1	*Ornithorhynchus anatinus*	0	0
63	XM_008249692.3	*Oryctolagus cuniculus*	0	0
64	XM_004015280.6	*Ovis aries*	0.388889	0.166667

Sr.	Accession	*Specie names*	Human similarity score	Mice similarity score
65	XM_008964144.4	*Pan paniscus*	0.833333	0.333333
66	XM_063801767.1	*Pan troglodytes*	1	0.166667
67	XM_019434422.2	*Panthera pardus*	0.166667	0.166667
68	XM_007086642.2	*Panthera tigris*	0.222222	0.166667
69	XM_049621270.1	*Panthera uncia*	0.166667	0.277778
70	XM_003915629.4	*Papio Anubis*	0.777778	0.222222
71	XM_059253485.1	*Peromyscus eremicus*	0.055556	0.444444
72	XM_028860001.2	*Peromyscus leucopus*	0.055556	0.388889
73	XM_042267688.1	*Peromyscus maniculatus*	0.055556	0.388889
74	XM_051207508.1	*Phodopus roborovskii*	0.222222	0.444444
75	XM_028530369.2	*Phyllostomus discolor*	0.611111	0.222222
76	XM_024132797.3	*Physeter catodon*	0.444444	0.222222
77	XM_023229215.2	*Piliocolobus tephrosceles*	0.555556	0.111111
78	NM_001132423.1	*Pongo abelii*	0.722222	0.277778
79	XM_054463351.2	*Pongo pygmaeus*	0.722222	0.277778
80	XM_012653468.1	*Propithecus coquereli*	0.611111	0.5
81	XM_055620670.1	*Psammomys obesus*	0	0
82	XM_054567634.1	*Pteronotus mesoamericanus*	0	-0.11111
83	NM_001033860.1	*Rattus norvegicus*	0.111111	0.5
84	XM_032894175.1	*Rattus rattus*	0.388889	0.444444
85	XM_017849842.1	*Rhinopithecus bieti*	0.666667	0.166667
86	XM_010358723.2	*Rhinopithecus roxellana*	0.666667	0.166667

Sr.	Accession	*Specie names*	Human similarity score	Mice similarity score
87	XM_016129903.2	*Rousettus aegyptiacus*	0.5	0.277778
88	XM_066271982.1	*Saccopteryx bilineata*	0.722222	0.333333
89	XM_066373583.1	*Saccopteryx leptura*	0.555556	0.277778
90	XM_003943210.3	*Saimiri boliviensis*	0.444444	0.166667
91	XM_055145459.1	*Sorex araneus*	0	0
92	XM_056135625.1	*Sorex fumeus*	0	0
93	XM_005655889.2	*Sus scrofa*	0.5	0.166667
94	XM_037516693.2	*Talpa occidentalis*	0.444444	0.111111
95	XM_025366225.1	*Theropithecus gelada*	0.722222	0.055556
96	XM_033197411.1	*Trachypithecus francoisi*	0.666667	0.055556
97	XM_026403696.1	*Urocitellus parryii*	0.611111	0.388889
98	XM_026482217.4	*Ursus arctos*	0.388889	0.277778
99	XM_008709687.2	*Ursus maritimus*	0.444444	0.388889
100	XM_040634031.1	*Ursus maritimus*	0	0
101	XM_031681826.1	*Vicugna pacos*	0.5	0.611111
102	XM_027858153.1	*Vombatus ursinus*	0	0

While comparing the similarity with human and mice between residues that interact with the virus, no mammal specie under top ten and twenty showed resemblance with both human and mice residues. Under top thirty there were found seven mammals showing resemblance with human and mice residues highlighted as colour brown in table 4.11. *Manis pentadactyla* showed highest resemblance with both mice and human binding residues followed by two primates; *Nomascus leucogenys* and *Pan paniscus*.

A boxplot graph was created according to the similarity score (Figure 4.1)

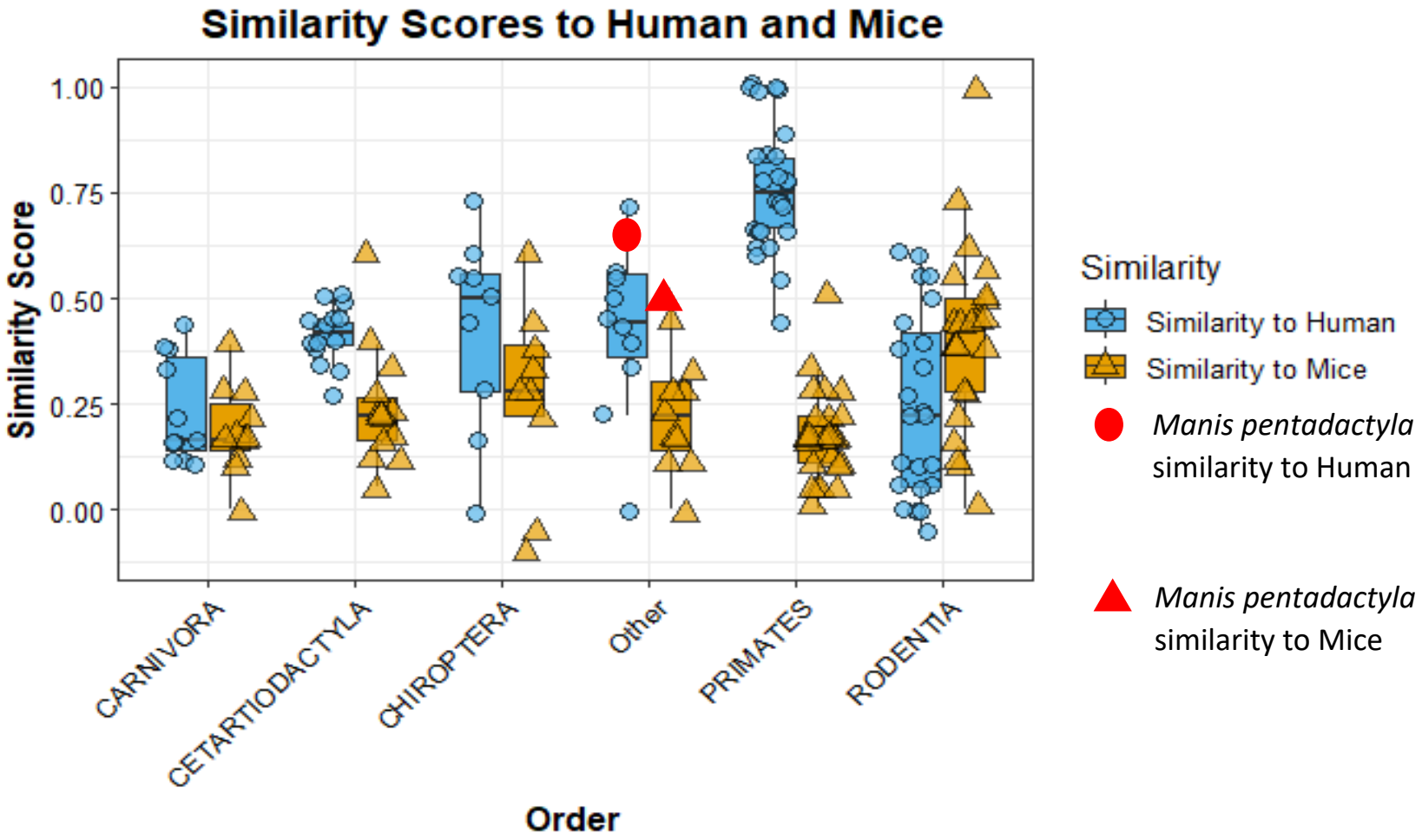

Figure 4.1: Similarity scores to human and mice.

Table 4.11: Comparative analysis of similarity to human and similarity to mouse scores. Yellow shade shows the specie with similarity to mice and human under position 40. Brown shade shows the specie with similarity to mice and human under position 30. ***Manis pentadactyla*** under shade blue shows the highest relevancy in both human and mice similarity score.

Accession	Specie	Order	Human score	Mice score	Position in Human Score	Position in Mice Score
XM_030797105.1	*Nomascus leucogenys* (White cheeked Gibbon)	PRIMATES	1.0000	0.2778	4	29
XM_008964144.4	*Pan paniscus* (Bonobo)	PRIMATES	0.8333	0.3333	9	25
XM_006871507.1	*Chrysochloris asiatica* (Mole)	AFROTHERIA	0.7222	0.2778	14	30
NM_001132423.1	*Pongo abelii*	PRIMATES	0.7222	0.2778	16	31
XM_054463351.2	*Pongo pygmaeus*	PRIMATES	0.7222	0.2778	17	32
XM_066271982.1	*Saccopteryx bilineata*	CHIROPTERA	0.7222	0.3333	18	26
XM_036896477.2	*Manis pentadactyla* (Pangolin)	PHOLIDOTA	0.6667	0.5000	21	7

Accession	Specie	Order	Human score	Mice score	Position in Human Score	Position in Mice Score
XM_012653468.1	*Propithecus coquereli* (Lemur)	PRIMATES	0.6111	0.5000	29	8
XM_026403696.1	*Urocitellus parryii*	RODENTIA	0.6111	0.3889	30	18
XM_053657140.1	*Artibeus jamaicensis*	CHIROPTERA	0.5556	0.6111	31	3
XM_007532949.3	*Erinaceus europaeus*	EULIPOTYPHLA	0.5556	0.3333	33	27
XM_015496801.2	*Marmota marmota*	RODENTIA	0.5556	0.4444	35	11
XM_066373583.1	*Saccopteryx leptura*	CHIROPTERA	0.5556	0.2778	37	33
XM_052168799.1	*Apodemus sylvaticus*	RODENTIA	0.5000	0.4444	38	12
XM_064489351.1	*Camelus dromedarius*	CETARTIODACTYLA	0.5000	0.2778	39	34
XM_005596278.3	*Equus caballus*	PERISSODACTYLA	0.5000	0.4444	40	13

4.3. MOLECULAR EVOLUTION ANALYSIS

4.3.1. Phylogenetic Trees

To accurately analyse evolutionary patterns and selection pressures on genes of interest, a time tree was employed. This tree provides a calibrated timeline for the evolutionary history of study organisms (66). By integrating this temporal information into HyPhy analyses, more precise estimates of the timing of key evolutionary events and rates of molecular evolution across lineages can be obtained.

Figure 4.2: MCC Phylogenetic of order Primates

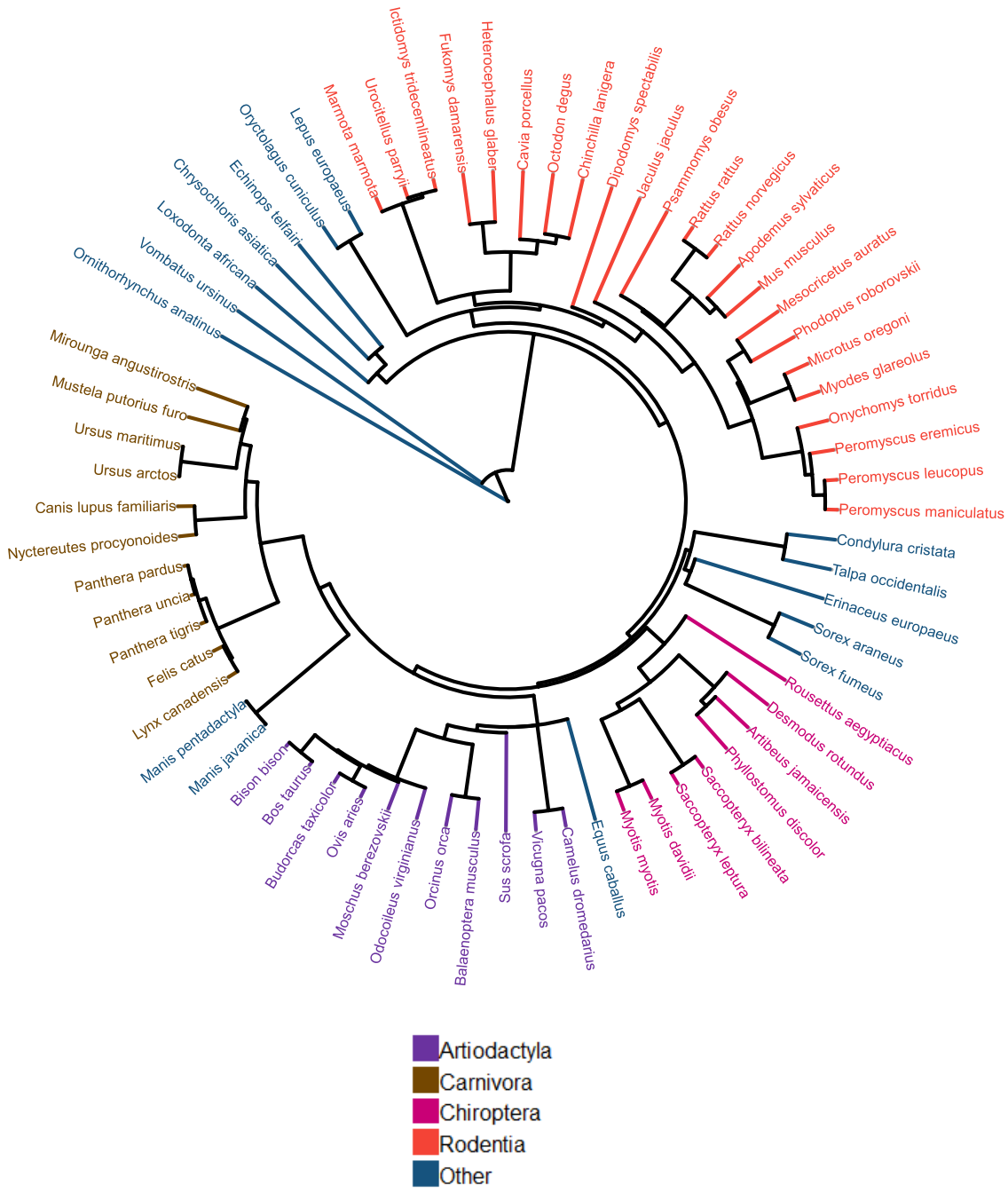

Figure 4.3: Phylogenetic tree of non-primate mammals

4.3.2. MEME analysis

A mixed-effects model of evolutionary analysis was conducted using HyPhy, with results analysed in R version 4.4.1. The following findings were observed:

For the "full dataset", 12 out of 18 contact residues were under positive selection at $p < 0.05$, while 76 out of 510 non-contact residues were also under positive selection at this threshold.

In the “non-primates” dataset, 11 of 18 contact residues and 70 of 510 non-contact residues showed positive selection at $p < 0.05$. In the “primates” dataset, 4 out of 18 contact residues and 27 out of 510 non-contact residues were under positive selection at $p < 0.05$ (Table 4.12).

Table 4.12: Positive selection analysis of residues at p-value<0.05. contact_sel_0.05;Contact residues under p-selection at p-value <0.05, contact_notsel_0.05; Contact residues under no p-selection at p-value <0.05, noncontact_sel_0.05; other than contact residues under p-selection at p-value <0.05, noncontact_notsel_0.05;other than contact residues under no p-selection at p-value <0.05, fisher_p_0.05; Fisher's Exact test at p-value<0.05, contact residues under p-selection and no p-selection versus non-contact residues under p-selection and no p-selection, fisher_odds_0.05; Fisher's Odd's Ratio at p-value<0.05

	contact_sel_0.05	**contact_notsel_0.05**	**noncontact_sel_0.05**	**noncontact_notsel_0.05**	**fisher_p_0.05**	**fisher_odds_0.5**
All Mammals	12	6	76	434	1.83E-06	11.4
Non-primates	11	7	70	440	7.59E-06	9.877551
Primates	4	14	27	483	0.0167	5.111111

At $p < 0.1$, for the “full dataset”, 13 out of 18 contact residues showed positive selection along with 104 out of 510 non-contact residues. In the “non-primates” dataset, 11 of 18 contact residues and 94 of 510 non-contact residues were under positive selection at this threshold. In the “primates” dataset, 7 out of 18 contact residues and 41 of 510 non-contact residues were under positive selection at $p < 0.1$ (Table 4.13).

Table 4.13: Positive selection analysis of residues at p-value<0.1. contact_sel_0.1; Contact residues under p-selection at p-value <0.1, contact_notsel_0.1; Contact residues under no p-selection at p-value <0.1, noncontact_sel_0.1; other than contact residues under p-selection at p-value <0.1, noncontact_notsel_0.1; other than contact residues under no p-selection at p-value <0.1, fisher_p_0.1; Fisher's Exact test at p-value<0.1, contact residues under p-selection and no p-selection versus non-contact residues under p-selection and no p-selection, fisher_odds_0.1; Fisher's Odd's Ratio at p-value<0.1

	contact_sel_0.1	**contact_notsel_0.1**	**noncontact_sel_0.1**	**noncontact_notsel_0.1**	**fisher_p_0.1**	**fisher_odds_0.1**
All Mammals	13	5	104	406	5.47E-06	10.15
Non-primates	11	7	94	416	0.00011	6.954407

	contact_sel_0.1	contact_notsel_0.1	noncontact_sel_0.1	noncontact_notsel_0.1	fisher_p_0.1	fisher_odds_0.1
Primates	7	11	41	469	0.000489	7.279379

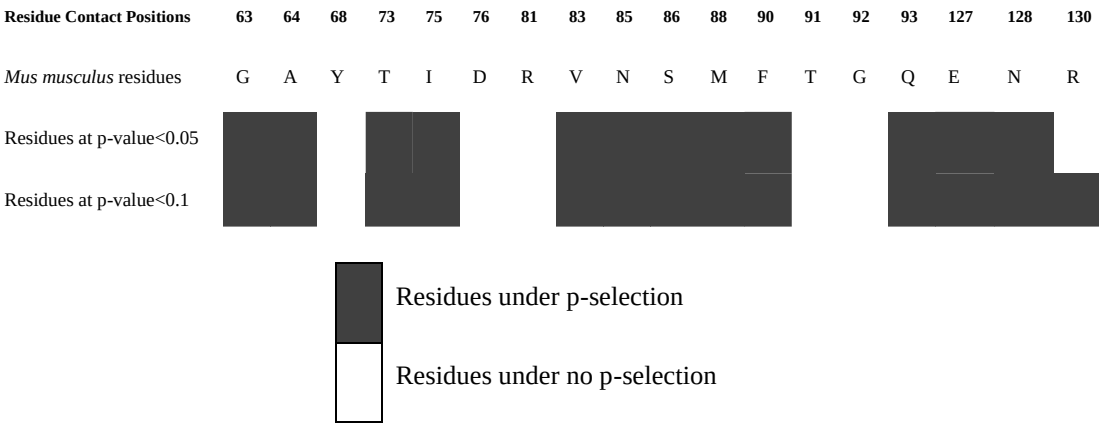

Figure 4.4: Contact residues with positive selection status. Black colour represents residues under positive selection and white colour represents residues under no positive selection.

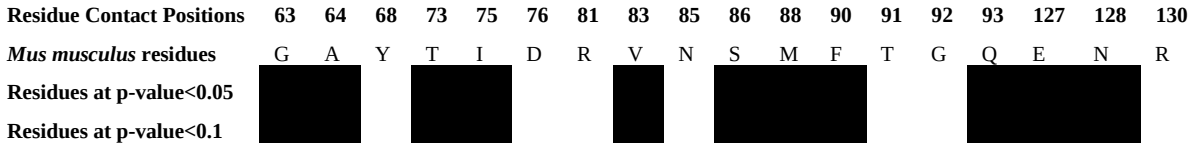

Figure 4.5: Contact residues of non-primate mammals with positive selection status. Black colour represents residues under positive selection and white colour represents residues under no positive selection.

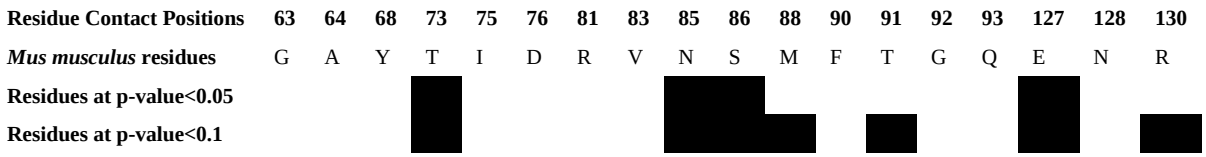

Figure 4.6: Contact residues of primates with positive selection status. Black colour represents residues under positive selection and white colour represents residues under no positive selection.

Complete MEME analysis files in CSV table format are given In Appendix10, appendix 11, and appendix 12 for the whole data set, non-primate dataset and primate dataset respectively.

4.3.2. aBSREL analysis

Adaptive branch site random effect likelihood analysis was performed on HyPhy and results were analysed on R 4.4.1, without p-value correction at p-value<0.05. Table 4.14 and figure 4.7 were made to visualize the results.

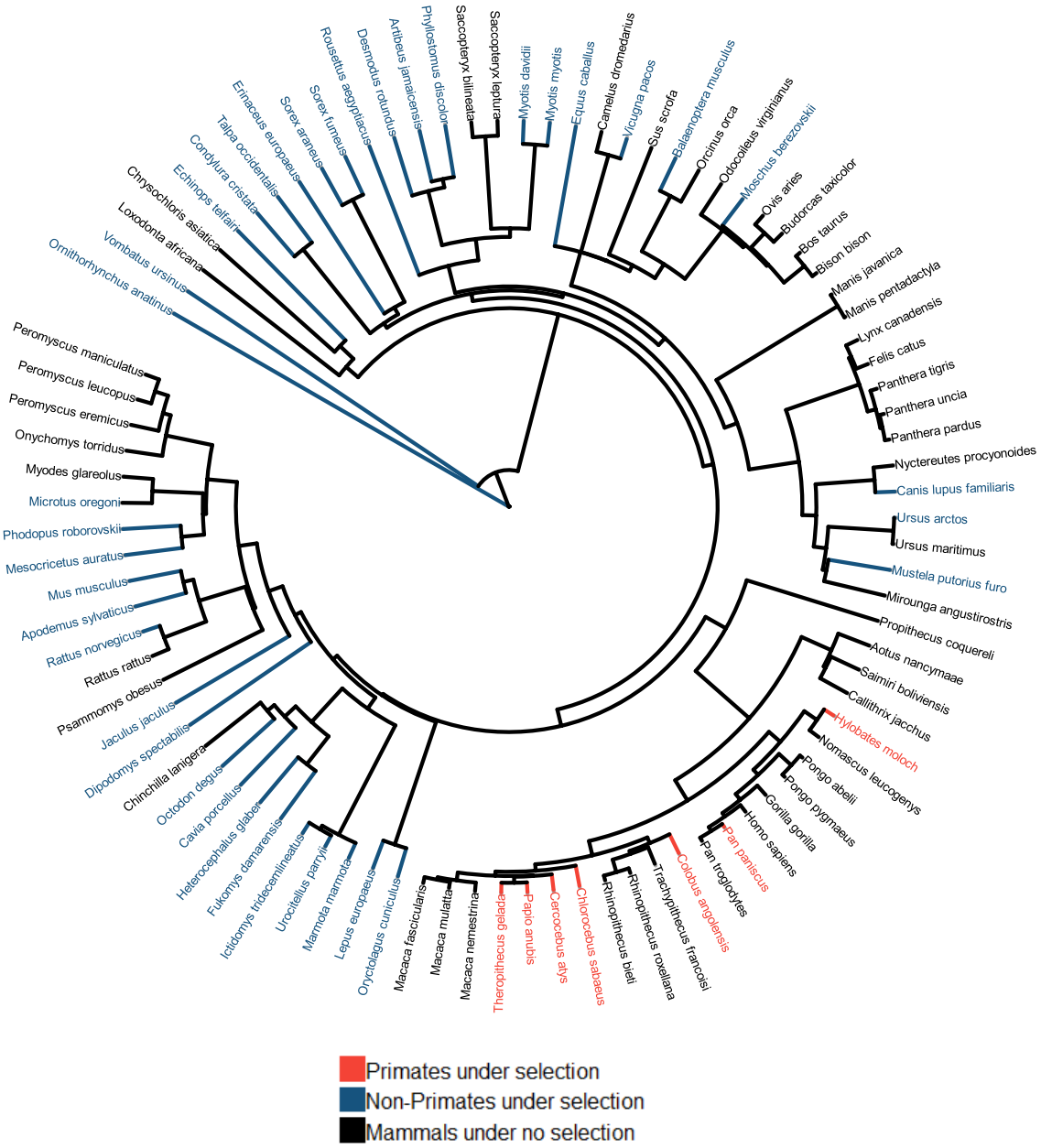

Figure 4.7: aBSREL analysis at p-value<0.05 showing terminal branches with their selection pressure status

Primates LRT and p-value at base node of clade Primates in all specie aBSREL analysis was found 24.345 and 1.71E-06 respectively. Results of fisher exact test are given in the table 4.14

Table 4.14: Number of branches with their selection pressure status and Fisher's Exact test values. Pri_sel; Primates under selection pressure, Pri_notsel; Primates under no selection pressure, Npri_sel; non-primates under selection pressure, Npri_notsel; non-primates under no selection pressure, fisher_p_prinonpri; Fisher's Exact test, Primates versus non-primates, fisher_odds_prinonpri; Fisher's Odd's Ratio, TermPri_sel; Terminal branches of primates under selection pressure, TermPri_notsel; Terminal branches of primates under no selection pressure, TermNpri_sel; terminal branches of non-primates under selection pressure, TermNpri_notsel; terminal branches of non-primates under no selection pressure, fisher_p_terminal; Fisher's Exact test, Terminal branches of primates versus terminal branches of non-primates, fisher_odds_terminal; Fisher's Odd's Ratio

Pri_sel	Pri_notsel	Npri_sel	Npri_notsel	fisher_p_prinonpri	fisher_odds_prinonpri
12	31	68	65	0.008495	0.370019
TermPri_sel	**TermPri_notsel**	**TermNpri_sel**	**TermNpri_notsel**	**fisher_p_terminal**	**fisher_odds_terminal**
7	16	39	29	0.03137	0.325321

4.4. SAMPLE STATISTICAL ANALYSIS

Independent distribution of codon base-pairs along primates and non-primate mammal species were analysed using Wilcoxon Rank-Sum Test and Shapiro-wilk test was performed to assess the distribution shape along all the dataset using R 4.4.1. Very low p-value (p-value=2.991e-08; much less than 0.05) of Wilcoxon Rank-Sum Test indicates a significant difference between the number of amino acid residues in two groups, primates and non-primate mammals, rejecting the null-hypothesis that two groups are identical. A higher W value (W=268) indicates a greater difference between primates and non-primate mammals.

As for Shapiro-wilk test, given the very small p-value (7.308e-14), we reject the null hypothesis that the data are normally distributed. This suggests that sample data (the number of amino acid residues of CEACAM1 in mammals) do not follow a normal distribution.

Table 4.15: Wilcoxon Rank-sum and Shapiro-Wilk test. W-value shows the test statistics and p-value shows the significance of results.

	W	**p-value**
Wilcoxon Rank-sum test	268	2.991e-08
Shapiro-Wilk test	0.66785	7.308e-14

Mean value of sample data in primates and non-primates are 504 and 477 respectively. To visualize the completeness and variability of sample a boxplot graph was made (Figure 4.8).

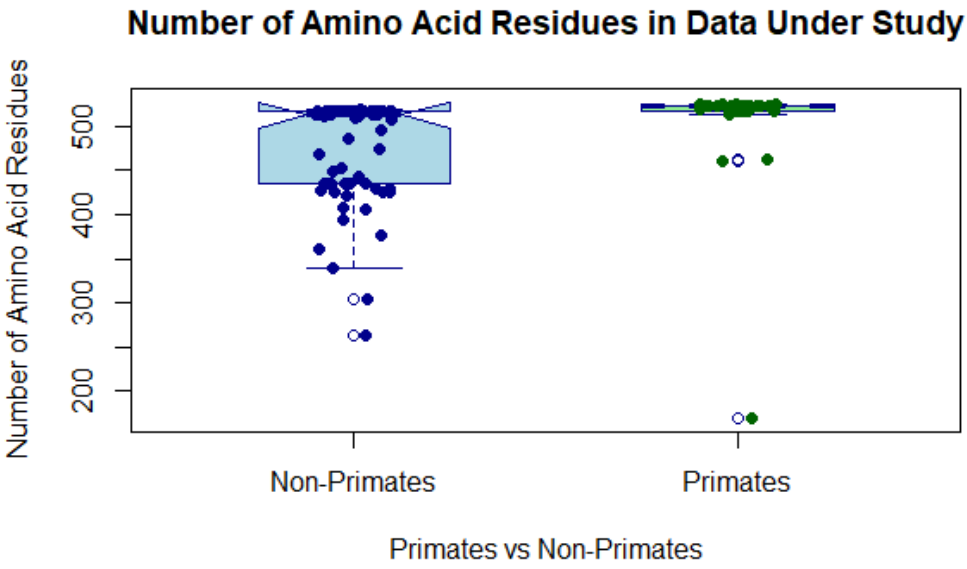

Figure 4.8: Sample completeness and variability. Mean value of sample data in primates and non-primates are 504 and 477 respectively. To visualize the completeness and variability of sample a boxplot graph was made

CHAPTER 5

DISCUSSION

Receptor genetic diversity plays a crucial role in functional adaptations across species. Various studies on receptor genetic diversity have been conducted over time. For example, the high variability in canine olfactory receptors affects receptor-ligand binding, potentially explaining why breeds like Labrador Retrievers and German Shepherds outperform Pekingese and Greyhounds as sniffer dogs. This variation highlights the importance of receptor flexibility in shaping behaviour and specialized functions (67). Similarly, variations in Fc gamma receptors (FCGR) showcase distinct regulatory mechanisms despite highly similar sequences. These differences are essential for host immune defences and disease outcomes, reflecting the impact of genetic diversity on critical biological processes (68).

Research on dopamine and glutamate receptors further emphasizes the significance of receptor diversity. Discovering multiple dopamine receptor subtypes has expanded our understanding of their complex roles in neural circuits and regulatory mechanisms (69). Likewise, glutamate receptor subtypes such as AMPA-kainate, NMDA, and metabotropic receptors demonstrate how diversity facilitates specific roles in synaptic transmission, neuroprotection, and brain function. Their cooperative actions and localization underline their importance in both normal physiology and therapeutic contexts (70).

The diversity of viral host receptors reflects evolutionary processes that adapt these receptors for viral exploitation. Viral receptors often evolve from host-derived proteins, undergoing structural changes driven by mutagenesis and natural selection of more virulent strains (71). For instance, viral seven-transmembrane (7TM) receptors differ significantly from endogenous ones, illustrating how receptor adaptations optimize viral attachment, entry, and replication. Studying these adaptations sheds light on receptor-mediated mechanisms across diverse viral families, advancing our understanding of host-pathogen dynamics (71).

Interactions between viruses and host receptors reveal intricate adaptability in processes like attachment, entry, and targeting. Examples include CD4 as the receptor for HIV and ICAM-1 for rhinoviruses, which highlight receptor specificity and diversity in function. Mechanisms such as the "canyon hypothesis" for rhinoviruses

demonstrate structural adaptations that facilitate host-cell recognition (72). Additionally, diverse receptor-ligand pairings, such as those involving natural killer (NK) cells and MHC class I molecules, showcase how viral systems counteract immune defences. This receptor diversity allows viruses to thrive in different environments, influencing infection progression and disease outcomes (73,74).

The diversity of viral host receptors is pivotal in shaping the dynamics of zoonosis, influencing the ability of viruses to infect and adapt to multiple species. The ACE2 receptor, a key entry point for SARS-CoV-2, exemplifies this phenomenon. While ACE2 orthologs are highly conserved across mammals, their greater diversity among bat species positions bats as significant reservoirs for SARS-CoV-2 and its progenitors. Studies have demonstrated that ACE2 orthologs from a variety of mammals; including pets, livestock, and zoo animals; can mediate SARS-CoV-2 infection, revealing a broad host range for zoonotic viruses (17,75). This diversity in host receptors underscores the critical role of robust animal surveillance systems in identifying potential zoonotic spill overs and preventing future outbreaks.

The correlation between receptor sequence similarity and zoonotic potential further underscores the importance of receptor diversity in cross-species transmission. Research has shown that infection likelihood increases with greater receptor similarity between primary and potential hosts, enabling predictive insights even before structural data are available (76). Advanced tools like REPOST refine these predictions by pinpointing key receptor residues and identifying potential viral hosts, enhancing efforts to pre-empt zoonotic events. Concurrently, the discovery of novel paramyxoviruses has highlighted the challenges of classification and the need to integrate receptor tropism and structural analysis into predictive frameworks. For example, discrepancies in receptor conservation have led to the reclassification of certain viruses, expanding our understanding of their host range and zoonotic potential (77). These insights highlight the intricate relationship between viral host receptor diversity and zoonosis, advocating for interdisciplinary research to address the complexities of viral evolution and transmission.

To examine the diversity analysis of residues that interact with the virus, 102 CEACAM1 sequences were analysed from 98 species, including 26 primates and 72 non-primate mammals, representing 14 mammalian orders. Eighteen amino acid sites

in CEACAM1 contribute to stabilizing its interaction with the receptor-binding domain of MHV. Among these 18 sites, identified based on their positions in human CEACAM1, at least 89 unique amino acid combinations were observed (23 in primates and 66 in non-primate mammals). All CEACAM1 residues were polymorphic across species, but residue position 92 was found to be monomorphic in both primates and non-primate mammals. While positions 64, 73, 92, and 93 of amino acid residues are monomorphic only in primates.

Shannon's diversity index of binding residues shows position 127 of CEACAM1 is most diverse (Shannon's diversity index= 2.42) with 15 different amino acid species in all mammals showing evenness 0.89. While conducting diversity analysis on non-primate mammals, again position-127 was found most diverse (Shannon's diversity index= 2.38) with 14 different amino acid species showing evenness 0.90, but in primates, position-85 was found most diverse (Shannon's diversity index= 1.68) with 7 different amino acid species showing evenness 0.86. Overall binding residues of CEACAM1 in primates found to be conserved as compared to non-primate mammals.

While comparing the similarity with human and mice between residues that interact with the virus, no mammal specie under top ten and twenty showed resemblance with both human and mice residues. Under top thirty there were found seven mammals showing resemblance with human and mice residues; three primates, one bat (*Saccopteryx bilineata)*, one rodent (*Urocitellus parryii)*, one mole (*Chrysochloris asiatica*) and one pangolin. Pangolin (*Manis pentadactyla*) showed highest resemblance with both mice and human binding residues followed by two primates; white cheeked gibbon (*Nomascus leucogenys*) and bonobo (*Pan paniscus)*. As these three mammals showed similarity with both human and mice binding sequence, they might be the potential source for the inter-transmission of MHV from rodents to humans. They might serve as intermediate host in the cross-specie transmission of MHV in future.

Pangolins, heavily trafficked wild animals, have been widely studied for potential intermediate host for the transmission of SARS-CoV-2 in many previous studies. It have drawn attention due to the discovery of SARS-CoV-2-related coronaviruses in Malayan pangolins. An evolutionary analysis of viral and host

phylogenies, alongside ACE2 and TMPRSS2 amino acid sequences, suggests that pangolins could serve as intermediate hosts in the cross-species transmission of SARS-CoV-2. The study found that SARS-like coronaviruses infecting pangolins and bats are genetically close to SARS-CoV-2. Notably, the pangolin ACE2 sequence shows lower evolutionary divergence from humans compared to bats, and even less divergence than civets, previously identified as the main intermediate host for SARS-CoV. These findings position pangolins as potential facilitators for the bat-to-human transmission of SARS-CoV-2, highlighting their relevance in understanding the virus's origins (78).

Another study found that coronaviruses in pangolins, known as pangolin-CoV-2020, share a high genetic similarity with SARS-CoV-2 and some bat coronaviruses. However, phylogenetic analysis indicates that SARS-CoV-2 likely did not originate directly from pangolin-CoV-2020. Further research revealed that the S gene of pangolin-CoV-2020 contains fragments from bat coronaviruses, highlighting its complex evolutionary history. Coronaviruses were consistently detected in pangolins from separate batches, showing remarkable genetic similarity and suggesting that pangolins might naturally harbor these viruses. While the spike protein of pangolin-CoV-2020 is structurally similar to that of SARS-CoV-2 and can bind to ACE2 receptors, its potential to infect humans remains unclear. Though this study does not support pangolins as intermediate hosts for SARS-CoV-2, it emphasizes the importance of monitoring coronaviruses in wildlife, conserving endangered species, and limiting human contact with wildlife to reduce the risk of zoonotic spill-overs (79).

The intermediate host for SARS-CoV-2 is still unknown, but it’s a crucial link between bats, the natural hosts, and humans, allowing the virus to evolve and replicate along the way. Pangolins are considered strong candidates because of the genetic similarity between pangolin-nCoV and SARS-CoV-2, which is second only to bat-nCoV. Other animals like snakes, minks, and turtles, often found in wildlife markets, could also be involved, along with domestic and companion animals, which adds another layer of complexity due to potential social implications. Identifying the intermediate host is essential to stopping the virus's spread, preventing future pandemics, and informing wildlife conservation and management efforts. More

research on a variety of animal species is needed to clarify these uncertainties and develop effective ways to prevent further outbreaks (80).

Another study analyzed ACE2 binding with the SARS-CoV-2 RBD and found that while *Rhinolophus sinicus* bats are likely the natural host, pangolins are unlikely to serve as intermediate hosts despite earlier suspicions. Other animals, such as primates, civets, and goats, showed greater susceptibility, offering more plausible candidates for facilitating transmission to humans. These findings redirect focus from pangolins to other species in understanding the transmission chain of SARS-CoV-2 (81). Another research examines the ACE2-interacting residues of the SARS-CoV-2 spike protein and compares isolates from bats and pangolins in Guangdong and Guangxi, South China. The analysis supports the idea that Guangdong pangolins likely acted as intermediate hosts, facilitating the adaptation and transmission of SARS-CoV-2. This positions them as a crucial evolutionary link in the virus's transmission pathway, with further discussion on the potential role of intermediate hosts in the emergence of variants like Omicron (82).

Another analysis of the viromes of 161 smuggled pangolins revealed 28 vertebrate-associated viruses, including 21 previously unreported in vertebrates. Among these, pangolin-associated viruses belonging to *Hunnivirus, Pestivirus,* and *Copiparvovirus* were identified, with notable genomic features such as the loss of the l-protein in *hunniviruses*. Additionally, human-associated viruses like respiratory syncytial virus and mammalian *orthoreovirus* were detected, alongside a coronavirus related to HKU4-CoV, originally found in bats. These findings underscore pangolins as potential reservoirs for emerging pathogenic viruses, emphasizing their relevance in zoonotic transmission pathways (83).

In this study, selection analyses were conducted on mammalian maximum clade credibility (MCC) tree. MEME analysis was conducted on HyPhy and results were analysed for p-value< 0.05 and p-value < 0.1. At $p < 0.1$, for the “full dataset”, 13 out of 18 contact residues showed positive selection along with 104 out of 510 non-contact residues. In the non-primate dataset, 11 out of 18 contact residues and 94 of 510 non-contact residues were under positive selection at this threshold. In the primate dataset, 7 out of 18 contact residues and 41 of 510 non-contact residues were under positive selection at $p < 0.1$. For the “full dataset”, 12 out of 18 contact residues were under

positive selection at $p < 0.05$, while 76 out of 510 non-contact residues were also under positive selection at this threshold. In the "non-primate" dataset, 11 of 18 contact residues and 70 of 510 non-contact residues showed positive selection at $p < 0.05$. In the primate dataset, 4 out of 18 contact residues and 27 out of 510 non-contact residues were under positive selection at $p < 0.05$.

This MEME analysis shows that a greater proportion of binding residues of CEACAM1 was under positive selection than other residues of CEACAM1. 12 binding residues; 63, 64, 73, 75, 83, 85, 86, 89, 90 , 93, 127 and 128; were under selection at p-value <0.05 but position 130 was under selection for $p < 0.1$. Out of 18 contact residue sites, 11 sites were under selection at $p < 0.05$ in non-primate mammals. Out of 18 contact residue sites, 8 sites were under selection at $p < 0.1$ and 6 sites were under selection at $p < 0.05$.

Residue positions 75 and 86 were under selection (MEME, $p<0.05$) in all mammal dataset, but position 91 was additionally under positive selection in primates (MEME, $p<0.1$). Residue position 63, 64, 73, 83, 90, 93 and 128 (MEME, $p<0.05$) were under selection in non-primate mammals but not in primates. While residue position 85, 91 and 130 were under selection in primates but not in non-primate mammals.

One hypothesis that primates are more under selection pressure in CEACAM1, was made. This hypothesis was tested using aBSREL on HyPhy. aBSREL analysis showed that primate clade is under positive selection with p value 1.71E-06. When aBSREL analysis was conducted without prior setting of foreground branch, primate branches were more likely to be selected than other branches of the mammalian phylogeny (Fisher exact test $p = 0.0085$).

While conducting MEME analysis, increased selection was observed in binding residues of CEACAM1 in non-primates (Fisher exact test p=7.59E-06) than in primate residues (Fisher exact test p= 0.0167).

CHAPTER 6

CONCLUSION

This study explored the evolutionary landscape of CEACAM1 across 98 mammalian species, with a focus on the residues involved in viral binding. Diversity analysis revealed that CEACAM1 contact residues are highly variable across species, with non-primates showing greater variability than primates. Notably, residue 127 emerged as the most diverse site in both overall and non-primate datasets, while residue 85 was most diverse among primates. A conserved site, position 92, remained monomorphic across all species.

Comparison of CEACAM1 binding residues revealed that no species matched both human and mouse sequences within the top ten most similar. However, seven species; including pangolin, one rodent, one bat, and three primates; showed overlap in the top thirty, with pangolin displaying the highest similarity. These species may represent potential intermediate hosts in the cross-species transmission of MHV.

Selection analyses showed that a significantly higher proportion of contact residues were under positive selection compared to non-contact sites. Non-primate mammals exhibited stronger selection signals at CEACAM1 binding sites than primates, although aBSREL analysis revealed that the primate clade as a whole was under strong episodic diversifying selection. Specific residues also showed distinct selection patterns between the two groups.

Overall, this study highlights the evolutionary plasticity of CEACAM1 and its relevance to virus–host interactions. Future studies should focus on experimental validation of binding differences, structural modelling of CEACAM1 variants, and surveillance of potential intermediate hosts to better understand cross-species viral transmission.

REFERENCES

1. Verta JP, Jacobs A. The role of alternative splicing in adaptation and evolution. Trends Ecol Evol [Internet]. 2022;37(4):299–308. Available from: https://doi.org/10.1016/j.tree.2021.11.010

2. Kim CH. Sars-cov-2 evolutionary adaptation toward host entry and recognition of receptor o-acetyl sialylation in virus–host interaction. Int J Mol Sci. 2020;21(12):1–34.

3. Salajegheh Tazerji S, Magalhães Duarte P, Rahimi P, Shahabinejad F, Dhakal S, Singh Malik Y, et al. Transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to animals: An updated review. J Transl Med [Internet]. 2020;18(1):1–11. Available from: https://doi.org/10.1186/s12967-020-02534-2

4. Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) origin and animal reservoir Susanna Lau. Virol J [Internet]. 2016;13(1):1–7. Available from: http://dx.doi.org/10.1186/s12985-016-0544-0

5. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The Evolution and Genetics of Virus Host Shifts. PLoS Pathog. 2014;10(11).

6. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet. 2015;16(4):224–36.

7. Bárcena M, Oostergetel GT, Bartelink W, Faas FGA, Verkleij A, Rottier PJM, et al. Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. Proc Natl Acad Sci U S A. 2009;106(2):582–7.

8. Peng G, Yang Y, Pasquarella JR, Xu L, Qian Z, Holmes K V., et al. Structural and molecular evidence suggesting coronavirus-driven evolution of mouse receptor. J Biol Chem [Internet]. 2017;292(6):2174–81. Available from: http://dx.doi.org/10.1074/jbc.M116.764266

9. Maginnis MS. Virus–Receptor Interactions: The Key to Cellular Invasion. J Mol Biol. 2018;430(17):2590–611.

10. Li J, Lai S, Gao GF, Shi W. The emergence, genomic diversity and global spread of SARS-CoV-2. Nature. 2021;600(7889):408–18.

11. Chen X, Chughtai AA, Dyda A, Macintyre CR. Comparative epidemiology of Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia and South Korea. Emerg Microbes Infect [Internet]. 2017;6(6). Available from: http://dx.doi.org/10.1038/emi.2017.40

12. Widjaja L, Krauss SL, Webby RJ, Xie T, Webster RG. Matrix Gene of Influenza A Viruses Isolated from Wild Aquatic Birds: Ecology and Emergence of Influenza A Viruses. J Virol. 2004;78(16):8771–9.

13. Berche P. The Spanish flu. Press Medicale. 2022;51(3).

14. Nickol ME, Kindrachuk J. A year of terror and a century of reflection: Perspectives on the great influenza pandemic of 1918-1919. BMC Infect Dis. 2019;19(1):1–10.

15. Al-Qahtani WS, Alneghery LM, Alqahtani AQS, Alkahtani MD, Alkahtani S. A review of comparison study between corona viruses (Sars-cov, mers-cov) and novel corona virus (COVID-19). Rev Mex Ing Quim. 2020;19(September 2020):201–12.

16. Smith I, Wang LF. Bats and their virome: An important source of emerging viruses capable of infecting humans. Curr Opin Virol [Internet]. 2013;3(1):84–91. Available from: http://dx.doi.org/10.1016/j.coviro.2012.11.006

17. Frank HK, Enard D, Boyd SD. Exceptional diversity and selection pressure on coronavirus host receptors in bats compared to other mammals. Proc R Soc B Biol Sci. 2022;289(1979).

18. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci U S A. 2020;117(21).

19. Phetsouphanh C, Xu Y, Zaunders J. CD4 T cells mediate both positive and negative regulation of the immune response to HIV infection: Complex role of T follicular helper cells and regulatory T cells in pathogenesis. Front Immunol. 2015;6(JAN):1–14.

20. Piot P, Bartos M, Ghys PD, Walker N, Schwartländer B. The global impact of

HIV/AIDS. Nature. 2001;410(6831):968–73.

21. Kamorudeen RT, Adedokun KA, Olarinmoye AO. Ebola outbreak in West Africa, 2014 – 2016: Epidemic timeline, differential diagnoses, determining factors, and lessons for future response. J Infect Public Health [Internet]. 2020;13(7):956–62. Available from: https://doi.org/10.1016/j.jiph.2020.03.014

22. Ochani RK, Batra S, Shaikh A, Asad A. Nipah virus the rising epidemic: A review. Infez Med. 2019;27(2):117–27.

23. Hess IMR, Massey PD, Walker B, Middleton DJ, Wright TM. Hendra virus: what do we know? N S W Public Health Bull. 2011;22(5–6):118–22.

24. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci. 2016;73(23):4433–48.

25. Worobey M, Holmes EC. Evolutionary aspects of recombination in RNA viruses. J Gen Virol. 1999;80(10):2535–43.

26. Brown IH. The epidemiology and evolution of influenza viruses in pigs. Vet Microbiol. 2000;74(1–2):29–46.

27. Yuan S, Jiang SC, Li ZL. Analysis of Possible Intermediate Hosts of the New Coronavirus SARS-CoV-2. Front Vet Sci. 2020;7(June):1–5.

28. Long JS, Mistry B, Haslam SM, Barclay WS. Host and viral determinants of influenza A virus species specificity. Nat Rev Microbiol [Internet]. 2019;17(2):67–81. Available from: http://dx.doi.org/10.1038/s41579-018-0115-z

29. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: How selection shapes host defence genes. Nat Rev Genet [Internet]. 2010;11(1):17–30. Available from: http://dx.doi.org/10.1038/nrg2698

30. Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol. 2012;12(10):687–95.

31. Brelot A, Chakrabarti LA. CCR5 Revisited: How Mechanisms of HIV Entry Govern AIDS Pathogenesis. J Mol Biol. 2018;430(17):2557–89.

32. Li Y, Zhang Z, Yang L, Lian X, Xie Y, Li S, et al. The MERS-CoV Receptor DPP4 as a Candidate Binding Target of the SARS-CoV-2 Spike. iScience. 2020;23(6).

33. De Silva E, Stumpf MPH. HIV and the CCR5-Δ32 resistance allele. FEMS Microbiol Lett. 2004;241(1):1–12.

34. Ayora-Talavera G. Sialic acid receptors: Focus on their role in influenza infection. J Receptor Ligand Channel Res. 2018;10:1–11.

35. Matthews AE, Weiss SR, Paterson Y. Murine hepatitis virus - A model for virus-induced CNS demyelination. J Neurovirol. 2002;8(2):76–85.

36. Pereira Oliveira G, Kroon EG. Mouse hepatitis virus: A betacoronavirus model to study the virucidal activity of air disinfection equipment on surface contamination. J Virol Methods [Internet]. 2021;297(June):114274. Available from: https://doi.org/10.1016/j.jviromet.2021.114274

37. Templeton SP, Perlman S. Pathogenesis of acute and chronic central nervous system infection with variants of mouse hepatitis virus, strain JHM. Immunol Res. 2007;39(1–3):160–72.

38. Choudhury PR, Saha T, Goel S, Shah JM, Ganjewala D. Cross-species virus transmission and its pandemic potential. Bull Natl Res Cent [Internet]. 2022;46(1). Available from: https://doi.org/10.1186/s42269-022-00701-7

39. Bajenova O, Tolkunova E, Malov S, Thomas P, Tomilin A, Brien SO. The Role of the Carcinoembryonic Antigen Receptor in Colorectal Cancer Progression. J Integr Oncol. 2017;06(02).

40. Kim WM, Huang YH, Gandhi A, Blumberg RS. CEACAM1 structure and function in immunity and its therapeutic implications. Semin Immunol [Internet]. 2019;42(August):101296. Available from: https://doi.org/10.1016/j.smim.2019.101296

41. Gu A, Zhang Z, Zhang N, Tsark W, Shively JE. Generation of human CEACAM1 transgenic mice and binding of Neisseria Opa protein to their neutrophils. PLoS One. 2010;5(4):1–14.

42. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion

molecules (CEACAMs) in cancer progression and metastasis. Cancer Metastasis Rev. 2013;32(3–4):643–71.

43. Tan K, Zelus BD, Meijers R, Liu JH, Bergelson JM, Duke N, et al. Crystal structure of murine sCEACAM1a[1, 4]: A coronavirus receptor in the CEA family. EMBO J. 2002;21(9):2076–86.

44. Tchoupa AK, Schuhmacher T, Hauck CR. Signaling by epithelial members of the CEACAM family - Mucosal docking sites for pathogenic bacteria. Cell Commun Signal. 2014;12(1):1–10.

45. Zhuo Y, Yang JY, Moremen KW, Prestegard JH. Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell Adhesion Molecule 1 (CEACAM1). J Biol Chem [Internet]. 2016;291(38):20085–95. Available from: http://dx.doi.org/10.1074/jbc.M116.740050

46. Gandhi AK, Sun ZYJ, Huang YH, Kim WM, Yang C, Petsko GA, et al. Structural analysis of human CEACAM1 oligomerization. Commun Biol. 2022;5(1):1–12.

47. Gandhi AK, Sun ZYJ, Kim WM, Huang YH, Kondo Y, Bonsor DA, et al. Structural basis of the dynamic human CEACAM1 monomer-dimer equilibrium. Commun Biol. 2021;4(1):1–14.

48. Deangelis AM, Heinrich G, Dai T, Bowman TA, Patel PR, Lee SJ, et al. Carcinoembryonic antigen-related cell adhesion molecule 1: A link between insulin and lipid metabolism. Diabetes. 2008;57(9):2296–303.

49. Horst AK, Ito WD, Dabelstein J, Schumacher U, Sander H, Turbide C, et al. Carcinoembryonic antigen-related cell adhesion molecule 1 modulates vascular remodeling in vitro and in vivo. J Clin Invest. 2006;116(6):1596–605.

50. Hosomi S, Chen Z, Baker K, Chen L, Huang YH, Olszak T, et al. CEACAM1 on activated NK cells inhibits NKG2D-mediated cytolytic function and signaling. Eur J Immunol. 2013;43(9):2473–83.

51. Nagaishi T, Pao L, Lin SH, Iijima H, Kaser A, Qiao SW, et al. SHP1 Phosphatase-Dependent T Cell Inhibition by CEACAM1 Adhesion Molecule

Isoforms. Immunity. 2006;25(5):769–81.

52. Lu R, Pan H, Shively JE. CEACAM1 negatively regulates IL-1β production in LPS activated neutrophils by recruiting SHP-1 to a SYK-TLR4-CEACAM1 complex. PLoS Pathog. 2012;8(4):12–3.

53. Heckt T, Bickert T, Jeschke A, Seitz S, Schulze J, Ito WD, et al. Increased osteoclastogenesis in mice lacking the carcinoembryonic antigen-related cell adhesion molecule 1. PLoS One. 2014;9(12).

54. Chen L, Chen Z, Baker K, Halvorsen EM, Pires da Cunha A, Flak MB, et al. The Short Isoform of the CEACAM1 Receptor in Intestinal T Cells Regulates Mucosal Immunity and Homeostasis via Tfh Cell Induction. Immunity [Internet]. 2012;37(5):930–46. Available from: http://dx.doi.org/10.1016/j.immuni.2012.07.016

55. Chen Z, Chen L, Qiao SW, Nagaishi T, Blumberg RS. Carcinoembryonic Antigen-Related Cell Adhesion Molecule 1 Inhibits Proximal TCR Signaling by Targeting ZAP-70. J Immunol. 2008;180(9):6085–93.

56. Hoda Badr, Cindy L. Carmack, Deborah A. Kashy, Massimo Cristofanilli and TAR, Kantrowitz, Hoda Badr, Cindy L. Carmack, Deborah A. Kashy, Massimo Cristofanilli and TAR. CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. Bone. 2011;23(1):1–7.

57. Klaile E, Müller MM, Kannicht C, Singer BB, Lucka L. CEACAM1 functionally interacts with filamin A and exerts a dual role in the regulation of cell migration. J Cell Sci. 2005;118(23):5513–24.

58. Cheung PH, Thompson NL, Earley K, Culic O, Hixson D, Lin SH. Cell-CAM105 isoforms with different adhesion functions are coexpressed in adult rat tissues and during liver development. J Biol Chem [Internet]. 1993;268(9):6139–46. Available from: http://dx.doi.org/10.1016/S0021-9258(18)53230-0

59. Poy MN, Yang Y, Rezaei K, Fernström MA, Lee AD, Kido Y, et al. CEACAM1 regulates insulin clearance in liver. Nat Genet. 2002;30(3):270–6.

60. Abou-Rjaily GA, Lee SJ, May D, Al-Share QY, DeAngelis AM, Ruch RJ, et

al. CEACAM1 modulates epidermal growth factor receptor–mediated cell proliferation. J Clin Invest. 2004;114(7):944–52.

61. Miura HS, Nakagaki K, Taguchi F. N-Terminal Domain of the Murine Coronavirus Receptor CEACAM1 Is Responsible for Fusogenic Activation and Conformational Changes of the Spike Protein. J Virol. 2004;78(1):216–23.

62. Shang J, Wan Y, Liu C, Yount B, Gully K, Yang Y, et al. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. PLoS Pathog [Internet]. 2020;16(3):1–20. Available from: http://dx.doi.org/10.1371/journal.ppat.1008392

63. Villullas S, Hill DJ, Sessions RB, Rea J, Virji M. Mutational analysis of human CEACAM1: The potential of receptor polymorphism in increasing host susceptibility to bacterial infection. Cell Microbiol. 2007;9(2):329–46.

64. Dveksler GS, Basile AA, Cardellichio CB, Holmes K V. Mouse hepatitis virus receptor activities of an MHVR/mph chimera and MHVR mutants lacking N-linked glycosylation of the N-terminal domain. J Virol. 1995;69(1):543–6.

65. Peng G, Sun D, Rajashankar KR, Qian Z, Holmes K V., Li F. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc Natl Acad Sci U S A. 2011;108(26):10696–701.

66. Upham NS, Esselstyn JA, Jetz W. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. Vol. 17, PLoS Biology. 2019. 1–44 p.

67. Robin S, Tacher S, Rimbault M, Vaysse A, Dréano S, André C, et al. Genetic diversity of canine olfactory receptors. 2009;16:1–16.

68. Geraghty DE, Thorball CW, Fellay J, Thomas R. Effect of Fc receptor genetic diversity on HIV-1 disease pathogenesis. Front Immunol. 2019;10(MAY).

69. Civelli O, Bunzow JR, Grandy DK. MOLECULAR DIVERSITY OF. 1992;(February).

70. Nakanishi S, Science S, Series N, Oct N. Molecular Diversity of Glutamate Receptors and Implications for Brain Function Published by : American Association for the Advancement of Science olecul ar Diversity Of Glutamate

Receptors and Implications for Brain Function. Science (80-) [Internet]. 1992;258(5082):597–603. Available from: http://about.jstor.org/terms

71. Jensen ASM, Sparre-Ulrich AH, Davis-Poynter N, Rosenkilde MM. Structural diversity in conserved regions like the DRY-motif among viral 7TM receptors - A consequence of evolutionary pressure? Adv Virol. 2012;2012.

72. Norkin LC. Virus receptors: Implications for pathogenesis and the design of antiviral agents. Clin Microbiol Rev. 1995;8(2):293–315.

73. Bastidas-Legarda LY, Khakoo SI. Conserved and variable natural killer cell receptors: diverse approaches to viral infections. Immunology. 2019;156(4):319–28.

74. Blish CA. Natural killer cell diversity in viral infection: Why and how much? Pathog Immun. 2016;1(1):165–92.

75. Du G, Ding Y, Li H, Wang X, Wang J, Sun Y, et al. A framework for predicting potential host ranges of pathogenic viruses based on receptor ortholog analysis. bioRxiv [Internet]. 2020; Available from: https://www.biorxiv.org/content/early/2020/12/07/2020.12.07.414292

76. Bae SE, Son HS. Classification of viral zoonosis through receptor pattern analysis. BMC Bioinformatics. 2011;12:1–7.

77. Zeltina A, Bowden TA, Lee B. Emerging Paramyxoviruses: Receptor Tropism and Zoonotic Potential. PLoS Pathog. 2016;12(2):1–8.

78. Lopes LR, de Mattos Cardillo G, Paiva PB. Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host. Brazilian J Microbiol. 2020;51(4):1593–9.

79. Liu P, Jiang JZ, Wan XF, Hua Y, Li L, Zhou J, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog [Internet]. 2020;16(5):1–13. Available from: http://dx.doi.org/10.1371/journal.ppat.1008421

80. Zhao J, Cui W, Tian BP. The Potential Intermediate Hosts for SARS-CoV-2. Front Microbiol. 2020;11(September):1–11.

81. Тарелкина ТВ, Галибина НА, Мощенская ЮЛ, Новицкая ЛЛ. In silico анализ регуляторных цис -элементов в промоторах генов, кодирующих апопластную инвертазу и сахарозосинтазу у березы повислой . Онтогенез. 2020;51(5):386–400.

82. Gupta SK, Minocha R, Thapa PJ, Srivastava M, Dandekar T. Role of the Pangolin in Origin of SARS-CoV-2: An Evolutionary Perspective. Int J Mol Sci. 2022;23(16).

83. Shi W, Shi M, Que TC, Cui XM, Ye RZ, Xia LY, et al. Trafficked Malayan pangolins contain viral pathogens of humans. Nat Microbiol. 2022;7(8):1259–69.

APPENDICES

Appendix 1. Diversity_Script.R

```
rm(list = ls()) #to clear environment
library(readr)
CEACAM1 <-
read_csv("C:/Users/DELL/Desktop/CEACAM1_Diversity/CEACAM
1.csv")
View(CEACAM1)
library(plyr)
library(dplyr)
CEACAM1full <-CEACAM1[,-1]
nrow(CEACAM1full)
str(CEACAM1full)

CEACAM1full<-CEACAM1full[CEACAM1full[,6]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,7]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,8]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,9]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,10]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,11]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,12]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,13]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,14]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,15]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,16]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,17]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,18]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,19]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,20]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,21]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,22]!="",]
CEACAM1full<-CEACAM1full[CEACAM1full[,23]!="",]

Unique_types <- CEACAM1full %>%
```

```
  count( `63`, `64`, `68`, `73`, `75`, `76`,`81`, `83`,
`85`, `86`, `88`, `90`, `91`, `92`, `93`,`127`
,`128`,`130`)
Unique_types$Seq <-rep("",nrow(Unique_types))  # worked
with dplyr
for (i in 1:18){
  for (j in 1:nrow(Unique_types)){
    Unique_types[j,20]<-
paste(Unique_types[j,20],Unique_types[j,i],sep="")
  }


}
CEACAM1full$Seq <-rep("",nrow(CEACAM1full))
for (i in 6:23){
  for (j in 1:nrow(CEACAM1full)){
    CEACAM1full[j,24]<-
paste(CEACAM1full[j,24],CEACAM1full[j,i],sep="")
  }


}
Unique_types$type <-seq(1:nrow(Unique_types))
head(Unique_types)
head(CEACAM1full)
CEACAM1full$type <-rep("",nrow(CEACAM1full))


for (i in 1:nrow(CEACAM1full)) {
  Seq <- as.character(CEACAM1full[i, 24])
  type <- Unique_types[Unique_types$Seq == Seq, "type"]
  if (nrow(type) > 0) {
    CEACAM1full[i, 25] <- as.character(type[1, 1])
  } else {
    CEACAM1full[i, 25] <- NA
  }
}
```

```
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/Unique_types_20
-10-24_1.csv"

write.csv(Unique_types, file = file_path, row.names =
TRUE)

file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/CEACAM1full_20-
10-24_1.csv"

write.csv(CEACAM1full, file = file_path, row.names =
TRUE)

# Diversity Index

Residues <-CEACAM1full

Residues[is.na(Residues)] <- "NA"

Rodentia <-Residues[Residues$Order=="RODENTIA",]

NotRodents <-Residues[Residues$Order!="RODENTIA",]

library(vegan)

Thing <- xtabs(~`63`, data=Residues)

AminoCount <-matrix(nrow=18, ncol=20) # to make new
matrix named AminoCount

colnames(AminoCount)<-
c("A","C","D","E","F","G","H","I","K","L","M","N","P","Q
","R","S","T","V","W","Y")

rownames(AminoCount)<-colnames(Residues[6:23])


for(i in 6:23){
  AminoCount[i-5,1]<-nrow(Residues[Residues[,i]=="A",])
  AminoCount[i-5,2]<-nrow(Residues[Residues[,i]=="C",])
  AminoCount[i-5,3]<-nrow(Residues[Residues[,i]=="D",])
  AminoCount[i-5,4]<-nrow(Residues[Residues[,i]=="E",])
  AminoCount[i-5,5]<-nrow(Residues[Residues[,i]=="F",])
  AminoCount[i-5,6]<-nrow(Residues[Residues[,i]=="G",])
  AminoCount[i-5,7]<-nrow(Residues[Residues[,i]=="H",])
  AminoCount[i-5,8]<-nrow(Residues[Residues[,i]=="I",])
  AminoCount[i-5,9]<-nrow(Residues[Residues[,i]=="K",])
  AminoCount[i-5,10]<-nrow(Residues[Residues[,i]=="L",])
  AminoCount[i-5,11]<-nrow(Residues[Residues[,i]=="M",])
  AminoCount[i-5,12]<-nrow(Residues[Residues[,i]=="N",])
```

```
  AminoCount[i-5,13]<-nrow(Residues[Residues[,i]=="P",])
  AminoCount[i-5,14]<-nrow(Residues[Residues[,i]=="Q",])
  AminoCount[i-5,15]<-nrow(Residues[Residues[,i]=="R",])
  AminoCount[i-5,16]<-nrow(Residues[Residues[,i]=="S",])
  AminoCount[i-5,17]<-nrow(Residues[Residues[,i]=="T",])
  AminoCount[i-5,18]<-nrow(Residues[Residues[,i]=="V",])
  AminoCount[i-5,19]<-nrow(Residues[Residues[,i]=="W",])
  AminoCount[i-5,20]<-nrow(Residues[Residues[,i]=="Y",])
}
colSums(AminoCount)
rowSums(AminoCount)
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/AminoCount_12-
12-24.csv"
write.csv(AminoCount, file = file_path, row.names =
TRUE)


# These values provide a measure of diversity for the
amino acid
# +distributions in different columns of the Residues`
data frame.
Shannons <-diversity(AminoCount, index="shannon")


# Shannon’s Diversity Index**: Reflects the diversity
accounting for abundance and evenness.
# +Species Richness: Shows the count of unique species
(amino acids) in each row.
# +Species Evenness: Indicates how evenly the species
are distributed in each row, providing insights into the
relative abundance distribution.
# these commands measures together give a comprehensive
picture of biodiversity in your samples, showing not
only how many different types there are but also how
they are distributed.


Species <-specnumber(AminoCount)
Evenness <-Shannons/log(Species)
```

```
# This command is used to create a single, organized
matrix that

# + contains the biodiversity measures (Shannon
diversity index, species richness, and species evenness)
for the subset of data related to "Bat" (Chiroptera).

cbind(Shannons, Species, Evenness)

combined_data <- cbind(Shannons, Species, Evenness)

file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/Shannons_Index_
08-01-25.csv"

write.csv(combined_data, file = file_path, row.names =
TRUE)


#Diversity Index of Rodents


Thing <- xtabs(~`63`, data=Residues)

AminoCountRodent <-matrix(nrow=18, ncol=20)

colnames(AminoCountRodent)<-
c("A","C","D","E","F","G","H","I","K","L","M","N","P","Q
","R","S","T","V","W","Y")

rownames(AminoCountRodent)<-colnames(Rodentia[6:23])


for(i in 6:23){

  AminoCountRodent[i-5,1]<-
nrow(Rodentia[Rodentia[,i]=="A",])

  AminoCountRodent[i-5,2]<-
nrow(Rodentia[Rodentia[,i]=="C",])

  AminoCountRodent[i-5,3]<-
nrow(Rodentia[Rodentia[,i]=="D",])

  AminoCountRodent[i-5,4]<-
nrow(Rodentia[Rodentia[,i]=="E",])

  AminoCountRodent[i-5,5]<-
nrow(Rodentia[Rodentia[,i]=="F",])

  AminoCountRodent[i-5,6]<-
nrow(Rodentia[Rodentia[,i]=="G",])

  AminoCountRodent[i-5,7]<-
nrow(Rodentia[Rodentia[,i]=="H",])

  AminoCountRodent[i-5,8]<-
nrow(Rodentia[Rodentia[,i]=="I",])
```

```
  AminoCountRodent[i-5,9]<-
nrow(Rodentia[Rodentia[,i]=="K",])
  AminoCountRodent[i-5,10]<-
nrow(Rodentia[Rodentia[,i]=="L",])
  AminoCountRodent[i-5,11]<-
nrow(Rodentia[Rodentia[,i]=="M",])
  AminoCountRodent[i-5,12]<-
nrow(Rodentia[Rodentia[,i]=="N",])
  AminoCountRodent[i-5,13]<-
nrow(Rodentia[Rodentia[,i]=="P",])
  AminoCountRodent[i-5,14]<-
nrow(Rodentia[Rodentia[,i]=="Q",])
  AminoCountRodent[i-5,15]<-
nrow(Rodentia[Rodentia[,i]=="R",])
  AminoCountRodent[i-5,16]<-
nrow(Rodentia[Rodentia[,i]=="S",])
  AminoCountRodent[i-5,17]<-
nrow(Rodentia[Rodentia[,i]=="T",])
  AminoCountRodent[i-5,18]<-
nrow(Rodentia[Rodentia[,i]=="V",])
  AminoCountRodent[i-5,19]<-
nrow(Rodentia[Rodentia[,i]=="W",])
  AminoCountRodent[i-5,20]<-
nrow(Rodentia[Rodentia[,i]=="Y",])
}
colSums(AminoCountRodent)
rowSums(AminoCountRodent)
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/AminoCountRoden
t_08-01-25.csv"
write.csv(AminoCountRodent, file = file_path, row.names
= TRUE)
ShannonsRodent <-diversity(AminoCountRodent,
index="shannon")
SpeciesRodent <-specnumber(AminoCountRodent)
EvennessRodent <-ShannonsRodent/log(SpeciesRodent)
cbind(ShannonsRodent, SpeciesRodent, EvennessRodent)
combined_data1 <- cbind(ShannonsRodent, SpeciesRodent,
EvennessRodent)
```

```
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/Shannons_Index_
Rodent_08-01-25.csv"
write.csv(combined_data1, file = file_path, row.names =
TRUE)


#Diversity Index of Non_Rodents


Thing <- xtabs(~`63`, data=Residues)
AminoCountNonRodent <-matrix(nrow=18, ncol=20)
colnames(AminoCountNonRodent)<-
c("A","C","D","E","F","G","H","I","K","L","M","N","P","Q
","R","S","T","V","W","Y")
rownames(AminoCountNonRodent)<-
colnames(NotRodents[6:23])


for(i in 6:23){
  AminoCountNonRodent[i-5,1]<-
nrow(NotRodents[NotRodents[,i]=="A",])
  AminoCountNonRodent[i-5,2]<-
nrow(NotRodents[NotRodents[,i]=="C",])
  AminoCountNonRodent[i-5,3]<-
nrow(NotRodents[NotRodents[,i]=="D",])
  AminoCountNonRodent[i-5,4]<-
nrow(NotRodents[NotRodents[,i]=="E",])
  AminoCountNonRodent[i-5,5]<-
nrow(NotRodents[NotRodents[,i]=="F",])
  AminoCountNonRodent[i-5,6]<-
nrow(NotRodents[NotRodents[,i]=="G",])
  AminoCountNonRodent[i-5,7]<-
nrow(NotRodents[NotRodents[,i]=="H",])
  AminoCountNonRodent[i-5,8]<-
nrow(NotRodents[NotRodents[,i]=="I",])
  AminoCountNonRodent[i-5,9]<-
nrow(NotRodents[NotRodents[,i]=="K",])
  AminoCountNonRodent[i-5,10]<-
nrow(NotRodents[NotRodents[,i]=="L",])
  AminoCountNonRodent[i-5,11]<-
nrow(NotRodents[NotRodents[,i]=="M",])
```

```
  AminoCountNonRodent[i-5,12]<-
nrow(NotRodents[NotRodents[,i]=="N",])
  AminoCountNonRodent[i-5,13]<-
nrow(NotRodents[NotRodents[,i]=="P",])
  AminoCountNonRodent[i-5,14]<-
nrow(NotRodents[NotRodents[,i]=="Q",])
  AminoCountNonRodent[i-5,15]<-
nrow(NotRodents[NotRodents[,i]=="R",])
  AminoCountNonRodent[i-5,16]<-
nrow(NotRodents[NotRodents[,i]=="S",])
  AminoCountNonRodent[i-5,17]<-
nrow(NotRodents[NotRodents[,i]=="T",])
  AminoCountNonRodent[i-5,18]<-
nrow(NotRodents[NotRodents[,i]=="V",])
  AminoCountNonRodent[i-5,19]<-
nrow(NotRodents[NotRodents[,i]=="W",])
  AminoCountNonRodent[i-5,20]<-
nrow(NotRodents[NotRodents[,i]=="Y",])
}
colSums(AminoCountNonRodent)
rowSums(AminoCountNonRodent)
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/AminoCountNonRo
dent_08-01-25.csv"
write.csv(AminoCountNonRodent, file = file_path,
row.names = TRUE)
ShannonsNonRodent <-diversity(AminoCountNonRodent,
index="shannon")
SpeciesNonRodent <-specnumber(AminoCountNonRodent)
EvennessNonRodent <-
ShannonsNonRodent/log(SpeciesNonRodent)
cbind(ShannonsNonRodent, SpeciesNonRodent,
EvennessNonRodent)
combined_data2 <- cbind(ShannonsNonRodent,
SpeciesNonRodent, EvennessNonRodent)
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/Shannons_Index_
Non_Rodent_08-01-25.csv"
write.csv(combined_data2, file = file_path, row.names =
TRUE)
```

```
#Similarity Score
Residues$humanorsimilar <-rep(0,nrow(Residues))
Residues$bad <-rep(0,nrow(Residues))
Residues$count <-rep(0,nrow(Residues))
Residues$micesimilar<-rep(0,nrow(Residues))
Residues$micebad<-rep(0,nrow(Residues))
Residues$micecount<-rep(0,nrow(Residues))
names(Residues)

for (i in 1:nrow(Residues)){

  if(Residues[i,6] %in% c("F","Y","L","I"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,6] %in% c("R","E","G","P"))
{Residues[i,27] <- (Residues[i,27] + 1)}
  if(Residues[i,6] %in% c("G","A","V","L","I"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,6] %in% c("F","Y","E","P","R"))
{Residues[i,30] <- (Residues[i,30] + 1)}

  if(Residues[i,7] %in% c("G","A","V","L","I"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,7] %in% c("F","K","R")) {Residues[i,27]
<- (Residues[i,27] + 1)}
  if(Residues[i,7] %in% c("G","A","V","L","I"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,7] %in% c("K","R")) {Residues[i,30] <-
(Residues[i,30] + 1)}

  if(Residues[i,8] %in% c("Y","F")) {Residues[i,26] <-
(Residues[i,26] + 1)}
  if(Residues[i,8] %in% c("H")) {Residues[i,27] <-
(Residues[i,27] + 1)}
  if(Residues[i,8] %in% c("Y","F")) {Residues[i,29] <-
(Residues[i,29] + 1)}
  if(Residues[i,8] %in% c("H")) {Residues[i,30] <-
(Residues[i,30] + 1)}
```

```
  if(Residues[i,9] %in% c("V","A","I","T","L"))
{Residues[i,26] <- (Residues[i,26] + 1)}

  if(Residues[i,9] %in% c("P","E","Q")) {Residues[i,27]
<- (Residues[i,27] + 1)}

  if(Residues[i,9] %in% c("A","I","T","S"))
{Residues[i,29] <- (Residues[i,29] + 1)}

  if(Residues[i,9] %in% c("P","E","Q")) {Residues[i,30]
<- (Residues[i,30] + 1)}


  if(Residues[i,10] %in% c("G","A","V","L","I"))
{Residues[i,26] <- (Residues[i,26] + 1)}

  if(Residues[i,10] %in%
c("F","D","E","P","H","K")){Residues[i,27] <-
(Residues[i,27] + 1)}

  if(Residues[i,10] %in% c("F","G","A","V","L","I"))
{Residues[i,29] <- (Residues[i,29] + 1)}

  if(Residues[i,10] %in% c("K","D","E","H"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,11] %in% c("N","T","S")) {Residues[i,26]
<- (Residues[i,26] + 1)}

  if(Residues[i,11] %in% c("R","K","D","H","E"))
{Residues[i,27] <- (Residues[i,27] + 1)}

  if(Residues[i,11] %in% c("D","E","S","N","H","T"))
{Residues[i,29] <- (Residues[i,29] + 1)}

  if(Residues[i,11] %in% c("K","R")) {Residues[i,30] <-
(Residues[i,30] + 1)}


  if(Residues[i,12] %in% c("G","A","L","S"))
{Residues[i,26] <- (Residues[i,26] + 1)}

  if(Residues[i,12] %in% c("Y","K","R")) {Residues[i,27]
<- (Residues[i,27] + 1)}

  if(Residues[i,12] %in% c("H","K","R")) {Residues[i,29]
<- (Residues[i,29] + 1)}

  if(Residues[i,12] %in% c("L","Y")) {Residues[i,30] <-
(Residues[i,30] + 1)}


  if(Residues[i,13] %in% c("A","V","L","I"))
{Residues[i,26] <- (Residues[i,26] + 1)}
```

```
  if(Residues[i,13] %in% c("R","K","E")) {Residues[i,27]
<- (Residues[i,27] + 1)}
  if(Residues[i,13] %in% c("V","A","L","T","I","M"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,13] %in% c("E","K","R")) {Residues[i,30]
<- (Residues[i,30] + 1)}


  if(Residues[i,14] %in% c("G","A","I","L"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,14] %in% c("R","D","E")) {Residues[i,27]
<- (Residues[i,27] + 1)}
  if(Residues[i,14] %in% c("N","Q","S","T"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,14] %in% c("R","E","D")) {Residues[i,30]
<- (Residues[i,30] + 1)}


  if(Residues[i,15] %in% c("N","T","S")) {Residues[i,26]
<- (Residues[i,26] + 1)}
  if(Residues[i,15] %in% c("D","K","E","P"))
{Residues[i,27] <- (Residues[i,27] + 1)}
  if(Residues[i,15] %in% c("S","T","N","I"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,15] %in% c("D","E","V","K","L"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,16] %in% c("Q","N","T","S"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,16] %in% c("V","K","I","L","R","E"))
{Residues[i,27] <- (Residues[i,27] + 1)}
  if(Residues[i,16] %in% c("M","L","V","I"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,16] %in% c("E","K","G","R"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,17] %in% c("N","A","I","T","S"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,17] %in% c("F","K")) {Residues[i,27] <-
(Residues[i,27] + 1)}
  if(Residues[i,17] %in% c("F","L","I")) {Residues[i,29]
<- (Residues[i,29] + 1)}
```

```
  if(Residues[i,17] %in% c("K")) {Residues[i,30] <-
(Residues[i,30] + 1)}


  if(Residues[i,18] %in% c("P")) {Residues[i,26] <-
(Residues[i,26] + 1)}
  if(Residues[i,18] %in% c("T","R","F","D","K"))
{Residues[i,27] <- (Residues[i,27] + 1)}
  if(Residues[i,18] %in% c("T","I","S","N"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,18] %in% c("K","D","E","F","R","Q","P"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,19] %in% c("G")) {Residues[i,26] <-
(Residues[i,26] + 1)}
  if(Residues[i,19] %in% c("G")) {Residues[i,29] <-
(Residues[i,29] + 1)}


  if(Residues[i,20] %in% c("S","P","A")) {Residues[i,26]
<- (Residues[i,26] + 1)}
  if(Residues[i,20] %in% c("R","F")) {Residues[i,27] <-
(Residues[i,27] + 1)}
  if(Residues[i,20] %in% c("S","Q")) {Residues[i,29] <-
(Residues[i,29] + 1)}
  if(Residues[i,20] %in% c("A","F","V","R","L"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,21] %in% c("N","Q","A","I","T","S"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,21] %in% c("D","E","R","K"))
{Residues[i,27] <- (Residues[i,27] + 1)}
  if(Residues[i,21] %in% c("D","E")) {Residues[i,29] <-
(Residues[i,29] + 1)}
  if(Residues[i,21] %in%
c("L","I","M","V","P","A","R","K")) {Residues[i,30] <-
(Residues[i,30] + 1)}


  if(Residues[i,22] %in% c("D","S","Q","T"))
{Residues[i,26] <- (Residues[i,26] + 1)}
  if(Residues[i,22] %in% c("H","N","R","K"))
{Residues[i,27] <- (Residues[i,27] + 1)}
```

```
  if(Residues[i,22] %in% c("N","Q","S","T"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,22] %in% c("D","H","R","K"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  if(Residues[i,23] %in%
c("V","L","I","T","A")){Residues[i,26] <-
(Residues[i,26] + 1)}
  if(Residues[i,23] %in%
c("R","E","Y","D","K")){Residues[i,27] <-
(Residues[i,27] + 1)}
  if(Residues[i,23] %in% c("R","K","Q","N"))
{Residues[i,29] <- (Residues[i,29] + 1)}
  if(Residues[i,23] %in% c("D","E","I","W","V"))
{Residues[i,30] <- (Residues[i,30] + 1)}


  Count <-Residues[i,6:23]!=""
  Residues[i,28]<-length(Count[Count==T])
  micecount <-Residues[i,6:23]!=""
  Residues[i,31]<-length(micecount[micecount==T])
}
# worked with plyr


Residues$count; Residues$micecount
names(Residues)
Residues$humanscore <-(Residues$humanorsimilar-
Residues$bad)/Residues$count
Residues$humanscore
Residues$micescore <-(Residues$micesimilar-
Residues$micebad)/Residues$micecount
file_path <-
"C:/Users/DELL/Desktop/CEACAM1_Diversity/CEACAM1_Likelin
ess_Score.csv"
write.csv(Residues, file = file_path, row.names = TRUE)
#all commands worked successfuly
```

Appendix 2. Pruned_MCC.R

```
rm(list = ls()) # to clear environment
require(phytools)
```

```
require(geiger)

require(ape)

require(jsonlite)

UphamMCC <-
read.nexus(file="C:/Users/DELL/Desktop/CEACAM1_SP/Upham_
MCC_no_fossil.tre.txt")

str(UphamMCC)

is.ultrametric(UphamMCC)

# if FALSE ;The evolutionary distances from the root to
the tips of the tree (leaves) are not equal.

Tips<-
read.csv("C:/Users/DELL/Desktop/CEACAM1_SP/CEACAM1_Upham
_Names.csv")

head(Tips)

tail(Tips)


rownames(Tips)<-Tips$Upham_name # This command assigns
the Upham_name column of the Tips data frame to the row
names of Tips.

Tips$Upham_name

Tester <-extract.clade(UphamMCC,5400) #Extracts a
specific clade (subtree) from the phylogenetic tree
(UphamMCC) that contains node or tip 5400.

plot(Tester) #This command will plot the extracted clade
(Tester) as a phylogenetic tree.


drop <-name.check(UphamMCC, Tips) # Checks if the names
in the phylogenetic tree (UphamMCC) match the names in
the Tips data frame.

drop$data_not_tree #  This command prints or accesses
the vector of names that are in the Tips data frame but
not in the phylogenetic tree (UphamMCC).


PrunedUphamMCC <-drop.tip(UphamMCC, drop$tree_not_data)

plot(PrunedUphamMCC)


# rename

PrunedUphamMCC$tip.label

Pruned2 <-PrunedUphamMCC
```

```
head(Tips)
tail(Tips)


for(j in 1:nrow(Tips)){
  Temp<-Pruned2$tip.label[j]
  Pruned2$tip.label[Pruned2$tip.label==Temp]<-
as.character(Tips[Tips$Upham_name==Temp,2])
}


Pruned2$tip.label
layout(c(1))
par(mar=c(1,1,1,1))
plot(Pruned2)


write.nexus(Pruned2,
file="C:/Users/DELL/Desktop/CEACAM1_SP/Upham_MCC_renamed
_91spp_forCEACAM1_2024-10-20.nexus")
write.tree(Pruned2,
file="C:/Users/DELL/Desktop/CEACAM1_SP/Upham_MCC_renamed
_91spp_forCEACAM1_2024-10-20.newick")


# now getting a mice versus not mice tree
Mices <-Tips[Tips$Order=="RODENTIA",]
rownames(Mices)<-Mices$Species
head(Mices[,1:3])
Micesdrop <-name.check(Pruned2, Mices)
MiceCEACAM1tree <-
drop.tip(Pruned2,Micesdrop$tree_not_data)
plot(MiceCEACAM1tree)


write.tree(MiceCEACAM1tree,
file="C:/Users/DELL/Desktop/CEACAM1_SP/Upham_MCC_CEACAM1
_Rodentsonly_2024-10-20.newick")
length(MiceCEACAM1tree$tip.label)


NotMice <- Tips[Tips$Order != "RODENTIA", ]
rownames(NotMice) <- NotMice$Species
```

```
NotMicedrop <-name.check(Pruned2, NotMice)

NotMiceCEACAM1tree <-
drop.tip(Pruned2,NotMicedrop$tree_not_data)

plot(NotMiceCEACAM1tree)


write.tree(NotMiceCEACAM1tree,
file="C:/Users/DELL/Desktop/CEACAM1_SP/Upham_MCC_CEACAM1
_NotRodents_2024-10-20.newick")

length(NotMiceCEACAM1tree$tip.label)

#all worked
```

Appendix 3. Extraction_JSONS_HyPhy.R

```
rm(list = ls()) #remove datasets

graphics.off() #remove plots

#Directory
(file="C:/Users/DELL/Desktop/CEACAM1_SP/Scores_for_plott
ing.csv")

## Extracting the aBSREL results from the json


library(jsonlite)


JSONS <-
list.files(path="C:/Users/DELL/Desktop/HyPhy_Reproduce/"
, pattern = ".json", full.names=F,recursive=F)


for (j in 1:length(JSONS)) {

  # Try reading the JSON file

  Thing <-
tryCatch(read_json(paste("C:/Users/DELL/Desktop/HyPhy_Re
produce/", JSONS[j], sep=""), simplifyVector = F),
                     error = function(e) {

                       warning(paste("Error reading:",
JSONS[j]))

                       return(NULL)

                     })


  # Skip to next if Thing is NULL (file couldn't be
read)

  if (is.null(Thing)) next
```

```
  OfInterest <- Thing$`branch attributes`$`0`
  if (is.null(OfInterest)) next

  ResultTable <- data.frame(matrix(ncol=4,
nrow=length(OfInterest)))

  for (i in 1:length(OfInterest)) {
    Temp <- OfInterest[[i]]
    ResultTable[i, 1] <- names(OfInterest)[i]
    ResultTable[i, 2] <- Temp$LRT
    ResultTable[i, 3] <- Temp$`Uncorrected P-value`
    ResultTable[i, 4] <- Temp$`Corrected P-value`
  }

  colnames(ResultTable) <- c("Name", "LRT", "Uncorrected
p-value", "Corrected p-value")
  write.csv(ResultTable,
file=paste("C:/Users/DELL/Desktop/HyPhy_Reproduce/",
strsplit(JSONS[j], ".json")[[1]][1],
"_ABSRELsummary.csv", sep=""), row.names = FALSE)
}
# Worked

## Extracting the MEME results from the JSON

library(jsonlite)
JSONS <-
list.files(path="C:/Users/DELL/Desktop/HyPhy_Reproduce/"
, pattern = "*.json", full.names=F,recursive=F)

for (i in 1:length(JSONS)) {
  MEME1 <-
read_json(paste("C:/Users/DELL/Desktop/HyPhy_Reproduce/"
, JSONS[i], sep=""), simplifyVector = T)
  CodonTable <- data.frame(MEME1$MLE$content)
  colnames(CodonTable) <- MEME1$MLE$headers[,1]
```

```
  write.csv(CodonTable,
file=paste("C:/Users/DELL/Desktop/HyPhy_Reproduce/",
strsplit(JSONS[i], ".json")[[1]][1], "_MEMEsummary.csv",
sep=""))
}
#worked
```

Appendix 4. MEME_Result_Analysis_Script.R

```
#clear environment
rm(list = ls())
# Load necessary libraries
library(dplyr)
library(readr)


# Step 1: Load your MEME results CSV
meme_data<-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/HyP
hy_23_MEME_Extracted.csv")


# Step 2: Define contact residue positions (add your
actual positions)
contact_residues <- c(63,  64,  68,  73,  75,  76,
     81,  83,  85,  86,  88,  90,  91,  92,  93,
     127, 128, 130)
# positions of contact residues


# Step 3: Categorize each site as contact or non-contact
meme_data <- meme_data %>%
  mutate(is_contact = ifelse(X %in% contact_residues,
"contact", "non-contact"))
#'X' for first column


# Step 4: Filter sites under selection based on p-values
under_selection_0.05 <- meme_data %>% filter(p.value <
0.05)
under_selection_0.1 <- meme_data %>% filter(p.value <
0.1)
```

```
# Step 5: Count contact and non-contact sites under
selection for p < 0.05 and p < 0.1

contact_sel_0.05 <- under_selection_0.05 %>%
filter(is_contact == "contact") %>% nrow()

contact_notsel_0.05 <- meme_data %>% filter(is_contact
== "contact", p.value >= 0.05) %>% nrow()

noncontact_sel_0.05 <- under_selection_0.05 %>%
filter(is_contact == "non-contact") %>% nrow()

noncontact_notsel_0.05 <- meme_data %>%
filter(is_contact == "non-contact", p.value >= 0.05) %>%
nrow()


contact_sel_0.1 <- under_selection_0.1 %>%
filter(is_contact == "contact") %>% nrow()

contact_notsel_0.1 <- meme_data %>% filter(is_contact ==
"contact", p.value >= 0.1) %>% nrow()

noncontact_sel_0.1 <- under_selection_0.1 %>%
filter(is_contact == "non-contact") %>% nrow()

noncontact_notsel_0.1 <- meme_data %>% filter(is_contact
== "non-contact", p.value >= 0.1) %>% nrow()


# Step 6: Create a summary table

summary_table <- data.frame(

  Test = "MEME",

  contact_sel_0.05 = contact_sel_0.05,

  contact_notsel_0.05 = contact_notsel_0.05,

  noncontact_sel_0.05 = noncontact_sel_0.05,

  noncontact_notsel_0.05 = noncontact_notsel_0.05,

  contact_sel_0.1 = contact_sel_0.1,

  contact_notsel_0.1 = contact_notsel_0.1,

  noncontact_sel_0.1 = noncontact_sel_0.1,

  noncontact_notsel_0.1 = noncontact_notsel_0.1

)


# Step 7: Write the summary table to a CSV

write.csv(summary_table,
file="C:/Users/DELL/Desktop/HyPhy_Reproduce/meme_23_summ
ary_analysis_23-10-24.csv")
```

```
meme_data1 <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/mem
e_91_summary_analysis_23-10-24.csv")

meme_data2 <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/mem
e_68_summary_analysis_23-10-24.csv")

meme_data3 <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/mem
e_23_summary_analysis_23-10-24.csv")


merged_meme_data <- bind_rows(meme_data1, meme_data2,
meme_data3)

write.csv(merged_meme_data,
file="C:/Users/DELL/Desktop/HyPhy_Reproduce/merged_meme_
analysis_23-10-24.csv")


#Worked

## Fisher's exact tests -------

# Fisher's Exact Test for whole data set; p>0.05

fisher.test(matrix(c(12, 6, 76, 434), nrow = 2, byrow =
TRUE)) #p-value = 1.832e-06

print(odds_ratio_whole <- (12 * 434) / (6 * 76))
#11.42105


# Fisher's Exact Test for not_rodents data set; p>0.05

fisher.test(matrix(c(10, 8, 69, 441), nrow = 2, byrow =
TRUE)) #p-value = 5.259e-05

print(odds_ratio_not_rodents <- (10 * 441) / (8 * 69))
#7.98913


# Fisher's Exact Test for Rodents ; p>0.05

fisher.test(matrix(c(6, 12, 42, 468), nrow = 2, byrow =
TRUE)) #p-value = 0.003304

print(odds_ratio_rodents <- (6 * 468) / (12 * 42))
#5.571429


# Fisher's Exact Test for whole data set; p>0.1
```

```
fisher.test(matrix(c(13, 5, 104, 406), nrow = 2, byrow =
TRUE)) #p-value = 5.468e-06

print(odds_ratio_whole_p1 <- (13 * 406) / (5 * 104))
#10.15


# Fisher's Exact Test for not_rodents data set; p>0.1

fisher.test(matrix(c(10, 8, 91, 419), nrow = 2, byrow =
TRUE)) #p-value = 0.0004905

print(odds_ratio_not_rodents_p1 <- (10 * 419) / (8 *
91)) #5.755495


# Fisher's Exact Test for Rodents ; p>0.1

fisher.test(matrix(c(8, 10, 61, 449), nrow = 2, byrow =
TRUE)) #p-value = 0.0008557

print(odds_ratio_rodents_p1 <- (8 * 449) / (10 * 61)) #
5.888525


#Worked
```

Appendix 5. aBSREL_Result_Analysis_Script.R

```
rm(list = ls())

# Load necessary libraries

install.packages("ape")

library(ape)

##to identify terminal branches

#not_rodents

tree <-
read.tree(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/Up
ham_MCC_CEACAM1_NotRodents_2024-10-20.Newick")

terminal_branches <- tree$tip.label

absrel_data <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/non
_rodents_absrel.csv")

absrel_data$Branch_Type <- ifelse(absrel_data$Name %in%
terminal_branches, "Terminal", "Non-terminal")

write.csv(absrel_data,
file="C:/Users/DELL/Desktop/HyPhy_Reproduce/not_rodents_
absrel_with_branch_type.csv")

head(absrel_data)

# rodents_only
```

```
rm(list = ls())

tree <-
read.tree(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/Upham_MCC_CEACAM1_RodentsOnly_2024-10-20.Newick")

terminal_branches <- tree$tip.label

absrel_data <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/rodents_absrel.csv")

absrel_data$Branch_Type <- ifelse(absrel_data$Name %in%
terminal_branches, "Terminal", "Non-terminal")

write.csv(absrel_data,
file="C:/Users/DELL/Desktop/HyPhy_Reproduce/rodents_absrel_with_branch_type.csv")

head(absrel_data)

# whole_data_set

rm(list = ls())

tree <-
read.tree(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/Upham_MCC_renamed_91spp_forCEACAM1_2024-10-20.Newick")

terminal_branches <- tree$tip.label

absrel_data <-
read.csv(file="C:/Users/DELL/Desktop/HyPhy_Reproduce/whole_gene_absrel.csv")

absrel_data$Branch_Type <- ifelse(absrel_data$Name %in%
terminal_branches, "Terminal", "Non-terminal")

write.csv(absrel_data,
file="C:/Users/DELL/Desktop/HyPhy_Reproduce/whole_gene_absrel_with_branch_type.csv")

head(absrel_data)

# Load necessary libraries

library(dplyr)

# Load the datasets

rodents_data <-
read.csv("C:/Users/DELL/Desktop/HyPhy_Reproduce/rodents_absrel_with_branch_type.csv")

non_rodents_data <-
read.csv("C:/Users/DELL/Desktop/HyPhy_Reproduce/not_rodents_absrel_with_branch_type.csv")

whole_gene_data <-
read.csv("C:/Users/DELL/Desktop/HyPhy_Reproduce/whole_gene_absrel_with_branch_type.csv")
```

```
# Define function to categorize selection status
categorize_selection <- function(data) {
  data %>%
    mutate(Selection_Status =
ifelse(`Uncorrected.p.value` < 0.05, "Under Selection",
"Not Under Selection"))
}

# Apply categorization to each dataset
rodents_data <- categorize_selection(rodents_data)
non_rodents_data <-
categorize_selection(non_rodents_data)
whole_gene_data <- categorize_selection(whole_gene_data)

# Combine all results into a single data frame
combined_results <- bind_rows(
  rodents_data %>% mutate(Group = "Rodents"),
  non_rodents_data %>% mutate(Group = "Non-Rodents"),
  whole_gene_data %>% mutate(Group = "Whole Gene
(Terminal/Non-Terminal)")
)

# Write combined results to a new CSV file
write.csv(combined_results,
"C:/Users/DELL/Desktop/HyPhy_Reproduce/absrel_combined_s
election_results.csv", row.names = FALSE)

# Preview the combined results
head(combined_results)

# Load necessary libraries
library(dplyr)
library(readr)

# Read in your dataset
```

```
data <- combined_results

# Load necessary library

library(dplyr)


data <-
read.csv("C:/Users/DELL/Desktop/HyPhy_Reproduce/absrel_c
ombined_selection_results.csv")


# Summary for Rodents

rodents_summary <- data %>%

  filter(Group == "Rodents") %>%  # Use 'Group' as per
your data

  summarise(

    Rodents_Under_Selection = sum(`Selection_Status` ==
"Under Selection"),

    Rodents_Not_Under_Selection = sum(`Selection_Status`
== "Not Under Selection")

  )

# Summary for Non-Rodents

non_rodents_summary <- data %>%

  filter(Group == "Non-Rodents") %>%

  summarise(

    Non_Rodents_Under_Selection = sum(Selection_Status
== "Under Selection"),

    Non_Rodents_Not_Under_Selection =
sum(Selection_Status == "Not Under Selection")

  )


# Summary for Terminal Branches excluding Rodents and
Not Rodents

terminal_summary <- data %>%

  filter(!(Group %in% c("Rodents", "Not Rodents"))) %>%

  filter(Branch_Type == "Terminal") %>%

  summarise(

    Terminal_Under_Selection = sum(Selection_Status ==
"Under Selection"),

    Terminal_Not_Under_Selection = sum(Selection_Status
== "Not Under Selection")
```

```
  )


# Summary for Non-Terminal Branches excluding Rodents
and Not Rodents
non_terminal_summary <- data %>%
  filter(!(Group %in% c("Rodents", "Not Rodents"))) %>%
  filter(Branch_Type == "Non-terminal") %>%
  summarise(
    Non_Terminal_Under_Selection = sum(Selection_Status
== "Under Selection"),
    Non_Terminal_Not_Under_Selection =
sum(Selection_Status == "Not Under Selection")
  )


# Combine all summaries into one data frame
combined_summary <- bind_rows(
  terminal_summary,
  non_terminal_summary,
  rodents_summary,
  non_rodents_summary
)


write.csv(combined_summary,
"C:/Users/DELL/Desktop/HyPhy_Reproduce/S5_absrel_combine
d_selection_results.csv", row.names = FALSE)
## Fisher exact test; under selction vs no_selction
#rodents versus not-rodents
fisher.test(matrix(c(25, 18, 53, 80), nrow = 2, byrow =
TRUE)) # p-value = 0.05139
print(odds_ratio <- (25 * 80) / (18 * 53))  # odds ratio
= 2.096436
# Terminal rodents versus terminal not-rodents
fisher.test(matrix(c(16, 7, 30, 38), nrow = 2, byrow =
TRUE)) # p-value = 0.0527
print(odds_ratio <- (16 * 38) / (7 * 30)) # odds ratio =
2.895238
```

Appendix 6. Variable_sites_and_Completeness.R

```
rm(list = ls())

rm(Sample3)

#Checking for variable sites and completeness ###

AllRes <-
read.csv(file="C:/Users/DELL/Desktop/CEACAM1_SP/CEACAM1_
residues_alignment.csv")

head(AllRes)

names(AllRes)

length(unique(AllRes$X1)[!is.na(unique(AllRes$X1))]) #
gets the number of unique not NA characters

length(unique(AllRes$X82)[!is.na(unique(AllRes$X82))])

length(unique(AllRes$X1))

Test <-unique(AllRes$X315)

Test <-Test[Test!=""]

Test <-Test[!is.na(Test)]

length(Test)

Changing <-matrix(,nrow=ncol(AllRes), ncol=2)

for (i in 1:ncol(AllRes)){
  Changing[i,1] <-colnames(AllRes)[i]
  Test <-unique(AllRes[,i])
  Test <-Test[Test!=""]
  Test <-Test[!is.na(Test)]
  Changing[i,2] <-length(Test)
}

head(Changing)

tail(Changing)

colnames(Changing) <- c("Variable", "Unique_residues")

ChangingDF <-data.frame(Changing)

nrow(ChangingDF[ChangingDF$Unique_residues==1,])

Changing[1:100,]


write.csv(Changing,
file="C:/Users/DELL/Desktop/CEACAM1_SP/CEACAM1_Variant_s
ites_2024-10-20.csv")

tAllRes <-t(AllRes)

tAllResDF <-data.frame(tAllRes)
```

```
Sample <- tAllResDF[, 16]
Sample2 <- as.character(Sample)
Sample2 <- Sample2[!is.na(Sample2) & trimws(Sample2) !=
""]
length(Sample2)
Sample3 <- Sample2[4:length(Sample2)]
#Worked
Sampling <-matrix(,nrow=ncol(tAllRes),ncol=3)
for (i in 1:ncol(tAllRes)){
  Sample <-tAllResDF[,i]
  Sample2 <-Sample; length(Sample)
  Sample2 <-Sample2[!is.na(Sample2)]; length(Sample2)
  Sample2 <-Sample2[Sample2!=""]; length(Sample2)
  Sample3 <-Sample2[4:length(Sample2)]
  Sampling[i,1]<-as.character(Sample2[1])
  Sampling[i,2]<-as.character(Sample2[2])
  Sampling[i,3]<-length(Sample3)
}
#Not Worked
#from Ai
Sampling <- matrix(, nrow = ncol(tAllRes), ncol = 3)
for (i in 1:ncol(tAllRes)) {
  Sample <- tAllResDF[, i]
  Sample2 <- as.character(Sample)
  Sample2 <- Sample2[!is.na(Sample2) & trimws(Sample2)
!= ""]

  if (length(Sample2) >= 2) {
    Sampling[i, 1] <- as.character(Sample2[1])
    Sampling[i, 2] <- as.character(Sample2[2])
  } else {
    Sampling[i, 1] <- NA
    Sampling[i, 2] <- NA
  }
```

```
  Sample3 <- Sample2[4:length(Sample2)]
  Sampling[i, 3] <- length(Sample3)
}
#Worked
head(Sampling)
write.csv(Sampling,
file="C:/Users/DELL/Desktop/CEACAM1_SP/Sample_coverage_2
024-10-20.csv")
Sampling <-
read.csv(file="C:/Users/DELL/Desktop/CEACAM1_SP/Sample_c
overage_2024-10-20.csv")
head(Sampling)
wilcox.test(V3 ~ V2, data = Sampling)
# p < 0.7549 ; W = 1003
boxplot(V3~V2, data=Sampling)
#Shapiro-Wilk Test
shapiro.test(Sampling[,3]) #W = 0.5341, p-value < 2.2e-
16
mean(Sampling[Sampling$V2==0,4]) # Not Rodents
mean==477.5325
# the mean of the values in the 4th column (V4) of the
Sampling dataset, but only for the rows where the value
in the 2nd column (V2) is equal to 0.
mean(Sampling[Sampling$V2==1,4]) #  Rodents mean==501.88
# mean of the values in the 4th column (V4) of the
Sampling dataset, but only for the rows where the value
in the 2nd column (V2) is equal to 1.
sd(Sampling[Sampling$V2==0,4])/sqrt(length(Sampling[Samp
ling$V2==0,4])) # 7.625047
# the standard error of the mean (SEM) for the group
where V2 == 0
sd(Sampling[Sampling$V2==1,4])/sqrt(length(Sampling[Samp
ling$V2==1,4])) #  6.362473
#  standard error of the mean (SEM) for the group where
V2 == 1
#Worked
```

Apendix 7: Figure.R

```
rm(list = ls()) #remove datasets

graphics.off() #remove plots

rm(humancores)

#Figure_1 Similarity scores to human and mice

Score_data <-
read.csv(file="C:/Users/DELL/Desktop/CEACAM1_SP/Scores_f
or_plotting.csv")

head(Score_data)

plot(humanscore~micescore, data=Score_data)

Score_data_original <-Score_data

str(Score_data_original)

Score_data2 <-Score_data[is.na(Score_data$Duplicate),]

Score_data2 <-Score_data

humanscores <-Score_data2[Score_data2$Count>17,c(1,3,4)]

micescores <-Score_data2[Score_data2$Count>17,c(1,3,5)]

humanscores$Similarity <-rep("Similarity to Human");
names(humanscores)<-c("Species", "Order",
"Score","Similarity")

head(humanscores)

micescores$Similarity <-rep("Similarity to Mice");
names(micescores)<-c("Species", "Order",
"Score","Similarity")

head(micescores)

Scores <-rbind(humanscores,micescores)

Scores <-Scores[!is.na(Scores$Score),]


Scores$Order <- ifelse(Scores$Order %in%
c("CETARTIODACTYLA", "RODENTIA", "PRIMATES",
"CARNIVORA", "CHIROPTERA"),
                       Scores$Order,
                       "Other")

library(ggplot2)

library(gridExtra)

library(ggbeeswarm)

plot1 <- ggplot(Scores, aes(x=Order, y=Score,
fill=Similarity, shape=Similarity)) +
  geom_boxplot(outlier.color=NA) +
```

```
  theme_bw() +
  labs(y="Similarity Score",
       x="Order",
       title="Similarity Scores to Human and Mice") +
  scale_fill_manual(values=c("gray85", "gray55")) +

geom_point(position=position_jitterdodge(jitter.width=0.
2, jitter.height=0.005), size=2, alpha=0.5) +
  scale_shape_manual(values=c(21, 24)) +
  theme(axis.text.x = element_text(angle=45, hjust=1))
print(plot1)
#All worked


### Wilcoxon signed-rank test for Rodents compared to
other orders


# for Rodents
wilcox.test(
  Score_data[Score_data$Order == "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 4],
  Score_data[Score_data$Order == "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 5],
  paired = TRUE,  # Assuming scores are paired (the same
species)
  alternative = "two.sided"
) # V = 62.5, p-value = 0.06795


#for not Rodents
wilcox.test(
  Score_data[Score_data$Order != "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 4],
  Score_data[Score_data$Order != "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 5],
  paired = TRUE,  # Assuming scores are paired (the same
species)
  alternative = "two.sided"
) # V = 2141, p-value = 3.766e-11
```

```
###
# For Rodents
wilcox.test(
  Score_data[Score_data$Order == "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 4],
  Score_data[Score_data$Order == "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 5],
  paired = FALSE,  # Not assuming paired samples
  alternative = "two.sided"
) # W = 159, p-value = 0.05175


# For Orders Other Than Rodents
wilcox.test(
  Score_data[Score_data$Order != "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 4],
  Score_data[Score_data$Order != "RODENTIA" &
Score_data$Count > 17 & is.na(Score_data$Duplicate), 5],
  paired = FALSE,  # Not assuming paired samples
  alternative = "two.sided"
) #W = 4066, p-value = 5.988e-13
```

Appendix 8: First and last 15 residues of CEACAM1

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Aotus nancymaae	M	G	H	L	S	A	R	L	H	R	V	C	V	P	W	S	P	T	T	T	D	I	I	Y	S	E	V	K	K	K
Apodemus sylvaticus	M	E	L	A	S	A	H	R	H	R	G	Q	I	P	W	S	P	K	A	T	E	T	L	Y	S	E	V	K	K	K
Artibeus jamaicensis	M	E	S	P	S	A	P	A	H	R	G	R	V	P	W	S	P	V	A	T	E	M	I	Y	S	E	V	K	K	K
Balaenoptera musculus	M	K	P	P	S	A	H	A	R	R	G	C	I	P	W	X	X	X	A	T	E	T	V	Y	S	E	V	K	N	K
Bison bison	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	S	P	T	H	T	E	T	I	Y	S	E	V	K	K	Q
Bos taurus	M	G	T	P	S	G	P	S	R	R	R	C	I	P	W	S	P	T	N	T	E	T	I	Y	S	E	V	K	K	Q
Bubalus bubalis	M	G	T	P	S	G	P	S	R	R	R	C	I	P	W	S	P	T	K	T	E	T	I	Y	S	E	V	K	K	Q
Bubalus kerabau	M	G	T	P	S	G	P	S	R	R	R	C	I	P	W	S	P	T	K	T	E	T	I	Y	S	E	V	K	K	Q
Budorcas taxicolor	M	E	P	P	S	G	P	A	S	R	R	H	V	P	W	S	P	T	D	T	E	T	I	Y	S	E	V	K	K	Q
Callithrix jacchus	M	G	H	L	S	A	R	L	H	R	V	C	V	P	W	S	P	T	T	T	D	I	I	Y	S	E	V	K	K	K
Camelus dromedarius	M	E	P	S	S	A	P	A	H	R	G	H	V	L	W	S	L	T	A	S	E	T	V	Y	S	E	V	K	K	K
Canis lupus familiaris	M	E	P	P	S	A	S	P	R	A	G	R	G	P	W	S	P	A	A	T	E	T	V	Y	S	E	V	Q	K	K
Cavia porcellus	M	E	R	R	T	A	T	S	Y	R	G	F	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cavia porcellus	M	E	P	R	T	T	T	S	H	R	G	F	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cercocebus atys	M	G	H	L	S	T	P	L	H	R	V	H	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Cercopithecus aethiops	X	X	X	X	X	X	X	X	H	R	V	R	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Chinchilla lanigera	M	E	P	R	S	A	T	S	H	G	G	C	A	P	W	F	S	T	V	P	E	T	V	Y	S	E	V	K	K	K

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Chlorocebus sabaeus	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	L	S	P	G	Q	G	Y	S	K	E	E	E	T
Chrysochloris asiatica	M	X	X	X	X	X	X	X	X	G	G	R	X	X	X	S	M	R	E	T	E	P	D	I	P	E	V	L	L	E
Colobus angolensis	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Condylura cristata	M	E	A	P	S	A	H	G	H	R	G	R	V	P	W	X	X	A	A	T	G	T	I	D	X	X	V	C	P	X
Desmodus rotundus	M	Q	S	L	S	A	P	A	H	R	G	P	I	P	W	S	P	V	A	T	E	M	I	Y	S	E	V	K	K	K
Dipodomys spectabilis	M	E	P	R	S	A	S	P	R	R	G	C	V	P	W	T	S	V	A	T	E	T	V	Y	S	E	V	K	K	K
Echinops telfairi	M	E	A	P	S	V	T	P	S	K	G	H	S	S	W	A	P	V	A	T	E	T	I	Y	T	E	V	K	K	K
Equus caballus	M	E	L	P	T	A	P	A	G	K	G	C	V	P	W	P	P	P	A	T	E	T	V	Y	S	E	V	K	K	K
Equus caballus	M	V	P	P	S	A	P	A	H	R	G	P	V	P	R	L	P	S	A	T	E	T	V	Y	S	E	V	K	K	K
Erinaceus europaeus	M	V	P	P	S	A	P	S	H	R	G	S	V	P	W	F	P	I	S	T	E	A	V	Y	S	E	V	K	K	K
Felis catus	M	E	P	P	S	A	L	P	R	G	G	R	V	P	W	P	L	T	A	T	E	T	V	Y	S	E	V	Q	K	K
Fukomys damarensis	M	E	T	R	S	A	T	S	H	R	G	L	L	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gorilla gorilla	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	X
Heterocephalus glaber	M	E	P	G	S	A	T	S	H	R	G	F	V	S	W	S	S	T	I	T	E	T	V	Y	S	E	V	K	K	K
Homo sapiens	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Homo sapiens	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	L	T	A	T	E	I	I	Y	S	E	V	K	K	Q
Hylobates moloch	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Ictidomys tridecemlineatus	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	S	P	T	A	T	E	T	V	Y	S	E	L	K	K	K
Jaculus jaculus	M	E	L	F	S	D	R	A	G	G	R	H	V	P	L	T	P	A	A	S	E	T	I	Y	S	E	V	K	K	K
Lepus europaeus	M	E	P	S	S	A	P	P	C	R	M	C	V	P	R	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Loxodonta africana	M	E	P	P	S	A	S	P	R	R	G	G	V	H	W	S	P	T	A	T	E	T	I	Y	S	E	V	K	K	K
Lynx canadensis	M	E	P	P	S	A	L	S	R	G	G	R	V	P	W	P	L	T	A	T	E	T	V	Y	S	E	V	Q	K	K
Macaca fascicularis	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Macaca mulatta	M	G	Y	L	S	T	P	L	H	R	V	R	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Macaca nemestrina	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Manis javanica	L	A	P	P	S	A	P	A	P	R	G	P	V	P	W	S	P	K	A	T	E	T	V	Y	S	E	V	K	R	S
Manis pentadactyla	L	A	P	P	S	A	P	A	P	R	G	L	V	P	W	S	P	K	A	K	E	T	V	Y	S	E	V	K	R	S
Marmota marmota	M	R	P	P	S	A	C	P	C	R	G	G	I	P	W	S	P	T	A	T	E	T	V	Y	S	E	L	K	K	K
Mesocricetus auratus	M	E	L	T	S	A	P	L	H	K	G	Q	V	P	W	S	P	R	A	I	E	A	V	Y	S	E	V	K	K	K
Microtus oregoni	M	E	L	S	S	T	L	L	H	K	G	Q	L	P	W	S	P	T	P	T	E	T	V	Y	S	E	V	K	R	K
Mirounga angustirostris	M	E	P	P	S	A	P	P	R	G	G	R	V	P	W	S	P	A	D	T	E	T	V	Y	T	E	V	Q	K	K
Moschus berezovskii	M	E	A	P	S	G	P	A	S	R	R	H	V	A	W	S	P	T	D	T	E	T	I	Y	S	E	V	K	K	P
Mus musculus	M	E	L	A	S	A	H	L	H	K	G	Q	V	P	W	S	P	R	A	T	E	T	V	Y	S	E	V	K	K	K
Mustela putorius furo	M	K	P	T	S	A	P	P	Q	G	G	R	V	P	W	S	P	A	D	T	E	T	V	Y	T	E	V	Q	K	K

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Myodes glareolus	M	E	L	S	S	A	P	L	H	K	G	Q	L	P	W	S	P	R	S	T	E	T	V	Y	S	E	V	K	K	Q
Myotis davidii	M	E	S	P	S	A	P	A	H	R	G	C	V	P	W	S	P	T	A	T	E	T	L	Y	S	E	V	K	K	K
Myotis myotis	M	E	S	P	S	A	P	A	H	R	G	C	V	S	W	S	P	T	T	T	E	T	L	Y	S	E	V	K	K	K
Nannospalax galili	T	E	P	L	S	A	P	P	R	T	G	S	I	C	W	A	X	X	X	X	E	T	V	Y	S	Q	V	K	K	N
Nomascus leucogenys	M	G	H	L	S	A	P	L	H	R	V	R	V	L	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Nyctereutes procyonoides	M	E	P	P	S	A	S	P	R	A	G	R	G	P	W	S	P	A	A	T	E	T	V	Y	S	E	V	Q	K	K
Octodon degus	M	E	P	R	S	A	T	S	H	R	G	L	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Odocoileus virginianus	M	E	P	P	S	G	P	S	R	R	R	Y	V	P	W	S	P	T	H	T	E	T	I	Y	S	E	V	K	N	Q
Onychomys torridus	M	E	L	T	S	A	P	L	H	K	G	Q	V	S	W	S	P	R	A	T	E	E	V	Y	S	E	V	K	K	R
Orcinus orca	M	K	P	P	L	A	H	A	H	R	G	C	S	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ornithorhynchus anatinus	M	A	X	S	P	A	P	P	G	R	A	R	S	S	W	C	P	R	D	X	X	T	P	T	T	R	X	N	H	Q
Oryctolagus cuniculus	M	E	P	S	S	A	P	P	R	R	T	R	V	P	R	S	P	R	A	T	E	T	V	Y	S	E	V	K	K	K
Ovis aries	M	E	S	P	S	G	P	A	S	R	R	H	V	P	W	T	P	T	D	T	E	T	I	Y	S	E	V	K	K	Q
Pan paniscus	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	L	T	A	T	E	I	I	Y	S	E	V	K	K	X
Pan troglodytes	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	X
Panthera pardus	M	E	P	P	S	A	L	P	R	G	G	R	V	P	W	S	L	T	A	T	E	T	V	Y	S	E	V	Q	K	K
Panthera tigris	M	E	P	P	S	A	L	P	R	G	G	R	V	P	W	S	L	T	A	T	E	T	V	Y	S	E	V	Q	K	K
Panthera uncia	M	E	P	P	A	A	L	P	R	G	G	R	V	P	W	S	L	T	A	T	E	T	V	Y	S	E	V	Q	K	K
Papio anubis	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Peromyscus eremicus	M	E	L	T	S	A	P	F	N	K	G	Q	V	P	W	S	P	R	A	T	E	E	V	Y	S	E	V	K	K	K
Peromyscus leucopus	M	E	L	T	S	A	P	L	H	K	E	Q	X	X	W	S	P	R	A	T	E	E	V	Y	S	E	V	K	K	K
Peromyscus maniculatus	M	E	L	T	S	A	P	L	H	N	A	Q	V	S	W	S	P	R	V	T	E	E	V	Y	S	E	V	K	K	K
Phodopus roborovskii	M	E	L	T	S	A	P	L	H	K	G	Q	L	P	W	S	P	R	A	T	E	A	V	Y	S	E	V	K	K	K
Phyllostomus discolor	M	E	S	L	S	A	P	A	H	R	G	R	V	P	L	S	P	V	A	T	E	M	I	Y	S	E	V	K	K	K
Physeter catodon	M	E	P	R	L	A	P	A	C	R	G	C	I	P	W	X	X	X	A	T	E	T	V	Y	S	E	V	K	N	K
Piliocolobus tephrosceles	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Pongo abelii	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Pongo pygmaeus	M	G	H	L	S	A	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Propithecus coquereli	M	E	P	P	S	A	P	P	R	R	G	R	V	P	W	S	S	T	A	T	E	T	V	Y	S	E	V	K	K	K
Psammomys obesus	M	E	R	A	S	A	S	P	R	E	G	R	V	R	W	S	P	S	A	T	E	I	I	Y	S	E	V	K	K	K
Pteronotus mesoamericanus	M	G	S	P	S	V	P	A	R	R	G	R	A	P	W	S	S	T	A	T	E	M	I	Y	S	E	V	K	K	K
Rattus norvegicus	M	E	L	A	S	A	R	L	L	R	G	Q	I	P	W	X	S	S	P	T	E	T	V	Y	S	V	V	K	K	K
Rattus rattus	M	E	L	A	S	A	R	L	L	R	G	Q	I	P	W	X	S	S	P	T	E	T	V	Y	S	E	V	K	K	K
Rhinopithecus bieti	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Rhinopithecus roxellana	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K

Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**514**	**515**	**516**	**517**	**518**	**519**	**520**	**521**	**522**	**523**	**524**	**525**	**526**	**527**	**528**
Rousettus aegyptiacus	M	E	L	P	S	R	P	N	H	R	E	R	T	P	W	S	P	T	A	T	E	T	V	Y	S	E	V	K	K	K
Saccopteryx bilineata	M	E	S	P	S	A	P	T	S	R	G	R	G	P	W	S	S	T	A	T	E	T	V	Y	L	E	V	K	I	K
Saccopteryx leptura	M	E	S	P	S	A	P	T	S	R	G	R	G	P	W	S	S	T	A	T	E	T	V	Y	L	E	V	K	I	K
Saimiri boliviensis	M	G	H	L	S	A	Q	L	H	R	V	C	V	P	W	S	P	T	T	T	D	V	I	Y	S	E	V	K	M	K
Sorex araneus	M	E	P	P	S	A	P	A	H	I	G	C	V	P	W	P	S	T	A	T	Q	T	I	Y	A	E	V	K	K	T
Sorex fumeus	M	E	S	P	S	A	P	A	H	R	G	Y	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sus scrofa	M	E	P	P	S	A	P	A	H	R	G	H	V	P	W	S	P	S	A	T	E	T	V	Y	S	E	V	K	K	K
Talpa occidentalis	M	E	A	P	S	V	L	G	H	R	S	R	V	P	W	S	P	A	A	T	E	T	V	Y	S	E	V	K	R	K
Theropithecus gelada	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	P	T	A	T	E	I	I	Y	S	E	V	K	K	K
Trachypithecus francoisi	M	G	H	L	S	T	P	L	H	R	V	R	V	P	W	S	A	T	A	T	E	I	I	Y	S	E	V	K	K	K
Urocitellus parryii	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	S	P	T	A	T	E	T	V	Y	S	E	L	K	K	K
Ursus arctos	M	E	P	P	S	A	P	P	R	G	G	H	V	P	W	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ursus maritimus	M	E	P	P	S	A	P	P	R	G	G	H	V	P	W	S	P	A	A	T	E	T	V	Y	T	E	V	Q	R	K
Ursus maritimus	M	E	P	P	S	A	P	P	R	G	G	R	A	P	W	S	P	S	G	L	X	X	X	X	X	X	X	X	X	X
Vicugna pacos	M	E	P	S	S	A	P	A	H	R	G	H	V	L	W	S	L	T	A	S	E	T	V	Y	S	E	V	K	K	K
Vombatus ursinus	M	E	S	P	S	Q	A	P	H	S	G	S	S	L	W	A	P	S	H	T	D	T	V	Y	S	E	I	K	K	K

Appendix - 9: All contact residues with common residues at each position

Accession	*Specie Name*	63	64	68	73	75	76	81	83	85	86	88	90	91	92	93	127	128	130
XM_012461510.3	*Aotus nancymaae*	T	G	Y	V	S	T	G	V	T	T	Q	I	P	G	P	A	D	V
XM_052168799.1	*Apodemus sylvaticus*	Q	V	Y	T	V	N	R	V	Q	T	R	T	M	G	P	P	D	N
XM_053657140.1	*Artibeus jamaicensis*	V	G	Y	T	N	N	S	V	Q	T	K	T	P	G	S	E	S	Q
XM_036833342.1	*Balaenoptera musculus*	L	G	Y	V	K	N	S	R	D	V	T	V	T	G	P	N	D	Q
XM_010842547.1	*Bison bison*	L	G	Y	V	N	S	S	R	A	T	L	T	K	G	P	D	D	Q
XM_024977969.2	*Bos taurus*	L	G	Y	V	N	T	S	R	D	T	A	T	K	G	P	D	D	Q
XM_006041224.4	*Bubalus bubalis*	L	G	H	V	N	T	S	R	A	T	L	T	K	G	P	D	D	Q
XM_055551361.1	*Bubalus kerabau*	L	G	H	V	N	T	S	R	A	T	L	T	K	G	P	D	D	Q
XM_052655777.1	*Budorcas taxicolor*	L	G	H	V	N	A	S	R	G	N	A	T	N	G	P	E	D	Q
XM_002762196.3	*Callithrix jacchus*	N	G	Y	V	S	T	G	V	A	S	L	T	P	G	P	A	D	V
XM_064489351.1	*Camelus dromedarius*	Q	G	F	V	N	S	S	I	S	S	A	V	P	G	P	D	K	I
NM_001097557.1	*Canis lupus familiaris*	A	S	F	V	P	T	L	V	D	T	I	T	P	G	P	Q	K	E
XM_063252825.1	*Cavia porcellus*	L	G	F	T	S	D	R	V	N	N	I	V	T	G	K	E	T	D
XM_063252826.1	*Cavia porcellus*	V	G	F	T	P	N	R	V	T	N	V	V	P	G	K	E	T	D
XM_012086961.1	*Cercocebus atys*	I	G	H	V	A	K	A	V	G	T	Q	T	P	G	P	E	D	V
AF259566.1	*Cercopithecus aethiops*	L	G	H	V	A	N	A	V	G	T	Q	T	P	G	P	Q	D	V
XM_005412355.1	*Chinchilla lanigera*	I	G	F	S	S	S	A	V	S	T	V	T	P	G	P	D	N	W
XM_007997002.2	*Chlorocebus sabaeus*	L	G	H	V	A	K	A	V	G	T	Q	T	P	G	P	Q	D	V
XM_006871507.1	*Chrysochloris asiatica*	Y	G	Y	A	L	S	A	L	E	N	S	A	L	G	V	N	D	Q
XM_011936539.1	*Colobus angolensis*	L	G	Y	V	N	K	A	V	R	T	Q	T	P	G	P	E	D	V
XM_012733660.1	*Condylura cristata*	G	I	Y	V	G	E	S	K	G	S	A	G	T	G	P	R	D	E
XM_024577349.3	*Desmodus rotundus*	G	R	Y	V	S	K	S	M	E	T	A	V	P	G	P	K	D	S
XM_042668120.1	*Dipodomys spectabilis*	E	L	Y	T	E	S	A	S	N	S	Q	T	P	G	P	R	D	A

Accession	*Specie Name*	63	64	68	73	75	76	81	83	85	86	88	90	91	92	93	127	128	130
XM_045292270.1	*Echinops telfairi*	A	F	Y	V	K	N	T	A	S	S	I	T	R	G	P	A	D	V
MF564057.1	*Equus caballus*	A	A	Y	V	H	H	S	K	G	P	E	T	P	G	P	T	S	Q
XM_005596278.3	*Equus caballus*	A	A	Y	V	P	H	S	V	N	T	Q	T	P	G	P	A	N	Q
XM_007532949.3	*Erinaceus europaeus*	I	G	F	V	S	A	G	E	A	T	Q	T	T	G	P	S	G	Q
XM_011289901.4	*Felis catus*	I	G	F	I	P	R	S	A	D	S	E	T	L	G	F	R	N	Q
XM_010623328.3	*Fukomys damarensis*	I	A	F	V	S	N	A	L	A	D	T	T	Q	G	P	P	N	L
XM_031004080.3	*Gorilla gorilla*	F	G	Y	V	G	N	G	V	A	T	Q	T	P	G	P	S	D	V
XM_004873005.2	*Heterocephalus glaber*	R	A	F	T	E	D	T	M	D	E	T	L	E	G	P	R	N	L
NM_001024912.3	*Homo sapiens*	F	G	Y	V	G	N	G	A	G	T	Q	T	P	G	P	S	D	V
NM_001712.5	*Homo sapiens*	F	G	Y	V	G	N	G	A	G	T	Q	T	P	G	P	S	D	V
XM_032172816.2	*Hylobates moloch*	I	G	Y	V	A	N	A	V	E	N	Q	S	L	G	P	L	D	V
XM_005338246.4	*Ictidomys tridecemlineatus*	I	G	F	A	N	R	S	M	S	T	L	T	P	G	S	A	N	R
XM_004670324.2	*Jaculus jaculus*	Y	G	Y	V	E	S	R	I	S	T	I	V	T	G	P	E	V	Q
XM_062175861.1	*Lepus europaeus*	P	V	Y	A	P	N	T	V	S	L	N	A	P	G	P	I	L	V
XM_064293748.1	*Loxodonta africana*	F	V	Y	T	N	N	S	A	Q	T	T	V	L	G	P			W
XM_030297609.1	*Lynx canadensis*	I	G	F	I	P	H	S	A	D	L	E	T	P	G	F	R	N	Q
XM_005589370.4	*Macaca fascicularis*	I	G	Y	V	A	K	A	V	E	S	Q	T	P	G	P	G	D	V
MG874670.1	*Macaca mulatta*	I	G	H	V	A	K	A	V	R	T	Q	T	P	G	P	G	D	V
XM_011764531.2	*Macaca nemestrina*	F	G	Y	V	A	K	A	V	G	T	Q	T	P	G	P	G	D	V
XM_037013906.1	*Manis javanica*	L	G	Y	V	F	N	S	E	N	K	A	T	P	G	I	G	S	L
XM_036896477.2	*Manis pentadactyla*	L	G	Y	V	F	N	S	V	N	T	A	T	P	G	I	G	S	L
XM_015496801.2	*Marmota marmota*	V	G	Y	T	Q	N	S	S	L	T	S	I	Q	G	P	P	N	Q
XM_040734318.1	*Mesocricetus auratus*	Q	V	Y	L	D	S	R	I	S	S	T	S	T	G	P	A	N	D
XM_041654886.1	*Microtus oregoni*	Q	V	H	T	N	S	R	V	S	D	V	K	T	G	P	Q	G	D
XM_064581140.1	*Mirounga angustirostris*	I	G	F	V	P	R	L	A	D	T	V	T	P	G	P	R	D	Q

Accession	*Specie Name*	63	64	68	73	75	76	81	83	85	86	88	90	91	92	93	127	128	130
XM_055412081.1	*Moschus berezovskii*	L	G	H	V	N	S	S	R	D	T	V	T	R	G	P	V	D	Q
NM_001039185.1	*Mus musculus*	G	A	Y	T	I	D	R	V	N	S	M	F	T	G	Q	E	N	R
XM_004780425.3	*Mustela putorius*	G	R	Y	T	E	N	S	I	D	T	K	V	I	G	P	R	T	E
XM_048440439.1	*Myodes glareolus*	Q	V	Y	Q			R	I	S	P	E	T	T	G	P	E	N	D
XM_015562203.1	*Myotis davidii*	R	G	Y	I	S	E	S	V	D	T	T	T	L	G	P	E	N	Q
XM_036346012.1	*Myotis myotis*	L	L	Y	I	S	N	S	V	D	T	A	N	F	G	P	K	S	K
XM_029570052.1	*Nannospalax galili*	N	I	Y	T	D	D	R	L	S	T	E	V	E	G	L	S	R	N
XM_030797105.1	*Nomascus leucogenys*	I	G	Y	V	G	N	A	A	G	T	Q	T	P	G	P	S	D	V
XM_055311974.1	*Nyctereutes procyonoides*	A	S	F	V	P	A	L	V	E	T	I	T	P	G	P	R	K	E
XM_023704520.1	*Octodon degus*	I	A	F	V	S	S	T	E	D	T	R	T	P	G	R	S	N	Q
XM_020871207.1	*Odocoileus virginianus*	L	G	Y	V	N	S	S	R	D	T	V	T	T	G	S	N	D	Q
XM_036168114.1	*Onychomys torridus*	Q	A	Y	A			R	V	S	T	T	K	T	G	P	E	N	D
XM_049703839.1	*Orcinus orca*	L	G	Y	V	N	N	S	R	D	V	T	T	T	G	L	N	D	Q
XM_039912179.1	*Ornithorhynchus anatinus*	G	V	Y	V			K	L	T	P	N	T	L	G	K	K	N	T
XM_008249692.3	*Oryctolagus cuniculus*	P	A	Y	A	Q	N	T	T	N	P	Q	T	P	G	P		N	Y
XM_004015280.6	*Ovis aries*	L	G	H	V	N	A	S	R	D	N	A	T	N	G	S	N	D	Q
XM_008964144.4	*Pan paniscus*	I	G	Y	V	G	N	A	A	Q	N	Q	T	R	G	P	S	D	V
XM_063801767.1	*Pan troglodytes*	F	G	Y	V	G	N	G	V	G	T	Q	T	P	G	P	S	D	V
XM_019434422.2	*Panthera pardus*	L	G	F	I	P	R	S	A	D	S	E	T	P	G	F	R	N	Q
XM_007086642.2	*Panthera tigris*	L	G	F	I	P	R	S	A	D	S	E	T	P	G	L	R	N	Q
XM_049621270.1	*Panthera uncia*	L	G	F	I	P	H	S	A	D	S	E	T	P	G	F	R	N	Q
XM_003915629.4	*Papio anubis*	I	G	Y	V	A	K	A	V	G	T	Q	T	P	G	P	E	D	V
XM_059253485.1	*Peromyscus eremicus*	Q	A	Y	V	D	S	R	L	S	I	T	K	T	G	P	E	N	D
XM_028860001.2	*Peromyscus leucopus*	R	A	Y	V	D	S	R	L	S	S	T	K	T	G	P	E	N	D

Accession	*Specie Name*	63	64	68	73	75	76	81	83	85	86	88	90	91	92	93	127	128	130
XM_042267688.1	*Peromyscus maniculatus*	R	A	Y	V	D	S	R	L	S	S	T	K	T	G	P	E	N	D
XM_051207508.1	*Phodopus roborovskii*	Q	V	Y	G	E	S	R	I	S	T	M	T	T	G	P	I	N	D
XM_028530369.2	*Phyllostomus discolor*	L	T	F	A	S	S	A	V	Q	T	K	T	P	G	P	K	D	L
XM_024132797.3	*Physeter catodon*	L	G	Y	V	S	N	S	R	D	G	A	A	T	G	P	N	D	Q
XM_023229215.2	*Piliocolobus tephrosceles*	L	G	Y	V	A	K	A	V	E	T	R	T	P	G	P	E	D	V
NM_001132423.1	*Pongo abelii*	L	G	Y	V	A	N	G	V	S	D	L	T	P	G	P	S	D	V
XM_054463351.2	*Pongo pygmaeus*	L	G	Y	V	A	N	G	V	S	D	L	T	P	G	P	S	D	V
XM_012653468.1	*Propithecus coquereli*	L	G	Y	V	S	N	A	V	S	T	V	N	T	G	P	A	D	Q
XM_055620670.1	*Psammomys obesus*	Q	V	Y	P			H	I	S	S	Q	V	D	G	A	A	D	S
XM_054567634.1	*Pteronotus mesoamericanus*	G	K	F	V	T	S	S	R	D	N	E	T	R	G	P	K	H	S
NM_001033860.1	*Rattus norvegicus*	Q	V	Y	L	P	D	R	I	S	D	M	K	T	G	P	Q	Q	N
XM_032894175.1	*Rattus rattus*	Q	V	Y	T	L	N	R	I	S	D	T	K	T	G	P	Q	D	N
XM_017849842.1	*Rhinopithecus bieti*	L	G	Y	V	A	K	A	V	R	T	Q	T	P	G	P	E	D	V
XM_010358723.2	*Rhinopithecus roxellana*	L	G	Y	V	A	K	A	V	R	T	Q	T	P	G	P	E	D	V
XM_016129903.2	*Rousettus aegyptiacus*	V	G	Y	V	S	N	S	V	E	N	T	V	P	G	P	K	N	L
XM_066271982.1	*Saccopteryx bilineata*	I	G	Y	V	S	N	S	A	A	T	E	T	P	G	P	Q	N	L
XM_066373583.1	*Saccopteryx leptura*	A	G	Y	V	S	N	S	V	A	T	E	T	P	G	P	K	N	L
XM_003943210.3	*Saimiri boliviensis*	T	G	H	V	S	T	G	V	A	T	L	T	R	G	P	A	N	V
XM_055145459.1	*Sorex araneus*	V	R	C	I	S	K	T	K	T	P	N	T	N	G	P	M	G	D
XM_056135625.1	*Sorex fumeus*	L	R	Y	E		K	T	K	N	P	A	V	N	G	P	L	G	D
XM_005655889.2	*Sus scrofa*	L	G	Y	V	N	N	T	R	D	T	A	T	Q	G	P	N	D	Q
XM_037516693.2	*Talpa occidentalis*	L	K	Y	V	K	S	S	R	D	T	N	I	P	G	P	N	N	L
XM_025366225.1	*Theropithecus gelada*	I	G	H	V	A	K	A	V	G	T	Q	T	P	G	P	G	D	V
XM_033197411.1	*Trachypithecus francoisi*	F	G	Y	V	A	K	A	V	R	T	Q	T	P	G	P	E	D	V
XM_026403696.1	*Urocitellus parryii*	I	G	F	T	D	S	S	S	L	T	N	N	T	G	P	N	D	R

Accession	*Specie Name*	63	64	68	73	75	76	81	83	85	86	88	90	91	92	93	127	128	130
XM_026482217.4	*Ursus arctos*	L	G	F	V	A	S	S	V	D	P	E	T	P	G	P	K	N	Q
XM_008709687.2	*Ursus maritimus*	L	G	F	V	T	S	S	V	D	T	V	T	P	G	P	K	N	Q
XM_040634031.1	*Ursus maritimus*	L	G	F	S	P	S	S	T		T	G	S	S	G	P	T	N	Q
XM_031681826.1	*Vicugna pacos*	L	G	F	V	N	S	S	V	S	S	L	I	P	G	P	D	N	S
XM_027858153.1	*Vombatus ursinus*	P	S	Y	I	S	N	Y	V	N	G	L	A	S			T	S	E
		63 - T, Q, V, L, N, A, I, Y, G, E, F, R, P	64 - G, V, S, I, R, L, F, A, T, K	68 - Y, H, F, C	73 - V, T, S, A, I, L, Q, G, P, E	75 - S, V, N, K, P, A, L, G, E, H, F, Q, D, I, T	76 - T, N, S, A, D, K, E, H, R	81 - G, R, S, L, A, T, K, H, Y	83 - V, R, I, L, K, M, S, A, E, T	85 - T, Q, D, A, G, S, N, E, R, L	86 - T, V, N, S, P, D, E, L, K, I, G	88 - Q, R, K, T, L, A, I, V, S, E, N, M, G	90 - I, T, V, A, G, L, S, K, F, N	91 - P, M, T, K, N, L, R, Q, E, I, F, D, S	92 - G	93 - P, S, K, V, F, I, Q, L, R, A	127 - A, P, E, N, D, Q, R, K, T, S, L, I, G, V, M,	128 - D, S, K, T, N, G, V, L, R, H, Q	130 - V, N, Q, I, E, D, W, S, A, L, R, K, T, Y

Appedix-10 : MEME analysis results of all species

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
1	1	0	0	0	1	0	0	0	0	1	0
2	1.00E-08	5.842047	1	2.261752	0.158123	7	0	-67.0825	-66.3713	0.491054	0.718356
3	1.00E-08	3.244059	1	0.256844	0.506166	0	0	-109.385	-109.385	1	2.531954
4	0.684482	2.594024	0.315518	0	0.666667	0	0	-103.03	-103.03	0.999979	2.711564
5	0.62804	0.800421	0.37196	0	0.666667	0	0	-54.8678	-54.868	1	2.151592
6	0.55518	11.32886	0.44482	3.283287	0.092006	7	0	-81.1027	-78.0261	0.046114	2.596992
7	1.00E-08	3.493853	1	6.641076	0.016239	0	0	-98.3839	-98.3845	1	0.785258
8	0.48741	12.07612	0.51259	0.592186	0.405169	1	0	-146.406	-144.986	0.241547	5.696186
9	1.00E-08	3.790206	1	0.923065	0.332118	1	0	-118.964	-118.964	1	2.254186
10	0.962279	2.028766	0.037721	0	0.666667	0	0	-86.908	-86.9084	1	2.58796
11	0.502264	4.967046	0.497736	0.199243	0.528499	1	0	-100.579	-100.016	0.569511	3.317055
12	0.835545	6.060029	0.164455	0	0.666667	0	0	-175.679	-175.679	1	9.631627
13	0.25339	8.371486	0.74661	1.552177	0.232585	7	0	-126.966	-125.763	0.300205	3.65529

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
14	1.00E-08	1.981849	1	0.288447	0.494787	0	0	-79.4463	-79.4465	1	1.448417
15	0.856289	12.5286	0.143711	0.959302	0.325175	4	0	-33.7014	-32.9489	0.471209	11.7545
16	0.678721	138.1973	0.321279	2.663911	0.12758	0	0	-110.263	-108.837	0.240222	4.937523
17	0.944359	1.052953	0.055641	0	0.666667	0	0	-63.3441	-63.3445	1	1.73811
18	0.909122	1.86318	0.090878	0	0.666667	0	0	-98.0173	-98.0177	1	3.595483
19	0.735507	0.834488	0.264493	0.008328	0.643329	1	0	-29.7433	-29.7332	0.989942	0.622564
20	1	0.365439	0	0	0.666667	0	0	-53.4318	-53.4326	1	2.92872
21	1.00E-08	1.619083	1	0.014432	0.635344	0	0	-76.6464	-76.6469	1	1.530883
22	1.00E-08	1.539029	1	0.018944	0.630368	0	0	-69.8318	-69.8319	1	1.413209
23	0.726247	1.970235	0.273753	0	0.666667	0	0	-72.4073	-72.4083	1	2.692836
24	0.626584	0.498004	0.373416	0	0.666667	0	0	-66.896	-66.8967	1	3.912237
25	0.815369	0.870083	0.184631	0	0.666667	0	0	-72.0663	-72.0668	1	3.963702

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
26	1.00E-08	1.940488	1	0.033854	0.616718	1	0	-76.665	-76.6657	1	1.703639
27	1.00E-08	3.200371	1	0.797526	0.357638	0	0	-72.2421	-72.1376	0.900767	1.135367
28	0.956731	7.520757	0.043269	0.448392	0.443955	4	0	-37.2696	-34.78	0.082938	0
29	1.00E-08	3.462947	1	2.804207	0.118434	1	0	-111.545	-111.545	1	1.090887
30	0.879815	3.704182	0.120185	0	0.666667	0	0	-128.974	-128.976	1	4.536372
31	0.935804	1.594123	0.064196	0	0.666667	0	0	-100.27	-100.27	1	3.688716
32	1	4.2161	0	-0.00406	0.666667	0	0	-110.832	-110.834	1	4.077825
33	1.00E-08	2.758159	1	0.422815	0.451457	1	0	-99.153	-99.1525	0.999533	1.987132
34	0.770751	0.586793	0.229249	0	0.666667	0	0	-52.2784	-52.2786	1	2.819046
35	0.646591	4.115765	0.353409	0	0.666667	0	0	-98.4136	-98.4089	0.995294	9.584636
36	0.8581	2.219892	0.1419	0	0.666667	0	0	-103.427	-103.428	1	2.860855
37	0.586992	1.161104	0.413008	0	0.666667	0	0	-86.2804	-86.2807	1	3.322047
38	1.00E-08	3.35325	1	0.279838	0.497833	0	0	-126.377	-126.377	0.999804	2.611921

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
39	0.755579	3.390867	0.244421	0	0.666667	0	0	-62.3213	-61.8475	0.622633	5.236224
40	0.73866	38.01696	0.26134	0.829136	0.350986	0	0	-131.149	-130.436	0.490194	7.538942
41	0.456276	9.333341	0.543724	0.870678	0.34248	0	0	-105.265	-104.822	0.642005	2.653729
42	0.949111	3.722068	0.050889	0	0.666667	0	0	-79.232	-79.2618	1	9.416334
43	1.00E-08	15.87003	1	6.866204	0.014474	2	0	-140.991	-138.036	0.052081	3.059998
44	0.413547	22.78199	0.586453	4.786769	0.042088	1	0	-131.618	-130.072	0.212936	2.246657
45	0.696176	2.135951	0.303824	0	0.666667	0	0	-122.362	-122.363	1	5.108847
46	0.963312	2.237153	0.036688	0	0.666667	0	0	-101.526	-101.527	1	5.027827
47	0.885918	1.126969	0.114082	0	0.666667	0	0	-66.579	-66.5815	1	5.498677
48	0.842782	23.24118	0.157218	2.04376	0.177819	4	0	-73.9306	-72.0032	0.145532	4.037736
49	1.00E-08	5.514651	1	0.1205	0.563775	1	0	-153.934	-153.934	1	4.694423
50	1	2.081049	0	0	0.666667	0	0	-96.1191	-96.1193	1	8.163897

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
51	0.768381	2.147585	0.231619	0	0.666667	0	0	-66.369	-66.3609	0.991844	2.845095
52	0.011522	6.517136	0.988478	2.665841	0.127449	2	0	-59.4685	-58.2303	0.289902	1.097194
53	1	0.669631	0	0	0.666667	0	0	-70.1062	-70.1063	1	3.57566
54	0.503608	50.38967	0.496392	1.957861	0.186284	2	0	-146.633	-144.706	0.145558	6.990579
55	1.00E-08	3.13061	1	1.302283	0.267337	1	0	-106.503	-106.503	1	1.779156
56	1.00E-08	2.579955	1	2.527897	0.137151	3	0	-91.5846	-91.5854	1	0.633288
57	0.835529	4.536817	0.164471	0.039086	0.612554	0	0	-74.7332	-74.6935	0.961162	2.800282
58	1.00E-08	5.080859	1	0.719316	0.374813	7	0	-129.738	-129.709	0.971442	3.224885
59	0.181494	9.630931	0.818506	1.229593	0.278521	1	0	-142.02	-141.179	0.431393	3.302295
60	0.581825	26.14284	0.418175	2.371168	0.14912	0	0	-159.223	-157.926	0.273291	5.853872
61	0.729281	66.35838	0.270719	7.681278	0.00955	1	0	-169.386	-165.589	0.022446	5.17816
62	1.00E-08	12.33055	1	11.70544	0.00124	3	0	-208.86	-208.788	0.93017	3.325168

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
63	1.00E-08	40.38672	1	18.12431	4.87E-05	0	0	-233.796	-231.372	0.088512	3.438181
64	0.604221	18.24602	0.395779	10.24702	0.002594	1	0	-125.547	-123.995	0.211885	1.051067
65	0.669875	5.086798	0.330125	0	0.666667	0	0	-104.974	-104.881	0.911037	7.259403
66	0.042648	23.20852	0.957352	20.14746	1.76E-05	2	0	-183.501	-181.255	0.105745	0.895854
67	1	0	0	0	1	0	0	0	0	1	0
68	0.260085	12.06368	0.739915	2.868357	0.114482	7	0	-100.991	-99.3607	0.195917	2.350591
69	0.851556	1.276415	0.148444	0	0.666667	0	0	-65.1622	-65.1645	1	3.856383
70	0.741657	5.018808	0.258343	0.32237	0.483139	5	0	-73.2806	-71.9755	0.271148	3.010887
71	1.00E-08	7.800908	1	7.200042	0.012206	8	0	-134.174	-134.002	0.841894	0.244046
72	0.320943	19.3773	0.679057	9.750709	0.003336	8	0	-184.361	-178.928	0.004371	3.278072
73	1.00E-08	34.0663	1	10.09364	0.002804	2	0	-172.894	-168.592	0.013543	4.276694
74	0.409737	150.0987	0.590263	2.954409	0.109394	1	0	-184.49	-182.154	0.096666	8.586281
75	0.494728	161.6065	0.505272	25.0219	1.52E-06	3	0	-218.733	-207.124	9.09E-06	6.123408

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
76	0.292594	15.50883	0.707406	0.953553	0.326265	0	0	-180.416	-179.293	0.325447	7.643933
77	0.891916	91.60214	0.108084	11.12696	0.001662	3	0	-184.142	-180.177	0.018987	3.418006
78	1.00E-08	22.94017	1	6.604488	0.016545	6	0	-192.154	-190.168	0.137305	4.783197
79	0.830788	3.81914	0.169212	0.498708	0.429765	2	0	-39.9906	-38.5016	0.225587	1.573354
80	1.00E-08	5.998754	1	6.975946	0.013685	1	0	-140.834	-140.834	1	1.935395
81	1.00E-08	5.938482	1	0.577559	0.40888	1	0	-152.634	-152.635	1	4.377967
82	0.406149	6.235591	0.593851	0.180431	0.536337	2	0	-96.3306	-95.757	0.563472	3.949102
83	0.238882	17.06151	0.761118	9.739726	0.003355	0	0	-166.933	-165.959	0.377321	1.9799
84	1.00E-08	10.9754	1	12.00634	0.001065	7	0	-165.582	-165.091	0.611889	0.978983
85	0.421873	67.33558	0.578127	7.973504	0.008229	1	0	-231.109	-227.344	0.023172	8.780736
86	0.789894	78.97243	0.210106	13.23922	0.000571	1	0	-216.496	-211.601	0.007487	4.372079
87	0.424819	101.6692	0.575181	8.158004	0.007492	4	0	-182.553	-176.245	0.001821	7.054037

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
88	1.00E-08	36.4171	1	34.0598	1.63E-08	4	0	-214.315	-213.085	0.292199	0.397861
89	1.00E-08	7.024181	1	1.056583	0.30739	0	0	-177.53	-177.53	1	4.806286
90	1.00E-08	20.70679	1	7.708434	0.009419	5	0	-179.791	-176.494	0.037016	5.768124
91	0.232511	26.98419	0.767489	2.001942	0.181888	0	0	-188.403	-186.928	0.228879	7.444689
92	1	0.877662	0	0	0.666667	0	0	-46.7165	-46.7165	1	4.149045
93	0.403852	26.27087	0.596148	8.929064	0.005062	13	0	-127.015	-121.196	0.002971	3.502921
94	0.593724	20.51374	0.406276	7.578101	0.010065	9	0	-96.6528	-90.4124	0.001949	2.934066
95	1.00E-08	4.770653	1	3.512349	0.081599	2	0	-127.781	-127.78	0.998469	1.41772
96	0.699921	3.760294	0.300079	0	0.666667	0	0	-104.947	-104.299	0.52345	28.14597
97	0.603027	8.604025	0.396973	0.002296	0.65478	1	0	-124.562	-123.247	0.268662	6.619262
98	0.875254	1.204307	0.124746	0	0.666667	0	0	-58.8641	-58.8695	1	2.539301
99	0.737845	66.02957	0.262155	5.008412	0.037539	1	0	-27.53	-24.1481	0.033983	0.548281
100	0.532621	39.48258	0.467379	3.23451	0.094393	1	0	-138.486	-135.786	0.067196	5.269622

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
101	0.458306	23.07432	0.541694	4.990004	0.037897	17	0	-135.601	-131.906	0.024852	4.446111
102	1.00E-08	4.00386	1	2.605188	0.131624	0	0	-75.7914	-75.5976	0.823868	0.672717
103	1.00E-08	12.1118	1	21.67909	8.16E-06	9	0	-127.336	-125.264	0.125957	0
104	1	1.261669	0	0	0.666667	0	0	-59.0846	-59.0851	1	4.217772
105	0.935824	2.421361	0.064176	0	0.666667	0	0	-55.8067	-55.7436	0.938884	3.719683
106	0.819885	0.514536	0.180115	0	0.666667	0	0	-47.089	-47.0894	1	1.790525
107	1	0.833906	0	0	0.666667	0	0	-40.6328	-40.633	1	1.715916
108	0.527259	4.414507	0.472741	0.298724	0.491199	1	0	-73.0334	-72.4051	0.533518	2.437747
109	1.00E-08	2.662918	1	0.194097	0.530613	0	0	-98.2722	-98.2723	1	2.021687
110	1.00E-08	11.2926	1	6.17953	0.020565	1	0	-97.247	-95.3404	0.14859	1.156914
111	0.89709	32.76237	0.10291	0.82541	0.351762	2	0	-106.64	-104.294	0.095798	12.28033
112	1.00E-08	1.854777	1	0.279308	0.498021	0	0	-73.6136	-73.6139	1	1.328625

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
113	0.458342	22.23889	0.541658	3.312227	0.090619	1	0	-109.038	-106.491	0.078315	3.638654
114	1.00E-08	4.34141	1	1.764944	0.206914	0	0	-124.958	-124.959	1	2.362039
115	0.75811	7.083649	0.24189	-0.00048	0.666667	0	0	-165.05	-165.051	1	7.032013
116	0.770216	18.97322	0.229784	1.312776	0.265766	5	0	-54.5192	-53.3708	0.317149	2.893743
117	0.670393	10.54152	0.329607	0.291136	0.493843	0	0	-96.7637	-96.5499	0.807516	3.595593
118	0.632476	8.865504	0.367524	1.296097	0.268269	2	0	-54.8465	-53.7759	0.342804	1.583929
119	1.00E-08	13.63219	1	13.97256	0.000395	2	0	-144.468	-143.395	0.341702	0.973059
120	1	0.729456	0	0	0.666667	0	0	-17.4562	-17.4554	0.99922	1.278032
121	0.681978	15.29145	0.318022	0.749872	0.367977	9	0	-143.168	-139.294	0.020768	9.194008
122	0.951289	3.10386	0.048711	0	0.666667	0	0	-132.805	-132.806	1	3.796135
123	0.398906	29.25281	0.601094	4.144821	0.058685	2	0	-152.546	-148.683	0.02101	5.748957
124	0.677731	5.517895	0.322269	0	0.666667	0	0	-152.112	-152.112	1	5.656888

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
125	0.006109	6.604312	0.993891	0.853474	0.345971	5	0	-165.623	-165.445	0.836859	3.808829
126	1.00E-08	7.187204	1	0.05026	0.604351	0	0	-172.2	-172.201	1	6.467641
127	1.00E-08	23.90612	1	14.44409	0.000311	1	0	-218.087	-217.546	0.582289	2.200272
128	0.833222	37.99776	0.166778	5.335046	0.031726	0	0	-148.304	-146.507	0.165731	2.800028
129	0.025635	12.52427	0.974365	9.973606	0.00298	10	0	-164.785	-162.378	0.090091	2.319062
130	0.382024	26.84743	0.617976	3.728227	0.072896	8	0	-202.344	-198.22	0.016171	11.28422
131	1.00E-08	19.0602	1	20.62566	1.39E-05	0	0	-194.286	-193.279	0.365217	1.269752
132	0.302488	21.35189	0.697512	3.799054	0.070252	3	0	-161.004	-158.135	0.05675	5.352191
133	1.00E-08	13.19212	1	17.20968	7.72E-05	0	0	-199.394	-199.394	1	0.887278
134	1.00E-08	6.201156	1	2.532944	0.136783	1	0	-166.33	-166.33	0.999994	3.314206
135	0.227138	49.18511	0.772862	2.627405	0.130078	1	0	-159.12	-157.833	0.275961	4.479738
136	0.908366	1.515513	0.091634	0	0.666667	0	0	-97.6874	-97.6872	0.999797	2.485742
137	0.89406	2.640074	0.10594	0	0.666667	0	0	-82.289	-82.2919	1	4.12429

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
138	1	1.910447	0	0	0.666667	0	0	-78.6821	-78.6821	1	3.030729
139	0.846191	5.637762	0.153809	0	0.666667	0	0	-173.949	-173.949	1	8.203324
140	0.823809	2.760968	0.176191	0	0.666667	0	0	-68.0566	-68.0379	0.981495	4.850911
141	0.798652	32.13882	0.201348	0.510332	0.426587	0	0	-116.548	-116.206	0.710066	6.213341
142	1.00E-08	12.97287	1	2.874524	0.114109	1	0	-135.364	-134.118	0.28752	3.569822
143	1.00E-08	5.947868	1	3.676379	0.074895	2	0	-132.535	-132.523	0.988359	2.245476
144	0.92815	1716.642	0.07185	7.333834	0.0114	2	0	-75.8603	-70.448	0.004461	4.742445
145	0.96529	1.722713	0.03471	0	0.666667	0	0	-94.5361	-94.5366	1	3.400023
146	1	1.676674	0	0	0.666667	0	0	-81.5357	-81.5357	1	4.130274
147	0.921413	15.09524	0.078587	2.294799	0.155345	1	0	-37.4183	-34.1825	0.039328	1.82059
148	1.00E-08	6.392334	1	11.24337	0.001567	0	0	-140.495	-140.495	1	0.470048
149	1.00E-08	0.684181	1	0.19219	0.531402	0	0	-34.7561	-34.7562	1	0.431824

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
150	0.884824	4.845255	0.115176	0.025794	0.62369	0	0	-113.24	-113.225	0.984834	3.569335
151	0.44731	5.7368	0.55269	1.595027	0.227147	10	0	-81.9575	-80.3865	0.207834	1.9424
152	1.00E-08	0.695288	1	0.021913	0.627367	0	0	-36.7167	-36.7169	1	0.597336
153	1.00E-08	2.227597	1	0.560818	0.413187	0	0	-53.9876	-53.8843	0.901795	0.667183
154	0.783142	18.15924	0.216858	3.450246	0.084295	6	0	-86.7619	-80.8324	0.00266	4.973326
155	1.00E-08	4.559537	1	0.394786	0.459923	3	0	-124.493	-124.493	1	3.266213
156	1	0.973164	0	0	0.666667	0	0	-71.5012	-71.5013	1	5.577294
157	1.00E-08	3.351241	1	8.140579	0.007558	5	0	-100.581	-100.581	1	0.365417
158	1	0.760973	0	0	0.666667	0	0	-22.3299	-22.3299	1	3.170781
159	0.752464	74.39177	0.247536	2.399919	0.146845	0	0	-109.287	-107.705	0.205677	6.408649
160	0.111002	15.07803	0.888998	3.313771	0.090546	1	0	-107.927	-106.533	0.24815	1.826247
161	0.767796	4.2145	0.232204	0	0.666667	0	0	-66.9318	-65.9551	0.376578	4.288452

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
162	0.862659	5.528074	0.137341	0.019479	0.629813	0	0	-111.036	-111.018	0.982044	3.751713
163	0.93128	2.464017	0.06872	0	0.666667	0	0	-65.0556	-65.1073	1	5.514877
164	0.751303	2.620984	0.248697	0	0.666667	0	0	-113.191	-113.191	1	4.262056
165	0.924709	1.304788	0.075291	0	0.666667	0	0	-67.0672	-67.0671	0.999947	2.442638
166	0.781097	8.360339	0.218903	0.116043	0.566007	7	0	-97.144	-94.45	0.067608	6.491604
167	1	0.91908	0	0	0.666667	0	0	-19.5036	-19.5038	1	1.9415
168	0.87447	2.605881	0.12553	0	0.666667	0	0	-97.6576	-97.6572	0.999678	3.696856
169	1.00E-08	9.712706	1	0.640188	0.393322	0	0	-121.351	-121.077	0.760951	2.304297
170	0.882039	37.67852	0.117961	6.082449	0.021614	2	0	-87.4321	-84.9096	0.080258	1.071224
171	1.00E-08	2.589638	1	2.84396	0.115969	1	0	-89.6296	-89.6296	0.999995	0.962773
172	0.689992	19.51092	0.310008	0.590165	0.405679	0	0	-111.143	-110.843	0.741259	4.011088
173	0.812389	5.65847	0.187611	0.113573	0.567259	0	0	-88.2933	-88.2478	0.955506	2.991434
174	0.512272	1.600472	0.487728	0	0.666667	0	0	-93.8704	-93.87	0.999662	4.76438

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
175	0.711123	17.36557	0.288877	5.091721	0.035961	1	0	-83.7212	-82.4734	0.28715	0.739296
176	0.730484	15.39832	0.269516	2.576734	0.133631	5	0	-53.4016	-50.5838	0.059741	2.873608
177	1.00E-08	3.307035	1	3.840292	0.068758	0	0	-101.266	-101.266	0.999964	1.123366
178	1	1.023989	0	0	0.666667	0	0	-4.86277	-4.86277	1	1.549934
179	0.452252	18.41491	0.547748	1.106248	0.298758	1	0	-93.2808	-92.2555	0.358685	3.797954
180	0.938463	2.725976	0.061537	-0.00372	0.666667	0	0	-98.3966	-98.3992	1	2.479004
181	0.926103	8.902914	0.073897	0.001327	0.657707	3	0	-57.7591	-57.4122	0.706909	7.905239
182	0.496493	32.17711	0.503507	11.76828	0.001201	7	0	-142.557	-134.221	0.00024	4.483966
183	0.319484	3.803112	0.680516	0.388882	0.461741	5	0	-55.4353	-55.0589	0.686337	1.650282
184	0.3233	3.540798	0.6767	0.383691	0.46335	11	0	-84.3301	-83.8887	0.643148	2.14742
185	0.803916	3.162289	0.196084	0.006889	0.645562	3	0	-51.7414	-51.5138	0.796449	2.953673
186	0.925593	6.320545	0.074407	-0.00241	0.666667	0	0	-129.318	-129.319	1	4.089345

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
187	1.00E-08	2.998879	1	0.63957	0.393472	1	0	-98.099	-98.0992	1	1.639663
188	0.971631	2.173655	0.028369	0	0.666667	0	0	-91.8252	-91.8257	1	5.111854
189	1.00E-08	12.41292	1	7.707048	0.009425	7	0	-136.114	-133.474	0.071369	2.392782
190	0.689024	14.87971	0.310976	2.098939	0.172599	1	0	-51.3625	-50.2947	0.343737	1.0262
191	0.118509	30.40053	0.881491	5.40784	0.03056	3	0	-54.4323	-52.1597	0.103044	0.578656
192	1.00E-08	4.352395	1	0.363546	0.469686	4	0	-113.826	-113.73	0.908498	2.71907
193	0.679912	2.591114	0.320088	-0.00972	0.666667	0	0	-68.8486	-68.8526	1	2.140721
194	0.923285	0.368684	0.076715	0	0.666667	0	0	-51.9735	-51.9754	1	2.879738
195	1.00E-08	16.32541	1	8.402011	0.006617	2	0	-143.299	-142.225	0.341397	1.764395
196	1.00E-08	1.500919	1	0.048945	0.605275	1	0	-61.8653	-61.8648	0.999562	1.243445
197	0.654896	6.231236	0.345104	0.943513	0.328178	4	0	-58.5268	-57.169	0.257235	2.056922
198	0.553706	0.514977	0.446294	0	0.666667	0	0	-53.8538	-53.8557	1	1.618614

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
199	0.939314	1.695228	0.060686	0	0.666667	0	0	-82.1637	-82.1648	1	2.321549
200	1	0.590074	0	0	0.666667	0	0	-19.4456	-19.4459	1	1.0449
201	1.00E-08	1.042642	1	0.110949	0.5686	0	0	-50.3327	-50.333	1	0.836226
202	0.829216	1.562478	0.170784	0	0.666667	0	0	-79.0929	-79.0932	1	2.54669
203	1.00E-08	3.006729	1	0.022151	0.627134	2	0	-101.827	-101.825	0.998587	2.759159
204	0.833218	3.997172	0.166782	0	0.666667	0	0	-137.313	-137.314	1	4.172303
205	0.957806	1.65757	0.042194	0	0.666667	0	0	-68.8895	-68.8905	1	1.824666
206	0.547469	58.93999	0.452531	9.887026	0.003113	4	0	-107.306	-100.516	0.001126	2.940111
207	0.85995	2.059777	0.14005	0	0.666667	0	0	-44.1733	-44.1309	0.958485	3.038119
208	0.86757	2.117466	0.13243	0	0.666667	0	0	-92.5743	-92.574	0.999656	2.685218
209	0.84001	0.406323	0.15999	0	0.666667	0	0	-35.9795	-35.9801	1	2.102599
210	0.568707	7.20686	0.431293	0.256494	0.506295	11	0	-109.806	-108.706	0.332609	4.701271

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
211	1	0.795474	0	0	0.666667	0	0	-22.4969	-22.497	1	1.468988
212	0.344161	172.7796	0.655839	8.638917	0.005866	3	0	-81.49	-76.7578	0.008807	1.782349
213	1	0.583186	0	0	0.666667	0	0	-19.8804	-19.8804	0.999998	2.140395
214	1.00E-08	1.765484	1	0.043931	0.608895	0	0	-79.9481	-79.949	1	1.517283
215	0.875217	9.985364	0.124783	0.24934	0.508951	1	0	-23.6309	-23.485	0.864229	0.824179
216	0.565754	4.031357	0.434246	0.043835	0.608966	0	0	-70.0752	-69.7262	0.705399	2.780012
217	0.834312	10.39003	0.165688	0.270989	0.501005	0	0	-108.332	-108.179	0.858517	3.501655
218	1.00E-08	4.831066	1	6.426047	0.018127	0	0	-112.131	-112.135	1	1.191192
219	1	1.482853	0	0	0.666667	0	0	-32.9189	-32.9192	1	6.407647
220	1.00E-08	8.526056	1	3.288509	0.091754	1	0	-85.8081	-84.816	0.370806	1.409837
221	0.922092	1.251036	0.077908	0.003552	0.651758	0	0	-47.0333	-47.0319	0.998583	0.946434
222	0.788639	18.14012	0.211361	1.820633	0.200717	2	0	-67.7155	-65.4028	0.09899	3.737434

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
223	0.602494	7.484292	0.397506	0.097705	0.575565	0	0	-129.934	-129.883	0.950251	3.754643
224	1.00E-08	5.730726	1	0.963762	0.324333	1	0	-143.069	-143.069	1	3.572959
225	0.301717	5.897974	0.698283	1.230253	0.278417	6	0	-80.5912	-79.7083	0.413587	1.903774
226	1	1.180334	0	0	0.666667	0	0	-20.3437	-20.3437	1	2.253091
227	0.877901	0.257707	0.122099	0	0.666667	0	0	-22.258	-22.2584	1	1.353624
228	0.760191	0.609334	0.239809	0	0.666667	0	0	-60.2417	-60.2419	1	2.594126
229	1.00E-08	1.698257	1	0.298757	0.491188	0	0	-69.6873	-69.6875	1	1.104822
230	0.021846	27.51336	0.978154	6.973225	0.013704	3	0	-105.505	-102.091	0.032911	2.448416
231	1	0.584273	0	0	0.666667	0	0	-49.0226	-49.0242	1	1.921029
232	0.732425	3.072732	0.267575	0.068076	0.592644	3	0	-49.4253	-48.6703	0.47004	2.261342
233	0.785724	0.906346	0.214276	0	0.666667	0	0	-83.6438	-83.6442	1	4.961853
234	0.040831	7.052117	0.959169	9.923228	0.003057	0	0	-86.3493	-86.0368	0.731658	0.366448

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
235	0.670238	1.24987	0.329762	0.003979	0.650848	2	0	-28.2854	-28.2792	0.993776	0.871955
236	0.801267	0.769013	0.198733	0.000301	0.662457	0	0	-21.1533	-21.1273	0.974342	0.745196
237	1	1.60893	0	0	0.666667	0	0	-51.2226	-51.2226	0.999957	4.896949
238	0.891521	3.002813	0.108479	0	0.666667	0	0	-54.3725	-54.3833	1	12.38679
239	1.00E-08	4.487986	1	1.018077	0.314286	1	0	-111.543	-111.544	1	2.8706
240	0.935466	7.672286	0.064534	0.320939	0.483619	1	0	-46.4307	-44.446	0.137421	3.024962
241	1.00E-08	2.094949	1	0.046294	0.60717	0	0	-70.6199	-70.6202	1	1.835464
242	1	1.152859	0	0	0.666667	0	0	-22.8693	-22.8693	1	2.563639
243	0.384196	10.36394	0.615804	3.062028	0.103358	5	0	-75.3686	-73.7428	0.196755	2.393349
244	0.843005	1.231289	0.156995	0	0.666667	0	0	-31.9684	-31.9472	0.979059	1.579578
245	1.00E-08	2.681315	1	4.578283	0.046877	0	0	-61.6141	-61.6142	1	0.631437
246	0.108754	4.772729	0.891246	0.430312	0.449236	7	0	-97.3946	-97.2025	0.825203	2.88744

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
247	0.70377	18.3316	0.29623	0.199023	0.528589	0	0	-94.4787	-94.3612	0.889124	3.949289
248	1.00E-08	4.152761	1	1.487694	0.241042	0	0	-98.1637	-98.164	1	2.290915
249	0.749674	0.689722	0.250326	0	0.666667	0	0	-39.3171	-39.3172	1	2.218227
250	1.00E-08	8.11119	1	7.737649	0.00928	1	0	-152.381	-152.381	0.999808	1.865879
251	1.00E-08	4.257857	1	1.420954	0.250157	0	0	-103.801	-103.801	1	2.498792
252	0.67594	0.432519	0.32406	0	0.666667	0	0	-40.8894	-40.8896	1	1.049926
253	0.912993	1.701399	0.087007	0	0.666667	0	0	-80.1899	-80.1902	1	2.80535
254	0.808754	24.43211	0.191246	4.258888	0.055312	6	0	-67.6453	-61.8834	0.003145	4.72169
255	0.461308	5.596614	0.538692	1.849116	0.197625	3	0	-30.548	-29.3435	0.299849	0.821675
256	0.83312	3.462642	0.16688	0	0.666667	0	0	-108.756	-108.768	1	5.910811
257	1	0.365628	0	0	0.666667	0	0	-12.0875	-12.0876	1	0.760216
258	1.00E-08	2.520468	1	2.696049	0.125422	3	0	-41.7453	-41.622	0.884044	0.097873

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
259	0.950208	125.7104	0.049792	6.721287	0.015586	1	0	-33.1382	-26.9742	0.002104	2.478853
260	0.742272	2.574661	0.257728	0	0.666667	0	0	-104.244	-104.243	0.998985	15.58357
261	0.773524	0.989184	0.226476	0	0.666667	0	0	-71.3916	-71.3808	0.989258	4.06954
262	0.846974	2.115542	0.153026	0	0.666667	0	0	-73.8015	-73.802	1	2.150044
263	0.775532	54.60755	0.224468	1.818767	0.200921	0	0	-35.6657	-34.6285	0.35446	1.565372
264	0.40169	2.21389	0.59831	3.229276	0.094653	4	0	-27.0833	-26.632	0.636761	0
265	1	1.701483	0	0	0.666667	0	0	-52.24	-52.2399	0.999905	5.58172
266	0.811307	0.704534	0.188693	0	0.666667	0	0	-60.7761	-60.7761	1	3.830897
267	0.804292	0.610963	0.195708	0	0.666667	0	0	-41.9234	-41.9052	0.981928	2.457546
268	0.859764	1.787141	0.140236	-0.00074	0.666667	0	0	-48.6806	-48.6814	1	1.267012
269	1	1.682598	0	0	0.666667	0	0	-50.8759	-50.8759	0.999966	5.727688
270	0.448356	3.735952	0.551644	0	0.666667	0	0	-87.1959	-86.9096	0.751052	5.883781

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
271	1	1.023989	0	0	0.666667	0	0	-4.86277	-4.86277	1	1.549934
272	1.00E-08	2.662608	1	0.63188	0.395338	0	0	-89.3278	-89.3279	1	1.745253
273	1.00E-08	2.162849	1	3.270452	0.092628	0	0	-62.6908	-62.6908	1	0.458942
274	0.998054	0.449573	0.001946	0	0.666667	0	0	-31.9395	-31.9396	1	2.397257
275	0.44614	4.790562	0.55386	0.310609	0.487115	7	0	-67.0583	-66.4667	0.553433	2.271221
276	0.858627	7.253745	0.141373	0.003998	0.65081	0	0	-126.854	-126.85	0.995883	5.285788
277	1.00E-08	4.482166	1	1.013552	0.315109	0	0	-103.556	-103.556	1	2.776076
278	0.057776	21.55541	0.942224	14.97214	0.000238	8	0	-73.3121	-67.5857	0.003259	0.303314
279	1.00E-08	18.23722	1	2.148927	0.168011	0	0	-102.505	-101.454	0.34968	3.547759
280	0.206607	14.33507	0.793393	7.411256	0.010959	2	0	-80.8997	-78.9713	0.145377	0.905469
281	0.023885	1.633831	0.976115	-0.0249	0.666667	0	0	-53.8131	-53.8241	1	1.08922
282	0.784026	5.224389	0.215974	0.049025	0.605219	3	0	-49.8909	-48.8868	0.36635	4.414257

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
283	0.783258	1.846761	0.216742	0	0.666667	0	0	-64.8473	-64.8233	0.976295	2.298245
284	0.840972	1.242949	0.159028	0	0.666667	0	0	-30.0675	-30.0475	0.980211	1.398506
285	0.68104	3.116876	0.31896	0	0.666667	0	0	-67.8422	-67.78	0.939625	4.346854
286	0.486399	0.877418	0.513601	0	0.666667	0	0	-50.5556	-50.5559	1	2.766099
287	1.00E-08	2.196757	1	0.060594	0.597392	0	0	-81.123	-81.1237	1	1.907959
288	0.901846	1.538236	0.098154	0	0.666667	0	0	-61.7751	-61.7752	1	2.312421
289	0.907831	1.53112	0.092169	0	0.666667	0	0	-60.377	-60.3776	1	2.070148
290	1.00E-08	1.539541	1	0.033902	0.616678	0	0	-62.9699	-62.9699	0.999991	1.380634
291	0.892426	59.2605	0.107574	2.485675	0.140273	0	0	-104.566	-103.97	0.551065	2.494655
292	1.00E-08	2.422664	1	0.179052	0.536924	4	0	-66.0397	-65.9897	0.951221	1.501494
293	0.877994	3.535323	0.122006	0	0.666667	0	0	-105.849	-105.849	1	6.567341
294	1	1.916987	0	0	0.666667	0	0	-49.8155	-49.8154	0.999974	5.87344

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
295	1	0.627186	0	0	0.666667	0	0	-28.6092	-28.6092	0.999997	2.357909
296	0.506305	18.66362	0.493695	5.181041	0.034343	4	0	-76.0707	-73.2796	0.061351	2.153639
297	1	1.724598	0	0	0.666667	0	0	-28.9104	-28.9104	1	5.771207
298	1.00E-08	2.656559	1	0.314824	0.485683	1	0	-80.1591	-80.1595	1	1.960237
299	1	1.180526	0	0	0.666667	0	0	-30.3121	-30.3122	1	3.322867
300	0.262522	10.26947	0.737478	0.100029	0.574318	0	0	-138.265	-138.21	0.946296	5.539913
301	1.00E-08	2.544592	1	1.537172	0.234524	0	0	-75.2115	-75.2116	1	1.183458
302	1.00E-08	5.006952	1	2.495289	0.139556	0	0	-114.036	-114.036	1	2.184347
303	0.890021	22.95248	0.109979	1.165884	0.288768	2	0	-36.1124	-32.6695	0.031971	2.519473
304	0.895429	49.73276	0.104571	5.972957	0.022862	2	0	-100.905	-98.6792	0.108019	2.063579
305	1.00E-08	3.215942	1	0.789861	0.359276	0	0	-74.5056	-74.4678	0.962917	1.196107
306	1	20.34486	0	0.535064	0.419943	0	0	-33.9523	-33.6849	0.765374	1.782807

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
307	0.530769	16.15943	0.469231	5.991498	0.022646	4	0	-82.2356	-77.8232	0.012127	2.536545
308	0.55384	13.85367	0.44616	2.458314	0.142336	6	0	-68.2133	-66.2444	0.139603	2.187862
309	0.763464	6.212914	0.236536	0.023937	0.625418	5	0	-65.2846	-64.5184	0.464811	5.313216
310	0.416256	35.59141	0.583744	2.952275	0.109517	1	0	-93.9253	-92.1271	0.165597	3.754972
311	0.786323	70.93211	0.213677	11.85285	0.001151	2	0	-67.2665	-62.7099	0.010497	0.44109
312	0.977393	1.77672	0.022607	0	0.666667	0	0	-66.3552	-66.3559	1	2.475878
313	1.00E-08	7.912484	1	3.221102	0.09506	6	0	-71.786	-70.35	0.237884	1.495532
314	0.239383	3.096369	0.760617	0.427488	0.450071	6	0	-57.8217	-57.5761	0.782211	1.549868
315	0.036054	10.34787	0.963946	1.189217	0.284965	0	0	-103.708	-103.143	0.568204	2.895802
316	0.899411	0.711892	0.100589	0	0.666667	0	0	-37.9274	-37.928	1	0.922454
317	0.539741	13.6672	0.460259	3.018589	0.105752	7	0	-90.8398	-87.8261	0.04911	3.599864
318	0.894855	0.857677	0.105145	0	0.666667	0	0	-35.5038	-35.506	1	0.836167
319	0.087252	17.36708	0.912748	3.241877	0.094028	0	0	-96.302	-95.7591	0.581064	1.462765

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
320	0.658592	1.610705	0.341408	0	0.666667	0	0	-64.4383	-64.4385	1	4.887224
321	0.593194	13.17423	0.406806	0.860855	0.344468	5	0	-103.757	-102.304	0.233857	5.307614
322	0.362206	34.90356	0.637794	4.286295	0.054531	4	0	-45.5755	-44.0467	0.216792	0.128635
323	0.173729	1007.464	0.826271	14.06495	0.000377	3	0	-39.3827	-32.3058	0.000844	14.93231
324	1.00E-08	129.3224	1	9.080142	0.004689	2	0	-35.3675	-32.1583	0.040391	3.491034
325	0.446855	0.707908	0.553145	0	0.666667	0	0	-29.9408	-29.9434	1	1.072781
326	0.879478	4.82469	0.120522	0.001351	0.657624	0	0	-117.439	-117.439	0.999777	4.632019
327	1.00E-08	14.40875	1	1.995518	0.182521	5	0	-139.775	-138.984	0.453304	4.60395
328	1	0.996079	0	0	0.666667	0	0	-43.6724	-43.6725	1	3.303054
329	1.00E-08	29.19881	1	19.13584	2.93E-05	6	0	-116.417	-109.226	0.000754	1.385785
330	0.917528	1.824093	0.082472	0	0.666667	0	0	-67.7413	-67.7437	1	1.917297
331	0.21692	9.966459	0.78308	0.831154	0.350567	0	0	-127.623	-127.4	0.800693	4.111473

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
332	0.193581	36.06271	0.806419	1.546125	0.233365	1	0	-68.3104	-67.5246	0.455771	1.923275
333	1.00E-08	2.95863	1	9.416364	0.003953	4	0	-86.2766	-86.2766	1	0
334	0.801484	1.18376	0.198516	0	0.666667	0	0	-62.6756	-62.677	1	3.881063
335	1.00E-08	1.452432	1	0.875133	0.341583	0	0	-61.9181	-61.9169	0.99882	0.708141
336	0.746883	6.583662	0.253117	0.238171	0.513161	4	0	-78.3549	-77.3799	0.377188	4.443746
337	0.578239	2.758251	0.421761	0	0.666667	0	0	-75.8092	-75.7774	0.968639	2.800573
338	0.25669	9.526692	0.74331	0.850215	0.346637	0	0	-99.7801	-99.0103	0.463122	3.137455
339	0.619276	5.35528	0.380724	1.167225	0.288547	4	0	-40.6169	-39.3957	0.294891	1.144866
340	0.868326	8.248643	0.131674	0.020193	0.629083	0	0	-65.6619	-65.6363	0.974678	5.585673
341	1.00E-08	4.45874	1	0.187582	0.533321	0	0	-119.416	-119.416	1	3.586596
342	0.672626	4.776343	0.327374	0	0.666667	0	0	-87.0295	-86.8183	0.809627	5.608071
343	0.316816	12.52331	0.683184	1.255028	0.274548	1	0	-137.642	-136.291	0.258934	5.354077

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
344	0.748256	0.329065	0.251744	0	0.666667	0	0	-40.5308	-40.5313	1	1.80592
345	1.00E-08	18.2288	1	10.60479	0.002164	6	0	-118.324	-115.387	0.05305	1.870411
346	0.529927	0.989052	0.470073	0	0.666667	0	0	-63.9119	-63.9125	1	2.247477
347	0.240832	128.5227	0.759168	1.321668	0.264442	1	0	-70.7707	-70.0967	0.509654	1.798725
348	1	0.788085	0	0	0.666667	0	0	-24.5123	-24.5125	1	3.036874
349	1.00E-08	18.02823	1	10.37053	0.002437	1	0	-116.075	-113.727	0.095552	1.297081
350	0.692639	5.991093	0.307361	0.06939	0.591831	6	0	-92.4581	-91.356	0.332156	4.892812
351	1.00E-08	4.196812	1	0.297746	0.491539	7	0	-114.412	-114.371	0.959489	2.906532
352	0.649682	3.3842	0.350318	0	0.666667	0	0	-73.2153	-73.1676	0.953457	3.709772
353	0.8665	3.123146	0.1335	0	0.666667	0	0	-112.837	-112.837	1	4.147267
354	1	1.534087	0	0	0.666667	0	0	-81.7412	-81.7421	1	3.65137
355	1.00E-08	2.886197	1	1.240036	0.276882	0	0	-82.2474	-82.2474	1	1.427458

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
356	0.016413	33.22342	0.983587	4.879264	0.040125	1	0	-55.394	-52.9904	0.090398	1.226734
357	0.651441	10.31413	0.348559	1.654662	0.21981	4	0	-53.3752	-51.9548	0.241625	2.083877
358	0.025148	7.293505	0.974852	0.198933	0.528626	3	0	-152.018	-151.93	0.915454	5.288764
359	1	0.547744	0	0	0.666667	0	0	-4.86277	-4.86277	1	1.549934
360	1.00E-08	2.714297	1	0.906738	0.335305	0	0	-87.9455	-87.946	1	1.514174
361	0.692838	11.77835	0.307162	2.598185	0.132115	2	0	-55.2675	-52.3355	0.053293	2.085203
362	0.790468	5.841058	0.209532	0.067363	0.593087	0	0	-88.1045	-88.0659	0.962095	2.526263
363	0.615055	23.34323	0.384945	9.290563	0.004214	10	0	-110.003	-102.241	0.000426	3.260639
364	0.881088	2.193885	0.118912	0	0.666667	0	0	-60.561	-60.5588	0.997824	2.501706
365	0.276233	8.701938	0.723767	1.97356	0.184705	6	0	-122.062	-120.468	0.203168	3.766328
366	0.917071	0.673554	0.082929	0	0.666667	0	0	-58.8591	-58.8604	1	2.400224
367	1.00E-08	7.704552	1	2.295807	0.155261	5	0	-140.361	-140.301	0.941682	3.353282

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
368	0.331778	56.59639	0.668222	5.851017	0.024337	1	0	-66.5719	-63.4042	0.042103	1.549148
369	0.755432	1.678923	0.244568	0	0.666667	0	0	-111.714	-111.714	1	8.080456
370	0.234792	14.90307	0.765208	4.097123	0.060158	1	0	-91.4115	-89.5431	0.154366	1.742972
371	0.626047	30.12018	0.373953	3.027627	0.105249	1	0	-67.4241	-65.4953	0.145333	2.477866
372	1.00E-08	1.549189	1	1.345098	0.26099	7	0	-58.0055	-58.0055	0.999999	0
373	0.158224	6.764627	0.841776	0.784599	0.360405	7	0	-108.876	-108.386	0.612883	3.560206
374	1	0.935433	0	0	0.666667	0	0	-46.8881	-46.8881	1	1.911712
375	0.539152	0.68709	0.460848	0	0.666667	0	0	-43.3235	-43.3238	1	1.529137
376	0.89512	2.676678	0.10488	0	0.666667	0	0	-123.513	-123.505	0.99258	5.948765
377	1.00E-08	6.936922	1	1.623029	0.223669	1	0	-67.4916	-66.9582	0.586593	0.865999
378	0.863098	0.596454	0.136902	0	0.666667	0	0	-44.9743	-44.9748	1	4.219731
379	0.790721	9.111533	0.209279	1.800184	0.202969	0	0	-86.1143	-85.4935	0.53753	1.055428

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
380	0.226972	3.36476	0.773028	0.405647	0.456611	4	0	-70.0315	-69.6834	0.705984	1.923877
381	1	0.61023	0	0	0.666667	0	0	-20.6293	-20.6293	1	1.0449
382	1.00E-08	2.747877	1	0.615676	0.39931	2	0	-95.7949	-95.7928	0.99795	1.684304
383	0.889373	2.496016	0.110627	0	0.666667	0	0	-59.6556	-59.6507	0.99509	7.947202
384	1.00E-08	4.217538	1	1.186426	0.285417	0	0	-107.485	-107.485	1	2.106524
385	0.741325	5.282642	0.258675	0	0.666667	0	0	-89.0616	-88.6434	0.658261	5.326846
386	0.876922	1.035735	0.123078	0	0.666667	0	0	-45.4299	-45.4314	1	1.125348
387	1.00E-08	2.929472	1	0.010798	0.639863	1	0	-105.563	-105.564	1	2.768989
388	0.879509	3.450476	0.120491	0	0.666667	0	0	-94.2179	-94.2292	1	5.861065
389	0.582279	7.542577	0.417721	0.07005	0.591426	0	0	-103.136	-102.445	0.500709	6.178778
390	1	1.025483	0	0	0.666667	0	0	-35.5436	-35.5434	0.99986	2.701553
391	0.750523	6.89618	0.249477	0	0.666667	0	0	-119.847	-119.545	0.739553	10.62187

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
392	1	0.739576	0	0	0.666667	0	0	-32.1567	-32.1567	1	1.626898
393	0.727492	24.2438	0.272508	5.611121	0.027527	0	0	-137.828	-136.163	0.18921	2.386255
394	1	0.923374	0	0	0.666667	0	0	-18.9457	-18.944	0.99835	1.939555
395	0.652906	13.31739	0.347094	0.740606	0.370033	5	0	-106.919	-104.489	0.088069	7.756175
396	1	0.627169	0	0	0.666667	0	0	-22.8784	-22.8784	0.999998	2.35782
397	0.802185	5.086851	0.197815	0.001484	0.657177	3	0	-58.5806	-57.6134	0.380124	4.807213
398	1.00E-08	1.629979	1	1.233782	0.277862	0	0	-53.4192	-53.4193	1	0.730005
399	0.042317	6.456411	0.957683	1.702239	0.214142	7	0	-78.7585	-77.9207	0.432627	1.753349
400	0.853725	0.68486	0.146275	0	0.666667	0	0	-46.0554	-46.0561	1	3.481307
401	0.600494	11.57902	0.399506	0.549455	0.416148	7	0	-123.276	-121.099	0.113294	7.078783
402	1.00E-08	2.574528	1	2.998472	0.10688	1	0	-78.8063	-78.8071	1	0.776866
403	0.510549	19.19674	0.489451	2.178083	0.165394	2	0	-73.0209	-71.47	0.212056	2.27175

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
404	0.937734	111.4565	0.062266	8.910043	0.005112	0	0	-105.511	-101.766	0.023632	2.07019
405	0.42662	12.14867	0.57338	3.159289	0.0982	0	0	-151.44	-150.845	0.551643	3.114017
406	1.00E-08	3.63685	1	0.035729	0.615198	0	0	-118.16	-118.16	0.999564	3.180649
407	1	1.315051	0	0	0.666667	0	0	-43.1541	-43.1543	1	6.695542
408	0.412183	7.861526	0.587817	1.912592	0.190918	1	0	-81.0583	-80.0665	0.370905	2.194367
409	0.435175	23.2801	0.564825	2.443334	0.143478	3	0	-116.397	-113.867	0.079672	4.474
410	1.00E-08	2.623117	1	0.471111	0.437458	1	0	-85.6914	-85.6915	1	1.749747
411	1.00E-08	6.021058	1	0.76103	0.365522	0	0	-150.391	-150.391	1	3.682574
412	0.606258	0.409656	0.393742	0	0.666667	0	0	-36.6815	-36.6815	1	0.985579
413	1.00E-08	7.038088	1	0.145321	0.551884	0	0	-173.099	-173.099	1	5.960536
414	0.804757	1.759154	0.195243	0	0.666667	0	0	-42.2056	-42.1505	0.946322	2.189336
415	0.009825	5.862197	0.990175	0.383974	0.463262	4	0	-124.369	-124.329	0.960366	3.526643
416	1.00E-08	13.06442	1	22.4166	5.63E-06	4	0	-89.0815	-87.5793	0.222633	0

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
417	1.00E-08	3.098792	1	2.711493	0.124398	0	0	-40.5073	-40.5076	1	0
418	0.90343	5.551399	0.09657	0	0.666667	0	0	-96.4102	-96.413	1	5.790287
419	1	34582.86	0	38.3629	1.89E-09	0	0	-174.961	-156.805	1.30E-08	3.139102
420	1.00E-08	64.2051	1	5.582061	0.027942	0	0	-138.004	-136.129	0.153366	3.129475
421	1.00E-08	6.364787	1	4.771782	0.042415	1	0	-129.704	-129.705	1	1.909938
422	0.012054	204.1239	0.987946	30.40355	1.02E-07	2	0	-133.837	-119.481	5.82E-07	2.439485
423	0.36279	16.69553	0.63721	6.337298	0.018969	6	0	-120.035	-115.996	0.017616	3.457228
424	0.267854	17.94415	0.732146	1.314852	0.265456	3	0	-63.6352	-62.7513	0.413186	5.410552
425	0.298942	25.52511	0.701058	4.869904	0.040319	0	0	-68.4271	-67.8572	0.565595	1.16862
426	0.359563	83.97551	0.640437	3.98522	0.063761	1	0	-87.4082	-85.0387	0.093524	1.944211
427	1	64.29282	0	2.688309	0.125938	2	0	-47.3718	-46.028	0.260853	2.478696
428	0.69394	11.55025	0.30606	0.327828	0.481315	7	0	-115.352	-114.252	0.332782	6.776725

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
429	1.00E-08	39.64252	1	6.994398	0.013556	4	0	-176.69	-174.416	0.102829	2.975284
430	1	1.030594	0	0	0.666667	0	0	-48.4708	-48.4718	1	2.887535
431	0.772189	5.95262	0.227811	0.193293	0.530945	3	0	-50.6522	-50.44	0.808803	1.95391
432	0.783521	1.426134	0.216479	0	0.666667	0	0	-51.8924	-51.9271	1	2.794985
433	0.751507	6.886878	0.248493	0.017632	0.631757	5	0	-99.474	-98.029	0.235754	6.312107
434	0.673698	14.94164	0.326302	0.875975	0.341414	2	0	-31.7258	-31.2517	0.622438	0.663405
435	1	0.760973	0	0	0.666667	0	0	-29.9966	-29.9966	1	3.170781
436	0.587974	6.331022	0.412026	1.837788	0.198849	3	0	-37.5778	-36.0674	0.220816	0.889261
437	0.778279	0.909234	0.221721	0	0.666667	0	0	-65.1577	-65.159	1	2.995867
438	0.669104	9.589979	0.330896	0.695688	0.380214	2	0	-58.6257	-58.0888	0.584563	1.757418
439	0.628865	85.48725	0.371135	4.118308	0.059499	2	0	-105.294	-103.044	0.105411	3.569897
440	1.00E-08	2.325861	1	1.894775	0.192775	0	0	-91.7878	-91.7879	1	1.095741
441	1	0.844002	0	0	0.666667	0	0	-78.6981	-78.6989	1	5.34561

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
442	0.740811	12.79829	0.259189	1.968418	0.185221	6	0	-84.5664	-81.761	0.060482	3.513958
443	0.773849	1.021057	0.226151	0	0.666667	0	0	-80.7534	-80.7556	1	3.24553
444	1.00E-08	8.633968	1	1.482182	0.241781	1	0	-98.4821	-97.7575	0.484517	1.910553
445	0.691751	0.708125	0.308249	0	0.666667	0	0	-69.4343	-69.4345	1	1.814501
446	0.630029	1.967583	0.369971	0	0.666667	0	0	-69.6968	-69.6776	0.980975	3.284594
447	1	174.8369	0	8.571355	0.006071	2	0	-116.865	-112.885	0.018677	1.976521
448	0.703412	14.0333	0.296588	0.221775	0.519488	0	0	-45.9276	-45.8059	0.885417	2.970112
449	1.00E-08	3.845507	1	0.541615	0.418209	0	0	-88.4153	-88.4157	1	2.604727
450	0.821354	11.02352	0.178646	0.427774	0.449986	2	0	-52.8139	-52.5743	0.786968	1.278931
451	0.865565	5.274979	0.134435	0	0.666667	0	0	-141.576	-141.58	1	6.745416
452	0.794876	2.566074	0.205124	0	0.666667	0	0	-49.2218	-49.1849	0.963815	3.581099
453	0.798277	2.877074	0.201723	0	0.666667	0	0	-57.7933	-57.7039	0.914515	4.226555

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
454	0.178296	22.67875	0.821704	2.1315	0.169595	5	0	-60.6642	-59.5996	0.344876	0.87647
455	0.597998	4.488694	0.402002	0	0.666667	0	0	-106.025	-105.918	0.897956	6.315712
456	1.00E-08	5.151269	1	4.540036	0.047814	10	0	-124.233	-124.111	0.885212	1.009629
457	0.738172	7.142837	0.261828	0.138858	0.554899	0	0	-90.2589	-90.1788	0.923005	2.151442
458	0.775661	170.3225	0.224339	6.70203	0.01574	2	0	-80.8069	-77.5186	0.037318	1.453649
459	1.00E-08	2.3906	1	0.134088	0.557159	0	0	-110.627	-110.628	1	1.99524
460	0.302034	49.02241	0.697966	5.152272	0.034856	1	0	-87.3037	-84.5229	0.061989	1.751803
461	1.00E-08	1.906899	1	5.926839	0.023409	5	0	-67.8542	-67.8542	1	0
462	1	45.29463	0	5.784758	0.025179	1	0	-59.1295	-56.6083	0.080366	0.79435
463	0.632625	4.021097	0.367375	0	0.666667	0	0	-96.8581	-96.7129	0.864914	4.110252
464	1.00E-08	1.044579	1	1.359766	0.258855	5	0	-39.2685	-39.2686	1	0
465	1.00E-08	6.877668	1	2.063644	0.175919	2	0	-77.6775	-76.8001	0.415865	1.444582

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
466	0.33128	4.385311	0.66872	0	0.666667	0	0	-120.781	-120.438	0.709559	7.32822
467	0.572966	7.991955	0.427034	1.869462	0.195448	7	0	-76.5521	-74.0794	0.084354	2.634133
468	0.379906	5.393121	0.620094	1.03221	0.311734	1	0	-55.4314	-54.6321	0.449676	1.389669
469	1.00E-08	8.642958	1	11.45289	0.001409	4	0	-40.934	-38.0176	0.054128	0
470	0.808342	2.305232	0.191658	0	0.666667	0	0	-45.1653	-44.5462	0.538446	3.545109
471	0.262774	4.804859	0.737226	1.088215	0.301859	1	0	-51.9103	-51.3842	0.590905	1.070335
472	0.800652	2.579942	0.199348	0	0.666667	0	0	-45.274	-44.4614	0.44373	3.137397
473	1.00E-08	2.730187	1	4.356459	0.052583	0	0	-84.4405	-84.4404	0.999942	0.684734
474	1.00E-08	0.793747	1	1.003252	0.316991	0	0	-29.0372	-29.0373	1	0.286801
475	0.757705	1319.326	0.242295	5.201719	0.03398	1	0	-81.6279	-78.9633	0.069629	2.584252
476	1.00E-08	14.68924	1	4.268696	0.055031	1	0	-44.4072	-42.8588	0.212591	0.348476
477	0.558672	1.205113	0.441328	0	0.666667	0	0	-64.5477	-64.4271	0.886382	3.756339

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
478	0.342032	33.69516	0.657968	3.286406	0.091855	1	0	-56.3883	-54.5788	0.163733	1.26864
479	1.00E-08	5.918582	1	3.96553	0.064418	8	0	-65.2395	-63.5558	0.185674	0.814393
480	0.236597	24.01316	0.763403	5.640932	0.027109	2	0	-68.3576	-65.7502	0.073724	1.22302
481	0.815249	7.473861	0.184751	0.014007	0.635845	0	0	-124.528	-124.494	0.966686	5.321142
482	0.608234	3.477327	0.391766	0	0.666667	0	0	-105.424	-105.411	0.986988	6.487279
483	1.00E-08	0.921869	1	0.001676	0.656564	0	0	-47.5578	-47.5581	1	0.902328
484	1	2.250813	0	0	0.666667	0	0	-69.5884	-69.5871	0.998675	5.782157
485	0.949343	2.315732	0.050657	0	0.666667	0	0	-87.5205	-87.5219	1	6.41936
486	0.425046	11.20182	0.574954	1.213593	0.281054	8	0	-122.998	-121.393	0.200796	5.482599
487	0.820548	2.770251	0.179452	0	0.666667	0	0	-98.6917	-98.6813	0.989731	2.701262
488	0.776611	20.34277	0.223389	1.654889	0.219782	1	0	-120.785	-119.79	0.369612	5.732404
489	0.695614	7.597898	0.304386	2.113949	0.171207	0	0	-60.5454	-59.9038	0.526493	0.867278

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
490	1.00E-08	4.684753	1	1.028471	0.312407	1	0	-126.459	-126.46	1	2.743567
491	1.00E-08	2.809155	1	0.422393	0.451583	0	0	-96.703	-96.7034	1	1.855964
492	0.442647	5.174651	0.557353	3.004271	0.106553	4	0	-45.7055	-44.0184	0.185055	0.585987
493	0.435251	13.46343	0.564749	1.334455	0.262552	1	0	-96.9975	-96.0822	0.400401	3.281715
494	0.885234	20.8844	0.114766	2.154717	0.167487	0	0	-118.344	-117.941	0.66847	2.607954
495	0.902303	12.66786	0.097697	0.908692	0.334922	2	0	-46.1143	-43.6215	0.08268	4.065852
496	0.48603	5.705924	0.51397	0.16396	0.543467	1	0	-80.2277	-79.8787	0.705397	3.162795
497	0.351578	7.442668	0.648422	1.206379	0.282205	4	0	-88.4276	-87.3631	0.344875	2.682657
498	0.070253	2.783289	0.929747	2.975153	0.108203	1	0	-39.8986	-39.6704	0.79601	0.251729
499	0.891184	2.539234	0.108816	0	0.666667	0	0	-91.3685	-91.3687	1	4.807277
500	0.78352	0.690204	0.21648	0	0.666667	0	0	-45.1555	-45.1554	0.999924	2.042774
501	0.752137	4.165363	0.247863	-0.02853	0.666667	0	0	-105.972	-105.987	1	3.33288

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
502	0.335489	5.510978	0.664511	0.112418	0.567848	0	0	-116.001	-115.931	0.931679	3.602179
503	0.870386	3.060248	0.129614	0	0.666667	0	0	-78.0426	-78.036	0.993495	3.53305
504	0.897009	39.50497	0.102991	6.384584	0.018515	2	0	-85.9035	-83.7653	0.117859	0.824731
505	0.839373	21.47276	0.160627	2.697206	0.125345	0	0	-99.6202	-98.9809	0.527618	2.050919
506	0.307272	4.30462	0.692728	0.013194	0.636822	0	0	-110.635	-110.63	0.995065	3.404077
507	0.193246	9.211459	0.806754	1.263578	0.273227	0	0	-103.166	-102.571	0.551268	3.35359
508	0.847087	11.03776	0.152913	0.079759	0.585619	0	0	-103.415	-103.375	0.960889	3.370965
509	0.531321	17.626	0.468679	2.501294	0.13911	1	0	-110.068	-108.786	0.277564	3.266022
510	0.123137	473.2247	0.876863	10.57435	0.002198	2	0	-60.6917	-55.4254	0.005163	1.568345
511	1	55.70869	0	1.815691	0.201259	1	0	-81.6878	-80.7803	0.403521	3.78363
512	0.224325	2.713495	0.775675	-0.00046	0.666667	0	0	-84.0213	-84.0225	1	2.635118
513	0.884727	3.076473	0.115273	0	0.666667	0	0	-114.328	-114.329	1	4.715938

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
514	1	2.771839	0	0	0.666667	0	0	-88.1329	-88.1332	1	3.261268
515	0.299384	8.11162	0.700616	2.091282	0.173314	9	0	-87.2057	-85.7507	0.233393	2.497952
516	1.00E-08	4.934336	1	9.883825	0.003119	4	0	-101.469	-101.47	1	0.534206
517	1.00E-08	13.85699	1	1.6946	0.215041	0	0	-114.217	-113.476	0.47687	2.810373
518	0.955965	1.338193	0.044035	0	0.666667	0	0	-68.086	-68.0866	1	2.485596
519	1.00E-08	0.717643	1	0.021816	0.627462	0	0	-34.7736	-34.7735	0.99996	0.608058
520	0.678308	38.17933	0.321692	2.089574	0.173473	2	0	-100.252	-97.8539	0.090898	5.729026
521	1.00E-08	2.488792	1	0.051359	0.603585	1	0	-91.6375	-91.6371	0.999607	2.291079
522	0.789883	11.32241	0.210117	3.511468	0.081636	2	0	-32.3359	-28.6892	0.026077	1.284201
523	0.976032	1.655915	0.023968	0	0.666667	0	0	-68.1434	-68.1445	1	3.480373
524	1.00E-08	2.862726	1	3.488888	0.082607	3	0	-24.868	-23.9725	0.408373	0
525	0.991175	0.946326	0.008825	0	0.666667	0	0	-77.8734	-77.8744	1	7.623808

	p¹	β⁺	p⁺	LRT	p-value	# branches under selection	Total branch length	MEME LogL	FEL LogL	FEL α	FEL β
526	0.359024	11.48614	0.640976	1.704233	0.213908	1	0	-34.9066	-34.0414	0.420981	0.633637
527	0.447443	11.99291	0.552557	3.401827	0.08646	7	0	-81.8586	-79.3034	0.077676	2.140469
528	0.52549	18.46363	0.47451	1.364007	0.258241	1	0	-73.1768	-72.2214	0.384673	2.982056

Appedix-11 : aBSREL analysis results

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
1	*Apodemus sylvaticus*	12.27833	0.000739	0.022899	Terminal	Under Selection	Rodents
2	*Cavia porcellus*	71.36178	1.11E-16	4.77E-15	Terminal	Under Selection	Rodents
3	*Chinchilla lanigera*	4.275097	0.043118	0.819243	Terminal	Under Selection	Rodents
4	*Dipodomys spectabilis*	17.22282	6.13E-05	0.002389	Terminal	Under Selection	Rodents
5	*Fukomys damarensis*	13.44086	0.000411	0.013565	Terminal	Under Selection	Rodents
6	*Heterocephalus glaber*	21.57104	6.89E-06	0.000276	Terminal	Under Selection	Rodents
7	*Ictidomys tridecemlineatus*	5.691686	0.020835	0.458363	Terminal	Under Selection	Rodents
8	*Jaculus jaculus*	12.54112	0.000647	0.020703	Terminal	Under Selection	Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
9	*Marmota marmota*	14.39333	0.000254	0.008904	Terminal	Under Selection	Rodents
10	*Mesocricetus auratus*	6.82629	0.011678	0.280266	Terminal	Under Selection	Rodents
11	*Microtus oregoni*	15.78154	0.000126	0.00468	Terminal	Under Selection	Rodents
12	*Mus musculus*	26.60888	5.50E-07	2.31E-05	Terminal	Under Selection	Rodents
13	*Myodes glareolus*	0	1	1	Terminal	Not Under Selection	Rodents
14	*Node11*	0	1	1	Non-terminal	Not Under Selection	Rodents
15	*Node13*	0	1	1	Non-terminal	Not Under Selection	Rodents
16	*Node16*	9.509303	0.002992	0.086778	Non-terminal	Under Selection	Rodents
17	*Node18*	16.85166	7.38E-05	0.002805	Non-terminal	Under Selection	Rodents
18	*Node2*	25.88104	7.92E-07	3.25E-05	Non-terminal	Under Selection	Rodents
19	*Node20*	3.70861	0.057795	0.982515	Non-terminal	Not Under Selection	Rodents
20	*Node21*	0.056632	0.454005	1	Non-terminal	Not Under Selection	Rodents
21	*Node22*	5.302826	0.025424	0.533896	Non-terminal	Under Selection	Rodents
22	*Node23*	0	1	1	Non-terminal	Not Under Selection	Rodents
23	*Node26*	0	1	1	Non-terminal	Not Under Selection	Rodents
24	*Node30*	7.747368	0.00731	0.197379	Non-terminal	Under Selection	Rodents
25	*Node31*	0.077546	0.444721	1	Non-terminal	Not Under Selection	Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
26	*Node33*	0	1	1	Non-terminal	Not Under Selection	Rodents
27	*Node35*	0	1	1	Non-terminal	Not Under Selection	Rodents
28	*Node38*	3.780601	0.055678	1	Non-terminal	Not Under Selection	Rodents
29	*Node39*	6.10814	0.016841	0.387347	Non-terminal	Under Selection	Rodents
30	*Node4*	7.799221	0.00712	0.199369	Non-terminal	Under Selection	Rodents
31	*Node42*	4.40707	0.040282	0.805634	Non-terminal	Under Selection	Rodents
32	*Node7*	2.422138	0.11329	1	Non-terminal	Not Under Selection	Rodents
33	*Node8*	13.8528	0.000334	0.011356	Non-terminal	Under Selection	Rodents
34	*Octodon degus*	7.215911	0.009577	0.239434	Terminal	Under Selection	Rodents
35	*Onychomys torridus*	0	1	1	Terminal	Not Under Selection	Rodents
36	*Peromyscus eremicus*	0	1	1	Terminal	Not Under Selection	Rodents
37	*Peromyscus leucopus*	0	1	1	Terminal	Not Under Selection	Rodents
38	*Peromyscus maniculatus*	2.918668	0.087229	1	Terminal	Not Under Selection	Rodents
39	*Phodopus roborovskii*	7.333948	0.009019	0.234504	Terminal	Under Selection	Rodents
40	*Psammomys obesus*	0	1	1	Terminal	Not Under Selection	Rodents
41	*Rattus norvegicus*	15.45767	0.000149	0.005359	Terminal	Under Selection	Rodents
42	*Rattus rattus*	0	1	1	Terminal	Not Under Selection	Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
43	*Urocitellus parryii*	10.35377	0.001952	0.058563	Terminal	Under Selection	Rodents
1	*Aotus nancymaae*	0	1	1	Terminal	Not Under Selection	Non-Rodents
2	*Artibeus jamaicensis*	13.68475	0.000364	0.042531	Terminal	Under Selection	Non-Rodents
3	*Balaenoptera musculus*	5.098743	0.028228	1	Terminal	Under Selection	Non-Rodents
4	*Bison bison*	0	1	1	Terminal	Not Under Selection	Non-Rodents
5	*Bos taurus*	0	1	1	Terminal	Not Under Selection	Non-Rodents
6	*Budorcas taxicolor*	0	1	1	Terminal	Not Under Selection	Non-Rodents
7	*Callithrix jacchus*	3.254749	0.073177	1	Terminal	Not Under Selection	Non-Rodents
8	*Camelus dromedarius*	0.154702	0.416486	1	Terminal	Not Under Selection	Non-Rodents
9	*Canis lupus familiaris*	7.083772	0.010243	0.983357	Terminal	Under Selection	Non-Rodents
10	*Cercocebus atys*	13.15868	0.000474	0.054969	Terminal	Under Selection	Non-Rodents
11	*Chlorocebus sabaeus*	84.7743	0	0	Terminal	Under Selection	Non-Rodents
12	*Chrysochloris asiatica*	9.703673	0.002712	0.295608	Terminal	Under Selection	Non-Rodents
13	*Colobus angolensis*	9.481329	0.003035	0.32778	Terminal	Under Selection	Non-Rodents
14	*Condylura cristata*	25.77678	8.35E-07	0.000108	Terminal	Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
15	*Desmodus rotundus*	6.400543	0.014507	1	Terminal	Under Selection	Non-Rodents
16	*Echinops telfairi*	25.45837	9.79E-07	0.000124	Terminal	Under Selection	Non-Rodents
17	*Equus caballus*	20.67082	1.08E-05	0.001354	Terminal	Under Selection	Non-Rodents
18	*Erinaceus europaeus*	7.582516	0.007949	0.802839	Terminal	Under Selection	Non-Rodents
19	*Felis catus*	0	1	1	Terminal	Not Under Selection	Non-Rodents
20	*Gorilla gorilla*	0	1	1	Terminal	Not Under Selection	Non-Rodents
21	*Homo sapiens*	0	1	1	Terminal	Not Under Selection	Non-Rodents
22	*Hylobates moloch*	0.133503	0.423594	1	Terminal	Not Under Selection	Non-Rodents
23	*Lepus europaeus*	6.597559	0.013121	1	Terminal	Under Selection	Non-Rodents
24	*Loxodonta africana*	2.52053	0.107548	1	Terminal	Not Under Selection	Non-Rodents
25	*Lynx canadensis*	0	1	1	Terminal	Not Under Selection	Non-Rodents
26	*Macaca fascicularis*	0	1	1	Terminal	Not Under Selection	Non-Rodents
27	*Macaca mulatta*	0	1	1	Terminal	Not Under Selection	Non-Rodents
28	*Macaca nemestrina*	0.082367	0.442719	1	Terminal	Not Under Selection	Non-Rodents
29	*Manis javanica*	0	1	1	Terminal	Not Under Selection	Non-Rodents
30	*Manis pentadactyla*	0	1	1	Terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
31	*Mirounga angustirostris*	0	1	1	Terminal	Not Under Selection	Non-Rodents
32	*Moschus berezovskii*	11.84423	0.00092	0.104827	Terminal	Under Selection	Non-Rodents
33	*Mustela putorius furo*	27.37991	3.74E-07	4.86E-05	Terminal	Under Selection	Non-Rodents
34	*Myotis davidii*	5.241999	0.026229	1	Terminal	Under Selection	Non-Rodents
35	*Myotis myotis*	12.18508	0.000774	0.089038	Terminal	Under Selection	Non-Rodents
36	*Node10*	6.649455	0.012778	1	Non-terminal	Under Selection	Non-Rodents
37	*Node100*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
38	*Node105*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
39	*Node106*	10.74976	0.001598	0.180577	Non-terminal	Under Selection	Non-Rodents
40	*Node107*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
41	*Node109*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
42	*Node113*	1.825521	0.155756	1	Non-terminal	Not Under Selection	Non-Rodents
43	*Node114*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
44	*Node115*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
45	*Node117*	-0.00141	0.5	1	Non-terminal	Not Under Selection	Non-Rodents
46	*Node12*	37.41809	2.44E-09	3.22E-07	Non-terminal	Under Selection	Non-Rodents
47	*Node120*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
48	*Node121*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
49	*Node126*	4.808684	0.032761	1	Non-terminal	Under Selection	Non-Rodents
50	*Node129*	15.2499	0.000165	0.019667	Non-terminal	Under Selection	Non-Rodents
51	*Node130*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
52	*Node15*	0.514857	0.32652	1	Non-terminal	Not Under Selection	Non-Rodents
53	*Node16*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
54	*Node17*	8.962074	0.003948	0.418465	Non-terminal	Under Selection	Non-Rodents
55	*Node18*	19.32286	2.13E-05	0.0026	Non-terminal	Under Selection	Non-Rodents
56	*Node21*	0.187272	0.406224	1	Non-terminal	Not Under Selection	Non-Rodents
57	*Node22*	5.744284	0.020282	1	Non-terminal	Under Selection	Non-Rodents
58	*Node23*	15.69411	0.000132	0.015859	Non-terminal	Under Selection	Non-Rodents
59	*Node26*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
60	*Node27*	3.78725	0.055486	1	Non-terminal	Not Under Selection	Non-Rodents
61	*Node3*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
62	*Node30*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
63	*Node33*	8.274533	0.005594	0.570591	Non-terminal	Under Selection	Non-Rodents
64	*Node34*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
65	*Node36*	1.315239	0.205715	1	Non-terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
66	*Node39*	5.785576	0.019858	1	Non-terminal	Under Selection	Non-Rodents
67	*Node4*	1.30736	0.206613	1	Non-terminal	Not Under Selection	Non-Rodents
68	*Node42*	3.499241	0.064431	1	Non-terminal	Not Under Selection	Non-Rodents
69	*Node44*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
70	*Node45*	4.233801	0.044047	1	Non-terminal	Under Selection	Non-Rodents
71	*Node48*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
72	*Node5*	10.417	0.001891	0.207975	Non-terminal	Under Selection	Non-Rodents
73	*Node50*	14.99548	0.000188	0.022167	Non-terminal	Under Selection	Non-Rodents
74	*Node51*	-0.07935	0.5	1	Non-terminal	Not Under Selection	Non-Rodents
75	*Node54*	6.068391	0.017186	1	Non-terminal	Under Selection	Non-Rodents
76	*Node56*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
77	*Node58*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
78	*Node59*	0.117509	0.429232	1	Non-terminal	Not Under Selection	Non-Rodents
79	*Node6*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
80	*Node62*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
81	*Node65*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
82	*Node67*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
83	*Node68*	2.024626	0.139967	1	Non-terminal	Not Under Selection	Non-Rodents
84	*Node7*	10.67233	0.001662	0.18612	Non-terminal	Under Selection	Non-Rodents
85	*Node70*	7.031168	0.010521	0.999513	Non-terminal	Under Selection	Non-Rodents
86	*Node73*	-0.02984	0.5	1	Non-terminal	Not Under Selection	Non-Rodents
87	*Node74*	5.470287	0.023334	1	Non-terminal	Under Selection	Non-Rodents
88	*Node77*	9.01484	0.003844	0.411271	Non-terminal	Under Selection	Non-Rodents
89	*Node80*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents
90	*Node81*	0.202565	0.401635	1	Non-terminal	Not Under Selection	Non-Rodents
91	*Node83*	19.43422	2.02E-05	0.002479	Non-terminal	Under Selection	Non-Rodents
92	*Node84*	3.044481	0.081669	1	Non-terminal	Not Under Selection	Non-Rodents
93	*Node86*	0.894679	0.260641	1	Non-terminal	Not Under Selection	Non-Rodents
94	*Node89*	10.56923	0.001751	0.194324	Non-terminal	Under Selection	Non-Rodents
95	*Node90*	0.234638	0.392412	1	Non-terminal	Not Under Selection	Non-Rodents
96	*Node91*	5.770103	0.020016	1	Non-terminal	Under Selection	Non-Rodents
97	*Node94*	1.098271	0.232164	1	Non-terminal	Not Under Selection	Non-Rodents
98	*Node95*	25.76344	8.40E-07	0.000108	Non-terminal	Under Selection	Non-Rodents
99	*Node98*	0.650726	0.30065	1	Non-terminal	Not Under Selection	Non-Rodents
100	*Node99*	0	1	1	Non-terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
101	*Nomascus leucogenys*	-0.00443	0.5	1	Terminal	Not Under Selection	Non-Rodents
102	*Nyctereutes procyonoides*	0.013312	0.479639	1	Terminal	Not Under Selection	Non-Rodents
103	*Odocoileus virginianus*	1.327234	0.204355	1	Terminal	Not Under Selection	Non-Rodents
104	*Orcinus orca*	0	1	1	Terminal	Not Under Selection	Non-Rodents
105	*Ornithorhynchus anatinus*	8.614833	0.004707	0.489566	Terminal	Under Selection	Non-Rodents
106	*Oryctolagus cuniculus*	7.153899	0.009884	0.958776	Terminal	Under Selection	Non-Rodents
107	*Ovis aries*	3.909782	0.052075	1	Terminal	Not Under Selection	Non-Rodents
108	*Pan paniscus*	8.771814	0.004347	0.456468	Terminal	Under Selection	Non-Rodents
109	*Pan troglodytes*	0	1	1	Terminal	Not Under Selection	Non-Rodents
110	*Panthera pardus*	0	1	1	Terminal	Not Under Selection	Non-Rodents
111	*Panthera tigris*	3.653312	0.059477	1	Terminal	Not Under Selection	Non-Rodents
112	*Panthera uncia*	0.802867	0.274856	1	Terminal	Not Under Selection	Non-Rodents
113	*Papio anubis*	17.89164	4.38E-05	0.005295	Terminal	Under Selection	Non-Rodents
114	*Phyllostomus discolor*	21.76322	6.26E-06	0.000788	Terminal	Under Selection	Non-Rodents
115	*Pongo abelii*	0	1	1	Terminal	Not Under Selection	Non-Rodents
116	*Pongo pygmaeus*	0	1	1	Terminal	Not Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
117	*Propithecus coquereli*	0	1	1	Terminal	Not Under Selection	Non-Rodents
118	*Rhinopithecus bieti*	0.627254	0.30491	1	Terminal	Not Under Selection	Non-Rodents
119	*Rhinopithecus roxellana*	0	1	1	Terminal	Not Under Selection	Non-Rodents
120	*Rousettus aegyptiacus*	7.576691	0.007972	0.789274	Terminal	Under Selection	Non-Rodents
121	*Saccopteryx bilineata*	0	1	1	Terminal	Not Under Selection	Non-Rodents
122	*Saccopteryx leptura*	0.169644	0.411689	1	Terminal	Not Under Selection	Non-Rodents
123	*Saimiri boliviensis*	0.895054	0.260584	1	Terminal	Not Under Selection	Non-Rodents
124	*Sorex araneus*	7.172517	0.009791	0.959529	Terminal	Under Selection	Non-Rodents
125	*Sorex fumeus*	7.582345	0.00795	0.794959	Terminal	Under Selection	Non-Rodents
126	*Sus scrofa*	0	1	1	Terminal	Not Under Selection	Non-Rodents
127	*Talpa occidentalis*	19.5851	1.87E-05	0.002317	Terminal	Under Selection	Non-Rodents
128	*Theropithecus gelada*	8.35331	0.005375	0.553615	Terminal	Under Selection	Non-Rodents
129	*Trachypithecus francoisi*	0.049998	0.457207	1	Terminal	Not Under Selection	Non-Rodents
130	*Ursus arctos*	5.198398	0.026821	1	Terminal	Under Selection	Non-Rodents
131	*Ursus maritimus*	0	1	1	Terminal	Not Under Selection	Non-Rodents
132	*Vicugna pacos*	4.248568	0.043712	1	Terminal	Under Selection	Non-Rodents

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
133	*Vombatus ursinus*	30.55079	7.62E-08	9.98E-06	Terminal	Under Selection	Non-Rodents
1	*Aotus nancymaae*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
2	*Apodemus sylvaticus*	12.63851	0.000616	0.091778	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
3	*Artibeus jamaicensis*	13.51521	0.000396	0.059785	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
4	*Balaenoptera musculus*	5.217679	0.026558	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
5	*Bison bison*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
6	*Bos taurus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
7	*Budorcas taxicolor*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
8	*Callithrix jacchus*	3.285525	0.072012	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
9	*Camelus dromedarius*	0.174357	0.410208	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
10	*Canis lupus familiaris*	7.164627	0.00983	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
11	*Cavia porcellus*	69.24417	3.33E-16	5.96E-14	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
12	*Cercocebus atys*	13.32381	0.000436	0.065404	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
13	*Chinchilla lanigera*	3.312805	0.070996	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
14	*Chlorocebus sabaeus*	98.58671	0	0	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
15	*Chrysochloris asiatica*	9.850425	0.002518	0.357547	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
16	*Colobus angolensis*	9.559109	0.002918	0.408498	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
17	*Condylura cristata*	26.78885	5.02E-07	8.69E-05	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
18	*Desmodus rotundus*	6.235546	0.015781	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
19	*Dipodomys spectabilis*	7.356278	0.008918	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
20	*Echinops telfairi*	27.34734	3.80E-07	6.65E-05	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
21	*Equus caballus*	20.86988	9.80E-06	0.001617	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
22	*Erinaceus europaeus*	7.802286	0.007109	0.924201	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
23	*Felis catus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
24	*Fukomys damarensis*	12.21658	0.000762	0.112782	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
25	*Gorilla gorilla*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
26	*Heterocephalus glaber*	23.41524	2.73E-06	0.000459	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
27	*Homo sapiens*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
28	*Hylobates moloch*	0.142522	0.420523	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
29	*Ictidomys tridecemlineatus*	5.574321	0.022124	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
30	*Jaculus jaculus*	7.846808	0.00695	0.910491	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
31	*Lepus europaeus*	5.068352	0.028671	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
32	*Loxodonta africana*	2.560605	0.105297	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
33	*Lynx canadensis*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
34	*Macaca fascicularis*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
35	*Macaca mulatta*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
36	*Macaca nemestrina*	0.088898	0.440074	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
37	*Manis javanica*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
38	*Manis pentadactyla*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
39	*Marmota marmota*	17.11966	6.45E-05	0.010258	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
40	*Mesocricetus auratus*	6.57957	0.013242	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
41	*Microtus oregoni*	13.99782	0.00031	0.047812	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
42	*Mirounga angustirostris*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
43	*Moschus berezovskii*	12.10565	0.000806	0.118468	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
44	*Mus musculus*	26.04819	7.29E-07	0.000125	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
45	*Mustela putorius furo*	24.45492	1.62E-06	0.000275	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
46	*Myodes glareolus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
47	*Myotis davidii*	4.59624	0.036543	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
48	*Myotis myotis*	13.57529	0.000384	0.058387	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
49	*Node10*	6.184506	0.016197	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
50	*Node100*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
51	*Node105*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
52	*Node106*	11.0728	0.001357	0.19411	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
53	*Node107*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
54	*Node109*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
55	*Node113*	1.820641	0.156166	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
56	*Node114*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
57	*Node115*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
58	*Node117*	0.001344	0.493949	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
59	*Node12*	37.04927	2.93E-09	5.19E-07	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
60	*Node120*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
61	*Node121*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
62	*Node126*	3.210559	0.074883	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
63	*Node127*	4.452635	0.039347	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
64	*Node130*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
65	*Node131*	6.928655	0.011085	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
66	*Node132*	26.86645	4.83E-07	8.41E-05	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
67	*Node134*	6.725445	0.012293	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
68	*Node137*	8.551839	0.00486	0.660977	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
69	*Node138*	13.89944	0.000326	0.049915	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
70	*Node141*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
71	*Node143*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
72	*Node146*	15.16289	0.000173	0.027282	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
73	*Node148*	24.34468	1.71E-06	0.000289	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
74	*Node15*	0.273183	0.381939	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
75	*Node150*	3.532036	0.063343	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
76	*Node151*	0.010952	0.481692	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
77	*Node152*	5.827632	0.019436	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
78	*Node153*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
79	*Node156*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
80	*Node16*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
81	*Node160*	8.442	0.005138	0.693696	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
82	*Node161*	0.039006	0.462893	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
83	*Node163*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
84	*Node165*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
85	*Node168*	4.60221	0.036431	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
86	*Node169*	6.300181	0.015269	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
87	*Node17*	14.49311	0.000242	0.03774	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
88	*Node172*	4.43275	0.039752	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
89	*Node175*	11.33322	0.00119	0.17138	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
90	*Node176*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
91	*Node18*	18.83405	2.73E-05	0.004442	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
92	*Node21*	0.331074	0.367237	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
93	*Node22*	6.738997	0.012209	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
94	*Node23*	17.13711	6.40E-05	0.010296	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
95	*Node26*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
96	*Node27*	3.576513	0.061896	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
97	*Node3*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
98	*Node30*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
99	*Node33*	8.290486	0.005549	0.738009	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
100	*Node34*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
101	*Node36*	1.38344	0.198117	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
102	*Node39*	5.882703	0.018896	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
103	*Node4*	2.2593	0.123507	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
104	*Node42*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
105	*Node44*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
106	*Node45*	4.44056	0.039593	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
107	*Node48*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
108	*Node5*	12.09618	0.00081	0.118225	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
109	*Node50*	15.10156	0.000178	0.02796	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
110	*Node51*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
111	*Node54*	6.390963	0.014578	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
112	*Node56*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
113	*Node58*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
114	*Node59*	0.129329	0.42504	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
115	*Node6*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
116	*Node62*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
117	*Node65*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
118	*Node67*	9.367839	0.003214	0.446811	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
119	*Node68*	2.550323	0.10587	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
120	*Node7*	17.12968	6.42E-05	0.010271	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
121	*Node70*	6.962451	0.010896	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
122	*Node73*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
123	*Node74*	4.894042	0.031355	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
124	*Node77*	9.307694	0.003314	0.457312	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
125	*Node80*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
126	*Node81*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
127	*Node83*	21.09089	8.77E-06	0.001456	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
128	*Node84*	2.604911	0.102866	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
129	*Node86*	0.936074	0.254527	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
130	*Node89*	11.82298	0.000929	0.13477	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
131	*Node90*	0.185164	0.406867	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
132	*Node91*	5.549683	0.022405	1	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
133	*Node94*	1.113184	0.230226	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
134	*Node95*	25.90925	7.81E-07	0.000134	Non-terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
135	*Node98*	0.706955	0.290764	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
136	*Node99*	0	1	1	Non-terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
137	*Nomascus leucogenys*	0.033278	0.4661	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
138	*Nyctereutes procyonoides*	0.016143	0.477368	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
139	*Octodon degus*	7.153408	0.009887	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
140	*Odocoileus virginianus*	1.360875	0.200596	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
141	*Onychomys torridus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
142	*Orcinus orca*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
143	*Ornithorhynchus anatinus*	8.0589	0.006241	0.823792	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
144	*Oryctolagus cuniculus*	7.202593	0.009642	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
145	*Ovis aries*	3.964614	0.050619	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
146	*Pan paniscus*	8.847081	0.004185	0.573294	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
147	*Pan troglodytes*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
148	*Panthera pardus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
149	*Panthera tigris*	3.727497	0.057232	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
150	*Panthera uncia*	0.899683	0.259892	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
151	*Papio anubis*	18.05147	4.04E-05	0.006542	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
152	*Peromyscus eremicus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
153	*Peromyscus leucopus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
154	*Peromyscus maniculatus*	3.035009	0.082075	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
155	*Phodopus roborovskii*	6.787223	0.011912	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
156	*Phyllostomus discolor*	22.24145	4.92E-06	0.000822	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
157	*Pongo abelii*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
158	*Pongo pygmaeus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
159	*Propithecus coquereli*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
160	*Psammomys obesus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
161	*Rattus norvegicus*	14.08166	0.000298	0.046132	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
162	*Rattus rattus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
163	*Rhinopithecus bieti*	0.632363	0.303976	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
164	*Rhinopithecus roxellana*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
165	*Rousettus aegyptiacus*	4.979006	0.030017	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
166	*Saccopteryx bilineata*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
167	*Saccopteryx leptura*	0.185574	0.406742	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
168	*Saimiri boliviensis*	0.873937	0.263771	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
169	*Sorex araneus*	7.288457	0.00923	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)

X	*Name*	LRT	Uncorrected.p.value	Corrected.p.value	Branch_Type	Selection_Status	Group
170	*Sorex fumeus*	7.390186	0.008765	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
171	*Sus scrofa*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
172	*Talpa occidentalis*	20.42164	1.23E-05	0.002013	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
173	*Theropithecus gelada*	8.432502	0.005163	0.691882	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
174	*Trachypithecus francoisi*	0.051525	0.456456	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
175	*Urocitellus parryii*	9.659546	0.002773	0.391027	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
176	*Ursus arctos*	5.248985	0.026135	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
177	*Ursus maritimus*	0	1	1	Terminal	Not Under Selection	Whole Gene (Terminal/Non-Terminal)
178	*Vicugna pacos*	4.361889	0.041231	1	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)
179	*Vombatus ursinus*	32.33	3.12E-08	5.50E-06	Terminal	Under Selection	Whole Gene (Terminal/Non-Terminal)