

**EVOLUTION OF MAMMALIAN SCALES AND IMMUNITY
WITHIN PANGOLIN GENOMES AND DATABASE
DEVELOPMENT**

TAN TZE KING

**FACULTY OF DENTISTRY
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2016

**EVOLUTION OF MAMMALIAN SCALES AND IMMUNITY
WITHIN PANGOLIN GENOMES AND DATABASE
DEVELOPMENT**

TAN TZE KING

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY**

**FACULTY OF DENTISTRY
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2016

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: TAN TZE KING

Registration/Matric No: DHA130010

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

EVOLUTION OF MAMMALIAN SCALES AND IMMUNITY WITHIN PANGOLIN
GENOMES AND DATABASE DEVELOPMENT

Field of Study:

BIOINFORMATICS

I do solemnly and sincerely declare that:

I am the sole author/writer of this Work;

This Work is original;

Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Pangolins (Order, Pholidota) are unique mammals that are edentulous, have poor vision, an acute olfaction system and they possess scales that cover most of their body. Among eight known pangolin species, the Malayan (*Manis javanica*) and Chinese pangolins (*M. pentadactyla*) are critically endangered species. In this study, I performed the first comparative genome analysis for the Malayan and Chinese pangolins to investigate pangolin biology and evolution. Strikingly, I found that interferon-epsilon (*IFNE*), which is exclusively expressed in epithelial cells and important in skin and mucosal immunity, is pseudogenised in all Asian and African pangolin species examined, perhaps impacting resistance to infection.

Gene family expansion/contraction analysis clearly showed that Olfactory Receptor (OR) gene families are significantly expanded in pangolins, reflecting their well-developed olfaction system. The significantly contracted IFN gene family was observed, probably reflects poor immunity in pangolins. Interestingly, I also found the pseudogenization of genes responsible for tooth development such as enamelins (*ENAM*), ameloblastins (*AMBN*), and amelogenins (*AMELX*), suggesting that the presumed loss of function of these genes might be linked to the edentulous characteristics of pangolins. Furthermore, I have identified mutations that cause the loss of Guanylate cyclase activator 1C (*GUCA1C*) and intermediate filament-like protein (*BFSP2*) gene functions, potentially affect pangolin vision acuity. In general, my findings reflect the unique traits of pangolins.

I hypothesize that the reduced IFN-mediated immunity from the loss of *IFNE* and the contraction of interferon gene family in pangolins imposed strong selective pressure on immunity-related genes. Identification of positively-selected genes that involving in a

wide range of immunity-related pathways including hematopoietic cell lineage, cytosolic DNA-sensing pathway, complement and coagulation cascades, cytokine-cytokine receptor interaction, and the phagosome pathway. This positive selection analysis also provided substantial evidence of pangolin specialized adaptations associated with the energy storage and metabolism, muscular and nervous systems, and scale/hair development.

To facilitate research in pangolins, I have developed a Pangolin Genome Database (PGD), aiming to be a future pangolin genome hub for hosting pangolin genomic and transcriptomic data and annotations, and supported by useful analysis tools for the research communities. Currently, the PGD provides the reference pangolin genome and transcriptome sequences, gene and functional information, expressed transcripts, pseudogenes, genomic variations, organ-specific expression data and other useful annotations. Therefore, it will be an invaluable platform for researchers, who are interested in pangolin and mammalian research in future.

This study provides insights into mammalian adaptation and functional diversification, new research tools and questions, and perhaps a new natural *IFNE*-deficient animal model for studying mammalian immunity in future. Lastly, I propose a potential relationship between scales development and *IFNE* pseudogenization. The development of pangolin scales could hypothetically be a counter mechanism for an inherently poor immune response and provide physical protection against injuries or stress and reduce pangolin's vulnerability to infection.

ABSTRAK

Tenggiling-Tenggiling (Order, Pholidota) merupakan mamalia-mamalia yang unik dengan sisik-sisik di atas sebahagian besar badan mereka, tiada gigi, mempunyai penglihatan yang lemah, sistem pencium yang kuat. Antara lain spesies tenggiling yang diketahui, tenggiling Malayan (*Manis javanica*) dan tenggiling Cina (*Manis pentadactyla*) merupakan spesies-spesies terancam yang amat kritikal. Saya telah menjalankan analisa genom secara perbandingan yang pertama bagi tenggiling-tenggiling Malayan dan Cina demi menyiasat biologi dan evolusi tenggiling. Sungguh menarik bahawa interferon-epsilon (*IFNE*), semata-mata dinyatakan di dalam sel-sel epitelium dan penting dalam kulit dan imuniti mukosa, adalah dipseudogenisasi dalam kesemua spesies tenggiling Asian dan Afrika yang telah saya periksa, dan mungkin menyebabkan rintangan terhadap jangkitan.

Analisis perkembangan dan pengecutan keluarga gen jelas menunjukkan bahawa keluarga-keluarga gen pencium (OR) berkembang secara ketara di dalam kalangan tenggiling-tenggiling, menggambarkan sistem pencium yang menyeluruh. Pengecutan keluarga gen-gen IFN yang ketara diperhatikan dan berkemungkinan manandakan imuniti yang lemah di dalam tenggiling-tenggiling. Sungguh menariknya, saya juga mendapati pseudogenisasi gen-gen yang bertanggungjawab bagi pertumbuhan gigi seperti pseudogenisasi enamelin (*ENAM*), ameloblastin (*AMBN*), dan amelogenin (*AMELX*), mencadangkan kemungkinan kehilangan fungsi-fungsi gen ini adalah berhubung dengan ciri-ciri ketiada gigian tenggiling-tenggiling. Selain itu, saya telah mengenalpasti mutasi-mutasi yang menyebabkan kehilangan fungsi-fungsi gen Guanylate cyclase aktivator 1C (*GUCA1C*) dan intermediate filament-like protein (*BFSP2*), berpotensi menyebabkan penglihatan yang lemah. Secara umum, pendapat saya mendedahkan ciri-ciri unik tenggiling-tenggiling.

Saya berhipotesis bahawa kekurangan immuniti pengantaraan-IFN daripada kehilangan *IFNE* dan pengecutan keluarga gen interferon di dalam tenggiling-tenggiling mengenakan tekanan terpilih yang kuat terhadap gen-gen yang berkaitan dengan immuniti. Pengenalpastian gen-gen pemilihan positif yang terlibat di dalam pelbagai laluan yang berkaitan dengan imuniti termasuk keturunan hematopoietic sel, laluan cytosolic DNA-sensing, sistem pembekuan koagulasi, interaksi reseptor cytokine-cytokine, dan laluan phagosome. Analisis pemilihan positif ini juga memberikan bukti yang substansial adaptasi-adaptasi khusus tenggiling berkaitan dengan penyimpanan tenaga dan metabolisme, sistem otot dan saraf, dan pertumbuhan sisik/rambut.

Untuk memudahkan penyelidikan dalam tenggiling-tenggiling, saya telah membangunkan Pangolin Genome Database (PGD), bertujuan untuk menjadi hab genom tenggiling pada masa depan untuk menjadi perumah data genomik dan transkriptomik dan anotasi-anotasi tenggiling, dan disokong oleh peralatan analisa yang berguna untuk komuniti-komuniti penyelidikan. Pada masa kini, PGD menyediakan rujukan genom tenggiling dan transkriptomik, penjujukan-penjujukan gen dan maklumat berfungsi, transkrip-transkrip bernyata, pseudogen-pseudogen, variasi-variasi genomik, organ data ungkapan tertentu dan anotasi-anotasi yang lain yang berguna. Oleh yang demikian, PGD akan menjadi platform yang berharga untuk para penyelidik-penyelidik yang berminat dalam penyelidikan tenggiling dan mamalia pada masa depan.

Kesimpulannya, kajian ini memberi pandangan yang baru dalam penyesuaian dan kepelbagaian fungsi mamalia serta alat-alat penyelidikan dan persoalan-persoalan yang baru, dan mungkin suatu model haiwan kekurangan gen *IFNE* semula jadi yang baru untuk kajian sistem imuniti mammalia pada masa depan. Natiujahnya, saya mencadangkan bahawa terdapat hubungan yang berpotensi antara penumbuhan sisik-sisik dan pseudogenisasi *IFNE*. Penumbuhan sisik-sisik tenggiling boleh secara hipotesis menjadi

mekanisme kaunter untuk tindak balas imun yang sememangnya lemah dan menyediakan perlindungan fizikal terhadap kecederaan-kecederaan dan tekanan dan mengurangkan pendedahan tenggiling kepada jangkitan.

University of Malaya

ACKNOWLEDGEMENTS

I would like to thank everyone who had contributed to the successful completion of this project. I would preferably like to thank my supervisors, Dr. Lawrence Choo Siew Woh and Professor Dr. Ian Charles Paterson for their invaluable advice, guidance, and their enormous patience to guide me during the project. I appreciate their willingness to teach, supervise and encourage me throughout my Ph.D. study and the preparation of the project reports and papers.

In addition, I would also like to thank my colleagues in Genome Informatics Research Laboratory for their supports, opinions and suggestions to me throughout this project. Special thanks to International Pangolin Research Consortium (IPaRC) members for assisting me in my research.

I would like to thanks to Bright Sparks Scholarship (SBSUM) for financially supported me for my Ph.D. program. I would like to take this opportunity to thank the staff from the High Impact Research (HIR) secretariat and the Faculty of Dentistry for their excellent administrative works.

TABLE OF CONTENTS

Abstract	iii
Abstrak	v
Acknowledgements	viii
Table of Contents	ix
List of Figures	xiv
List of Tables.....	xvi
List of Symbols and Abbreviations.....	xvii
List of Appendices	xix
CHAPTER 1: INTRODUCTION	1
1.1 Overview.....	1
1.2 Field of Research	3
1.3 Project Objectives.....	3
CHAPTER 2: LITERATURE REVIEW.....	4
2.1 Pholidota Order (Pangolin).....	4
2.1.1 <i>Manis Javanica</i> (Malayan Pangolin)	8
2.1.2 <i>Manis Pentadactyla</i> (Chinese Pangolin).....	10
2.2 Conservation Efforts on Pangolins	11

2.3	Unique Pangolin Traits	12
2.3.1	Edentulous.....	12
2.3.2	Poor Vision and Hearing.....	13
2.3.3	Strong Sense of Smell	13
2.3.4	Poor Immunity	14
2.3.5	Pangolin Scales	14
2.4	Sequencing Study of Pangolin Genomes.....	15
2.5	Comparative Genomic Analyses.	16
2.6	Pseudogene Analysis	17
2.7	Gene Gain and Loss Analyses	21
2.8	Positive Selection Analysis	22
CHAPTER 3: MATERIALS AND METHODOLOGY		25
3.1	Pangolin Specimen	25
3.2	Pseudogene Screening	25
3.3	Pangolin Trait Study.....	26
3.4	Sanger Sequencing Validation of Pseudogenes.....	26
3.5	RNA-Seq Expression Analysis.....	28
3.6	Comparative Pseudogene Analysis of Predicted Pseudogenes	29

3.7	Functional Enrichment Analysis	29
3.8	Positive Selection Analysis	30
3.9	Functional Impact of Substitutions in Positive Selected Genes	30
3.10	Gene Family Expansion and Contraction Analysis	30
3.11	Development of Pangolin Genome Database (PGD) Hub	31
3.12	Pangolin Genome Browser in The PGD Database	33
 CHAPTER 4: RESULTS FOR COMPARATIVE ANALYSIS OF <i>MANIS</i>		
<i>JAVANICA</i> AND <i>MANIS PENTADACTYLA</i>		34
4.1	Identification of Pseudogenes	34
4.2	Unique Traits and Adaptations of Pangolins	34
4.2.1	Edentulous Characteristic of Pangolin Species	35
4.2.2	Poor Vision	36
4.2.3	Immunity of Pangolins	41
4.3	Comparative Analysis Among Pangolin and Mammals	45
4.3.1	Comparative Gene Family Clustering	45
4.3.2	Gene Family Expansion and Contraction Analysis	48
4.3.3	Positive Selection Analysis	54

CHAPTER 5: DESIGN AND DEVELOPMENT OF PANGOLIN GENOME

DATABASE (PGD)	65
5.1 Introduction To Pangolin Genome Database (PGD).....	65
5.2 Database Sources	65
5.3 Genome Data and Functional Annotation	66
5.4 Transcriptome Data	67
5.5 Pseudogene Annotation Data.....	68
5.6 Database Graphical User Interface (GUI) Design.....	69
5.6.1 Database Overview	69
5.6.2 Data Browsing Interface	70
5.6.3 Keyword and Sequence Search.....	71
5.7 Multiple Mammalian Genome Sequence Alignments.....	72
5.8 Pangolin Genome Data Visualization (JBrowse)	74
5.8.1 Genes And Annotations Tracks	75
5.8.2 Repeat Tracks.....	75
5.8.3 Transcriptome Tracks.....	75
5.8.4 Transcript Read Mapping Tracks.....	76
5.8.5 Transcriptome Profile Tracks.....	76

5.9	Assembly Hub for UCSC Genome Browser	77
5.10	Data Download	78
CHAPTER 6: DISCUSSION		80
6.1	Unique Traits of Pangolins and Their Genetics.....	80
6.2	The Presumable Loss of <i>IFNE</i> Gene Function in Pangolins	83
6.3	Possible Relationship Between <i>IFNE</i> Pseudogenization and Scales Formation ...	85
6.4	Unique Adaptations and Behaviors of Pangolins	86
6.5	Pangolin Genome Database (PGD) Development.....	87
6.6	Potential Impact and Future Works	88
CHAPTER 7: CONCLUSION.....		89
References		90
List of Publications and Papers Presented		105
List of Publications and Papers Presented		106
Appendix		108

LIST OF FIGURES

Figure 2.1: Taxonomy tree of Pholidota order.....	5
Figure 2.2: Phylogeny of Pholidota based on osteological characteristics.	6
Figure 2.3: Distribution of four pangolin species in Asia (<i>M. javanica</i> , <i>M. pentadactyla</i> , <i>M. crassicaudata</i> and <i>M. culionensis</i>) and Africa (<i>M. tricuspis</i> , <i>M.</i> <i>tetradactyla</i> , <i>M. gigantean</i> and <i>M. temminckii</i>).....	8
Figure 2.4: The Malayan pangolin.	9
Figure 2.5: The Chinese pangolin.	11
Figure 2.6: Skull of <i>M. gigantea</i> , The Giant Pangolin.....	12
Figure 2.7: Pangolin scales aligned and overlapped each other accordingly.....	15
Figure 2.8: Diagram of pseudogene classification.	20
Figure 3.1: PGD 4-tier web application architecture.	33
Figure 4.1: Case studies of pseudogenised genes in pangolins.....	37
Figure 4.2: Sanger validation of <i>AMBN</i> genes in African pangolin species.....	38
Figure 4.3: Sanger validation of <i>AMELX</i> genes in African pangolin species.....	39
Figure 4.4: Sanger validation of <i>GUCA1C</i> genes in African pangolin species.	40
Figure 4.5: Multiple sequence alignment of all mammalian <i>IFNE</i> genes.	43
Figure 4.6: Sanger validation of <i>IFNE</i> genes in African pangolin species.....	44
Figure 4.7: Comparison between pangolin and human (reference) <i>IFNE</i> protein sequences.	45
Figure 4.8: Venn diagram showing the unique and shared gene families among pangolins and their closest relatives (cat, dog, and giant panda).	46
Figure 4.9: Heatmap shows the expression of pangolin-specific genes across different pangolin organs, which represented by FPKM values.	47
Figure 4.10: GO enrichment analysis of 1,152 pangolin-specific genes.	48
Figure 4.11: Phylogenetic tree and gene family expansion and contraction.....	49
Figure 4.12: Functional enrichment of 147 significantly expanded gene families.	50

Figure 4.13: Phylogenetic tree shows the significant contraction of the interferon gene family in pangolins.	51
Figure 4.14: Expansion of the Septin gene family in the pangolin lineage.	52
Figure 4.15: Expansion of olfactory gene family 2 in the pangolin lineage.	53
Figure 4.16: Evolution in the immunity-related pathways.	56
Figure 4.17: Evolution in the immunity-related pathways.	58
Figure 4.18: Evolution in the immunity-related genes.....	60
Figure 4.19: Critical pangolin-specific amino acid changes were detected in hair/scale-related keratin proteins.....	61
Figure 4.20: The keratin-related proteins in African pangolin species.....	62
Figure 4.21: Positively selected LTF.	64
Figure 5.1: Schematic structure of the PGD.	70
Figure 5.2: A screenshot of gene details page.....	71
Figure 5.3: The phylogenetic tree of species involved into the structural alignment.	74
Figure 5.4: Pangolin genome browser.	77
Figure 5.5: Interface for UCSC Genome Browser (<i>M. javanica</i> genome).	78
Figure 5.6: Web interfaces for data download in PGD.....	79

LIST OF TABLES

Table 3.1: Primer sequences used for the validation of gene mutations.....	28
Table 4.1: Pseudogene identification across different mammalian species.....	34
Table 4.2: Interferon family gene number across the different animals.	42
Table 4.3: List of the selected genes under positive selection in pangolins with their assigned functions or pathways.....	55
Table 4.4: Positively-selected genes associated with known diseases.....	64
Table 5.1: Summary statistics of two pangolin genomes and transcriptome assembly datasets in PGD.	69
Table 5.2: Genome assembly version for each mammalian genome used for multiple sequence alignment.	73

LIST OF SYMBOLS AND ABBREVIATIONS

AJAX	:	Asynchronous JavaScript and XML
BLAST	:	Basic local alignment search tool
BLAT	:	Basic local alignment tool
cDNA	:	Complementary Deoxyribonucleic acid
dN/dS	:	Non-synonymous over synonymous rate ratio
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxynucleotides triphosphates
DWNP	:	Department of Wildlife and National Parks
EMP	:	Extracellular matrix protein
EN	:	Endangered Species
FDR	:	False Discovery Rate
FPKM	:	Fragments Per Kilobase of transcript per Million mapped reads
FTP	:	File Transfer Protocol
GIRG	:	Genome Informatics Research Group
GO	:	Gene ontology
Indels	:	Insertion or deletion of bases
IPaRC	:	International Pangolin Research Consortium
IUCAC	:	International Union for Conservation of Nature
MRCA	:	Most recent common ancestor
MSA	:	Multiple Sequence Alignment
MVC framework	:	Model-view-controller framework
MySQL	:	My Software Query Language
NCBI	:	National Center for Biotechnology Information

NGS	:	Next-generation Sequencing
OR	:	Olfactory Receptor
PAML	:	Phylogenetic Analysis by Maximum Likelihood
PCR	:	Polymerase Chain Reaction
PE	:	Pair-end
PGD	:	Pangolin Genome Database
PHP	:	Hypertext Preprocessor
PROVEAN	:	Protein Variation Effect Analyzer
RNA	:	Ribonucleic acid

LIST OF APPENDICES

Appendix A: Summary of selected genes for gene traits analysis.	108
Appendix B: List of genes under positive selection.	112
Appendix C: <i>ENAM</i> alignment of pangolin species with mammals.....	125
Appendix D: <i>AMELX</i> alignment of pangolin species with mammals.	125
Appendix E: <i>AMBN</i> alignment of pangolin species with mammals.....	139
Appendix F: <i>BFSP2</i> alignment of pangolin species with mammals.....	135
Appendix G: <i>GUCA1C</i> alignment of pangolin species with mammals.....	146
Appendix H: <i>IFNE</i> alignment with 72 mammals and African pangolin species.....	150
Appendix I: Gene distribution of Olfactory Receptor (OR) gene families.....	172
Appendix J: PCR assays to detect the presence of <i>Burkholderia fungorum</i>	173

CHAPTER 1: INTRODUCTION

1.1 Overview

Pangolins (also known as scaly anteaters) are mammals of the order Pholidota. Their name derived from the Malay word "pengguling", meaning "something that rolls up." Pangolins are ancient creatures whose ancestors are thought to be members of a suborder of Palaeanodonta from some 60 million years ago (Kondrashov & Agadjanian, 2012). Eight pangolin species are recognized, four from Asia (*Manis javanica*, *M. pentadactyla*, *M. crassicaudata* and *M. culionensis*) and four from Africa (*M. tricuspis*, *M. tetradactyla*, *M. gigantea* *M. temminckii*) (Timothy J. Gaudin, 2009). Although pangolin morphology is similar to the South-American anteaters and armadillos (superorder Xenarthra), they are phylogenetically distinct with the Pholidota order and Carnivora order being a sister group within the superorder Laurasiatheria (Murphy *et al.*, 2001).

Pangolins are rare, toothless nocturnal burrowing mammals except for Black-bellied pangolin (*M. tetradactyla*) which regularly active during the day. Distinct from other placental mammals, pangolin skins are covered by broad and overlapping keratinized scales (S. C. Chin *et al.*, 2015; Meyer *et al.*, 2013). These hard scales adopted as a protective layer; when threatened, pangolins quickly roll up into a tight ball, and the scales act as a shield (Kawashima *et al.*, 2015). The selective forces underlining the origin of this unique mammalian trait remains a mystery, although observations of the eight modern pangolin species suggest defensive armor function against predators. They all have hard keratinous scales covered their back by having the similar protection strategy, roll into tight ball and isolate predators by the strong keratinous scales barrier. Furthermore, pangolins are edentulous, eating mostly ants and termites (Lin *et al.*, 2015) captured using long and muscular tongue and had a well-developed muscular system for fossorial or arboreality and a remarkable olfactory system. Pangolins are the most

poached and trafficked mammal in the world due to the huge demand for their meat as a delicacy and their scales and blood for use in traditional medicines (Luo *et al.*, 2013; Zhou *et al.*, 2014). Among all pangolin species, the Malayan pangolin (*M. javanica*) and the Chinese pangolin (*M. pentadactyla*) are classified as critically endangered by the IUCN Red List of Threatened Species (D. Challender, Nguyen The Truong A, Nguyen Van Q, Chung Y, 2014) because of deforestation and hunting. The Malayan pangolin is mainly found in Southeast Asia, whereas the Chinese pangolins mostly found in China, Taiwan and in some Northern of Southeast Asia.

The Genome Informatics Research Group at the University of Malaya has recently sequenced the genomes of two pangolin species, the Malayan pangolin and the Chinese pangolin from Malaysia and Taiwan, respectively, using high-throughput Next-Generation Sequencing (NGS) approaches. Using the NGS technology, the group has also sequenced the transcriptomes of different organs of the Malayan pangolin such as heart, cerebrum, cerebellum, liver, and kidney.

In this study, I have presented the first comparative genome analyses of these two unique pangolin species of the Pholidota, providing better understanding of mammalian genome evolution and fundamental knowledge for further research in pangolin biology and conservation. To better understanding the biology of pangolins, particularly their unique traits and evolution, I compared and analyzed both genomes (as well as other mammalian species) at their genomic level in an attempt to link their genomics/genetics with pangolin traits such as toothless, poor vision, strong sense of smells and well-established muscular system, etc. To further facilitate research into pangolin biology, I have also designed and developed a Pangolin Genome Database (PGD), aiming to be a central hub for hosting genomic resources and information of pangolins, together with useful analysis tools for the research community. PGD provides interactive and user-friendly web interfaces for

users to search, browse, retrieve, visualize and analyze pangolin genomic and transcriptomic data and information. I anticipate that the PGD will serve as a computational platform for researchers to advance research in this field, particularly in the area of conservation of this unique endangered mammalian species and also to provide a useful resource for research in mammalian evolution and biology, in general.

1.2 Field of Research

Bioinformatics

1.3 Project Objectives

The following are the main objectives of this project:

- To perform comparative analyses of pangolins genomes
- To understand relationships between pangolins' unique gene traits with their genomics/genetics
- To design and develop a pangolin genomic resource for research communities

CHAPTER 2: LITERATURE REVIEW

2.1 Pholidota order (Pangolin)

The Pholidota order comprising of only eight living species of pangolins is one of the smallest extant placental mammal orders (Gaubert & Antunes, 2005). Belonging to the Pholidota, all eight species of extant pangolins classified within the single family Manidae. Four pangolin species are from Asia (*M. javanica*, *M. pentadactyla*, *M. crassicaudata* and *M. culionensis*), whereas the other four are from Africa (*M. tricuspis*, *M. tetradactyla*, *M. gigantean* and *M. temminckii*) (Timothy J. Gaudin, 2009). The different pangolins range in size from 30cm to 100cm and vary in size from 1.6kg to a maximum of 33kg (Abayomi *et al.*, 2009). They are distinctive by the color of their scales from light to yellowish brown through dark olive brown. At rest, the tongue will be kept into a sheath in its chest cavity. Pangolins have muscular stomach and keratinized spines projecting into its interior. They have strong limbs and well-adapted for digging with sharp claws. Interestingly, although pangolin morphology shares some analogies with South-American anteaters and armadillos (superorder Xenarthra), they are phylogenetically distinct compared to the Pholidota order and the Carnivora order being a sister group (Ferae) within the superorder Laurasiatheria (Figure 2.1) (Murphy *et al.*, 2001).

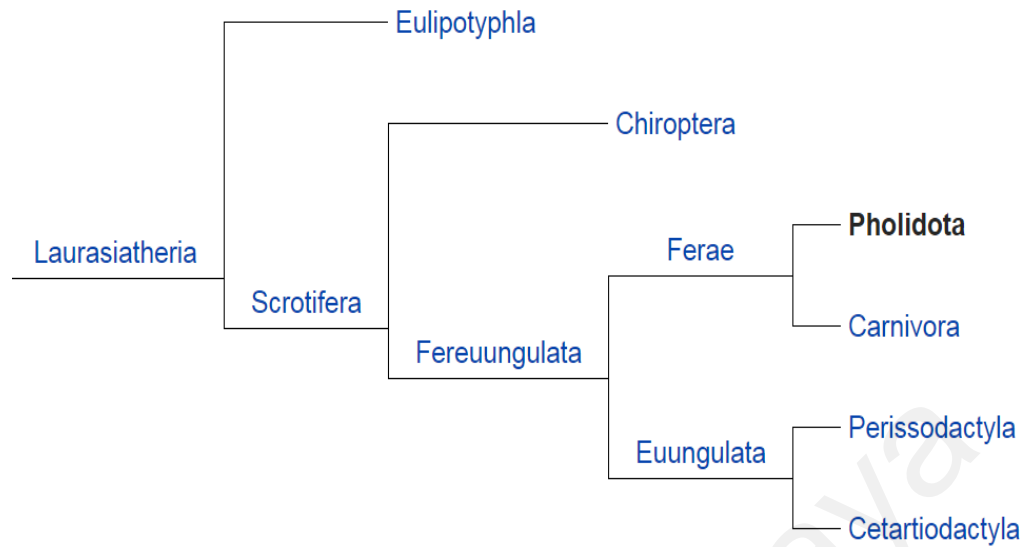


Figure 2.1: Taxonomy tree of Pholidota order.

Pangolins are also believed to be one of the ancient creatures whose ancestors are thought to be members of a suborder Palaenodonta from some 60 million years ago (Kondrashov & Agadjanian, 2012). Figure 2.2 shows the phylogenetic tree based on osteological characteristics. The Pholidota order is very near to the Pholidota-morph order to which all the extinct Palaenondonta species belong.

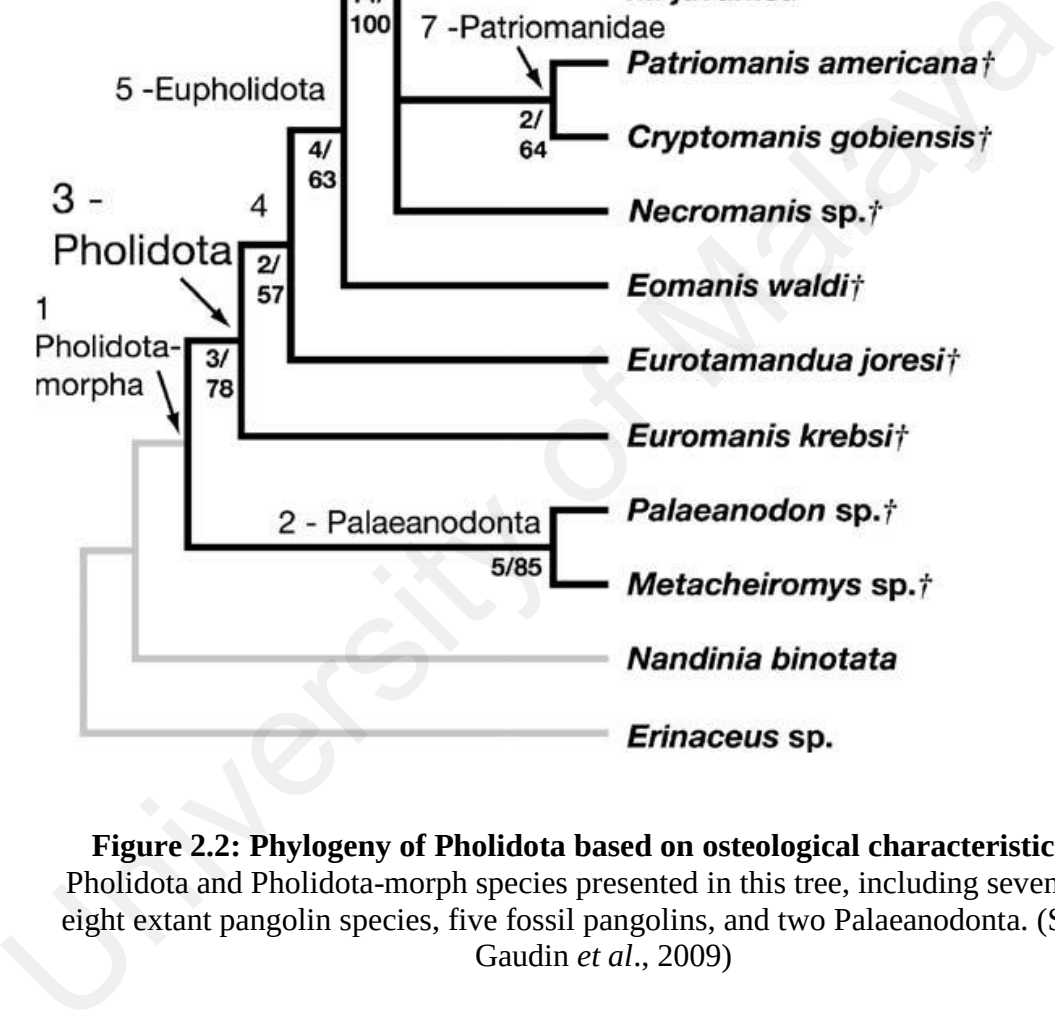


Figure 2.2: Phylogeny of Pholidota based on osteological characteristics. 17
Pholidota and Pholidota-morph species presented in this tree, including seven of the
eight extant pangolin species, five fossil pangolins, and two Palaeonodonta. (Source:
Gaudin *et al.*, 2009)

Unlike other placental mammals, pangolin skins are covered and protected by large and overlapping keratinized scales (Meyer *et al.*, 2013). The functional reason for these unique mammalian scales remains a mystery. Pangolins protect themselves by rolling up into a sphere and use their hard body armor to protect from predators. Furthermore, pangolins are edentulous hence they will ingest small stones for grinding up their food in the stomach. They are insectivorous and feed mostly on ants and termites captured using

their exceptionally long and muscular tongue covered with sticky mucous secreted enabling pangolins to remove ants and termites from their nests. They can consume 140g to 200g of insects per day on average (Grosshuesch, 2012). Pangolins have a rather weak sense of vision and rely on their smell and hearing for predating their food and protecting themselves from predators. They have a pair of sharp front claws that are powerful enough to tear open anthills and have long tails to counterbalance their long bodies during walking and climbing (Durojaye A Soewu & Ibukun A Ayodele, 2009).

Unfortunately, all modern pangolin species around the world are under threat, from being vulnerably threatened (African pangolin) to critically endangered (e.g. Malayan and Chinese pangolins) (Figure 2.3). They are the most poached and trafficked mammal in the world due to a huge demand for their meat as a delicacy and their scales for traditional medicines (Guo *et al.*, 1997). The Malayan and Chinese pangolins are the two most critically endangered species, and their population numbers are still in decline; the complete extinction of these two pangolin species may be just a matter of time if conservation efforts are unsuccessful (D. Challender, Nguyen The Truong A, Nguyen Van Q, Chung Y, 2014).

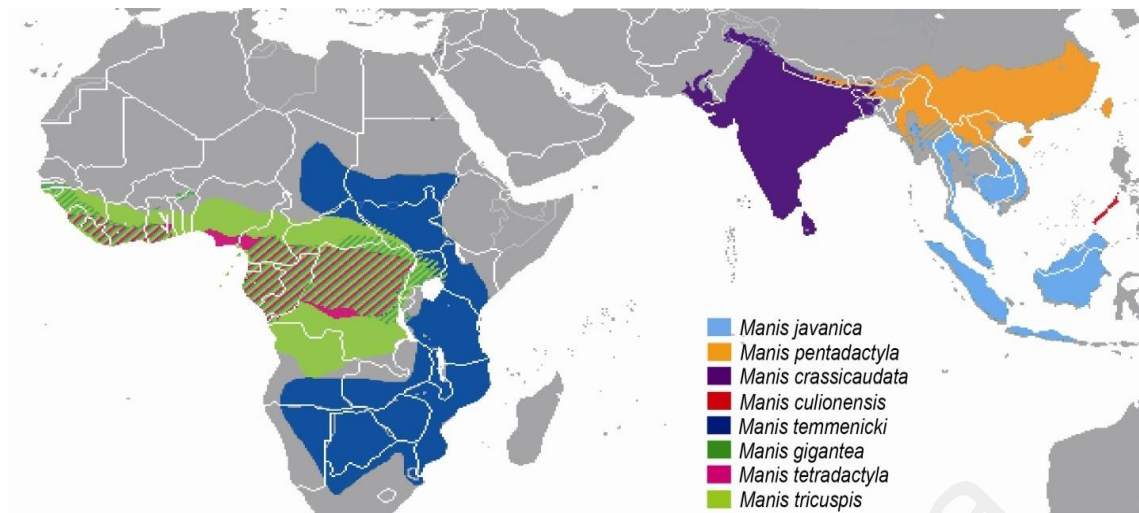


Figure 2.3: Distribution of four pangolin species in Asia (*M. javanica*, *M. pentadactyla*, *M. crassicaudata* and *M. culionensis*) and Africa (*M. tricuspis*, *M. tetradactyla*, *M. gigantea* and *M. temminckii*) (Source: whatspecies.com/blog/world-pangolin-day-2015)

2.1.1 *Manis javanica* (Malayan Pangolin)

Manis javanica named by Desmarest at 1822 and it is classified under the Mammalia class, Pholidota order and Manidae family. It is also known as the Malayan pangolin or Sunda Pangolin (Figure 2.4), classified as an Endangered (EN) species in the Red List of Threatened Species. The Malayan Pangolin is mostly found in tropical or subtropical forests in mainland South East Asia from the south of Myanmar through the central and southern part of Laos, Thailand, Vietnam, Cambodia, Peninsular Malaysia, Sumatra, Java, and Borneo (D. Challender, Nguyen The Truong A, Nguyen Van Q, Chung Y, 2014; D. W. Challender *et al.*, 2012). In Malaysia, the species is distributed primarily in the forests (as well as gardens and rubber plantations) of Peninsular Malaysia and Borneo (Medway, 1969).

Malayan pangolins have a small conical face and deformed edentulous jaw, and has an incredibly long, muscular and sticky tongue that specially adapted for reaching and gathering termites from deep cavities. They have poor eyesight but a strong sense of smelling (Thapa, 2014). Malayan pangolins are more arboreal animal compares to Chinese pangolins, there are adept climbers with prehensile tails. They often climb trees

to access ant nests in trees. Malayan pangolins are thought to breed in the autumn and give birth in winter with a gestation period of estimated 130 days. Usually, they will be one or two offspring reproduced. The population of Malayan pangolin remains unknown due to their secretively nocturnal habits. Unfortunately, the number of Malayan pangolins in the wild is declining for several reasons, and the species is characterized as being critically endangered in the International Union for Conservation of Nature and Natural Resources (IUCN) Red List of Threatened Species (2015, 2015; D. Challender, Nguyen The Truong A, Nguyen Van Q, Chung Y, 2014). One of the major threats for its declining numbers is the rapid loss and deterioration of their natural habitat due to deforestation activities and human agricultural expansion(D. Challender, Nguyen The Truong A, Nguyen Van Q, Chung Y, 2014; CITES, 2000). Malayan pangolins (or pangolins in general) are also heavily hunted for their meat, skin, and scales, as illegal trade in live animals has become a severe threat to pangolins(D. W. S. Challender, 2011; Pantel & Anak, 2010; Pantel & Chin, 2009).



Figure 2.4: The Malayan pangolin.

In China, pangolin meat is consumed as an exotic delicacy while the scales are used for traditional medicinal purposes such as skin diseases and cancer remedies, among other

illnesses. There has also been an attempt to relocate and breed the species in captivity while imitating its natural habitat, but with little success as pangolins do not survive and breed well in captivity (Ci Wen Yang *et al.*, 2007).

2.1.2 *Manis pentadactyla* (Chinese Pangolin)

Chinese pangolin is another species in the Pholidota order and bears a striking resemblance to the Malayan Pangolin (Figure 2.5). They are found in Nepal, Bhutan, northern India, Bangladesh, Myanmar, Vietnam, Thailand, China, and Taiwan. The IUCN Red List reported the number of Chinese pangolins has declined considerably over the past fifteen years (Duckworth, 2014).

The Chinese pangolins have slightly longer head and body measure compared to Malayan pangolins, which is around 60 cm and its tail measures about 20 cm (Heath & Vanderlip, 1988). A mature Chinese pangolin weighs 2-7 kg. They have small and narrow mouths, very long and sticky tongue. They usually reproduce in April and May when the weather is warm. Chinese pangolins never found on trees and they are more terrestrial than the arboreal Malayan pangolin (Newton *et al.*, 2008). However, they are fully capable of climbing trees and swims well like other pangolins (Heath & Vanderlip, 1988).

Like the Malayan pangolins, Chinese pangolins also consume on ants and termites. It has been reported that there is a close correlation between Chinese pangolins distribution and the distribution of two termite species which are *Coptotermes formosanus* and *Termes formosanus* suggest this two termite species a major food source for Chinese pangolin (Heath & Vanderlip, 1988).

Chinese pangolins are also considered a delicacy especially in Viet Nam and Hong Kong. They hunted in broad scale for human consumption and the scales and blood are in vast demand for their alleged medical functions (Guo *et al.*, 1997). As a result of ongoing

exploitation, over the time frame of three generations which approximate 21 years (The generation length of Chinese pangolin is seven years (Duckworth, 2014), it has assumed that Chinese pangolins are commercially extinct in China (S. Wu *et al.*, 2004).



Figure 2.5: The Chinese pangolin.

2.2 Conservation efforts on Pangolins

Many efforts have been carried out to conserve or maintain pangolins (Heath & Vanderlip, 1988), but most of these efforts failed likely due to lack of understanding of the biology and genetics of pangolins. For instance, there were more than 100 zoos or organizations have attempted to sustain pangolins (Hua *et al.*, 2015), but most captive pangolins died within a few months to years, probably due to pangolin specialized food source, poor adaptability, and weak immunity (Hua *et al.*, 2015. Although there are cases of newborn of captive pangolins, it is still very challenging to maintain pangolins in captivity (C. W. Yang *et al.*, 2007).

2.3 Unique pangolin traits

Pangolins are mammals that have evolved with different unique traits possibly due to selection pressure and living habits. Here I will discuss some unique adaptations of pangolins.

2.3.1 Edentulous

Pangolins are edentulous and feed off termites and ants (Redford, 1987; Wilson, 1994). Figure 2.6 shows the specialized, tubular smooth skull formation of a Giant pangolin. Pangolin skulls are made of dense bone that helps to protect against predators. Pangolins zygomatic arch structure is incomplete and has an extremely reduced mandible. The narrow lower horizontal jaw lacks chewing muscles and most importantly, lack of proper dentition development. As a result, they use their strong muscular sticky tongue to hunt for their prey, ants, and termites.



Figure 2.6: Skull of *M. gigantea*, The Giant Pangolin. (Source: Natural History Collections <http://www.nhc.ed.ac.uk/index.php?page=493.496>)

Tooth formation is a complicated process that includes complex series of epithelial-mesenchymal interactions involving several hundred genes associated with tooth development (Thesleff, 2006). Previous studies have suggested that edentulous

characteristics are due to molecular decay of extracellular matrix protein (EMP) genes (R. W. Meredith *et al.*, 2009; Robert W Meredith *et al.*, 2013). Mammals usually secrete multiple extracellular matrix proteins in ameloblast, enamelin (*ENAM*) which is the largest EMP in mammals. This protein, together with amelogenin (*AMELX*) and ameloblastin (*AMBN*), directs the formation of hydroxyapatite crystals during the secretory phase of enamel formation. These EMP genes mutations are known to cause defects in enamel known as amelogenesis imperfecta (Wright *et al.*, 2008).

2.3.2 Poor vision and hearing

Visual perception plays a major role in the animal kingdom, as vision is necessary for foraging, mate recognition, and predator avoidance. However, certain nocturnal or subterranean animals have impaired visual acuity and rely on other sensory organs. All existing pangolins species are known to have poor eyesight and poor hearing capability likely due to pangolins nocturnal behavior and structural morphology of their head, with small eyes and no external ears resulting in poor hearing and vision, similar to the situation in bats.

Nonfunctional visual genes usually associated with animals that inhabit low light environments such as nocturnal, aquatic and subterranean animals. A previous study suggested that such genes lost function through relaxed selection acting on the visual system (Zhao *et al.*, 2009).

2.3.3 Strong sense of smell

Pangolins have strong olfaction, probably to counterbalance their poor vision and hearing (Mohapatra & Panda, 2014) . For instance, pangolins may need remarkable olfactory ability that aids in the location of food and mates, and to monitor their environment for predator detection (Sweeney, 1956). Pangolins are like many mammals that are driven by

scent, borne out by the large olfactory bulb of the brain and they possess large olfactory turbinate that enhanced the sense of smell.

Olfactory communications play a major role in pangolins' behavior (Rylander, 2015). Strong scent produced from paired anal glands of pangolin, together with feces and urine that deposited along trails and trees. There are similar to skunks that eject unpleasant smelling secretion from the anal glands (Feldhamer *et al.*, 2007).

2.3.4 Poor immunity

It is believed that poor immunity in pangolins could be one of the reasons why pangolins are hard to maintain in captivity (Hua *et al.*, 2015). For instance, it has been reported that skin diseases, gastrointestinal diseases, and parasites are the common causes of death for captive pangolins (Chang, 2004; Clark L, 2008; Lihua *et al.*, 2010). Besides that, over half of the captive pangolin death cases are reported with pneumonia and hemorrhagic gastric ulcers due to captivity stress and failure artificial diet (S.-C. Chin & Yang, 2009).

2.3.5 Pangolin scales

Pangolins are the only animals in the world that have scales covering most of their body (Durojaye A Soewu & Ibukun A Ayodele, 2009). Their scales are famous for their alleged capability in improving blood circulation and in tumor treatment (Yu, 1992). Pangolin scales have been traded all over the world especially in China which is the largest importing country (Rosen & Smith, 2010) because pangolin scales are used in traditional Chinese medicine to treat some diseases (Shepherd, 2009).

Since pangolins have no teeth and move slowly, the scales may play important role in protecting pangolins against predators (W. Yang *et al.*, 2013). Their scales aligned and overlapped each other accordingly, providing an extra protective barrier (Figure 2.7). Interestingly, they are different to fish scales that are made of connective tissue covered

with calcium, nor the armadillo scales-like structure which made of skin, pangolin scales are made up of keratin filament structures similar to human hair and nails (W. Yang *et al.*, 2013). These scales are movable and made up of hairs that have been fused together resulting in durable and hard structures. Pangolin scales are dense and contribute up to 20% of pangolin's body weight. Other than passive protection, these scales are sharp and pangolins can swing their long tails to attack their enemy.



Figure 2.7: Pangolin scales aligned and overlapped each other accordingly. The alignment of the scales serves as an extra protective barrier to protect pangolin from physical injury. (Source: The Tikki Hywood Thrust)

2.4 Sequencing study of pangolin genomes

Genetic studies of endangered species have become widespread during the last several decades but more recently, the genomes and transcriptomes of endangered species have been sequenced (Cho *et al.*, 2013; Dobrynin *et al.*, 2015; R. Li *et al.*, 2010; Liu *et al.*, 2013) (Ge *et al.*, 2013). With the emergence of high-throughput Next-Generation Sequencing (NGS) technologies, more molecular information about the endangered species can be acquired and studied in-depth. The endangered pangolins have raised our concern to the conservation effort on these animals. To the best of our knowledge, the

Malayan pangolin and Chinese pangolin genome analysis reported in this thesis is the first Pholidota genome ever sequenced. By further understanding the genomics and genetics of the pangolins, it is possible that it would enhance the conservation of the pangolins. Moreover, pangolin genomes research will be a valuable resource for the detailed study of this species that may assist the future management and conservation of this critically endangered mammal.

2.5 Comparative genomic analyses.

Comparative genomics analysis is a powerful approach to compare multiple organisms. Comparative genomics virtually started at 1995 when the first two organisms' genome available (*Haemophilus influenzae* and *Mycoplasma genitalium*) and later this approach has become a standard component of every new genome sequenced (Fleischmann *et al.*, 1995).

The key principle for the comparative genomics is the common genomic features of two or more usually phylogenetically similar species genome will often be encoded within the DNA that is evolutionarily conserved between these species (Hardison, 2003). Therefore, in normal circumstances, the comparative genomics start with multiple sequence alignment of genome sequences, look for orthologous sequences in the aligned genomes and analyze to what extent these sequences are conserved. Based on these comparisons, the phenotypic evolution of different species could be identified.

Comparative genomics analysis is able to disclose high levels of similarities between closely related animals, such as human and chimpanzee (Hazkani-Covo & Graur, 2007); lion, tiger and snow leopard (Cho *et al.*, 2013), and amazingly, between species that are phenotypically distinct such as human and yeast genomes (Tolstorukov *et al.*, 2009). The point is the sequences that stay conserved across similar or distant species, is likely to be constrained due to evolutionary pressures and have a similar biological functions (Alfoldi

& Lindblad-Toh, 2013). A previous study that compared human genome with 29 other placental mammalian genomes have successfully identified 3.6 million constrained elements within the human genome (Lindblad-Toh *et al.*, 2011).

In addition, the comparative genomics approach has successfully been used to reveal the relationships between the unique traits and the genetics/genomic of some organisms. For instance, by comparing the genome of the white lion and the genomes of lion, tiger and other big cat, Yun Sung Cho and colleagues have revealed a potential link between the mutations in the Tyrosinase (*TYR*) gene and albinism in the white lion (Cho *et al.*, 2013). As Xiao Xu *et al.* has reported that the genetic basis of the white tiger with black stripe revealed as an evolutionary change in the transporter protein SLC45A2 by comparing with other different tiger genomes (Xu *et al.*, 2013). Moreover, comparative analysis of black flying fox (*Pteropus alecto*) and the bat (*M. davidii*) genomes has also revealed the evidence of genetic changes that might contribute to the bat immunity that can serve as the reservoir for multi-viruses (G. J. Zhang *et al.*, 2013). A previous comparative genomics study using differentially expressed genes revealed a similar mechanism of genetic control among distantly related species, suggesting genetic phenotyping may be affected by changes in gene regulation (Villanueva-Cañás *et al.*, 2014).

2.6 Pseudogene analysis

The word 'Pseudogene' was first describe as a DNA sequence resemblance to gene coding for RNA that contained multiple mutation that rendered the RNA product nonfunctional (Jacq *et al.*, 1977). Currently, 'pseudogene' was used to describe all DNA sequence that similar to normal gene but lost of their gene expression or ability to code protein due to function disability or structural modification (Vanin, 1984). The structural modification that results in pseudogenes involves multiple kinds of alterations, which include deletion, insertion and point mutation. Pseudogenes are primarily thought of as

junk DNA that is to be removed from the genome and the last stop for genomic materials (Balakirev & Ayala, 2003; Zheng *et al.*, 2007). Moreover, pseudogenes may contain evolutionary and biological histories within their sequences and are well known as genomic fossils because they share common ancestry with its associated functional gene copy that had diverged as separate genetic entities over millions of years mainly due to mutation (Harrison *et al.*, 2002). There are three distinct classes of pseudogenes namely “Unitary” pseudogene, “Processed” pseudogene and “Duplicated” pseudogene (Figure 2.8). The unitary pseudogenes often refer to unprocessed pseudogenes that lack functional counterparts in genomes (Z. D. Zhang *et al.*, 2010). Besides that, they usually have single gene copies in the genomes and mutations induced into the genes, resulted in gene deactivation and loss of gene function before duplication. For instance, a previous study has linked the deactivation of caspase twelve genes to positive selection in humans, suggesting that the positive selection mechanism may act to counter-balance the loss of certain gene function (Xue *et al.*, 2006). The unitary pseudogenes would not affect the organism’s lifestyle as they typically do not depend on that particular gene. However, the pseudogenised of single copy gene may indicate the potential permanent loss of certain gene functions that might relate to the host phenotype and even resulting in certain single gene disorder (R. W. Meredith *et al.*, 2011; F. J. Smith *et al.*, 2006).

The other category of pseudogenes known as processed pseudogenes or retrotransposed pseudogenes. Processed pseudogenes occur when the reverse transcriptases generate complementary DNA (cDNA) backward from the RNA sequences (Z. Zhang *et al.*, 2003). The newly reverse transcribed cDNA gets inserted back into the genome and forms a processed pseudogene. The processed pseudogenes resemble the mRNA transcript sequence but presumably lose function due to the lack of the promoter region that initiates transcription. Host machinery may modify the RNA sequences after transcription including removing intron regions by splicing, alternately spliced exons and additional

poly “A” tail at the end of the RNA sequences. All these modifications will be reinserted into the host genome as part of the processed pseudogene (Vanin, 1985). The mechanism of this cDNA reinsertion into the host genome refers to as retro-transposition mechanism. Retro-transposition is a common event that severely impacts upon genome composition (Cordaux & Batzer, 2009). For example, approximately 42% of the human genome and almost half the genomes of mammals (45%-48%) are made of transposable elements (Lander *et al.*, 2001).

Duplicated pseudogenes are also known as “non-processed pseudogenes,” are another major category of pseudogenes. Duplicated pseudogenes are normally the largest pseudogene class among mammals arising when genes undergo duplication in the genome (Rana, 2008). The duplicated genes contain all the same characteristics as their parent genes, including intact exon-intron structure and even promoter sequences. However, the duplicated gene copies may experience irreversible mutations that render the loss of function of the duplicated genes (hereafter referred as duplicated pseudogene). The loss of function for duplicated pseudogenes are minimal as actual intact functional copies exist (Rana, 2008) (Max, 2003; Ploos van Amstel *et al.*, 1990).

In the current bioinformatics research, the pseudogene screening and identification is not as challenging as before (Rouchka & Cha, 2009; Zheng & Gerstein, 2006). Although pseudogenes are similar regarding gene sequences and structures to their parent genes, molecular biologists can classify pseudogenes based on two criteria: (1) gene sequence similarity to a known gene, and (2) mutational event that results in a non-functional pseudogene. *In silico* pseudogene screening and annotation in human, animal and plant genomes are consistently updated by Deyou Zheng, Baikang Pei and Mark Gerstein group from Yale University using their pipeline, Pseudopipe and GENCODE resources (Pei *et al.*, 2012; Z. Zhang *et al.*, 2006).

Pseudogene study is also one of the useful approaches to understand animal molecular evolutions (Balakirev & Ayala, 2003; Bensasson *et al.*, 2001). Pseudogenes record changes in genomic DNA, which can be used as models for determining underlying rates of nucleotide substitution, insertion, and deletion in the animal genomes. Comparative pseudogene analyses across animals can provide vital information on evolution, the rate of divergence and speciation of different animals. Gene loss through pseudogenization is increasingly recognized as an important factor in lineage and adaptive phenotypic diversification (Albalat & Canestro, 2016; Wang *et al.*, 2006).

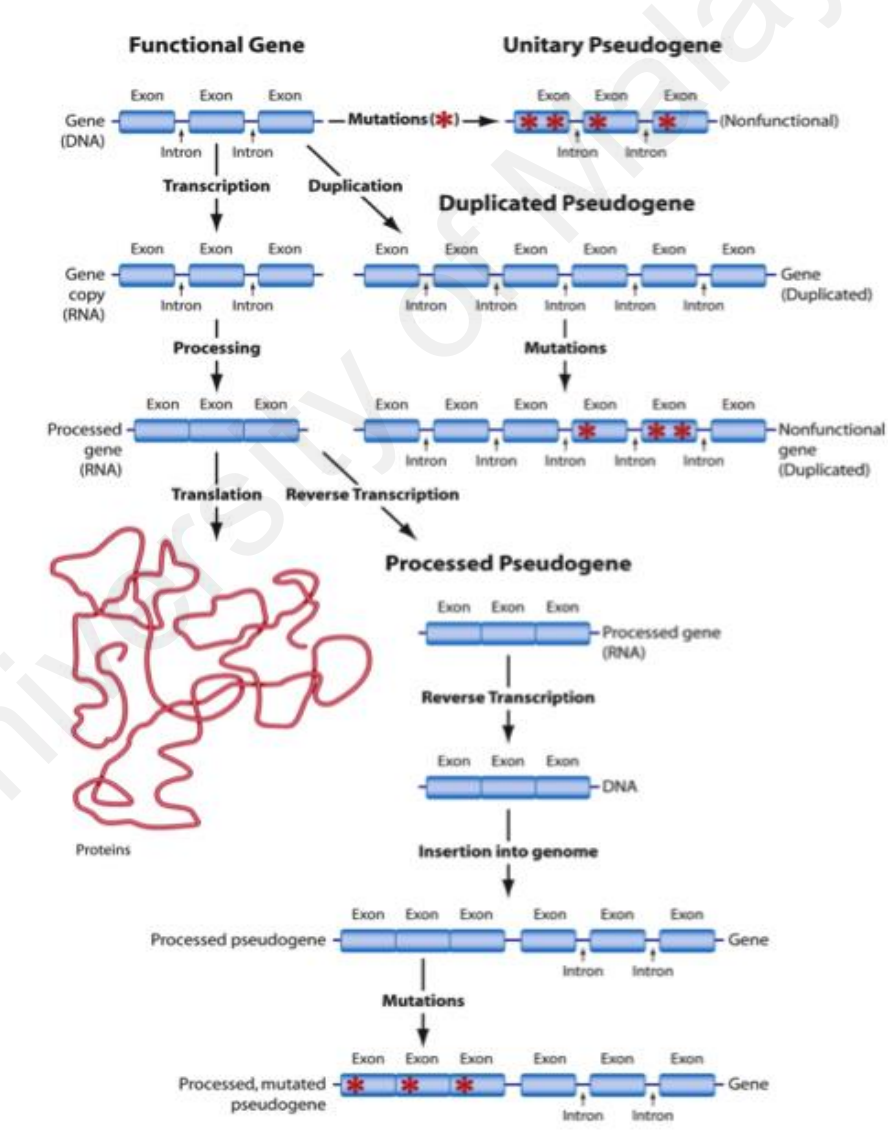


Figure 2.8: Diagram of pseudogene classification. Three distinct categories of pseudogenes “Unitary pseudogenes”, “Processed pseudogenes” and “Duplicated pseudogenes” formation described in the diagram. (Source: www.reasons.org)

2.7 Gene gain and loss analyses

Gene gain and loss are both crucially vital for genome evolution processes as gene families expand or contract due to this mechanism (Hahn *et al.*, 2007; Kulmuni *et al.*, 2013). Gene family is a set of gene members or gene groups that are acquired from a common ancestor within given time or sequence similarity threshold (Graur & Li, 2000). Gene families can be gained by horizontal transfer, gene duplications, and other evolutionary events, whereas the gene loss can be due to deletion, sequence evolution and pseudogenization (Kulmuni *et al.*, 2013).

Gene family gain and loss were initially assumed to be of equal rate, and the same rates applied to all gene families in the first phylogenetic model (Hao & Golding, 2006) that related to Jukes-Cantor model of nucleotide evolution (Nei & Gojobori, 1986). After that, the rate of gene gain and loss are allowed to be different, and the unequal gain and loss rates fit better to genome data (Cohen *et al.*, 2008). The mixture models developed by Cohen *et al.* (2008) and Hao and Golding (2008) using several categories of gene families, each having the same relative gain and loss rates, but different absolute rates, resulting in vast improvement in fit to genome data.

The evolutionary events of particular gene families usually relate to their gene gain and loss. A previous study from Yoshihito Niimura and Masatoshi Nei described the gene gain and loss for the Olfactory Receptor (OR) genes in mammalian evolution (Niimura & Nei, 2007). They reported that the evolution of OR gene number was exceptionally high in mammals, proposing that their OR genes repertoire are highly variable although the number of functional OR genes in several mammals is similar. By understanding the pangolin and mammalian evolution in the aspect of the gene gain and loss, it is possible for us to better understand the evolution of pangolin gene families related to their functions.

2.8 Positive selection analysis

Positive selection is the evolutionary process by a population where new advantageous genetic variants become fixed in the population (Mitchell-Olds *et al.*, 2007). The positive selection is also known as a directional selection or Darwinian selection, which is a mode of evolution where natural selection on particular phenotypes is favored over other phenotypes, resulting the shift of allelic frequency over time (Sabeti *et al.*, 2006). The positive selection pressure may even encourage the recessive allele to be fixed based on the environment requirement for population survival and better adaptations (Teshima & Przeworski, 2006).

Although it is highly desirable to have these advantageous mutations, they are few difficulties that need to be overcome because neutral and deleterious mutations predominate positive selection in their frequency (A. L. Hughes, 2008). Currently, there are two major approaches to detect positive selection: (1) population study method to analyze positive selection based on the nature and frequency of allele diversity within species, and (2) a codon analysis method by comparing patterns of synonymous and non-synonymous changes in protein-coding sequences (S. L. Chen *et al.*, 2006; Suzuki & Gojobori, 1999). By performing comparative genomics, the selective pressure in the protein-coding genes can be studied. The selective pressure is defined by the ratio of non-synonymous rate over the synonymous rate of amino acid changes, dN/dS (ω). In the absence of evolutionary pressure, the dN/dS ratio are equal, which is equal to one. Whereas under purifying selection or negative selection, natural selection prevents the replacement of amino acids. Therefore, the synonymous rate will be lower than non-synonymous rate, where dN/dS ratio is lower than one. For positive selection, the replacement of amino acid is favored by selection. Therefore, dN/dS is greater than one (Delpont *et al.*, 2009; J. Zhang *et al.*, 2005). In summary, purifying selection conserved

of existing phenotype, whereas positive selection promotes the emergences of new phenotype (Vallender & Lahn, 2004).

CodeML program from the PAML package based on Maximum Likelihood Model (Z. Yang, 2007) is normally used to detect the positively selected genes. CodeML estimates various parameters such as the ratio of nonsynonymous over synonymous (dN/dS) and branch length on the codon alignment or protein alignment based on a given topology (phylogenetic tree). Three different categories of models exist in CodeML. 1) The model estimates a unique dN/dS ratio for the whole genome alignment (Model=0); 2) The site model estimates different dN/dS among sites (Model=1); 3) The branch-site model estimates different dN/dS among sites and among branches (Model=2).

It will be essential to study the selective forces that give rise to the specific traits (Vallender & Lahn, 2004). A recent natural selection study has explained population genetics by identifying empirical evidence of natural selection that had constrained the level of neutral genetic diversity across many species (Corbett-Detig *et al.*, 2015). In addition, human genome-wide positive selection signals also suggested frequent adaptation in human evolution (Enard *et al.*, 2014). The first positive selection in humans was on the Major Histocompatibility Complex (MHC) gene, which showed high levels of allelic diversity in human (Austin L Hughes & Nei, 1988). A previous study also reported that human taste receptor for sweetness can show different perspective among primates, where humans capable of recognizing sweeter compound than prosimians (Glaser *et al.*, 1995). Moreover, the positive selection of *MC1R* genes can effect manipulation in the human and animal skin and hair coloration (Mundy & Kelly, 2003). A recent study has even explored the positive selection in the genetic basis of pork quality through whole-genome sequencing (Jeong *et al.*, 2015). Therefore, the positive selection

analysis would definitely provide a useful approach to give better insights into the evolutionary adaptations of pangolins.

University of Malaya

CHAPTER 3: MATERIALS AND METHODOLOGY

3.1 Pangolin specimen

A female wild Malayan pangolin, provided by the Department of Wildlife and National Parks (DWNP) Malaysia under a Special Permit No. 003079 (KPM 49) for endangered animals, was used in this study. The Chinese pangolin originated from an island of Taiwan. Both the pangolin genomes used in this study were previously sequenced, assembled and functionally annotated by the Genome Informatics Research Group (GIRG), University Malaya.

3.2 Pseudogene screening

Pseudogene screenings for the genomes of both Malayan pangolin and Chinese pangolin were performed using Pseudopipe pipeline (Z. Zhang *et al.*, 2006). The protein-coding genes of the Malayan (23,448 genes) and Chinese (20,298 genes) pangolins were used in this analysis. All protein sequences were BLAST-searched against their own pangolin genome sequences. The e-value cut-off ($\leq 1E-4$) was used to identify significant homologous hits. The BLAST hits were then partitioned according to the scaffold ID and strand direction. BLAST hits that had overlap >30bps with the functional genes were discarded. The partitioned BLAST hits were then categorized into different sets based on the match of the hits on the similar or different query protein. Each disjoint set was further merged into single super-hit or pseudo-exon. The pseudo-exon was further extended in both directions for 30 nucleotides to achieve optimal alignment by using the tfasty (Lipman & Pearson, 1985) program from the FASTA suite that refines the BLAST results for the identification of frameshift, stop codons, indels and calculates accurate sequence similarity (W. R. Pearson & Lipman, 1988). To obtain high-quality pseudogene predictions, the output results were further filtered based on the sequence similarity cut-off (> 40%), BLAST e-value < 1E-10 and the predicted pseudogenes must cover at least

70% of the parent genes. The pseudogenes were classified into either processed or duplicated pseudogenes.

3.3 Pangolin trait study

To identify potential gene mutations that could be associated with the unique traits of pangolins, I compiled 405 different known genes from the literature (Appendix A). 107 of genes were related to tooth formation and dental morphological development, 217 genes were related to vision, eye development, light sensing protein and other were selected for pangolin poor vision traits study, and 81 were immune response genes. The protein sequences of these genes of interest were retrieved from the National Center for Biotechnology Information (NCBI) and subjected for the pseudogene screening in the genome sequences of both Malayan pangolin and Chinese pangolin, as well as closely related organisms such as cat (*Felis catus*) genome (FelCat6.2/ Ensembl release 69) and dog (*Canis familiaris*) genome (CanFam3.1/ Ensembl release 67) by running the Pseudopipe pipeline. A confident pseudogene threshold was used (e-value < 1E-10 and 40% parent gene identity) to select the candidate pseudogenes present in these genomes.

3.4 Sanger sequencing validation of pseudogenes

To validate mutations found in the pseudogenised genes, I performed Polymerase Chain Reaction (PCR) and Sanger sequencing for each gene, Sanger sequencing was performed using DNA extracted from seven unrelated adult Malayan pangolins and one fetus sample from one of the adult pangolins that were used in this validation. Further validation was done by using four African pangolin genomes, which are Ground pangolin (*M. temminckii*), Black-bellied pangolin (*M. tetradactyla*), White-bellied pangolin (*M. tricuspis*) and Giant pangolin (*M. gigantea*).

Primer sequences used in the validation as shown in Table 3.1. Two sets of primers were used to examine two different mutated regions in the *ENAM* genes. The first primer pair

and second primer pair were used to validate the frameshift mutation and premature stop codon found in the *ENAM* gene. The *AMELX* gene was amplified with a forward primer and a reverse primer targeting 552 bps mutation site. Whereas *AMBN* was amplified with a forward primer and a reverse primer targeting 792 bps of mutation site. *BFSP2* amplified with a forward primer and a reverse primer targeting 437 bps of the mutation site. Whereas *GUCA1C* was amplified with a forward primer and a reverse primer targeting 476 bps of the mutation site. The *IFNE* gene was amplified using a forward primer and a reverse primer. As for positive selected genes validation, two sets of primers were used to amplify *KRT36* gene, the first forward primer and reverse primer targeting 399 bps of the mutation site. The second forward primer and reverse primer targeting 354 bps of the mutation site. The other keratin *KRT75* amplified with a forward primer and a reverse primer targeting 370 bps of the mutation site.

All the primers were used for PCR using the following described protocol. The total reaction volume of 25 μ L contained 60ng purified gDNA, 0.3 pmol of each primer, deoxynucleotides triphosphates (dNTP, 400 μ M each), 0.5 U Taq DNA polymerase and supplied buffer were used. The PCR was performed as follow: 1 cycle (94 $^{\circ}$ C for 2 minutes) for initial denaturation; 30 cycles (98 $^{\circ}$ C for 10 seconds; 65 $^{\circ}$ C for 30 seconds; 68 $^{\circ}$ C for 30 seconds) for annealing and extension for DNA amplification. The PCR products were purified by standard methods and directly sequenced with the same primers using BigDye $^{\circ}$ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

Table 3.1: Primer sequences used for the validation of gene mutations. The mutations were validated using Sanger sequencing.

Gene	Primers	Primer Length (bps)	Annealing Temperature
<i>ENAM_1</i>	5'-AGCAAGTTCAAAGGGTTTCTCAGC-3' 3'-TTCAGCTTGTTTCATCAGAATTTGG-5'	464	62 °C
<i>ENAM_2</i>	5'- CCTTATTTTCAGTAAGTCCCAAGCT-3' 3'-CTTGCAATTCTTAGTCTGGGTATCTT-5'	348	62 °C
<i>AMELX</i>	5' CAAAGAAAGCATTGCTACTTCTCC 3' 3'-GGGTCTAGAGTTTCAGTAACCAGAG-5'	552	62 °C
<i>AMBN</i>	5'-GTCTCTCACTTTGTTACGGTTTCT-3' 3'-CAAGTGTTTGGAATAATATAGGTCCC-5'	792	62 °C
<i>BFSP2</i>	5'-GAGTGCCCAGAGTCTATGTAGGGATGG-3' 3'-TTCCTGCTCATCGTCCTTCCCAGAG-5'	437	68 °C
<i>GUCA1C</i>	5'-CAGCTGTAAGAGATTTGGAGTAGC-3' 3'-GTTTAATGACTCACTTACCTACAAGC-5'	476	62 °C
<i>IFNE</i>	5'-GAGGAAATGTCCCATGAACACTAGG-3' 3'-GGGTAGGTGGATGGGAGAGAGAACT-5'	878	65 °C
<i>KRT75</i>	5'-GAGCATCTAGGGCACAATGC -3' 3'-GCATTCCCTGAGCTGACAGA -5'	370	66 °C
<i>KRT36_1</i>	5'-CAGTTAGCATGGGTAGCAGGTG -3' 3'-CTCTGCTTCAACCTCCCTGAG -5'	399	63 °C
<i>KRT36_2</i>	5'-AGAGATGCGGTGTCAGTATG -3' 3'-CATCCAGAAAGAAAGAGGAGAG -5'	354	60 °C

3.5 RNA-seq expression analysis

RNA-seq raw data were retrieved from the Malayan pangolin transcriptome sequencing project (Yusoff A.M *et al.*, In-press 2016). Briefly, transcriptomes from the cerebellum, cerebrum, lung, heart, kidney, liver, spleen and thymus were sequenced using Illumina HiSeq technology platform (100bp Paired End (PE) strategy). The data from each sample ranged from 41-53 million of PE reads or ~8.2-10.6Gbp. For each organ, the PE reads were mapped to the genome assembly using TopHat 2.0.11 (Trapnell *et al.* 2009). The mapped PE reads with the best match were sorted and indexed with SAMtools (Li *et al.* 2009). The expression of each gene was calculated and represented in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) based on the coordinates of the

MAKER-generated gene models. Heat map showing expression of pangolin-specific genes across the eight different pangolin organs were generated using in-house R scripts. Using the same approach, the expression level of pseudogenised *IFNE* gene in the transcriptomes of three organs (lung, cerebrum, and cerebellum) of Malayan pangolin were calculated, as well as a transcriptome of skin tissue which was sequenced using Illumina MiSeq technology.

3.6 Comparative pseudogene analysis of predicted pseudogenes

The selected pseudogene candidates were manually aligned for sequential comparative analysis. The list of potential pseudogenes was aligned with other animal genomes such as dog (*Canis familiaris*), cat (*Felis catus*) and human (*Homo sapiens*). These selected gene sequences were BLAST searched from Malayan pangolin genome and Chinese pangolin genome, and then sent to MultAlin (Corpet, 1988) online multiple sequence alignment server for Multiple Sequence Alignment (MSA). The alignments were performed using their default parameters and scoring method, for example, DNA-5-0 (Gap penalty for opening is -5 and gap penalty for extension is 0) for nucleotide base MSA, BLOSUM-12-2 for protein base MSA (Gap penalty for opening is -12 and gap penalty for extension is -2). The MSA results were then exported to BioEdit (Hall, 2004) tools for careful manual examination of frameshift indels mutation and premature stop codon as the significant variation, which might lead to losing of the function at the protein level.

3.7 Functional enrichment analysis

Enrichment analyses of the pangolin-specific and the significantly expanded gene families were performed using Blast2GO (Conesa & Gotz, 2008), with all pangolin genes used as background for comparisons (Conesa, Gotz, *et al.*, 2005). Fisher's exact test was used with a cut-off of a False Discovery Rate (FDR) of 0.05 after multiple test correction.

3.8 Positive selection analysis

To find the signatures of positive selection in the pangolin lineage, ten different mammal species (horse, dog, cat, panda, cow, mouse, human, megabat, plus the two pangolin species) were selected and used in this analysis. In total, 8,498 orthologs (1:1) between the species were found using Poff/proteinortho tool (Lechner *et al.*, 2014). Orthologs were aligned using Prank F+ codon model and filtered out using Gblocks software (Castresana, 2000). The final set consisted of 8,250 genes. Codeml software as part of PAML 4.8 package (Z. Yang, 2007) was used to test branch-site model of positive selection for pangolin lineage. A total of 8250 1:1 orthologs were tested. For each orthologs alignment, codon model of molecular evolution was fitted with pangolin lineage as foreground branch and other lineages as background branches. A null model of neutral evolution of foreground branch with fixed omega = 1 (model =2, NSsites=2) was tested using Likelihood-Ratio Test (with $p < 0.05$ significance level) against the alternative model with estimated omega value (model =2, NSsites=2).

3.9 Functional impact of substitutions in positive selected genes

The potential functional impact of pangolin-specific amino acid changes in the positively selected genes of interest in the pangolin ancestor was assessed using Protein Variation Effect Analyzer (PROVEAN) (Choi and Chan 2015). Amino acid substitutions were considered “deleterious” if the PROVEAN score was ≤ -2.5 and “neutral replacements” if the score was > -2.5 .

3.10 Gene family expansion and contraction analysis

Correct assignment of genes to families is highly relevant for this analysis. Firstly, 40 mammalian species present in TreeFam database (H. Li *et al.*, 2006) were retrieved and tested. Proteins of these species assigned to families annotated in the TreeFam database were extracted and mapped to hmm profiles of known families present in this database

(<http://www.treefam.org/static/download/treefam9.hmm3.tar.gz>) by HMMER3 (hmmsearch with default parameters) (Johnson *et al.*, 2010). Each protein was assigned to a top-hit family if it was statistically significant ($E\text{-value} < 1E-6$). Surprisingly, twenty-one proteins were unmapped to any family, and six more proteins have no significant hits. However, 8,885 of 10,306 (86.2%) families were correctly assembled for all species, and 9,213 (89.4%) families were assembled with errors in less than 5% species. This set of families was used in following expansion/contraction analysis. The assignment of pangolin genes to families was performed in the same manner. Expansion and contraction analysis of gene families were conducted by CAFE v3.1 (De Bie *et al.*, 2006). Eight reference species were used: *Ailuropoda melanoleuca* (giant panda), *Bos taurus* (cow), *Canis familiaris* (dog), *Equus caballus* (horse), *Felis catus* (cat), *Homo sapiens* (human), *Mus musculus* (mouse), *Pteropus vampyrus* (megabat). One of CAFE assumptions is that most recent common ancestor (MRCA) of all analyzed species had at least one protein in each family. Because of this reason, families present in less than three species were removed. Therefore 8,438 families were retained in the final dataset.

3.11 Development of Pangolin Genome Database (PGD) hub

The PGD hub was implemented under the Linux system using various conventional software packages including Apache, MySQL, PHP, and Perl. PGD was designed in PHP-HTML5 using CodeIgniter and Twitter Bootstrap as the back-end and front-end frameworks, respectively (Figure 3.1). Additionally, the website was separated into logic, presentation and application data into three interconnected parts following the Model-view-controller (MVC) framework. Whilst for the client-side, it was achieved with jQuery, which is a feature-rich JavaScript library that may enhance user interaction with the web pages through AJAX (Asynchronous JavaScript and XML) that is used to transfer data between the client-side and server-side. The genomic annotations data that are

generated from in-house scripts and published software are stored in the MySQL database management system.

PGD contains four main features/tabs: (i) “Browse” tab for users to browse all pangolin genome annotation and transcripts information; (ii) “Tools” tab, which contains a list of analysis tools for users to analyze pangolin genomic data; (iii) “Genome Browser” tab which has incorporated JBrowse for visualizing pangolin genome sequences and annotation; and (iv) Data download tab which allows users to download all pangolin genome/transcript data and annotations. Additionally, PGD also has implemented a quick search function, allowing users to rapidly search for genes of interest stored in the PGD. Furthermore, users can perform sequence searches against the pangolin genome sequences with input sequences using the provided web-based BLAT (Kent, 2002) and BLAST (Mount, 2007) tools. Our web architecture allows users to access and submit their tasks using the front-end provided securely with the web server via the Internet. Normalized and optimized of the database schema were carried out in order to reduce the data redundancy of pangolin's genomic data. Therefore, the database was well-designed in such a way that MySQL would be performing optimally without the loss of ability to maintain data integrity. As such, a user can submit tasks, search, browse, and retrieve the genomic data through PGD in a robust and efficient manner.

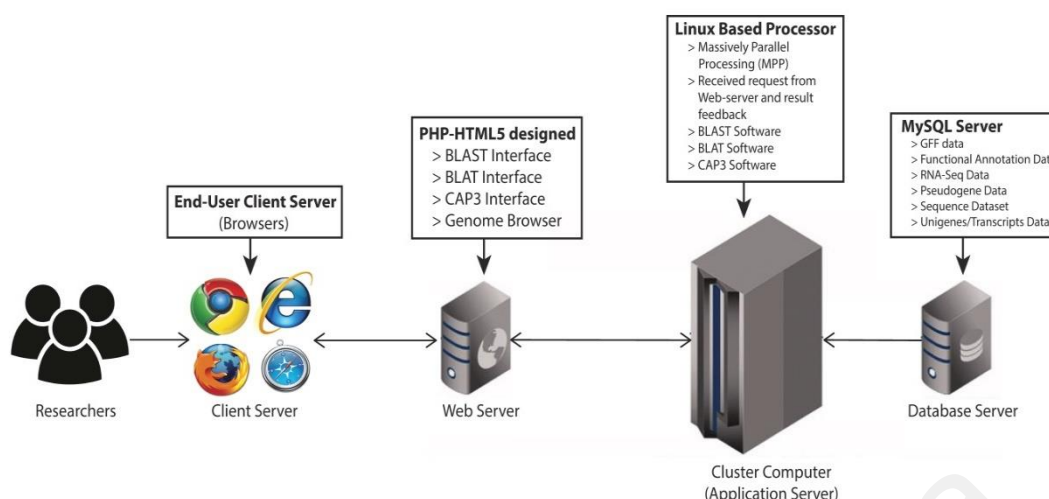


Figure 3.1: PGD 4-tier web application architecture. (Client workstation, web server, application server and database server)

3.12 Pangolin genome browser in the PGD database

To allow the visualisation of the pangolin genome sequence and annotations on the fly, I incorporated an entirely fully dynamic AJAX-based real-time genome browser, JBrowse (Skinner *et al.*, 2009) into the PGD.

The genome browser was separated into two distinct browsers according to the pangolin species. Users are able to choose whether to visualize the pangolin data in Malayan pangolin genome or Chinese pangolin genome. Once they chose which genome to browse, all relevant annotation tracks in the genome browser will be available for users. Genome sequence track of each pangolin was set as default for users, different annotation tracks that are able to trigger on/off for easier visualization and analysis were implemented. These tracks are classified into different categories such as ‘Genes and Annotations,’ ‘Repeat,’ ‘Transcriptome,’ ‘Transcript Reads Mapping,’ ‘Organ-Specific Expression,’ and ‘Variation’ which represent our annotations from different analyses.

CHAPTER 4: RESULTS FOR COMPARATIVE ANALYSIS OF *MANIS JAVANICA* AND *MANIS PENTADACTYLA*

4.1 Identification of pseudogenes

The pseudogenes were classified into processed pseudogene (retrotransposed pseudogene) and duplicated pseudogenes. Genome-wide screening of pseudogenes revealed a total of 3,316 pseudogenes in the Malayan pangolin, which includes 2,649 processed pseudogenes (79.8%) and 667 duplicated pseudogenes (20.1%) (Table 4.1). On the other hand, a total of 1,577 pseudogenes were found in the Chinese pangolin genome including 1,227 processed pseudogenes (77.8%) and 350 duplicated pseudogenes (22.1%). The number of pseudogenes in both pangolin species is relatively lowered than the number of predicted pseudogenes in human and dog.

Table 4.1: Pseudogene identification across different mammalian species. The number of processed and duplicated genes is shown for each species. The percentage of pseudogenes is shown in parentheses. The human and dog pseudogene datasets were from www.pseudogene.org (Filter threshold: Pseudogene completeness $\geq 70\%$, e-value $< 1E-10$, Sequence identity $\geq 40\%$)

Pseudogene Type	Malayan pangolin	Chinese pangolin	Dog	Human
Processed	3637 (78.05%)	1859 (76.95%)	6075 (86.23%)	8522 (82.25%)
Duplicated	1023 (21.95%)	557 (23.05%)	970 (13.77%)	1839 (17.75%)
Total	4660	2416	7045	10361

4.2 Unique traits and adaptations of pangolins

To gain insight into the possible genomic patterns associated with the unique traits of pangolins, I identified potential pseudogenised genes (presumably loss-of-function) that are potentially linked with the unique traits. Gene loss through pseudogenization is increasingly recognized as an important factor in the lineage and adaptive phenotypic diversification (Albalat & Canestro, 2016; Wang *et al.*, 2006).

4.2.1 Edentulous characteristic of pangolin species

Given pangolins are edentulous mammals, 107 tooth development-related genes manually compiled from the literature were screened through pangolins genomes, and three pseudogenised candidate genes were commonly found in both pangolins. Most notable of these genes was the *ENAM* gene, the largest protein found in enamel matrix, the enamelin protein essential for regular tooth development (R. W. Meredith *et al.*, 2009). Mutations in this gene can lead to enamel defectiveness (Hart *et al.*, 2003). Several frameshift indels and premature stop codons were observed in the *ENAM* genes of both the Malayan and Chinese pangolins (confirmed by Sanger sequencing; sample size=8; Appendix C) (Figure 4.1a). Interestingly, the *ENAM* pseudogenization was also reported in the African tree pangolin, *M. tricuspis* (R. W. Meredith *et al.*, 2009).

Other pseudogenised genes that are related to tooth development are amelogenin (*AMELX*) and ameloblastin (*AMBN*). As shown in Figure 4.1a, frameshift indels were observed in both amelogenin (Appendix D) and ameloblastin (Appendix E). There was a frameshift deletion in positions c1311-c1324, a single base pair deletion in c1476 and an insertion of AGAT at position 1621, resulting in another premature stop codon in the ameloblastin gene. The amelogenin gene contains a large frameshift deletion at position c396-c455. The mutations found in the two known genes may indicate the loss-of-function of these genes which likely contributed to the loss of teeth in pangolins (Ghodsi *et al.*, 2012; Landin *et al.*, 2012). These genes shared common mutations in both Asian and all four African species (*M. tricuspis*, *M. tetradactyla*, *M. gigantean*, and *M. temminckii*) that later examined, provide evidence to suggest that pseudogenization occurred early in the evolutionary history of the Pholidota lineage (Figure 4.2 and Figure 4.3). Moreover, all of these pseudogenised genes have also been reported to be associated with other edentulous vertebrates such as toothless baleen whales, birds, and turtles (R. W. Meredith *et al.*, 2014).

4.2.2 Poor vision

Most Pangolins are nocturnal (the only exception being the long-tailed pangolin) and have poor vision (D. A. Soewu & I. A. Ayodele, 2009). 217 vision-related genes were screened and identified several candidate genes that were pseudogenised in both pangolins (Figure 4.1b). One of these genes, *BFSP2* encodes a lens-specific intermediate filament-like protein that is a unique cytoskeletal element in the ocular lens of vertebrates (Sandilands *et al.*, 2003). Interestingly, one frameshift insertion and three premature stop codons were consistently identified in the *BFSP2* genes of both pangolin species (confirmed by Sanger sequencing; sample size=8; Appendix F), suggesting a possible loss of optical clarity, which may lead to progressive cataracts (Alizadeh *et al.*, 2004). Another candidate gene, Guanylate cyclase activator 1C (*GUCA1C*) stimulates photoreceptor guanylate cyclase (GC) in the rod outer segment at low concentration of calcium ion. The *GUCA1C* proteins stimulate both photoreceptors GC1 and GC2 when the concentration of the free calcium ions is low, but inhibit GCs when the concentration of the free calcium ions elevated which is important for regulating the recovery of the dark state of rod photoreceptors following light exposure (Haeseleer *et al.*, 1999). Identical frameshift mutations and premature stop codons were found in the *GUCA1C* gene in both pangolin species (confirmed by Sanger sequencing; sample size=8) (Figure 4.1b), suggesting a reduced function of the simulations of GC1 and GC2 and also reduced the rate of phototransduction in pangolins (Haeseleer *et al.*, 1999). The *GUCA1C* gene was also pseudogenised in all African species with identical mutations (Figure 4.4; Appendix G). Interestingly, the pseudogenization of *BFSP2* and *GUCA1C* have been reported in mice with poor vision, supporting the hypothesis view that the two genes are likely associated with the poor vision of pangolins (Imanishi *et al.* 2002; Sandilands *et al.* 2004; Song *et al.* 2009).

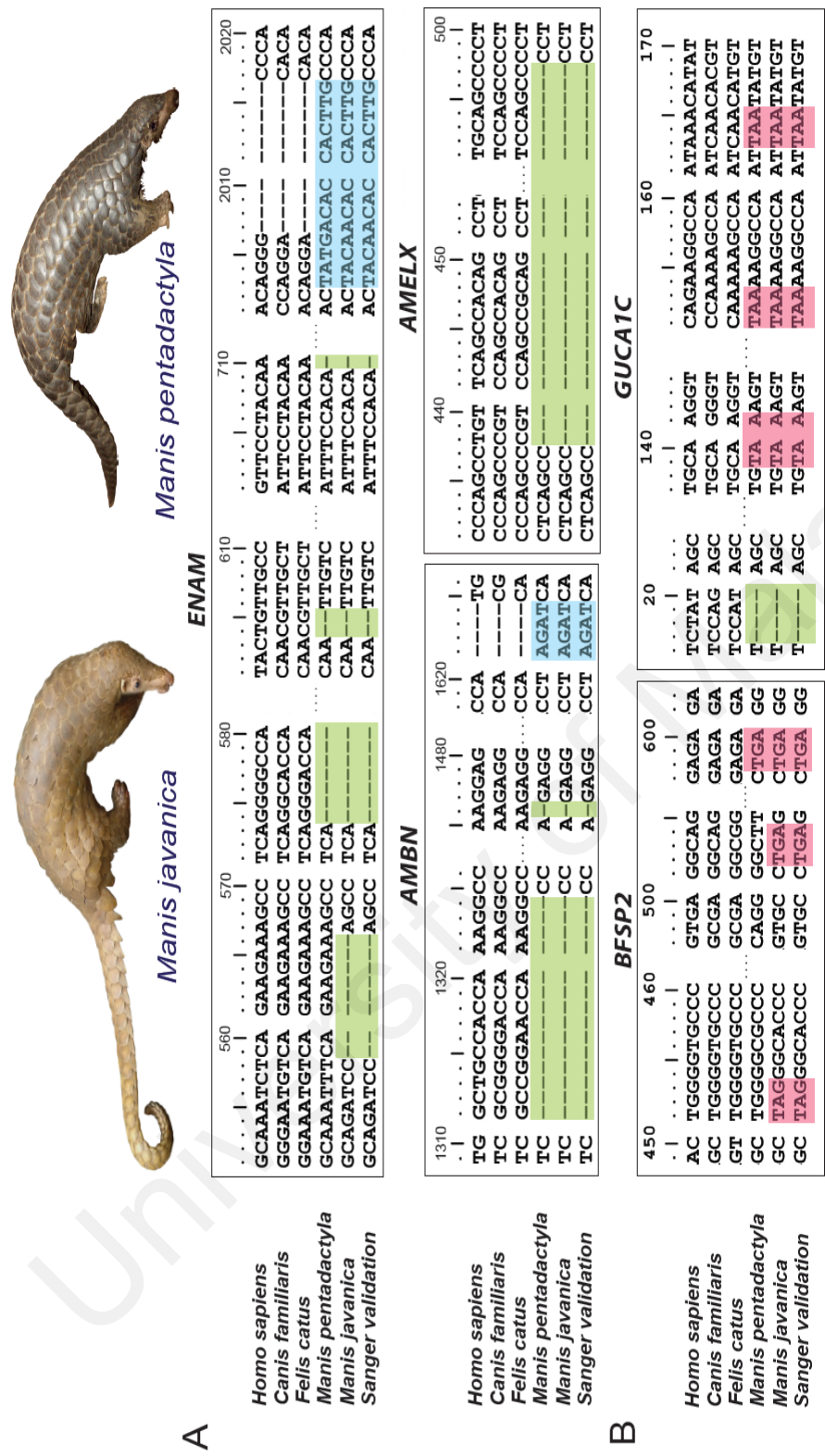


Figure 4.1: Case studies of pseudogenised genes in pangolins. (A) Three genes involving in the tooth development were pseudogenised and may be related to the loss of teeth in pangolins. (B) Two genes were pseudogenised and may be related to the poor vision of pangolins. Blue=insertion; green=deletion; and Pink=stop codon.

AMB	1310	1320	1480	1620
<i>Homo sapiens</i>	TG	GCTGCCACCA AAGGCC	AAGGAG	CCA
<i>Canis familiaris</i>	TC	GCGGGGACCA AAGGCC	AAGAGG	CCA
<i>Felis catus</i>	TC	GCCGGAACCA AAGGCC	AAGAGG	CCA
<i>Manis temminckii</i> (5)	TC	-----CC	A-GAGG	CCT
<i>Manis tricuspis</i> (5)	TC	-----CC	A-GAGG ...	CCT
<i>Manis tetradactyla</i> (4)	TC	-----CC	A-GAGG	CCT
<i>Manis gigantea</i> (1)	TC	-----CC	A-GAGG	CCT
<i>Manis javanica</i>	TC	-----CC	A-GAGG	CCT
<i>Manis pentadactyla</i>	TC	-----CC	A-GAGG	CCT

Figure 4.2: Sanger validation of AMB genes in African pangolin species. The numbers in the

brackets showed the number of samples examined.

AMELX	440	450	485	500
<i>Homo sapiens</i>	CCCAGCCTGT	TCAGCCACAG CCT	CCACCT GTGCACCCCA	TGCAGCCCCCT
<i>Canis familiaris</i>	CCCAGCCCCT	CCAGCCACAG CCT	CCACCC GTGCACCCCA	TCCAGCCCCCT
<i>Felis catus</i>	CCCAGCCCCT	CCAGCCGCAG CCT	CCACCC GTGCACCCCA	TCCAGCCCCCT
<i>Manis temminckii</i> (5)	CCCAGCC	---	---ACCC ATGTGCCCA	TTCAGCCCCCT
<i>Manis tricuspis</i> (5)	CCCAGCC	---	---ACCC GTGTGCCCA	TTCAGCCCCCT
<i>Manis tetradactyla</i> (5)	CCCAGCC	---	---ACCC ATGTGCCCA	TTCAGCCCCCT
<i>Manis gigantea</i> (1)	CCCAGCC	---	---ACCC ATGTGCCCTA	TTCAGCCCCCT
<i>Manis javanica</i>	CTCAGCC	---	---	---CCT
<i>Manis pentadactyla</i>	CTCAGCC	---	---	---CCT

Figure 4.3: Sanger validation of AMELX genes in African pangolin species. The numbers in the brackets indicated the number of samples examined.

GUCA1C	20	140	160	165
		...		
<i>Homo sapiens</i>	TCTAT AGC	TGCA AGGT	CAGAAAGGCCA	ATAAACA
<i>Canis familiaris</i>	TCCAG AGC	TGCA GGGT	CCAAAAGCCA	ATCAACA
<i>Felis catus</i>	TCCAT AGC	TGCA AGGT	CAAAAAGCCA	ATCAACA
<i>Manis temminckii</i> (5)	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA
<i>Manis tricuspis</i> (5)	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA
<i>Manis tetradactyla</i> (5)	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA
<i>Manis gigantea</i> (1)	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA
<i>Manis javanica</i>	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA
<i>Manis pentadactyla</i>	T--- AGC	TGTA AAGT	TAAAAGGTCA	ATTAAATA

Figure 4.4: Sanger validation of *GUCA1C* genes in African pangolin species. The numbers in parentheses indicate the number of samples examined.

4.2.3 Immunity of pangolins

It is plausible that the immunity of pangolins evolved differently than in other mammals. Interferons (IFNs) are a cluster of highly conserved gene families that encode for cytokines expressed by host cells for communication between cells, leading to the activation of the immune system in the presence of pathogens (De Andrea *et al.*, 2002; Fensterl & Sen, 2009). The *IFNE* gene is a unique interferon exclusively expressed in skin epithelial cells and inner mucosa-protected tissues (e.g. lung, intestines and reproductive tissues), establishing a first line of defense against pathogens in other placental mammals (Day *et al.*, 2008; Demers *et al.*, 2014; Fung *et al.*, 2013; Ponten *et al.*, 2008; Uhlen *et al.*, 2015; Xi *et al.*, 2012).

Strikingly, I found that the single copy intron-less *IFNE* gene is pseudogenised in both pangolin species (confirmed by Sanger sequencing; sample size=8; Appendix H), but it is intact in 71 other mammalian species that I examined (Figure 4.5). An insertion from positions 195 to position 219 was observed, and a point mutation at position c235C>T causing a premature stop codon. Other two frameshift deletions at positions 264 and 540-546 would cause the loss of function. Although an alternate splicing pattern might avoid the premature stop codon and preserve the protein, high-throughput RNA sequencing of lung, cerebellum, cerebrum and skin transcriptomes (Illumina MiSeq; 250bp PE; # of PE reads=~17 millions) failed to detect *IFNE* expression (Fragments Per Kilobase Million (FPKM) = 0.0) although it is expressed in these organs in other mammals (Demers *et al.*, 2014; Uhlen *et al.*, 2005). Sequencing reads that mapped to the remainder of the gene after the premature stop codon were unable to identify, which is consistent with the hypothesis of a loss of function even if the stop codon were rescued. To test whether the *IFNE* was also pseudogenised in the African pangolins, the gene in all African species were sequenced using same primers designed based on the genome sequence of Malayan pangolin. My data showed

that the IFN gene was also pseudogenised in all African species and shared ancestral mutations with the Asian pangolin species (Figure 4.6).

Interestingly, many putative functional domains or key signatures are deleted in the pangolin *IFNE* protein due to the premature stop codon (Figure 4.7). This suggests that the *IFNE* protein is unlikely to function properly if expressed, indicating that resistance to infection may be impacted in pangolins. Other IFN families were examined (*IFNB*, *IFNK*, *IFNA*, *IFNG* and *IFNL*) and found that all families have intact gene copies in both pangolin species (Table 4.2). It should be noted that the *IFNA* family consists of a cluster of 13 functional genes in human. Although several intact genes found from the IFN family in both pangolin species were retained, the number of *IFNA* genes is relatively low compared to human, suggesting diminished function of this gene family in pangolins.

Subfamily	Gene Number						
	Human	Dog	Cat	Mouse	Cow	Malayan Pangolin	Chinese Pangolin
<i>IFNA</i>	13	3	19	14	13	1	3
<i>IFNB</i>	1	1	1	1	6	1	1
<i>IFNE</i>	1	1	1	1	1	0	0
<i>IFNK</i>	1	1	1	1	1	1	1
<i>IFNG</i>	1	1	1	1	1	1	0
<i>IFNL</i>	3	1	1	3	0	3	3

Table 4.2: Interferon family gene number across the different animals. Pangolins have had intact copy of different IFN family genes examined (*IFNB*, *IFNK*, *IFNA*, *IFNG* and *IFNL*), except *IFNE* gene which is absent in both pangolins.

	Nucleotide sequences						Protein sequences		
IFNE	195	210	220	265	540	545	65	75	95
<i>Homo sapiens</i>									
<i>Pan paniscus</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAGCCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Pan troglodytes</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAGCCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Gorilla gorilla</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAGCCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Pongo abelii</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Nomascus leucogenys</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Rhinopithecus roxellana</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Macaca fascicularis</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Colobus angolensis palliatus</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Macaca mulatta</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Mandrillus leucophaeus</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Cercocebus atys</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Macaca nemestrina</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Papio anubis</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Chlorocebus sabaeus</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Aotus nancymae</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACTG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Callithrix jacchus</i>	TCAGAAG--	-----	-----	TCTCC	GAA ATCAACCG	-----	FLLPQKS--	----LSPQQY	EMLQQ
<i>Galeopterus variegatus</i>	CCAGAAG--	-----	-----	TTTCC	GAA ATCAACAG	-----	FLLPQKS--	----ANPQQY	EMLQQ
<i>Pteropus vampyrus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----LNPPLY	ETLQQ
<i>Pteropus alecto</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----LNPPLY	ETLQQ
<i>Ceratotherium simum</i>	ACAGAAA--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Propithecus coquereli</i>	GCAGATG--	-----	-----	TCTCC	GAA ATCAGTCA	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Balaenoptera acutorostrata sca</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----MNPQY	EMLQQ
<i>Physeler catodon</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Equus caballus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Lipotes vexillifer</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----MNPQY	EMLQQ
<i>Orcinus orca</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----MNPQY	EMLQQ
<i>Otlemur gametii</i>	CGAGAG--	-----	-----	GCTCC	GAA ATCAATCG	-----	FLLPQKS--	----VSPQY	ETLQQ
<i>Sus scrofa</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----MNPQY	EMLQQ
<i>Ovis canadensis canadensis</i>	CCAGAAG--	-----	-----	WCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Camelus ferus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----MDPQY	EMLQQ
<i>Ovis aries musimon</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Bos taurus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Bison bison bison</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Epitescus fuscus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Vicugna pacos</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Bubalus bubalis</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Tarsius syrichta</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Bos mutus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Pantholops hodgsonii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Trichechus manatus latirostris</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Capra hircus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Canis lupus familiaris</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Orycteropus afer afer</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Dasyatis novemcinctus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Tupaia chinensis</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Myotis brandtii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Myotis lucifugus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Leptonyx chotes weddellii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Loxodonta africana</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Oryctolagus cuniculus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Felis catus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Panthera tigris altaica</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Odobenus rosmarus divergens</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Ursus maritimus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Alouropoda melanoleuca</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Mustela putorius furo</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Myotis davidii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Condylura cristata</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Nannospalax galii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Chrysocloris asiatica</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Eriacus europaeus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Ictidomys tridecemlineatus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Sorex araneus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Dipodomys ordii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Ochotona princeps</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Echinops telfairi</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Elephantulus edwardii</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Cricetus griseus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Mesocricetus auratus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Mus musculus</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Manis pentadactyla</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Manis javanica</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ
<i>Sanger validation</i>	CCAGAAG--	-----	-----	TCTTC	GAA ATCAACCG	-----	FLLPQKS--	----VNPQY	EMLQQ

Figure 4.5: Multiple sequence alignment of all mammalian *IFNE* genes. 73 mammalian species had available *IFNE* sequences and used for alignment. Nucleotide sequence alignment of *IFNE* genes across different mammalian species. Blue=insertion and green=deletion. Protein sequence alignment of *IFNE* genes across different mammalian species is also shown. An insertion (blue) in *IFNE* was observed starting from nucleotide position of 195th but not in other mammalian species, indicating that this insertion is specific to pangolins and possibly a marker to differentiate pangolins and others mammalian species. A premature stop codon or frameshift (pink) in the *IFNE* gene was consistently detected in both pangolin genomes. These mutations were validated by Sanger sequencing in eight Malayan pangolins. Protein sequences highlighted in red are the frameshift mutations and premature stop codon.

<i>IFNE</i>	190	220	240	270
<i>Homo sapiens</i>	TTTCTGCTTC CTCAGAAG--	-----T CTTTGAGTCC	TCAGCAGTAC	CATTCTCCAT
<i>Felis catus</i>	TTCCTGCTTC CCCAGCGG--	-----T CTGTGAATCC	TCGCCAGTAC	CATTCTTCAC
<i>Canis familiaris</i>	TTCCTGCTTC CCCAGCAG--	-----T CTGTGAATCG	TCACCAGTAC	CATTCTTCAT
<i>Manis temminckii</i> (5)	TTCCTACTTC CCCACCAGTA	CCAGAATCCT CCTGAATCTT	TCCTGAATCC	TCGCTAGTAC CAT-CTTCCT
<i>Manis tricuspis</i> (4)	TTCCTACTTC CCCACCAGTA	CCAGAATCCT CCTGAATCTT	TCCTGAATCC	TCGCTAGTAC CAT-CTTCCT
<i>Manis tetradactyla</i> (2)	TTCCTACTTC CCCACCAGTA	CCAGAATCCT CCTGAATCTT	TCCTGAATCC	TCGCTAGTAC CAT-CTTCCT
<i>Manis gigantea</i> (1)	TTCCTGCTTC CCCACCAGTA	CCAGAATCCT CCTGAATCTT	TCCTGAATCC	TTGCTAGTAC CAT-CTTCCT
<i>Manis javanica</i>	TTCCTGCTTC CCCACCAATA	CCAGGATCCT CCTGTATCTT	TCCTGAATCC	TCACTAGTAC CAT-CTTCCT
<i>Manis pentadactyla</i>	TTCCTGCTTC GCCACCAATA	CCAGGATCCT CCTGTATCTT	TCCTGAATCC	TCACTAGTAC CAT-CTTCCT

	65	75	85
<i>Homo sapiens</i>	FLLPQK----	---SLSPQQY	QKGHTLA
<i>Felis catus</i>	FLLPQR----	---SVNPRQY	QKGQALA
<i>Canis familiaris</i>	FLLPQQ----	---SVNRHQY	QKGQALA
<i>Manis temminckii</i> (5)	FLLPHQYQNP	PESFLNPR*Y	QKRHTIT
<i>Manis tricuspis</i> (4)	FLLPHQYQNP	PESFLNPR*Y	QKRHTIT
<i>Manis tetradactyla</i> (2)	FLLPHQYQNP	PESFLNPR*Y	QKRHTIT
<i>Manis gigantea</i> (1)	FLLPHQYQNP	PESFLNPC*Y	QKRHTIT
<i>Manis javanica</i>	FLLPHQYQDP	PVSFLNPH*Y	QKGHTIT
<i>Manis pentadactyla</i>	FLLRHQYQDP	PVSFLNPH*Y	QKGHTIT

Figure 4.6: Sanger validation of *IFNE* genes in African pangolin species. (Top) *IFNE* nucleotide sequences. (Bottom) *IFNE* protein sequences. The numbers in the brackets indicated the number of samples examined.

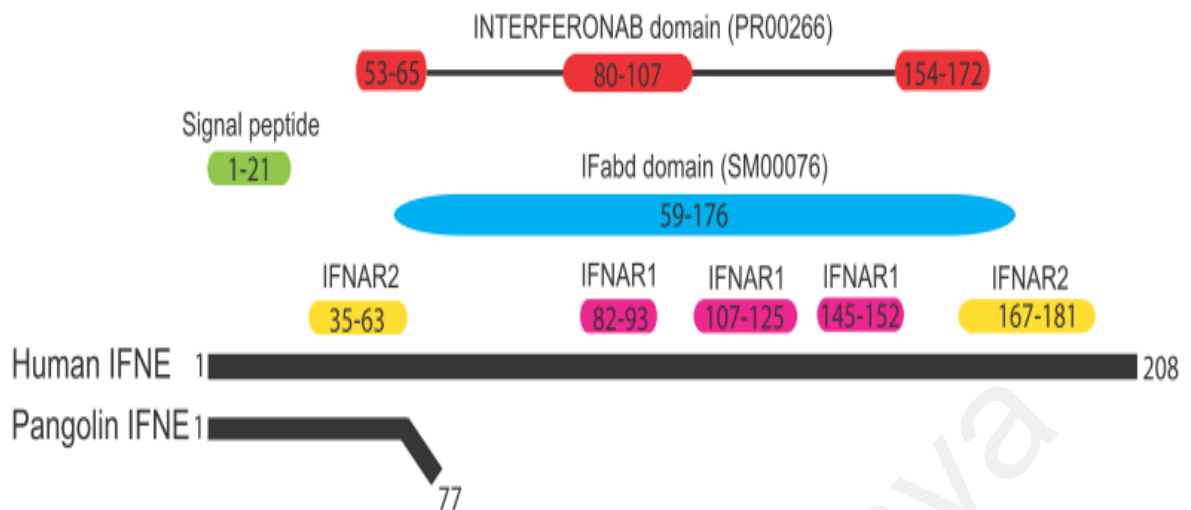


Figure 4.7: Comparison between pangolin and human (reference) IFNE protein sequences. Pangolins have a short putative protein sequence because of a premature stop codon. Predicted functional domains and signatures in the human *IFNE* gene are represented by color boxes. Yellow and pink boxes represent the predicted binding residues to Ifnar2 and Ifnar1, respectively, and which are bounded by interferons, which obtained from a previously published paper (Fung *et al.*, 2013). IFabd (SM00076) is a conserved functional domain in known interferons. The main conserved structural feature of interferons is a disulphide bond. INTERFERONAB (PR00266) is a 3-element fingerprint that provides a signature for alpha, beta and omega-interferons. The elements 1 and 3 contain Cys residues involved in disulphide bond formation.

4.3 Comparative analysis among pangolin and mammals

4.3.1 Comparative gene family clustering

To further explore genetic differences between pangolins and their closest relatives, the overlap of gene families among the two pangolin species and closely related species, namely the dog (*Canis familiaris*), cat (*Felis catus*), and giant panda (*Ailuropoda melanoleuca*) were compared. The analysis identified 8,325 ancestral gene families common to all five species (Figure 4.8).

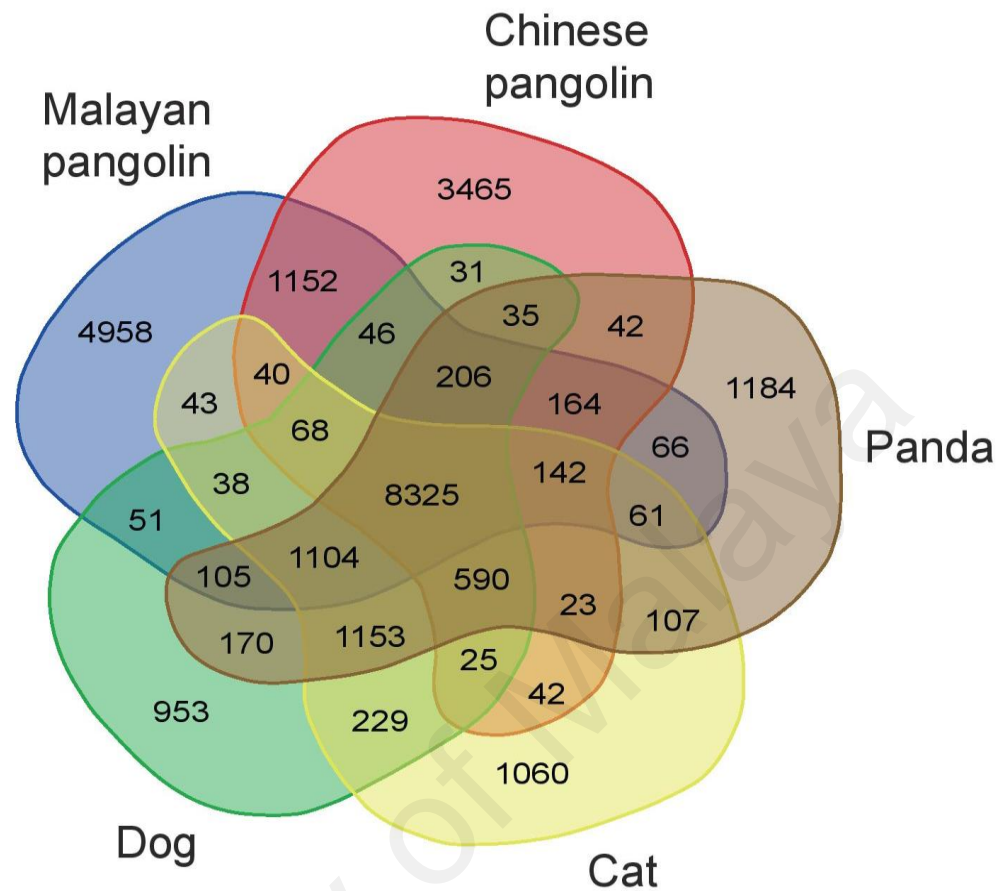


Figure 4.8: Venn diagram showing the unique and shared gene families among pangolins and their closest relatives (cat, dog, and giant panda).

Interestingly, many gene families were exclusive to either Malayan pangolin (4,958) or Chinese pangolin (3,465), probably reflecting the high level of divergence within pangolins likely resulting from evolutionary adaptations to different ecological environments. 1,152 pangolin-specific gene families were found, of which 61% are expressed in at least one of the eight organs assessed, further suggesting that at least these genes are not assembly or annotation artefacts (Figure 4.9).

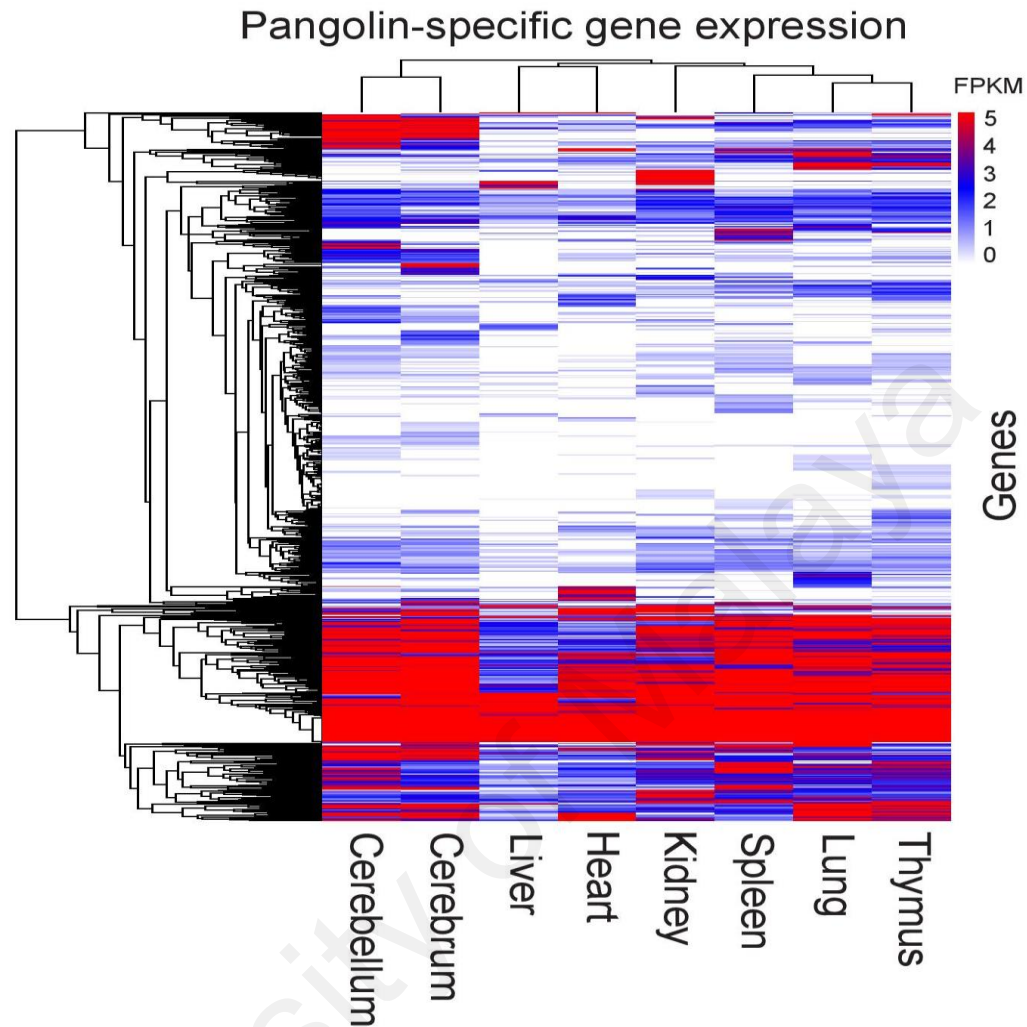


Figure 4.9: Heatmap shows the expression of pangolin-specific genes across different pangolin organs, which represented by FPKM values. Any FPKM values >5 were set to 5 for visualization purpose. Only genes expressed ($\text{FPKM} \geq 0.3$) in at least one organ were shown. Any FPKM values >5 were set to 5 in the heatmap for visualization purpose. Only genes expressed ($\text{FPKM} \geq 0.3$) in at least one organ were shown.

Functional enrichment analyses suggest that the pangolin-specific genes are significantly over-represented in signal transduction (496 genes; GO:0007165; FDR p-value=2.53E-09), neurological system processes (262 genes; GO:0050877; FDR p-value=2.51E-20), cytoskeleton organization (99 genes; GO:0007010; FDR p-value=1.53E-03) and cell junction organization (28 genes; GO:0034330; FDR p-value= 1.32E-02) (Figure 4.10). A large number of pangolin-specific genes significantly enriched in signal transduction and

neurological system processes suggest that pangolins evolved complex and unique signal transduction and neurological system networks. A large number of cytoskeleton organization associated genes may be linked with the assembly, formation and maintenance of the cytoskeletal elements involved in cell shape, structural integrity and motility of the pangolin scales.

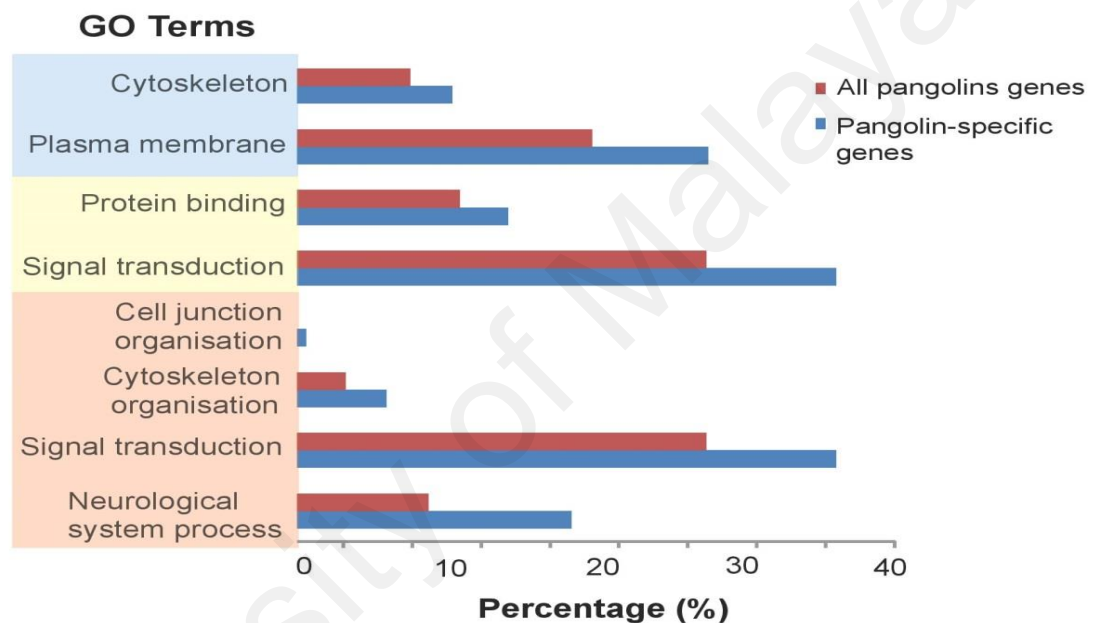


Figure 4.10: GO enrichment analysis of 1,152 pangolin-specific genes. Significantly enriched GO terms are shown for the categories of Cellular Compartment (blue), Molecular Function (yellow) and Biological Process (red).

4.3.2 Gene family expansion and contraction analysis

Gene family analysis identified 147 families that are significantly expanded and 18 that are contracted in the pangolin lineage, indicating that gene expansion and contraction events played a major role in the functional diversification of pangolins (Figure 4.11). GO term enrichment analysis of all significantly 147 expanded gene families in pangolins was performed using Blast2GO (Conesa, Gotz, et al., 2005) (Figure 4.12). Among the significantly enriched GO terms are “translation” (174 genes; GO: 0006412; FDR p-

value=1.13E-60) and “ribosome biogenesis” (38 genes; GO: 0042254; FDR p-value=1.02E-10). The significant expansion of ribosome-associated gene families may be an advantageous selection for global upregulation of protein synthesis.

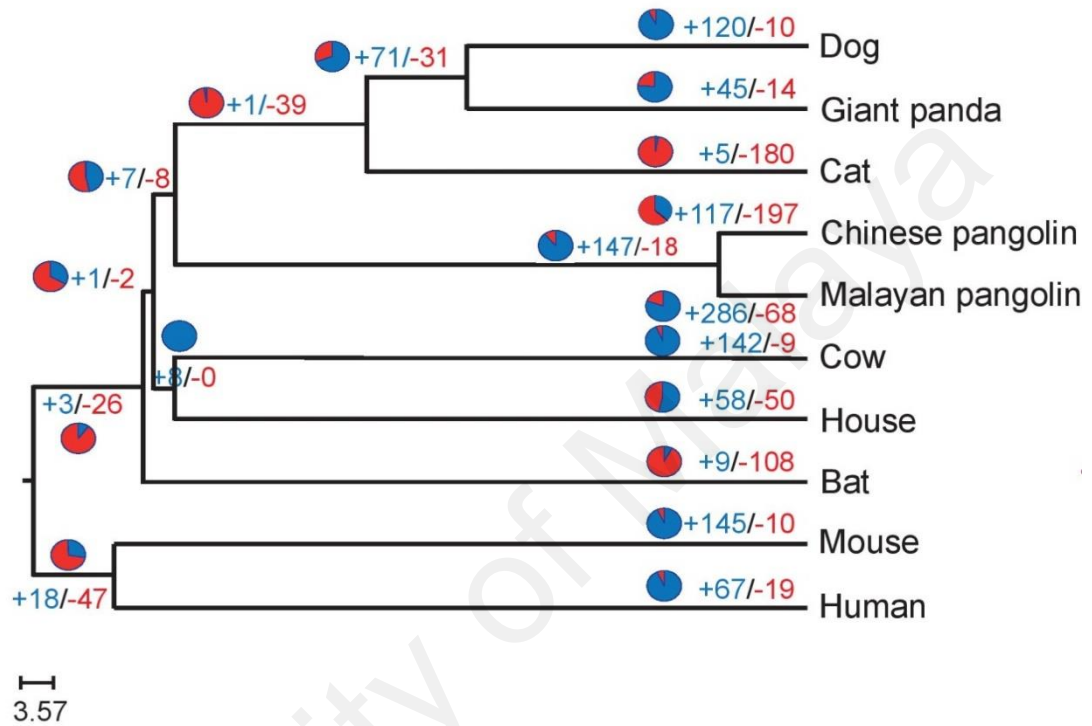


Figure 4.11: Phylogenetic tree and gene family expansion and contraction. Expanded gene families are indicated in blue, whereas contracted gene families are indicated in red. The proportion of expanded and contracted gene families is also shown in pie charts.

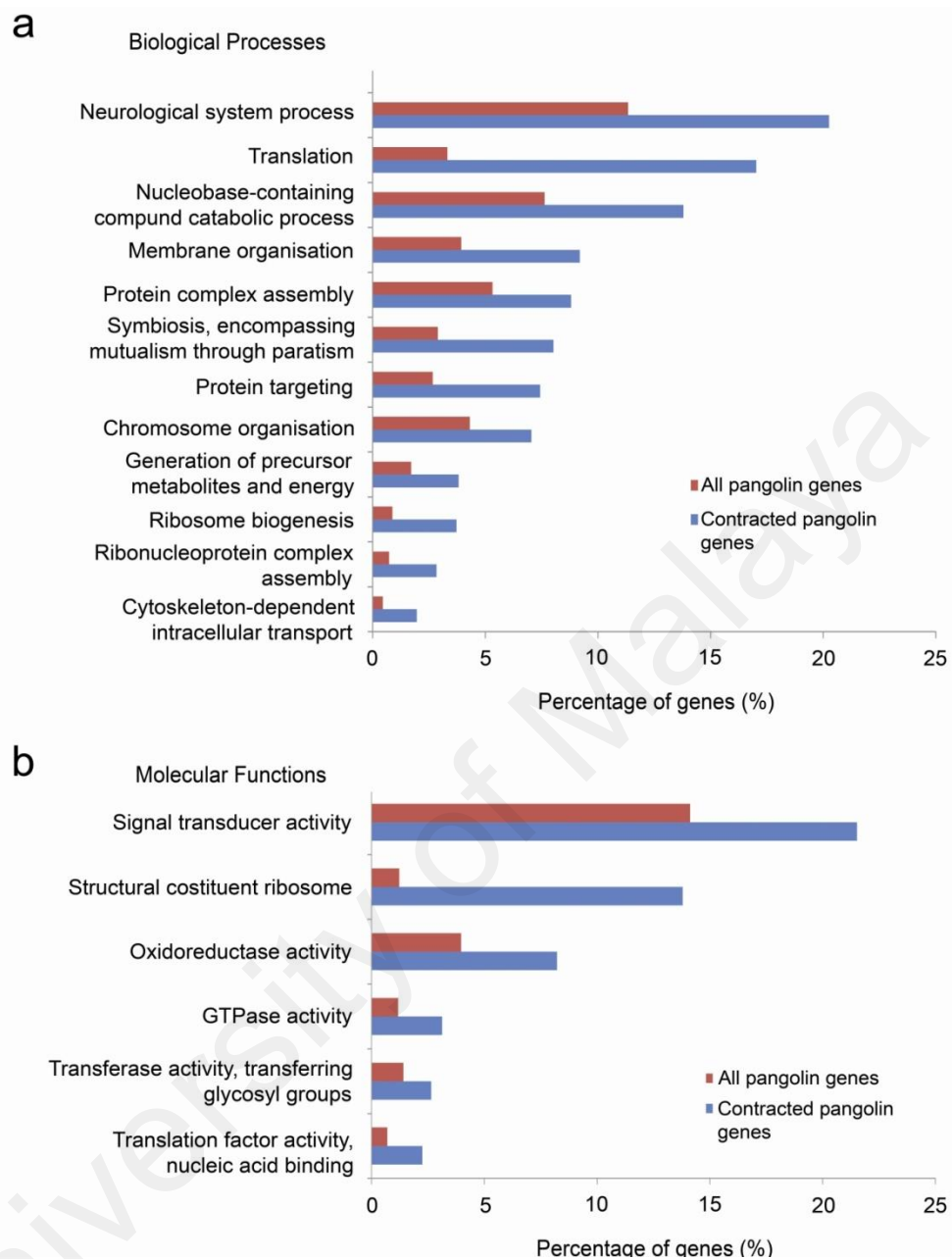


Figure 4.12: Functional enrichment of 147 significantly expanded gene families. Only Gene Ontology (GO) terms with a significant enrichment (Fisher Exact Test; FDR p-value<0.05) are shown. (a) Significantly over-represented biological processes. (b) Significantly over-represented molecular functions.

Notably, pangolins have a significantly reduced number of interferon genes, which have a major role in responding to infections, inflammation and healing of skin (Fung *et al.*, 2013; Xi *et al.*, 2012). Of the ten genes in the interferon family, only three were found in Malayan pangolin and two in Chinese pangolin (Figure 4.13). The heat shock gene family also

contracted significantly, possibly contributing to the sensitivity of pangolins to stress (Hua *et al.* 2015). However, gene expansion in several important families were documented, including ribosomal genes (17 families), olfactory receptor (OR) genes (6 families), cathepsin genes (1 family), and septin genes (1 family) (Figure 4.14), which were recently shown to deter microbial pathogens and preventing them from invading other cells (Mascarelli, 2011).

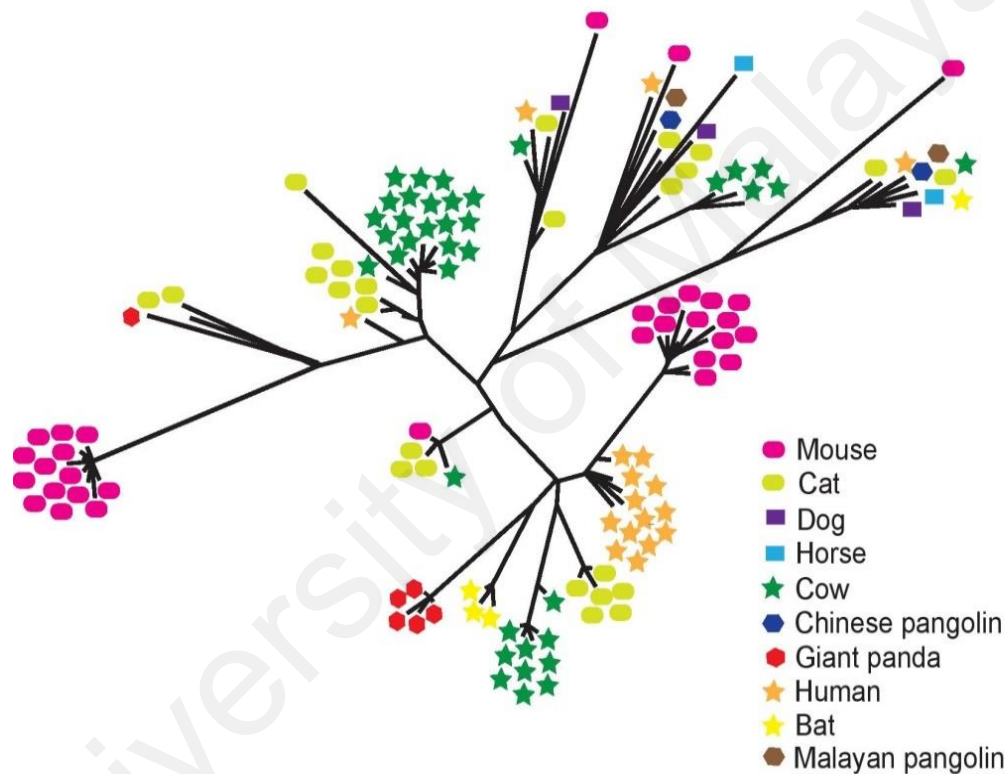


Figure 4.13: Phylogenetic tree shows the significant contraction of the interferon gene family in pangolins. Two interferon gene families found in both Malayan pangolin and Chinese pangolin indicates significant gene family contraction compared to other mammals.

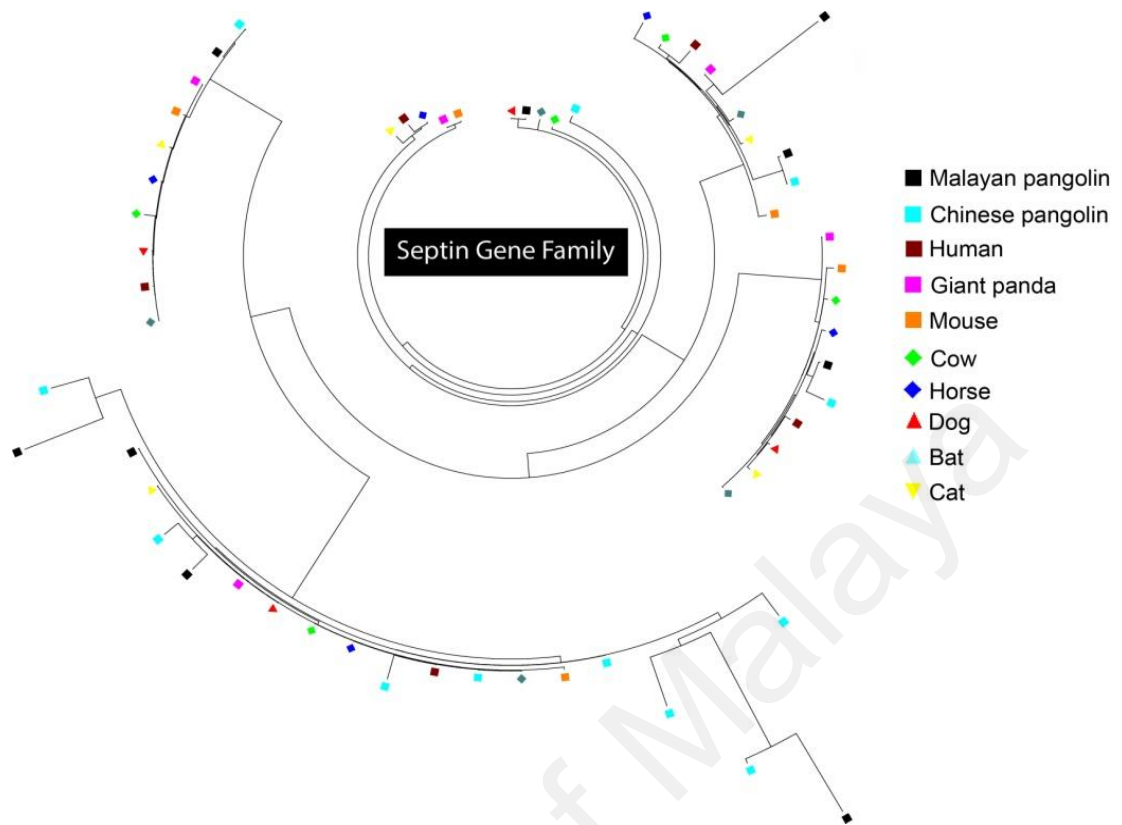


Figure 4.14: Expansion of the Septin gene family in the pangolin lineage. Septin was expanded in the Malayan pangolin (9 genes; Black Square) and Chinese pangolin (12 genes; light Blue Square).

The significantly expanded OR gene families suggest that pangolins have an enhanced sense of smell, possibly helping locate prey and counterbalancing poor vision (Figure 4.15). Strikingly, functional enrichment analysis of the 147 expanded gene families reveals these genes are significantly over-represented in neurological system processes (including neuromuscular process and sensory perception) (207 genes; GO:0050877; FDR p-value=6.43E-14), which is consistent with our finding in the functional analysis of pangolin-specific genes (Figure 4.12). The gene families associated with symbiosis and host-parasite interactions were also over-represented (82 genes; GO: 0044403; FDR p-value=2.67E-13).

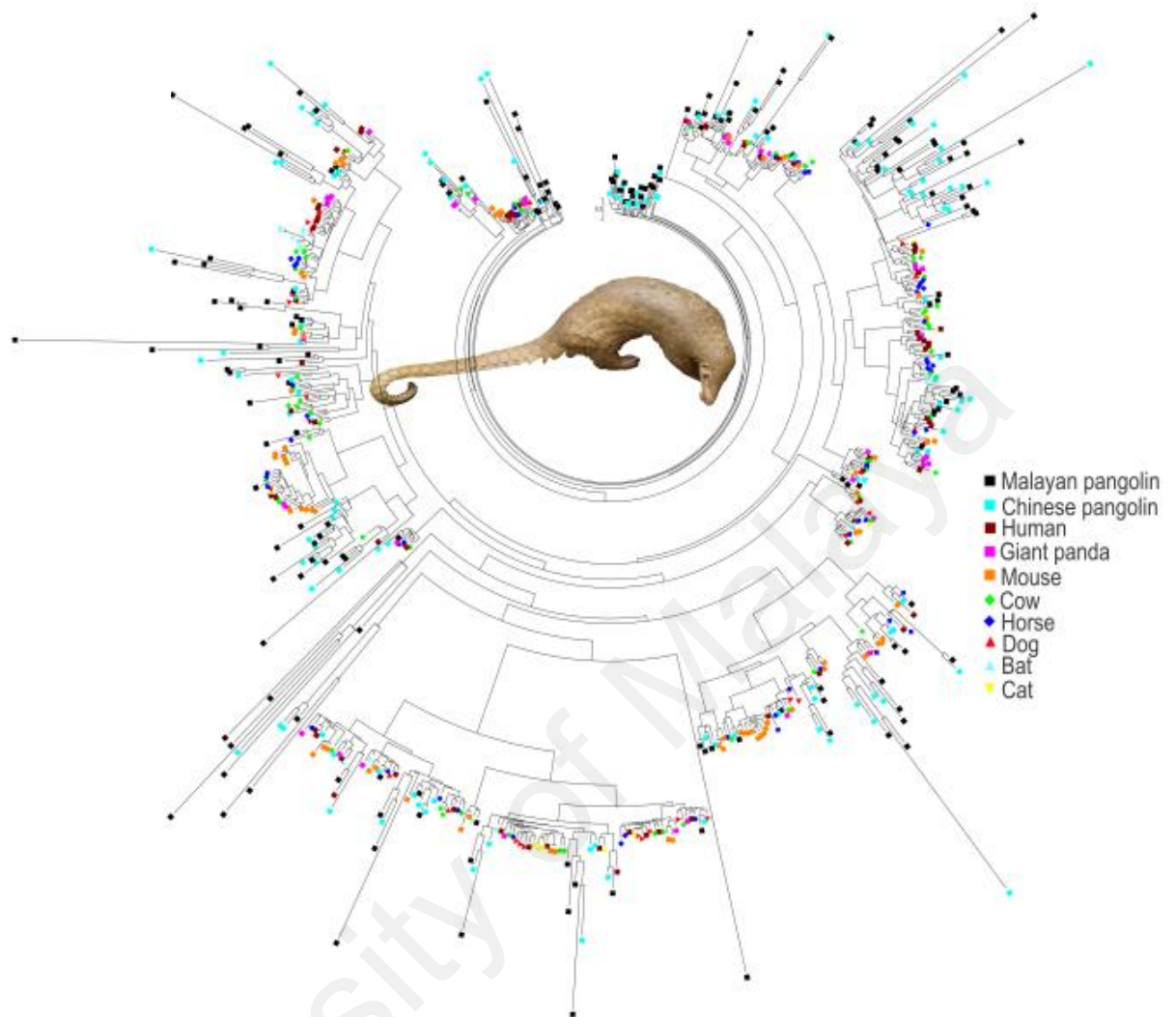


Figure 4.15: Expansion of Olfactory Receptor gene family 2 in the pangolin lineage. The protein sequences of the Olfactory Receptor (OR) gene family 2 used in this analysis were derived from expanded orthologous gene families in the pangolin lineage. The OR genes in the Malayan pangolin and Chinese pangolin are represented by a black square and light blue square, respectively. The olfactory receptor genes in both pangolin species were significantly expanded compared to others mammalian species including human, giant panda and mouse.

4.3.3 Positive selection analysis

To identify signatures of natural selection, a set of 8,250 protein-coding orthologs shared among the dog, cat, panda, cow, horse, mouse, human and megabat genomes were selected. Branch site tests identified evidence of positive selection along the pangolin lineage in 427 genes that had significant signals ($p < 0.05$ adjusted; Appendix B).

Hypothetically, the reduced of IFN-mediated immunity from the loss of *IFNE* and the contraction of interferon gene family in pangolins imposed strong selective pressure on immunity-related genes. A large proportion of genes were identified under selection in the pangolin lineage, which involve a broad range of immunity-related pathways including hematopoietic cell lineage, cytosolic DNA-sensing pathway, complement and coagulation cascades, cytokine-cytokine receptor interaction and the phagosome pathway (Table 4.3).

For instance, in the hematopoietic cell lineage, the colony stimulating factor 3 receptor (granulocyte) (*CSF3R*) gene encodes a transmembrane receptor for a cytokine (granulocyte colony stimulating factor 3) (Figure 4.16). *CSF3R* proteins present on precursor cells in the bone marrow, playing important roles in the proliferation and differentiation into mature neutrophilic granulocytes and macrophages which are essential for combating infection (Liongue & Ward, 2014). As predicted by Protein Variation Effect Analyzer (PROVEAN), *CSF3R* protein contains a critical pangolin-specific mutation at position 247 (G->H) in the functionally relevant Fibronectin Type III domain that was likely affecting its protein function in pangolins (Figure 4.18).

Table 4.3: List of the selected genes under positive selection in pangolins with their assigned functions or pathways.

Function	Genes	Function	Genes
Immunity	<u>Hematopoietic cell lineage</u> <i>CD44, ITGA3, ITGAM, CD2, ANPEP, CSF3R</i> <u>Cytokine-cytokine receptor interaction</u> <i>LTBR, IL2RB, TNFRSF13B, IL13, BMPR1B, CSF3R, OSM</i> <u>Complement and coagulation cascades</u> <i>F2, F2R, C8A, A2M, MASP2</i> <u>Phagosome pathway</u> <i>ATP6V0A2, MPO, RAB5B, ITGAM, CYBB</i> <u>Ameobiasis pathway</u> <i>LAMB2, COL1A2, RAB5B, COL4A4, ITGAM, C8A</i> <u>Cytosolic DNA-sensing pathway</u> <i>TMEM173, RIPK3, POLR3G</i> <u>Lysosome pathway</u> <i>FUCA1, ATP6V0A2, ACP2, NPC2, ABCA2</i> <u>Others</u> <i>LTF, LY86, TNFRSF13B, MASP2, TIGIT, AKNA, HCLS1, TREM1, MAPK9, SIAE, ITGB7</i>	Inflammation	<i>LPCAT2, RAB18, CAPN5, CORO2A, LBP, ADAM17, REL, ITGAM, RIPK3, IL13, P2RX7, PTGER2, IL2RB, MPO, LTF, AHSG, LTBR, TREM1, LBP, F2R, ANGPT2, OSM, MASP2</i>
Energy storage and metabolism	<i>GP2, ASXL1, SLC19A1, SLC27A4, ACSM3, SLC35A2, HK3, CYP7A1, AHSG, CTBS, TPH1, ANPEP</i>	Nervous system	<i>KIF1B, DOCK6, UGT8, PRX, MAP2, KCNH4, MERTK, LAMB2, PNKP, RTN4</i>
Mitochondrial metabolism	<i>ISCA2, SDHC, BCAT2, LETM1, ACAT1, GTPBP3, NDUFS4, FOXJ3, MRPS31, SLC25A19, HOGA1</i>	Osteogenesis	<i>COL1A2, MSH3, BMPR1B, SPP2, ACAN</i>
Muscular	<i>SGCG, NEBL, LDB3, FLNC, NEK8, LGALS2, RTCA, CACNG1, MYOZ3, TXLNB</i>	Apoptosis	<i>BIRC5, BIRC7, TOPBP1, APAF1</i>
Hair/Scale	<i>KRT75, KRT82, KRTAP3-1, SYNE4, KRT36</i>	Skin	<i>P2RX7, KLK5</i>
Eye	<i>KRT12, ADAMTS14</i>	Olfaction	<i>ARHGEF40, NCDN</i>
Stress	<i>USP15, RPA3</i>	ECM-receptor interaction pathway	<i>LAMB2, ITGB7, COL1A2, CD44, ITGA3, COL4A4</i>
Regulation of actin cytoskeleton pathway	<i>ITGB7, F2, ITGA3, WASF1, FGF6, ITGAM, ENAH, F2R</i>	Homologous recombination	<i>POLD3, RPA3, BRCA2, RAD51B</i>
Mismatch repair	<i>RFC5, POLD3, RPA3, MSH3</i>		

Another gene, Integrin alpha M (*ITGAM*) mediates leukocyte activation and migration, phagocytosis and neutrophil apoptosis. It has been reported that *ITGAM* can migrate neutrophils across intestinal epithelium to maintain intestinal homeostasis and eliminate pathogens that have translocated across the single layer of mucosal epithelial cells (Parkos *et al.*, 1991). Two pangolin-specific amino acid changes were identified in the functional Integrin alpha-2 domain of *ITGAM* which likely affect protein function (Figure 4.18).

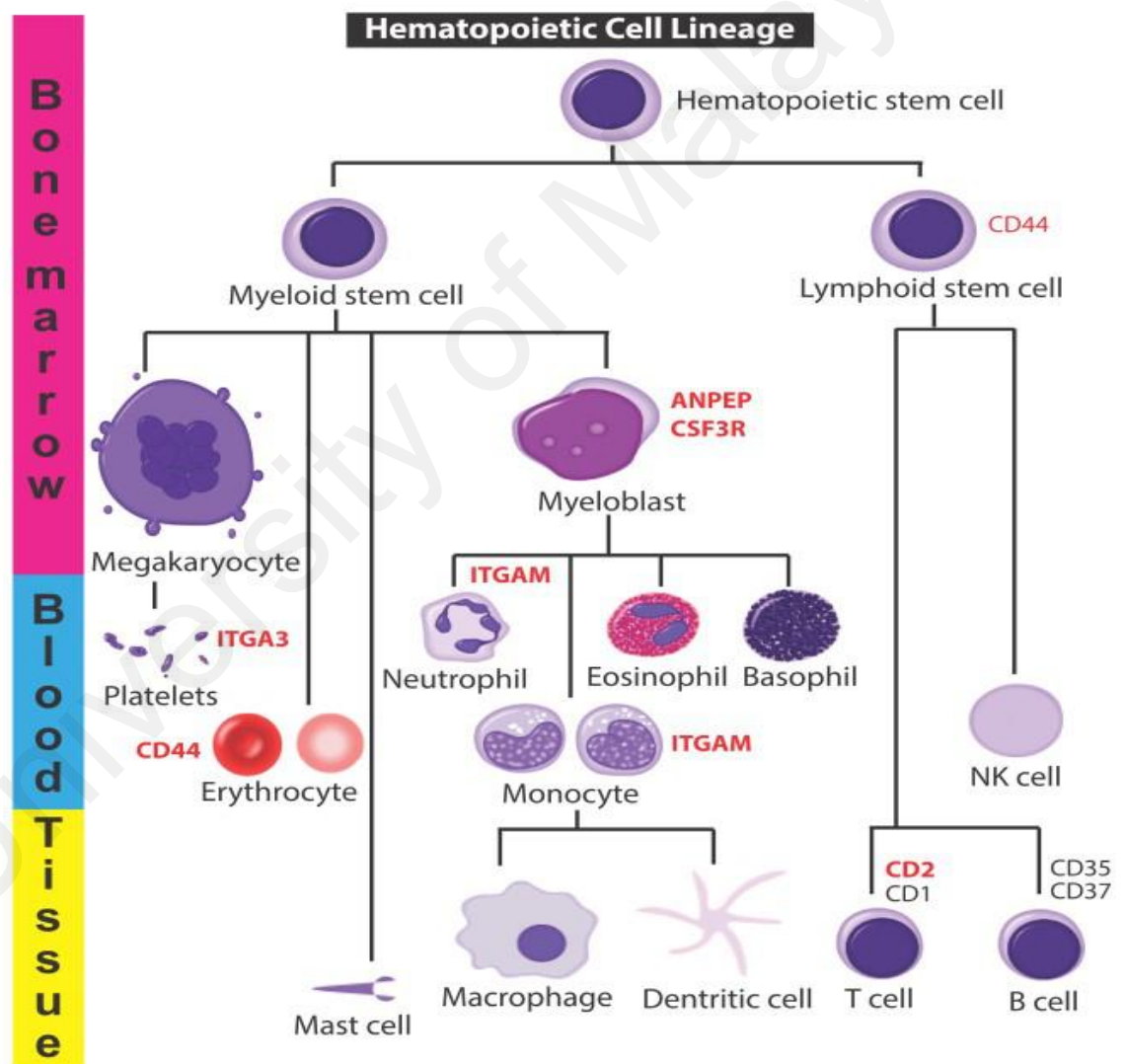


Figure 4.16: Evolution in the immunity-related pathways. Genes under positive selection at the pangolin lineage in the hematopoietic cell lineage are highlighted in red.

In the cytosolic DNA-sensing pathway, transmembrane protein 173 (*TMEM173*), also known as Stimulator of interferon genes, plays a major role in innate immunity by sensing cytosolic foreign DNA and inducing type I interferon production when cells are infected with pathogens (Figure 4.17) (Ran *et al.*, 2014). Two critical amino acid changes were observed in pangolins at positions 216(D->G) and 217(P->L) in the functionally relevant region, suggesting that they may have functional an impact on *TMEM173* (Figure 4.18). For the *CSF3R*, *ITGAM* and *TMEM173*, p-values of the detection of positive selection were 1.58E-002, 3.75E-002 and 2.92E-005, respectively.

Cytosolic DNA-sensing pathway

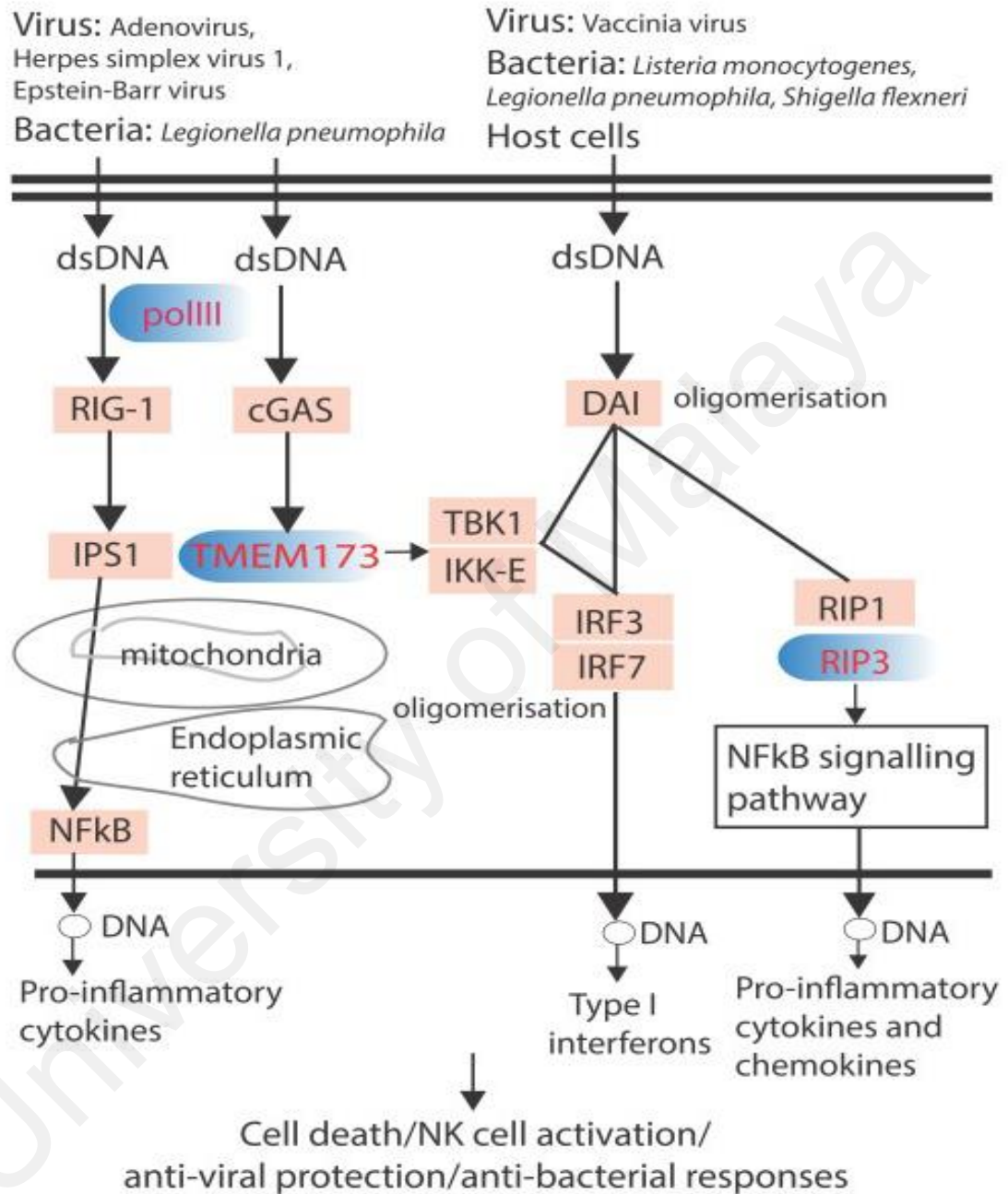


Figure 4.17: Evolution in the immunity-related pathways. Genes under positive selection at the pangolin lineage in cytosolic DNA-sensing pathway are highlighted in red.

The genetic changes during the evolution of hair-derived pangolin scales are likely to involve genes related to hair formation in general, and particularly keratins, which are an essential component of scales and hairs (Meyer *et al.*, 2013). Among the positively-selected candidate genes associated with skin formation are *KRT36*, *KRT75*, *KRT82* and *KRTAP3-1*. For instance, Type II Keratin 75 (*KRT75*), a hair follicle-specific keratin, has an essential role in hair and nail integrity (J. Chen *et al.*, 2008). Type I Keratin 36 (*KRT36*), a hair or “hard” sulfur-rich keratin, is mainly responsible for the extraordinarily high degree of filamentous cross-linking by specialized keratin-associated proteins (Moll *et al.*, 2008). Many critical pangolin-specific amino acid changes were identified in keratin genes in functionally relevant domains Figure 4.19. These key amino changes are also present in both *KRT36* and *KRT75* genes in the distantly related African pangolin species, suggesting that they potentially contributed to the development of pangolin scales (Figure 4.19 and Figure 4.20).

Genes related to energy storage and metabolism (12 genes) and mitochondrial metabolism (11 genes) were observed positively selected, perhaps an adaptive response to reduce their metabolic rate given their large body size, but low energy diet of ants and termites (da Fonseca *et al.*, 2008). Other candidate genes are associated with the development of the robust muscular system of pangolins which assists its fossorial lifestyle (10 genes) and the nervous system (10 genes) (Table 4.3).

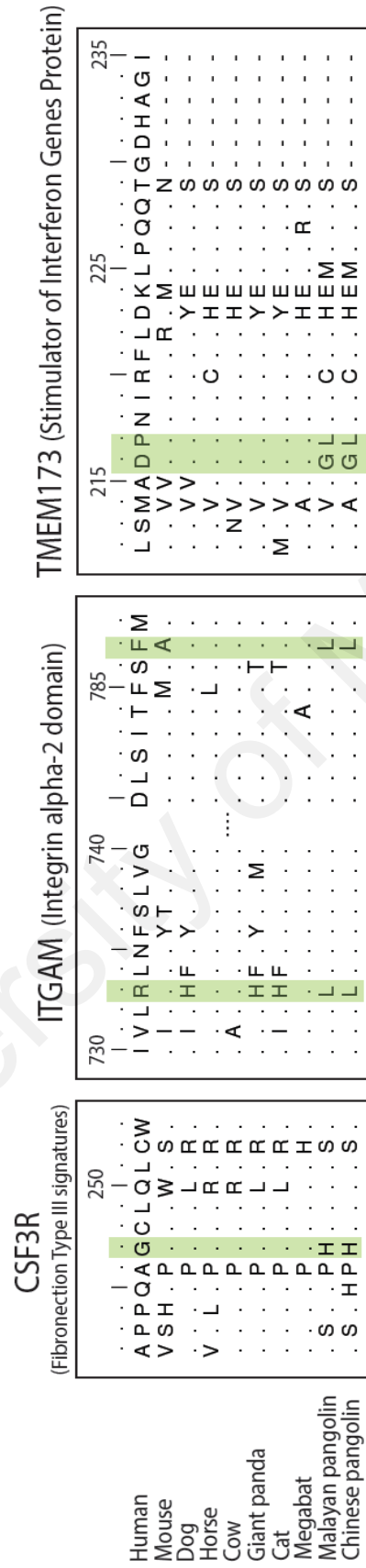


Figure 4.18: Evolution in the immunity-related genes. Several critical pangolin-specific mutations were identified in the functionally relevant Fibronectin Type III signatures (INTERPRO ID=IPR003961) of CSF3R, the Integrin alpha-2 signatures (IPR013649) of ITGAM and the Stimulator of Interferon Genes Protein region (INTERPRO ID=IPR029158) of TMEM173.

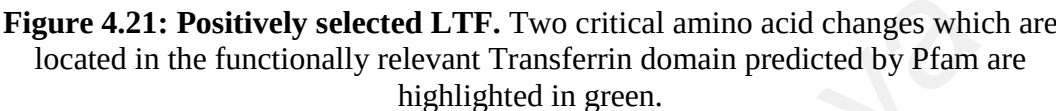
	KRT36	KRT75
<i>Homo sapiens</i>	LNVSSEQLQCC	DVDAAYMNKVELEAKVKSLPEEINF
<i>Manis temminckii</i> (4)	LNVMSEQLQSD	DTDAAY-NKVE-----SLTYEINF
<i>Manis tetradactyla</i> (2)	LNVMSEQLQSY	DADSAYVNKVELESRVNSLTDEINF
<i>Manis tricuspis</i> (1)	LNVMSEQLQSY	DADSAYVNKVELESRVNSLTDEINF
<i>Manis javanica</i>	LSVMSEQLQSY	DADAAAYVDKVELESRVNSLTDEINF
<i>Manis pentadactyla</i>	LSVMSEQLQSY	DADAAAYVNKVELESRVNSLTDEINF

Figure 4.20: The keratin-related proteins in African pangolin species. (Left) KRT36 protein sequence alignment. At least two critical amino acid changes observed in the Malayan and Chinese pangolins were identical in three African species examined. In *M. temminckii*, the C->D amino acid change likely affect the biological function of KRT36 (score=-8.461) as predicted by PROVEAN (Choi & Chan, 2015). (Right) Two critical amino acid changes observed in the Malayan and Chinese pangolins were identical in *M. tricuspis* and *M. temminckii*, but not *M. tetradactyla*. In *M. tetradactyla*, the changes of V->T and M->deletion are also likely to affect the function of KRT75 as predicted by PROVEAN with scores of -3.459 and -12.790, respectively. No data are available for *M. gigantea*.

Furthermore, a significant number of genes were observed associated with inflammation (18 genes), cancer or viral infection (25 genes), bacterial infections (10 genes), pneumonia and gastrointestinal diseases (21 genes) and skin diseases (15 genes) (Table 4.4).

Interestingly, alterations in the Lactotransferrin (*LTF*) gene was found. *LTF* encodes a multifunctional immune protein normally presence at mucosal surfaces, providing the first line of defense to the host against inflammation and infection (Ward *et al.*, 2002). It has been shown that *LTF* has antimicrobial activities and induces both systematic and mucosal immune responses, for example, to against lung and gut-related systemic infections (Debbabi *et al.*, 1998). Two pangolin-specific amino acid changes were detected in the *LTF* protein sequence, potentially having an impact on the LTF function (Figure 4.21).

Together, evidence of selection on these genes provides possible links with the unique phenotypes and physiological requirements of pangolins. Importantly, the immunity- or inflammation/infection- related genes provide avenues to explore the evolution of mammalian immunity further.



Function	Genes	Function	Genes
Cancer or viral infection	<i>KLK5, ABCB5, TOPBP1, SDHC, ST14, RET, MYBL2, HYAL2, KIAA1524, SLC19A1, CYP1A1, SH2D1A, BRCA2, BIRC7, ITGA3, MST1R, BIRC5, KLF4, MSH3, RAD51B, APAF1, CD44, POLH, DDIT3, CA9</i>	Inflammation	<i>ADAM17 REL ITGAM RIPK3 IL13 P2RX7 PTGER2 IL2RB MPOLTF AHSG LTBR TREM1 LBP F2R ANGPT2 OSM MASP2</i>
Bacterial infection	<i>SLC37A4, TREM1, ITGAM, IL13, P2RX7, LBP, MPO, LTF, MASP2, CYBB</i>	Pneumonia	<i>LBP, TREM1, MPO, SFTPC, IL13, CYBB, MASP2</i>
Skin diseases	<i>KLK5, BRCA2, CYP1A1, MST1R, ST14, THEMIS2, HPS4, IL13, RAD51B, GJB3, ASIP, KRT75, CD44, POLH, DSG3</i>	Gastrointestinal diseases	<i>CYP1A1, MST1R, BIRC5, KLF4, RET, MSH3, DKK4, GUCY2C, PTGER2, CD44, MST1, ANO1, MUC13, CA9</i>

CHAPTER 5: DESIGN AND DEVELOPMENT OF PANGOLIN GENOME DATABASE (PGD)

5.1 Introduction to Pangolin Genome Database (PGD)

The GIRG has sequenced the genomes of the Malayan and the Chinese pangolin from Malaysia and Taiwan, respectively, using High-throughput next-generation sequencing (NGS) approaches. Further, the transcriptomes of different organs of the female Malayan pangolin were also sequenced. To facilitate future research, I have implemented a Pangolin Genome Database (PGD), to function as a central hub for hosting genomic resources and information of pangolins, together with useful analysis tools for the research community. PGD provides interactive and user-friendly web interfaces for users to search, browse, retrieve, visualize and analyze pangolin genomic and transcriptomic data and information. The PGD can serve as a computational platform for researchers to advance research in this field particularly in the area of conservation of this unique endangered mammalian species and also to provide a useful resource for research in mammalian evolution and biology, in general. The PGD database can be freely access using this URL: <http://pangolin-genome.um.edu.my>

5.2 Database sources

PDG currently hosts all data, results and information about pangolins generated by the International Pangolin Research Consortium (IPaRC) (website: <http://pangolin.um.edu.my>). Currently, IPaRC has the reference genome sequences of Malayan pangolin and Chinese pangolin. The Malayan pangolin genome was sequenced using Illumina HiSeq 2000 at BGI, Hong Kong and assembled using SGA-0.10.10 (Simpson & Durbin, 2012) (coverage=146X). The assembled contigs were scaffolded using SOAPdenovo2 scaffolder and achieved N50 of 204,525bp, assembled scaffold

genome size of 2.5 Gbps (Choo *et al.*, 2016). For the Malayan pangolin, GIRG has sequenced the transcriptomes from different organs including cerebrum, cerebellum, liver, heart, lung and thymus. The sequenced Malayan pangolin specimen was provided by the Department of Wildlife and National Parks Malaysia and sequenced and assembled by the Genome Informatics Research Group (GIRG), University of Malaya, Malaysia. The Chinese pangolin specimen was from Taiwan and sequenced and assembled by the Warren Research Group from the Genome Institute of Washington University, USA and analyzed by Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg, Russia. The Chinese pangolin genome (coverage=56X) sequenced using Illumina platform, assembled using SOAPdenovo v1.0.5 and scaffolded using L_RNA_scaffolder approach achieving N50 of 157,892, assembled scaffold genome size of 2.2Gbp (Choo *et al.*, 2016). The whole-genome sequences have also been deposited in GenBank under the accession numbers JSZB000000000.1 (*M. javanica*) and JPTV000000000.1 (*M. pentadactyla*) whereas the RNA-Seq reads are accessible via the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) under the accession number SRP064341.

5.3 Genome data and functional annotation

The whole-genome of both pangolin species were annotated using a well-established MAKER pipeline, which was designed for the annotation of emerging model genomes using evidence from *ab initio* gene prediction, sequence homology approach and the transcriptomic data from Malayan pangolin (Cantarel *et al.*, 2008). MAKER is a well-established annotation pipeline for eukaryotic genomes. Briefly, repetitive sequences in both pangolin genomes were identified and masked using RepeatMasker(Shirak *et al.*, 2010) and RepeatRunner(C. D. Smith *et al.*, 2007).

After the repeat masking process, the RNA-seq data from Malayan pangolin were mapped to the genomes as first evidence and the canine protein sequences from Ensembl (Cunningham *et al.*, 2015) used as a reference for protein homology. This automated pipeline integrates the evidence from its gene prediction algorithm and produces high-quality gene models in subsequent runs. The gene models were *de novo* predicted by Augustus (Stanke & Morgenstern, 2005), GenScan (Burge & Karlin, 1997) and SNAP (Korf, 2004) and integrates with the homologs and RNA-seq evidence to generate the final set of gene models. By using a stringent procedure, 23,446 genes were identified in the Malayan pangolin genome and 20,298 genes in the Chinese pangolin genome. These genes were further annotated using BLAST2GO, allowing a better understanding of the function of each gene. BLAST2GO uses BLAST algorithm on the annotated protein sequence query to find homologs (Conesa, Göt, *et al.*, 2005). Then, a straightforward mapping is performed to retrieve the gene ontology (GO) term with the obtained hits. In order to annotate the functional domains/signatures, different biological databases were used such as GO modulate, GoSlim (Ashburner *et al.*, 2000), Enzyme Code annotation with KEGG (Kanehisa & Goto, 2000) maps and also InterPro (Hunter *et al.*, 2009) annotation with the default parameters. BLAST2GO was assigned functions to 21,451 (91%) genes of Malayan pangolin and 19,287(95%) genes of Chinese pangolin (Table 5.1).

5.4 Transcriptome data

PGD stores the sequences of expressed genes from different organs, namely the cerebellum, cerebrum, liver, heart, kidney, lung, thymus, and spleen of the Malayan pangolin. Briefly, these transcriptomes were sequenced using an Illumina HiSeq 100bp Paired End (PE) sequencing strategy. The sequencing reads from all organs were pooled to generate a representative catalog of pangolin genes, preprocessed and *de novo*

assembled using three different approaches: SOAPdenovo (Xie *et al.*, 2014), Velvet (Zerbino & Birney, 2008) and Trinity (Haas *et al.*, 2013) software. To generate a high-quality set of pangolin genes, common assembled transcripts or genes predicted by the three different assemblers will only be considered. This resulted in a set of 89,751 genes in the Malayan pangolin transcriptome (Table 5.1).

5.5 Pseudogene annotation data

To the best of our knowledge, the Malayan and Chinese pangolin genomes are the first whole-genome sequencing efforts for the Pholidota order. Therefore, information regarding pseudogenes are still not available. Here, PGD offers information about the putative pangolin pseudogenes. The whole-genome sequences were screened using MAKER-generated protein-coding genes (parent genes) as query sequences through the whole-genome using Pseudopipe pipeline (Z. L. Zhang *et al.*, 2006) to predict pseudogenes in the pangolin genomes. Pseudopipe is a well-established pipeline for whole-genome pseudogene screening (Karro *et al.*, 2007; Pei *et al.*, 2012; Sisu *et al.*, 2014). Protein sequences derived from MAKER annotation were used as a query to BLASTALL (Altschul *et al.*, 1990) through the whole-pangolin genomes. After all homologous BLAST hits had been reported, Pseudopipe removed all hits that belong to its parent genes and merged small gaps between two short hits and output as potential pseudogene candidates. These candidate pseudogenes were then screened using the tFasty (William R Pearson, 1999) tool to report the insertion, deletion, pre-mature stop codon and frameshift mutation events that may disrupt the normal functions of the genes and generates the final set of pseudogene dataset. To obtain reliable sets of pseudogenes, stringent thresholds were used to filter all the false positive candidates (1E-10 e-value, 70% parent gene coverage, 40% gene identity). All pseudogenes were reported as duplicated or processed pseudogenes based on the nature of the reported hits. There are

4,660 pseudogenes reported in Malayan pangolin and 2,416 pseudogenes Table 5.1) in the Chinese pangolin genome after the stringent filtering criteria.

Table 5.1: Summary statistics of two pangolin genomes and transcriptome assembly datasets in PGD.

Genome	Malayan Pangolin	Chinese Pangolin
Number of Scaffolds	81,732	87,621
Estimated Coverage (X)	146	56
Estimated Genome Size (bp)	2,492,544,425	2,205,289,822
N50 (bp)	204,525	157,892
# of protein-coding genes	23,446	20,298
# of annotated genes	21,451 (91%)	19,287 (95%)
# of pseudogenes	4660	2416
# of transcripts	89,751	NA

5.6 Database Graphical User Interface (GUI) Design

5.6.1 Database Overview

The homepage of PGD contains general descriptions of pangolins in the main panel and manually compiled information about pangolins such as latest news & conferences, blogs and published articles in the right side panel. On the top of the PGD homepage, several option tabs are provided, allowing users to access different features of PGD. For instance, the “Browse” tab allows users quickly to browse the detailed annotation data of Malayan and Chinese pangolins, as well as transcriptomic data for the Malayan pangolin. In addition, the “Tools” tab allows access to analysis tools such as BLAST (Altschul *et al.*, 1990) and BLAT (Kent, 2002) packages were incorporated into PGD. Users also can visualize the genome and genomic features using the unique real-time keyword search feature for fast and smooth searching of genes of interest (Figure 5.1).

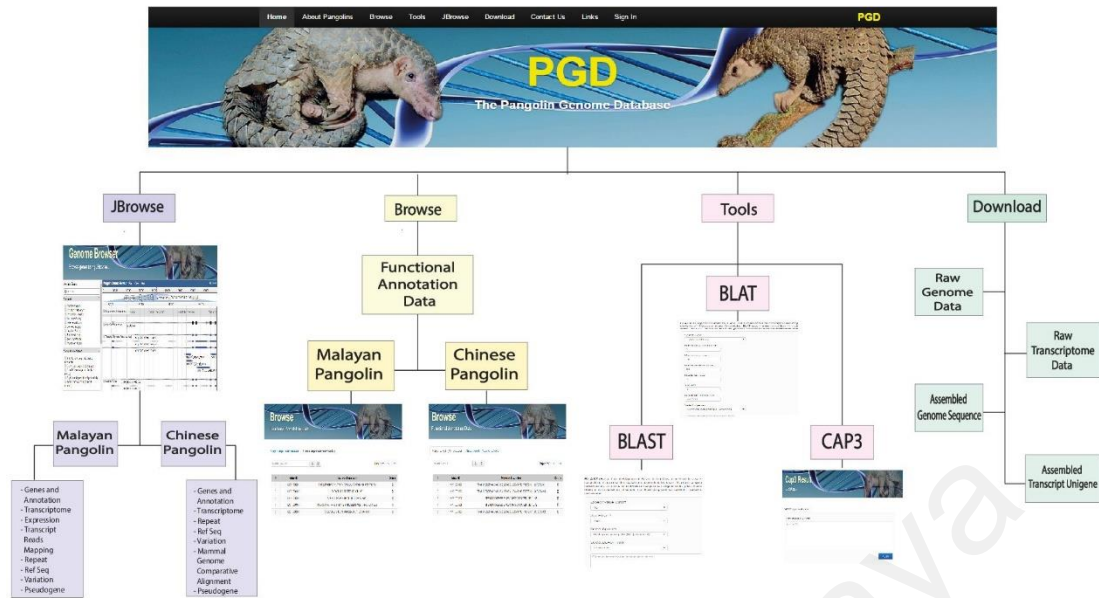


Figure 5.1: Schematic structure of the PGD.

5.6.2 Data browsing interface

By clicking on the 'Browse' tab on the top of PGD homepage, users can access all gene information of the both pangolin species (Figure 5.2). All genes will be displayed after users click on the tab. Users can access the gene details page containing the gene annotation and functional information of a particular gene by clicking on the "details" button associated with the gene. This gene details page will display all information about the gene of interest including sequence name, scaffold ID, start and stop position of the genes, putative functions, gene sequences and Blast2GO alignment results such as 'Top Hit Species', 'Min E-value', annotated 'GOs' information, 'InterPro Scan' domain information. JBrowse is also incorporated on this page, allowing users to visualize and explore the locus and neighboring regions of the gene within the pangolin genome.

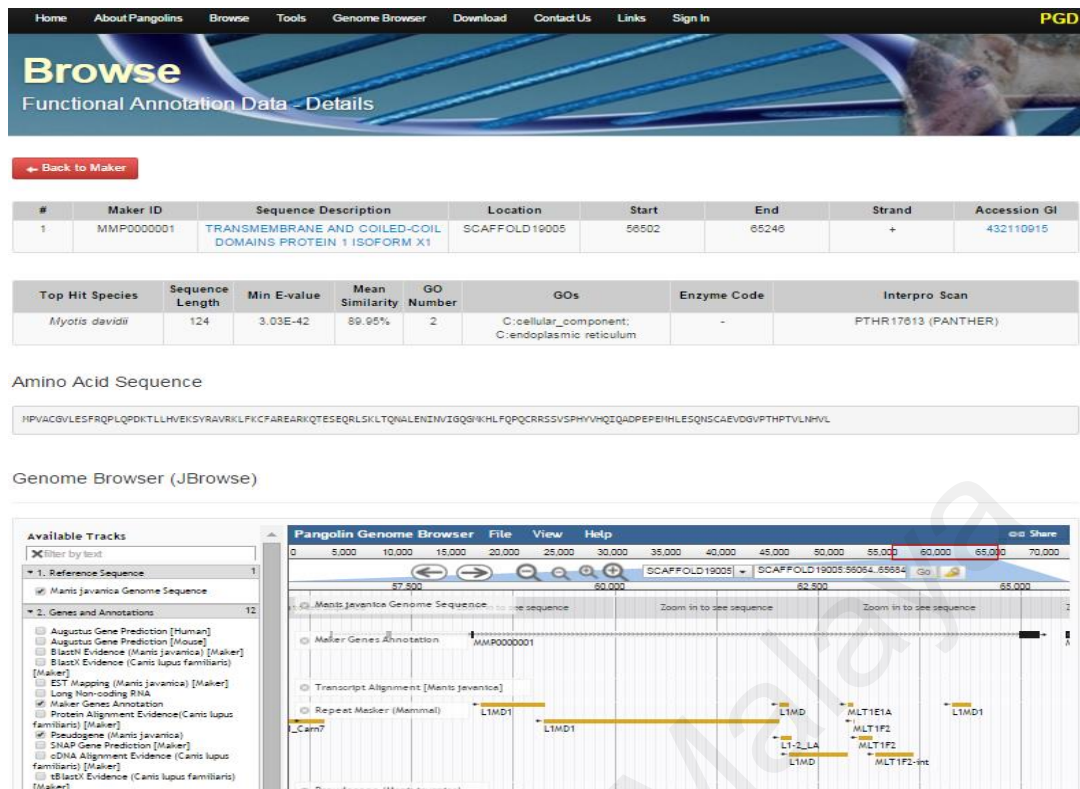


Figure 5.2: A screenshot of gene details page. This page will display information of a gene including its sequences and functional annotation.

5.6.3 Keyword and sequence search

PGD contains a vast amount of pangolin genomic, transcriptomic data and annotations. Therefore, it is necessary to provide intuitive web-based GUI allowing users to search a large volume of data rapidly. I have implemented a keyword search system in PGD to allow users to search genes of interest in a spontaneous manner. Alternatively, users can search by sequences. PGD built-in with BLAST (Altschul *et al.*, 1990) and BLAT (Kent, 2002) tools which allow users to search or compare a query sequence against the pangolin genome and transcriptomic data. Using the BLAST tools, users can perform nucleotide searches to each annotated sequence (BLASTN), whole genome nucleotide searches (BLAST Whole Genome), protein searches of each annotated sequence (BLASTP), and nucleotide searches of each annotated protein sequence (BLASTX). Users can choose the

pangolin species of interest (either Malayan pangolin or Chinese pangolin) and also set the desired cutoffs (e.g. expect value and enable to search for low compositional complexity regions) for their sequence searches. Alternatively, users can perform a sequence search using BLAT which is structured differently from BLAST to search the similarity in a query sequence, but it accepts an exact or nearly exact match to find the hit. Unlike BLAST which is a local alignment software, BLAT is also designed to map transcripts/ESTs that have no introns back to the reference pangolin genomes. In this case, users would be able to examine the location of this transcript and also its gene structure within the genome. By using the provided web-based BLAST, users are able to set the parameters such as (i) the minimum number of matches (ii) minimum score (iii) minimum identity percentage (iv) maximum gap (v) tile size (vi) maximum intron size. The BLAT outputs will be linked with the genome browser, allowing instant visual comparison of each match hit.

5.7 Multiple mammalian genome sequence alignments

Multispecies genome sequence alignments between two pangolin genomes and other mammalian genomes (dog, cat, cow, horse, human and mouse) were run using the Progressive Cactus software (Paten *et al.*, 2011). The animal genome data used in the alignment are shown in Table 5.2. Prior to performing the structural alignment procedure, repetitive regions of the assemblies were masked with RepeatMasker (Smit *et al.*, 1996). The phylogenetic tree used to perform the alignment is given in Figure 5.3.

Table 5.2: Genome assembly version for each mammalian genome used for multiple sequence alignment.

Animal	Scientific name	Genome assembly
dog	<i>Canis familiaris</i>	CanFam3.1
cat	<i>Felis catus</i>	Felis_catus_8.0
cow	<i>Bos taurus</i>	Bos_taurus_3.1
horse	<i>Equus caballus</i>	EquCab_2.0
human	<i>Homo sapiens</i>	GRCh37.p5
mouse	<i>Mus musculus</i>	GRCm38.p4

Base-wise conservation scores were obtained for the Malayan and Chinese pangolin genomes from the multispecies structural alignment using the HAL-phyloP tool from the HAL package (Hickey *et al.*, 2013). HAL-phyloP wraps the original phyloP algorithm (Pollard *et al.*, 2010) and is designed to process multispecies structural alignments efficiently by Progressive Cactus.

Coding DNA sequences (CDS) of Malayan pangolin genes were used to train a neutral evolution model for phyloP. Next, phyloP was launched in the conservation and acceleration (“CONACC”) mode to obtain the corresponding p-values for each position in both pangolin genomes. The common logarithms of the base-wise p-values were reported; positions under acceleration were flagged by making their values negative while positions under conservation had their values positive.

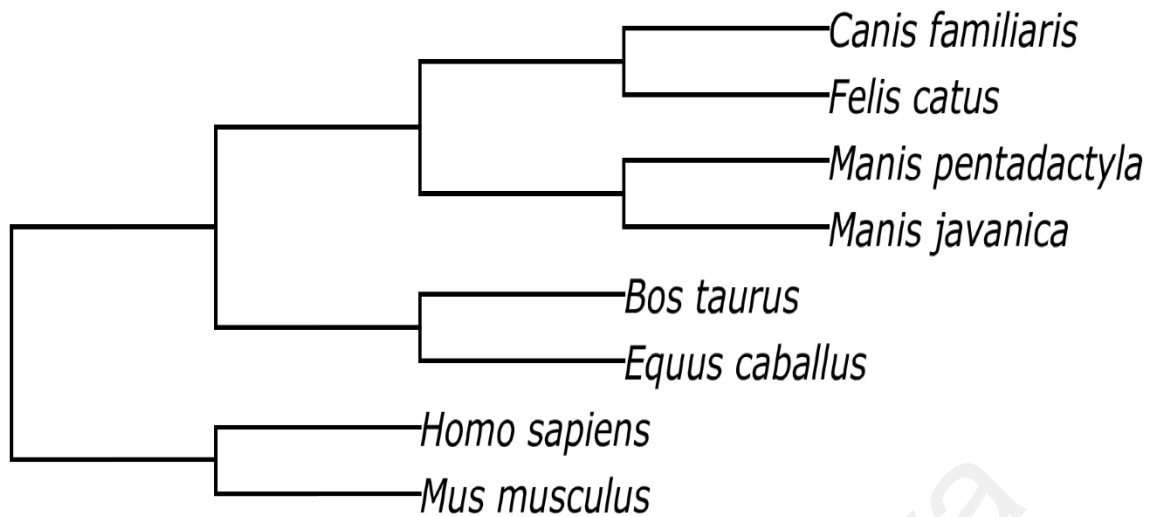


Figure 5.3: The phylogenetic tree of species involved into the structural alignment.

5.8 Pangolin Genome Data Visualization (JBrowse)

A fully dynamic AJAX-based real time genome browser, JBrowse (Skinner *et al.*, 2009), was incorporated into PGD, allowing users to browse instantly and visualize pangolin genome sequence and annotation data on the fly. JBrowse provides a few useful features: (i) supports fast and smooth genome navigation (ii) utilize multiple types of data format (iii) real-time genome browsing with light server resource required (iv) high speed visualizing the data, sequences and genome annotation results.

Using the pangolin genome browser (Figure 5.4), users can choose whether to visualize the pangolin data in Malayan pangolin genome or Chinese pangolin genome. Currently, more annotations (e.g. transcriptomic data) are available for the Malayan pangolin than the Chinese pangolin because many research projects are ongoing under the IPaRC of the Malayan pangolin species. Once users choose which genome to explore, they can see all relevant annotation tracks in the genome browser. For instance, a basic track as the reference genome sequence of the selected genome is available by default. Different annotation tracks that can be triggered on/off are available for easier visualization and analysis. These tracks are classified into various categories such as ‘Genes and

Annotations', 'Repeat', 'Transcriptome', 'Transcript Reads Mapping', 'Organ-Specific Expression' and 'Variation' which represent our annotations from different analyses.

5.8.1 Genes and Annotations tracks

This category will show the results of MAKER gene models including the evidence-based and de novo gene predictions, together with the predicted pseudogenes. In the pseudogene track, users can visualize the predicted pseudogenes in the selected pangolin genome. These pseudogenes are labeled with 'PSSD' and 'DUP' as categorized by PseudoPipe pipeline (Z. Zhang *et al.*, 2006). The pseudogene tracks report all useful features found in the pseudogenes, including parent gene coverage 'Frac,' identity 'Ident,' insertion 'INS,' deletion 'Del,' frameshift 'Shift,' and stop codon 'Stop.' This information will give a better understanding of pseudogene attributes found in both pangolins.

5.8.2 Repeat tracks

The repetitive element information is provided by different software which displayed in three different tracks: RepeatMasker, Repeat Modeler, and Structural Simple Repeat tracks. RepeatMasker track shows the repetitive elements screened in the pangolin genomes using two different repeat libraries: the carnivore repeat library and mammal repeat library. Meanwhile, I used SSRIT (Temnykh *et al.*, 2001) for scanning the simple sequence repeat or structural sequence repeat present in both pangolin genome. The SSRIT reports the entire simple sequence repeats and allows the user to study microsatellites in pangolin genomes.

5.8.3 Transcriptome tracks

Users can explore and visualize expressed genes in Malayan pangolin through the tracks under this category. The expressed transcripts generated by pooling all reads from the eight pangolin organs and assembled using three different assemblers are available here.

These common transcripts were mapped onto the reference genome using GMAP (T. D. Wu & Watanabe, 2005). Moreover, users can also explore and examine which transcripts/genes are expressed in a specific pangolin organ such as cerebrum, cerebellum, and liver through the provided organ-specific transcripts tracks.

5.8.4 Transcript Read Mapping tracks

Users can also examine the expression level of each gene of Malayan pangolin across different organs. Briefly, for each organ or organ-specific transcriptome, RNA-seq reads were mapped to the Malayan pangolin genome sequence and the reads mapping data of eight different organs (cerebrum, cerebellum, liver, heart, kidney, thymus, spleen, and lung) were available in the 'Transcript Reads Mapping' category.

5.8.5 Transcriptome Profile tracks

The raw transcript reads mapping results of eight different pangolin organs were normalized using RPKM value and converted into a graphical bigwig format track that visualized the expression profile of each organ. Users are able to examine the expression profiles of a gene of interest across eight different organs simultaneously. Of note, the PGD genome browser also links some features (e.g. genes and repeats) to external resources. Among of these resources are RepeatMasker database for each repeats elements, BLASTX evidence tracks is linked to well-establish databases to check each hit on ENSEMBL database, PANTHER database(Thomas *et al.*, 2003), Dog Genome SNP database (Bai *et al.*, 2015) and UniProt (Chan & Consortium, 2010) by simply right clicking on the features in the tracks. These external database interlinks may help users understand further information of the features.



Figure 5.4: Pangolin genome browser. Users can turn on and off the annotation tracks in the left panel

5.9 Assembly hub for UCSC Genome Browser

The UCSC Genome Browser (Kent *et al.*, 2002) is a well-known and widely used tool for bioinformatics analysis. Alongside with the PGD database, I integrated an assembly hub (Raney *et al.*, 2014) for the UCSC Genome Browser that contains datasets from PGD and provided means for viewing them in the genome browser. The hub also presents the structural alignment between genomes of both pangolin species and other mammals in the form of the snake tracks (Nguyen *et al.*, 2014) and the genomic conservation tracks obtained from the alignment with the HAL-phyloP tool.

In addition to presenting the pangolin genome annotation tracks, the hub provides access to other features and tools available in the UCSC Genome Browser, including Table Browser (Karolchik *et al.*, 2004) for manipulating the tracks and obtaining nucleotide sequences of the features.

JBrowse and UCSC genome browser are the dual-genome browser system that integrated into PGD as JBrowse is excellent in real-time genome data visualization that provides smooth and fast experience to users for their data comparative analysis (Buels *et al.*, 2016). However, the short-hand of JBrowse for not compatible for some of the large alignment data in a certain format. Hence the UCSC genome browser may help to visualize those data format where JBrowse not supported. Therefore, PGD provides dual-genome browser system for best users' experiences in genome visualization and comparative analysis.

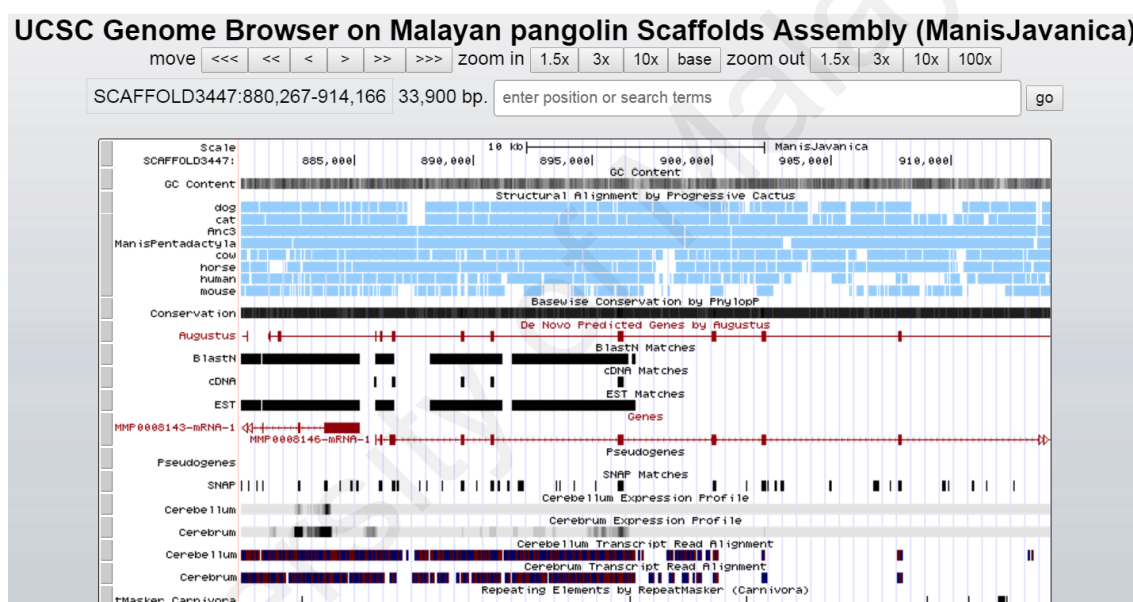


Figure 5.5: Interface for UCSC Genome Browser (*M. javanica* genome).

5.10 Data download

PGD provides a user-friendly interface for downloading pangolin genome and transcriptomic data and annotations. The raw sequencing reads from different library sizes are available for download. Users also can download the assembled reference genome sequences of the Malayan and Chinese pangolin for downstream analyses. Other available data or annotations are available for download including the MAKER-generated

gene sequences (nucleotide CDS, exons and protein sequences), expressed genes from Malayan pangolin transcriptomes and the raw RNA-seq data.

There are two ways users can download these data and annotations. First, users can download these data through the ‘Download’ page using the provided web interface. Users can select the data or annotations of interest to download. Second, users can download the raw data and annotations through the FTP (File Transfer Protocol) interface provided on the download page (Figure 5.6).

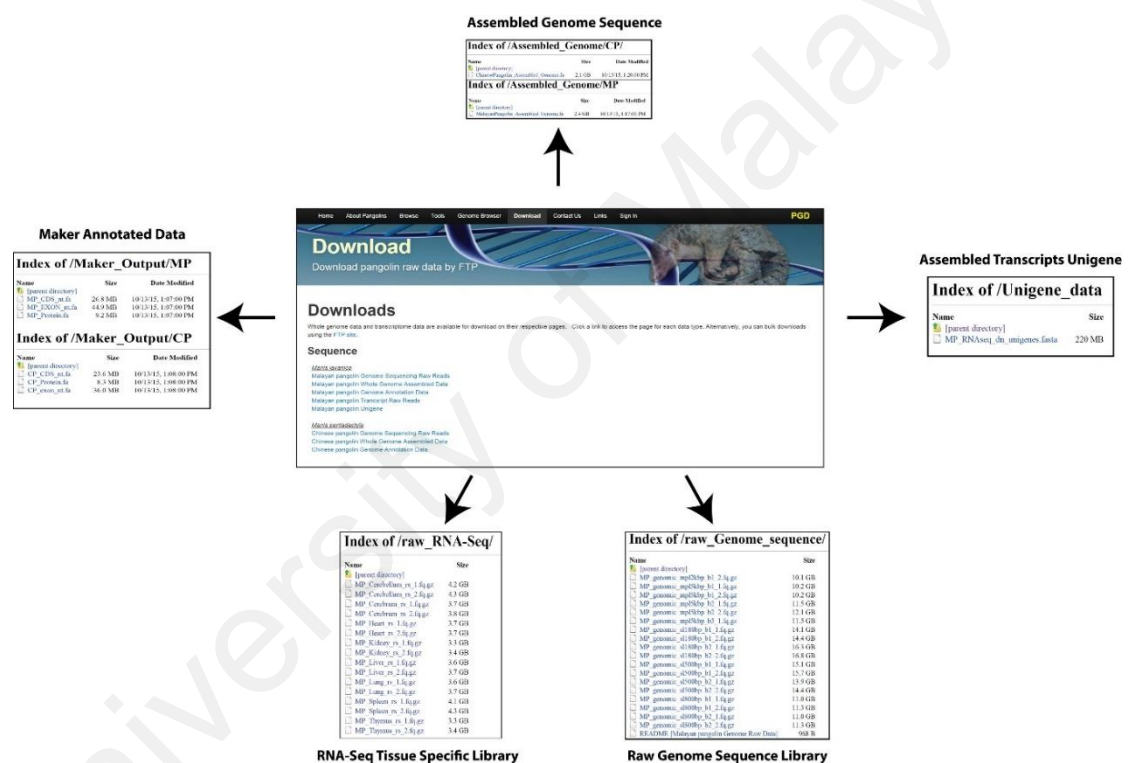


Figure 5.6: Web interfaces for data download in PGD.

CHAPTER 6: DISCUSSION

6.1 Unique traits of pangolins and their genetics

Pangolins are unique mammals that are characterized as being edentulous and covered with keratinized scales that resemble human hair and nails. Pangolins do not usually survive in captivity and it has been suggested that this could be due to a poor immune system (Hua *et al.*, 2015). In the present study, genomic analyses have revealed that gene loss through pseudogenization has had a significant impact on the evolution and diversification of pangolin genomes and their biology, similar to observations in other lineages (Wang *et al.* 2006; Meredith *et al.* 2014). Specifically, pseudogenes were identified and validated in pathways related to multiple morphological and/or physiological functions including dentition, vision, and immunity.

Several single-copy genes related to tooth development were pseudogenised in pangolins. One of these genes is enamelin (*ENAM*), which is a large extracellular matrix protein (*EMPs*) crucial for tooth development, and previously suggested to be responsible for molecular decay in pangolins (R. W. Meredith *et al.*, 2009). Meredith *et al.*, 2009 previously reported pseudogenised *ENAM* in the Chinese pangolin and the Tree pangolin. In the present study, similar mutation sites plus a number of additional mutations were found in the Malayan pangolins (Figure 4.1).

Interestingly, in addition to *ENAM*, two other EMP genes were observed to be pseudogenised, namely amelogenin (*AMELX*) and ameloblastin (*AMBN*). The mutational sites observed in these two genes were further verified in four African Pangolins species (Figure 4.2 and Figure 4.3). These mutations highlight the occurrence of mutation events before the divergence of the Asian and African pangolin species. It is noteworthy that the loss of *ENAM*, *AMBN* and *AMELX* are known to cause amelogenesis imperfecta in

humans (Wright *et al.*, 2008) and knock-out mice (Gibson *et al.*, 2001), which most likely explains the edentulous phenotype in pangolins.

Most pangolins are nocturnal (the only exception being the long-tailed pangolin) and are thought to have poor vision (D. A. Soewu & I. A. Ayodele, 2009). The visual acuity of pangolins is poor compared to their phylogenetically closest order Carnivora and other mammals. My results identified that the *BFSP2* and *GUCA1C* genes were pseudogenised in both Asian pangolins and an identical mutation site in *GUCA1C* was observed in all African pangolins species. Interestingly, the pseudogenization of *BFSP2* and *GUCA1C* has been reported in mice with reduced vision, supporting the hypothesis that the two genes are likely associated with the poor vision of pangolins (Imanishi *et al.* 2002; Sandilands *et al.* 2004; Song *et al.* 2009). Although pangolins have poor vision, they have developed a strong sense of smell (Thapa, 2014), which may have arisen as a result of strong selection pressure. Here, I identified a significant increase or expansion of the olfactory gene family 2 in pangolins compared to other closely related mammals (Figure 4.15). Other olfactory receptor (OR) gene families also showed a significant expansion in both pangolin species studied (Appendix I) compared to other mammals. Taken together, these results support the hypothesis that the expansion of olfactory gene families compensates for poor visual acuity.

Pangolins are the only mammals covered in keratinous scales (Tong *et al.*, 2007) and this may have been associated with the positive selection. Positive selection analyses revealed that several keratin genes, *KRT75*, *KRT82*, *KRTAP3-1*, and *KRT36* were positively selected in both pangolin species. The molecular composition of pangolin scales is similar to human hair and nails, but they cover the outermost skin layer of pangolins to provide a physical barrier. I hypothesize that the critical amino acid modifications predicted by

PROVEAN in these positively selected keratin genes could be associated with the molecular modification of keratin protein structure resulting in strong and hard keratinous scale formation. Interestingly, these critical amino acid changes are also present in both the *KRT36* and *KRT75* genes in the far diverged African pangolin species that were also examined in the present study, suggesting that they also contributed to the development of pangolin scales (Figure 4.20).

Pangolins are known for their poor immunity. The results of the present study clearly showed that the interferon (IFN) family genes esp. *IFNA* in pangolins is significantly reduced/contracted compared to other closely related mammals (Table 4.2). Although five other IFN-gene families except *IFNG* in the Chinese pangolin remain intact in both pangolins, the number of copies for *IFNA* in both pangolins are significantly reduced. Manual validation by BLAST-search further confirmed the reduced copy of *IFNA*. Draft genome assembly may result in reduced number of *IFNA* genes due to misassembly or misannotations. However, the pangolin genes that were annotated using the well-established MAKER pipeline based on several sources of evidence, and other highly conserved gene families are able to annotate correctly. Hence, the reduced *IFNA* copies in pangolins are likely due to the result of natural selection (Demuth & Hahn, 2009). The significantly reduced number of IFN genes might reduce the immunity of pangolins in general, as IFN genes are a group of signaling proteins that play an essential role in the immune defense in many complex organisms by activating host immune responses (Theofilopoulos *et al.*, 2005). Furthermore, in my pseudogene screening analysis, I found that the pseudogenization of the highly conserved *IFNE* gene in pangolins, but not in other mammals, suggesting that pangolins may have poor epidermal and mucosal immunity (Figure 4.6).

It is possible that the poor immune response likely due to the contraction or loss of IFN gene family members might have imposed a strong selective pressure in pangolins to acquire an alternative compensatory mechanism. Interestingly, I found evidence for the positive selection of other genes involved in immunity pathways, such as 1) the Colony Stimulating Factor 3 Receptor (*CSF3R*) which is essential for proliferation and differentiation of precursor cells into mature neutrophilic granulocytes and macrophages; 2) the integrin alpha M (*ITGAM*) gene that mediates neutrophil activation, leukocyte activation, and migration, macrophages phagocytosis that strengthens host immune responses and; 3) the transmembrane protein 173 (*TMEM173*) gene that is crucial for innate immune response and initiates type 1 interferon production during pathogen infection. Therefore, I postulate that the modifications in these positively selected immunity-related genes might compensate for the reduction of the immunity in pangolins. However, further experiments are required to test this hypothesis in the future.

6.2 The presumable loss of *IFNE* gene function in pangolins

One of the most striking observations of the present study was that the *IFNE* genes were pseudogenised in pangolins, but were otherwise functional in all other mammal genomes surveyed. *IFNE* is a single-copy gene known for its role in protection against viral infections, reproductive tract protection and providing essential mucosal immunity (Xi *et al.*, 2012). Therefore, the loss of *IFNE* gene function, for example, due to the premature stop codon (Figure 4.7), might result in the production of incomplete *IFNE* proteins that could greatly compromise the skin and mucosal immunity in pangolins.

In mice, a deficiency of *IFNE* in epithelial cells of the female reproductive tract can increase susceptibility to microbial infection (Fung *et al.*, 2013). Therefore, the pseudogenization of this interferon gene in pangolins supports the view that immunity

might be compromised these animals, resulting in an increased susceptibility to infection, particularly in the skin and mucosa-protected organs. This is further supported by evidence that the Malayan pangolin used here had been significantly colonized by *Burkholderia* sp. in tissues that express *IFNE* (cerebrum, cerebellum, and lung) in other species, but not in tissues (liver, kidney, thymus, spleen and heart) where the gene is usually not expressed in other mammals (Appendix J) (Demers *et al.*, 2014; Ponten *et al.*, 2008; Uhlen *et al.*, 2015). Furthermore, captive pangolins are prone to fatal gastrointestinal disease, pneumonia, and skin maladies (Clark L, 2008; Hua *et al.*, 2015). Pangolins are notoriously difficult to maintain in captivity, and the stress of captivity and poor husbandry might render pangolins even more vulnerable to infections or diseases by suppressing immune responses.

Interestingly, pangolins curl into nearly impregnable balls covering their scale-less abdomen using their well-developed neuromuscular system during sleep or when threatened, which would also support the hypothesis that these adaptations serve to protect pangolins from skin injuries (Durojaye A Soewu & Ibukun A Ayodele, 2009; Wilson, 1994). It is noteworthy that besides the anti-microbial effects, plasmacytoid dendritic cell (pDC)-produced type 1 *IFNA/IFNB* help heal wounds through re-epithelization of injured skin (Gregorio *et al.*, 2010). It is unknown whether keratinocyte-produced Type-I *IFNE* is also involved in skin wound healing (e.g. recruiting pDCs to the wound). Pangolins have a marked contraction of interferon genes that likely facilitates wound healing. Future studies are needed to test the relationship between *IFNE*, skin healing and scale development.

6.3 Possible relationship between *IFNE* pseudogenization and scales formation

My study raises the fundamental evolutionary question of how pangolin adaptations occurred given their seemingly increased risk to infection if their skin and mucosal surfaces are constantly being exposed to pathogens. The analyses suggest several evolutionary changes in genes that may respond or interact with pathogens, including immunity-related genes/pathways, the significant expansion of the Septin gene family and the enrichment of the genes in symbiosis. These changes may enhance the immune system of pangolins, although the relative efficiency of these adaptations compared to interferon-mediated immunity is unknown. However, I propose that the pangolin scales may be an important morphological innovation to act as defensive armor to protect pangolins against injuries (or stress) which would make pangolins even more vulnerable to infection or the invasion of pathogens. To further study the correlation between *IFNE* pseudogene and the development of pangolin scales, I tried to determine whether the formation of scales preceded the pseudogenization of *IFNE* or vice versa. Therefore, I sequenced and examined the *IFNE* sequences in all 4 modern African species. Encouragingly, I found that the *IFNE* gene was also pseudogenised in the African species. Since both Asian and African pangolins shared the common mutations in the *IFNE* gene, I suggest that this gene was pseudogenised before the divergence of both Asian and African species, which should be about 19-47 MYA based on previous studies (Du Toit *et al.*, 2014; R. W. Meredith *et al.*, 2009). I also estimated the time point of the gene inactivation using the dN/dS approach, which suggested that the time of the *IFNE* pseudogenization event was about 10 MYA. However, the predicted age of *IFNE* pseudogenization might not be reliable (probably due to high conservation of this pseudogene or insufficient number of *IFNE* sequences used in this estimation) because it does not fit with our above observation that the *IFNE* pseudogenization likely occurs

before the divergence of Asian and African species (e.g. at least 19-47 MYA) based on their shared mutations.

Additionally, looking at some known Pholidota fossils, the fossil record of *Eomanis waldi* and *Eomanis krebsi*, which is the earliest true pangolin (~50 MYA) from the middle Eocene of Europe, reported a lack of scales on its legs and tail (Maclean-Henry, 2012). Later, *Eurotamandua joresi* fossil was discovered in the middle Eocene of Messel (~47 MYA), which previously known closer to Xenarthra than Pholidota. *Eurotamandua joresi* has closed osteological characters to Pholidota and the lack of keratinous scale formation, suggesting the potential timeline for the initial formation of Pholidota scales (Maclean-Henry, 2012). Based on these fossil records, I hypothesize that the scale formation might occur nearly 47-50 MYA in the Pholidota order, whereas the *IFNE* was pseudogenised before the divergence of the Asian and African pangolins, which is approximately 19-47 MYA, suggesting a close potential time frame or relationship for the evolution of these two distinct traits of the Pholidota. The actual relationship between these two traits remain unclear, hence it would be interesting if we could examine the *IFNE* genes in the ancient pangolin fossils in order to find more evidence to support the potential connection between the *IFNE* pseudogenization and the scale formation in future.

6.4 Unique adaptations and behaviors of pangolins

My analyses showed that pangolin-specific genes were significantly over-represented in the GO categories governing cytoskeleton organization, neurological system process, and signal transduction. Further, certain genes involved in muscular and neuron system showed evidence of positive selection. These could be linked to the unique adaptations and behavior of pangolins, particularly those related to their highly sophisticated musculoskeletal system. For example, pangolins have evolved and adapted special

muscles allowing them to close their ears and nostrils to protect them from the attack of insects and preventing ants and termites from escaping in their mouth. Secondly, pangolins have extremely elongated and muscular tongues that can flick in and out quickly for capturing ants and termites. Thirdly, the toothless pangolins have strong stomach muscles with keratinous spines to digest its food through strong muscle contraction. Fourthly, Pangolins can roll into tight balls with their strong muscles, when threatened by predators.

6.5 Pangolin Genome Database (PGD) Development

The PGD database will be an important hub for hosting pangolin resources that will be of benefit to research communities working on pangolin biology. Currently, the PGD provides genome data and annotation of two Asian pangolins supported with dual-genome browsers for genome visualization. I will continue to expand and update the PGD database with the data and analysis results including pseudogene screening data and transcriptome expression data generated from our IPaRC consortium in the future. Furthermore, more analysis tools will also be added into the PGD, for example, Gene Ontology (GO) tools (Ashburner *et al.*, 2000) to find and map annotation of gene groups to GO terms, Primer3 (Untergasser *et al.*, 2012) enable users for rapid primer design and In-Silico PCR (Kent *et al.*, 2002) that searches the sequence database with a pair of PCR primers. To make the PGD a comprehensive pangolin genome hub, it would be useful to include the genome data and annotation of other pangolin species particularly the African species into this hub for the research community. It is doable since our collaborators have recently sequenced the genomes of African pangolins. I can include these data once it is available to the public. Finally, I hope this genome hub would facilitate research in pangolins particularly in the conservation of this critically endangered species and will also enhance our understanding of mammalian biology and evolution.

6.6 Potential impact and future works

Over the past century, more than 100 zoological organizations have tried to maintain pangolins in captivity to maintain their population, but most of the animals died within 6-36 months and only a few cases survived for more than 10 years (C. W. Yang *et al.*, 2007). The captive pangolins usually die of infections including lung infection, stomach infection and skin diseases (Hua *et al.*, 2015). I believe that this could be related to the reduced immunity in pangolins likely due to lack of *IFNE* (as well as the contraction of IFN gene family). Based on this observation, I would like to propose several suggestions that may be important to improve the conservation of pangolins. Firstly, we should keep pangolins in very clean environments to avoid possible microbial infections; Secondly, we should regularly monitor the health of pangolins in captivity, for example, to perform regular blood tests to make sure pangolins are not infected by bacteria; Thirdly, it would be helpful to enhance the pangolins' immunities, for example, by providing suitable food sources and even antibiotics; and lastly, it would be also important to reduce the environmental stresses on pangolins in captivity, which could help to maintain strong good immune responses of the pangolins.

Moreover, the *IFNE* genes of both Asian and African pangolin species shared a common sequence insertion, which is absent in the other 72 known mammalian species that I examined. It is possible that this mutation/pangolin-specific region could be used as a biomarker to identify pangolin species in future. However, further experiments may be needed to confirm this unique genomic region in the other pangolin species especially the Indian (*Manis crassicaudata*) and Philippine (*Manis culionensis*) pangolins that I did not examine in this study.

CHAPTER 7: CONCLUSION

In this study, I have successfully performed the first comparative genome analyses of Malayan and Chinese pangolins. This study has provided many new insights into the biology, evolution, and diversification of the critically endangered pangolins. The data revealed several pseudogenised genes that may be associated with the poor vision, the edentulous phenotype and poor immunity of pangolins. Significant expansion of the olfactory receptor genes is likely associated with the strong sense of smell of pangolins.

Unexpectedly, I have reported for the first time the loss of *IFNE* in pangolins, suggesting that the pseudogenization of this gene (as well as significant gene family contraction of IFN families) is likely associated with the poor immunity of pangolins. This finding may facilitate efforts to enhance conservation programs of these critically endangered animals. This study also provides new directions for further research in pangolin biology, immunity, and conservation. Moreover, it also provides better insights into mammalian genome evolution and possibly a natural animal model for studying mammalian immunity in the future.

To further facilitate research into pangolin biology, I have also designed and developed the PGD database, which can be a central hub for hosting genomic resources and information of pangolins, together with useful analysis tools for the research community. PGD provides interactive and user-friendly web interfaces for users to search, browse, retrieve, visualize and analyze pangolin genomic and transcriptomic data and information. I anticipate that the PGD will serve as a computational platform for researchers to advance research in this field, particularly in the area of conservation of this unique endangered mammalian species and also to provide a useful resource for research in mammalian evolution and biology, in general.

REFERENCES

- Abayomi, T. A., Ofusori, D. A., Ayoka, O. A., Odukoya, S. A., Omotoso, E. O., Amegor, F. O., . . . Oluwayinka, O. P. (2009). A Comparative Histological Study of the Tongue of Rat (*Rattus Norvegicus*), Bat (*Eidolon Helvum*) and Pangolin (*Manis Tricuspis*). *International Journal of Morphology*, 27(4), 1112-1120.
- Albalat, R., & Canestro, C. (2016). Evolution by gene loss. *Nat Rev Genet*.
- Alfoldi, J., & Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Res*, 23(7), 1063-1068.
- Alizadeh, A., Clark, J., Seeberger, T., Hess, J., Blankenship, T., & FitzGerald, P. G. (2004). Characterization of a mutation in the lens-specific CP49 in the 129 strain of mouse. *Invest Ophthalmol Vis Sci*, 45(3), 884-891.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), 403-410.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Consortium, G. O. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25-29.
- Bai, B., Zhao, W. M., Tang, B. X., Wang, Y. Q., Wang, L., Zhang, Z., . . . Zhang, Y. P. (2015). DoGSD: the dog and wolf genome SNP database. *Nucleic Acids Res*, 43(D1), D777-D783.
- Balakirev, E. S., & Ayala, F. J. (2003). Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet*, 37(1), 123-151.
- Bensasson, D., Zhang, D., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol*, 16(6), 314-321.
- Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., . . . Holmes, I. H. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biology*, 17(1), 66.
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1), 78-94.
- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., . . . Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*, 18(1), 188-196.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*, 17(4), 540-552.

- Challender, D., Nguyen The Truong A, Nguyen Van Q, Chung Y. (2014). *Manis javanica*. The IUCN Red List of Threatened Species. Version 2014.3.
- Challender, D. W., Thai, N. V., Jones, M., & May, L. (2012). Time-budgets and activity patterns of captive Sunda pangolins (*Manis javanica*). *Zoo Biology*, 31(2), 206-218.
- Challender, D. W. S. (2011). Asian pangolins: Increasing affluence driving hunting pressure. *TRAFFIC Bulletin*, 23, 92-93.
- Chan, W. M., & Consortium, U. (2010). The UniProt Knowledgebase (UniProtKB): a freely accessible, comprehensive and expertly curated protein sequence database. *Genetics Research*, 92(1), 78-79.
- Chang, C. (2004). Study on the apparent digestibility of diet on Formosan pangolin. *National Taiwan University, Taipei, Taiwan.[In Chinese]*.
- Chen, J., Jaeger, K., Den, Z., Koch, P. J., Sundberg, J. P., & Roop, D. R. (2008). Mice expressing a mutant Krt75 (K6hf) allele develop hair and nail defects resembling pachyonychia congenita. *J Invest Dermatol*, 128(2), 270-279.
- Chen, S. L., Hung, C.-S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., . . . Ozersky, P. (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proceedings of the National Academy of Sciences*, 103(15), 5977-5982.
- Chin, S.-C., & Yang, C.-W. (2009). *Formosan pangolin rescue, rehabilitation and conservation*. Paper presented at the WORKSHOP ON TRADE AND CONSERVATION OF PANGOLINS NATIVE TO SOUTH AND SOUTHEAST ASIA.
- Chin, S. C., Lien, C. Y., Chan, Y. T., Chen, C. L., Yang, Y. C., & Yeh, L. S. (2015). Hematologic and Serum Biochemical Parameters of Apparently Healthy Rescued Formosan Pangolins (*Manis Pentadactyla Pentadactyla*). *Journal of Zoo and Wildlife Medicine*, 46(1), 68-76.
- Cho, Y. S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., . . . Bhak, J. (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat Commun*, 4, 2433.
- Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2745-2747.
- Choo, S. W., Rayko, M., Tan, T. K., Hari, R., Komissarov, A., Wee, W. Y., . . . Antunes, A. (2016). Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Res*, 26(10), 1312-1322.

- CITES. (2000). Prop. 11.13. *Manis crassicaudata*, *Manis pentadactyla*, *Manis javanica*. Transfer from Appendix II to Appendix I (India, Nepal, Sri Lanka, United States). Available at: <http://www.cites.org/eng/cop/11/prop/13.pdf>.
- Clark L, V. T. N., Phuong TQ. (2008). *A long way from home: the health status of Asian pangolins confiscated from the illegal wildlife trade in Viet Nam*. In: Pantel S, Chin SY. (Eds). Paper presented at the Proceedings of the Workshop on Trade and Conservation of Pangolins Native to South and Southeast Asia. TRAFFIC Southeast Asia, Singapore Zoo, Singapore.
- Cohen, O., Rubinstein, N. D., Stern, A., Gophna, U., & Pupko, T. (2008). A likelihood framework to analyse phyletic patterns. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1512), 3903-3911.
- Conesa, A., & Gotz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*, 2008, 619832.
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676.
- Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol*, 13(4), e1002112.
- Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet*, 10(10), 691-703.
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 16(22), 10881-10890.
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Res*, 43(D1), D662-D669.
- Da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, 9, 119.
- Day, S. L., Ramshaw, I. A., Ramsay, A. J., & Ranasinghe, C. (2008). Differential effects of the type I interferons alpha4, beta, and epsilon on antiviral activity and vaccine efficacy. *J Immunol*, 180(11), 7158-7166.
- De Andrea, M., Ravera, R., Gioia, D., Gariglio, M., & Landolfo, S. (2002). The interferon system: an overview. *Eur J Paediatr Neurol*, 6 Suppl A, A41-46; discussion A55-48.

- De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22(10), 1269-1271.
- Debbabi, H., Dubarry, M., Rautureau, M., & Tome, D. (1998). Bovine lactoferrin induces both mucosal and systemic immune response in mice. *J Dairy Res*, 65(2), 283-293.
- Delport, W., Scheffler, K., & Seoighe, C. (2009). Models of coding sequence evolution. *Briefings in bioinformatics*, 10(1), 97-109.
- Demers, A., Kang, G., Ma, F., Lu, W., Yuan, Z., Li, Y., . . . Li, Q. (2014). The mucosal expression pattern of interferon-epsilon in rhesus macaques. *J Leukoc Biol*, 96(6), 1101-1107.
- Demuth, J. P., & Hahn, M. W. (2009). The life and death of gene families. *Bioessays*, 31(1), 29-39.
- Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A. A., Krasheninnikova, K., . . . O'Brien, S. J. (2015). Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biology*, 16, 277.
- Du Toit, Z., Grobler, J. P., Kotze, A., Jansen, R., Brettschneider, H., & Dalton, D. L. (2014). The complete mitochondrial genome of temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904). *Gene*, 551(1), 49-54.
- Duckworth, J. W., Challender, D., Baillie, J., Ades, G., Kaspal, P., Chan, B., Khatiwada, A., Xu, L., Chin, S., KC, R., Nash, H. & Hsieh, H. (2014). *Manis pentadactyla*. In: IUCN 2014. *IUCN Red List of Threatened Species*. Version 3.1. Retrieved 19 November, 2015, from <http://www.iucnredlist.org/details/12764/0>
- Enard, D., Messer, P. W., & Petrov, D. A. (2014). Genome-wide signals of positive selection in human evolution. *Genome Res*, 24(6), 885-895.
- Feldhamer, G., Drickamer, L., Vessey, S., Merritt, J., & Krajewski, C. (2007). Mammalogy: Adaptation, Diversity. *Ecology*, 346.
- Fensterl, V., & Sen, G. C. (2009). Interferons and viral infections. *Biofactors*, 35(1), 14-20.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., . . . Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *science*, 269(5223), 496-512.
- Fung, K. Y., Mangan, N. E., Cumming, H., Horvat, J. C., Mayall, J. R., Stifter, S. A., . . . Hertzog, P. J. (2013). Interferon-epsilon protects the female reproductive tract from viral and bacterial infection. *science*, 339(6123), 1088-1092.

- Gaubert, P., & Antunes, A. (2005). Assessing the taxonomic status of the Palawan pangolin *Manis culionensis* (Pholidota) using discrete morphological characters. *Journal of Mammalogy*, 86(6), 1068-1074.
- Ge, R. L., Cai, Q., Shen, Y. Y., San, A., Ma, L., Zhang, Y., . . . Wang, J. (2013). Draft genome sequence of the Tibetan antelope. *Nat Commun*, 4, 1858.
- Ghodsi, S., Rasaeipour, S., & Vojdani, M. (2012). Oral rehabilitation of a patient with amelogenesis imperfecta using removable overlay denture: a clinical report. *J Contemp Dent Pract*, 13(2), 227-231.
- Gibson, C. W., Yuan, Z. A., Hall, B., Longenecker, G., Chen, E., Thyagarajan, T., . . . Kulkarni, A. B. (2001). Amelogenin-deficient mice display an amelogenesis imperfecta phenotype. *J Biol Chem*, 276(34), 31871-31875.
- Glaser, D., Tinti, J. M., & Nofre, C. (1995). Evolution of the sweetness receptor in primates. I. Why does alitame taste sweet in all prosimians and simians, and aspartame only in Old World simians? *Chem Senses*, 20(5), 573-584.
- Graur, D., & Li, W. (2000). Gene duplication, exon shuffling, and concerted evolution. *Fundamentals of Molecular Evolution*, Ed, 2, 249-322.
- Gregorio, J., Meller, S., Conrad, C., Di Nardo, A., Homey, B., Lauerma, A., . . . Gilliet, M. (2010). Plasmacytoid dendritic cells sense skin injury and promote wound healing through type I interferons. *J Exp Med*, 207(13), 2921-2930.
- Grosshuesch, C. (2012). Diet and Nutrition - What's for Dinner? Retrieved 6 November, 2015, from http://bioweb.uwlax.edu/bio203/s2012/grosshue_crai/diet.htm
- Guo, Y., Zou, X., Chen, Y., Wang, D., & Wang, S. (1997). Sustainability of wildlife use in traditional Chinese medicine. *Conserving China's Biodiversity*, 190-220.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494-1512.
- Haeseleer, F., Sokal, I., Li, N., Pettenati, M., Rao, N., Bronson, D., . . . Palczewski, K. (1999). Molecular characterization of a third member of the guanylyl cyclase-activating protein subfamily. *J Biol Chem*, 274(10), 6526-6535.
- Hahn, M. W., Han, M. V., & Han, S. G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet*, 3(11), e197.
- Hall, T. (2004). BioEdit version 7.2.5.
- Hao, W., & Golding, G. B. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res*, 16(5), 636-643.

- Hardison, R. C. (2003). Comparative genomics. *PLoS Biol*, 1(2), E58.
- Harrison, P. M., Hegyi, H., Balasubramanian, S., Luscombe, N. M., Bertone, P., Echols, N., . . . Gerstein, M. (2002). Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res*, 12(2), 272-280.
- Hart, T. C., Hart, P. S., Gorry, M. C., Michalec, M. D., Ryu, O. H., Uygur, C., . . . Firatli, E. (2003). Novel ENAM mutation responsible for autosomal recessive amelogenesis imperfecta and localised enamel defects. *J Med Genet*, 40(12), 900-906.
- Hazkani-Covo, E., & Graur, D. (2007). A comparative analysis of numt evolution in human and chimpanzee. *Mol Biol Evol*, 24(1), 13-18.
- Heath, M. E., & Vanderlip, S. L. (1988). Biology, Husbandry, and Veterinary Care of Captive Chinese Pangolins (Manis-Pentadactyla). *Zoo Biology*, 7(4), 293-312.
- Hickey, G., Paten, B., Earl, D., Zerbino, D., & Haussler, D. (2013). HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*, btt128.
- Hua, L., Gong, S., Wang, F., Li, W., Ge, Y., Li, X., & Hou, F. (2015). Captive breeding of pangolins: current status, problems and future prospects. *Zookeys*(507), 99-114.
- Hughes, A. L. (2008). Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann N Y Acad Sci*, 1133, 162-179.
- Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 335(6186), 167-170.
- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., . . . Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res*, 37, D211-D215.
- Jacq, C., Miller, J. R., & Brownlee, G. G. (1977). A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell*, 12(1), 109-120.
- Jeong, H., Song, K.-D., Seo, M., Caetano-Anollés, K., Kim, J., Kwak, W., . . . Cho, S. (2015). Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. *BMC genetics*, 16(1), 104.
- Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *Bmc Bioinformatics*, 11, 431.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28(1), 27-30.

- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*, 32(suppl 1), D493-D496.
- Karro, J. E., Yan, Y. P., Zheng, D. Y., Zhang, Z. L., Carriero, N., Cayting, P., . . . Gerstein, M. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res*, 35, D55-D60.
- Kawashima, T., Thorington, R. W., Bohaska, P. W., Chen, Y. J., & Sato, F. (2015). Anatomy of Shoulder Girdle Muscle Modifications and Walking Adaptation in the Scaly Chinese Pangolin (*Manis Pentadactyla* *Pentadactyla*: Pholidota) Compared with the Partially Osteoderm-Clad Armadillos (*Dasypodidae*). *Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology*, 298(7), 1217-1236.
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res*, 12(4), 656-664.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res*, 12(6), 996-1006.
- Kondrashov, P., & Agadjanian, A. K. (2012). A Nearly Complete Skeleton of *Ernanodon* (Mammalia, Palaeonodonta) from Mongolia: Morphofunctional Analysis. *Journal of Vertebrate Paleontology*, 32(5), 983-1001.
- Korf, I. (2004). Gene finding in novel genomes. *Bmc Bioinformatics*, 5.
- Kulmuni, J., Wurm, Y., & Pamilo, P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity (Edinb)*, 110(6), 538-547.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.
- Landin, M. A., Shabestari, M., Babaie, E., Reseland, J. E., & Osmundsen, H. (2012). Gene Expression Profiling during Murine Tooth Development. *Front Genet*, 3, 139.
- Lechner, M., Hernandez-Rosales, M., Doerr, D., Wieseke, N., Thevenin, A., Stoye, J., . . . Stadler, P. F. (2014). Orthology detection combining clustering and synteny for very large datasets. *PLoS One*, 9(8), e105015.
- Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J. K., Osmotherly, L., . . . Durbin, R. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res*, 34(Database issue), D572-580.

- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., . . . Wang, J. (2010). The sequence and de novo assembly of the giant panda genome. *Nature*, 463(7279), 311-317.
- Lihua, C., Liang, X., Hui, X., Yunya, X., & Zhongfeng, X. (2010). Protection and Cultivated Technology of Pangolin. *Asia-Pacific Traditional Medicine*, 2, 006.
- Lin, M. F., Chang, C. Y., Yang, C. W., & Dierenfeld, E. S. (2015). Aspects of digestive anatomy, feed intake and digestion in the Chinese pangolin (*Manis pentadactyla*) at Taipei zoo. *Zoo Biology*, 34(3), 262-270.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., . . . Univ, W. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370), 476-482.
- Liongue, C., & Ward, A. C. (2014). Granulocyte colony-stimulating factor receptor mutations in myeloid malignancy. *Front Oncol*, 4, 93.
- Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *science*, 227(4693), 1435-1441.
- Liu, Z., Gui, D., Jia, K., Xia, J., Yang, L., Chen, J., . . . Yi, M. (2013). De novo Assembly of the Indo-Pacific Humpback Dolphin Leucocyte Transcriptome to Identify Putative Genes Involved in the Aquatic Adaptation and Immune Response. *PLoS One*, 8(8), e72417.
- Luo, J.-y., Yan, D., Song, J.-y., Zhang, D., Xing, X.-y., Han, Y.-m., . . . Chen, S.-l. (2013). A strategy for trade monitoring and substitution of the organs of threatened animals. *Scientific reports*, 3.
- Maclean-Henry, G. (2012). Review of fossil pangolins (Mammalia: Pholidota) in respect to their phylogenetic relationships.
- Mascarelli, A. (2011). Septin proteins take bacterial prisoners: A cellular defence against microbial pathogens holds therapeutic potential. *Nature. Do i*, 10.
- Max, E. E. (2003). Plagiarized errors and molecular genetics. *On-line*], URL: <http://talkorigins.org/faqs/molgen>.
- Medway, L. (1969). The wild mammals of Malaya and offshore islands including Singapore. *The wild mammals of Malaya and offshore islands including Singapore*.
- Meredith, R. W., Gatesy, J., Cheng, J., & Springer, M. S. (2011). Pseudogenization of the tooth gene enamelysin (MMP20) in the common ancestor of extant baleen whales. *Proc Biol Sci*, 278(1708), 993-1002.
- Meredith, R. W., Gatesy, J., Murphy, W. J., Ryder, O. A., & Springer, M. S. (2009). Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet*, 5(9), e1000634.

- Meredith, R. W., Gatesy, J., & Springer, M. S. (2013). Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC evolutionary biology*, 13(1), 1.
- Meredith, R. W., Zhang, G., Gilbert, M. T., Jarvis, E. D., & Springer, M. S. (2014). Evidence for a single loss of mineralized teeth in the common avian ancestor. *science*, 346(6215), 1254390.
- Meyer, W., Liamsiricharoen, M., Suprasert, A., Fleischer, L. G., & Hewicker-Trautwein, M. (2013). Immunohistochemical demonstration of keratins in the epidermal layers of the Malayan pangolin (*Manis javanica*), with remarks on the evolution of the integumental scale armour. *Eur J Histochem*, 57(3), e27.
- Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, 8(11), 845-856.
- Mohapatra, R. K., & Panda, S. (2014). Behavioural Descriptions of Indian Pangolins (*Manis crassicaudata*) in Captivity. *International Journal of Zoology*, 2014.
- Moll, R., Divo, M., & Langbein, L. (2008). The human keratins: biology and pathology. *Histochem Cell Biol*, 129(6), 705-733.
- Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc*, 2007, pdb top17.
- Mundy, N. I., & Kelly, J. (2003). Evolution of a pigmentation gene, the melanocortin-1 receptor, in primates. *Am J Phys Anthropol*, 121(1), 67-80.
- Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A., & O'Brien, S. J. (2001). Molecular phylogenetics and the origins of placental mammals. *Nature*, 409(6820), 614-618.
- Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*, 3(5), 418-426.
- Newton, P., Van Thai, N., Robertson, S., & Bell, D. (2008). Pangolins in peril: using local hunters' knowledge to conserve elusive species in Vietnam. *Endangered Species Research*, 6(1), 41-53.
- Nguyen, N., Hickey, G., Raney, B. J., Armstrong, J., Clawson, H., Zweig, A., . . . Paten, B. (2014). Comparative assembly hubs: web-accessible browsers for comparative genomics. *Bioinformatics*, 30(23), 3293-3301.
- Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One*, 2(8), e708.

- Pantel, S., & Anak, N. A. (2010). A preliminary assessment of the pangolin trade in Sabah. *TRAFFIC Southeast Asia*.
- Pantel, S., & Chin, S. Y. (2009). Proceedings of the Workshop on Trade and Conservation of Pangolins native to South and Southeast Asia. *TRAFFIC Southeast Asia*.
- Parkos, C. A., Delp, C., Arnaout, M. A., & Madara, J. L. (1991). Neutrophil migration across a cultured intestinal epithelium. Dependence on a CD11b/CD18-mediated event and enhanced efficiency in physiological direction. *J Clin Invest*, 88(5), 1605-1612.
- Paten, B., Diekhans, M., Earl, D., John, J. S., Ma, J., Suh, B., & Haussler, D. (2011). Cactus graphs for genome comparisons. *Journal of Computational Biology*, 18(3), 469-481.
- Pearson, W. R. (1999). Flexible sequence similarity searching with the FASTA3 program package *Bioinformatics methods and protocols* (pp. 185-219): Springer.
- Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, 85(8), 2444-2448.
- Pei, B. K., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X. J., . . . Gerstein, M. B. (2012). The GENCODE pseudogene resource. *Genome Biology*, 13(9).
- Ploos van Amstel, H. K., Reitsma, P. H., Van der Logt, C. P. E., & Bertina, R. M. (1990). Intron-exon organization of the active human protein S gene PS. alpha. and its pseudogene PS. beta.: Duplication and silencing during primate evolution. *Biochemistry*, 29(34), 7853-7861.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, 20(1), 110-121.
- Ponten, F., Jirstrom, K., & Uhlen, M. (2008). The Human Protein Atlas--a tool for pathology. *J Pathol*, 216(4), 387-393.
- Ran, Y., Shu, H. B., & Wang, Y. Y. (2014). MITA/STING: a central and multifaceted mediator in innate immune response. *Cytokine Growth Factor Rev*, 25(6), 631-639.
- Rana, F. (2008). *The Cell's Design (Reasons to Believe): How Chemistry Reveals the Creator's Artistry*: Baker Books.
- Raney, B. J., Dreszer, T. R., Barber, G. P., Clawson, H., Fujita, P. A., Wang, T., . . . Karolchik, D. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics*, 30(7), 1003-1005.

- Redford, K. H. (1987). Ants and termites as food *Current mammalogy* (pp. 349-399): Springer.
- Rosen, G. E., & Smith, K. F. (2010). Summarizing the evidence on the international trade in illegal wildlife. *EcoHealth*, 7(1), 24-32.
- Rouchka, E. C., & Cha, I. E. (2009). Current Trends in Pseudogene Detection and Characterization. *Current Bioinformatics*, 4(2), 112-119.
- Rylander, P. (2015). Rediscovering the Pangolin. Retrieved 7-6-2016, 2016, from <http://nrmnh.typepad.com/100years/2015/05/rediscovering-the-pangolin.html>
- Sabeti, P., Schaffner, S., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., . . . Lander, E. (2006). Positive natural selection in the human lineage. *science*, 312(5780), 1614-1620.
- Sandilands, A., Prescott, A. R., Wegener, A., Zoltoski, R. K., Hutcheson, A. M., Masaki, S., . . . Quinlan, R. A. (2003). Knockout of the intermediate filament protein CP49 destabilises the lens fibre cell cytoskeleton and decreases lens optical quality, but does not induce cataract. *Exp Eye Res*, 76(3), 385-391.
- Shepherd, C. R. (2009). *Overview of pangolin trade in Southeast Asia*. Paper presented at the WORKSHOP ON TRADE AND CONSERVATION OF PANGOLINS NATIVE TO SOUTH AND SOUTHEAST ASIA.
- Shirak, A., Grabherr, M., Di Palma, F., Lindblad-Toh, K., Hulata, G., Ron, M., . . . Seroussi, E. (2010). Identification of Repetitive Elements in the Genome of *Oreochromis niloticus*: Tilapia Repeat Masker. *Marine Biotechnology*, 12(2), 121-125.
- Simpson, J. T., & Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. *Genome Res*, 22(3), 549-556.
- Sisu, C., Pei, B. K., Leng, J., Frankish, A., Zhang, Y., Balasubramanian, S., . . . Gerstein, M. B. (2014). Comparative analysis of pseudogenes across three phyla. *Proc Natl Acad Sci U S A*, 111(37), 13361-13366.
- Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., & Holmes, I. H. (2009). JBrowse: a next-generation genome browser. *Genome Res*, 19(9), 1630-1638.
- Smit, A. F., Hubley, R., & Green, P. (1996). RepeatMasker Open-3.0.
- Smith, C. D., Edgar, R. C., Yandell, M. D., Smith, D. R., Celniker, S. E., Myers, E. W., & Karpen, G. H. (2007). Improved repeat identification and masking in Dipterans. *Gene*, 389(1), 1-9.
- Smith, F. J., Irvine, A. D., Terron-Kwiatkowski, A., Sandilands, A., Campbell, L. E., Zhao, Y., . . . McLean, W. H. (2006). Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nature Genetics*, 38(3), 337-342.

- Soewu, D. A., & Ayodele, I. A. (2009). Utilisation of pangolin (*Manis* spp) in traditional Yorubic medicine in Ijebu province, Ogun State, Nigeria. *J Ethnobiol Ethnomed*, 5(1), 1.
- Soewu, D. A., & Ayodele, I. A. (2009). Utilisation of Pangolin (*Manis* spp) in traditional Yorubic medicine in Ijebu province, Ogun State, Nigeria. *J Ethnobiol Ethnomed*, 5, 39.
- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*, 33, W465-W467.
- Suzuki, Y., & Gojobori, T. (1999). A method for detecting positive selection at single amino acid sites. *Mol Biol Evol*, 16(10), 1315-1328.
- Sweeney, R. (1956). CXV.—Some notes on the feeding habits of the Ground Pangolin, *Smutsia Temminckii* (Smuts). *Journal of Natural History*, 9(108), 893-896.
- Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res*, 11(8), 1441-1452.
- Teshima, K. M., & Przeworski, M. (2006). Directional positive selection on an allele of arbitrary dominance. *Genetics*, 172(1), 713-718.
- Thapa, P. (2014). An Overview of Chinese Pangolin (*Manis pentadactyla*): Its General Biology, Status, Distribution and Conservation Threats in Nepal. *The Initiation*, 5, 164-170.
- Theofilopoulos, A. N., Baccala, R., Beutler, B., & Kono, D. H. (2005). Type I interferons (α/β) in immunity and autoimmunity. *Annu. Rev. Immunol.*, 23, 307-335.
- Thesleff, I. (2006). The genetic basis of tooth development and dental defects. *American Journal of Medical Genetics Part A*, 140(23), 2530-2535.
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H. Y., Karlak, B., Daverman, R., . . . Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res*, 13(9), 2129-2141.
- Timothy J. Gaudin, R. J. R. J. E. (2009). The Phylogeny of Living and Extinct Pangolins (Mammalia, Pholidota) and Associated Taxa: A Morphology Based Analysis. *J Mammal Evol*, 16, 235-305.
- Tolstorukov, M. Y., Kharchenko, P. V., Goldman, J. A., Kingston, R. E., & Park, P. J. (2009). Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes. *Genome Res*, 19(6), 967-977.

- Tong, J., Lü T.-b., Ma, Y.-h., Wang, H.-k., Ren, L.-q., & Arnell, R. (2007). Two-body abrasive wear of the surfaces of pangolin scales. *Journal of Bionic Engineering*, 4(2), 77-84.
- Uhlen, M., Bjorling, E., Agaton, C., Szigartyo, C. A., Amini, B., Andersen, E., . . . Ponten, F. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics*, 4(12), 1920-1932.
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *science*, 347(6220), 1260419.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res*, 40(15), e115.
- Vallender, E. J., & Lahn, B. T. (2004). Positive selection on the human genome. *Hum Mol Genet*, 13 Spec No 2, R245-254.
- Vanin, E. F. (1984). Processed pseudogenes. Characteristics and evolution. *Biochim Biophys Acta*, 782(3), 231-241.
- Vanin, E. F. (1985). Processed pseudogenes: characteristics and evolution. *Annu Rev Genet*, 19, 253-272.
- Villanueva-Cañás, J. L., Faherty, S. L., Yoder, A. D., & Albà M. M. (2014). Comparative genomics of mammalian hibernators using gene networks. *Integrative and comparative biology*, icu048.
- Wang, X., Grus, W. E., & Zhang, J. (2006). Gene losses during human origins. *PLoS Biol*, 4(3), e52.
- Ward, P. P., Uribe-Luna, S., & Conneely, O. M. (2002). Lactoferrin and host defense. *Biochem Cell Biol*, 80(1), 95-102.
- Wilson, A. E. (1994). Husbandry of pangolins *Manis* spp. *International Zoo Yearbook*, 33(1), 248-251.
- Wright, J. T., Hart, T. C., Hart, P. S., Simmons, D., Suggs, C., Daley, B., . . . Li, Y. (2008). Human and mouse enamel phenotypes resulting from mutation or altered expression of AMEL, ENAM, MMP20 and KLK4. *Cells Tissues Organs*, 189(1-4), 224-229.
- Wu, S., Liu, N., Zhang, Y., & Ma, G. (2004). Assessment of threatened status of Chinese Pangolin (*Manis pentadactyla*). *Chinese Journal of Applied and Environmental Biology*, 10(4), 456-461.
- Wu, T. D., & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), 1859-1875.

- Xi, Y., Day, S. L., Jackson, R. J., & Ranasinghe, C. (2012). Role of novel type I interferon epsilon in viral infection and mucosal immunity. *Mucosal Immunol*, 5(6), 610-622.
- Xie, Y. L., Wu, G. X., Tang, J. B., Luo, R. B., Patterson, J., Liu, S. L., . . . Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, 30(12), 1660-1666.
- Xu, X., Dong, G. X., Hu, X. S., Miao, L., Zhang, X. L., Zhang, D. L., . . . Luo, S. J. (2013). The genetic basis of white tigers. *Curr Biol*, 23(11), 1031-1035.
- Xue, Y., Daly, A., Yngvadottir, B., Liu, M., Coop, G., Kim, Y., . . . Huckle, E. (2006). Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *The American Journal of Human Genetics*, 78(4), 659-670.
- Yang, C. W., Chen, S., Chang, C.-Y., Lin, M. F., Block, E., Lorentsen, R., . . . Dierenfeld, E. S. (2007). History and dietary husbandry of pangolins in captivity. *Zoo Biology*, 26(3), 223-230.
- Yang, C. W., Chen, S., Chang, C. Y., Lin, M. F., Block, E., Lorentsen, R., . . . Dierenfeld, E. S. (2007). History and dietary husbandry of pangolins in captivity. *Zoo Biology*, 26(3), 223-230.
- Yang, W., Chen, I. H., Gludovatz, B., Zimmermann, E. A., Ritchie, R. O., & Meyers, M. A. (2013). Natural flexible dermal armor. *Advanced Materials*, 25(1), 31-48.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 24(8), 1586-1591.
- Yu, R. (1992). Clinical application of chinese herbs remedies as immunomodulators in the treatment of cancer. *Chinese Journal of Cancer Research*, 4(3), 66-78.
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*, 18(5), 821-829.
- Zhang, G. J., Cowled, C., Shi, Z. L., Huang, Z. Y., Bishop-Lilly, K. A., Fang, X. D., . . . Wang, J. (2013). Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity. *science*, 339(6118), 456-460.
- Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*, 22(12), 2472-2479.
- Zhang, Z., Carriero, N., Zheng, D., Karro, J., Harrison, P. M., & Gerstein, M. (2006). PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*, 22(12), 1437-1439.

- Zhang, Z., Harrison, P. M., Liu, Y., & Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res*, 13(12), 2541-2558.
- Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J., & Gerstein, M. (2010). Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biology*, 11(3), R26.
- Zhang, Z. L., Carriero, N., Zheng, D. Y., Karro, J., Harrison, P. M., & Gerstein, M. (2006). PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*, 22(12), 1437-1439.
- Zhao, H., Rossiter, S. J., Teeling, E. C., Li, C., Cotton, J. A., & Zhang, S. (2009). The evolution of color vision in nocturnal mammals. *Proceedings of the National Academy of Sciences*, 106(22), 8980-8985.
- Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Choo, S. W., . . . Gerstein, M. B. (2007). Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res*, 17(6), 839-851.
- Zheng, D., & Gerstein, M. B. (2006). A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biology*, 7 Suppl 1, S13 11-10.
- Zhou, Z.-M., Zhou, Y., Newman, C., & Macdonald, D. W. (2014). Scaling up pangolin protection in China. *Frontiers in Ecology and the Environment*, 12(2), 97-98.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, et al. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome research*. 2016;26(10):1312-22. Epub 2016/08/12. doi: 10.1101/gr.203521.115. PubMed PMID: 27510566.

Tan TK, Tan KY, Hari R, Mohamed Yusoff A, Wong GJ, Siow CC, et al. PGD: a pangolin genome hub for the research community. *DATABASE: the journal of biological databases and curation*. 2016;2016. Epub 2016/09/13. doi: 10.1093/database/baw063. PubMed PMID: 27616775; PubMed Central PMCID: PMCPmc5018392.

Mohamed Yusoff A, Tan TK, Hari R, Koepfli KP, Wee WY, Antunes A, et al. De novo sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin. *Scientific reports*. 2016;6:28199. Epub 2016/09/14. doi: 10.1038/srep28199. PubMed PMID: 27618997; PubMed Central PMCID: PMCPmc5020319.

LIST OF PUBLICATIONS AND PAPERS PRESENTED





Certificate of Presentation

This is to certify that

TANTZE KING

Participated and presented the research paper titled **“VIBRIOBASE: A MODEL FOR NEXT-GENERATION GENOME AND ANNOTATION DATABASE DEVELOPMENT”** in the 2nd international conference **“Trends in Multidisciplinary Business and Economic Research (TMBER-2015)”** held at **HOLIDAY INN Bangkok Silom, Thailand** on **March 25-26, 2015.**

Tariq Iqbal Khan
Conference Co- Chair
TMBER-2015 Secretariat
Director- Global operations
Global Illuminators



APPENDIX

Appendix A: Summary of selected genes for gene traits analysis.

TOOTH RELATED GENES		VISION RELATED GENES				IMMUNE RELATED GENES	
GENE	ENTREZ ID	GENE	ENTREZ ID	GENE	ENTREZ ID	GENE	ENTREZ ID
CTNNB1	1499	ABCA4	24	NYX	60506	ADORA2A	135
BMP2	650	ABCC6	368	OAT	4942	C5	727
BMP4	652	ABLIM1	3983	OCLM	10896	C8A	731
DLX2	1746	AIPL1	23746	OPA1	4976	CAMP	820
EDA	1896	AOC2	314	OPA3	80207	CASP1	834
EDAR	10913	ARL6	84100	OPN1LW	5956	CASP4	837
PAX9	5083	ARR3	407	OPN1MW	2652	CCL2	6347
PITX2	5308	ATP8A2	51761	OPN1MW2	728458	CCR3	1232
RUNX2	860	ATXN7	6314	OPN1SW	611	CD14	929
SHH	6469	BBS1	582	OPN3	23596	CD1D	912
WNT7B	7477	BBS10	79738	OPN4	94233	CD55	1604
HOXA2B	30325	BBS2	583	OPN5	221391	CHUK	1147
HOXA5A	58055	BBS4	585	PAX2	5076	COLEC12	81035
HOXB2A	30338	BBS5	129880	PAX6	5080	CRP	1401
HOXB5B	58052	BBS9	27241	PCDH15	65217	CYBB	1536
HOXB6B	58053	BEST1	7439	PDC	5132	DEFB4A	1673
HOXC6A	30346	BFSP2	8419	PDCL	5082	DMBT1	1755
HOXD4A	30329	C2ORF71	388939	PDE6A	5145	FN1	2335
EVE1	30335	CABP4	57010	PDE6B	5158	HMOX1	3162
LHX6	26468	CACNA1F	778	PDE6C	5146	IFNA1	3439
LHX7	431707	CACNA2D4	93589	PDE6D	5147	IFNB1	3456
FGF3	2248	CACNB2	783	PDE6G	5148	IFNGR1	3459
FGF10	2255	CACNB4	785	PDE6H	5149	IFNGR2	3460
NOTCH2	4853	CDH23	64072	PITPNA	5306	IKBKB	3551
PAX9	5083	CDH3	1001	POU4F3	5459	IL10	3586
MSX1	4487	CDHR1	92211	POU6F2	11281	IL12RB2	3595
AXIN2	8313	CHM	1121	PPEF2	5470	IL1A	3552
MSX2	4488	CHRNA2	1141	PPT1	5538	IL1B	3553
BMP4	652	CLDN19	149461	PRCD	768206	IL1F10	84639
WNT7B	7477	CLN5	1203	PRPH2	5961	IL1R1	3554
NGFR	4804	CLN6	54982	PRR4	11272	IL1R2	7850
ITGB4	3691	CLN8	2055	RABGGTA	5875	IL1RAP	3556

<i>LRP6</i>	4040	<i>CLN8</i>	2055	<i>RABGGTB</i>	5876	<i>IL1RAPL2</i>	26280
<i>FAM20A</i>	54757	<i>CLRN1</i>	7401	<i>RAX</i>	30062	<i>IL1RL2</i>	8808
<i>AMBN</i>	258	<i>CNGA1</i>	1259	<i>RAX2</i>	84839	<i>IL1RN</i>	3557
<i>AMEL</i>	396940	<i>CNGA3</i>	1261	<i>RB1</i>	5925	<i>IL36A</i>	27179
<i>ENAM</i>	10117	<i>CNGB1</i>	1258	<i>RBP3</i>	5949	<i>IL36B</i>	27177
<i>COL1A1</i>	1277	<i>CNGB3</i>	54714	<i>RBP4</i>	5950	<i>IL36G</i>	56300
<i>ACVR2B</i>	2028	<i>CNNM4</i>	26504	<i>RCVRN</i>	5957	<i>IL36RN</i>	26525
<i>ADM</i>	2047	<i>COL11A1</i>	1301	<i>RD3</i>	343035	<i>IL37</i>	27178
<i>ALPL</i>	2100	<i>COL18A1</i>	80781	<i>RDH10</i>	157506	<i>IL6</i>	3569
<i>AMBN</i>	2101	<i>COL1A1</i>	1277	<i>RDH12</i>	145226	<i>IRAK1</i>	3654
<i>AMELX</i>	2107	<i>COL2A1</i>	1280	<i>RDH5</i>	5959	<i>IRAK2</i>	3656
<i>AMTN</i>	1592356	<i>CRX</i>	1406	<i>RDH8</i>	50700	<i>IRF1</i>	3659
<i>ANKRD11</i>	1306679	<i>CRX</i>	1406	<i>RGR</i>	5995	<i>LALBA</i>	3906
<i>ASPN</i>	1549776	<i>CRYAA</i>	1409	<i>RGS16</i>	6004	<i>LBP</i>	3929
<i>BAX</i>	2192	<i>CRYBA1</i>	1411	<i>RGS9</i>	8787	<i>LTF</i>	4057
<i>BCL11B</i>	1311128	<i>CRYBA4</i>	1413	<i>RGS9BP</i>	388531	<i>LY96</i>	23643
<i>BCL2L11</i>	628774	<i>CRYBB1</i>	1414	<i>RHO</i>	6010	<i>LYZ</i>	4069
<i>BCOR</i>	1562735	<i>CRYBB1</i>	1414	<i>RLBP1</i>	6017	<i>MAPK14</i>	1432
<i>BMP2</i>	2211	<i>CRYBB2</i>	1415	<i>ROM1</i>	6094	<i>MAPK8</i>	5599
<i>BMP4</i>	2213	<i>CRYBB3</i>	1417	<i>RORB</i>	6096	<i>MIF</i>	4282
<i>BMP7</i>	620743	<i>CRYGA</i>	1418	<i>RP1</i>	6101	<i>MYD88</i>	4615
<i>BMPRI1A</i>	70989	<i>CRYGB</i>	1419	<i>RP1L1</i>	94137	<i>NCF4</i>	4689
<i>BSG</i>	2220	<i>CRYGC</i>	1420	<i>RP2</i>	6102	<i>NFKB1</i>	4790
<i>CAR2</i>	2240	<i>CRYGD</i>	1421	<i>RPE65</i>	6121	<i>NFKB2</i>	4791
<i>CHUK</i>	1306661	<i>CRYZ</i>	1429	<i>RPGR</i>	6103	<i>NFKBIA</i>	4792
<i>COL1A1</i>	61817	<i>CYP1A2</i>	1544	<i>RPGRIP1</i>	57096	<i>NLRC4</i>	58484
<i>CSF1</i>	621063	<i>CYP1B1</i>	1545	<i>RRH</i>	10692	<i>NOS2</i>	4843
<i>CSF3R</i>	1309114	<i>CYP4V2</i>	285440	<i>RS1</i>	6247	<i>PGLYRP1</i>	8993
<i>CTNNA1</i>	1359485	<i>DHDDS</i>	79947	<i>SAG</i>	6295	<i>PGLYRP2</i>	114770
<i>CTNNB1</i>	70487	<i>DHRS3</i>	9249	<i>SDR16C5</i>	195814	<i>PGLYRP3</i>	114771
<i>DICER1</i>	1309381	<i>DLAT</i>	1737	<i>SFRP5</i>	6425	<i>PPBP</i>	5473
<i>DLX1</i>	1309593	<i>DLL4</i>	54567	<i>SIX3</i>	6496	<i>PROC</i>	5624
<i>DLX2</i>	1304853	<i>DNAJC19</i>	131118	<i>SIX6</i>	4990	<i>PTAFR</i>	5724
<i>DLX3</i>	1304875	<i>EFEMP1</i>	2202	<i>SLC24A1</i>	9187	<i>S100A12</i>	6283
<i>DMRT3</i>	1306043	<i>EML2</i>	24139	<i>SLC24A2</i>	25769	<i>SERPINA1</i>	5265
<i>DSPP</i>	2525	<i>EPAS1</i>	2034	<i>SLC45A2</i>	51151	<i>SERPINE1</i>	5054
<i>EDA</i>	1563178	<i>EYA3</i>	2140	<i>SLC4A7</i>	9497	<i>SFTPD</i>	6441
<i>EDAR</i>	1561714	<i>EYA4</i>	2070	<i>SLITRK6</i>	84189	<i>TGFB1</i>	7040
<i>EDARADD</i>	1564010	<i>EYS</i>	346007	<i>SOX14</i>	8403	<i>TLR1</i>	7096
<i>ENAM</i>	1306796	<i>FAM161A</i>	84140	<i>SOX9</i>	6662	<i>TLR10</i>	81793

<i>FAM20A</i>	1306364	<i>FSCN2</i>	25794	<i>SPATA7</i>	55812	<i>TLR2</i>	7097
<i>FAM20C</i>	1311980	<i>GABRR1</i>	2569	<i>T</i>	6862	<i>TLR3</i>	7098
<i>FGF10</i>	2606	<i>GABRR2</i>	2570	<i>TACSTD2</i>	4070	<i>TLR4</i>	7099
<i>FGF4</i>	620127	<i>GJA10</i>	84694	<i>TCTN3</i>	26123	<i>TLR6</i>	10333
<i>FGF8</i>	70891	<i>GJA3</i>	2700	<i>TGFB1</i>	7045	<i>TLR8</i>	51311
<i>FOXC1</i>	1589718	<i>GJA8</i>	2703	<i>TH</i>	7054	<i>TLR9</i>	54106
<i>FST</i>	2633	<i>GJC1</i>	10052	<i>TIMP3</i>	7078	<i>TNF</i>	7124
<i>GATA6</i>	2666	<i>GJD2</i>	57369	<i>TRPM1</i>	4308	<i>TNFRSF1A</i>	7132
<i>GLI2</i>	1309270	<i>GLRA1</i>	2741	<i>TULP1</i>	7287	<i>TOLLIP</i>	54472
<i>GLI3</i>	620272	<i>GLRB</i>	2743	<i>TULP2</i>	7288	<i>TRAF6</i>	7189
<i>HAND1</i>	621206	<i>GNAT1</i>	2779	<i>TYR</i>	7299	<i>TREM1</i>	54210
<i>HAND2</i>	621207	<i>GNAT2</i>	2780	<i>UNC119</i>	9094		
<i>HDAC1</i>	1309799	<i>GNAT3</i>	346562	<i>UNC119B</i>	84747		
<i>HDAC2</i>	619976	<i>GPR143</i>	4935	<i>USH2A</i>	7399		
<i>HTRA1</i>	69235	<i>GPR179</i>	440435	<i>VAX2</i>	25806		
<i>IFT88</i>	1309717	<i>GPR98</i>	84059	<i>VSX1</i>	30813		
<i>ITGA6</i>	621633	<i>GRK1</i>	6011	<i>VSX2</i>	338917		
<i>ITGB4</i>	2928	<i>GRK7</i>	131890	<i>WDR36</i>	134430		
<i>JAG2</i>	2938	<i>GRM6</i>	2916	<i>WFS1</i>	7466		
<i>KLK4</i>	1303228	<i>GUCA1A</i>	2978	<i>ZIC2</i>	7546		
<i>LAMA5</i>	621023	<i>GUCA1B</i>	2979	<i>ZNF513</i>	130557		
<i>LEF1</i>	620241	<i>GUCA1C</i>	9626	<i>RDH5</i>	5959		
<i>LHX8</i>	1308749	<i>GUCY2D</i>	3000	<i>RDH8</i>	50700		
<i>LRP4</i>	619731	<i>GUCY2F</i>	2986	<i>RGR</i>	5995		
<i>LRP6</i>	1304749	<i>GUCY2F</i>	2986	<i>RGS16</i>	6004		
<i>MMP20</i>	1308730	<i>HMCN1</i>	83872	<i>RGS9</i>	8787		
<i>MSX1</i>	620929	<i>HPS1</i>	3257	<i>RGS9BP</i>	388531		
<i>MSX2</i>	3116	<i>HPS1</i>	3257	<i>RHO</i>	6010		
<i>NF2</i>	3169	<i>IKZF1</i>	10320	<i>RIMS1</i>	0		
<i>NFIC</i>	62016	<i>IMPG1</i>	3617	<i>RLBP1</i>	6017		
<i>NGFR</i>	3177	<i>IMPG2</i>	50939	<i>ROM1</i>	6094		
<i>NKX2-3</i>	1308521	<i>IRX5</i>	10265	<i>RORB</i>	6096		
<i>ODAM</i>	1561883	<i>KCNJ10</i>	3766	<i>RP1</i>	6101		
<i>PDGFRA</i>	3284	<i>KERA</i>	11081	<i>RP1L1</i>	94137		
<i>PERP</i>	1310294	<i>KIFC3</i>	3801	<i>RP2</i>	6102		
<i>PITX2</i>	3331	<i>KRT12</i>	3859	<i>RPE65</i>	6121		
<i>PPARA</i>	3369	<i>LAMB2</i>	3913	<i>RPGR</i>	6103		
<i>PRKCB</i>	3396	<i>LAMC1</i>	3915	<i>RPGRIP1</i>	57096		
<i>PVRL1</i>	620791	<i>LAMC3</i>	10319	<i>RRH</i>	10692		
<i>RUNX2</i>	2282	<i>LRAT</i>	9227	<i>RS1</i>	6247		

<i>SCN10A</i>	3629	<i>LUM</i>	4060	<i>SAG</i>	6295		
<i>SCN5A</i>	3637	<i>MFRP</i>	83552	<i>SDR16C5</i>	195814		
<i>SHH</i>	3673	<i>MIP</i>	4284	<i>SFRP5</i>	6425		
<i>SLC24A4</i>	1305926	<i>MKKS</i>	8195	<i>SIX3</i>	6496		
<i>SLC34A1</i>	3708	<i>MYO3A</i>	53904	<i>SIX6</i>	4990		
<i>SMO</i>	3726	<i>MYO3B</i>	140469	<i>SLC24A1</i>	9187		
<i>SOSTDC1</i>	628877	<i>MYO5A</i>	4644	<i>SLC24A2</i>	25769		
<i>TBX1</i>	1307734	<i>MYO7A</i>	4647	<i>SLC45A2</i>	51151		
<i>TCF7L2</i>	1583621	<i>MYO9A</i>	4649	<i>SLC4A7</i>	9497		
<i>TFAP2A</i>	1310267	<i>MYOC</i>	4653	<i>SLITRK6</i>	84189		
<i>TNC</i>	621057	<i>NDP</i>	4693	<i>SOX14</i>	8403		
<i>TNFRSF11B</i>	619802	<i>NOB1</i>	28987	<i>SOX9</i>	6662		
<i>TP63</i>	620863	<i>NPHP3</i>	27031	<i>SPATA7</i>	55812		
<i>TRAF6</i>	1306853	<i>NR2E1</i>	7101	<i>T</i>	6862		
<i>WNT10A</i>	1307015	<i>NR2E3</i>	10002	<i>TACSTD2</i>	4070		
<i>WNT6</i>	1304559	<i>NRL</i>	4901	<i>TCTN3</i>	26123		
<i>WNT7B</i>	1311441	<i>NXNL2</i>	158046	<i>TGFBI</i>	7045		
		<i>USH2A</i>	7399	<i>TH</i>	7054		
		<i>VAX2</i>	25806	<i>TIMP3</i>	7078		
		<i>VSX1</i>	30813	<i>TRPM1</i>	4308		
		<i>VSX2</i>	338917	<i>TULP1</i>	7287		
		<i>WDR36</i>	134430	<i>TULP2</i>	7288		
		<i>WFS1</i>	7466	<i>TYR</i>	7299		
		<i>ZIC2</i>	7546	<i>UNC119</i>	9094		
		<i>ZNF513</i>	130557	<i>UNC119B</i>	84747		

Appendix B: List of genes under positive selection. Totally 427 genes with raw p-values <0.05 according to standard Likelihood Ratio Tests.

ENSEMBL transcript ID	Chinese pangolin gene ID	Malayan pangolin gene ID	Gene symbol	Description	p-value
ENST00000612742	MCP0013542	MMP0007444	<i>METTL8</i>	methyltransferase like 8	4.5E-19
ENST00000396625	MCP0014112	MMP0015713	<i>COL4A4</i>	collagen, type IV, alpha 4	8.8E-19
ENST00000424952	MCP0014989	MMP0001974	<i>ZDHHC3</i>	zinc finger, DHH C-type containing 3	6E-11
ENST00000620295	MCP0011698	MMP0006416	<i>KIF1B</i>	kinesin family member 1B	1.6E-10
ENST00000404158	MCP0013765	MMP0006992	<i>GATAD2A</i>	GATA zinc finger domain containing 2A	2.4E-10
ENST00000262134	MCP0002597	MMP0013384	<i>LPCAT2</i>	lysophosphatidylcholine acyltransferase 2	6.8E-10
ENST00000438552	MCP0013319	MMP0005369	<i>SNRPB</i>	small nuclear ribonucleoprotein polypeptides B and B1	8.3E-10
ENST00000319327	MCP0005590	MMP0004890	<i>SERINC4</i>	serine incorporator 4	9.6E-09
ENST00000465301	MCP0003852	MMP0011385	<i>RGAG1</i>	retrotransposon gag domain containing 1	2.1E-07
ENST00000256997	MCP0003917	MMP0002374	<i>ACP2</i>	acid phosphatase 2, lysosomal	4.7E-07
ENST00000170150	MCP0006225	MMP0004578	<i>BPIFB2</i>	BPI fold containing family B, member 2	5.6E-07
ENST00000262288	MCP0018898	MMP0015215	<i>SCPEP1</i>	serine carboxypeptidase 1	6.4E-07
ENST00000399518	MCP0004479	MMP0016031	<i>PLA2G4E</i>	phospholipase A2, group IVE	8.7E-07
ENST00000417439	MCP0006067	MMP0003610	<i>LTF</i>	lactotransferrin	1.8E-06
ENST00000220812	MCP0008149	MMP0010347	<i>DKK4</i>	dickkopf WNT signaling pathway inhibitor 4	2.2E-06
ENST00000311772	MCP0015226	MMP0017399	<i>PPP1R8</i>	protein phosphatase 1, regulatory subunit 8	2.7E-06
ENST00000556816	MCP0000554	MMP0002391	<i>ISCA2</i>	iron-sulfur cluster assembly 2	4E-06
ENST00000297268	MCP0005262	MMP0004567	<i>COL1A2</i>	collagen, type I, alpha 2	7.8E-06
ENST00000218867	MCP0010675	MMP0004325	<i>SGCG</i>	sarcoglycan, gamma (35kDa dystrophin-associated glycoprotein)	7.9E-06
ENST00000350051	MCP0008180	MMP0019042	<i>BIRC5</i>	baculoviral IAP repeat containing 5	1.1E-05
ENST00000307921	MCP0013631	MMP0009333	<i>ADAT1</i>	adenosine deaminase, tRNA-specific 1	1.3E-05
ENST00000252245	MCP0002974	MMP0022581	<i>KRT75</i>	keratin 75, type II	1.3E-05
ENST00000039989	MCP0006888	MMP0011614	<i>TTC17</i>	tetratricopeptide repeat domain 17	1.5E-05
ENST00000399080	MCP0016897	MMP0011951	<i>RAD51AP2</i>	RAD51 associated protein 2	1.6E-05
ENST00000372754	MCP0009233	MMP0007886	<i>MATN4</i>	matrilin 4	1.9E-05
ENST00000435159	MCP0011655	MMP0001856	<i>TMEM132C</i>	transmembrane protein 132C	2.8E-05
ENST00000430500	MCP0007111	MMP0011114	<i>SLC22A8</i>	solute carrier family 22 (organic anion transporter), member 8	2.8E-05
ENST00000330794	MCP0003498	MMP0002794	<i>TMEM173</i>	transmembrane protein 173	2.9E-05
ENST00000302555	MCP0003040	MMP0003207	<i>GP2</i>	glycoprotein 2 (zymogen granule membrane)	3.4E-05
ENST00000230568	MCP0016099	MMP0019937	<i>LY86</i>	lymphocyte antigen 86	3.5E-05
ENST00000229243	MCP0001799	MMP0002450	<i>ACRBP</i>	acrosin binding protein	4.1E-05
ENST00000318602	MCP0015520	MMP0022433	<i>A2M</i>	alpha-2-macroglobulin	4.4E-05
ENST00000355661	MCP0002865	MMP0018609	<i>PLEKHA7</i>	pleckstrin homology domain containing, family A member 7	4.7E-05

ENST00000217169	MCP0011196	MMP0023102	<i>BIRC7</i>	baculoviral IAP repeat containing 7	4.8E-05
ENST00000398189	MCP0011417	MMP0007405	<i>APOF</i>	apolipoprotein F	4.9E-05
ENST00000367843	MCP0012903	MMP0002856	<i>DCAF6</i>	DDB1 and CUL4 associated factor 6	0.00005
ENST00000407315	MCP0003536	MMP0017719	<i>THAP4</i>	THAP domain containing 4	5.3E-05
ENST00000527372	MCP0019410	MMP0019410	<i>MYO18A</i>	myosin XVIIIa	6.3E-05
ENST00000305949	MCP0004057	MMP0021541	<i>MTBP</i>	MDM2 binding protein	6.8E-05
ENST00000621805	MCP0006948	MMP0015704	<i>RAB18</i>	RAB18, member RAS oncogene family	7.9E-05
ENST00000265081	MCP0014405	MMP0008354	<i>MSH3</i>	mutS homolog 3	9.1E-05
ENST00000377122	MCP0012659	MMP0012306	<i>NEBL</i>	nebulin	0.0001
ENST00000298694	MCP0010292	MMP0010258	<i>ARHGEF40</i>	Rho guanine nucleotide exchange factor (GEF) 40	0.0001
ENST00000286428	MCP0002071	MMP0019114	<i>VBP1</i>	von Hippel-Lindau binding protein 1	0.0001
ENST00000367975	MCP0011264	MMP0021735	<i>SDHC</i>	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	0.0001
ENST00000260810	MCP0016712	MMP0022386	<i>TOPBP1</i>	topoisomerase (DNA) II binding protein 1	0.0001
ENST00000306058	MCP0016146	MMP0001730	<i>ASXL1</i>	additional sex combs like transcriptional regulator 1	0.0001
ENST00000394518	MCP0006774	MMP0016921	<i>AKAP13</i>	A kinase (PRKA) anchor protein 13	0.0002
ENST00000394931	MCP0015501	MMP0006535	<i>BMPRI1B</i>	bone morphogenetic protein receptor, type IB	0.0002
ENST00000598162	MCP0018058	MMP0004019	<i>BCAT2</i>	branched chain amino-acid transaminase 2, mitochondrial	0.0002
ENST00000535274	MCP0012621	MMP0021425	<i>B3GNT4</i>	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 4	0.0002
ENST00000380010	MCP0012645	MMP0004724	<i>SLC19A1</i>	solute carrier family 19 (folate transporter), member 1	0.0002
ENST00000389758	MCP0017570	MMP0014493	<i>MROH2A</i>	maestro heat-like repeat family member 2A	0.0002
ENST00000360351	MCP0018405	MMP0012150	<i>MAP2</i>	microtubule-associated protein 2	0.0002
ENST00000332578	MCP0008778	MMP0014322	<i>TMPPSS9</i>	transmembrane protease, serine 9	0.0002
ENST00000261652	MCP0001724	MMP0007814	<i>TNFRSF13B</i>	tumor necrosis factor receptor superfamily, member 13B	0.0002
ENST00000532211	MCP0002130	MMP0006949	<i>PIH1D2</i>	PIH1 domain containing 2	0.0003
ENST00000493960	MCP0006786	MMP0016372	<i>FAM208A</i>	family with sequence similarity 208, member A	0.0003
ENST00000265992	MCP0019338	MMP0001436	<i>CCNJ</i>	cyclin J	0.0003
ENST00000400897	MCP0009359	MMP0017922	<i>MASP2</i>	mannan-binding lectin serine peptidase 2	0.0003
ENST00000371253	MCP0001163	MMP0015162	<i>ADGRF1</i>	adhesion G protein-coupled receptor F1	0.0004
ENST00000404938	MCP0020085	MMP0019810	<i>ABCB5</i>	ATP-binding cassette, sub-family B (MDR/TAP), member 5	0.0005
ENST00000361373	MCP0019329	MMP0001401	<i>LDB3</i>	LIM domain binding 3	0.0005
ENST00000338005	MCP0000142	MMP0008507	<i>CCDC183</i>	coiled-coil domain containing 183	0.0005
ENST00000390683	MCP0006516	MMP0006681	<i>RAD51B</i>	RAD51 paralog B	0.0005
ENST00000370378	MCP0007955	MMP0008286	<i>KIAA1107</i>	KIAA1107	0.0005
ENST00000222286	MCP0001801	MMP0002446	<i>GAPDHS</i>	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	0.0006

ENST00000295500	MCP0018223	MMP0014289	<i>GPR155</i>	G protein-coupled receptor 155	0.0006
ENST00000357590	MCP0018531	MMP0009444	<i>SLC37A4</i>	solute carrier family 37 (glucose-6-phosphate transporter), member 4	0.0006
ENST00000270115	MCP0012832	MMP0023303	<i>LRRC56</i>	leucine rich repeat containing 56	0.0007
ENST00000436874	MCP0004214	MMP0002213	<i>VEZT</i>	vezatin, adherens junctions transmembrane protein	0.0007
ENST00000353364	MCP0018742	MMP0014729	<i>USP15</i>	ubiquitin specific peptidase 15	0.0007
ENST00000486257	MCP0011128	MMP0017391	<i>TIGIT</i>	T cell immunoreceptor with Ig and ITIM domains	0.0007
ENST00000371421	MCP0011485	MMP0018057	<i>ARRDC1</i>	arrestin domain containing 1	0.0008
ENST00000374075	MCP0008783	MMP0003826	<i>AKNA</i>	AT-hook transcription factor	0.0008
ENST00000381527	MCP0000833	MMP0021053	<i>CDK8</i>	cyclin-dependent kinase 8	0.0008
ENST00000529629	MCP0017036	MMP0011201	<i>CAPN5</i>	calpain 5	0.0008
ENST00000301391	MCP0015374	MMP0013367	<i>CYB5D2</i>	cytochrome b5 domain containing 2	0.0009
ENST00000325888	MCP0015344	MMP0001388	<i>FLNC</i>	filamin C, gamma	0.001
ENST00000346736	MCP0004001	MMP0016910	<i>C19orf57</i>	chromosome 19 open reading frame 57	0.001
ENST00000302787	MCP0003872	MMP0005668	<i>LETM1</i>	leucine zipper-EF-hand containing transmembrane protein 1	0.001
ENST00000360299	MCP0013647	MMP0008458	<i>RAB5B</i>	RAB5B, member RAS oncogene family	0.001
ENST00000311601	MCP0018930	MMP0009659	<i>SH3PXD2B</i>	SH3 and PX domains 2B	0.001
ENST00000414749	MCP0010652	MMP0012397	<i>MLXIPL</i>	MLX interacting protein-like	0.001
ENST00000427836	MCP0007865	MMP0007255	<i>PLEKHM3</i>	pleckstrin homology domain containing, family M, member 3	0.0011
ENST00000314583	MCP0001012	MMP0006283	<i>HCLS1</i>	hematopoietic cell-specific Lyn substrate 1	0.0012
ENST00000251643	MCP0012401	MMP0022286	<i>KRT12</i>	keratin 12, type I	0.0012
ENST00000261847	MCP0000241	MMP0014830	<i>SECISBP2L</i>	SECIS binding protein 2-like	0.0012
ENST00000373095	MCP0001404	MMP0017002	<i>FAM102A</i>	family with sequence similarity 102, member A	0.0012
ENST00000257974	MCP0013193	MMP0017552	<i>KRT82</i>	keratin 82, type II	0.0012
ENST00000250018	MCP0010835	MMP0010087	<i>TPH1</i>	tryptophan hydroxylase 1	0.0013
ENST00000483725	MCP0006513	MMP0017318	<i>KIAA0408</i>	KIAA0408	0.0014
ENST00000394479	MCP0003938	MMP0014594	<i>REL</i>	v-rel avian reticuloendotheliosis viral oncogene homolog	0.0014
ENST00000289359	MCP0005854	MMP0008461	<i>MITD1</i>	MIT, microtubule interacting and transport, domain containing 1	0.0014
ENST00000375077	MCP0017522	MMP0015837	<i>CORO2A</i>	coronin, actin binding protein, 2A	0.0015
ENST00000249776	MCP0002618	MMP0013567	<i>KNSTRN</i>	kinetochore-localized astrin/SPAG5 binding protein	0.0015
ENST00000371605	MCP0017168	MMP0005817	<i>ABCA2</i>	ATP-binding cassette, sub-family A (ABC1), member 2	0.0016
ENST00000324366	MCP0017142	MMP0001601	<i>PRMT5</i>	protein arginine methyltransferase 5	0.0016
ENST00000372236	MCP0004577	MMP0018467	<i>POLH</i>	polymerase (DNA directed), eta	0.0016
ENST00000228887	MCP0017874	MMP0014991	<i>GPRC5D</i>	G protein-coupled receptor, class C, group 5, member D	0.0017
ENST00000294618	MCP0018019	MMP0003901	<i>DOCK6</i>	dedicator of cytokinesis 6	0.0018
ENST00000269025	MCP0000730	MMP0007037	<i>LRRC46</i>	leucine rich repeat containing 46	0.002

ENST00000373368	MCP0006150	MMP0004180	<i>SPP2</i>	secreted phosphoprotein 2, 24kDa	0.002
ENST00000391588	MCP0012404	MMP0014927	<i>KRTAP3-1</i>	keratin associated protein 3-1	0.0022
ENST00000549091	MCP0003001	MMP0006493	<i>WDR90</i>	WD repeat domain 90	0.0023
ENST00000381821	MCP0002417	MMP0012852	<i>TEX33</i>	testis expressed 33	0.0023
ENST00000295897	MCP0000306	MMP0019724	<i>ALB</i>	albumin	0.0023
ENST00000361756	MCP0018455	MMP0017079	<i>RNF121</i>	ring finger protein 121	0.0023
ENST00000264661	MCP0012356	MMP0006932	<i>KCNH4</i>	potassium channel, voltage gated eag related subfamily H, member 4	0.0024
ENST00000278742	MCP0016769	MMP0018597	<i>ST14</i>	suppression of tumorigenicity 14 (colon carcinoma)	0.0026
ENST00000355898	MCP0015864	MMP0018417	<i>ZNF507</i>	zinc finger protein 507	0.0027
ENST00000251481	MCP0018685	MMP0005848	<i>SULT1C2</i>	sulfotransferase family, cytosolic, 1C, member 2	0.0027
ENST00000265838	MCP0006597	MMP0020696	<i>ACAT1</i>	acetyl-CoA acetyltransferase 1	0.0028
ENST00000328278	MCP0018196	MMP0011129	<i>LRRC14B</i>	leucine rich repeat containing 14B	0.0028
ENST00000611114	MCP0007699	MMP0004874	<i>ZNF804B</i>	zinc finger protein 804B	0.0028
ENST00000431016	MCP0013950	MMP0008195	<i>PCYT1A</i>	phosphate cytidyltransferase 1, choline, alpha	0.0029
ENST00000616727	MCP0009477	MMP0020481	<i>MUC13</i>	mucin 13, cell surface associated	0.0029
ENST00000268766	MCP0019423	MMP0018895	<i>NEK8</i>	NIMA-related kinase 8	0.0029
ENST00000306534	MCP0018265	MMP0009051	<i>ROBO4</i>	roundabout, axon guidance receptor, homolog 4 (Drosophila)	0.003
ENST00000295408	MCP0007017	MMP0010239	<i>MERTK</i>	MER proto-oncogene, tyrosine kinase	0.003
ENST00000394810	MCP0005684	MMP0005627	<i>SYNPO2L</i>	synaptopodin 2-like	0.0032
ENST00000324444	MCP0013626	MMP0013848	<i>SYNE4</i>	spectrin repeat containing, nuclear envelope family member 4	0.0032
ENST00000431282	MCP0007907	MMP0014887	<i>APOBR</i>	apolipoprotein B receptor	0.0032
ENST00000543976	MCP0014448	MMP0022930	<i>TMF1</i>	TATA element modulatory factor 1	0.0033
ENST00000215886	MCP0018753	MMP0012866	<i>LGALS2</i>	lectin, galactoside-binding, soluble, 2	0.0033
ENST00000506113	MCP0010371	MMP0021546	<i>ABLIM3</i>	actin binding LIM protein family, member 3	0.0033
ENST00000354666	MCP0017215	MMP0008867	<i>ELOVL2</i>	ELOVL fatty acid elongase 2	0.0033
ENST00000354042	MCP0003926	MMP0004701	<i>SLC13A4</i>	solute carrier family 13 (sodium/sulfate symporter), member 4	0.0034
ENST00000356575	MCP0016969	MMP0011779	<i>MEGF6</i>	multiple EGF-like-domains 6	0.0034
ENST00000244709	MCP0009903	MMP0013921	<i>TREM1</i>	triggering receptor expressed on myeloid cells 1	0.0034
ENST00000314400	MCP0015810	MMP0019471	<i>C3orf17</i>	chromosome 3 open reading frame 17	0.0035
ENST00000305544	MCP0004733	MMP0013542	<i>LAMB2</i>	laminin, beta 2 (laminin S)	0.0035
ENST00000260563	MCP0018133	MMP0011335	<i>RTCA</i>	RNA 3'-terminal phosphate cyclase	0.0035
ENST00000322344	MCP0004077	MMP0013763	<i>PNKP</i>	polynucleotide kinase 3'-phosphatase	0.0036
ENST00000313961	MCP0014297	MMP0017195	<i>RGS5</i>	regulator of G-protein signaling 5	0.0037
ENST00000328963	MCP0019554	MMP0016407	<i>P2RX7</i>	purinergic receptor P2X, ligand gated ion channel, 7	0.0038

ENST00000439706	MCP0017942	MMP0007807	<i>SLC38A1</i>	solute carrier family 38, member 1	0.0038
ENST00000452135	MCP0016083	MMP0006968	<i>MAPK9</i>	mitogen-activated protein kinase 9	0.0038
ENST00000324894	MCP0003934	MMP0001358	<i>GTPBP3</i>	GTP binding protein 3 (mitochondrial)	0.004
ENST00000226021	MCP0000027	MMP0021116	<i>CACNG1</i>	calcium channel, voltage-dependent, gamma subunit 1	0.0041
ENST00000380752	MCP0019069	MMP0003494	<i>SLC7A1</i>	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	0.0043
ENST00000342203	MCP0000920	MMP0003716	<i>SSX2IP</i>	synovial sarcoma, X breakpoint 2 interacting protein	0.0043
ENST00000263681	MCP0001224	MMP0016227	<i>POLD3</i>	polymerase (DNA-directed), delta 3, accessory subunit	0.0043
ENST00000539925	MCP0003183	MMP0002442	<i>LTBR</i>	lymphotoxin beta receptor (TNFR superfamily, member 3)	0.0045
ENST00000343629	MCP0010938	MMP0000384	<i>TLDC1</i>	TBC/LysM-associated domain containing 1	0.0045
ENST00000415318	MCP0018547	MMP0009442	<i>CCDC153</i>	coiled-coil domain containing 153	0.0045
ENST00000394511	MCP0011838	MMP0000988	<i>UGT8</i>	UDP glycosyltransferase 8	0.0046
ENST00000296684	MCP0017370	MMP0000981	<i>NDUFS4</i>	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	0.0046
ENST00000407426	MCP0014869	MMP0002002	<i>WDR43</i>	WD repeat domain 43	0.0047
ENST00000542230	MCP0007969	MMP0011163	<i>TMEM50B</i>	transmembrane protein 50B	0.0048
ENST00000274811	MCP0003067	MMP0003667	<i>RNF44</i>	ring finger protein 44	0.0048
ENST00000334418	MCP0018528	MMP0009454	<i>CCDC84</i>	coiled-coil domain containing 84	0.0048
ENST00000355303	MCP0013532	MMP0002716	<i>ANO1</i>	anoctamin 1, calcium activated chloride channel	0.0049
ENST00000264852	MCP0009814	MMP0019635	<i>SIDT1</i>	SID1 transmembrane family, member 1	0.005
ENST00000309863	MCP0005702	MMP0007961	<i>GCC2</i>	GRIP and coiled-coil domain containing 2	0.005
ENST00000215912	MCP0003785	MMP0000371	<i>PIK3IP1</i>	phosphoinositide-3-kinase interacting protein 1	0.0054
ENST00000545068	MCP0019773	MMP0005122	<i>FOXJ3</i>	forkhead box J3	0.0055
ENST00000330550	MCP0016133	MMP0015659	<i>SLC22A16</i>	solute carrier family 22 (organic cation/carnitine transporter), member 16	0.0055
ENST00000240364	MCP0005737	MMP0018438	<i>FAM117A</i>	family with sequence similarity 117, member A	0.0055
ENST00000300456	MCP0004129	MMP0016727	<i>SLC27A4</i>	solute carrier family 27 (fatty acid transporter), member 4	0.0055
ENST00000374566	MCP0018084	MMP0015061	<i>EPB41L4B</i>	erythrocyte membrane protein band 4.1 like 4B	0.0056
ENST00000301740	MCP0005104	MMP0003365	<i>SRRM2</i>	serine/arginine repetitive matrix 2	0.0056
ENST00000344846	MCP0003787	MMP0001893	<i>SYNGR4</i>	synaptogyrin 4	0.0058
ENST00000319211	MCP0019438	MMP0013162	<i>F2R</i>	coagulation factor II (thrombin) receptor	0.0059
ENST00000356090	MCP0001098	MMP0018684	<i>NCDN</i>	neurochondrin	0.0059
ENST00000300060	MCP0014718	MMP0010943	<i>ANPEP</i>	alanyl (membrane) aminopeptidase	0.006
ENST00000340413	MCP0004303	MMP0020777	<i>NUP43</i>	nucleoporin 43kDa	0.006
ENST00000245457	MCP0001670	MMP0009897	<i>PTGER2</i>	prostaglandin E receptor 2 (subtype EP2), 53kDa	0.006

ENST00000289416	MCP0013726	MMP0011504	ACSM3	acyl-CoA synthetase medium-chain family member 3	0.0062
ENST00000324001	MCP0017442	MMP0020847	PRX	periaxin	0.0064
ENST00000323563	MCP0007589	MMP0004153	MRPS31	mitochondrial ribosomal protein S31	0.0064
ENST00000402418	MCP0020037	MMP0010176	SLC25A19	solute carrier family 25 (mitochondrial thiamine pyrophosphate carrier), member 19	0.0066
ENST00000367500	MCP0018845	MMP0007546	SWT1	SWT1 RNA endoribonuclease homolog (S. cerevisiae)	0.0066
ENST00000247138	MCP0017955	MMP0017948	SLC35A2	solute carrier family 35 (UDP-galactose transporter), member A2	0.0067
ENST00000292432	MCP0003074	MMP0003660	HK3	hexokinase 3 (white cell)	0.007
ENST00000266943	MCP0004222	MMP0021379	SLC46A3	solute carrier family 46, member 3	0.007
ENST00000517768	MCP0004190	MMP0010549	MYOZ3	myozenin 3	0.0072
ENST00000391809	MCP0012341	MMP0013799	KLK5	kallikrein-related peptidase 5	0.0073
ENST00000449910	MCP0011050	MMP0010488	ADAM15	ADAM metallopeptidase domain 15	0.0074
ENST00000308278	MCP0017882	MMP0004783	FAM57A	family with sequence similarity 57, member A	0.0074
ENST00000245663	MCP0000136	MMP0003285	ZBTB46	zinc finger and BTB domain containing 46	0.0075
ENST00000330498	MCP0005710	MMP0005291	SLC5A2	solute carrier family 5 (sodium/glucose cotransporter), member 2	0.0079
ENST00000217407	MCP0019043	MMP0006507	LBP	lipopolysaccharide binding protein	0.0079
ENST00000354753	MCP0016915	MMP0004855	GPSM1	G-protein signaling modulator 1	0.0079
ENST00000301645	MCP0014225	MMP0001776	CYP7A1	cytochrome P450, family 7, subfamily A, polypeptide 1	0.008
ENST00000263593	MCP0001387	MMP0009049	SIAE	sialic acid acetyltransferase	0.008
ENST00000621632	MCP0001103	MMP0017093	DDX24	DEAD (Asp-Glu-Ala-Asp) box helicase 24	0.0081
ENST00000447092	MCP0003849	MMP0023062	HYAL2	hyaluronoglucosaminidase 2	0.0082
ENST00000359088	MCP0007332	MMP0017878	ST6GALNAC1	ST6 (alpha-N-acetyl-neuraminy-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1	0.0083
ENST00000233615	MCP0019960	MMP0021807	WBP1	WW domain binding protein 1	0.0083
ENST00000361840	MCP0010459	MMP0000525	SPRYD7	SPRY domain containing 7	0.0085
ENST00000389722	MCP0013161	MMP0018690	SPTB	spectrin, beta, erythrocytic	0.0091
ENST00000411641	MCP0015400	MMP0000222	AHSG	alpha-2-HS-glycoprotein	0.0091
ENST00000616898	MCP0003819	MMP0005509	HEMGN	hemogen	0.0091
ENST00000370630	MCP0000919	MMP0003715	CTBS	chitinase, di-N-acetyl-	0.0092
ENST00000550527	MCP0014906	MMP0012839	APAF1	apoptotic peptidase activating factor 1	0.0092
ENST00000327042	MCP0011082	MMP0012434	TMEM86B	transmembrane protein 86B	0.0092
ENST00000373574	MCP0012203	MMP0007019	WDR38	WD repeat domain 38	0.0092
ENST00000373921	MCP0015227	MMP0017398	THEMIS2	thymocyte selection associated family member 2	0.0092
ENST00000593360	MCP0015953	MMP0001364	HAUS8	HAUS augmin-like complex, subunit 8	0.0093
ENST00000325207	MCP0015257	MMP0001745	RIC8A	RIC8 guanine nucleotide exchange factor A	0.0097

ENST00000360586	MCP0000997	MMP0019124	<i>WDHD1</i>	WD repeat and HMG-box DNA binding protein 1	0.0099
ENST00000358430	MCP0003232	MMP0002103	<i>TXLNB</i>	taxilin beta	0.0099
ENST00000337526	MCP0008816	MMP0001505	<i>RTN4</i>	reticulon 4	0.01
ENST00000301919	MCP0011580	MMP0014551	<i>MSANTD4</i>	Myb/SANT-like DNA-binding domain containing 4 with coiled-coils	0.01
ENST00000352105	MCP0017673	MMP0016505	<i>ACAN</i>	aggrecan	0.0103
ENST00000209929	MCP0019534	MMP0017754	<i>FMO2</i>	flavin containing monooxygenase 2 (non-functional)	0.011
ENST00000334186	MCP0018062	MMP0004021	<i>PPFIA3</i>	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 3	0.0114
ENST00000223129	MCP0002345	MMP0014069	<i>RPA3</i>	replication protein A3, 14kDa	0.0115
ENST00000267082	MCP0009066	MMP0008491	<i>ITGB7</i>	integrin, beta 7	0.0116
ENST00000602404	MCP0005556	MMP0004720	<i>NDUFA6</i>	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	0.0117
ENST00000452691	MCP0004730	MMP0013547	<i>CCDC36</i>	coiled-coil domain containing 36	0.0117
ENST00000366844	MCP0007354	MMP0015889	<i>ENAH</i>	enabled homolog (Drosophila)	0.0117
ENST00000269373	MCP0013449	MMP0021667	<i>FN3KRP</i>	fructosamine 3 kinase related protein	0.0121
ENST00000296161	MCP0016527	MMP0006292	<i>DTX3L</i>	deltex 3 like, E3 ubiquitin ligase	0.0122
ENST00000311907	MCP0003090	MMP0018178	<i>F2</i>	coagulation factor II (thrombin)	0.0123
ENST00000308304	MCP0009102	MMP0017064	<i>PROP1</i>	PROP paired-like homeobox 1	0.0125
ENST00000457542	MCP0006577	MMP0016026	<i>MAPKBP1</i>	mitogen-activated protein kinase binding protein 1	0.0125
ENST00000307017	MCP0002694	MMP0011274	<i>USP38</i>	ubiquitin specific peptidase 38	0.0127
ENST00000398955	MCP0007453	MMP0013031	<i>MGARP</i>	mitochondria-localized glutamic acid-rich protein	0.013
ENST00000560491	MCP0007085	MMP0020598	<i>LYSMD2</i>	LysM, putative peptidoglycan-binding, domain containing 2	0.013
ENST00000354955	MCP0014950	MMP0002579	<i>FMOD</i>	fibromodulin	0.0131
ENST00000370646	MCP0003657	MMP0022557	<i>HOGA1</i>	4-hydroxy-2-oxoglutarate aldolase 1	0.0132
ENST00000260970	MCP0004809	MMP0009231	<i>PPIG</i>	peptidylprolyl isomerase G (cyclophilin G)	0.0132
ENST00000216223	MCP0002419	MMP0012854	<i>IL2RB</i>	interleukin 2 receptor, beta	0.0133
ENST00000379144	MCP0000296	MMP0013153	<i>PCYT1B</i>	phosphate cytidylyltransferase 1, choline, beta	0.0134
ENST00000370978	MCP0000772	MMP0021084	<i>ZNF280C</i>	zinc finger protein 280C	0.0135
ENST00000368662	MCP0011383	MMP0016855	<i>TUBE1</i>	tubulin, epsilon 1	0.0136
ENST00000337508	MCP0010776	MMP0010042	<i>NRIP2</i>	nuclear receptor interacting protein 2	0.0137
ENST00000522447	MCP0002806	MMP0011283	<i>LACTB2</i>	lactamase, beta 2	0.0139
ENST00000351989	MCP0013788	MMP0020268	<i>DGCR8</i>	DGCR8 microprocessor complex subunit	0.0142
ENST00000263383	MCP0008985	MMP0002425	<i>ILVBL</i>	ilvB (bacterial acetolactate synthase)-like	0.0142
ENST00000374672	MCP0006831	MMP0010874	<i>KLF4</i>	Kruppel-like factor 4 (gut)	0.0144
ENST00000428726	MCP0000978	MMP0023338	<i>CD44</i>	CD44 molecule (Indian blood group)	0.0144
ENST00000589872	MCP0016178	MMP0021992	<i>NBR1</i>	neighbor of BRCA1 gene 1	0.0148

ENST00000358242	MCP0005976	MMP0007904	<i>DMTN</i>	dematin actin binding protein	0.0151
ENST00000378313	MCP0009768	MMP0009843	<i>C19orf54</i>	chromosome 19 open reading frame 54	0.0153
ENST00000266754	MCP0019383	MMP0005009	<i>GAS2L3</i>	growth arrest-specific 2 like 3	0.0156
ENST00000361632	MCP0004387	MMP0008043	<i>CSF3R</i>	colony stimulating factor 3 receptor (granulocyte)	0.0158
ENST00000336431	MCP0011090	MMP0002047	<i>GGT7</i>	gamma-glutamyltransferase 7	0.0159
ENST00000356517	MCP0001853	MMP0019384	<i>AADACL2</i>	arylacetamide deacetylase-like 2	0.016
ENST00000342427	MCP0020078	MMP0006212	<i>ZNF341</i>	zinc finger protein 341	0.0162
ENST00000378079	MCP0017117	MMP0012259	<i>FBXO47</i>	F-box protein 47	0.0163
ENST00000361249	MCP0010977	MMP0013244	<i>C8A</i>	complement component 8, alpha polypeptide	0.0165
ENST00000373256	MCP0019426	MMP0011949	<i>GLP1R</i>	glucagon-like peptide 1 receptor	0.0165
ENST00000318663	MCP0006710	MMP0003096	<i>ORAI3</i>	ORAI calcium release-activated calcium modulator 3	0.0165
ENST00000491431	MCP0011328	MMP0020561	<i>ZNF786</i>	zinc finger protein 786	0.0166
ENST00000373795	MCP0002472	MMP0011315	<i>SRSF4</i>	serine/arginine-rich splicing factor 4	0.0167
ENST00000368842	MCP0006994	MMP0008481	<i>LHPP</i>	phospholysine phosphohistidine inorganic pyrophosphate phosphatase	0.0167
ENST00000338560	MCP0019120	MMP0016525	<i>TRPV2</i>	transient receptor potential cation channel, subfamily V, member 2	0.0169
ENST00000599564	MCP0006528	MMP0000806	<i>GRAMD1A</i>	GRAM domain containing 1A	0.017
ENST00000343484	MCP0016587	MMP0003292	<i>TCEA2</i>	transcription elongation factor A (SII), 2	0.0171
ENST00000253801	MCP0016167	MMP0021988	<i>G6PC</i>	glucose-6-phosphatase, catalytic subunit	0.0171
ENST00000341500	MCP0002795	MMP0008821	<i>INSR</i>	insulin receptor	0.0171
ENST00000378357	MCP0015565	MMP0022380	<i>CA9</i>	carbonic anhydrase IX	0.0175
ENST00000359128	MCP0000445	MMP0006146	<i>NLRC3</i>	NLR family, CARD domain containing 3	0.0175
ENST00000007722	MCP0005741	MMP0000160	<i>ITGA3</i>	integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	0.0177
ENST00000324464	MCP0009769	MMP0002064	<i>ADCK4</i>	aarF domain containing kinase 4	0.0179
ENST00000370611	MCP0012259	MMP0003720	<i>LPAR3</i>	lysophosphatidic acid receptor 3	0.018
ENST00000344657	MCP0009992	MMP0007014	<i>SPHKAP</i>	SPHK1 interactor, AKAP domain containing	0.0181
ENST00000436063	MCP0002089	MMP0018374	<i>DNTTIP2</i>	deoxynucleotidyltransferase, terminal, interacting protein 2	0.0181
ENST00000347842	MCP0000655	MMP0005502	<i>ASIC4</i>	acid sensing (proton gated) ion channel family member 4	0.0182
ENST00000320027	MCP0004993	MMP0018827	<i>HSPH1</i>	heat shock 105kDa/110kDa protein 1	0.0185
ENST00000521315	MCP0005965	MMP0008119	<i>SFTPC</i>	surfactant protein C	0.0187
ENST00000344206	MCP0006746	MMP0023042	<i>MST1R</i>	macrophage stimulating 1 receptor	0.0188
ENST00000487519	MCP0006585	MMP0009797	<i>TFAM</i>	transcription factor A, mitochondrial	0.0194
ENST00000396091	MCP0013441	MMP0012082	<i>ANO10</i>	anoctamin 10	0.0194
ENST00000372838	MCP0004130	MMP0012292	<i>CERCAM</i>	cerebral endothelial cell adhesion molecule	0.0195
ENST00000072869	MCP0014531	MMP0010736	<i>ADCK2</i>	aarF domain containing kinase 2	0.0195
ENST00000274605	MCP0010862	MMP0016560	<i>N4BP3</i>	NEDD4 binding protein 3	0.0195

ENST00000378750	MCP0001807	MMP0006321	<i>PEX16</i>	peroxisomal biogenesis factor 16	0.0197
ENST00000262753	MCP0007362	MMP0003307	<i>POF1B</i>	premature ovarian failure, 1B	0.0201
ENST00000358807	MCP0008894	MMP0008203	<i>MICAL1</i>	microtubule associated monooxygenase, calponin and LIM domain containing 1	0.0204
ENST00000290472	MCP0017819	MMP0016029	<i>PLA2G4D</i>	phospholipase A2, group IVD (cytosolic)	0.0204
ENST00000486442	MCP0011737	MMP0018623	<i>KLHL29</i>	kelch-like family member 29	0.0207
ENST00000268379	MCP0014411	MMP0019697	<i>UQCRC2</i>	ubiquinol-cytochrome c reductase core protein II	0.0207
ENST00000370759	MCP0000461	MMP0021023	<i>GIPC2</i>	GIPC PDZ domain containing family, member 2	0.0209
ENST00000504930	MCP0006210	MMP0017961	<i>POLR3G</i>	polymerase (RNA) III (DNA directed) polypeptide G (32kD)	0.0209
ENST00000357175	MCP0017012	MMP0000275	<i>MUM1L1</i>	melanoma associated antigen (mutated) 1-like 1	0.0213
ENST00000225174	MCP0011163	MMP0010105	<i>PIIF</i>	peptidylprolyl isomerase F	0.0213
ENST00000281722	MCP0018684	MMP0001866	<i>RBM46</i>	RNA binding motif protein 46	0.0214
ENST00000402105	MCP0013036	MMP0000963	<i>HPS4</i>	Hermansky-Pudlak syndrome 4	0.0215
ENST00000301908	MCP0005275	MMP0003962	<i>PNOC</i>	prepronociceptin	0.0216
ENST00000356443	MCP0016738	MMP0011564	<i>MYOM1</i>	myomesin 1	0.0219
ENST00000257909	MCP0015119	MMP0018357	<i>TROAP</i>	trophinin associated protein	0.022
ENST00000342315	MCP0019800	MMP0010055	<i>OAS2</i>	2'-5'-oligoadenylate synthetase 2, 69/71kDa	0.0221
ENST00000256190	MCP0005252	MMP0019435	<i>SBF2</i>	SET binding factor 2	0.0223
ENST00000328119	MCP0012382	MMP0015051	<i>KRT36</i>	keratin 36, type I	0.0224
ENST00000273183	MCP0002165	MMP0015548	<i>STAC</i>	SH3 and cysteine rich domain	0.0225
ENST00000370853	MCP0013876	MMP0001106	<i>MBNL3</i>	muscleblind-like splicing regulator 3	0.0226
ENST00000551812	MCP0011596	MMP0006972	<i>BAZ2A</i>	bromodomain adjacent to zinc finger domain, 2A	0.0226
ENST00000263816	MCP0016283	MMP0000802	<i>LRP2</i>	low density lipoprotein receptor-related protein 2	0.0227
ENST00000382751	MCP0010551	MMP0010509	<i>URB1</i>	URB1 ribosome biogenesis 1 homolog (S. cerevisiae)	0.0228
ENST00000617259	MCP0012431	MMP0003984	<i>IL13</i>	interleukin 13	0.023
ENST00000253686	MCP0001504	MMP0007745	<i>MRPS25</i>	mitochondrial ribosomal protein S25	0.0231
ENST00000378043	MCP0007191	MMP0016825	<i>BEST1</i>	bestrophin 1	0.0232
ENST00000287295	MCP0018314	MMP0021081	<i>AIFM1</i>	apoptosis-inducing factor, mitochondrion-associated, 1	0.0235
ENST00000228936	MCP0012592	MMP0023387	<i>ART4</i>	ADP-ribosyltransferase 4 (Dombrock blood group)	0.0238
ENST00000216274	MCP0014040	MMP0022997	<i>RIPK3</i>	receptor-interacting serine-threonine kinase 3	0.024
ENST00000320005	MCP0002024	MMP0018307	<i>CNGB3</i>	cyclic nucleotide gated channel beta 3	0.0242
ENST00000229729	MCP0014390	MMP0018839	<i>SLC44A4</i>	solute carrier family 44, member 4	0.0243
ENST00000294818	MCP0016138	MMP0014854	<i>LRRC52</i>	leucine rich repeat containing 52	0.0244
ENST00000318315	MCP0006161	MMP0022978	<i>C5orf46</i>	chromosome 5 open reading frame 46	0.0244
ENST00000369478	MCP0009172	MMP0001338	<i>CD2</i>	CD2 molecule	0.0245
ENST00000316428	MCP0003115	MMP0002258	<i>LRRC31</i>	leucine rich repeat containing 31	0.0248
ENST00000392588	MCP0008467	MMP0016019	<i>WASF1</i>	WAS protein family, member 1	0.025

ENST00000279873	MCP0015156	MMP0010966	<i>ARID5B</i>	AT rich interactive domain 5B (MRF1-like)	0.0251
ENST00000301671	MCP0012355	MMP0006933	<i>GHDC</i>	GH3 domain containing	0.0253
ENST00000358241	MCP0015861	MMP0011975	<i>RTP2</i>	receptor (chemosensory) transporter protein 2	0.0254
ENST00000252032	MCP0012682	MMP0003973	<i>C20orf194</i>	chromosome 20 open reading frame 194	0.0254
ENST00000254928	MCP0019425	MMP0018894	<i>ERAL1</i>	Era-like 12S mitochondrial rRNA chaperone 1	0.0256
ENST00000216471	MCP0014469	MMP0019215	<i>SAMD15</i>	sterile alpha motif domain containing 15	0.0256
ENST00000223026	MCP0013849	MMP0019748	<i>HYAL4</i>	hyaluronoglucosaminidase 4	0.0259
ENST00000255039	MCP0006306	MMP0000774	<i>HAPLN2</i>	hyaluronan and proteoglycan link protein 2	0.026
ENST00000366518	MCP0017619	MMP0021968	<i>KIF26B</i>	kinesin family member 26B	0.0262
ENST00000297596	MCP0018348	MMP0016530	<i>GEM</i>	GTP binding protein overexpressed in skeletal muscle	0.0265
ENST00000371901	MCP0006938	MMP0003865	<i>CYP4X1</i>	cytochrome P450, family 4, subfamily X, polypeptide 1	0.0269
ENST00000367025	MCP0002075	MMP0019880	<i>TRAF3IP3</i>	TRAF3 interacting protein 3	0.0269
ENST00000594369	MCP0007755	MMP0022304	<i>ZNF446</i>	zinc finger protein 446	0.0271
ENST00000268124	MCP0017667	MMP0007595	<i>POLG</i>	polymerase (DNA directed), gamma	0.0274
ENST00000371980	MCP0012607	MMP0003592	<i>LURAP1</i>	leucine rich adaptor protein 1	0.0275
ENST00000389194	MCP0003387	MMP0020809	<i>LTN1</i>	listerin E3 ubiquitin protein ligase 1	0.0275
ENST00000544455	MCP0016017	MMP0012495	<i>BRCA2</i>	breast cancer 2, early onset	0.0278
ENST00000372479	MCP0002921	MMP0017770	<i>RBM41</i>	RNA binding motif protein 41	0.0282
ENST00000454584	MCP0012178	MMP0018514	<i>GAS2</i>	growth arrest-specific 2	0.0282
ENST00000264499	MCP0009711	MMP0007162	<i>BBS7</i>	Bardet-Biedl syndrome 7	0.0283
ENST00000310823	MCP0013431	MMP0011525	<i>ADAM17</i>	ADAM metalloproteinase domain 17	0.0285
ENST00000447906	MCP0016536	MMP0011234	<i>OTUD4</i>	OTU deubiquitinase 4	0.0286
ENST00000548660	MCP0016668	MMP0013267	<i>GLT8D2</i>	glycosyltransferase 8 domain containing 2	0.0291
ENST00000408910	MCP0011260	MMP0012771	<i>UMODL1</i>	uromodulin-like 1	0.0291
ENST00000330342	MCP0005412	MMP0009538	<i>ATP6V0A2</i>	ATPase, H ⁺ transporting, lysosomal V0 subunit a2	0.0294
ENST00000315939	MCP0014652	MMP0002522	<i>WNK1</i>	WNK lysine deficient protein kinase 1	0.0295
ENST00000409652	MCP0009467	MMP0010667	<i>APOL6</i>	apolipoprotein L, 6	0.0298
ENST00000552570	MCP0009953	MMP0022593	<i>TNS2</i>	tensin 2	0.0299
ENST00000441366	MCP0013779	MMP0015144	<i>EPB42</i>	erythrocyte membrane protein band 4.2	0.03
ENST00000613904	MCP0000800	MMP0022899	<i>TTI2</i>	TELO2 interacting protein 2	0.0301
ENST00000531316	MCP0000865	MMP0022729	<i>C11orf63</i>	chromosome 11 open reading frame 63	0.0302
ENST00000619684	MCP0013903	MMP0004499	<i>KIAA1524</i>	KIAA1524	0.0305
ENST00000263663	MCP0000943	MMP0018392	<i>TAF1B</i>	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa	0.0307
ENST00000266971	MCP0013648	MMP0004763	<i>SUOX</i>	sulfite oxidase	0.0309
ENST00000202556	MCP0003446	MMP0012094	<i>PPP1R13B</i>	protein phosphatase 1, regulatory subunit 13B	0.0311
ENST00000254508	MCP0009571	MMP0006747	<i>NUP210</i>	nucleoporin 210kDa	0.0312

ENST00000302250	MCP0002176	MMP0009244	<i>FAM151A</i>	family with sequence similarity 151, member A	0.0313
ENST00000382492	MCP0019195	MMP0001867	<i>TAS2R1</i>	taste receptor, type 2, member 1	0.0317
ENST00000441564	MCP0020157	MMP0018382	<i>PSD4</i>	pleckstrin and Sec7 domain containing 4	0.0317
ENST00000223836	MCP0017617	MMP0007936	<i>AK1</i>	adenylate kinase 1	0.0318
ENST00000358755	MCP0005575	MMP0017408	<i>FZD6</i>	frizzled class receptor 6	0.032
ENST00000225275	MCP0016854	MMP0005707	<i>MPO</i>	myeloperoxidase	0.0322
ENST00000395048	MCP0017309	MMP0003875	<i>CYP1A1</i>	cytochrome P450, family 1, subfamily A, polypeptide 1	0.0323
ENST00000395343	MCP0011403	MMP0023107	<i>DIDO1</i>	death inducer-obliterator 1	0.0325
ENST00000547303	MCP0013065	MMP0004933	<i>DDIT3</i>	DNA-damage-inducible transcript 3	0.0331
ENST00000371941	MCP0005304	MMP0009152	<i>PREX1</i>	phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 1	0.0334
ENST00000399645	MCP0000449	MMP0006148	<i>C16orf90</i>	chromosome 16 open reading frame 90	0.0334
ENST00000271636	MCP0018789	MMP0006678	<i>CGN</i>	cingulin	0.0335
ENST00000378910	MCP0003244	MMP0000620	<i>NPHS1</i>	nephrosis 1, congenital, Finnish type (nephrin)	0.0336
ENST00000373347	MCP0014149	MMP0011869	<i>DLGAP3</i>	discs, large (Drosophila) homolog-associated protein 3	0.0337
ENST00000373362	MCP0014152	MMP0011866	<i>GJB3</i>	gap junction protein, beta 3, 31kDa	0.0337
ENST00000339777	MCP0012322	MMP0017512	<i>LRRC43</i>	leucine rich repeat containing 43	0.0339
ENST00000340990	MCP0015698	MMP0021217	<i>ADIPOR1</i>	adiponectin receptor 1	0.0339
ENST00000378673	MCP0016480	MMP0008644	<i>GDF9</i>	growth differentiation factor 9	0.0343
ENST00000228837	MCP0014289	MMP0021398	<i>FGF6</i>	fibroblast growth factor 6	0.0343
ENST00000361256	MCP0011630	MMP0008295	<i>C9orf114</i>	chromosome 9 open reading frame 114	0.0345
ENST00000216487	MCP0006109	MMP0021879	<i>RIN3</i>	Ras and Rab interactor 3	0.0348
ENST00000320634	MCP0011619	MMP0004344	<i>FAIM2</i>	Fas apoptotic inhibitory molecule 2	0.035
ENST00000374954	MCP0013496	MMP0001124	<i>ASIP</i>	agouti signaling protein	0.0353
ENST00000573584	MCP0010752	MMP0006662	<i>NUP88</i>	nucleoporin 88kDa	0.0353
ENST00000409991	MCP0018538	MMP0009448	<i>NLRX1</i>	NLR family member X1	0.0354
ENST00000340611	MCP0001629	MMP0023234	<i>BRAT1</i>	BRCA1-associated ATM activator 1	0.0355
ENST00000331456	MCP0006745	MMP0023043	<i>TRAIP</i>	TRAF interacting protein	0.0356
ENST00000401399	MCP0019651	MMP0023154	<i>NFASC</i>	neurofascin	0.0358
ENST00000370017	MCP0012422	MMP0018767	<i>FNDC7</i>	fibronectin type III domain containing 7	0.0358
ENST00000263314	MCP0016354	MMP0005860	<i>P2RX3</i>	purinergic receptor P2X, ligand gated ion channel, 3	0.0359
ENST00000398712	MCP0008237	MMP0020956	<i>SHARPIN</i>	SHANK-associated RH domain interactor	0.0359
ENST00000551568	MCP0015111	MMP0009618	<i>CPM</i>	carboxypeptidase M	0.036
ENST00000357089	MCP0004560	MMP0005830	<i>UBXN11</i>	UBX domain protein 11	0.0361
ENST00000380379	MCP0005453	MMP0017402	<i>BPHL</i>	biphenyl hydrolase-like (serine hydrolase)	0.0362
ENST00000495893	MCP0018184	MMP0007567	<i>PHC3</i>	polyhomeotic homolog 3 (Drosophila)	0.0363
ENST00000518444	MCP0006591	MMP0023036	<i>LARP4</i>	La ribonucleoprotein domain family, member 4	0.0363

ENST00000258403	MCP0009994	MMP0006068	<i>SLC19A3</i>	solute carrier family 19 (thiamine transporter), member 3	0.0366
ENST00000409423	MCP0015797	MMP0004754	<i>NCAPG2</i>	non-SMC condensin II complex, subunit G2	0.037
ENST00000287497	MCP0004385	MMP0004069	<i>ITGAM</i>	integrin, alpha M (complement component 3 receptor 3 subunit)	0.0375
ENST00000409655	MCP0019334	MMP0001742	<i>ATHL1</i>	ATH1, acid trehalase-like 1 (yeast)	0.0377
ENST00000374479	MCP0001536	MMP0020606	<i>FUCA1</i>	fucosidase, alpha-L- 1, tissue	0.0378
ENST00000239440	MCP0014237	MMP0002154	<i>ARAP3</i>	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3	0.0381
ENST00000355710	MCP0019667	MMP0010315	<i>RET</i>	ret proto-oncogene	0.0387
ENST00000376236	MCP0017424	MMP0007618	<i>APBB1IP</i>	amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein	0.0388
ENST00000376042	MCP0014429	MMP0012318	<i>CCNB3</i>	cyclin B3	0.0393
ENST00000371139	MCP0014209	MMP0015425	<i>SH2D1A</i>	SH2 domain containing 1A	0.0394
ENST00000226067	MCP0017959	MMP0023337	<i>HLF</i>	hepatic leukemia factor	0.0394
ENST00000392692	MCP0009724	MMP0009221	<i>ECT2</i>	epithelial cell transforming 2	0.0399
ENST00000457717	MCP0015300	MMP0000748	<i>MTTP</i>	microsomal triglyceride transfer protein	0.0401
ENST00000372626	MCP0010273	MMP0005152	<i>TCEAL1</i>	transcription elongation factor A (SII)-like 1	0.0406
ENST00000315251	MCP0005367	MMP0004513	<i>CHDH</i>	choline dehydrogenase	0.0411
ENST00000262415	MCP0010589	MMP0001013	<i>DHX8</i>	DEAH (Asp-Glu-Ala-His) box polypeptide 8	0.0411
ENST00000325680	MCP0008831	MMP0001371	<i>YLPM1</i>	YLP motif containing 1	0.0414
ENST00000328886	MCP0001924	MMP0010218	<i>TMIGD1</i>	transmembrane and immunoglobulin domain containing 1	0.0414
ENST00000449682	MCP0012616	MMP0015381	<i>MST1</i>	macrophage stimulating 1	0.0415
ENST00000378387	MCP0015555	MMP0022373	<i>ARHGEF39</i>	Rho guanine nucleotide exchange factor (GEF) 39	0.0416
ENST00000263577	MCP0007381	MMP0007092	<i>CDON</i>	cell adhesion associated, oncogene regulated	0.0422
ENST00000284288	MCP0001385	MMP0009043	<i>PANX3</i>	pannexin 3	0.0426
ENST00000261170	MCP0012594	MMP0011166	<i>GUCY2C</i>	guanylate cyclase 2C	0.0431
ENST00000217026	MCP0017044	MMP0018634	<i>MYBL2</i>	v-myb avian myeloblastosis viral oncogene homolog-like 2	0.0434
ENST00000261435	MCP0000775	MMP0019092	<i>N4BP2</i>	NEDD4 binding protein 2	0.0434
ENST00000427704	MCP0016402	MMP0009616	<i>PHACTR2</i>	phosphatase and actin regulator 2	0.0439
ENST00000406927	MCP0013369	MMP0020781	<i>METTL21A</i>	methyltransferase like 21A	0.0439
ENST00000555619	MCP0000558	MMP0002393	<i>NPC2</i>	Niemann-Pick disease, type C2	0.0442
ENST00000222598	MCP0019547	MMP0016736	<i>DLX5</i>	distal-less homeobox 5	0.0442
ENST00000340360	MCP0013583	MMP0011026	<i>XRR1</i>	X-ray radiation resistance associated 1	0.0443
ENST00000392542	MCP0009499	MMP0015985	<i>RFC5</i>	replication factor C (activator 1) 5, 36.5kDa	0.0449
ENST00000286096	MCP0010880	MMP0012089	<i>KDM8</i>	lysine (K)-specific demethylase 8	0.045
ENST00000447830	MCP0010287	MMP0008659	<i>SPATC1</i>	spermatogenesis and centriole associated 1	0.045
ENST00000216445	MCP0013018	MMP0019280	<i>C14orf105</i>	chromosome 14 open reading frame 105	0.0453
ENST00000265983	MCP0001047	MMP0002130	<i>HPX</i>	hemopexin	0.0453

ENST00000257789	MCP0012558	MMP0017456	<i>ORC3</i>	origin recognition complex, subunit 3	0.0461
ENST00000403389	MCP0007994	MMP0013200	<i>OSM</i>	oncostatin M	0.0464
ENST00000257189	MCP0014019	MMP0017373	<i>DSG3</i>	desmoglein 3	0.0467
ENST00000411463	MCP0005155	MMP0008143	<i>R3HCC1</i>	R3H domain and coiled-coil containing 1	0.0471
ENST00000297581	MCP0011582	MMP0010726	<i>DCSTAMP</i>	dendrocyte expressed seven transmembrane protein	0.0474
ENST00000281282	MCP0010015	MMP0015394	<i>CGNL1</i>	cingulin-like 1	0.0475
ENST00000378588	MCP0013522	MMP0014350	<i>CYBB</i>	cytochrome b-245, beta polypeptide	0.0475
ENST00000312777	MCP0009217	MMP0002230	<i>TCHP</i>	trichoplein, keratin filament binding	0.0479
ENST00000434748	MCP0001658	MMP0022723	<i>FBRSL1</i>	fibrosin-like 1	0.0484
ENST00000585618	MCP0004820	MMP0000552	<i>SEC14L1</i>	SEC14-like 1 (<i>S. cerevisiae</i>)	0.0485
ENST00000498508	MCP0004308	MMP0022478	<i>PROX1</i>	prospero homeobox 1	0.049
ENST00000325203	MCP0005437	MMP0008689	<i>ANGPT2</i>	angiopoietin 2	0.049
ENST00000389005	MCP0000413	MMP0022658	<i>C17orf85</i>	chromosome 17 reading frame 85	0.0493
ENST00000296137	MCP0016953	MMP0015637	<i>FYCO1</i>	FYVE and coiled-coil domain containing 1	0.0496

Appendix C: ENAM alignment of pangolin species with mammals.

<i>Homo sapiens</i>	ATGTTTGAACAAAGAT--TTTGAACAAACCCAAAGAAAGATCCTCTAAAGCAGAAAGTC	58
<i>Canis familiaris</i>	ATGTTTGAACAAAGAT--TTTGAACAAACCCAAAGAAAGATCCTCTAAAGCAGAGATC	58
<i>Felis catus</i>	ATGTTTGAACAAAGAT--TTTGAACAAACCCAAAGAAAGATCCTCTAAAGCAGAGATC	58
<i>Manis pentadactyla</i>	--ATGCTTGAACAAAGATTTTAAACCCAAAGAAAGATCCTCTAAAGCAGAGATC	58
<i>Manis javanica</i>	ATGCTTGAACAAAGATTTTAAATCCCAAGA--AAAAATATCCTCTCTAAAGCAGAGATC	58
	* * * * *	
<i>Homo sapiens</i>	CAGGCACAGAACCCACAGCTAATTCAACAGTCACTGAGACGAATTCTACCCCAACCAATC	118
<i>Canis familiaris</i>	CGGCCGTGAGCCCTCGGGTAATTCAACAGGTCCCGAGACTAATTCTACTCAATCAAAATC	118
<i>Felis catus</i>	CAGCCACTGAACCCCTCGGGTAATTCAACAGGTCTGAGACTAATTCTACTCAACCAAAATG	118
<i>Manis pentadactyla</i>	CAGCCCTGAACCCCTCAGGGGAATTCAACAGGTCTGAGACTAATTTTACCCGGATAAAATG	118
<i>Manis javanica</i>	CAGTCCCTGAACCCCTCAAGGAATTCAACAGGTCTGAGACTAATTTTACCCAGATAAAATG	118
	* * * * *	
<i>Homo sapiens</i>	---CTAAAGGGAGTCAAGGAGGAAATGACACCAGCCCCACAGGAAACAGTACCCAGGAC	175
<i>Canis familiaris</i>	---CTGAGGGAGTCAAGTGGAAATGACACCAGCCCCAACAGGAAACAGTGGCCCTGGCC	175
<i>Felis catus</i>	CACCTGGAGGAAATCAGGCGGAAATGACACCAGCCCCAACAGGAAACAGTGGCCATGGCC	178
<i>Manis pentadactyla</i>	CACCCAGTCCAAAGAAAGTCAAGTGGAAATG-----ATACCAGCCCAACTG	166
<i>Manis javanica</i>	CACCCAGTCCAAAGAAAGTCAAGTGGAAATG-----ATACCAGCCCAACTG	166
	* * * * *	
<i>Homo sapiens</i>	TAAACACTGGGAACAAACCCCTCCAGCTCAAAATGGGATTGGCCCACTCCCTGCAGTCAACG	235
<i>Canis familiaris</i>	CAAAACACTGTAGTAATCCTACGGTCAAAACGGGGTTATCTCACCCGTACAGTTAATA	235
<i>Felis catus</i>	CAAAACACTGTAGTAATCCTACAGTCAAAACGGGGTTATCTCACCCCTACAGTTAATG	238
<i>Manis pentadactyla</i>	CAAAACACTGTAGCAATCCTACAGTCAAAATGGGGTTATCCCTACAG-----TTACTAT	221
<i>Manis javanica</i>	GAAACACTGTAGCAATCCTACAGTCAAAATGGGACCATCCCTGCAG-----TTACCAT	221
	***** * * * * *	
<i>Homo sapiens</i>	CTTCAGGCCAGGGAGGGCCAGGAAGTCAAAATCCCATGGAGACCAAGTCAGCCAAATATTC	295
<i>Canis familiaris</i>	TTTCAGGCCAGGGAGTACCAAGAACTCAAAATCTCATGGGGACCAAAATCAGCCAAATATTC	295
<i>Felis catus</i>	TTTCAGGCCAGGGAGTACCAAGAACTCAAAATCTCATGGGGACCAAGTCAGCCAAATATTC	298
<i>Manis pentadactyla</i>	TTTCAGGCCCTGG-GA---GTGCCAAGTAGAATTCATGGGGACCAAGTCAGCCAAATATTC	277
<i>Manis javanica</i>	TTTCAGGCCCTGGGAGTGCCAAAGTAGAATTCATGGGGATCAAGTCAGCCAAATATTC	281

Homo sapiens
Canis familiaris
Felis catus
Manis pentadactyla
Manis javanica

GTTGCCCACTGGGTCCCAAACTGGCCCTGTTGTTGCGAATGAAAAATCCAAAAATCCA
 654
 GTTGCTCCTCTGGGTCCCAAAACATGGTACTGTTGTCCACAATGAAAAATCCAAAAATCCA
 654
 GTTGCTCCTCTGGGTCCCAAAACATGGTACTGTTGTCCGCAATGAAAAATCCAAAAATCCA
 657
 ATTGTCCTCTGGGTCCCAAAACATGGTACTGTTGTATCACAAAGAAAAACATCCAAAAATCCA
 627
 ATTGTCCTCTGGGTCCCAAAACATGGTACTGTTGTACCCACAAAGAAAAACATCCAAAAATCCA
 624
 *** ** * ***** ** * ***** * *** ***** *****

Homo sapiens
Canis familiaris
Felis catus
Manis pentadactyla
Manis javanica

AAGGAGAAAGCCCTGGGTCCAAAAAGAAACAAATAATAGTTTCTTACAAAGAAATCCAACCCAGC
 714
 GGAGAGAAACCCAGTAGGCCCAAAAGAAAGAAATAGTTCATTTCTTACAAAGGATCCATCTGGC
 714
 GCAGAGAAACCCAGTAGGTCCAAAAAGAAAGAAATAGTTCATTTCTTACAAAGAGATCCATCTGGC
 717
 AGACAGAAAGCCACTGGGTCAAAAAAGAAAGAAATGGTTCATTTCCAC-AGGATCCAACTGGC
 686
 GGACAGAAAGCCACTTGGTCAAAAAAGAAAGAAATGGTTCATTTCCAC-AGAGATCCAACTGGC
 683
 **** ** * ***** ** * ***** * *** * ***** * **

Homo sapiens
Canis familiaris
Felis catus
Manis pentadactyla
Manis javanica

CCCTGGAGAAACTCTCAACAGTATGAAGTTAATAAAATCAAAATTAATAACTGCCTCACTCT
 774
 CCCTGGAGAAACTCTCAAGATTATGGAGTTAAACAAATCAAACTATAAACTGTCTCCCCCT
 774
 CCCTGGAGAAACTCTCAAGATTATGGAGTTAATAAAATCAAACTATAAACTGCCTCATCCT
 777
 CCCTGGAGAAACTCTCAAGACTATGGAGTTAATAAAACCAAACTATAAACTGCCTCACCT
 746
 TGCTAGAGAAACTCTCAAGACTATGGAAATTAATAAAACCAAACTATAAACTGCCTCACCT
 743
 ** ***** ** * ***** * ***** * ***** * ***** *

Homo sapiens
Canis familiaris
Felis catus
Manis pentadactyla
Manis javanica

GAGGGTTATATGCCAGTCCCAAAATTTAAATTCGTGTT---GATCAACATGAAAACTCCTAT
 831
 GAGAGTAATGTGCTAGTCCCAAAATTTAAATTCGTGTT---GATCAACATGAAAACTCCTAT
 831
 GAGAGTAACATGTAGTCCCAAAATTTAAATTCGTGTT---GGTCAACATGAAAACTCCTAT
 834
 GAGGATAACATGTAGTCCCAAAATTTAAATTCGTGTT---TGATCAAAAGCAAACTCCTAT
 803
 AAGGGTAACATGTAGTCCCTAAATTTAAATTCGTGTTGATCAAAAGCAAACTCCTAT
 803
 ** * * ***** * ***** * ***** * ***** *

Homo sapiens
Canis familiaris
Felis catus
Manis pentadactyla
Manis javanica

TACCCAAAGAGGAGATTCCAGAAAAAGTCCCAAAATTCGTATGGACAAACCCAAAGCCAGAAAT
 891
 TACTCAAGAGGAGATTCCAGAAAGAGCCCCAAATTCGTATGGACAAACCCAAAGCCAGAAAT
 891
 TACCCAAAGAGGAGATTCCAAAAAGAGCCCCAAATTCGTATGGACAAACCCAAAGCCAGAAAT
 894
 TAGCCAAAGAGGAGATTCCAGAAAAAGCCCCAAATTCGTATGGACAACTGAAAGCCAGATT
 863
 TAGCCAAAGAGGAGATTCCAGAAAAAGCCCCAAATTCGTATGAACAAGCTGAAAGCCAGATT
 863
 ** ***** ** * ***** ** * ***** * ***** *

<i>Homo sapiens</i>	TTGCCCAAAGGGATTGTTTAGGGTCAAGAAGGATGCCATATGAATCAGAACTAATCAG	951
<i>Canis familiaris</i>	TTGCCCAAAGGGATTATTTTAGAGCCAAGAAGAAACCCGTATGAATCAGAACTAATCGG	951
<i>Felis catus</i>	TTGCCCAAAGGGATTATTTTGGAGCCAAGAAGAAATCCCTTATGAATCAGAACTAATCGA	954
<i>Manis pentadactyla</i>	TTGCCCAAAGGAATTATCTTAGAGTCAAGAAGAAATCCTCTATAAAATCAGAAATTAATCAT	923
<i>Manis javanica</i>	TTGCCCAAAGGAATTATCTTAGAGCCAAGAAGAAATCCTGTATATATCAGAAATTAATCAT	923
	***** *	
<i>Homo sapiens</i>	TCAGAAATTAAGCACAGCTCATATCAGCCTGCTGTATACCCCTGAGGAAATCCCTTCTCCT	1011
<i>Canis familiaris</i>	CCAGAAATTAAGCACAGTACATATCAAC---CTGTATTTCCCTGAGGAAATTCCTCCCCCT	1008
<i>Felis catus</i>	CCAGAAATTAAGCACAGTACATATCAAC---GTGTATACCCCTGAGGAAATCCCTTCCACT	1011
<i>Manis pentadactyla</i>	CCAAAATTGAACCATAGTACATACCAGC---CTGTATACCCCTGAAGAAATCCCTTCCCCCT	980
<i>Manis javanica</i>	CCAAAATTGAACCATATACATACCAGC---CTGTATACCCCTGAGGAAATCCCTTCCCCCT	980
	** **** ** ** * **** * * * **** * * * * * * * * * * * * * * * *	
<i>Homo sapiens</i>	GCAAAAGAACATTTTCCTGCTGGAAGAAATACTTTGGACCACCAAGAAATCTCTCCACCT	1071
<i>Canis familiaris</i>	GCAAGAGAACATTTTCCTGCTGGAAGAAATACTTTGGCATCACCAAGAAATCTCTCCATCT	1068
<i>Felis catus</i>	GCAAGAGAACATTTTCCTGCTGGAAGAAATATTTTGAATCATCAAGGAAATCTCTCCACCT	1071
<i>Manis pentadactyla</i>	GCAAGAGAACATTTTCCTGTTGGAAGAAACACTTTGAACCACCAAGAAACCTCTCTCCACCT	1040
<i>Manis javanica</i>	GCAAGAGAACATTTTCCTGTTGGAAGAAACA-TTGAACCACCAAGAAACCTCTCTCCACCT	1039
	**** ***** ** ***** * **** * * * **** * * ***** **	
<i>Homo sapiens</i>	TTTAAGGAAGATCCAGGGAGGCAAGAAGAACATTTACCCCATCCTTCCCATGGTTCTAGA	1131
<i>Canis familiaris</i>	TTTAAGGAAGATCCTGGGAGGCAGGAGGAACCTTTACCTCCTCCTTCCCAGGGCTTAGG	1128
<i>Felis catus</i>	ATTAAGGAAGATCCTGGGATGCAGGAAGAACACTTACTTCATCCTTCCCATGGCTCAAGG	1131
<i>Manis pentadactyla</i>	TATAAGGAAGATCCCAGGATGCAGGAAGAACACTTACCTCATCCTTCCCGTGACTTTAGA	1100
<i>Manis javanica</i>	TATAACGAAGATCCTGGAAATGCAGGAAGAACACCTACCTCATCCTTCCCGTGCTTTAGA	1099
	*** ***** * * * **** * * **** * * ***** * * * * *	
<i>Homo sapiens</i>	GGAAAGTGTTTTCTACCCCTGAATATAACCCATATGATCCCAGGGGAAAACTCACCATACCTT	1191
<i>Canis familiaris</i>	GGAGGTGTTTACTACCCCTGACTATAACTTTATGATCCCAGGGGAAAAATTCACCATACCTT	1188
<i>Felis catus</i>	GGAGGTGTTTACTACCCCTGACTATAACTTTATATATCCCAGGGGAAAACTCACCATACCTT	1191
<i>Manis pentadactyla</i>	GGAAGTGTTTACTACCGTGACTTTAACCCGATGATCCCAAGGAAAAATTCATCATACCTT	1160
<i>Manis javanica</i>	GGAAGTGTTTACTACTGTGACTTTTAACCTCTATGATCCCAAGGAAAAATTCACCATACCTT	1159

<i>Homo sapiens</i>	AGAGGCAATACATGGGATGAGAGAGATGATTCTCCCAATACATATGGGGCAAAAAGAAAGT	1251
<i>Canis familiaris</i>	AGAAAGCAATAGATGGGATGAGAGAGATGATTCTCCCAACACAAATAGGGCAACCCAGAAAAAT	1248
<i>Felis catus</i>	AGAAAGCAATACATGGGATGAGAGAGATGATTCTCCGAAATACATATGGGGCAACCCAAAAAAT	1251
<i>Manis pentadactyla</i>	AGAAAGCAATTATGATGAGAGAGAGATGATTCTCCAGTACTAAATAC-----CAAAAAAT	1214
<i>Manis javanica</i>	AGCAGCAATTATGATGAGAGAGAGATGATTCTCCAGTACTAAATAC-----CAAAAAAT	1213
	** ***** ** ***** ** ***** ** ***** ** ***** ** *****	
<i>Homo sapiens</i>	CCACTCTACCCCAATAAATACCCAGACCAGAGAGATAGTCCCTTTAATAATGAAGAGGAC	1311
<i>Canis familiaris</i>	TGCGTATATCCCATAAATACTCCAGAGCTGAAAGAAAACAGTCCCTTTACAATGAAGAGGAC	1308
<i>Felis catus</i>	TCACTGTACCCCATGAATACTCCAGACCTGAAAGAAAACAGTCCCTTTAATAATGAAGAGGAC	1311
<i>Manis pentadactyla</i>	CCAAATGTGCCCATGAATACTCCAAATCCAAATCCAAAGAGACAGTCCCTTTCTAATGAAGAGGAC	1274
<i>Manis javanica</i>	CCAAATGTACCCACGAATACTCCATATCCAAATCCAAAGAGACAGTCCCTTTCTAATGAAGAGGAC	1273
	* *	
<i>Homo sapiens</i>	CCAGTTGATCCAACTGGAGATGAAGTCTTTCTCTGGACAAAAATAGATGGGTGAA---GAG	1368
<i>Canis familiaris</i>	CCAAATTGATCCAACTGGAGATGAAAATTTCCAGGACAAAAATAGATGGGTATGGAAGAG	1368
<i>Felis catus</i>	CCAAATTGATGCAAACTGGAGATGAAAATTTTCCAGGACAAAAATAGATGGGTATGGAAGAG	1371
<i>Manis pentadactyla</i>	CCAAATTGATCCAACTGGAGATGAACCTTTCCAGGACTAAGTAGATGGAGTACAAAAGGAG	1334
<i>Manis javanica</i>	CCAAATTGATCCAACTGGAGATGAACCTTTCCAGGCAAAAATAGATGGAGTACAAAAGGAG	1333
	*** *	
<i>Homo sapiens</i>	TTGAGCTTCAAAAGGAGGCCCAACAGTTAGGCACATATGAAGGTGAACAAATATACCTCAAAAT	1428
<i>Canis familiaris</i>	CCAAAGCTTTAAAGAAAGTCCCAACAGTTAGGCACATATGAAGGCGAGCAATATACCGTAAAT	1428
<i>Felis catus</i>	CCGAGCTTTAAAGAAAGGCGCAACAGTTAGGCACATATGAAGGCGAGCAATATACCTCAAAAT	1431
<i>Manis pentadactyla</i>	TCAAAGCTTTCCAGAAAGTCCCAACAGTTAGG--CATATGAAGGTGAGCAATATACCTCAAAAT	1393
<i>Manis javanica</i>	TCAAAGCTTTCCAGAAAGTCCCAACAGTTAGG--CATATGAAGGTGAGCAATATACCTCAAAAT	1392
	***** ** ** ***** *	
<i>Homo sapiens</i>	CAGCCAAAAGGAATATCTTCCCTATTCTTTAGATAATCCATCAAAAAACCAAGGAGGATTTT	1488
<i>Canis familiaris</i>	CAACCAAAAGGAATACCTTCCCTATTCTTTAGATAATCCATCAAAAAACCAAGGAGGATTTT	1488
<i>Felis catus</i>	CAACCGAAAAGGAATACCTTCCCTATTCTTTAGATAATCCATCGAAAACCAAGGAGGATTTT	1491
<i>Manis pentadactyla</i>	CAACCAAAAGGAATACCTTCCCTATTCTTTAAATAATCCATCAAAAAACCAAGGGAATATTTT	1453
<i>Manis javanica</i>	CAACCAAAAGGAATACCTTCCCTATTCTTTAAATAATCCATCAAAAAACCAAGGGAATATTTT	1452
		

Homo sapiens 1548
Canis familiaris 1548
Felis catus 1551
Manis pentadactyla 1513
Manis javanica 1512
TATTACAGTGAAATTTTACCCATGGAGCCCGGATGAGAAATTTTCCATCATATAATACAGCT
CCTTATGGTGAGTTTACCCCTGGAACCCAGATGAGAAATTTTCCATCATATAATACAGCT
CCTTATGGTGAAATTTTATCCCTGGAACCAAGATGAGAAATTTTCCATCATATAATACAGCT
CCTTATGGTGAGTTTACCCCTGGAACCCAGATGAGAAATTTTCCATCATATAATACAGCT
CCTTATGGTGAGTTTACCCCTGGAACCCAGATGAGAAATTTTCTATCATATAATACAGTT
*** *** ***** ** ***** * * * * *

Homo sapiens 1607
Canis familiaris 1607
Felis catus 1610
Manis pentadactyla 1572
Manis javanica 1572
TCT-ACTATGCCACCACCTATAGAGAGCAGGGGCTACTACGTTAATAATGCCGCTGGACC
CCC-ACTGTACCACCACCTGTGGAGAGCAGGGGCTACTATGCTAATAATGCTGTCAGACA
CCC-AGTGCACCACCACCTGTGGAGAGCAGGGGCTATTATGCTAATAATGCTGTTGGACA
CCC-ACTGTAACACCACATATGGAGAGCAGGGGCTCTTATGCTGATAATGCTGTTGGACA
CCCACCTGTAACACCACATGTGGAGAGCAGGGGCTCTTATGCTGATAATGCTTTTGGACA
* * * * *

Homo sapiens 1667
Canis familiaris 1667
Felis catus 1670
Manis pentadactyla 1632
Manis javanica 1632
AGAAGAAAGCACTCTATTTTCCTTTCACGGAATTCCTGGGACCACAGGATACAAGCCCAAGG
AGAAGAAAGCCCTCTGATTCTCTTGTGAACTCCTGGGACCACAGGGTTCAAGCCCAAGG
GGAAGAAAGCACTCTGTTCCCTCTTGGAACTCCTGGGACCATGAGATTCAAGCCCAAGG
AGAAGAAAGCACTCTGATTCCCTTTTCTGAACTCCTGGGACCACAGGATTCAAGCCCAAGG
AGAAGAAAGCACTCTGATTCCCTTTTCTGAACTTCTGGGACCACAGCATTCAGGCCCAAGG
**** * * * * *

Homo sapiens 1727
Canis familiaris 1727
Felis catus 1730
Manis pentadactyla 1692
Manis javanica 1692
GCAGAGAGAAAGAGCGCGTATTTTAAACAGAAATATCTGGGATCAGGCAACACATTTACA
ACAGAAAGAAAGGACGCCATATTTTAAACAGAAATTAATCTGGGATCAGCCAACTTTACA
GCAGAAAGAAAGAGACCATATTTTAAACAGAACTTCTGGGATCGACCAACTTTACA
GCAGAAAGAAAGAGGCTATATTTGAAACAGAAATTTCTGAGATCAGCCAACTTTACA
GCAGAAAGAAAGAGGCTATATTTGAAATAGAAATTTCTGGGATCAACCAACTTTTACA
**** * * * * *

Homo sapiens 1787
Canis familiaris 1787
Felis catus 1790
Manis pentadactyla 1752
Manis javanica 1752
AAAAGCCCCAGTAGGCCACCAGACACAGAAAGGTAACAGCCCTATTACAGTAACACCCC
CAAAGCGCCCCCTAGTCCACCACACAGAAAGAGAACCCCTATCCTAGTAATTCCTCC
CAAAGCTCCCCCTAGTCCACCACATGAGAAAGAGAACCCCTTATCCTAGCAATTCCTCC
CAAAGACCTGCTAGTCCACCACACAGAAAGAGAAATCAGCCTTATTTTCAGTAACCTCCA
CAAACACCTGCTAGTCCACCACACAGAAAGAGAACCCCTTATTTTCAGTAACCTCCA

<i>Homo sapiens</i>	AGCTGGGCTTCAGAAAAATCCAAATATGGCATGAAGGTGAGAAATTTGAACTATGGCATGCA	1847
<i>Canis familiaris</i>	AGCTGGGCTTCAGAAAAATCCAAATATGGCATGAAGGTGAGAAATTTGAACTATGGCATGCA	1847
<i>Felis catus</i>	AGCTGGGCTTCAGAAAAAGTCCAAATATGGCATGAAGGTGAGAAATTTGAACTATGGCATGCA	1850
<i>Manis pentadactyla</i>	AGCTGAACCTTCAGAAAAATCCAAATATGGCATGAAGGTGAGAAATTTGAGTAATGGCATACA	1812
<i>Manis javanica</i>	AGCTGAACCTTCAGAAAAATCCAAATATGGCATGAAGGTGAGAAATTTGAACTATGGCATACA	1812
	***** ***** *	
<i>Homo sapiens</i>	AATAACTAGGATGAATTCCTCCAGAGAGAGAAACATTCATCTTTCCCTAACTTCATCCACC	1907
<i>Canis familiaris</i>	GATTACTAGGTAAATTCACCAAGAGAGAGAAACATTTGGCTTTCCCAAGATTTAAATTCCTCC	1907
<i>Felis catus</i>	CATTACTAGGATAAATTCACCAAGAGAGAGAAACAGTTGGCTTTCCCAAGATTTAAATTCCTCC	1910
<i>Manis pentadactyla</i>	AATAACTAGGTAAATTCACCCGAGGAGAGAAACATCTGACTGTCCAAAGACTTAAATTCCTCC	1872
<i>Manis javanica</i>	AATTACTAGGTAAATTCACCAAGAGGAGAAACATCTGGCTTTCCAAAGACTTAAATTCCTCT	1872
	** ***** * ***** ** *	
<i>Homo sapiens</i>	AAGTTACCCATCAGGTCAAAAAAGAGCACATTTATTTACCTAAAGCCAGAGAGGCTCTTG	1967
<i>Canis familiaris</i>	AAGTTACCCAGAGGTCAAAAAAGAGCACATGTATTTACCTAAAGCCAGAGAGGCTCTTG	1967
<i>Felis catus</i>	AAGTTACCCATCAGGTCAAAAAAGAGCACATGTATTTACCTAAAGCCAGAGAGGCTCTTG	1970
<i>Manis pentadactyla</i>	AAGTTACCCAGTAGGTCAAAAAAGAGCACATTTATCTCACCTAAAGCTACAGAGGTCATTG	1932
<i>Manis javanica</i>	AAGTTACCCAGTAGGTCAAAAAAGAGCACATTTATCTCACCTAAAGCTACAGAGGTCGTTG	1932
	***** * ***** *	
<i>Homo sapiens</i>	CTGTGCTGGTAGCTCCACAGG-----GCCCAAGGACAAATCCACTAGCTCTACAAG	2017
<i>Canis familiaris</i>	CTGTGCTGGTAGGCTCCACAGG-----ACACAAGGACAAATCCACTTGTCTTACAGG	2017
<i>Felis catus</i>	CTGTGCTGGAGCTCCACAGG-----ACACAAGGAAAGATCCACTTGTCTTACAGG	2020
<i>Manis pentadactyla</i>	CTATGCTGGTGTTCCTACTATGACACCACTTGCCCAATGACAAATCCACTTGTCTTACACAAG	1992
<i>Manis javanica</i>	CTATGCTGGTGTTCCTACTACAACCACTTGCCCAAGGACAAATCCACTTGTCTTACACAAG	1992
	** ***** *	
<i>Homo sapiens</i>	ACTACACTCCATCCTATGGTCTTGCACTTGGGAGAAACCAAGACACCAGTCTCTGTATA	2077
<i>Canis familiaris</i>	ACTACACTCCACCTTCGATCTTGCACTTGGCACTTGGGAGAAATGAAGACACCAGTCTCTGTATA	2077
<i>Felis catus</i>	ACTACACTCCATCCTTTGGTCTTGCACTTGGGAGAAACCAAGACACCAGTCTCTGTATA	2080
<i>Manis pentadactyla</i>	ACTACACTTATCTTTTGTCTTTGCACTTGGCACTTGGGAGAAACCAAGACACCAGTCTGTATA	2052
<i>Manis javanica</i>	ACCACACTTATCTTTTGTCTTTGCACTTGGCACTTGGGAGAAACCAAGACACCAGTCTGTATA	2052

<i>Homo sapiens</i>	CCAGCTCAATGAAAAAAGAACTGTTGACCTTACTCCTTGAGCAGAGCTTGTATTATGGTACACCTGA	2420
<i>Canis familiaris</i>	CCAGCTCAGTGAAGAAACCGTTGACCTTACTCCTTGAAACAACTTGTATTATGGTACACCTGA	2435
<i>Felis catus</i>	CCAGCTCAATGAAAAAGAACTGTTGACCTTACTCCTTGAGCAGAGCTTGTATTATGGGTACACCTGA	2438
<i>Manis pentadactyla</i>	CCAGCTCAATGAAAGAACTGTTGGCCCTACTCCTTGAGCAGAGCTTGTATTATCAGTACACCTGA	2411
<i>Manis javanica</i>	CCAGCTCAATGAAAGAACTGTTGGCCCTACTCCTTGAGCAGAGCTTGTATTATCAGTACACCTGA	2411

<i>Homo sapiens</i>	TGAAGGGCTCCAATCCAGAAGGCATCCAAAAGTCAAAGTCCAAAGAAAAATGAGAGTGAGAGGCA	2480
<i>Canis familiaris</i>	TGAAGGGCCCTGAGCCAGAAAGGAATCCAAAAGTCAAAGTCAAAGGAAATGAGGGTGACAGGCA	2495
<i>Felis catus</i>	TGAAGGCCCTTAATCCAGAAGGAATCCAAAAGTCAAAGTCAAAGGAAATGAGGGTGAAAGGCA	2498
<i>Manis pentadactyla</i>	TGAAGGGCCCAAGCCAGAAAGAAATCCAAAAGTCAAAGTCAAAGGAAATGAAAGGTGAAAGGCA	2471
<i>Manis javanica</i>	TGAAGGGCCCAAGCCAGAAAGAAATCCAAAAGTCAAAGTCAAAGGAAATGA - GGTGAAAGGCA	2470

<i>Homo sapiens</i>	ACAGCAAAAGACCATCTAAACATTTCTGCAATTTGOCATGCTTTGGCTCCAAAATTAGCAAAAGCA	2540
<i>Canis familiaris</i>	GCAACAAAAGACCGTCAAGGCATCATACAGTTACCATGCTTTGGCTCCAAAATAACAAAACTA	2555
<i>Felis catus</i>	GCAACAAAAGACCATCAAGGCATCCTGCAAGTTACCATGCTTTGGCTCCAAAATTGGCAAAAGTA	2558
<i>Manis pentadactyla</i>	A---CAAAAGACATCTAGCATCCTACAGTCACCTTGCTTTGGCTCCAAAATTGGCAAAAGTA	2528
<i>Manis javanica</i>	A---CAAAAGACATCTAGCATCCTACAGTCACCTTGCTTTGGCTCCAAAATTGGCAAAAGTA	2527

<i>Homo sapiens</i>	TCACTCTTCCACCACCGAACTCCATCTAGCATGGAAGGCAAAAGCCCATTTGATGGGGA	2600
<i>Canis familiaris</i>	CCACTCTCCAGCACAGGAACTCCATCTAGCATCGGAAGGCAAGGCCCATTTGATGAAGA	2615
<i>Felis catus</i>	TCAATCTCCAGCACTGGAACTCCATCTGCGCATTTGGAAGGCAAGGCCCATTTGATGAAGA	2618
<i>Manis pentadactyla</i>	TCACTCTCCAGCACTGGAACTCCATCTAGTGTGGAAGGCTAGGCCCATTTGATGGGGA	2588
<i>Manis javanica</i>	TCACTCTCCAGCACTGGAACTCCATCTAGTGTGGAAGGCTAGGCCCATTTGAAAGGGA	2587

<i>Homo sapiens</i>	TTCAATTACGCCCTACTGAAAAATCCTAAACATTTGGTTGAGTTAGCTACTGAGGAACAATT	2660
<i>Canis familiaris</i>	GCCGATTATGCCCTACCGAAAAATCCTAACTCATTTGCTAGGTTAGCTACTGGGGCACAGTT	2675
<i>Felis catus</i>	TCCGATTACGCCCTACTGAAAAGTCTAACTCATTTGCTAGGTTAGCTACTGGGGCACAGTT	2678
<i>Manis pentadactyla</i>	TCCAATTATACCTACTGA - ATTCCTAACCTCGTTGGTGGGTTAACTACTGGGGCACAGTT	2647
<i>Manis javanica</i>	TCCAATTATGCCCTACTGAAAATTCCTAACTCATTTGGTGGGTTAACTACTGGGGCACAAATT	2647

<i>Homo sapiens</i>	TAAGAGTATAAATGTAGACCCACTTGTATGCAGATGAACACACAGTCCATTGTGAATTCCCTTCA	2720
<i>Canis familiaris</i>	TCAGAGTATAAATGTAGACCCACTTAGTGTGATGAACACACACTCCATTGTGATTCCTCTTCA	2735
<i>Felis catus</i>	TCAGAGTATAAATGTTGACCCACTTAATGCAGATGAACACACATTTCAATTGATTCCTCTTCA	2738
<i>Manis pentadactyla</i>	TCAGAGTATAAATGTAGACCTACGTAATGCAGATGAACACACTGCTTTTGTATCCCCCTTCA	2707
<i>Manis javanica</i>	TCAGAGTATAAATGTAGACTTACTTAAATGCAGATGAACACACTGCTTTTGTATCCCCCTTCA	2707
	* ***** *	
<i>Homo sapiens</i>	AAGAGGGACCAATGTACAGGACCAGGTACAAGACTGCTTACTTACTTACGGCCTAG-----	2775
<i>Canis familiaris</i>	ATTAGGGACCAATCCACAGGACCATGTACAAGACTGCTTGTGCTTCAGGCCCTAGTGCTT	2795
<i>Felis catus</i>	AATAGAGACCAATCCACAGGACCATGTGCAAGACTGCTTACTTACTTACGGCCTAGTACTA	2798
<i>Manis pentadactyla</i>	AATAGGGACCAATCCACAGGACCAGACACAAGAATGCTTACTTACTTCA-----	2755
<i>Manis javanica</i>	AATGGGGACCAATCCACAGGACCAGACACAAGAATGCTTACTTACTTCA-----	2755
	* *	

Appendix D: AMELX alignment of pangolin species with mammals.

<i>Homo sapiens</i>	-----CCCCCTTTGAAAGTGG-----TACCAGAGACATA	135
<i>Canis familiaris</i>	-----GCTTACCCCTCTGAAGTGGTACCAGAACATGATA	138
<i>Felis catus</i>	ATCGATATTGACAGAGACTGCGATTAGTCTCACTCCTCTGAAGTGGTACCAGAACATGATA	180
<i>Manis temminckii</i>	ATTGCTACTTCTCCAGCTGCAAGTCAATG---GAGCTCATGGTAAACCTGACTCTTTGTT	68
<i>Manis tricuspis</i>	ATTGCTACTTCT-CCAGCTGCAAGTCAATGG---AGCGCATGGAACCCCAACTCTTTGTT	67
<i>Manis tetradactyla</i>	ATTGCTACTTCTCCAGCTGCAAGT-CAATGG---AGCGCATGGAACCCCAACTCTTTGTT	67
<i>Manis gigantea</i>	CGACTCTTTGTTTCTCACTGGTACCCCTTCTTATGTTACGGACCCATGGGTGGATGGCTG	82
<i>Manis javanica</i>	GGGGAT-----CTGGATTT-----TATCTGCCTGCCTCCTGGGAGCAGCCTTT	45
<i>Manis pentadactyla</i>	GGGGATCTGG-----ATTTTATC-----TGCCCTGCCT-CCTGGGAGCAGCCT-TCGC	47
<i>Homo sapiens</i>	AGGCCACCGTACCCCTTCCTATGGTTACGAGCCCATGGGTGGATGGGTGCACCACCAAAATC	195
<i>Canis familiaris</i>	AGGCATCCGTACCCCTTCCTATGGTTACGAAACCCCATGGGTGGATGGGTGCACCACCAAAATC	198
<i>Felis catus</i>	AGGCATCCGTACCCCTTCCTATGGTTACGAAACCCCATGGGTGGATGGGTGCACCACCAAAATC	240
<i>Manis temminckii</i>	TCTCACTGGTACCCCTTCCTATGGTTACGAAACCCCATGGGTGGATGGGTGCACCACCAATATC	128
<i>Manis tricuspis</i>	CCTCACTGGTACCCCTTCCTATGGTTACGAAACCCCATGGGTGGATGGGTGCACCACCAATATC	127
<i>Manis tetradactyla</i>	TCTCACTGGTACCCCTTCCTATGGTTACGAAACCCCATGGGTGGATGGGTGCACCACCAATATC	127
<i>Manis gigantea</i>	CACCAC-----CATATC	94
<i>Manis javanica</i>	GCCATGCCCTTACCCCTTCCTATGGATACGAAACCCCATGGGCGGATGGGTGCACCACCAAAATC	105
<i>Manis pentadactyla</i>	TATGCCCGTTACCCCTTCCTATGGATACGAAACCCCATGGGTGGATGGGTGCACCACCAAAATC	107
	** ** *	
<i>Homo sapiens</i>	ATCCCCGTGTGTCCCAACAGCACCCCGGACTCACACCCCTGCAGCCTCATCACCACATC	255
<i>Canis familiaris</i>	ATTCGCCGTGTGTCCAGCAAAATCCCTCGAATCACGCCCTGCAGCCTCATCACCACATC	258
<i>Felis catus</i>	ATTCGCCGTGTGTCCAGCAAAATCCCTCGAATCACGCCCTGCAGCCTCATCACCACATC	300
<i>Manis temminckii</i>	ATTCGCCGTGTGTCCCTGCAGCC-----TCATCCCCACATC	164
<i>Manis tricuspis</i>	ATTCGCCGTGTGTCCCTGCAGCC-----TCATCCCCACATC	163
<i>Manis tetradactyla</i>	ATTCGCCGTGTGTCCCTGCAGCCTAATCC-----CCTAATCCCCACATC	171
<i>Manis gigantea</i>	ATTCGCCGTGTGTCCCTGCAGCC-----TCATCCCCACATC	130
<i>Manis javanica</i>	ATTCGCCGTGTGTCCAGCAAAATCCCGGAATTATACCCCTGCAGCCTCATCACCACACC	165
<i>Manis pentadactyla</i>	ATTCGCCGTGTGTCCAGCAAAATCCCGGAATCACACCCCTGCAGCCTCATCACCACACC	167

<i>Homo sapiens</i>	CCAGTGGTGCCAGCTCAGCAGCCCGTGTATCCCCCAGCAACCAATGATGATGCCCGTTCCTGGC	315
<i>Canis familiaris</i>	CCCATGGTGCCAGCTCAGCAGCCCGTGGTCCCCCAGCAACCAATGATGATGCCAGTTCTCTGGC	318
<i>Felis catus</i>	CCCATGGTGCCAGCTCAGCAGCCCGTGGTCCCCCAGCAACCAATGATGATGCCAGTTCTCTGGC	360
<i>Manis temminckii</i>	CCCACGGTGCCAGCTCGGAGCCTGTGGT-CCCCAGGAATCAATGATGCCAGTTCTCTGGC	223
<i>Manis tricuspis</i>	CCCACGGTGCCAGCTCGGAGCCTGTGGT-CCCCAGGAATCAATGATGCCAGTTCTCTGGC	222
<i>Manis tetradactyla</i>	CCCACGGTGCCAGCTCGGAGCCTGTGGT-CCCCAGGAATCAATGATGCCAGTTCTCTGGC	230
<i>Manis gigantea</i>	CCCACGGTGCCAGCTCGGAGCCTGTGGT-CCCCAGGAATCAATGATGCCAGTTCTCTGGC	189
<i>Manis javanica</i>	CCCATGGTGCCAGCTCAGCAGCCTGTGGTCCCCCAGCAATCAGTATGCCAGTTCTCTGGC	225
<i>Manis pentadactyla</i>	CCCGTGGTGCCAGCTCAGCAGCCTGTAGTCCCCCAGCAATCAGTATGCCAGTTCTCTGGC	227
	** ***** ** * * ***** ** * * ***** * ***** *	
<i>Homo sapiens</i>	CAACACTCCATGACTCCAATCCAACACCACACAGCCAAACCTCCCTCCGCCCGCCAGCAG	375
<i>Canis familiaris</i>	CAACACTCCATGACTCCAACCCCAACACCACACAGCCAAACCTCCCTCTGCCGTGCCAGCAG	378
<i>Felis catus</i>	CAACACTCCATGACTCCAACCCCAACACCACACAGCCAAACCTCCCTCTGCCGTGCCAGCAG	420
<i>Manis temminckii</i>	CAACATTACATGACTCCAAACCCAACTCCACACAGACAAGCCTCCCTCTACATCCCCAGCAG	283
<i>Manis tricuspis</i>	CAACATTCCATGACTCCAAACCCAGCTCCACACAGACAAGCCTCCCTCTACAGCCCCAGCAG	282
<i>Manis tetradactyla</i>	CAACATTCCATGACTCCAAACCCAACTCCACACAGACAAGCCTCCCTCTACAGCCCCAGCAG	290
<i>Manis gigantea</i>	CAACATTCCATGACTCCAAACCCAACTCCACACAGACAAGCCTCCCTCTACAGCCCCAGCAG	249
<i>Manis javanica</i>	CAACATTCCATGACTCCGACCCAACTCCACACAGACAAGCCTCCCTCTACACCCCCAGCAG	285
<i>Manis pentadactyla</i>	CAACATTCTAGGACTCTGACCCAACTCCACACAGGCAAAACCTCCCTCTACACCCCCAGCAG	287
	***** * * ***** * * * * * ***** * * * * * ***** *	
<i>Homo sapiens</i>	CCCTACAGCCCCAGCCTGTTCAGCCACAGCCTCACAGCCCATGCAGCCCCAGCCACCT	435
<i>Canis familiaris</i>	CCCTTCAGCCCCAGCCCGTCCAGCCACAGCCTCACAGCCCATCCAGCCCCAGCCACCC	438
<i>Felis catus</i>	CCCTTCAGCCCCAGCCCGTCCAGCCGAGCCTCACAGCCCATCCAGCCCCAGCCACCC	480
<i>Manis temminckii</i>	CCCTTCAGCCCC-----AGCCACCC	304
<i>Manis tricuspis</i>	CCCTTCAGCCCC-----AGCCACCC	303
<i>Manis tetradactyla</i>	CCCTTCAGCCCC-----AGCCACCC	311
<i>Manis gigantea</i>	CCCTTCAGCCCC-----AGCCACCC	270
<i>Manis javanica</i>	CCCTTCAGCCTCAGCC-----	303
<i>Manis pentadactyla</i>	CCCTTCAGCCTCAGCC-----	305
	**** * ***** *	

<i>Homo sapiens</i>	GTGACCCCATGAGCCCTGCGGCACAGCCACCTCTGCGCTCCGATGTTCCCCATGCAG	495
<i>Canis familiaris</i>	GTGACCCCATCCAGCCCTGCGGCACAGCCACCTCTGCGCTCCGATGTTCCCCATACAG	498
<i>Felis catus</i>	GTGACCCCATCCAGCCCTGCGGCACAGCCACCTCTGCGCTCCGATGTTCCCCATACAG	540
<i>Manis temminckii</i>	ATGTGCCCCATTCA-----G	319
<i>Manis tricuspis</i>	GTGTGCCCCATTACGCCCTGCGGCACAGCCACCTCTGCGCTCCAAATGTTCCCCATGCAG	363
<i>Manis tetradactyla</i>	ATGTGCCCCATTACGCCCTGCGGCACAGCCACCTCCGCTCCGATGTTCCCCATGCAG	371
<i>Manis gigantea</i>	ATGTGCCCCATTACGCCCTGCTGCCACAGCCACCTCTGCGCTCCGATGTTCCCCATGCAG	330
<i>Manis javanica</i>	-----CTGCCACACAGCCACCTCTGCGCTCCGATGTTCCCCATGCAG	345
<i>Manis pentadactyla</i>	-----CTGCCGCACAGCCACCTCTGCGCTCCGATGTTCCCCATGCAG	347
	*	
<i>Homo sapiens</i>	CCCCTGCTCCCATGCTTCTGATCTGACTCTGGAAGCTTGGCCATCAACAGACAAGACC	555
<i>Canis familiaris</i>	CCCCTGCCCCCATGCTTCTGACCTGCGCTCTGGAAGCTTGGCCAGCAACAGACAAGACC	558
<i>Felis catus</i>	CCCCTGCCCCCATGCTTCTGACCTGCGCTCTGGAAGCTTGGCCAGCAACAGACAAGACC	600
<i>Manis temminckii</i>	CCCCTGCTCTCCATGCTTCTGACCTGCGCTCTGGAAGCTTGGCCAGCAACAGACAAGACC	379
<i>Manis tricuspis</i>	CCCCTGCCCCCGTGTCTCTGACCTGCGCTCCAGAAAGCTTGGCCAGCAACAGACAAGACC	423
<i>Manis tetradactyla</i>	CCCCTGCTCCCCGTGTCTCTGACCTGCGCTCCAGAAAGCTTGGCCAGCAACAGACAAGACC	431
<i>Manis gigantea</i>	CCCCTGCTCCCCCATGCTTCTGACCTGCGCTCCAGAAAGCTTGGCCAGCAACAGACAAGACC	390
<i>Manis javanica</i>	CCCCTGCTCCCCGTGTCTCTGACCTGCGCTCCAGAAAGCTTGGCCAGCAACAGACAAGACC	405
<i>Manis pentadactyla</i>	CG-----TCCCCCGTGTCTCTGACCTGCGCTCCAGAAAGCTTGGCCAGCAACAGACAAGACC	403
	* * * * * * * * * * * * * * * * *	
<i>Homo sapiens</i>	AAGCGGAGGAAGTGGATTAA-----	576
<i>Canis familiaris</i>	AAGCGGAGGAAGTGGATTAA-----	579
<i>Felis catus</i>	AAGCGGAGGAAGTGGATTAA-----	621
<i>Manis temminckii</i>	AAGCGGGA-CAAGTGTGAGTATACCTTGAAGCCACTAGAACATCTGTGAGAAATAGTAAA	438
<i>Manis tricuspis</i>	AAGCGGACCAAGTGTGACTATACCTTGAAGCCACTAGAACATCTGTGAGAAATAGTAAA	483
<i>Manis tetradactyla</i>	AAGCGGACCAAGTGTGAGTATACCTTGAAGCCACTAGAACATCTGTGAGAAATAGTAAA	491
<i>Manis gigantea</i>	AAGCGGGA-CAAGTGTGAGTATACCTTGAAGCCACTAGAACATCTGTGAGAAATAGTAAA	449
<i>Manis javanica</i>	AAGCGGAGCTAGTG-----	420
<i>Manis pentadactyla</i>	AAGCGGAGCGAGTG-----	418
	***** ****	

<i>Homo sapiens</i>	-----	576
<i>Canis familiaris</i>	-----	579
<i>Felis catus</i>	-----	621
<i>Manis temminckii</i>	GCAAAACTGACCCCTAGAGTCTAAGCTCTCGAAACAAATCTAGGGTCTAGAGTTTCAGTAAC	498
<i>Manis tricuspis</i>	GCAAAACTGGCCCTAGAGTCTAAGTTCTCGAAACAAATCTAGGGTCTAGAGTTTCAGTAAC	543
<i>Manis tetradactyla</i>	GCAAAACTGGCCCTAGAGTCTAAGTTCTCGAAACAAATCTAGGGTCTAGAGTTTCAGTAAC	551
<i>Manis gigantea</i>	GCAAAACTGACCCCTAGAGTCTAAGTTCTCGAAACAAATCTAGGGTCTAGAGTTTCAGTAAC	509
<i>Manis javanica</i>	-----	420
<i>Manis pentadactyla</i>	-----	418

<i>Homo sapiens</i>	-----	576
<i>Canis familiaris</i>	-----	579
<i>Felis catus</i>	-----	621
<i>Manis temminckii</i>	CAGAGA 504	
<i>Manis tricuspis</i>	CAGAGA 549	
<i>Manis tetradactyla</i>	CAGAGA 557	
<i>Manis gigantea</i>	CAGAGA 515	
<i>Manis javanica</i>	-----	420
<i>Manis pentadactyla</i>	-----	418

Appendix E: AMBN alignment of pangolin species with mammals.

<i>Homo sapiens</i>	GCCTATGGAGCCATGTTTCCAGGATTTGGAGGATGAGGCCCGGCTTTGAGGGAATGCCCC	879
<i>Canis familiaris</i>	GCCTATGGAGCCATGTTTCCAGGATTTGGAGGATGAGGCCCTAACCTTTGGAGGGATGCCCC	801
<i>Felis catus</i>	GCCTATGGAGCCATGTTTCCAGGATTTGGAGGATGAGGCCCTAACCTTTGGCGGGATACCC	1181
<i>Manis temminckii</i>	GCCTATGGAGCCGTTTCCAGGATTTGGAGGATGAGGCCCTAGCCTTTGGAGGGATGCCCC	293
<i>Manis tricuspis</i>	GCCTATGGAGCCATGTTTCCAGGATTTGGAGGATGAGGCCCTAGCCTTTGGAGGGATGCCCC	298
<i>Manis tetradactyla</i>	GCCTATGGAGCCATGTTTCCAGGATTTGGAGGATGAGGCCCTAACCTTTGGAGGGATGCCCC	294
<i>Manis gigantea</i>	GCCTATGGAGCCGTTTCCAGGATTTGGAGGATGAGGCCCTAGCCTTTGGAGGGATGCCCC	293
<i>Manis javanica</i>	GCCTATGGAGCTGTGTTTCCAGGATTTGGAGGATGAGGCCCTAGCCTTTGGAGGGATGCCCC	259
<i>Manis pentadactyla</i>	GCCTATGGAGCTGTGTTTCCAGGATTTGGAGGATGAGGCCCTAGCCTTTGGAGGGATGCCCC	266
	***** *	
<i>Homo sapiens</i>	CACAAACCAGCTATGGGCGGTGACTTCACTCTGGAATTTGACTCCCCAGTGGCTGCCACC	939
<i>Canis familiaris</i>	CACAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCCCAGTCGCGGGGACC	861
<i>Felis catus</i>	CACAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCCCAGTCGCGGGGACC	1241
<i>Manis temminckii</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	346
<i>Manis tricuspis</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	351
<i>Manis tetradactyla</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	347
<i>Manis gigantea</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	346
<i>Manis javanica</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	312
<i>Manis pentadactyla</i>	CCAAATCCAGGCATGGGCGGGGACTTTACTCTGGAATTTGACTCCTCCATCCC-----	319
	* *	
<i>Homo sapiens</i>	AAAGGCCCTGAGAACGAAAGAGGTGCACAAAGGTCCTCCCTATGCCGGAGGCCAACCCA	999
<i>Canis familiaris</i>	AAAGGCCCTGAGAACGAGAGAGAGGGGCACAAAGGTCCTCCCTATGCCCGATGTCAACCCA	921
<i>Felis catus</i>	AAAGGCCCTGAGAACGAGAGAGAGGTACACAAAGGTCCTCCCTATGCCCGATGTCAACCCG	1301
<i>Manis temminckii</i>	-----CAGAGAGGAGAAAGGAGGTACACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCG	399
<i>Manis tricuspis</i>	-----CAGAGAGGAGAAAGGAGGTACACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCCT	404
<i>Manis tetradactyla</i>	-----CAGAGAGGAGAAAGGAGGTACACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCCT	400
<i>Manis gigantea</i>	-----CAGAGAGGAGAAAGGAGGTACACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCG	399
<i>Manis javanica</i>	-----CAGAGAGGAGAGGAGGTGCACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCA	365
<i>Manis pentadactyla</i>	-----CAGAGAGGAGAGGAGGTGCACAAAGGTCCTCCCTATTTGCCAGAGGCCAACCCA	372

<i>Homo sapiens</i>	GACAAATCTAGAAAAACCCAGCTTTCTTTACAGAGCTAGAAACCTGCTCTCCACGCAGGGCTC	1059
<i>Canis familiaris</i>	GCCAAATCCAGAAAAACCCGGTCTCTCTTACCAGAGTAGACACCTGGTGCCTTAGGAGGGCTT	981
<i>Felis catus</i>	GCCGATCCGGAAAAACCCAGCTCTCTTCCAGAGGTAGACACCTGGTGCCTTGGAGGGCTT	1361
<i>Manis temminckii</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCAGAGGAGCAAAATATCTTAAGAGGGCTT	459
<i>Manis tricuspis</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCGAGGGAGCAACCAATATCTTAGGAGGGCTT	464
<i>Manis tetradactyla</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCGAGGGAGCAACCAATATCTTAGGAGGGCTT	460
<i>Manis gigantea</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCAGAGGAGCAACCAATATCTTAGGAAGGCTT	459
<i>Manis javanica</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCAGAGGAGCAACCAATTTCTTAGGAGGGCTT	425
<i>Manis pentadactyla</i>	GCCAGTCTGAAAAACCCAGCTCACCTTCCAGAGGAGCAACCAATTTCTTAGGAGGGCTT	432
	* * * * * ***** *	
<i>Homo sapiens</i>	CTTGCTCTCCCTAAAGGATGACATTTCCCGGCTTGCCAAGGAGGCCCTTCAGGGAAGATGAAG	1119
<i>Canis familiaris</i>	CTCGCTCATCCAAAAGGACAAATGATCCAGCCTGGCAAGAGGCCCTGCAGGCGAGAGCGGG	1041
<i>Felis catus</i>	CTTGCTCACCCAAAAGGGCAACGACCCAGCCTGGCAAGAGGTCCTGCAGGGCAGAGTGGG	1421
<i>Manis temminckii</i>	CTTGCTCTCCCTAAAGGGCAAGGTTCTTGCCCTGGCA-GAGGACCTGCAGCACAGAGTGGC	518
<i>Manis tricuspis</i>	CTTGCTCTCCCTAAAGGGCAAGGTTCTTGCCCTGGCA-GAGGACATGCAGCACAGAAATGGC	523
<i>Manis tetradactyla</i>	CTTGCTCTCCCTAAAGGACAAAGGTTCTTGCCCTGGCA-GAGGACCTGCAGCACAGAAATGGC	519
<i>Manis gigantea</i>	CTTGCTCTCCCTAAAGGGCAAGGTTCTTGCCCTGGCA-GAGGACCTGCAGCACAGAGTGGC	518
<i>Manis javanica</i>	CTTGTTCTCCCTAAAGGGCAAGGTTCTTGACCCAGCA-GAGGACCTGCAGCACAGAGTGGC	484
<i>Manis pentadactyla</i>	CTTGTTCTCCCTAAAGGGCAAGGTTCTTGACCCAGCA-GAGGACCTGCAGCACAGAGTTGC	491
	** * * * * ** *	
<i>Homo sapiens</i>	GGACTCCCGAGGTCACCCAGCAGCTGCTGACCCACTGATGACCCCTGAATTAGCTGAT	1179
<i>Canis familiaris</i>	GGACCCCGAGGGTCACCCCGGACGCTGACCCACTGATGACCCCTGAATTAGCTGAT	1101
<i>Felis catus</i>	GGACCCCGAGGGTCACCCAGCAGAGAGCTGACCCACTGATGACCCCTGGATTAGCAGAT	1481
<i>Manis temminckii</i>	AAACACCTCAGGGTCACTCCAGCCACTGTGTACCCCGAGATGACACCTGGATTGCAGAT	578
<i>Manis tricuspis</i>	AAACATCTCAGGGTCACTCCAGCCACTGTGTATCCCGAGATGACACCTGGATTACGGAT	583
<i>Manis tetradactyla</i>	AAACACCTCAGGGTCACTCCAGCCACTGTGTACCCCGAGATGACACCTGGATTACAGAT	579
<i>Manis gigantea</i>	AAACACCTCAGGGTCACTCCAGCCACTGTGTACCCCGAGATGACACCTGGATTGTGGAT	578
<i>Manis javanica</i>	AGACACCTCAGGGTCACTTCAGCCGCTGCGGCCCCCAGATGACACCTGGATTCACTGAT	544
<i>Manis pentadactyla</i>	AGACACCTCAGGGTCACTTCAGCCACTGCAGACCCCAATGACTCCTGGATTCACTGAT	551
	** * * * * ***** *	

<i>Homo sapiens</i>	GTTTATAGGACCTACGATGCTGACATGACACATCCGTGGATTTCAGGAAGAAAGCAACC	1239
<i>Canis familiaris</i>	ATTTATGAGACCTACGGTGTGATGTGACACACACCCCTGGA-----AGAAACGCCC	1152
<i>Felis catus</i>	ATTTATGAGACCTATGGTGCCGACGTGACACACCCCTGGA-----AGAAACGCCC	1532
<i>Manis temminckii</i>	GCTTCTGAGACCTTCAGTG-----ACCACCCCAAGGTTTCCAGGAAGAAACAACC	630
<i>Manis tricuspis</i>	GCTTCTGAGACCTTCAGTGCTGACATGACACACCCCAAGGTTTCCAGGAAGAAACAACC	643
<i>Manis tetradactyla</i>	GCTTCTGAAAACCTTCAGTGCTGACATGACACACCCCAAGGTTTCCAGGAAGAAACAACC	639
<i>Manis gigantea</i>	GCTTCTGAGACCTTCAGTG-----ACCACCCCAAGGTTTCCAGGAAGAAACAACC	630
<i>Manis javanica</i>	ACTTCTGAGACCTTCAGTGCTGACATGACACACCCCAAGGTTTCCAGGAAGAAACAACC	604
<i>Manis pentadactyla</i>	GCTTCTGAGACCTTCAGTGCTGACATGACACACCCCAAGGTTTCCAGGAAGAAACAACC	611
	*** ** ***** ** ** ** *	
<i>Homo sapiens</i>	A-----TGGATACCAAGATGGCCCAAACTCTCTGCAAAACATCCATCCAGGAAACAAGC	1295
<i>Canis familiaris</i>	A-----CGGATACCAAGATGATCCAGACACTCAGCAAAACATTTGATCCAGAAACAAGGC	1208
<i>Felis catus</i>	A-----CAGATACCAAGCATCCAGACACTCAGCAAAACATCCATCCAGAAACAAGGC	1588
<i>Manis temminckii</i>	TGGATCAGATACCACAAAATGATCCAGAAAATTGAGCAAAATACTGATCCAGGAAACAAGGT	690
<i>Manis tricuspis</i>	TGGAGCAGATACCACAAAATGATCCAGAAAATTGAGCAAAACACTGACGCCAGGAAACAAGGT	703
<i>Manis tetradactyla</i>	TGGATCAGATACCACAAAATGATCCAGAAAATTGAAACAAAACACTGATGCCAGGAAACAAGGT	699
<i>Manis gigantea</i>	TGGATCAGACACCACAAAATGATCCAGAAAATTGAGCAAAACACTGATGCCAGGAAACAAGGT	690
<i>Manis javanica</i>	TAGATCAGATACCACAAAATGATCCAGACATTTGAGTAAAACAATGATGCCAGCAAAACAAGGT	664
<i>Manis pentadactyla</i>	TAGATCAGATACCACAAAATGATCCAGACATTTGTAAAACACTGATGCCAGCAAAACAAGGT	671
	** ** * * * * * * * * * * * * * * * * *	
<i>Homo sapiens</i>	CCAGAGCCCGAGATGATGCATGCATGCATGGCATTTCCAAGAGCCC-----	1341
<i>Canis familiaris</i>	CCAGAGCCCGAGATTATGCATGCATGCATGGCATTTCCAAGAGCCC-----	1254
<i>Felis catus</i>	CCAGAGCCCGAGATGCATGCATGCATGGCATTTCCAAGAGCCCCTGA-----	1637
<i>Manis temminckii</i>	CCAGCATCCCAGATAAATGCATGCATGCATGGCATTTCCAAGAGCCTTGAGCTACCGGCTG	750
<i>Manis tricuspis</i>	CCAGCATCCCAGATAAATGCATGCATGCATGGCATTTCCAAGAGCCTTGAGCTTCTGGCTG	763
<i>Manis tetradactyla</i>	CCAGCATCCCAGATAAATGCATGCATGCATGGCATTTCCAAGAGCCTTGAGCTTCTGGCTG	759
<i>Manis gigantea</i>	CCAGCATCCCAGATAAATGCATGCATGCATGGTATTTCCAAGAGTCTTGAGCTACCGGCTG	750
<i>Manis javanica</i>	CCAGCAACCCAGATAAATGCATGCATGCATGGTATTTCCAAGAGCCTTGAGCTACCGGCTG	724
<i>Manis pentadactyla</i>	CCAGCATCCCAGATAAATGCATGCATGCATGGCATTTCCAAGAGCCTTGAGCTACCGGCTG	731
		

Homo sapiens ----- 1341
Canis familiaris ----- 1254
Felis catus ----- 1637
Manis temminckii ----- 755
Manis tricuspis ----- 823
Manis tetradactyla ----- 819
Manis gigantea ----- 810
Manis javanica ----- 784
Manis pentadactyla ----- 791
CTTTC
CTTTCGTGTGCACAAAGCTCCCCAGTTTTGTCCCCATAGTATACCTTTTTTGTAAAAACAC
CTTTCGTGTGCACAAAGCTCCCCAGTTTTGTCCCCATAGTATACCTTTTTTGTAAAAACAC
CTTTCGTGTGCACAAAGCTCCCCAGTTTTGTCCCCATAGTATACCTTTTTTGTAAAAACAC
CTTTCGTGTGCACAAAGCTCCCCAAATTTTGTCCCCATAGCATACCTTTTTTGTGCTGAAAAACAC
CTTTCGTGTGCACAAAGCTCCCCAGTTTTTGTCCCCATAGCATACCTTTTTTGTGCTAAAAACAC

Homo sapiens ----- 1341
Canis familiaris ----- 1254
Felis catus ----- 1637
Manis temminckii ----- 755
Manis tricuspis ----- 883
Manis tetradactyla ----- 869
Manis gigantea ----- 870
Manis javanica ----- 842
Manis pentadactyla ----- 851
TTATCACCCCTTCTGCAGTGAAGGCATTTAAAAAGCAGCTAAACCTATATTAATAAATACATGT
TTATCACCCCTTCTGCAGTGAAGGCATTTAAAAAGCAGCTAAACCTATATTAAT-----
TTATCACCCCTTCTGCAGTGAAGGCATTTAAAAAGCAGCTAAACCTATATTAATAAATACACGT
TTATTACCTTTCTGCAGCAAAAGGTGTTAAAAAGCAGCTAAACCTATATTAATAAATACAAA--
TTATTACCTTTCTGCAGCAAAAGGTGTTAAAAAGCAGCTAAACCTATATTAATAAATACAAAGT

Appendix F: *BFSP2* alignment of pangolin species with mammals.

<i>Homo sapiens</i>	CCAGCCTCCAGGACCAATGCCATGAGTGGCCCTTGTCCGAGCACCCGGGGTCTATGTAGGA	177
<i>Canis familiaris</i>	CCAGCCTCCAGGACCAATGCCCTTGGGTGGTCTCATCCGAGCACCCAGAGTCTATGTAGGG	177
<i>Felis catus</i>	CCAGCCTCCAGGACCAATGCCATGGGTGGTCTCGTCGAGGAGCCAGAGTCTATGTAGGG	180
<i>Manis temminckii</i>	-----GAGTGCCCAAGAGTCTATGTAGGG	23
<i>Manis tricuspis</i>	-----GAGTGCCCAAGAGTCTATGTAGGG	23
<i>Manis tetradactyla</i>	-----GAGTGCCCAAGAGTCTATGTAGGG	23
<i>Manis javanica</i>	TCGGCCTCCAGGACCAAGTGTGTGGGTGGTCTCGAGTCCGAGAGTCTATGTAGGG	177
<i>Manis pentadactyla</i>	TCGGCCTCCAGGACCAAGTGTGTGGGTGGTCTCGAGTCCGAGAGTCTATGTAGGG	177
	*** * *****	
<i>Homo sapiens</i>	ACAGCACCAAGTGGGTGCATAGGTGGCTTGGGTGCCGTGTGACCCGCCGGGCCCTCGGC	237
<i>Canis familiaris</i>	ATGGCGCCAGTGGGTTCACAGGTGGCTGGGCACCCGTGTGACACGCCGGGCTCTGGGC	237
<i>Felis catus</i>	ATGGTGCCAGTGGGTACACAGGTGGCTGGGCACCCGTGTGACACGCCGGGCCCTGGGC	240
<i>Manis temminckii</i>	ATGGTGCCAGCGGGTGTGAGCGGCTTGGATGCCGTGTGACCCGCCGGGCCCTAGGC	83
<i>Manis tricuspis</i>	ATGGTGCCAGCGGGTGTGAGGTGGCTTGGATGCCGTGTGACCCGCCGGGCCCTAGGC	83
<i>Manis tetradactyla</i>	ATGGTGCCAGCGGGTGTGAGGTGGCTTGGATGCCGTGTGACCCGCCGGGCCCTAGGC	83
<i>Manis javanica</i>	ATGGTGCCAGTGGGTGTGAGGTGGCTTGGATGCCATGTGACCTGTGGGCCCTAGGC	237
<i>Manis pentadactyla</i>	ATGGTGCCAGTGGGTGTGAGGTGGCTTGGATGCCCTGTGACCCGCCGGGTCCTAGGC	237
	* * ***** *	
<i>Homo sapiens</i>	ATCAGCAGTGTCTTCTTCAGGGCCCTGCGGAGCTCAGGCCTGGCCACCGTGCCGGCTCCA	297
<i>Canis familiaris</i>	ATCAGCAGTGTCTTCTTCAGGGCCCTGCGGAGCTCAGGTCTTGCCACAGCACAGCTCCA	297
<i>Felis catus</i>	ATCAGCAGTGTCTTCTTCAGGGCCCTGCGGAGCTCAGGTCTTGCCACAGCACAGCTCCA	300
<i>Manis temminckii</i>	ATCAGTCTGTCTTCTTCAGGGCCCTGAGGAGCTCAGGCCTAGCCACAGCCCCAGCTCCA	143
<i>Manis tricuspis</i>	ATCAGCACGGTCTTCTTCAGGGCCCTGAGGAGCTCAGGTCTTGCCCGCAGTACCAGCTCCA	143
<i>Manis tetradactyla</i>	ATCAGCACGGTCTTCTTCAGGGCCCTGAGGAGCTCAGGCCTGGCCGAGTACCAGCTCCA	143
<i>Manis javanica</i>	ATCAGCACTGTCTTCTTCAGGGCCCTGAGGAGCTCAGGCCTAGCCACAGTACCAGCTCCA	297
<i>Manis pentadactyla</i>	ATCAGCACTGTCTTCTTCAGGGCCCTGAGGAGCTCAGGCCTAGCCGAGTGCAGGCTCCA	297
	***** *	

Homo sapiens 357 GGTTTGGAGAGGACCATGGTGTGTTGAGGACCTAGGGGGCTGCCTGGTGGAAATATATG
Canis familiaris 357 GGTCTGGAGAGAACTCAGTACAGTTGAGGACCTGGGGGGTTCCCTGGTGGAAATACATG
Felis catus 360 GGCCCTGAGAGAACTCGGTGCCGTTGAGGACCTGGGGGGTTGCCTGGTGGAAATACATG
Manis temminckii 203 GGCCCTGAGAGGACCTCAGTGCCGCTGAGGACCTGGGGGGTTGCCTGGTGGAAATACATG
Manis tricuspis 203 GGCCCTGGCAGGACCTCAGTGCCGCCGAGGACCTGGGGGGTTGCCTGGTGGAAATACATG
Manis tetradactyla 203 GGCCCTGAGAGGGGGCCGAGTGCCGCCGAGGACCTGGGGGGTTGCCTGGTGGAGTTACG
Manis javanica 357 GGCCCTGAGAGGGGGCCCTCAGTGCCGCTGAGGACCTGGGGGGTTGCCTGGTGGAAATACATG
Manis pentadactyla 357 GGCCCTGAGAGGGGGCCCTCAGTGCCGCCGAGGACCTGGGGGGTTGCCTGGTGGAAATACATG
 ** *** ***** ** * * * * * ***** * * * * *

Homo sapiens 417 GCCAAAAGTGCACGCCCTTGAGCAAGTCAGTCAGGAGCTGGAACACAACTGCGGATGCAC
Canis familiaris 417 GCCAAGGTTTCATGCCCTCGAGAGCGTCAGCCAAGAGCTGGAGGCACAGCTGCGGATGCAC
Felis catus 420 GCCAAGGTTTCATGCCCTCGAGAACGTCAGCCAGGAGCTGGAGGCACAAATTGCCGTCAC
Manis temminckii 263 GCCAAGGTGCACGCCCTGGACAAAAGTCAGCCAGGAGCTGGAGGCACAGCTGCCGTCAC
Manis tricuspis 263 GCCAAGGTGCACGCCCTGGACAAAAGTCAGCCAGGAGCTGGAGGCACAGCTGCCACGCAC
Manis tetradactyla 263 GCCAAGGTGCACGCCCTGGACAAAAGTCAGCCAGGAGCTGGAGGCACAGCTGCCACGCAC
Manis javanica 417 ACCAAGGTGCATGCCCTGGACAAAAGTCAGCCAGGAGCTGGAGGCACAGCTGCCAGGCAC
Manis pentadactyla 417 GCCAAGGTGCACGCCCTGGATAGAGTCAGCCAGGAGCTGGAGGCACAGCTGCCAGGCAC
 ***** ** * * * * * ***** * * * * *

Homo sapiens 477 CTGGAGAGCAAAAGCACACGCTCGGAAAACCTGGGGTGCCCTACGGGCTTCCTGGGCCAGC
Canis familiaris 477 CTGGAGAGCAAAAGCACACGCTCAGAAAAGCTGGGGTGCCCTCAGGGCCTCCTGGGCCAGC
Felis catus 480 CTGGAGAGCAAGGCACACGCTCCGAAAAGTTGGGGTGCCCTCAGGGCTTCCTGGGCCAGC
Manis temminckii 323 CTGGAGAGCAAGGCCGCGCTCTGAGGGCTGGGGCGCCCTCAGGGCCTCCTGGGCCAGC
Manis tricuspis 323 CTGGAGAGCAAGGCCGCGCTCTGAGGGCTGGGGCGCCCTCAGGGCCTCCTGGGCCAGC
Manis tetradactyla 323 CTGGAGAGCAAGGCCGCGCTCTGAGGGCTGGGGCGCCCTCAGGGCCTCCTGGGCCAGC
Manis javanica 477 TCGGAGAGCAAGGCCGCGAGCTCTGAGGGCTAGGGCACCCCTCAGGGCCTCCTGGGCCAGC
Manis pentadactyla 477 CTGGAGAGCAAGGCCGCGAGCTCTGAGGGCTGGGGCGCCCTCAGGGCCTCCTGGGCCAGC
 ***** *** * * * * * ***** * * * * *

<i>Homo sapiens</i>	AGCTGCCAGCAGGTGGGTGAGGAGTCTTTGGAAAAATGCCCCGGCTCATGCTGCAGACAGAA	537
<i>Canis familiaris</i>	AGTTGCCAGCAGGTGGCGAGGAGTCTTTGGAAAAATGCTCGGCTCATGCTTCAGACAGAG	537
<i>Felis catus</i>	AGCTGCCAACAGGTGGCGAGGCGGTCTTTGGAAAAATGCCCCGGCTCATGCTTCAGACAGAG	540
<i>Manis temminckii</i>	AGCTGCCAGCAGGTGAGTGGCTGGGTACGCCCCAGGGGCTTTGGATGGAGCTGGGAGAGG	383
<i>Manis tricuspis</i>	AGCTGCCAGCAGGTGAGTGGCTGGGTACACCCAGGGGCTTTGGATGGAGCTGGGAGAGG	383
<i>Manis tetradactyla</i>	AGCTGCCAGCAGGTGAGTGGCTGGGTACGCCCCAGGGGCTTTGGATGGAGCTGGGAGAGG	383
<i>Manis javanica</i>	AGCTGCCAGCAGGTGAGTGGCTGAGTCATGCCCAAGGGGCTTTGGGTGGGGCTGGGAGAGG	537
<i>Manis pentadactyla</i>	AGCTGCCAGCAGGTGAGTACGGGGGTTCATGCCCAAGGGGCTTTGGGTGGGGCTGGGAGAGG	537
	** *	
<i>Homo sapiens</i>	ACTATCCAGGCCGCGGAGCAGATGACTTTTAAAGAGAGATATGAAAAATGACAGCCATTTCGA	597
<i>Canis familiaris</i>	AAATATCCAGCGGGGTGCTGACGACTTTTAAAGAGAGATATGAAAAATGACAGCCATTTCGA	597
<i>Felis catus</i>	AAATATCCAGCGGGGTGCGAGATGACTTTTAAAGAGAGATATGAAAAATGACAGCCATTTCGA	600
<i>Manis temminckii</i>	ATGCTGACAGCGCGCCTGTTGTACCATCATCTCTGGGAAAGGACGATGAGCAGGAAA-----	438
<i>Manis tricuspis</i>	ATGCGGACAGCTGCCTGTTGTATCATCATCTCTGGGAAAGGACGATGAGCAGGAAA-----	438
<i>Manis tetradactyla</i>	ATGCGGACAGCTGCCCTTGTATCATCATCTCTGGGAAAGGACGATGAGCAGGAAA-----	438
<i>Manis javanica</i>	GTGCTGACAGCTGCCTGTTGTACCATCATCTCTGGGAAAGGACGATGAGCAGGAAATGCTGA	597
<i>Manis pentadactyla</i>	GTGCTGACAGCTGCCTGTTGTACCATCATCTCTGGGAAAGGACGATGAGCAGGAAATGCTGA	597
	** *	
<i>Homo sapiens</i>	AAGGCAGCAGAAAGAGGAAATTAACCTCTCTGTATAAAAGTCATTGATGAGGCTAATTTGACT	657
<i>Canis familiaris</i>	AAGGCAGCGGAAGAGGAAATTAACCTCTTTGTATAAAAGTCATTGATGAAGCTAATTTGACA	657
<i>Felis catus</i>	AAGGCAGCAGAAAGAGGAAATTAACCTCTCTGTATAAAAGTCATCGATGAGGCTAATTTGACA	660
<i>Manis temminckii</i>	-----	438
<i>Manis tricuspis</i>	-----	438
<i>Manis tetradactyla</i>	-----	438
<i>Manis javanica</i>	GGCAGGAGAGTGAGTGGTGTGAGGCGCAGGAGACTGAGCATGGCCCTTACCCCTTTCTGCTC	657
<i>Manis pentadactyla</i>	GGGAGGGGAGTGAGTGGTGTGAG-----	620

Appendix G: GUCA1C alignment of pangolin species with mammals.

<i>Homo sapiens</i>	-----ATGGGGAATGGCAAACTATAGCTGGTGATCAGAAAAGCAGATTCCCTACACAAGAG	54
<i>Canis familiaris</i>	-----ATGGGAAATAACGAATCCAGAGCTGGTGGTCAGGCAGCAGTTTCTGTAAAGAG	54
<i>Felis catus</i>	-----ATGGGAAATAATGAATCCATAGCTGGTGGTCAGAAAAGCAGTTTCTGTAAAGAG	54
<i>Manis temminckii</i>	GTGAATCACAATGGGAAATAGTGATATAGCTGGTGGTTGGACAGCAATTCCTGTAAAGGAG	180
<i>Manis tricuspis</i>	GTGAATCACAATGGGAAATAGTGATATAGCTGGTGGTTGGCGGCAATTCCTGTAAAGGAG	179
<i>Manis tetradactyla</i>	GTGAATCACAATGGGAAATAGTGATATAGCTGGTGGTTGGACGCAATTCCTGTAAAGGAG	179
<i>Manis gigantea</i>	GTGAATCACAATGGGAAATAGTGATATAGCTGGTGGTTGGCGGCAATTCCTGTAAAGGAG	143
<i>Manis javanica</i>	-----AATGGGAAATAGTGATATAGCTGGTGGTTGGACAGAAAATTCCTGTAGGAG	51
<i>Manis pentadactyla</i>	-----AATGGGAAATAGTGATATAGCTGGTGGTTGGACAGAAAATTCCTGTAGGAG	51
	* * * * *	
<i>Homo sapiens</i>	ACCATGTGTGTACAGAAACATTTATGATGGAATATCCATCCGGCCTGCAAAACACTACAT	114
<i>Canis familiaris</i>	ACCTATGTGTGTACAGAAAATTTATGAGGGAATATCCATCTGGCCTGCAAAACACTTCAC	114
<i>Felis catus</i>	ACCTATGTGTGTACAGAAAATTTATGAGGGAATATCCATCCGGCCTGCAAAACACTGCAT	114
<i>Manis temminckii</i>	ACTTATGTGTGTACAGAAAATTTTGTGATGGACTGTACATCTGACCTTCAAAACACTGCAT	240
<i>Manis tricuspis</i>	ACTTATGTGTGTACAGAAAATTTTGTGATGGAAATATACATCTGACTTTCAAAACACTGCAT	239
<i>Manis tetradactyla</i>	ACTTATGTGTGTACAGAAAATTTTGTGATGGAAATATACATCTGACTTTCAAAACACTGCAT	239
<i>Manis gigantea</i>	ACTTATGTGTGTACAGAAAATTTTGTGATGGAAATATACATCTGACTTTCAAAACACTGCAT	203
<i>Manis javanica</i>	ACCTATGTGTGTATAGAAAATTTTGTGATGGAAATTTACATCTGGCCTTCAAAACACTGCAT	111
<i>Manis pentadactyla</i>	ACCTATGTGTGTATAGAAAATTTTGTGATGGAAATTTACATCTGGCCTTCAAAACACTGCAT	111
	** * * * * * * * * * *	
<i>Homo sapiens</i>	GAATTTAAGACACTTTTGGGTCTGCAAGGCTGAAATCAGAAAGGCCAATAAACA-TATTGA	173
<i>Canis familiaris</i>	GAATTTAAGACACTCTTTGGGTCTGCAAGGCTGAAACCCAAAAGCCAAATCAACA-CGTTGA	173
<i>Felis catus</i>	GAATTTAAGACACTCTTTGGGTCTGCAAGGCTGAAATCAAAAAGCCAAATCAACA-TGTTGA	173
<i>Manis temminckii</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAATGTTGA	300
<i>Manis tricuspis</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAAT-ATGTTGA	298
<i>Manis tetradactyla</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAAT-ATGTTGA	298
<i>Manis gigantea</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAAT-ATGTTGA	262
<i>Manis javanica</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAAT-TGTTAA	170
<i>Manis pentadactyla</i>	GAATTTAAGACACTCTTTGGGTCTGTAAGGCTGAAATTAATAAGGTCAAATTAAT-TGTTGA	170
	** * * * * * * * * * *	

Homo sapiens TCAAGTTTATAATACCTTTGACACGAAACAAGGATGGATTTGTTGACTTTT-----TTGGAGT 229
Canis familiaris TCAAGTCTATAATACCTTTGACATGAACAAGGATGGATTTATTGACTTTT-----TTAGAGT 229
Felis catus TCAAGTCTATAATACCTTTGACATGAACAAGGATGGATTTATTGACTTTT-----TTAGAGT 229
Manis temminckii TCAAAACTGTAATAACTTTGACATGAACAAGGATGGATTTTACTTTCCATTTGGCTAT 360
Manis tricuspis TCAAAATTGTAATAACTTTGACATGAACAAGGATGGATTTTACTTTCCATTTGGCTAT 358
Manis tetradactyla TCAAAACTGTAATAACTTTGACATGAACAAGGATGGATTTTACTTTCCATTTGGCTAT 358
Manis gigantea TCAAAATTGTAATAACTTTGACATGAACAAGGATGGATTTTACTTTCCATTTGGCTAT 322
Manis javanica TCAAAATCTATAATAACTTTGACATGAACAAGGATTTATTGACTTTGTAGAGTTTATTGC 230
Manis pentadactyla TCAAAATCTATAATAACTTTGACATGAACAAGGATTTGCAGGCTGTACAAACC-----CTCAATG 226
* * * * *

Homo sapiens TTATTGCTGCTGTAATCTAATCATGCA---AGAAAAAATGGAGCAAAAAATTAAAAATGGT 286
Canis familiaris TTATTGCTGCTGTAATCTGTTTATACG---AGGAAAAATGGATCAAAAAATTAAAAATGGT 286
Felis catus TTATTGCTGCTGTAATCTGTTTATACG---AGGAAAAATGGAGCAGAAAAATTAAAAATGGT 286
Manis temminckii GGTTCTGTAATTTCTCCTTTTAAAAAATATGAAATAGCCTGAAAGTAA-AATTTGTATGTC 419
Manis tricuspis GGTTCTGTAATTTCTCCTTTTAAAAAATATGAAATAGCCTGAAAGTAC-AATTTGTATGTC 417
Manis tetradactyla GGTTCTGTAATTTCTCCTTTTAAAAAATATGAAATAGCCTGAAAGTAC-AATTTGTATGTC 417
Manis gigantea GGTTCTGTAATTTCTCCTTTTAAAAAATATGAAATAGCCTGAAAGTAC-AATTTGTATGTC 381
Manis javanica TGCGGTAAATCATAAATCTACTTGTACA---AGCAAAAAATGGAACAAAAATTTTAAATGGT 287
Manis pentadactyla GCCAGCAAAACCTGTAGTCTCTAAAGAAATTT---CGCCAACTTGGT-----GTTTCCCTAAGAT 278
* * * * *

Homo sapiens ATTTTAAAGCTGTATGATGCTGATGGAAAATGGTTCTATTGACAAAAAATGAACTACTGGACA 346
Canis familiaris ATTTTAAAGCTGTATGATGCTGATGGAAAATGGTTCTATTGACAAAAAAGGAACTACTGAAACA 346
Felis catus ATTTTAAAGCTGTATGATGCTGATGGAAAATGGTTCTATTGACAAAAAAGGAACTACTGAAATA 346
Manis temminckii CCGCTGAACAGGAAGGCTTCAGAGAACCCGAGCTTGTAGGTAAGTGAGTCATTAAAC- 478
Manis tricuspis CCACTGAACAGGAAGGCTTCAGAGAACCCGAGCTTGTAGGTAAGTGAGTCATTAAAC- 476
Manis tetradactyla CCACTGAGCAGGAAGGCTTCAGAGAACCCGAGCTTGTAGGTAAGTGAGTCATTAAAC- 476
Manis gigantea CCACTGAACAGGAA----- 395
Manis javanica ATTTTAAAGCTGTATGATGCTGATGGAAAATGGTTCTATTGACAAAAAAGGAACTACTGAGC 347
Manis pentadactyla CGAT-ACAAAACAATGATGGTAAG----- 300
* * * * *

<i>Homo sapiens</i>	T-GTTCATGGGGTACAAAGCCCTCAATGGCCAGCAAACTCTGAGTCCTGAAAGAAATTCATC	405
<i>Canis familiaris</i>	T-CTTCATGGCTGTACAAAGCCCTCAATGGCCAGCAAACTCTGAGTCCTGAAAGAAATTTACT	405
<i>Felis catus</i>	T-CTTCCCTGGCTGTGCAAGCCCTCAATGGCCAAACAAACCCTGAGTCCTGAAAGAAATTTACC	405
<i>Manis temminckii</i>	-----	478
<i>Manis tricuspis</i>	-----	476
<i>Manis tetradactyla</i>	-----	476
<i>Manis gigantea</i>	-----	395
<i>Manis javanica</i>	ATATTCTTGGTAAATGAAGCAGGCAAGTAGCCCTTTGAAG--GACATTGTGTGTTAAATT	404
<i>Manis pentadactyla</i>	-----	300

<i>Homo sapiens</i>	AACCTGGTGTCCATAAAGATCGATATATAAAACAATGATGGTAATGGAGCTACTCTCTACTGG	465
<i>Canis familiaris</i>	AACCTGGTGTCTTAATAAAGATCGATATATAAAACAATGATGG--	443
<i>Felis catus</i>	AACCTGGTGTCTTCAATAAAGATCGATNAAAAACAATGATGG--	443
<i>Manis temminckii</i>	-----	478
<i>Manis tricuspis</i>	-----	476
<i>Manis tetradactyla</i>	-----	476
<i>Manis gigantea</i>	-----	395
<i>Manis javanica</i>	CA-----GTTTTCTCTCTATGGAC--	424
<i>Manis pentadactyla</i>	-----	300

<i>Homo sapiens</i>	ATCACTTTTCCCTACAAAGGGAATTGACTTTAGAAAGAAATTTATCAATGGCATGGCAAAAAGA	525
<i>Canis familiaris</i>	-----AGAAATTGACTTTAGAAAGAAATTTATCAATGGCACGGGAAAAAGA	485
<i>Felis catus</i>	-----AGAACTGACTTTAGAAAGAAATTTATCAATGGCACGGGAAAGA	485
<i>Manis temminckii</i>	-----	478
<i>Manis tricuspis</i>	-----	476
<i>Manis tetradactyla</i>	-----	476
<i>Manis gigantea</i>	-----	395
<i>Manis javanica</i>	-----	424
<i>Manis pentadactyla</i>	-----	300

<i>Homo sapiens</i>	TCAGGATCTCCTGGAGATTGTTTACAAAGAGCTTCGACTTCTCCAAATGTGTGCTGAGAGTAAT	585
<i>Canis familiaris</i>	TCAGGATCTTCTGGACATTGTTTCCAAAGAGCTTTGACTTTTCCAAATGTACTAAAAAGTAAT	545
<i>Felis catus</i>	TCAGGATCTTCTGGCGATTGTTTCCAAAGAGCTTTGATTTTCCAGTGTACAGAAAAATAAT	545
<i>Manis temminckii</i>	-----	478
<i>Manis tricuspis</i>	-----	476
<i>Manis tetradactyla</i>	-----	476
<i>Manis gigantea</i>	-----	395
<i>Manis javanica</i>	-----	424
<i>Manis pentadactyla</i>	-----	300

<i>Homo sapiens</i>	CTGTAATGGGAAGCAGCCAGACATGGAGACAGACT---CCTC AAAATCTCCTGACAAGGC	642
<i>Canis familiaris</i>	CTACAGTGGGAAAACAGCCAGACACAGAAANGGAGGCTTCCCTTCAAAACCATGTGAGAAGAC	605
<i>Felis catus</i>	CTATAATGGGAATCAGCCAGACACAGAAAGGAGGTTTTTC-TTCAAAACCTTGTGACAAGGT	604
<i>Manis temminckii</i>	-----	478
<i>Manis tricuspis</i>	-----	476
<i>Manis tetradactyla</i>	-----	476
<i>Manis gigantea</i>	-----	395
<i>Manis javanica</i>	-----	424
<i>Manis pentadactyla</i>	-----	300

Appendix H: IFNE alignment with 72 mammals and African pangolin species.

<i>Homo sapiens</i>	ATGATTATCAAGCACTTCTTTGGA---ACTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Pan paniscus</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Pan troglodytes</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Gorilla gorilla</i>	ATGATTATCAAGCACTTCTTTGGA---ATTTTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Nomascus leucogenys</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Pongo abelii</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Rhinopithecus roxellana</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Macaca mulatta</i>	ATGATTATCAAGCACTTCTTTGGA---ATTATGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Mandrillus leucophaeus</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Cercocebus atys</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Macaca nemestrina</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Papio anubis</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Macaca fascicularis</i>	ATGATTATCAAGCACTTCTTTGGA---ATTATGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Colobus angolensis</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Chlorocebus sabaues</i>	ATGATTATCAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTACCACTGTC	57
<i>Aotus nancymae</i>	ATGATTATCAAGCACTTCTTTGGA---ATTTTGTGTTGGTGTGCTGGCCTCTACCACTATC	57
<i>Callithrix jacchus</i>	ATGATCATCAAGCACTTCTTTGGA---ATTTTGTGTTGGTGTGCTGGCCTCTATCACTATC	57
<i>Pteropus vampyrus</i>	ATGATTAGCAAGTACTTCTTTGGA---GTTGTGTTGGTGTGCTGGCCTCTTCCACTGTC	57
<i>Pteropus alecto</i>	ATGATTAGCAAGTACTTCTTTGGA---GTTGTGTTGGTGTGCTGGCCTCTTCCACTGTC	57
<i>Galeopterus variegatus</i>	ATGATTAAACAAGCACTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTCTTCTACTATC	57
<i>Propithecus coquereli</i>	ATGATTAAACAAGCACTTCTTTGGA---ATTGTGTTGATGGACTGGCCTCTTCTGCAGTC	57
<i>Ceratotherium simum</i>	ATGATAAAAGAAGCAATCTTTGGA---ATTGTGTTGGTACTTCTGGCCTCTTCTCTATT	57
<i>Equus caballus</i>	ATGATAAAACAAGCAATCTTTGAG---ATTATGTTGGTACTGCTTGGCCTCTTCCACTATT	57
<i>Camelus ferus</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTTTCCACCATC	57
<i>Vicugna pacos</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACCGTC	57
<i>Balaenoptera acutorostrata</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCCTGTGC	57
<i>Physeter catodon</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCCTGTGC	57
<i>Lipotes vexillifer</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCCTGTGC	57
<i>Orcinus orca</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCCTGTGC	57
<i>Sus scrofa</i>	ATGATCAACAAGTCTTCTTTGGA---ATCATGTTGGTACTGCTGGCCTTCTTCCACTGGC	57
<i>Ovis canadensis</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Ovis aries musimon</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Pantholops hodgsonii</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Bos taurus</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Bison bison bison</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Bubalus bubalis</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTCTTCCACTGTT	57
<i>Bos mutus</i>	ATGATTAAACAAGCCTTCTTTGGA---ATTGTGTTGGTGTGCTGGCCTTATTCACACTGTT	57

<i>Capra hircus</i>	ATGATTAAACAAGGCTTTCTTTGAA---ATTGTGTTGGTTCTGTGGCTTCTTCCACTGTT	57
<i>Otolemur garnettii</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTTGGTGGGCTGGCCCTCTTCCACTATC	57
<i>Tupaia chinensis</i>	ATGGTTAAACAGGCACTTCTTTGAA---GCTGTGTTGATGCTGCTGGCCCTCTTCCACTATC	57
<i>Eptesicus fuscus</i>	ATGATTAGCAAGCACTTCTGCGAA---AGTGTGTTGGTGTGCTGAGACCTCTTCCACTATC	57
<i>Myotis brandtii</i>	ATGATTAAACAAGCACTTCCGTGAG---AGTGTGTTGGTGTGCTGAGACCTCTTCCACTATC	57
<i>Myotis lucifugus</i>	ATGATTAAACAAGGCTTCCGTGAG---AGTGTGTTGGTGTGCTGAGACCTCTTCCACTATC	57
<i>Myotis davidii</i>	ATGATTCAACAAGCACTTCCGTGAA---AGTGTGTTGGTGTGCTGATCTCTCCCACTATC	57
<i>Leptonychotes weddellii</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTTGGTGTGCTGGCCCTCTTCTACTCTC	57
<i>Odobenus rosmarus</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTTAGTGTGCTGGCCCTCTTCTACTCTT	57
<i>Ursus maritimus</i>	ATGATTCCCAAGCACTTCTTTGAA---ATCGTGTGGTGTGCTGGCCCTCTTCCACTCTC	57
<i>Ailuropoda melanoleu</i>	ATGATTCAACAAGCACTTCTTTGAA---ATCGTGTGGTGTGCTGGCCCTCTTCCACTCTC	57
<i>Felis catus</i>	ATGATTAAACAAGCACTTCTTTGCA---ATTGTGTCGGTGTGCTGGCCCTCTTCTACTATC	57
<i>Panthera tigris altaica</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTCGGTGTGCTGGCCCTCTTCTACTATC	57
<i>Mustela putorius</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTTGGTGTGCTTTGGCCCTCTTCTACTCTC	57
<i>Canis lupus familiaris</i>	ATGATTAAACAAGCACTTCTTTGAA---ATTGTGTTGGTGTGCTGGTGTGCTTTCTACTATC	57
<i>Condylura cristata</i>	ATGATTAGCAAGTATTTTGTGAA---ATTGTGTTGCTGTGCTAGCCCTCTTTTACTTTT	57
<i>Sorex araneus</i>	ATGATTAGCAAGCACTTCTTTGAA---GTTTTGCTGGTGTACTGACCTCTTCTACTGTC	57
<i>Trichechus manatus latirostris</i>	ATGATTAAACAAGCACTTCTCTGAG---ATTGTGTTGGTGTGCTGGCCCTCGTCCACGACC	57
<i>Loxodonta africana</i>	ATGATTAAACAAGTATTTCTTTGAG---ACTGTGGTGGTCTGCTGTCTCTCGTCCATGATC	57
<i>Orycteropus afer</i>	ATGATTAAACAAGCACTTATTTGAG---ATTATGTTGCTGTGTTTCTGGCTTCTTCCATGATC	57
<i>Chrysochloris asiatica</i>	ATGAATAATAAACAACATTTATTTGAG---ATTGTGTTGCTGTGCTGGCCCTCTTCCATGATC	57
<i>Tarsius syrichta</i>	ATGATTATCAAAAACACTTTTGTGAA---ATTGATTTTGTGTTGCTGTCTCTTCCACTATC	57
<i>Erinaceus europaeus</i>	-----ATGATTAGCAAGCAATATC---TTTTGTGTTGATACTACTGGCCCTCTTCCCTATC	51
<i>Dasyopus novemcinctus</i>	ATGATTAAATAAGCACTTCTTTGAA---ATTGTGTTGGTGTGCTGTGGCCCTCTTCCACCCCTC	57
<i>Oryctolagus cuniculus</i>	ATGACTTACAAGTACTTCTTTGAA---ATTGTGTTGGTGTGCTGTGGCCCTCATCTACTATC	57
<i>Ochotona princeps</i>	ATGCTTTTGCAAAATACTTCTTTTAAA---ATTGTGTTGGTGTGCTGTGGTCTCATCCACAATC	57
<i>Elephantulus edwardii</i>	ATGATTGACAAGCACTTCTTTGTG---ATTGCGTTGCTGTCTCTGGCCCTCTCCCGTGATC	57
<i>Echinops telfairi</i>	ATGATTAGCAAGCACTTCTTTGOG---TTTTGTGTTTCTGTGCTGGCTTCCCTTCCACGATC	57
<i>Dipodomys ordii</i>	ATGTTTCACAAGCACTTCTTTGAG---ATTGGGTTGGTGTGCTTCTGGCCCTCTT---CTGTC	54
<i>Ictidomys tridecemlineatus</i>	ATGATTCAACAAGCACTTCCCTGGC---CTTGTGTTAGTGGTAC-----TGGTC	45
<i>Nannospalax galili</i>	ATGATTGCAAAACCGCTCCTTGAAA---TTGTGTTGGTGTGCTGTGGCTTCTTCCACTGTC	57
<i>Cricetulus griseus</i>	ATGATTCAACAAGCACTCCTGAAATGG---TGTTGCTGTGTTGGCTTCTTCCACTGTC	57
<i>Mesocricetus auratus</i>	ATGATTCAACAAGCACTCCTGAAATGGTGTGCTGTGCTGTGGCTGTGTTCCACTGTC	60
<i>Mus musculus</i>	-----GTGTTGCTGTCTCTTGGTTTCTTCCACTATC	30
<i>Manis temminckii</i> (5)	-----	0
<i>Manis tricuspis</i> (4)	-----	0
<i>Manis tetradactyla</i> (2)	-----	0
<i>Manis gigantea</i> (1)	-----	0
<i>Manis javanica</i> (8)	ATGATTAAACAAGCCTTTCTTTGAA---ATTGTCTTGGTGTGCTGGCCCTTACCAGTGC	57
<i>Manis pentadactyla</i> (1)	ATGATTAAACAAGCCTTTCTTTGAA---ATTGTCTTGGTGTGCTGGCCCTCTTCCACTGTC	57

<i>Homo sapiens</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Pan paniscus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Pan troglodytes</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Gorilla gorilla</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Nomascus leucogenys</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Pongo abelii</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Rhinopithecus roxellana</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Macaca mulatta</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Mandrillus leucophaeus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Cercocebus atys</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Macaca nemestrina</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Papio anubis</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Macaca fascicularis</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Colobus angolensis</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Chlorocebus sabaeus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Aotus nancymae</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Callithrix jacchus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Pteropus vampyrus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Pteropus alecto</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Galeopterus variegatus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Propithecus coquereli</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Ceratotherium simum</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Equus caballus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Camelus ferus</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Vicugna pacos</i>	TTCTCTCTAGATTTTGAAAACCTGATTATCTTTCCAGCAAAAGACAAAGTGAATCAAGAAAGTTTA	117
<i>Balaenoptera acutorostrata</i>	TGCTCCCGAGAGCTGAAAACCTGGTTCTTTTCCAAACAAAGAGAGAGTGAACACAGAGAGAGTCTA	117
<i>Physeter catodon</i>	TGCTCCCGAGAGCTGAAAACCTGGTTCTTTTCCAAACAAAGAGAGAGTGAACACAGAGAGAGTCTA	117
<i>Lipotes vexillifer</i>	TGCTCCCGAGAGCTGAAAACCTGGTTCTTTTCCAAACAAAGAGAGAGTGAACACAGAGAGAGTCTA	117
<i>Orcinus orca</i>	TGCTCCCGAGAGCTGAAAACCTGGTTCTTTTCCAAACAAAGAGAGAGTGAACACAGAGAGAGTCTA	117
<i>Sus scrofa</i>	TTCTCCCGGGAACCTGAAAACCTGGTTCTTTTCCAAACAAAGAGAGAGTGAACACAGAGAGAGTCTA	117
<i>Ovis canadensis</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Ovis aries musimon</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Pantholops hodgsonii</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Bos taurus</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Bison bison bison</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Bubalus bubalis</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Bos mutus</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Capra hircus</i>	TGCTCCCAAGAGCTGAAAACCTGGTTCTTTTGCACAGCAAAAGAGAGTGAACCAAGAGAGAGTCTA	117
<i>Otolemur garnettii</i>	TTCTCTCTAGAAATTGAAATTTGGTTCTTTTCACAGAGAAAGAGTGAACACAGAGAGAGTATA	117
<i>Tupaia chinensis</i>	CTCTCTCTAGAACTGAAGCCATTTCTCTTTCCAAACAAAGGGGTGAACACAGAGAGAGTCTT	117

<i>Eptesicus fuscus</i>	TTCTCCCTCGAGCTGAAACTCATTTCTCTTCCAAACAGAGAAAGAGTGAACACAGAGAAAGTTTA	117
<i>Myotis brandtii</i>	CTCTCCCTAGAAACTGAAACTCGTTCTCTCTTCCAAACAGAGAAAGAGTGAACACAGAGAAAGTTCA	117
<i>Myotis lucifugus</i>	CTCTCCCTAGAAACTGAAACTCGTTCTCTCTTCCAAACAGAGAAAGAGTGAACACAGAGAAAGTTCA	117
<i>Myotis davidii</i>	CTCTCCCTAGAAACTGAAACTCGTTCTCTCTTCCAGCAGAGAAAGAGTGAACACAGAGAAAGTTCA	117
<i>Leptonychotes weddellii</i>	TTCTCCGTCGGACTGAAACTGGCTCTCATCCAA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Odobenus rosmarus</i>	TTCTCCCTCGGATTGAAACTGGCTCTCATCCAA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Ursus maritimus</i>	TTCTCCCTCGAACTGAAACTGGCTCTCTTTCCA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Ailuropoda melanoleu</i>	TTCTCCCTCAAACTGAAACTGGCTCTCTTTCCA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Felis catus</i>	TTCCCCCTAGAGTTGAAACTGGCTCTCTTTCCA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Panthera tigris altaica</i>	TTCTCCCTAGAGTTGAAACTGGCTCTCTTTCCA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Mustela putorius</i>	TTCTCCCTGAACTGAAACTGGCTCTCTTTCCA--ACAAAAGAGTAAACACAGAGAGNGTTTA	114
<i>Canis lupus familiaris</i>	TTCTCCCTAGAAATTGAAACTGGCTTTCTTTCCA--ACAAAAGAGTAAACACAGAGAGAGTTTA	114
<i>Condylura cristata</i>	TCCTCCCTAGAACTTAACTGGTTATCTTCCAGCAAAAGAGTGAACACAGAGAGAGCTTC	117
<i>Sorex araneus</i>	TTCTCCCTAGAACTGAAACTGGTTCTTTTCCAAACACGGAGAAATAAACACAGAGAGAGTTTA	117
<i>Trichechus manatus latirostris</i>	TTCTCCCTAGAACTGAAATTAGTTCACTTTCCA--AAGAAAGAGTAAACACAGAGAAAGTTTA	114
<i>Loxodonta africana</i>	TTCTCCCTAGAACTGAAATTAGTTCACTTTCCA--ACAAAAGATGAACACAGAGAAAGTTTA	114
<i>Orycteropus afer</i>	TTCTCCCTAGAACTGAAAGTCAGTTCACTTTCCAAACAAAGAGTGAACAAAGAGAGTTTA	117
<i>Chrysochloris asiatica</i>	TTCTCCTTAGAACTGAAATTGGTTCACTTTCCA--ACAAAAGATGCACACAGACAGAGTTTA	114
<i>Tarsius syrichta</i>	TTCTCGTAGAAATTGAAACTGGTTCTCTACAAGCAAAAGAGTGAACCAAGAAAGTTTA	117
<i>Erinaceus europaeus</i>	TTCTCCCTAGAAATTGAAACTGTTTCTTGTTCCAACTAAGAAATTGAACACAGAGAGAGTTTA	111
<i>Dasyopus novemcinctus</i>	TTCTCCCTGGAACAGAAATTGGTTCTCTTTTCAGCAAAAGACAAAGTGAATAGAGAGAGTTTA	117
<i>Oryctolagus cuniculus</i>	TTCTCCCTAAAACTGAAACGGGTTCCCTTTCCAAACAAAGAAAGAAATGAACACAAAGGAGTTTA	117
<i>Ochotona princeps</i>	TTTTCCCTGGAACAGAAAGTCGATTCCTCTCAACAAAGAAAGAGTGAATGAGGATTTTA	117
<i>Elephantulus edwardii</i>	TTCTCGTAGGACTGAAAGTCGATTCCTCTCAACAAAGAAAGAGTGAATGAGAGAGAGTCTG	117
<i>Echinops telfairi</i>	GTCTCCCGAGAACTGAAACCCAGTGCACCTTCCAGCAAAAGAGTGAACACAGAGAGAGCTTA	117
<i>Dipodomys ordii</i>	TTTTCCCTAGAAAGAGAAATCTGGTTCTTTTCCA--GCCAAGAAATGACCACAGAGAGAGTTTA	111
<i>Ictidomys tridecemlineatus</i>	TTCTTTTGTAGAACTGAAACTAGTTCTCTCTCTCC--AACAAAAGAAATAAAGGAAAGAGAAATTTA	102
<i>Nannospalax galili</i>	TTCTCCCTAGAAAGCGAAACTGATTCCTCTTCCA--ATCAAGAAATGAACACACAAAGAGTTTA	114
<i>Cricetulus griseus</i>	TTCTCCCTAGAAACCGAAATCGATTCCTCTTCCA--ATCAAGAAATGAAGAGAGAGAAAGCTTG	114
<i>Mesocricetus auratus</i>	TTCTCCCTAGAAACCGAAAGTTGATTCCTCTTCCA--GTCAAGAAATGAAGAGAGAGAAAGCTTG	117
<i>Mus musculus</i>	TTCTCCCTAGAAACCGAAACGGATTCCTCTTCCA--ATTGTGGATGAACACAGAGAGAAAGCCTTA	87
<i>Manis temminckii</i> (5)	-----GAGAGAGTTTA	11
<i>Manis tricuspis</i> (4)	-----GAGAGAGTTTA	11
<i>Manis tetradactyla</i> (2)	-----GAGAGAGTTTA	11
<i>Manis gigantea</i> (1)	-----GAGAGAGTTTA	11
<i>Manis javanica</i> (8)	TTCTCCCTGGAACAGAACTGGTTCTCTTTCCA--ACAAAAGAGTGAACACAGAGAGAGTTTA	114
<i>Manis pentadactyla</i> (1)	TTCTCCCTGGAACAGAACTGGTTCTCTTTCCA--ACAAAAGAGTGAACACAGAGAGAGTTTA	114

<i>Homo sapiens</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Pan paniscus</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Pan troglodytes</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Gorilla gorilla</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Nomascus leucogenys</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Pongo abelii</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Rhinopithecus roxellana</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Macaca mulatta</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Mandrillus leucophaeus</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Cercocebus atys</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Macaca nemestrina</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Papio anubis</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Macaca fascicularis</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Colobus angolensis</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Chlorocebus sabaeus</i>	AAACTCTTGAATAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Aotus nancymae</i>	AAACTCTT---TAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	174
<i>Callithrix jacchus</i>	AAACTCTT---TAAAGTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	174
<i>Pteropus vampyrus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Pteropus alecto</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Galeopterus variegatus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Propithecus coquereli</i>	AAATCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Ceratotherium simum</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Equus caballus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Camelus ferus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Vicugna pacos</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Balaenoptera acutorostrata</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Physeter catodon</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Lipotes vexillifer</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Orcinus orca</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Sus scrofa</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Ovis canadensis</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Ovis aries musimon</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Pantholops hodgsonii</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Bos taurus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Bison bison bison</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Bubalus bubalis</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Bos mutus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Capra hircus</i>	AAACTCTTGAATAAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177
<i>Otolemur garnettii</i>	AAACTCTTAAATGAATTTGCAAAACCTTGTCAATTCAGCAGTGTCTACCAACACACAGGAAAAAC	177

<i>Tupaia chinensis</i>	AAACTCTTGAATACACTACGGAGCTCATCAATTTCAACAGTGTCTTACCACACAGGAAAAAC	177
<i>Eptesicus fuscus</i>	ACACTCTTGAATAAATTTGGACCCCTCATCCATTCAGCAGTGCCTTACCACACAGGATGGAC	177
<i>Myotis brandtii</i>	GAAACTCTTGAATAAATTTGGCAACCCCTCATCCATTTCTTCAAGTGTCTTGCCACACAGGAAGGAC	177
<i>Myotis lucifugus</i>	GAAACTCTTGAATAAATTTGGCAACCCCTCATCCATTCATCAGTGTCTTGCCACACAGGAAGGAC	177
<i>Myotis davidii</i>	GAAACTCTTGAATAAATTTGGCAACCCCTCATCCATTCATCAGTGTCTTGCCACACAGGAAGGAC	177
<i>Leptonychotes weddellii</i>	AGACTCTGGAATAAATTTGCAAAAAATCATCAATTCAGCAGTGTCTTACCACACAGGACAAAAAC	174
<i>Odobenus rosmarus</i>	AAACTCTGGAATAAATTTGCAAAAAATCATCAATTCAGCAGTGTCTTAGCACAGAGGACAAAAAC	174
<i>Ursus maritimus</i>	CGACTCTTGAATAAATTTGCAAAAAAGCATCAATTCAGCAGTGTCTTACCACACAGGACAAAAAC	174
<i>Ailuropoda melanoleu</i>	CGACTCTTGAATAAATTTGCAAAAAAGCATCAATTCAGCAGTGTCTTACCACACAGGACCAAC	174
<i>Felis catus</i>	AAACTCTTGAGTACATTTGCAAAAAAGCTCATCAATTCAGCAGTGTCTTGCCACACAGGAAAAAC	174
<i>Panthera tigris altaica</i>	AAACTCTTGAGTACATTTGCAAAAAAGCTCATCAATTCAGCAGTGTCTTGCCACACAGGAAAAAC	174
<i>Mustela putorius</i>	AAACTCTTGAATAAATTTCCCAAAATCATCAGTTCAGCCGTGTCTTGCCACACAGGACAAAAAC	174
<i>Canis lupus familiaris</i>	AAACTCTTGAATAGATTTCAAAAAATCATCCATTCAGCAGTGTCTTAGCACACAGGAAAAAC	174
<i>Condylura cristata</i>	AAACTTGACAAATGCAATTTGCAAAACCTCCACCATCAATTCAGCAGTGTCTTACCACACAGGAAAGAC	177
<i>Sorex araneus</i>	AAACTCTTGAGTAGATTTCAAAAGTTCGTCAATTCAGCAGTGTCTTACAAACACAGGAAAAAC	177
<i>Trichechus manatus latirostris</i>	AAACTCTTGAATAAATTTGCGGAGTTCCTCAATTCAGCAGTGTCTTACCACACAGGAAAAAC	174
<i>Loxodonta africana</i>	AAACTCTTGAATACATTTGGAGCTCTTCGATTCAGCAGTGTCTTACCACATAGGCAAC	174
<i>Orycteropus afer</i>	AAACTCTTGAATGCTTTGCAGAGTTCCTTCAATTCAGCAGTGTCTTACCACACAGGAAAAAC	177
<i>Chrysochloris asiatica</i>	AAGCTTTTGGATACATTCGAGAGCTCTTCAATTCAGCAGTGTCTTGAGATACAGGAAAAAT	174
<i>Tarsius syrichta</i>	AAACTCTTGAATAAATTTGCAAAAGTTCATCAGTTCAGCAGTGTCTTACCACACAGGAAAAAG	177
<i>Erinaceus europaeus</i>	AAACTGTTGAATAAATTTGCAAAAGTTCATCAATTCAGCAGTGTCTTACCACACAGGAAAAAC	171
<i>Dasylops novemcinctus</i>	AAACTCTTGAATAAATTTGCACAC---TTCAATTCAGCAGTGTCTTACCACACAGGAAAAAC	174
<i>Oryctolagus cuniculus</i>	AAACCCCTGGAATAAATTTACAAAAAGTTGAGAAATCCAGCAATGTCTTACCACACAGGAATGAT	177
<i>Ochotona princeps</i>	AAATCCTGGAACAATTTTCAAAATGTCGACAATTCAGCAATGCCTACAATACAGGAAAGAT	177
<i>Elephantulus edwardii</i>	AAGTTCTTGACCACACTGCGAGAGTTCCTTGATTCGGCGGTGCTTACCCACAGGAAAAAC	177
<i>Echinops telfairi</i>	AAGCTCTTGGGCACGTTTGCTTACTTCTTTGATTCGGCAGTGTCTTGCCACACAGGAAAAAC	177
<i>Dipodomys ordii</i>	CAGCTACTGAATAAATTTGCAAAACCTCATCCATCCAGCAGTGTCTTACCACACAGGAAAGAAT	171
<i>Ictidomys tridecemlineatus</i>	CAACTACTGAATAAATTTGCAAAACCTCATCGATCCAGCAGTGTCTTGCCATACAGGAAAGAAT	162
<i>Nannospalax galili</i>	CAACTACTGAATAAATTTGCAAAATTCATCAATTCAGCAGTGTCTTACCCACAGGAAAAAT	174
<i>Cricetulus griseus</i>	CACCTACTGAAAACTTTGCAAAATTCGTGCGTGTCTTACCAGTGTCTTGCGTCAACAGGAAAAAT	174
<i>Mesocricetus auratus</i>	CACCCACCCACAATGTTGCCAGTCTCATCGGTGCCAGCAGTGTCTTGCTTCAACAGGAAAAAT	177
<i>Mus musculus</i>	CAACTACTGAAAACTTTGCCAAAGCTCGTCAAGTTCAGCAGTGTCTTAGCACACAGGAAAGAAT	147
<i>Manis temminckii</i> (5)	AAACTCTTGAATAAATTTGAAAAACCCCATCGATTCAGCAGCACCCTACCACACAGGAAAAAT	71
<i>Manis tricuspis</i> (4)	AAACTCTTGAATAAATTTGAAAAACCCCATCGATTCAGCAGCACCCTACCACACAGGAAAAAT	71
<i>Manis tetradactyla</i> (2)	AAACTCTTGAATAAATTTGAAAAACCCCATCGATTCAGCAGCACCCTACCACACAGGAAAAAT	71
<i>Manis gigantea</i> (1)	AAACTCTTGAATAAATTTGAAAAACCCCATCGATTCAGCAGTGTCTTACCACACAGGAAAAAT	71
<i>Manis javanica</i> (8)	AAACTCTTGAATAAATTTGAAAAACCCCATCAATTCAGCAGTGTCTTACCACACAGGAAAAATC	174
<i>Manis pentadactyla</i> (1)	AAACTCTTGAATAAATTTGAAAAACCCCATCAATTCAGCAGTGTCTTACCACACAGGAAAAATC	174

*

*

*

<i>Homo sapiens</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Pan paniscus</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCTTCAGC-----AGTAC	216
<i>Pan troglodytes</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Gorilla gorilla</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Nomascus leucogenys</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Pongo abelii</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Rhinopithecus roxellana</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Macaca mulatta</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Mandrillus leucophaeus</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Cercocebus atys</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Macaca nemestrina</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Papio anubis</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Macaca fascicularis</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Colobus angolensis</i>	TTTCTGCTTCCCTCAGAAAGTCTTTGAGTCCCTCAGC-----AGTAC	216
<i>Chlorocebus sabaeus</i>	TTTATGCTTCCCTCAGAAAGTCTGTGAGTCTCAAC-----AGCAC	213
<i>Aotus nancymae</i>	TTTATGCTTCCCTCAGAAAGTCTGTGAGTCTCAAC-----AGCAC	213
<i>Callithrix jacchus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Pteropus vampyrus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Pteropus alecto</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Galeopterus variegatus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Propithecus coquereli</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Ceratotherium simum</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAC	216
<i>Equus caballus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Camelus ferus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Vicugna pacos</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Balaenoptera acutorostrata</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Physeter catodon</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Lipotes vexillifer</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Orcinus orca</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Sus scrofa</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Ovis canadensis</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Ovis aries musimon</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Pantholops hodgsonii</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Bos taurus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Bison bison</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Bubalus bubalis</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Bos mutus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Capra hircus</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216
<i>Otolemur garnettii</i>	TTTCTGCTTCCCTCAGAAAGTCTGTGAAATCCTCACC-----TGTAT	216

<i>Tupaia chinensis</i>	TTCTGCTTCCCATAAAGTCTGTGAACCCCTCAGC-----AGTTC	216
<i>Eptesicus fuscus</i>	TTCTGCTTCCCCAGAAAGGCTGTGGATCCTCACC-----AGGAC	216
<i>Myotis brandtii</i>	TTCTGCTTCCCCAGGAGGCTGGCAATCCTCACC-----AGGAC	216
<i>Myotis lucifugus</i>	TTCTGCTTCCCCAGGAGGCTGGGAATCCTCACC-----AGGAC	216
<i>Myotis davidii</i>	TTCTGCTTCCCCAGGAGGCTGGGAATCCGCACC-----AGGAC	216
<i>Leptonychotes weddellii</i>	TTCTACTTCCCCAGCAGTCTCTGAATCGTCACT-----GGGAC	213
<i>Odobenus rosmarus</i>	TTCTACTTCCCCAGCAGTCTATGAATCATCACT-----GGTAC	213
<i>Ursus maritimus</i>	TTCTACTTCCCCAGCAGTCTGTGACTCGTCGCC-----GGTAC	213
<i>Ailuropoda melanoleu</i>	TTCTACTTCCCCAGCAGTCTGTGACTCGTCGCC-----GGTAC	213
<i>Felis catus</i>	TTCTGCTTCCCCAGCGGTCTGTGAATCCTCGCC-----AGTAC	213
<i>Panthera tigris altaica</i>	TTCTGCTTCCCCAGCGGTCTGTGAATCCTCGCC-----AGTAC	213
<i>Mustela putorius</i>	TTCTCTCTTCCCCAGCAGTCTGTGAATCGTCGCC-----GGTCC	213
<i>Canis lupus familiaris</i>	TTCTGCTTCCCCAGCAGTCTGTGAATCGTCACC-----AGTAC	213
<i>Condylura cristata</i>	TTCTGCTTCCCCAGAAAGTCTGTGAATCCTTGCC-----AGTAC	216
<i>Sorex araneus</i>	TTCTGCTTCCCTGAGGTCTATGAATCCTTGACC-----AAAAC	216
<i>Trichechus manatus latirostris</i>	TTTGCTTCCCCAGAAAGTCTATGAATCCTCACC-----ATTAC	213
<i>Loxodonta africana</i>	TTTGACTTCCCCAGAAAGTCTATGAATCCTCACC-----AGTAC	213
<i>Orycteropus afer</i>	TTCATGCTTCCCTCAAAAGTCTATGAATCCTCACC-----AGTAC	216
<i>Chrysochloris asiatica</i>	TTCATGCTTCCCCAGAAAGTCTATGAATCCTCACC-----AGTAC	213
<i>Tarsius syrichta</i>	TTTCTGCTTCCCTC-----AGC-----AGTAC	198
<i>Erinaceus europaeus</i>	TTCTGCTTCCCTGAAAGTCTCCAAGTCTCCTCACT-----GGTAT	210
<i>Dasylops novemcinctus</i>	TTTCTGCTTCCCCAGGAGTCTATGAATCCTTACC-----AGTAC	213
<i>Oryctolagus cuniculus</i>	TTCCGGCTTCCCTGAAAGTCTGTGAATCCTCATC-----AGTAC	216
<i>Ochotona princeps</i>	TTCCGGCTTCCCCAGATGTCTGTGAATCTTTATC-----AGTAC	216
<i>Elephantulus edwardii</i>	TTCCGGCTTCCCCACAAAGTCCATGAATTCCTCACC-----AGGAC	216
<i>Echinops telfairi</i>	TTCTGCTTCCCTCAGAAAGGCCATAAACCCCTCCTC-----GATGC	216
<i>Dipodomys ordii</i>	TTTTTGTTCCTTGGAATCCATGAGTCTCCTCACC-----AGTAC	210
<i>Ictidomys tridecemlineatus</i>	TTCCGGCTTCCCTCAGAAATTCCTGAATTTTACC-----AGTAC	201
<i>Nannospalax galili</i>	TTCTGCTTCCCTCAGAAAGTCTGTGGCTCCTTACC-----AGTAT	213
<i>Cricetulus griseus</i>	TTCCAGCTTCCCTTGGCAGTCTGTGAGTCGTCAACC-----AGTAC	213
<i>Mesocricetus auratus</i>	TTCTTGTTCCTCTGTGAGTCTGTGAGTCGTCAACC-----AGTAC	216
<i>Mus musculus</i>	TTCTGCTTTCCTCAGAGCTGTGAGTCCTCACC-----AGTAC	186
<i>Manis temminckii</i> (5)	TTCTACTTCCCCACAGTACAGAAATCCTCCTGAATCTTTCTGAATCCTCGTAGTAC	131
<i>Manis tricuspis</i> (4)	TTCTACTTCCCCACAGTACAGAAATCCTCCTGAATCTTTCTGAATCCTCGTAGTAC	131
<i>Manis tetradactyla</i> (2)	TTCTACTTCCCCACAGTACAGAAATCCTCCTGAATCTTTCTGAATCCTCGTAGTAC	131
<i>Manis gigantea</i> (1)	TTCTGCTTCCCCACAGTACAGAAATCCTCCTGAATCTTTCTGAATCCTCGTAGTAC	131
<i>Manis javanica</i> (8)	TTCTGCTTCCCCACCAATACCAGGATCCTCCTGTATCTTTCTGAATCCTCACTAGTAC	234
<i>Manis pentadactyla</i> (1)	TTCTGCTTCCGCCACCAATACCAGGATCCTCCTGTATCTTTCTGAATCCTCACTAGTAC	234

*
**

<i>Homo sapiens</i>	CAAAAAAGGACACACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Pan paniscus</i>	CAAAAAAGGACACACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Pan troglodytes</i>	CAAAAAAGGACACACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Gorilla gorilla</i>	CAAAAAAGGACACACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Nomascus leucogenys</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Pongo abelii</i>	CAAAAAAGGACACACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Rhinopithecus roxellana</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Macaca mulatta</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Mandrillus leucophaeus</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Cercocebus atys</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Macaca nemestrina</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Papio anubis</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Macaca fascicularis</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Colobus angolensis</i>	CAAAAAAGGACATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Chlorocebus sabaeus</i>	CAAAAAAGGATATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Aotus nancymae</i>	CGAAAAAGGATATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	273
<i>Callithrix jacchus</i>	TGAAAAAGGATATACTCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	273
<i>Pteropus vampyrus</i>	CAGAAAAGGATGTGCACCTGGCCATTCTTCATGAGACGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Pteropus alecto</i>	CAGAAAAGGATGTGCACCTGGCCATTCTTCATGAGACGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Galeopterus variegatus</i>	CAAAAAAGGACATGACCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Propithecus coquereli</i>	CAAAAAAGGACATGACCTGGCCATTCTCCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Ceratotherium simum</i>	CAGAAAAGGACACGACCTGGCTGTCTTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Equus caballus</i>	CAGAAAAGGACACGACCTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Camelus ferus</i>	CAGAAAAGGACACATACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Vicugna pacos</i>	CAGAAAAGGACACATACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Balaenoptera acutorostrata</i>	CAGAAAAGGACAAAGCAGCTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Physeter catodon</i>	CAGAAAAGGACAAAGCAGCTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Lipotes vexillifer</i>	CAGAAAAGGACAAAGCAGCTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Orcinus orca</i>	CAGAAAAGGACAAAGCAGCTGACCATCTTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Sus scrofa</i>	CAGAAAACAAAGCAGCTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Ovis canadensis</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Ovis aries musimon</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Pantholops hodgsonii</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Bos taurus</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Bison bison bison</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Bubalus bubalis</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Bos mutus</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Capra hircus</i>	CAGAAAAGGACAAAGTACTGGCCATTCTTCATGAGATGCTTTCAGCAGATCTTTCAGCCTCTTTC	276
<i>Otolemur garnettii</i>	CAGGAAAGGACATGCACTGGCCCTGCTCCATGAGACACTGCAGACATCTTTCAGCCTCTTTC	276

<i>Tupaia chinensis</i>	CAAAAAAGACATACACTGGCCATTCTCCACGAGATGCTTCAGCAGATCTTCAACCTCTTC	276
<i>Eptesicus fuscus</i>	GAGAAAGGATACACACTGATCATTTCTACATGAGATGCTTCAGCAGATCTTCAAGCTCTTC	276
<i>Myotis brandtii</i>	GAGAAAGGATACGCACTGACCATTTCTACATGAGATGCTTCAGCAGATCTTCAAGCTCTTT	276
<i>Myotis lucifugus</i>	GAGAAAGGATACGCACTGACCATTTCTACATGAGATGCTTCAGCAGATCTTCAAGCTCTTT	276
<i>Myotis davidii</i>	CAGAAAGGATCGCGCTGACCATTTCTACATGAGATGCTTCAGCAGATCTTCAAGCTCTTT	273
<i>Leptonychotes weddellii</i>	CAGAGAGGACAGGCACTGGTCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Odobenus rosmarus</i>	CAGAGAGGACAAAGCACTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Ursus maritimus</i>	CAGAGAGGCAAGCCCTGGCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Ailuropoda melanoleu</i>	CAGAGAGGCAAGCACTGGGNATTTCTCATGAGATGCTTCAGCAAAATCTTCAACCTCTTC	273
<i>Felis catus</i>	CAGAAAGCAAGCACTGGCCATTTCTACGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Panthera tigris altaica</i>	CAGAAAGCAAGCACTGGCCATTTCTACGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Mustela putorius</i>	CAGAGTGACAGGCCCTGGCCATTTCTCATGAGATGCTTCGCAAACTTCAGCCTCTTC	273
<i>Canis lupus familiaris</i>	CAGAAAGCAAGCACTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Condylura cristata</i>	CAGACCAGACGGCACTAGCGATTTCTCATGAGATGCTTCAGCAGATCTTCAAGCTCTTC	276
<i>Sorex araneus</i>	CAGACGGACATGCAATCGCCATTTCTCAGGAAAGTCTTCAACAGATCTTCCATCTTTTC	276
<i>Trichechus manatus latirostris</i>	CATAAAGCACACAGTGGCCATTTCTCATGAGATACTCCAGCAGATCTTCAACCTCTTC	273
<i>Loxodonta africana</i>	CATAAAGCACACAGTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Orycteropus afer</i>	CATAAAGCACACAGTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Chrysochloris asiatica</i>	AGTAAAGCACACAGTGGCTATTCTCCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	276
<i>Tarsius syrichta</i>	CAAAAAGGACACACAATGGCAATTTCTCCACGACATGCTTCTGCAGATATTCAACCTCTTC	258
<i>Erinaceus europaeus</i>	CAGACAGAACATGCACTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAATCTCTTT	270
<i>Dasyopus novemcinctus</i>	CAGAAAGGACACGAGTGGCCATTTCTCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	273
<i>Oryctolagus cuniculus</i>	CAAAAAGGACAAAGTACTGGCTGTAATCCATGAGACACTTCAGCAGATCTTCAAGCTCTTC	276
<i>Ochotona princeps</i>	CAGAAAGGACAAAGTACTGGCTGTTCTCCATGAGACATTTCAAGCAGTACTCATCTCATC	276
<i>Elephantulus edwardii</i>	CACGAAAGGACACGAGTGGCCATTTCTCCATGAGATGCTTCAGCAGATCTTCAAGCTCTTC	276
<i>Echinops telfairi</i>	AGCAGAGGGGATGACAGGGCTGTTCTCCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	276
<i>Dipodomys ordii</i>	CAGAAAGAAACATATACTAGCCTTTATTTATGAGGTGCTACAACAGATCTTCAAGCTCTTC	270
<i>Ictidomys tridecemlineatus</i>	CCAAAAAGGGCATGTACTGGCCATTTCTCCATGAGATGCTTCAGCAGATCTTCAACCTCTTC	261
<i>Nannospalax galili</i>	CAAAAAAGGACATGTACTTGAGTTATTCAAGAGATTTCTTCAACAAATCTTCAAGCTCTTC	273
<i>Cricetulus griseus</i>	CAAAAAAGGACATGTACTGGCTGTTCTGCACGAGTCTTTCAGCAGATCTTCAAGCTCTTC	273
<i>Mesocricetus auratus</i>	CAAAAAAGGACATGTACTGGCTGTTCTGCACGAGATCTTTCAGCAGATCTTCAAGCTCTTC	276
<i>Mus musculus</i>	CAAGAGGACAGGTGGCTGTTCTGCACGAGATCTTTCAGCAGATCTTCAAGCTCTTC	246
<i>Manis temminckii</i> (5)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	190
<i>Manis tricuspis</i> (4)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	190
<i>Manis tetradactyla</i> (2)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	190
<i>Manis gigantea</i> (1)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	190
<i>Manis javanica</i> (8)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	293
<i>Manis pentadactyla</i> (1)	CAGAAAGAC-ACACAAATAACCATCTTCTGAGATGCTTCAGCAGATCTTCAACCTGTTTC	293

* * * * *

<i>Homo sapiens</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAACACACGAGAAAAATTCCTCATTTCAA	336
<i>Pan paniscus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAACACATGGAGAAAATTCCTCATTTCAA	336
<i>Pan troglodytes</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAACACATGGAGAAAATTCCTCATTTCAA	336
<i>Gorilla gorilla</i>	AGGGCAGATATTTCTCTGGATGGTTGGAGGAAAAACACATGGAGAAAATTCCTCATTTCAA	336
<i>Nomascus leucogenys</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAACACACGAGAAAATTCCTCATTTCAA	336
<i>Pongo abelii</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAACACATGGAGAAAATTCCTCATTTCAA	336
<i>Rhinopithecus roxellana</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Macaca mulatta</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Mandrillus leucophaeus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Cercocebus atys</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Macaca nemestrina</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Papio anubis</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Macaca fascicularis</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Colobus angolensis</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Chlorocebus sabaeus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Aotus nancymae</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	333
<i>Callithrix jacchus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	333
<i>Pteropus vampyrus</i>	GGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Pteropus alecto</i>	GGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Galeopterus variegatus</i>	GAAGCAAAATTTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Propithecus coquereli</i>	CGGGTAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Ceratotherium simum</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Equus caballus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Camelus ferus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Vicugna pacos</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Balaenoptera acutorostrata</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Physeter catodon</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Lipotes vexillifer</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Orcinus orca</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Sus scrofa</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Ovis canadensis</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Ovis aries musimon</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Pantholops hodgsonii</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Bos taurus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Bison bison</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Bubalus bubalis</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Bos mutus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Capra hircus</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336
<i>Otolemur garnettii</i>	AGGGCAAAATATTTCTCTGGATGGTTGGAGGAAAAATCACATGGAGAAAATTCCTCATTTCAA	336

<i>Tupaia chinensis</i>	AGTGCAAAATGTTTCTCCGGATGATTGGGAAGAACACACTACAAAGAGAAAAATTCCTCATTTGAG	336
<i>Eptesicus fuscus</i>	CAGGCAAAATATTTCTCTGGAAGGTTTGAAGGAAAGCCACATGGAGAAGTTTCCTCATTTGAG	336
<i>Myotis brandtii</i>	CAGGCAAAATACTTCTCTGGAAGGTTTGAAGGAGAGCCACATGGAGAAGTTTCCTCATTTGAG	336
<i>Myotis lucifugus</i>	CAGGCAAAATACTTCTCTGGAAGGTTTGAAGGAAAGCCACATGGAGAAGTTTCCTCATTTGAG	336
<i>Myotis davidii</i>	CAGGCAAAACACTTCTCTGGAAGGTTTGAAGGAAAGCCGATGGAGAGTTTCCTCATTTGAG	336
<i>Leptonychotes weddellii</i>	AGGGCAGATATTTCTCTGATGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAA	333
<i>Odobenus rosmarus</i>	AGGGCAGATATTTCTCTGATGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAA	333
<i>Ursus maritimus</i>	AGGGCAGATATGTTCTCTGATGGCTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAA	333
<i>Ailuropoda melanoleu</i>	AGGGCAGATATGTTCTCTGATGGCTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAA	333
<i>Felis catus</i>	AGGGCAAAATACTTCTCTGATGGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAG	333
<i>Panthera tigris altaica</i>	AGGGCAAAATACTTCTCTGATGGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAG	333
<i>Mustela putorius</i>	AGGGCAGGTATTTCTCTGATGGTTGGAGGAAATTCGCGTGGAGAAATTTTCCTCACTGAA	333
<i>Canis lupus familiaris</i>	AGGGCAAAATATTTCTCTGATGGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCACTGAA	333
<i>Condylura cristata</i>	AAAGCAAAAGATTTCTCTGATGGTTGGAGGAAAGGCACATGGAGAAATTTTCCTCATTTGTT	336
<i>Sorex araneus</i>	AGGGCAAGAACTTCTCTGATGGTTGGAGGAAAGGCACACACGAGCAATTCCTCCTTGAA	336
<i>Trichechus manatus latirostris</i>	AGGACAAATCTTTCTCTGACAGTTGGAGGAAAGCCAGATGGAGAAATTCCTCATTTGAA	333
<i>Loxodonta africana</i>	AGGACAAATCTTTCTCTGACAGTTGGAGGAAAGCCAGATGGAGAAATTCCTCATTTGAA	333
<i>Orycteropus afer</i>	TGGACAAATCTTTCTCTGACAAATTTGGAGGAAAGCCACATGGAGAAATTCCTCATTTGAA	336
<i>Chrysochloris asiatica</i>	AGGACAAATCTTTCTCTGACAGTTGGAGGAAAGCCATGTGGAGAAATTCCTCATTTGAA	333
<i>Tarsius syrichta</i>	AGAAATAAATATTTCTCTGATGGTTGGAGGAAAGTTACATGGAGAAATTTCTCATTTGAA	318
<i>Erinaceus europaeus</i>	AGGGTAAATATTTCTCTGATGACTGTGAGGAAAGCTACATGGAGAAATTCCTCATTTGAG	330
<i>Dasyopus novemcinctus</i>	AGGGAAAAAGTTGTTCTTTAGCCATTTGGAGGAAAGCCAGGTGGAGAAATTCCTCATTTGAG	333
<i>Oryctolagus cuniculus</i>	AGAGAAAAATATTTCCCTGGTGGTGGAGGAAAGAAATGACTTGGAGAAATTCCTGATTTGAA	336
<i>Ochotona princeps</i>	AAAGAAAAATTTTGGCCCTGGATGGTGGAGGAAAGAAATACCTGGCAAAATTCCTTACTGAG	336
<i>Elephantulus edwardii</i>	AGGACAAACCACTCTCTGGGTAACCTGGAGCAAAAGCCACATGGAGAAATTCCTCACTGAA	336
<i>Echinops telfairi</i>	GGGGCAGATCAGTCTCTGATGATGGGATGAAAGCCCTTGTGGAGAAATTCCTCGTTGAG	336
<i>Dipodomys ordii</i>	CAGGCAAAAGACTTTTCTGGTCTTCAGAAAGGAAAGCCCA--TACAAAGTTCCCTCACTGAA	327
<i>Ictidomys tridecemlineatus</i>	AAGAAAAATATTTCTCTGGGAGTAGGGAGGAAAGCCACATTCCTCAATATTCCTCACTGAT	321
<i>Nannospalax galili</i>	CAAGCAAAATGTTTCTCTAGTCAGTTGGAGGAAAGAACTACATAGAGAAATTTCTCACTGTA	333
<i>Cricetulus griseus</i>	CAGGCACATGTTTCTCTCGGGCAATTTGGAGGAAAGAAACCATATAGAGAGAGTCTTAGGTGCA	333
<i>Mesocricetus auratus</i>	CAGGCACACCTTTTCTCTGGGCAATTTGGAGGAAAGAAACCATATAGAGAGAGTCTTAGGGCA	336
<i>Mus musculus</i>	CAGACACATGGACTATGGGCAATTTGGAGGAAAGAAACCATATAGAGAGAGTCTTAGGTGCG	306
<i>Manis temminckii</i> (5)	AGGACAAAGTATGTTCTTTGGACAGTTGGGAAAGAAAGCCATGTGGAGAAATTAATCAACGAA	250
<i>Manis tricuspis</i> (4)	AGGACAAAGTATGTTCTTTGGACAGTTGGGAAAGAAAGCCATGTGGAGAAATTAATCAACGAA	250
<i>Manis tetradactyla</i> (2)	AGGACAAAGTATGTTCTTTGGACAGTTGGGAAAGAAAGCCATGTGGAGAAATTAATCAACGAA	250
<i>Manis gigantea</i> (1)	AGGACAAAGTATGTTCTTTGGACAGTTGGGAAAGAAAGCCATGTGGAGAAATTAATCAACGAA	250
<i>Manis javanica</i> (8)	AGGACAAAGTATGCTTTGGACAGTTGGGAGAGAAAGCCATGTGGAGAAATTAATCAACGAA	353
<i>Manis pentadactyla</i> (1)	AGGACAAAGTATGTTCTTTGGACAGTTGGGAGAGAAAGCCATGTGGAGAAATTAATCAACGAA	353

*

<i>Homo sapiens</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Pan paniscus</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Pan troglodytes</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Gorilla gorilla</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Nomascus leucogenys</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Pongo abelii</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Rhinopithecus roxellana</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Macaca mulatta</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Mandrillus leucophaeus</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Cercocebus atys</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Macaca nemestrina</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Papio anubis</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Macaca fascicularis</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Colobus angolensis</i>	CTTCATCGACAGCTCGAATACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Chlorocebus sabaeus</i>	CTTCATCAACAGCTAGTAATAACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	393
<i>Aotus nancymae</i>	CTTCATCAACATCTGGAATACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	390
<i>Callithrix jacchus</i>	CTTCATCAACATCTGGAATACCTAGAA---GCACTCATGGGACTGGAAGCAGAGAAAGCTA	390
<i>Pteropus vampyrus</i>	CTTCATCAACAACTGGAATACCTAGAA---ACGCTCATGAGACAGCAAGCAGAAACAGAAAA	393
<i>Pteropus alecto</i>	CTTCATCAACAACTGGAATACCTAGAA---ACGCTCATGAGACAGCAAGCAGAAACAGAAAA	393
<i>Galeopterus variegatus</i>	CTTTATCAATAGCTGGAATACCTAGAA---GCACTGATGGAAGTGGAAAGCAGGACAGAAAA	393
<i>Propithecus coquereli</i>	CTTCATCAACAGCTGGAATACCTAGAA---GTACTCATAGGACTGGAAGCAGAGCAGAAAA	393
<i>Ceratotherium simum</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATTTGGACTGGAAGCAGAGCAGAAAA	393
<i>Equus caballus</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCGCTCATGGGACTGGAAGCAGAGCAGAAAA	393
<i>Camelus ferus</i>	CTTCATCAACAGCTGGAATACCTAGAA---ACACTCATACGACTGCAAGCTGAACACAGAGA	393
<i>Vicugna pacos</i>	CTTCATCAACAGCTGGAATACCTAGAA---ACACTCATACAACACTGCAAGCTGAACACAGAGA	393
<i>Balaenoptera acutorostrata</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAGGCAGAGCAGAAAA	393
<i>Physeter catodon</i>	CTTCATCAACAGCTGGAATACCTAGAA---GTGCTCATGAGACTGCAAGCAGAGCAGAAAA	393
<i>Lipotes vexillifer</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCGCTCATGAGACGGCAAGCAGAGCAGAAAA	393
<i>Orcinus orca</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCGCTCATGAGACGGCAAGCAGAGCAGAAAA	393
<i>Sus scrofa</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAGGCAGAGCAGAAAA	393
<i>Ovis canadensis</i>	CTTCTTCAACAGCTGGAATACCTAGAA---GCASTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Ovis aries musimon</i>	CTTCTTCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Pantholops hodgsonii</i>	CTTCTTCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Bos taurus</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Bison bison bison</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Bubalus bubalis</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Bos mutus</i>	CTTCATCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Capra hircus</i>	CTTCTTCAACAGCTGGAATACCTAGAA---GCACTCATGAGACTGCAAGCAAAAGAAAGAAAA	393
<i>Otolemur garnettii</i>	CTGCAATCAGCAGCTGGAATACCTAGAA---GCTCTCACGGGACTGGAAGCAGAGCAGAGA	393

<i>Tupaia chinensis</i>	CTTCATGAACAGCTAGACTATCTGGAA---GGACTCATGGAGCTGGAAGCAGAGCAGAAG	393
<i>Eptesicus fuscus</i>	CTTCATCAACAGCTAGAAATACTTAGAA---ATGCTCATGGGGCTGCAAAACAGATCAGAAA	393
<i>Myotis brandtii</i>	CTGCAATCAACAGCTAGAAATACCTAGAA---ACGCTCATGGGACTGCAAAACAGATCAGAAA	393
<i>Myotis lucifugus</i>	CTGCAATCAACAGCTGGAATACCTAGAA---ACGCTCATGGGACTGCAAAACAGATCAGAAA	393
<i>Myotis davidii</i>	CTGCAATCAACAGCTGGAATACCTAGAA---ACGCTCAGGGGACTGCAAAACAGATCAGAAA	393
<i>Leptonychotes weddellii</i>	CTTCATCAACAGCTGGAATACCTAGAA---GGGCTTAGAGGGCTGGAAGTAGAGCAGAAAA	390
<i>Odobenus rosmarus</i>	CTTCATCAACAGCTGGAATACCTAGAA---ATGCTTAAAGGCTGGAAGCAGAGCAGAAAA	390
<i>Ursus maritimus</i>	CTTCATCAACAGCTGGAACACCTAGAA---GCGCTTAGAGGGCTGCAAGCAGAGCAGAAAC	390
<i>Ailuropoda melanoleu</i>	CTCCATCAACAGCTGGAACACCTAGAA---GCGCTTAGAGGGCTGCAAGCAGAGCAGAAAC	390
<i>Felis catus</i>	CTTCATCAACAGCTGGAATACCTAGAA---GAACTCACAGGCCCGGAAAGCAGAGCAGGAC	390
<i>Panthera tigris altaica</i>	CTTCATCAACAGCTGGAATACCTAGAA---ACACTCACAGGCCCTGGAAGCAGAGCAGGAC	390
<i>Mustela putorius</i>	CTTCATCAACAGCTGGAATACCTAGGG---GCCCTCAGAGGCCCTGGAGCAGAGCAGAGAAC	390
<i>Canis lupus familiaris</i>	CTTCATCAACAGCTGGAATACCTAGAA---GTGCTGATGAGCCTGGAAGCAGAGCAGGAAAC	390
<i>Condylura cristata</i>	CTTCATCAACAGCTTGAATACCTAGAA---GCACTGACAGAGCTGGAAGCAGAGAACAGAAAA	393
<i>Sorex araneus</i>	ATTACCAACAACTGGAATATCTAGAA---ATGCTTCTGAAATCGGAAGCAGAGAACAGAAAA	393
<i>Trichechus manatus latirostris</i>	CTTTACCAACAGCTGGAACATCTGGAG---GCACTCATGGGACTGGAAGCAGAGAACAGAAAG	390
<i>Loxodonta africana</i>	CTTTATGGACAGCTGGAACATCTAGAGACAGCATCTATGGGACTGGAAGCAGAGAACAAAAAG	393
<i>Orycteropus afer</i>	CTTCATCAACAGCTAGAAATATCTACAG---GCACTCATTTGGACTAGAAAGCAGAGAACAGAAA	393
<i>Chrysochloris asiatica</i>	CTTCATCAACAGCTGGAATATCTAGAG---GCAGCAATGGGACTGGAAGCAGAGAACAGAAAT	390
<i>Tarsius syrichta</i>	CTTCATCAACAACTGGAATACCTAGAG---TCACCTCATGGGGCTGGAAGCAGAGCAGAGATA	375
<i>Erinaceus europaeus</i>	CTTCATCAACAGCTGAGCTCCTAGAA---GCACTCATAGGACTGGAAGCTGAACAGAAAT	387
<i>Dasyopus novemcinctus</i>	CTTCATCAACAGCTTGAACACCTAGAG---GCACTCATAAAGACAGGAACCCAGAGCTGAAA	390
<i>Oryctolagus cuniculus</i>	CTTCATCAACAGCTGAGTACCTAGAA---GGATTTCATGGAACCTGGAAGCGAAGCAGGAA	393
<i>Ochotona princeps</i>	CTTCATCAACAGCTACAATACCTAGAA---GGATTTCATAGAACTGGAACCAAAACAGAGA	393
<i>Elephantulus edwardii</i>	CTTCATCAGCAGCTGGACCATCTAGAGG-----CACTAGAAAGCAGAGAACAGAAAG	384
<i>Echinops telfairi</i>	CTTCACCAACAACTGGAACGCTAGAG---ACGGTTCATGGGGCTGGAAGCAGAGCAGAAAG	393
<i>Dipodomys ordii</i>	CTTCACCAAGCAGCTGGAATACCTGCAA---GTATTTCATGAGACTGGAAGCAGAGCAGCAA	384
<i>Ictidomys tridecemlineatus</i>	CTTCATCAACAACTAGAAATACCTGGAA---ACACTCTTTGGGATTGGAAGCAGAGAACAGAAA	378
<i>Nannospalax galili</i>	CTTCACCAACAGCTGGAATACACAGAA---TCACTCCTTGGAACTGGAGGCAGAGGAGAGA	390
<i>Cricetulus griseus</i>	CTTCACCAACAGCTGGAATACGTTGGAG---TTGCTAGCTGGAACCTAAAAGCAGAGTCGAAAG	390
<i>Mesocricetus auratus</i>	CTTCACCAACAGCTGGAATACATGGAG---TCGCTAGCTGGAACCTAAAAGCAGAGCAG---	390
<i>Mus musculus</i>	CTTCACCGGAGCTGGAATACGTTGGAG---TCACTGGTGGGACTGGAACGCAGCGCAGAAAG	363
<i>Manis temminckii</i> (5)	CTTCATCCACAGCTGGAATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	307
<i>Manis tricuspis</i> (4)	CTTCATCCACAGCTGGAATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	307
<i>Manis tetradactyla</i> (2)	CTTCATCCACAGCTGGAATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	307
<i>Manis gigantea</i> (1)	CTTCATCCACAACTGGAATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	307
<i>Manis javanica</i> (8)	CTTCATCCACAGCTGGTATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	410
<i>Manis pentadactyla</i> (1)	CTTCATCCACAGCTGGTATACCTTAGAA---GCTTTTCATGGGACTGGAAGCAGAGCAAAAAA	410
	* * *	

<i>Homo sapiens</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Pan paniscus</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Pan troglodytes</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Gorilla gorilla</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Nomascus leucogenys</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCTGAAGGATC	453
<i>Pongo abelii</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Rhinopithecus roxellana</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Macaca mulatta</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Mandrillus leucophaeus</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Cercocebus atys</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Macaca nemestrina</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Papio anubis</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Macaca fascicularis</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Colobus angolensis</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Chlorocebus sabaues</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Aotus nancymae</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCTGAAGAAATC	450
<i>Callithrix jacchus</i>	AGTGGTACTTTTGGGTAGTGATAACCTTAGATTACAAAGTTAAAAATGTACTTCCAAAGAAATC	450
<i>Pteropus vampyrus</i>	AGTGGCATCTTAGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCAAAGGATC	453
<i>Pteropus alecto</i>	AGTGGCACCTTAGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCAAAGGATC	453
<i>Galeopterus variegatus</i>	AGTGGTGGCTTGGGCAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Propithecus coquereli</i>	AATGGTACCTTGGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCGAAAGGATC	447
<i>Ceratotherium simum</i>	AGTGGCACATTTGGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Equus caballus</i>	TGTGGCCACTGGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCGAAAGGATC	453
<i>Camelus ferus</i>	AGTGGCATCTTGGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCAAAGGATC	453
<i>Vicugna pacos</i>	AGTGGCACCTTGGGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTACTTCCAAAGGATC	453
<i>Balaenoptera acutorostrata</i>	AGAGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATATATTTCCAAAGGATA	453
<i>Physeter catodon</i>	AGAGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATATATTTCCAAAGGATA	453
<i>Lipotes vexillifer</i>	AGAGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATATATTTCCAAAGGATA	453
<i>Orcinus orca</i>	AGAGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATATATTTCCAAAGGATA	453
<i>Sus scrofa</i>	AGTGACACCTTGTGTAGTGAGAACCTTAGGTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Ovis canadensis</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Ovis aries musimon</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Pantholops hodgsonii</i>	AGTAACACCTTGGGCAGTGAGAACCTTAGGTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Bos taurus</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Bison bison bison</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Bubalus bubalis</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Bos mutus</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Capra hircus</i>	AGTGACACCTTGGGCAGTGAGAACCTTAGATTACAGGTTAAAAATGTATTTCCAAAGGATC	453
<i>Otolemur garnettii</i>	AGTGGGACTGTGGGTATTGAGAAACTCAGGTTACAGGTTAAAAATGTACTTCCGAAAGGATC	453

<i>Tupaia chinensis</i>	AATGATGCCTTGGATAGTGAAGACCGTAGATTGCATGTTAAGAAAGTATTTTCGAAGGATC	453
<i>Eptesicus fuscus</i>	AGTAGCACTAGTGTAGTGTGAGAAACCTTAGGTTGCAGGTGAAAAATGTACTTCCGAAGGATC	453
<i>Myotis brandtii</i>	AGTAGCACTGTAGTGTGAGAAAGCTTAGGTTGCAGGTAAAAATGTACTTCCGAAGGATC	453
<i>Myotis lucifugus</i>	AGTAGCACCGTGTAGTGTGAGAAAGCTTAGGTTGCAGGTAAAAATGTACTTCCGAAGGATC	453
<i>Myotis davidii</i>	AGTAGCACCGTGTAGTGTGAGAAAGCTTAGGTTGCAGGTAAAAATGTACTTCCGAAGGATC	453
<i>Leptonychotes weddellii</i>	AGTGGCATCTGTAGTGTGAGAAACCCCTAGGTTGCAGGTAAAAATGTACTTCTGGAGGATC	450
<i>Odobenus rosmarus</i>	AGTGGCATCTGTAGTGTGAGAAACCCCTAGGTTGCAGGTAAAAATGTACTTCTGGAGGATC	450
<i>Ursus maritimus</i>	-----AGTAGTGAGAAACCCCTAGGTTGCAGGTAAAAATGTACTTCCGGAGGATC	438
<i>Ailuropoda melanoleu</i>	-----AGTAGTGAGAAACCCCTAGGTTGCAGGTAAAAATGTACTTCCGGAGGATC	438
<i>Felis catus</i>	AGTGCATCTTGGGAGTGTGAGAAACCTTAGGTTGCAGGTAAAAATGTACTTCCAAAGGATC	450
<i>Panthera tigris altaica</i>	AGCAGCATCTTGGGAGTGTGAGAAACCTTAGGTTGCAGGTAAAAATGTACTTCCAAAGGATC	450
<i>Mustela putorius</i>	AGCGGTGCTTGTAGTGTGGGAACTCTAGGATGCAGGTAAAAAGTACTTCCGGAGGATC	450
<i>Canis lupus familiaris</i>	AGTGGCATCTGTAGAAAGTGTGAGAAACCCCTAGATTACAGATTAAAAAGTACTTCCAAAGGATC	450
<i>Condylura cristata</i>	AGTGGCTCTTGTAGGAAATTAGAAACCTTAGGTTACAGGTAAAAATGTACTTCCGAAAGAAATC	453
<i>Sorex araneus</i>	AAAAACCTTGTAGTCATGAGAAACCTTGAGGTTACAGGTTAAAAATATATCTTTCAAAGGATC	453
<i>Trichechus manatus latirostris</i>	AGTGGCACCTTAGGTACTGTAGAAACCTTAGATTACAGGTTAAAAATGTACTTTCAAAGGATC	450
<i>Loxodonta africana</i>	AGTGGCTCTTGGGTACGGAGAACTTAGATTACAGGTTAAAAATGTACTTTCAAAGGATC	453
<i>Orycteropus afer</i>	AGTGGCGCTTGTGACACAGAGAACTTAGATTGCAGGTTAAAAATGTACTTTCAAAGGATC	453
<i>Chrysochloris asiatica</i>	AGTGATGCCCTTGGATACTGTAGAAACCTTAGACTGCAGATTAAAAATGTACTTTAGAAAGAAATC	450
<i>Tarsius syrichta</i>	AGTGGTATCTTTTGGTAGTGAGAAACCTTCAGAAACAGACAGGTTAAAAATGTACTTCCAAAAGTTTC	435
<i>Erinaceus europaeus</i>	AGTAACACCTTGTAGTAGTGAGAAACCTTGAGGATGCAGGTTAAAAATGTACTTTCAAAGGATC	447
<i>Dasyopus novemcinctus</i>	AGTGACACCTTGGATAGTGAGACTTTTCAGATTGCAGGTTAAAAACATATTTTCGAAGGATC	450
<i>Oryctolagus cuniculus</i>	AGTGGAGCCTTGGGTAGTGAGATCCTTAGGTTACAGATTAAAAACATATTTTCCTGAGGATC	453
<i>Ochotona princeps</i>	AATGGTGCTGTGGGTAGTGAGAAACCTCAGATTACAGATTAAAAATGTATTTTCACAAGGATC	453
<i>Elephantulus edwardii</i>	AGCCACCCCTAGGTACTGTACAGCTTCATGCTCAAGTGAAGTAAAAATGTACTTCCGGAGGATC	444
<i>Echinops telfairi</i>	AGTCACACCTTGGGTACTGTAGAAACATTCGATTGCAGGTTAAGATGTACTTTAGAAAGGATC	453
<i>Dipodomys ordii</i>	AATGGTGCTGGGTGGTGATTAACCTTAGATTACAGGTTAAAGCATATTTCCGAAGAAATC	444
<i>Ictidomys tridecemlineatus</i>	AGTCATGCCCAAGGGAGTAAAGAACTTAGATTGCAGGTTAAAGCATATTTTCGAAGGATC	438
<i>Nannospalax galili</i>	ACTGGTGCTTGTAGTGTGAGAAACCTTAGATTAAAGATTAAGCATATTTTCAGAAGGATC	450
<i>Cricetulus griseus</i>	AGTGGCGCCTGTAGTGTGCGAGAGCCTTAGGCTACAGATTAAAAATCATATTTTCAGGAGGATC	450
<i>Mesocricetus auratus</i>	AGTGGTGACCGAGCTGTGAGAGGCTTAGGATACAGATTAAAGCCTACTTTCAGGAGGATC	450
<i>Mus musculus</i>	AGTGGGGCTGTAGTGTGCGAGAAACCTTAGGTTACAGATTAAAAATGTATTTTCAGGAGGATC	423
<i>Manis temminckii</i> (5)	GGTGTACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTATTTTCAGGAGGATC	367
<i>Manis tricuspis</i> (4)	GGTGTACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTATTTTCAGGAGGATC	367
<i>Manis tetradactyla</i> (2)	GGTGTACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTATTTTCAGGAGGATC	367
<i>Manis gigantea</i> (1)	GGTGGCACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTATTTTCAGGAGGATC	367
<i>Manis javanica</i> (8)	GGTGGCACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTACTTCTTAATGGATC	470
<i>Manis pentadactyla</i> (1)	GGTGGCACATTGAGTAAGGAAAAAATTTAGATTACAGGTTAAAAATGTACTTCCAAATGGATC	470

* **

<i>Homo sapiens</i>	CATGATTACCTGGAAAAACCAAGACTACAGACACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Pan paniscus</i>	CATGATTACCTGGAAAAACCAAGACGACAGACACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Pan troglodytes</i>	CATGATTACCTGGAAAAACCAAGACGACAGACACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Gorilla gorilla</i>	CATGATTACCTGGAAAAACCAAGACTACAGACACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Nomascus leucogenys</i>	TGTGATTACCTGGAAAAACCAAGACTATGGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Pongo abelii</i>	TGTGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Rhinopithecus roxellana</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Macaca mulatta</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Mandrillus leucophaeus</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Cercocebus atys</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Macaca nemestrina</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Papio anubis</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Macaca fascicularis</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Colobus angolensis</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Chlorocebus sabaeus</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Aotus nancymae</i>	CGTGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	510
<i>Callithrix jacchus</i>	CATGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	510
<i>Pteropus vampyrus</i>	CGTGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Pteropus alecto</i>	CGTGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Galeopterus variegatus</i>	CATGATTACCTGGAAAGGCCAGGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Propithecus coquereli</i>	AATGATTATCTGAAAAACCAAGAAATACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	507
<i>Ceratotherium simum</i>	CTGGATTACCTGGAAATACCAAGAAATACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Equus caballus</i>	CGTGATTACCTGGAAAAACCAAGAAATACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Camelus ferus</i>	CATGATTACCTGGAAAGTCAGGAATACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Vicugna pacos</i>	CATGATTACCTGGAAAGTCAGGAATACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Balaenoptera acutorostrata</i>	CGTGATTACCTGGAAAAACCAAGACTATAGCACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Physeter catodon</i>	CGTGATTATCTGGAAAAACCAAGACTATAGCACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Lipotes vexillifer</i>	CGTGATTACCTGGAAAAACCAAGACTATAGCACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Orcinus orca</i>	CGTGATTACCTGGAAAAACCAAGACTATAGCACCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Sus scrofa</i>	CGTGATTACCTGGAAAAACCAAGACTACAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Ovis canadensis</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Ovis aries musimon</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Pantholops hodgsonii</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Bos taurus</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Bison bison bison</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Bubalus bubalis</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Bos mutus</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Capra hircus</i>	CATGATTACCTGGAAAGGCCAGGACTATAGCAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513
<i>Otolemur garnettii</i>	CGTGAGTACCTGGAAAAACCAAGAAATACAGTAGCTGTGCTCGGCCATTGTCCAAGTAGAA	513

<i>Tupaia chinensis</i>	CGCAATTACCTGGAAAAACCATAAATACAGCAGCTGTGCTCGACCATTTGTCCAAAGTAGAA	513
<i>Eptesicus fuscus</i>	CGGGATTACCTGGAAAAACCCAGGAATACAGCAGCTGTGCTCGACCATTTGTCCAGGTGGAA	513
<i>Myotis brandtii</i>	CGGGATTACCTGGAAAAACCCAGGAACACAGCAGCTGTGCTCGACCATTTGTCCAGGTGGAA	513
<i>Myotis lucifugus</i>	CGGGATTACCTGGAAAAACCCAGGAACACAGCAGCTGTGCTCGACCATTTGTCCAGGTGGAA	513
<i>Myotis davidii</i>	CGGGATTACCTGGAAAAACCCAGGAATACAGCGGTGCGCTGGACCGTTGTCCAGGTGGAA	513
<i>Leptonychotes weddellii</i>	CATGATTACCTGGAAAAACCCAGGAGTACAGCGCTGTGCTGGACCATTTGTTCGAGTAGAA	510
<i>Odobenus rosmarus</i>	CATGATTACCTGGAAAAACCCAGGAGTACAGCGCTGGCTGGACCATTTGTTCGAGTAGAA	510
<i>Ursus maritimus</i>	CATGATTACCTGGAAAAACCCAGGAGTACAGAGCTGTGCTGGACCATTTGTTCGAGTAGAA	498
<i>Ailuropoda melanoleu</i>	CATGATTACCTGGAAAAACCCAGGAGTACAGAGCTGAGCTGAGCTGGACCATTTGTTCGAGTAGAA	498
<i>Felis catus</i>	CATGATTACCTGGAAAAACCCAGGAATACAGCAGCTGTGCTGGACCATTTGTTCGAGTAGAA	510
<i>Panthera tigris altaica</i>	CACGATTACCTGGAAAAACCCAGGAGTACAGCAGCTGTGCTGGACCATTTGTTCGAGTAGAA	510
<i>Mustela putorius</i>	CACAAATTACCTGGAAAAACCCAGGAGTACAAACAGCTGTGCTGGACCATTTGTTCGAGTAGAA	510
<i>Canis lupus familiaris</i>	CATAATTACCTGGAAAAACCCAGGAGCAGCAGCTGTGCTGGACCATTTGTTCGAGTAGAA	510
<i>Condylura cristata</i>	TGGGGTTTCTGGAAAAATGAGAAAAATACAGCAGCTGTGCTGGACCATTCGTCCAAAGTAGAA	513
<i>Sorex araneus</i>	CGTGATTACCTAGAAAAACCCAGGAATACTGCAGCTGTGCTGGGGCCATTTTCATTCGCGGAA	513
<i>Trichechus manatus latirostris</i>	CGTAACTACCTGGAAAAACCCAGAAAAATACAGCAGCTGTGCTGGACCATTTGTCCGAAATAGAA	510
<i>Loxodonta africana</i>	CATAACTACCTGGAAAAACCCAGAAAAATACAGCAGCTGTGCTGGACCATTTGTCCGAGTAGAA	513
<i>Orycteropus afer</i>	CGTAACTACCTGGAAAAACCCAGAAAAATACAGCAGCTGTGCTGGACCATTTGTCCGAGTAGAA	513
<i>Chrysochloris asiatica</i>	CATAACTACCTGAAAAACCCATAAATACAGCAGCTGCGCTGGATTAATCGTCCGAGTAGAA	510
<i>Tarsius syrichta</i>	TGTAATTACCTGGAAAAACCCAGGAATATACCACTGTGCTGGACCATTCGTCCAAAGTAGAA	495
<i>Erinaceus europaeus</i>	CATAACTACCTGGAAAAACCCAGGGTACAGCAACTGTGCTGGACCATTCGTTCGAGTAGAA	507
<i>Dasyapus novemcinctus</i>	CGTGACTACCTGGAAAAACCCAGGAATACAGCAGCTGTGCTGGACCATTTGTCCACGTAGAA	510
<i>Oryctolagus cuniculus</i>	CGTGACTACCTGGAAAGCCAGGACTACAGTATCTGTGCTGGATCATTTGTCCAGGTGGAA	513
<i>Ochotona princeps</i>	CGTGACTACCTGGAAAAACCCAGGAGTACAGTAGCTGCGCTGGACCATTTGTCCAGGTGGAA	513
<i>Elephantulus edwardii</i>	CGTCGGTACCTGGAAAAACCCAGAAACACAGCACTGTGCTGGACCATTTGTCCGAGTTGAA	504
<i>Echinops telfairi</i>	CATAACTACCTGGAAAAACCCAGAAAAATACAGCAGCTGCGCTGGCTCCTTTGTCCGAGTTGAA	513
<i>Dipodomys ordii</i>	CATGCTTACCTGGAAAAACCCAGCTACAGCAGCTGTGCTGGACCATTTGTCCAGGTGGAA	504
<i>Ictidomys tridecemlineatus</i>	CATAATTACCTGGAAAAACCCAGCAGACAGCAGTTGTGCTGGACCATTTGTTCATGCAGAA	498
<i>Nannospalax galili</i>	CATGATTACCTTGGAAAAATCATGGGTACAGCAGCTGTGCTGGATCATTTGTTCAGGTAGAA	510
<i>Cricetulus griseus</i>	CACGACTACTTGGAAAAATCAGAGGTACAGCAGCTGTGCTGGATCGTTGTTCAGGTAGAA	510
<i>Mesocricetus auratus</i>	CACGACTACTTGGAAAAACCCAGAGGTACAGCAGCTGTGCTGGATTAATTTGTTCAGGTAGAA	510
<i>Mus musculus</i>	CACGATTACTTGGAAAAACCCAGGAATACAGCAGCTGTGCTGGATCATTTGTTCAGACAGAA	483
<i>Manis temminckii</i> (5)	CATGATTATCTGGAAAAACCCAGGAACACAGCAGCTGTGCTGGACCATTTGTCA-----	419
<i>Manis tricuspis</i> (4)	CATGATTATCTGGAAAAACCCAGGAACACAGCAGCTGTGCTGGACCATTTGTCA-----	419
<i>Manis tetradactyla</i> (2)	CATGATTATCTGGAAAAACCCAGGAACACAGCAGCTGTGCTGGACCATTTGTCA-----	377
<i>Manis gigantea</i> (1)	CATGATTATCTGGAAAAACCCAGGAACACAGCAGCTGTGCTGGACCATTTGTCA-----	419
<i>Manis javanica</i> (8)	CATGATTATCTGGAAAAACCCAGGAATACAGCAGCTGTGCTGGACCATTTGTCA-----	530
<i>Manis pentadactyla</i> (1)	CATGATTATCTGGAAAAACCCAGGAATACAGCAGCTGTGCTGGACCATTTGTCA-----	530

*

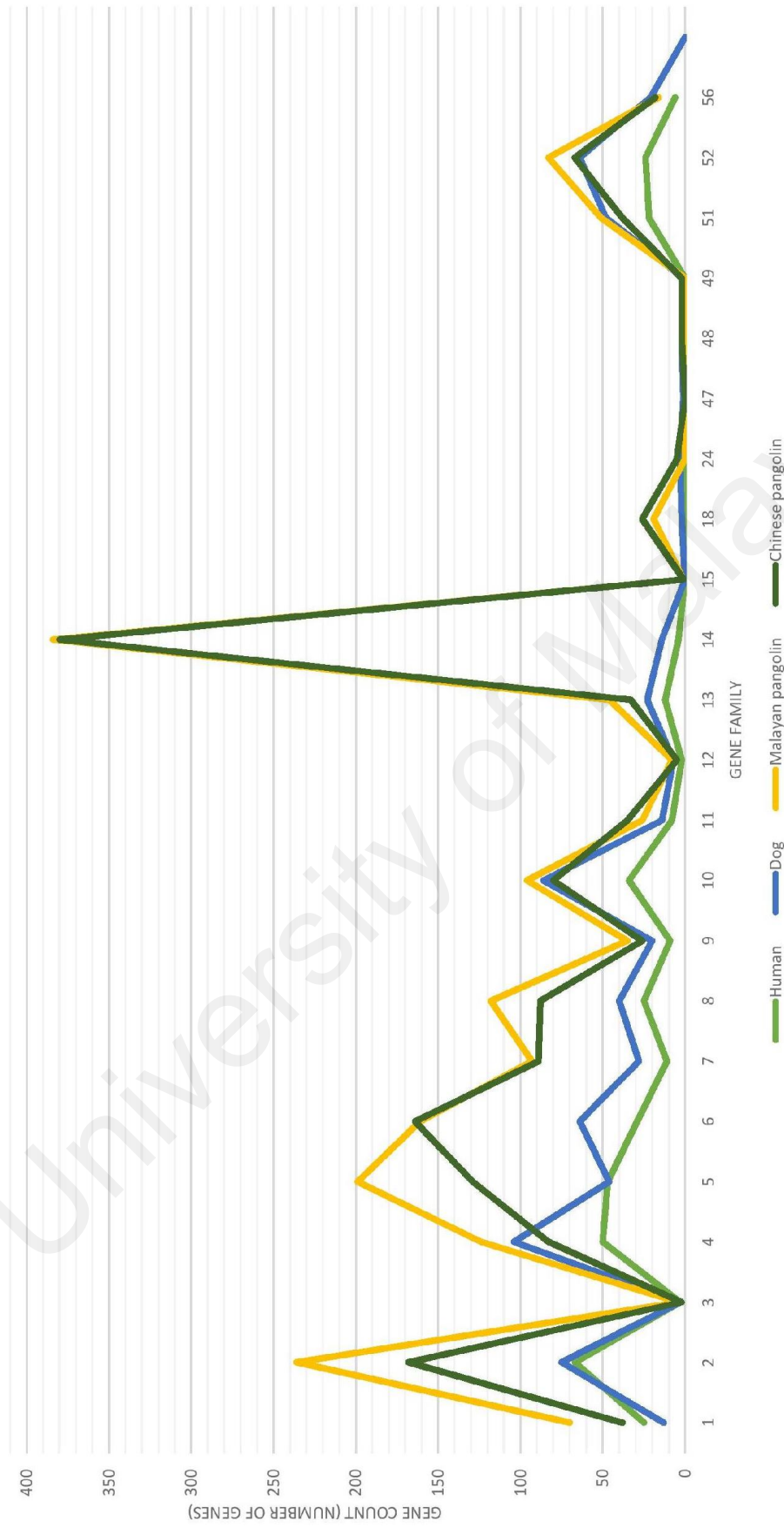
<i>Homo sapiens</i>	ATCAGCCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAAGA	573
<i>Pan paniscus</i>	ATCAGCCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAAGA	573
<i>Pan troglodytes</i>	ATCAGCCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAAGA	573
<i>Gorilla gorilla</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAAGA	573
<i>Nomascus leucogenys</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGA--A	571
<i>Pongo abelii</i>	ATCAACCGGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAACA	573
<i>Rhinopithecus roxellana</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAACA	573
<i>Macaca mulatta</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAACTGAGCAAAACAAGGAACA	573
<i>Mandrillus leucophaeus</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCATAGAAAAAGCTGAGAAAAACAAGGAACA	573
<i>Cercocebus atys</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	573
<i>Macaca nemestrina</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	573
<i>Papio anubis</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	573
<i>Macaca fascicularis</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGA--A	571
<i>Colobus angolensis</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	573
<i>Chlorocebus sabaeus</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	573
<i>Aotus nancymae</i>	ATCAACTGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAACA	570
<i>Callithrix jacchus</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAATCAAGGAAAA	570
<i>Pteropus vampyrus</i>	ATCAACCGGTGTCCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAATG	573
<i>Pteropus alecto</i>	ATCAACCGGTGTCCTTTCTTTTGTGTTCTCAGTCTCACAGAAAAAGCTGAGCAAAACAAGGAATG	573
<i>Galeopterus variegatus</i>	ATCAACAGGTGTCGTCTTTCTTTTGTGTTCTCAGTCTCATAGAAAAAGTTGAACAAACAAGGAATG	573
<i>Propithecus coquereli</i>	ATCAGTCAGTGTCTGTCCTTTGTTGTTCCAGTCTCAGACTCAGAGGAAAGCTGAGCAAAACAAGGAATG	567
<i>Ceratotherium simum</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCAGAGGAAAGCTGAGCAAAACAAGGAACG	573
<i>Equus caballus</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCAACTCAGAGGAAAGCTGAGCAAAACAAGGAATG	573
<i>Camelus ferus</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTATGATCCAACTCAGAGGAAAGCTGAGCAAAACAAGGAATG	573
<i>Vicugna pacos</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTGATCCAACTCAGAGGAAAGCTGAGCAAAACAAGGAATG	571
<i>Balaenoptera acutorostrata</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCAGAGGAAAGCTGAGCAAAACAAGGAA--	571
<i>Physeter catodon</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCAGAGGAAAGCTGAGCAAAACAAGGAA--	571
<i>Lipotes vexillifer</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCCGACTCAGAGGAAAGCTGAGCAAAACAAGGAA--	571
<i>Orcinus orca</i>	ATCAACCGATGTCGTCTTTCTTTTGTGTTCCAACTCAGAGGAAAGCTGAGCAAAACAAGGAA--	571
<i>Sus scrofa</i>	ATCAACCGGTGTCCTATTCTTTTGTGTTCCAACTCAGGGGAAAAAACTGAGCAAAACAAGGAATG	573
<i>Ovis canadensis</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGGCATG	573
<i>Ovis aries musimon</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGGCATG	573
<i>Pantholops hodgsonii</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCAACTCACAAGAAAAAGCTGAGTGAACAGGG--CA	571
<i>Bos taurus</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGGCATG	573
<i>Bison bison bison</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGGCATG	573
<i>Bubalus bubalis</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCAACTCACAAGAAAAAGCTGAGTGAACAGGGCGTG	573
<i>Bos mutus</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGG--	569
<i>Capra hircus</i>	ATCAACCGGTGTCGTCTTTCTTTTGTGTTCCGACTCACAAGAAAAAGCTGAGTGAACAGGG--	569
<i>Otolemur garnettii</i>	ATCAATCGGTGTCATTGTTTGTGTTCTCAGACTCACAGAAAAAAAGCTGAGCAAAACAAGGAATG	573

<i>Tupaia chinensis</i>	ATCAAGAGATGTCTGTTCTTAGTGCTCAGACTTATAGAAAGTTGAGCACAGAAAGGAATG	573
<i>Eptesicus fuscus</i>	GTCAACCGGTGTCTATTTCTTTGTTCCCACTCACAGGAAAACTGAGCAAAACAAGAAATG	573
<i>Myotis brandtii</i>	GTCAACCGGTGTCTACTCTTTGTTGTTCCCACTCACAGGAAAACTGAGCCAAACAAGAAATA	573
<i>Myotis lucifugus</i>	GTCAACCGGTGTCTACTCTTTGTTGTTCCCACTCACAGGAAAACTGAGCCAAACAAGAAATG	573
<i>Myotis davidii</i>	GTCAACCGGTGTCTATTTCTTTGTTGTTCCCACTCACAGGAAAACTGAGGCAACAAGAAAAAG	573
<i>Leptonychotes weddellii</i>	ATCAACCGGTGTCTGTTCTTTGCTTCCAACTGGTAAGAAAAAGCTAAAGAAAAACAAGAAATG	570
<i>Odobenus rosmarus</i>	ATCAACCGGTGTCTGTTCTTTGCTTCCAACTGATAAGAAAAAGCTAAAGAAAAACAAGAAATG	570
<i>Ursus maritimus</i>	ATCAGCCAGTGTCTGTTCTTTGCTTCCAACTGATAAGAGAGCTAAAGCAAAACAAGGAATG	558
<i>Ailuropoda melanoleu</i>	ATCAGCCAGTGTCTGTTCTTTGCTTCCAACTGATAAGAGAGCTAAAGCAAAACAAGGAATG	558
<i>Felis catus</i>	ATCAACCGGTGTCTCTTTGTCACCTCCAGCTGATAAGAAAAAGATAAGCAAAACGAGGAATG	570
<i>Panthera tigris altaica</i>	ATCAACCGGTGTCTCTTTGTCACCTCCAGCTGATAAGAAAAAGATAAGCAAAACGAGGAATG	570
<i>Mustela putorius</i>	ATCAACCGGTGTCTCTTTGTCATTCAGCTAGTAGTAAGAAAAAGCTAAAGCAAAATAAGAAAAACA	570
<i>Canis lupus familiaris</i>	ATCAACCAATGTCTGTTCTTTGCAATTCCAACTGATAAGAAAACTAAAGCAAAATAA-----	564
<i>Condylura cristata</i>	ATCAAGCGGATTCTGTTCTTTGTTGTTTCAGACTCACAGAAAAAACTGAGTGAGCAAGGAAGG	573
<i>Sorex araneus</i>	ATTCACCGATGTCTCTTTGTTGTTTCAGACTCACAGGCAATGCTGAGCAAAACAAGAAATG	573
<i>Trichechus manatus latirostris</i>	ATCAATTCGGTGCTGTTCTTTGTTGTTTTCAGACTCACGGGAAAAAGCTGAGCAAAATAA-----	564
<i>Loxodonta africana</i>	ATCAATTCGGTGCTGTTCTTTGTTGTTTTCAGACTCACAGGAAAAAGCTGAGCCAAATACAGAAATG	573
<i>Orycteropus afer</i>	ATCAATTCGGTGCTGTTCTTTGTTGTTTTCAGACTCACAGCAAAAGCTG--AGCAAAATGATGAATG	571
<i>Chrysochloris asiatica</i>	ATTATACGGTGCTGTTCTTTGTTGTTTTCAGACTCACAGCAAAAGCTGAGCAAAATAAGGAATG	570
<i>Tarsius syrichta</i>	ATCAACAGGTGTTGTTCCCTGTTGTTTCAGACTCATAGAAAAAGCTGAGCAAAACAAGAAATG	555
<i>Erinaceus europaeus</i>	ATCAACCGGTGTTTACTCTTCTCATGCTCAGACTCACACGAAAGCTAAAGCAAAACAAGGAA--	565
<i>Dasyapus novemcinctus</i>	ATCAACCGGTGCTGTTCTTTGTTTTCAGACTCACAGGA--AAGCTGA-----	555
<i>Oryctolagus cuniculus</i>	ATCAACCGGATGCTCTCTTCTTTGTTGTTTCAGACTCACAGAAATGCTGACAAAAAGAGCTAGA	573
<i>Ochotona princeps</i>	ATCAACCGGTGCTCTTCTTCTTTGTTGTTTCAGACTCACAGGAAAGCTCAGC-----	561
<i>Elephantulus edwardii</i>	GTCAATTCGATGCCTGTTCTTCTTTGTTGTTTCAGACTCACAGGAAAGCTGAGCAAA-----	564
<i>Echinops telfairi</i>	ATTAATCGGTGCCTGTTCTTCTTTGTTGTTTCAGACTCACAGGAAAGCTGAGCAAA-----	563
<i>Dipodomys ordii</i>	GTTAAACCGGTGCTGTTCTTTGTTGTTTTCAGACTCACACTTACAAGAAAGCTGAGGAAA-----	555
<i>Ictidomys tridecemlineatus</i>	ATCAACCGGTGCTGTTCTTCTTTGTTGTTTTCAGACTCATCAGACTCATCAGACAA-----	540
<i>Nannospalax galili</i>	ATCAACAGGTGCTGTTCTTCTTTGTTGTTTCAGACTGACAGGATGGCTGAGCAAAACAAGAAAAACA	570
<i>Cricetulus griseus</i>	ATCAACCGGTGATGTTCTTCTTTGTTGTTTCAGGCTAAACCGCATGGCTGAACAAACAAGAAAAACA	570
<i>Mesocricetus auratus</i>	ATCAACCGGTGATGTTCTTCTTTGTTGTTTCAGGCTCACTGGATGGCTGAGCAAAACAAGAAAAACA	570
<i>Mus musculus</i>	ATCCACCGGTGATGTTCTTCTTTGTTGTTTCAGTTTCAACAACATGGCTGAGCAGACAAGACCCA	543
<i>Manis temminckii</i> (5)	-----	419
<i>Manis tricuspis</i> (4)	-----	419
<i>Manis tetradactyla</i> (2)	-----	377
<i>Manis gigantea</i> (1)	-----	419
<i>Manis javanica</i> (8)	AGTG-----TCTGTCCTTTGTTGTTCCAACTCACAGGAAAAAGCTGAGCAAAACGAGGAGTG	583
<i>Manis pentadactyla</i> (1)	AGTG-----TCTGTCCTTTGTTGTTCCAACTCACAGGAAAAAGCTGAGCAAAACGAGGAGTG	583

<i>Homo sapiens</i>	--CCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCGAGGTAG-	627
<i>Pan paniscus</i>	--CCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCAAGGTAG-	627
<i>Pan troglodytes</i>	--CCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCAAGGTAG-	627
<i>Gorilla gorilla</i>	--CCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCAAGGTAG-	627
<i>Nomascus leucogenys</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCAAGGTAG-	627
<i>Pongo abelii</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTTAGAAGCCCCAAGGT---	627
<i>Rhinopithecus roxellana</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGG---	627
<i>Macaca mulatta</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGGTAG	630
<i>Mandrillus leucophaeus</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGGTAG	630
<i>Cercocebus atys</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGG---	627
<i>Macaca nemestrina</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGG---	627
<i>Papio anubis</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGG---	627
<i>Macaca fascicularis</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGG---	625
<i>Colobus angolensis</i>	GACCCCTTGAAACGACATGAAAGCAAG-AGCTTACTACAGAGTTGTAGAAGCCCCAAGGTAG	630
<i>Chlorocebus sabaeus</i>	GACCCCTTGAAACGACATGAAAGCAAG-AACTTAATACAGAGTTGTAAAAGCTCAAGG---	627
<i>Aotus nancymae</i>	GACCCCTTGAAAGACATGGAGCAAA-AGCTTACTACAGAGTCTAGAAGCCCCAAGGTAG-	626
<i>Callithrix jacchus</i>	GACCCCTTGAAAGACATGGAAACAAA-AGCTTACCACAGAGTCTAGAAGCCCCAAGGTAG-	626
<i>Pteropus vampyrus</i>	GACCCCTTGAAACCACTGTGG-AGCAAGAGGCAACCACAGACTTTAAAAGC-----	620
<i>Pteropus alecto</i>	GACCCCTTGAAACCACTGTGG-AGCAAGAGGCAACCACAGACTTTAAAAGC-----	620
<i>Galeopterus variegatus</i>	CCCTTGAA--TGATGTGAAACAAAG-AGCTGACTACAGGCTTTAAAAGCACAAAGGTAG-	627
<i>Propithecus coquereli</i>	ACCTTGAA--TGACATGGAGCAAG-AAACCATTCACAGACTTTAAAAGC-----	612
<i>Ceratotherium simum</i>	GACTCTTGA-----	582
<i>Equus caballus</i>	GACCCCTTGA-----	582
<i>Camelus ferus</i>	GATCCTTGA-----	582
<i>Vicugna pacos</i>	GATCCTTGA-----	582
<i>Balaenoptera acutorostrata</i>	-----	571
<i>Physeter catodon</i>	-----	571
<i>Lipotes vexillifer</i>	-----	571
<i>Orcinus orca</i>	-----	571
<i>Sus scrofa</i>	GAAACCTTGA-----	582
<i>Ovis canadensis</i>	GAAACTTGAAACCATTTGA-AGGAAAGAGCCAACTGCAGGCTTTAAAAGC-----	620
<i>Ovis aries musimon</i>	GAAACTTGAAACCATTTGA-AGGAAAGAGCCAACTGCAGGCTTTAAAAGC-----	620
<i>Pantholops hodgsonii</i>	TGGAATTGAAACCATTTGA-AGGAAAGAGCCAACTGCAGGCTTTAAAAGC-----	618
<i>Bos taurus</i>	GAAACTTGAAACCATTTGA-AGGAAACAGCCGACTGCAGGCTTTAAAAGC-----	620
<i>Bison bison bison</i>	GAAACTTGAAACCATTTGA-AGGAAACAGCCGACTGCAGGCTTTAAAAGC-----	620
<i>Bubalus bubalis</i>	GAAACTTGAAACCATTTGA-AGGAAAGAGCCGACTGCAGGCTTTAAAAGC-----	609
<i>Bos mutus</i>	-----	569
<i>Capra hircus</i>	-----	569
<i>Otolemur garnettii</i>	AGCCCTTGA-----	582

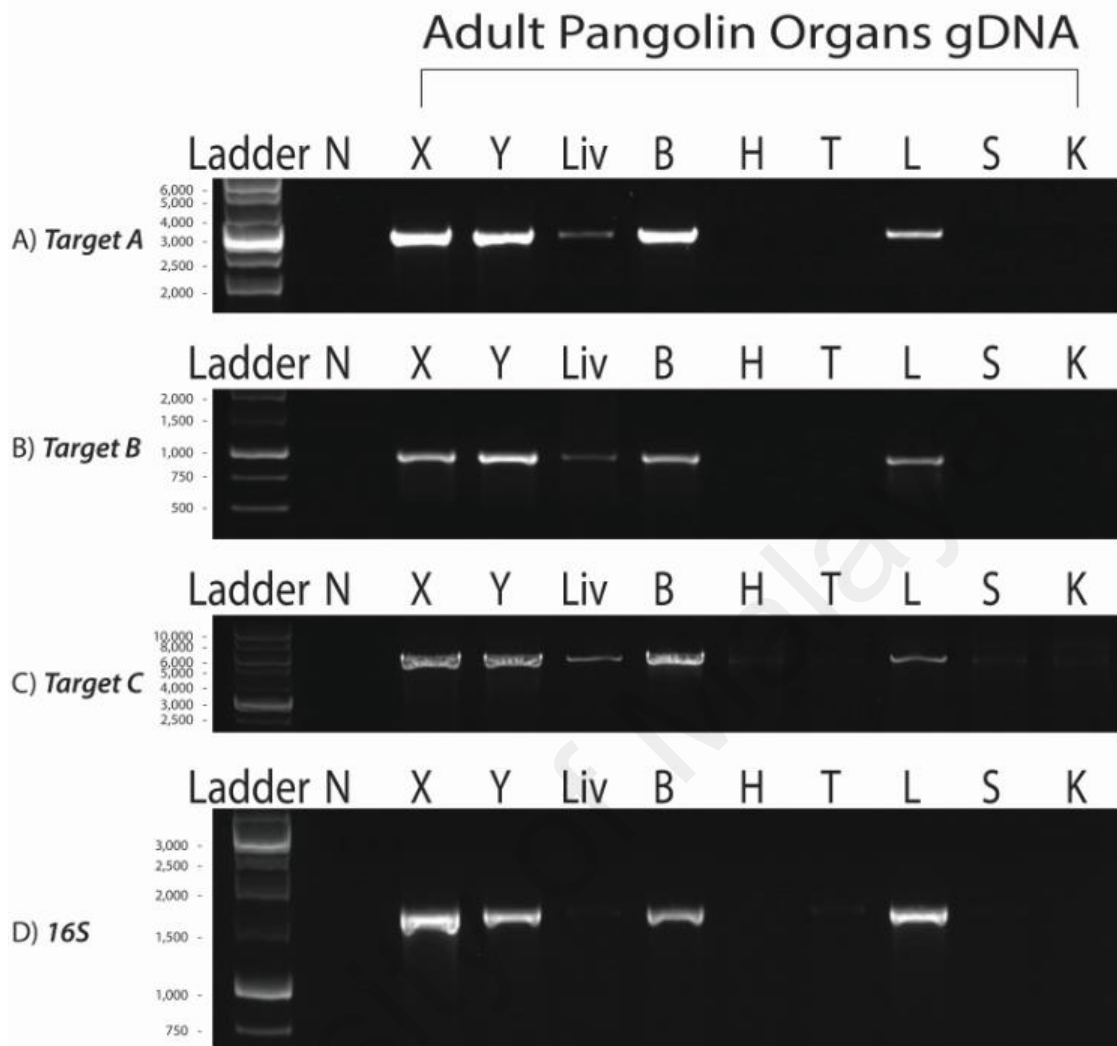
<i>Tupaia chinensis</i>	GACCTTGAATGATGACATGGAGCAAGTGCTGACTATGGGCTTTAAAAAGC-----	621
<i>Eptesicus fuscus</i>	GACCCCTGA-----	582
<i>Myotis brandtii</i>	GACCCCTTGA-----	582
<i>Myotis lucifugus</i>	GACCCCTTGA-----	582
<i>Myotis davidii</i>	GACCCCTTGA-----	582
<i>Leptonychotes weddellii</i>	GATTCCTTGAAGAATGTAGAGCAAGAGCCAACTGCAGGCTTTAAAAAGCATAGG-GTAG	627
<i>Odobenus rosmarus</i>	GATTCCTTGAAGAATGTAGAGCAAGAGCCAACTGCAGGCTTTAAAAAGCATAGG-GTAG	627
<i>Ursus maritimus</i>	GATTCCTTCAAGAATGTGGAACAAGAGCCAACTGCAGGCTTTAAAAAGCATAGG-GTAG	615
<i>Ailuropoda melanoleu</i>	GATTCNTGAAGAATGTGGAACAAGAGCCAACTGCAGGCTTTAAAAAGCATAGG-GTAG	615
<i>Felis catus</i>	CATTCCTCGAAAGAATGTGGAGCATGAGCCAAAGGGCAGACTTTAGAAAGCATAGG-GTAG	627
<i>Panthera tigris altaica</i>	CATTCCTCGAAAGAATGAGGAGCATGAGCCGAGGGCAGACTTTAGAAAGCATAGG-GTAG	627
<i>Mustela putorius</i>	GATTCC-----TTCCAAAGTGAGCAAGAGCCATCTCCAGGCTTTAAAAAGCATAGG-GTAG	624
<i>Canis lupus familiaris</i>	GACCTT-----	564
<i>Condylura cristata</i>	TACCTTGA-----	579
<i>Sorex araneus</i>	-----	582
<i>Trichechus manatus latirostris</i>	-----	564
<i>Loxodonta africana</i>	GACCTTGA-----	582
<i>Orycteropus afer</i>	GACCTTGAACCAGATGAAGCATGATCT-----	599
<i>Chrysochloris asiatica</i>	GACCTTGA-----	579
<i>Tarsius syrichta</i>	GACCTTGA-----	564
<i>Erinaceus europaeus</i>	-----	565
<i>Dasytus novemcinctus</i>	-----	555
<i>Oryctolagus cuniculus</i>	TCCTTGAACAGTGTGAGCAAG--AGCTGACTACAG-----	607
<i>Ochotona princeps</i>	-----	561
<i>Elephantulus edwardii</i>	CCCTCAAAACCCGATGAAGCGTG--ATCTGACCACAGCCGTTAAAAAGCATAGGGGTAG--	618
<i>Echinops telfairi</i>	-----	563
<i>Dipodomys ordii</i>	-----	555
<i>Ictidomys tridecemlineatus</i>	-----	540
<i>Nannospalax galili</i>	GACCCCTTGA-----	579
<i>Cricetulus griseus</i>	GACCCCTTGA-----	579
<i>Mesocricetus auratus</i>	GACCCCTTGA-----	579
<i>Mus musculus</i>	GACCCCTTGAACACTGAGAAAGCAAGAGC-----	570
<i>Manis temminckii</i> (5)	-----	419
<i>Manis tricuspis</i> (4)	-----	419
<i>Manis tetradactyla</i> (2)	-----	377
<i>Manis gigantea</i> (1)	-----	419
<i>Manis javanica</i> (8)	GATCCTGGAACCACTGG-AGCAGGAGCCAACTGCAGGCTTTAAAAAGCATAGGGGTAG--	639
<i>Manis pentadactyla</i> (1)	GATCCTGGAACCACTGG-AGCAGGAGCCAACTGCAGGCTTTAAAAAGCATAGGGGTAG--	639

Appendix I: Gene distribution of Olfactory Receptor (OR) gene families



Appendix I shows 23 Olfactory Receptor (OR) gene families observed in both Malayan pangolin and Chinese pangolin, significant gene expansion in OR2, OR5, OR6, OR8, OR10, OR14 and OR18 gene families observed.

Appendix J: PCR assays to detect the presence of *Burkholderia* sp.



Target A, target B, and target C are three different primers design unique to amplify *Burkholderia* sp. sequence, together with 16S universal bacteria primer confirmed the presence of the bacterial species in the pangolin cerebrum, cerebellum, blood, and lung. (X-cerebrum, Y-cerebellum, Liv-Liver, B-blood, H-heart, T-thymus, L-lung, S-spleen, K-kidney)