GENETIC DIVERSITY OF THE *RATTUS* COMPLEX (RODENTIA: MURIDAE) IN KWAZULU-NATAL

by

DEENADAYALAN NAIR

GENETIC DIVERSITY OF THE *RATTUS* COMPLEX (RODENTIA: MURIDAE) IN KWAZULU-NATAL

by

DEENADAYALAN NAIR

Submitted in fulfilment of the academic requirements for the degree of

Master of Science

in the

School of Biological & Conservation Sciences
University of KwaZulu-Natal
Durban
South Africa

December 2010

As the candidate's supervisor I have/have not approved this dissertation for submission.

Signed:_______________ Name:_____________________ Date:___________

ABSTRACT

The rodent genus *Rattus* is considered to be the single largest genus of mammals in the world. One species of *Rattus* is usually more dominant than another within a specific geographical area; however within the province of KwaZulu-Natal South Africa current observations indicate that Norway rats (*R. norvegicus*), black rats (*R. rattus*) and the indistinct Asian house rat (*R. tanezumi*) exist sympatrically.

DNA sequencing of the cytochrome *b* and D-loop regions of the mitochondrion were used in conjunction with karyotyping of bone marrow and tissue culture cells to analyse the genetic diversity of selected *Rattus* populations from KwaZulu-Natal.

Comparison of sequence data obtained during the study to reference sequences obtained from the NCBI GenBank revealed three well-supported monophyletic groups in maximum parsimony and Bayesian analyses. These three monophyletic groups indicated the existence of three species of the *Rattus* complex within KwaZulu-Natal, namely *Rattus rattus*, *Rattus norvegicus* and *Rattus tanezumi*.

Analysis of cytochrome *b* sequence data revealed the presence of 6, 3 and 2 haplotypes in 20 *R. norvegicus*, 8 *R. rattus* and 5 *R. tanezumi* specimens, respectively. The *R. norvegicus* haplotypes were separated from *R. rattus* and *R. tanezumi* haplotypes by 60 mutational steps, while *R. rattus* haplotypes were separated from *R. tanezumi* haplotypes by 24 mutational steps. Analysis of D-loop sequence data revealed the presence of 6, 2 and 1 haplotypes in 14 *R. norvegicus*, 4 *R. rattus* and 3 *R. tanezumi* specimens, respectively. *R. norvegicus* haplotypes were separated from *R. rattus* and *R. tanezumi* haplotypes by 15 mutational steps, while *R. rattus* haplotypes were separated from *R. tanezumi* haplotypes by 11 mutational steps.

Karyotype analysis of specimens revealed that: (1) *R. rattus* specimens sampled presented with a karyotype of either 2n = 38 or 2n = 40; (2) *R. tanezumi* specimens sampled presented with a karyotype of 2n = 42 and (3) *R. norvegicus* specimens sampled presented with a karyotype of 2n = 42 which was very distinct from that of *R. tanezumi*.

PREFACE

The experimental work described in this dissertation was carried out in the School of Biological and Conservation Sciences, Department of Biology, University of KwaZulu-Natal, Durban, from August 2005 to December 2006, under the supervision of Prof. J. M. Lamb, Dr. G. Contraffato and Prof. P.J. Taylor.

This study represents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

______________________ ________________

D. Nair Date

DECLARATION 1 – PLAGARISM

I, Deenadayalan Nair, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis in the References sections.

Signed:________________

DECLARATION 2 – PUBLICATIONS

Details of contribution to publications that form part and/or include research presented in this thesis:

Publication 1:

Armanda DS Bastos, Deenadayalan Nair, Peter J Taylor, Helene Brettschneider, Frikkie Kirsten, Elmarie Mostert, Emil von Maltitz, Jennifer M Lamb, Pim van Hooft, Steven R Belmain, Giancarlo Contrafatto, Sarah Downs and Christina T Chimimba (2011). Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive *Rattus* congeners in South Africa. *BMC Genetics* **12**. **DOI**: **10.1186/147-2156-12-26**

Signed:________________

Table of Contents

ABSTRACT I

PREFACE II

DECLARATION 1 – PLAGARISM III

DECLARATION 2 – PUBLICATIONS IV

ACKNOWLEDGEMENTS VII

CHAPTER 1 : INTRODUCTION 8

1.1 Background 8
1.2 The biology, taxonomy and distribution of the *Rattus* complex in South Africa 9
1.2.1 General Review 9
1.2.2 The Norway Rat (*Rattus norvegicus*) 10
1.2.3 The Black Rat (*Rattus rattus*) 11
1.2.4 The Asian House Rat (*Rattus tanezumi*) 12
1.2.5 Taxonomy 13
1.3 Review of species concepts 14
1.4 Molecular methods used in assessing genetic diversity 16
1.5 Phylogeography 18
1.6 Data analysis 19
1.6.1 Introduction to phylogenetic analyses 19
1.6.2 Character – state methods 19
1.6.3 Genetic distance models 20
1.6.4 Software packages used to analyze data for this study 21
1.7 Cytogenetic analyses 22
1.9 Aims, objectives and hypotheses of this study 23

CHAPTER 2 : MATERIALS AND METHODS 25

2.1 Sample collection and storage 25
2.2 Molecular analyses 28
2.2.1 DNA isolation, extraction and quantification 28
2.2.2 Polymerase Chain Reaction (PCR) amplification 28
2.2.3 DNA sequencing 32
2.2.4 Sequence editing and alignment 32
2.2.5 Sequence data analyses 36
2.3 Cytogenetic Analyses 37

2.3.1 Preparation of slides from bone marrow 37
2.3.2 Skin fibroblast culture cells 38
2.3.3 Banding techniques 40
2.3.3.1 G-Banding 40
2.3.3.2 C-Banding 40

CHAPTER 3 : RESULTS 42
3.1 Molecular Analyses 42
3.1.1 Haplotype Analyses 42
3.1.2 Genetic Distance Analyses 50
3.1.3 Phylogenetic Analyses 56
3.2 Cytogenetic Analyses 64

CHAPTER 4 : DISCUSSION 68
4.1 Molecular Analyses 68
4.1.1 *Rattus norvegicus* 68
4.1.2 *Rattus rattus* 70
4.1.2.1 *Rattus rattus* 70
4.1.2.2 *Rattus tanezumi* 71
4.2 Cytogenetic Analyses 71
4.3 Conclusion 74

REFERENCES 75

APPENDICES 82
APPENDIX 1: Preparation of 1.2% agarose gels 82
APPENDIX 2: Preparation of DMEM 82
APPENDIX 3: List of Published sequences 83
APPENDIX 4: 1146 nucleotide Cytochrome b consensus sequences 84
APPENDIX 5: 676 nucleotide Cytochrome b consensus sequences 94
APPENDIX 6: D-loop consensus sequences 102

ACKNOWLEDGEMENTS

I wish to thank Dr. J.M. Lamb and Dr. P. J. Taylor for allowing me the opportunity to conduct research into the topic of this study. I would like to express my sincerest gratitude towards my supervisors, Dr. Lamb for her constant guidance and assistance throughout the course of this study, Dr. P. J. Taylor for conducting field collections, Dr. G. Contraffato for his aid and supervision of my karyotyping exercises, and to Dr. A. D. Bastos for her aid in the sequencing of some of my sequences and her suggestions with regards to choice of primers and temperature profiles for the conduction of PCR for this study. I would also like to express my sincerest gratitude towards Mr. S. Moodley and his vector control team from the KwaZulu-Natal Department of Health whom aided with the capture of live specimens for this study.

I would also like to acknowledge my parents, without whose moral (and financial) support I would never have been able to complete any of my studies thus far and I would like to take this opportunity to thank them for their constant guidance over the years.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

CHAPTER 1
INTRODUCTION

1.1 Background

The rodent genus *Rattus* Fischer de Waldheim 1803, consisting of at least 61 valid species, is considered to be the single largest genus of mammals. Members of this genus are observed to be widely distributed throughout the world, thus earning them the reputation of being successful colonisers. Species belonging to the genus *Rattus* occur in all types of habitats including sandy desert (De Graaff, 1981; Aplin *et al.*, 2003; Chinen *et al.*, 2005, Yiğit *et al.*, 1998). Members of this ubiquitous genus are perhaps most known as pests, for their involvement in the spread of diseases such as plague during the 5th and subsequent centuries in Europe (Aplin *et al.*, 2003) and currently as reservoirs for at least 61 different emerging diseases including Hantavirus in America (Aplin *et al.*, 2003; Mills e*t al.*, 1999) and plague, leptospirosis and toxoplasmosis in Africa and worldwide (Begon, 2003; Singelton *et al.*, 2003; Taylor *et al.*, 2008; Meerburg *et al.*, 2009). Only 15 *Rattus* species are known to be significant agricultural pests, of which five are true commensals, while three species now exist solely within the confines of human environments (Aplin *et al.*, 2003).

Rats are also of considerable benefit to humans in the medical field where they are used as model organisms in the study of various diseases. The brown or Norway rat (*Rattus norvegicus* (Berkenhout, 1769)) has been much studied in relation to the formulation of cures for diseases, the workings of the mammalian brain and the effects of various substances on the mammalian body. In fact the Norway rat has been used as a model for learning about human physiology and disease from as early as the 1800s (Twigg, 1975).

Even though wild *Rattus* populations may be beneficial to humans, their negative impacts are thought to far outweigh the benefits that they provide (Twig, 1975; Aplin *et al.*, 2003). Five percent of *Rattus* species cause considerable damage to agricultural crops, thus creating major problems for farmers, as their crop yields are reduced (Singleton, 2003; Brown *et al.*, 2003). Aside from the damage caused to crops, introduced *Rattus* populations are also known to negatively affect the conservation status and population sizes of indigenous rodents (Goodman, 1995). Within urban environments, rodents, especially those belonging to the genera *Rattus* and *Mus*, are considered as significant pest species that normally co-inhabit human environments,

preferring to live within human homes whilst living off discarded human waste material (Singleton *et al.*, 2003; Begon, 2003).

Another major concern related to rodents that are capable of co-inhabiting human environments is the potential these animals have to behave as zoonotic vectors capable of actively transmitting disease, especially since members of the genus *Rattus* are known to serve as hosts to approximately 60 zoonotic diseases (Twigg, 1975; De Graaff, 1981; Aplin *et al.*, 2003; Belmain, 2003; Meerburg *et al*., 2009). In South Africa, rodents belonging to the genus *Rattus* are known to transmit diseases such as leptospirosis, plague and toxoplasmosis, with leptospirosis being of major concern, as it is viewed as an emerging infectious disease (Leclerc-Madlala and Janowski, 2004; Taylor *et al.* 2008). Plague is also of concern, as transmission of the plague bacterium from rats to humans via fleas usually results in death if treatment is not administered (Aplin *et al.*, 2003; Twigg, 1975; Duplantier *et al.*, 2003; Duplantier *et al.*, 2005). Toxoplasmosis is only fatal to immuno-compromised persons; within the context of South Africa, this is a major problem owing to the high proportion of immune-compromised HIV positive people in the population (Berdoy *et al.*, 2000; Leclerc-Madlala and Janowski, 2004).

The purpose of this study is to derive some insight into the genetic diversity of members of the *Rattus* complex within the province of KwaZulu–Natal, as knowledge of the species present and their distribution is a necessary first step towards controlling rodent–derived problems in this society.

1.2 The biology, taxonomy and distribution of the *Rattus* complex in South Africa

1.2.1 General Review

The spread of the black (*Rattus rattus* (Linnaeus, 1758)) and Norway (*Rattus norvegicus* (Berkenhout, 1769)) rats, which are commensal organisms by nature, has been mainly attributed to recent human migrations, trade and the higher acclimatisation rates of these species compared with other species in this genus (De Graaff, 1981; Aplin *et al.*, 2003; Chinen *et al.*, 2005, Yiğit *et al.*, 1998). Throughout the world one species of *Rattus* is usually more dominant than another within a specific geographical area (De Graaff, 1981; Aplin *et al.*, 2003). Within the province of KwaZulu-Natal current observations indicate that a *Rattus* complex comprising distinct

populations of Norway rats (*Rattus norvegicus*) and black rats (*Rattus rattus*) exists. Prior to this study, the presence of *Rattus tanezumi* Temminck, 1844 was independently reported in Limpopo and western KwaZulu-Natal provinces (Bastos *et al.*, 2005; Taylor *et al.*, 2008).

1.2.2 The Norway Rat (*Rattus norvegicus*)

Rattus norvegicus, more commonly called the brown or Norway rat, is very different in character and to some extent in appearance from the common house rat, *Rattus rattus*. The former attains a much larger size (captured specimens have been recorded at total body length approximately 200 – 400 mm) and can always be distinguished by its stouter body as well as by the shortness of its ears, length of tail (shorter than or approximately equal to length of head and body) and coarser fur. The most distinguishing feature of this species is the colour of its fur, which is usually reddish-brown to a grayish-brown on the back, often darker medially, and is woolly and harsher to touch than that of *R. rattus* (De Graaff, 1981).

R. norvegicus has a high reproductive rate. Wild populations, which have an average lifespan of two years, are able to breed and produce young throughout the year, with a decline in the rate of breeding and offspring production only occurring during cold seasons. The gestation period of *R. norvegicus* is 22 - 24 days; litter size varies from two to 14 individuals. These rats have the capability to breed up to seven times a year, as the young that are produced from a single breeding phase are independent from the adults at around four to five weeks, and gain sexual maturity at around three to four months (Twigg, 1975).

R. norvegicus is thought to be originally native to the Hondo region of Japan as it is represented there by Holocene and late Pleistocene fossils (Musser and Carleton, 2005). Late Pleistocene – Holocene fossils of *R. norvegicus* have also been found in cave deposits in the Sichuan-Guizhou region of China (Musser and Carleton, 2005). This discovery adds to the conventional view that the original home range of this species also included the northerly temperate regions of China and Southeast Siberia, with distributions limited to forest and brushy areas, woodlands and open fields (Twigg, 1975; De Graff, 1981; Musser and Carleton, 2005). In South Africa, this species is not present within the fossil record and it is speculated to have only been introduced within the last 100 years (De Graaff, 1981).

Following a number of introductions that began in the 1700s, the distribution of the Norway rat has spread to such an extent that it is found virtually everywhere in the world, including areas that are dominated by cold harsh climates such as Antarctica (Twigg, 1975; De Graaff, 1981;

Musser and Carleton, 2005). The widespread dispersal of this species is mainly due to its ability to survive in a variety of environments, probably acquired as a result of its adaptation to the harsh Asian environments where it evolved (De Graaff, 1981). However, it has been suggested that the tropics (and possibly temperate areas as well) are unsuitable for this species, and that it acclimatises less readily than *R. rattus* to countries with warmer climates (Twigg, 1975; De Graaff, 1981; Musser and Carleton, 2005; Borchert *et al.*, 2007).

Possibly due to its slower acclimatization rate and relative inability to survive in an arid environment, as well as its less arboreal habit, this species is observed to have a preference for ground-based dwellings in areas that are close to a water supply. Large *R. norvegicus* populations are normally found close to or alongside rapidly expanding human populations in areas containing garbage dumps or sewers, where there is easy access to an abundant source of discarded human food and an ample water supply (De Graaff, 1981; Twigg, 1975; Musser and Carleton, 2005; Borchert *et al.*, 2007).

1.2.3 The Black Rat (*Rattus rattus*)

Rattus rattus is more commonly referred to as the black or roof rat. It is clearly distinguishable from *R. norvegicus* as it is smaller in size (approximately 165–254 mm) and has a tail that is longer than its body. Other distinguishing features include fur that is blue-black to black on the upper parts and slate or light gray to white on the under parts. The ears and eyes are large and the tail is black or brown and almost hairless (Long, 2003; De Graaff, 1981).

R. rattus has a higher reproductive rate than *R. norvegicus* (De Graaff, 1981). Wild individuals, which have an average lifespan of 11–17 months, are able to breed and produce young throughout the year. The female oestrus cycle occurs every ten days, whilst the gestation period is between 21 and 30 days. Usually, five to six litters are produced in a year with as many as 17 pups per litter. There is an observed interval between litters of approximately 32 days (De Graaff, 1981). During this 32 day interval, 21–28 days are spent weaning the current litter, after which individuals from weaned litters become independent from the adults and usually reach sexual maturity at three to four months (Long, 2003; De Graaff, 1981).

R. rattus is native to the Indian Peninsula and has since been introduced worldwide (Long, 2003, De Graaff, 1981; Musser and Carleton, 2005; Hingston *et al.*, 2005; Matisoo-Smith and Robins, 2009). The species is unknown in Europe before the Holocene and may have reached here via trade routes established by the Roman Empire with Asia (Long, 2003; De Graaff, 1981;

Musser and Carleton, 2005; Robins *et al.*, 2007; Tollenaere *et al.*, 2009). In North Africa (viz. Egypt and Libya) the earliest *R. rattus* remains were dated to 2000 BC, whilst in East Africa no fossils or archaeological remains have been found. However, *R. rattus* is considered to have been present since the beginning of the Christain era, as commercial links between the Arabian Peninsula, the Middle East and the East African Coast were already important during this period (Tollenaere *et al.*, 2009). In South Africa, fossil evidence indicates that *R. rattus* has been present in KwaZulu-Natal since the eight century and in Limpopo since 1000 A.D. (Borchert *et al.*, 2007). However, major introduction of this species into South Africa have only occurred within the last 100 years with its spread attributed to the spread of European settlements (De Graaf, 1981; Borchert *et al.*, 2007; Tollenaere *et al.*, 2009).

R. rattus is observed to be an excellent climber and prefers to inhabit areas that are above the ground, such as ceilings (De Graaff, 1981; Twigg, 1975). It has been observed, however, that Black rats will inhabit terrestrial environments and create burrows if arboreal environments are not available (Long, 2003). Much like the Norway rat, Black rats, although they are able to acclimatise relatively faster to warmer climates, cannot survive in arid areas and require sufficient access to a water supply (De Graaff, 1981). In South Africa, it has been found that this species has spread to virtually all areas where the rapid expansion of human populations has occurred, such as cities, towns, villages and farming centres where abundant food and water are present. This species is generally not, however, found in human settlements located in the drier areas of the country (De Graaff, 1981; Twigg, 1975).

1.2.4 The Asian House Rat (*Rattus tanezumi*)

Rattus tanezumi is commonly referred to as the Asian type house rat. This *Rattus* species is considered to be indigenous to Southeast Asia, Japan, the Philippines, Island Southeast Asia, New Guinea and several Pacific islands (Musser and Carleton, 2005; Robins *et al.*, 2007; Hingston *et al.*, 2005; Matisoo-Smith and Robins, 2009). This species is similar in appearance, habitat preference and behaviour to the black rat, *R. rattus* (Long, 2003; Musser and Carleton, 2005) and has only recently been recognised as a distinct species as previously it was considered a sub-species of the black rat (Long 2003; Musser and Carleton, 2005). This re-allocation from sub-species to species was based primarily on biochemical, cytogenetic and to a lesser degree morphological traits (Long, 2003; Musser and Carleton, 2005).

It should be noted that this species displays parapatric or overlapping population distributions with *R. rattus* (Musser and Carleton, 2005), however, it has been found that if *R. rattus* is

introduced to locations where populations of *R. tanezumi* are already present, *R. rattus* populations will be restricted along port areas (Musser and Carleton, 2005; Robins *et al.*, 2007; Matisoo-Smith and Robins, 2009).

1.2.5 Taxonomy

The taxonomy of the genus *Rattus,* since its initial definition in 1803, has over the years presented a number of problems, especially regarding subdivisions of the genus (De Graaff, 1981; Musser and Carleton, 2005). Initial problems emerged when the great majority of rodents that were obviously either rats or mice were referred to the genus *Mus*. This policy continued until 1881, when the erection of the sub-genus *Epimys* was proposed to cover the more typical rat, as opposed to mice (De Graaff, 1981). With the implementation of this sub-genus, only 27 forms were left in *Mus*, while 140 forms were placed in *Epimys*. However, in 1916 the sub-genus *Epimys* was dropped and all members placed within the genus *Rattus*, as a means of differentiating rats from mice (De Graaff, 1981). Even though the genus *Rattus* had been in existence for a number of years and had provided a better morphological description of what a typical rat was, its taxonomic priority was over-looked (De Graaff, 1981). At one point the genus *Rattus* was considered to consist of more than 550 distinct forms. Over the years a number of species that were initially allocated to *Rattus* have been reallocated to genera such as *Maxomys, Berylmys, Leopoldamys* and *Niviventer.* However, even though a number of species have been removed from the genus *Rattus*, the boundaries of this genus have not finally been decided (Aplin *et al.*, 2003; Musser and Carleton, 2005).

The initial type species of *Rattus* is *Mus decumanus* Pallas 1778, which is a synonym for *R. norvegicus. R. norvegicus* is morphologically distinctive, highly aggressive and territorial and appears to have few close relatives (Aplin *et al.*, 2003; De Graaff, 1981; Musser and Carleton, 2005). In contrast *R. rattus* appears to have numerous close relatives and is often confused with other Asian *Rattus* species, thus making field identifications highly questionable (Aplin *et al.*, 2003; Musser and Carleton, 2005).

Other problems with the taxonomy of the genus *Rattus* relate to the fact that several groups of species may still need to be removed from this genus (Aplin *et al.*, 2003; Musser and Carleton, 2005). Evolutionary relationships of *Rattus* to other murid genera also remain somewhat enigmatic, and this problem is further compounded by the lack of attention given to this genus by taxonomists and geneticists (Aplin *et al.*, 2003; Chinen *et al.*, 2005; Musser and Carleton, 2005).

Taxonomic revisions within the genus *Rattus* are essential, as other species that could belong to this genus may be recognised once more detailed morphological and genetic studies have been conducted. Even though many authors have attempted to make sense of the morphological diversity among black rats, these attempts have been largely regionally–based, with authors placing too great an emphasis on belly fur colour, even though this characteristic is highly polymorphic within populations and under relatively simple genetic control (Aplin *et al.*, 2003; Musser and Carleton, 2005).

The most significant breakthrough in understanding the *R. rattus* complex came with Yosida's studies, from 1980 onwards, of chromosomal variation (Aplin *et al.*, 2003). He documented 11 chromosomal variants that resolve into five major groups viz. (i) Asian black rats with $2n = 42$ chromosomes; (ii) Japanese black rats with $2n = 42$ chromosomes; (iii) Ceylonese black rats with $2n = 40$ chromosomes; (iv) 'Oceanian' or European black rats with $2n = 38$ chromosomes; and (v) Mauritius black rats with $2n = 42$, where the notation $2n$ refers to the diploid chromosome number (Aplin *et al.*, 2003). In more recent taxonomic revisions by Musser and Carleton (2005), Yosida's five recognised *R. rattus* groups have been reduced to only two basic population groups, viz. the 'Oceanian' or European type having a chromosome number of $2n$ = 38/40, which has been assigned the name *R. rattus* by Musser and Carleton (2005), and the Asian type black rats having a chromosome number of $2n = 42$, which has been assigned the name *R. tanezumi* by Musser and Carleton (2005) (Hingston *et al.*, 2005; Robins *et al.*, 2007; Matisoo-Smith and Robins, 2009).

It is clear that further genetic, cytogenetic and morphometric studies are still required in order for us to fully understand the complexity of the relationships between the various species that make up the *R. rattus* complex.

1.3 Review of species concepts

Prior to the 1990s, the Biological Species Concept (BSC) was the dominant species concept (Agapow *et al.*, 2004). The BSC defines species as populations that cannot interbreed successfully, and implies that species are necessarily reproductively isolated, and represent separate evolutionary lineages. The appeal of the BSC is that it proposes that species boundaries are falsifiable by the natural (and substantial) production of fertile hybrids across them.

However, due to its simplicity, the BSC is flawed as a concept, as assessing reproductive barriers can be difficult and time consuming. Speciation often occurs via an allopatric route in the wild, thus making this process difficult to define objectively. The BSC cannot be applied to asexually reproducing organisms. Difficulties are also encountered in populations where hybridization occurs (Agapow *et al.*, 2004).

The specific-mate recognition species concept of Paterson (as reviewed by McNaught and Owens, 2002) dictates that species–specific pre-mating signals allow individuals to recognize members of their own species, thus preventing interbreeding between members of different species. The specific-mate recognition species concept is considered to contain a number of flaws, as it does not predict reproductive character displacement in sympatric populations, and is not falsifiable when sympatric species appear to be no more divergent than allopatric species (McNaught and Owens, 2002).

A cohesion species is an evolutionary lineage whose boundaries arise from the genetic and ecological forces that create cohesive reproductive communities. According to this species concept, an evolutionary linage is a reproducing population with sufficient historical continuity (ancestral – descendant relationships) to have its own evolutionary trajectories and tendencies. The use of the Cohesion Species concept is highly appealing as it can be used to define evolutionary lineages that occur within sexual and asexual taxa. This species concept can also be used to define hybrids produced from a stabilized or recombinant hybridization event between species in botanical taxa. The appeal of the Cohesion Species concept is that unlike other species concepts, it acknowledges asexual reproduction and hybridization as important and potentially creative forces in speciation (Templeton, 2001).

According to the Phylogenetic Species Concept (PSC), a species is defined as a group of organisms that share at least one uniquely–derived character, have a shared pattern of ancestry and descent, or exhibit monophyly. The PSC is widely applicable in practice, in comparison to other species concepts, due to its applicability to asexual organisms and allopatric populations. It has also been argued by many authors that the PSC is highly objective, has the ability to distinguish morphologically unremarkable but important populations and can be considered to be a good indicator of the biodiversity and conservation worth of a population. Criticism of the PSC is that zealous application can result in the creation of large numbers of taxonomic units (Agapow *et al.*, 2004).

The Genetic Species Concept utilizes DNA base composition and sequence data as well as biochemical and enzymatic data to classify and define a species, as opposed to the practice of splitting up genera into a number of species based on small phenotypic differences (Friedman and De Ley, 1965). Over the years this species concept has been continuously reviewed and to date the current definition of a genetic species is that of a group of genetically compatible interbreeding natural populations that is genetically isolated from other such groups. Under this revised definition of the Genetic Species Concept, speciation is considered as the accumulation of genetic changes in two lineages that produces genetic isolation and two gene pools that have independent evolutionary fates (Baker and Bradley, 2006).

The Phylogenetic and Genetic Species concepts appear to be the most relevant species concepts for use in interpreting results obtained in this study of genetic variation in *Rattus* in KZN. These two species concepts will be utilized primarily due to their widespread applicability and the objectivity offered by both concepts.

1.4 Molecular methods used in assessing genetic diversity

Since the last decade of the 20th century, we have seen a considerable impact of population-genetic studies on our understanding of evolutionary processes, population and species history. These advances have revolutionized population genetic research to such a degree that this discipline is no longer largely a debating field of mathematics and theory, but has become an explanatory science. The recent development of coalescent and phylogenetic theory has changed the way we analyse and interpret molecular data. The effects of such technical advances are so profound that they have transformed the mainstream of population-genetics research (Zhang *et al.*, 2003).

With recent advances in molecular technology, the number of DNA-based markers have greatly increased (Isabel *et al.*, 1999). This increase in DNA-based markers has resulted in DNA sequencing becoming the most commonly–utilized molecular method for assessing genetic variation, as it provides knowledge of the nucleic acid sequence that is specific for the target DNA (Gülbitti *et al.*, 2003; Gharizadeh *et al.*, 2003; Chan, 2005). This increase in the reliance on DNA sequencing has resulted in mitochondrial DNA (mtDNA) being identified as one of the most suitable markers for genetic studies where taxonomic reconsideration of closely related species or populations of a variety of species is required (Koh *et al.*, 2004).

Mitochondrial DNA has been chosen as a suitable marker for genetic studies as it contains relatively conserved sequences such as the cytochrome *b* gene, which in recent years has been utilized in many phylogenetic studies. Mitochondrial DNA also contains relatively variable non-coding regions such as the D-loop, which is often used in systematic studies in conjunction with the coding cytochrome *b* region. mtDNA has a small size, a highly conserved structure and a mean mutation rate of 2% of its nucleotide composition per one million years in its coding regions, while non-coding regions are quoted to have a nucleotide mutation rate in the range of 3.6% per million years (Gülbitti *et al.*, 2003; Ballard and Whitlock, 2004; Koh *et al*, 2004; Aboim et al., 2005).

While analysis of mitochondrial DNA (mtDNA) has proven powerful for genealogical and evolutionary studies of animal populations, it has a number of limitations. These include the presence of mitochondrial pseudogenes in the nucleus that may occur at high copy numbers and be mistakenly amplified, resulting in false comparisons (Gülbitti *et al.*, 2003). mtDNA is maternally-inherited, thus biasing inferences made about species/population history using mtDNA as a genetic marker (Gülbitti *et al.*, 2003; Zhang *et al*., 2003). Another consideration with regards to the use of mtDNA as a genetic marker is that the effective population size of mtDNA is only a fourth of that of nuclear autosomal sequences, thus mtDNA lineages have a much faster lineage sorting rate and higher allele extinction rate, resulting in (i) possible over-simplification of evolutionary relationships estimated from mtDNA data, (ii) under–estimation of genetic diversity and (iii) uncertainty in genealogical analysis due to the increased probability of missing links in mitochondrial haplotypes (Zhang *et al*., 2003).

Even though a number of problems have been encountered in the use of mtDNA, its popularity in genealogical and evolutionary studies on animals has not diminished (Freeland, 2005). The popularity of mtDNA in these studies is based on its small size and highly-conserved structure (Gülbitti *et al.*, 2003), as well as the ease with which it can be manipulated, its relatively rapid mutation rate, its presumed lack of recombination, which results in an effectively clonal inheritance, and the fact that universal animal mitochondrial primers are readily available (Freeland, 2005).

For this study it was decided to sequence the cytochrome *b* and D-loop regions of the mitochondrial DNA. The choice of the mitochondrial cytochrome *b* gene sequence was based on its widespread successful use in other phylogenetic studies on rodents, as well as the availability of previously–published cytochrome *b* sequences on DNA databases such as

GenBank, which allow for comparison with sequences obtained in this study (Robinson *et al.*, 1997; Martin *et al.*, 2000; Ballard and Whitlock, 2004; Koh *et al.*, 2004; Yang and Speller, 2006).

The use of D-loop sequences in this study was primarily to provide a greater level of resolution, as D-loop sequences usually evolve at a faster rate than cytochrome *b* sequences as they are non-coding (Martin *et al.,* 2000; Kerth *et al.*, 2000; Ballard and Whitlock, 2004; Koh *et al.*, 2004; Yang and Speller, 2006). Martin *et al*. (2000) also state that if the cytochrome *b* sequence analyses are in agreement with D-loop sequence analyses, the relationships between populations inferred from DNA samples indicate stable phylogenetic relationships.

1.5 Phylogeography

The concept of phylogeography was first introduced by Avise *et al.* in 1987 as a means of intergrating and uniting the disparate fields of phylogenetics and population genetics and to provide a means by which phylogenetic analysis of geographically contextualized data could be used to test hypotheses with regard to the causal relationship among geographic phenomena, species distributions, and the mechanisms driving speciation (Hickerson *et al.*, 2010). Phylogeography can thus be considered as the study of principles and processes governing the geographical distributions of genealogical lineages, including those at the intraspecific level (Ditchfield, 2000), or more simply put, it is defined as the phylogenetic analysis of organismal data in context of the geographic distribution of the organism (Hickerson *et al.*, 2010).

The initial and still-dominant infrastructure that governs the discipline of phylogeography is that of mitochondrial DNA (mtDNA) analyses at species level (Bermingham and Moritz, 1998; Hickerson *et al.*, 2010). Analysis of mtDNA has permitted genealogical traces to be followed across the genetic boundaries between populations, species and higher taxonomic levels due to its lack of recombination, putative neutrality and shorter time to reciprocal monophyly between geographic regions (Bermingham and Moritz, 1998; Hickerson *et al.*, 2010). The empirical success of mtDNA-based phylogeography has led to more precise descriptions of geographical distribution, phylogenetic relationships and genetic distances among evolutionary lineages of animals, thus allowing for a better understanding of biogeographical distributions and areas of endemism (Bermingham and Moritz, 1998).

In this this study, the concept of phylogeography will be applied to identify the geographical

distribution of the members of the *Rattus* complex in KwaZulu – Natal based on the genetic composition of the various populations sampled.

1.6 Data analysis

1.6.1 Introduction to phylogenetic analyses

Many types of data can be applied to the investigation of evolutionary relationships among organisms. The classical method of estimating the relationship between taxa was comparison of their morphological characters (the Morphological Species Concept). However with the advent of new molecular technologies and the increasing availability of nucleotide or amino-acid sequences, phylogenetic relationships can now be determined using molecular data as well as visible morphological characteristics. The use of molecular data to produce phylogenetic trees is considered by some to be a more precise and objective way of inferring phylogenetic relationships amongst species (Vandamme, 2003).

Molecular data can be analyzed in a variety of ways to produce trees; methods include phenetic/distance-based analyses and phylogenetic (cladistic) analyses (Vandamme, 2003).

1.6.2 Character – state methods

Character-state methods of analyzing molecular data include Maximum Parsimony analysis, Maximum Likelihood analysis and Bayesian Inference. Maximum Parsimony analysis aims to find the tree topology for a set of aligned sequences that can be explained with the smallest number of character changes (mutations). The Maximum Parsimony algorithm starts by considering a tree with a particular topology and then infers the minimum number of character changes required to explain all nodes of the tree at every sequence position. The Maximum Parsimony algorithm evaluates all possible tree topologies obtainable using a given data set, with the best tree being chosen by this algorithm as that tree which contains the minimum number of changes (Vandamme, 2003).

Maximum Likelihood analysis is similar to the analyses carried out using the Maximum Parsimony algorithm in that it examines every reasonable tree topology and evaluates the support for each by examining every sequence position. In principle, the Maximum Likelihood algorithm calculates the probability of expecting each possible nucleotide (amino acid) in the ancestral (internal) nodes and infers the likelihood of the tree structure from these probabilities.

The likelihood of all reasonable tree topologies is searched in this way, and the most likely tree is chosen as the best tree (Vandamme, 2003).

Bayesian Inference is a relatively new technique that is rapidly gaining popularity. In Bayesian analyses inferences are based upon the posterior probability of a parameter, which is the probability of the parameter conditional on the observed data. To define phylogenetic trees, Bayesian Inference uses the Markov Chain Monte Carlo method to estimate the posterior probability of phylogenies. A Bayesian analysis of phylogenies requires that the investigator specify any prior knowledge about the phylogeny before observing any data; this is considered to be a strength of this technique, as it takes advantage of any prior knowledge the investigator may have with regards to the phylogeny under study. A robust Bayesian analysis can usually be conducted in a relatively short time, for example, overnight, which makes Bayesian analysis easier to use than the more computer-intensive Maximum Parsimony or Maximum Likelihood methods (Huelsenbeck and Bollback, 2001).

1.6.3 Genetic distance models

Any two sequences, derived from a common ancestor, which evolve independently, will eventually diverge from each other; genetic distance is a measure of this divergence. As the divergence increases, there is an apparent saturation of the amount of sequence change. This is due primarily to the occurrence of multiple substitutions at the same site. A number of genetic distance models have been developed to correct for this; these include the Jukes Cantor, Kimura 2-parameter, HKY85 and General Time Reversible genetic distance models (Strimmer and von Haeseler, 2003). These models may take into account factors such as the transition/transversion ratio, nucleotide frequencies and the proportion of invariant sites in estimating corrected genetic distances, which are larger than uncorrected distances as they account for the occurrence of multiple substitutions at the same site. Computer programmes such as ModelTest (Posada and Crandall, 1998) and MrModelTest, a modification of ModelTest by Nylander (2004), are used to predict which model best fits a particular dataset based on the Akaike Information Criterion (AIC). The Akaike Information Criterion is an unbiased estimator of the expected relative Kullback-Leibler information quantity or distance (K-L) which represents the amount of information lost when we use one model to evaluate another to identify the best-fitting model to be used for analysis (Burnham and Anderson, 2002; Posada and Buckley, 2004).

1.6.4 Software packages used to analyze data for this study

There are a multitude of methods to assess and analyze molecular data, some of which include mathematical calculations that are very time consuming. Software programs have thus been developed to aid researchers in analyzing their data. The programs used in my study are BioEdit, ClustalX, MEGA, PAUP, MrBayes and ModelTest.

BioEdit is a user–friendly sequence alignment editor and analysis package (Hall, 1999) that may be used to visually edit and align sequence data. The program may also be used to perform Blast searches of DNA databases such as NCBI GenBank (Hall, 1999) to obtain highly similar sequences for reference purposes.

ClustalX (Thompson *et al.*, 1997) is a sequence alignment program which can be used to convert sequence data to alternate file formats which are used by other programs such as MEGA (Kumar *et al.*, 2004); PAUP (Swofford, 2001) and MrBayes (Ronquist and Huelsenbeck, 2003). Both PAUP and MEGA can be used to carry out standard phenetic and phylogenetic analyses, although PAUP is preferred. MrBayes may be used to conduct phylogenetic analyses using Bayesian Inference (Hall, 2001).

The program ModelTest (Posada and Crandall, 1998) is designed to compare different nested models of DNA substitution in a hierarchical hypothesis-testing framework. The program is used to determine which of the 56 genetic distance models available is the most appropriate for the analysis of a particular dataset.

MrModelTest is a modification of ModelTest by Nylander (2004). This program is used to determine which of 24 genetic distance models best fits sequence data (Nylander, 2004) and to determine the model parameters to be used in Bayesian analysis (Ronquist and Huelsenbeck, 2003).

The program DNASP 4.10 (Rozas *et al.* 2003) is a statistical software package that makes use of coalescent-based methods by Monte Carlo computer simulations to extensively analyze genetic differentiation and gene flow among populations. The use of this program in any genetic study is critical for the detection of any possible signatures of positive natural selection, the identification of haplotype blocks across the genome, and for inferring the effects of intragenic recombination (Rozas *et al.* 2003).

The program TCS v1.21 (Clement *et al.* 2000) estimates genealogical relationships amongst organisms using sequence data, by collapsing sample sequences into haplotypes and calculating the frequencies for these haplotypes. The frequencies are then used to estimate haplotype outgroup probabilities as well as calculate an absolute distance matrix. The absolute distance matrix is comprised of all pairwise comparisons of haplotypes which are calculated until the probability of parsimony exceeds 95% for all pairwise differences. All mutational differences associated with the probability just before the 95% cut-off are then considered as the maximum number of mutational connections between pairs of sequences and output as a haplotype network (Clement *et al.* 2000).

1.7 Cytogenetic analyses

Within a eukaryote cell, the genome is divided into a number of chromosomes. Each chromosome contains a very large single linear DNA molecule, which is supercoiled, and is normally associated with DNA-binding proteins. Such proteins, complexed with the DNA, compromise the chromatin. Different eukaryotic species contain widely–varying numbers of distinguishable chromosomes, from one in an Australian ant to 190 in a species of butterfly (Mathews *et al.*, 2000). Most eukaryotes have diploid cells and carry two copies of each chromosome, although instances of polyploidy, especially in plants, are common.

Past studies of chromosomal variability within rodents have revealed that members of the family Muridae exhibit a tendency for extensive variation within species and species complexes (see Taylor, 2000 for a full review of southern African rodent case studies). In his review, Taylor (2000) reported that there was a large amount of chromosomal variation in species belonging to genera such as *Otomys* and *Mastomys.* One species, *O. irroratus* displays karyotypes with diploid numbers ranging from 2n = 23 to 2n = 32 (Contrafatto *et al.,* 1992a, 1992c). Species belonging to the genus *Mastomys* are generally characterised as having a karyotype of either 2n = 32 or 2n = 36 (Taylor, 2000). Within *O. irroratus*, variation is attributed to the number of heterochromatic short arms (Contrafatto, 1992a, 1992c) as well as pericentric inversions and a complex tandem fusion (Rambau *et al.*, 2001; Engelbrecht *et al.*, 2006, Taylor *et al.,* 2009), whereas within *Mastomys*, karyotype variation is due to either peri- or para- centric inversions, heterochromatic changes, or Robertsonian fusions and fissions (Taylor, 2000).

Initial studies into the chromosome variability of members of the genus *Rattus* have reported

this genus as having karyotypes that are highly conserved (Baverstock *et al.*, 1983; Caldarini *et al.,* 1989; Yosida, 1985). However, more recent studies have revealed that chromosome variability amongst certain members of this genus is much higher that initially reported (Baverstock *et al.*, 1983; Yiğit *et al.,* 1998; Chinen *et al.*, 2005).

Within the *R. rattus* complex, observations indicate that a pattern of chromosome polymorphism exists, and that this polymorphism is either derived from one of the karyomorphs of *R. rattus* or differentiated progressively from that of the black rat (Baverstock *et al.*, 1983; Yosida, 1985). It is generally accepted that the basic karyotype of *R. rattus* has a diploid number of 2n = 42 and comprises 13 pairs of acrocentric chromosomes, 7 metacentric pairs and an acrocentric pair of XY or XX sex chromosomes (Baverstock *et al.*, 1983; Caldarini *et al.,* 1989; Yosida, 1985; Chinen *et al.*, 2005). Variations to this basic karyotype are due to either Robertsonian fusions or fissions (Yosida, 1985).

Polymorphisms have not been observed within the basic karyotype of *R. norvegicus*, and as such the karyotype of this species is considered to be conserved amongst all observed populations worldwide (Yiğit *et al.,* 1998). The basic karyotype for *R. norvegicus* is composed of 10 acrocentric chromosome pairs, 2 sub-telocentric chromosome pairs, 9 metacentric chromosome pairs and an acrocentric pair of XY or XX sex chromosomes, resulting in a diploid number of 2n = 42 (Yiğit *et al.,* 1998).

In this study, metaphase chromosomes of somatic cells were used to construct karyograms that were then compared to the basic karyotypes of *R. rattus* and *R. norvegicus* to identify *Rattus* specimens analyzed.

1.9 Aims, objectives and hypotheses of this study

The primary goals of this study were to elucidate and compare patterns of genetic diversity between different members of the *Rattus* complex in KwaZulu-Natal, to identify the number of species of *Rattus* occurring within this complex, and to obtain an estimate of their relatedness, phylogenetic and phylogeographic structure in KwaZulu-Natal.

To achieve these goals, the cytochrome *b* and D-loop regions of the mitochondrial DNA were sequenced from tissues obtained from specimens collected from various populations, and dendrograms and haplotype networks were created using phenetic and phylogenetic methods.

Karyology of the collected specimens were carried out to further substantiate relationships inferred from sequence data. It was hypothesized that there were two distinct species of *Rattus* within the complex found in KZN, but as the presence of *R. tanezumi* became increasingly apparent during this study, the study's aim was refocused to include an elucidation of the relationship within and between this species and its close sibling *R. rattus.*

CHAPTER 2

MATERIALS AND METHODS

2.1 Sample collection and storage

Sampling was conducted in various areas of KwaZulu-Natal, with emphasis on populated areas, as the study organisms are usually found in close proximity to human habitation.

Live specimens were captured using either multi–capture or Sherman traps. Once captured, specimens were identified in the field using morphological characteristics such as size and position of the ears, overall size of specimen and fur colour. These characteristics were chosen for preliminary field identification of captured *Rattus* specimens, as *R. norvegicus* is identified as a very large rodent with ears that are reduced in size and closer to the skull, with a brown fur colour (Twigg, 1975; De Graaff, 1981; Long, 2003), whereas *R. rattus* is identified as a rodent with large ears that protrude from the skull, a gray fur colour and a relatively longer tail (De Graaff, 1981; Long, 2003).

Once specimens were initially identified, they were transported to animal housing facilities at the School of Biological and Conservation Sciences and quarantined for two weeks. Specimens were then euthanazed within a sealed container that contained tissue paper doused with 100% chloroform which was in accordance with the protocol approved by the Animal Ethics subcommittee of the University of KwaZulu-Natal Research committee. Liver tissue was excised from each specimen and stored in 90% ethanol at 4ºC until needed for genetic analysis. Voucher specimens were prepared using standard natural museum procedures for mammal specimens and were deposited in the South African mammal reference collections of the Durban Natural Science Museum. Sample details are reported in Table 1.

Table 1: Sample names and collection localities of *Rattus* specimens. Identification numbers indicated in the Specimen/Sample ID column represent a single specimen. DM = Durban Natural Science Museum

Sample Origin		Coordinates		Specimen/ Sample ID	Museum Catalogue Number
		Latitude	Longitude		
KwaZulu-Natal Province					
Durban Central Business District					
	Cnr. Russell & St. George Street	29°51' 43.24"S	31° 00' 57.94"E	SA 213 SA 217	DM7783
	Cnr. Commercial & Soldiers Way	29°51' 24.53"S	31° 01' 26.04"E	SA 234 SA 235	DM7804 DM7805
	Browns Road	29°52' 16.65"S	31° 02' 52.49"E	SA 136 SA 141	DM7706 DM7711
	Warwick Avenue	29°51' 11.60"S	31° 00' 35.68"E	SA 75	DM7645
Sydenham					
	View Street	29°49' 49.75"S	31° 00' 2.97"E	SA 174 SA 175	DM7744 DM7745
Cato Crest					
	Nzuza CC1504	29°51' 36.37"S	30°58' 35.08"E	SA 220 SA 221	DM7802 DM7803
	MaKhanyile CC1514	29°51' 36.37"S	30°58' 35.08"E	SA 232 SA 233	DM7790 DM7791
	145 Cato Manor Rd	29°51' 36.37"S	30°58' 35.08"E	SA 252	DM7820
Montclair					
	Montclair Park	29°55' 26.67"S	30°57' 56.91"E	SA 218 SA 219	DM7788 DM7789

Sample Origin		Coordinates		Specimen/ Sample ID	Museum Catalogue Number
Shongweni					
	Pickford Estates	29°47' 51.24"S	30°44' 27.93"E	SA 253	DM8401
				SA 254	DM8402
Umkomaas					
	Axbridge Place, Widenham	30°13' 13.81"S	30°47' 39.07"E	SA 255	DM8403
				SA 256	DM8404
				SA 257	DM8405
				SA 258	DM8406
				SA 259	DM8407
Verulam					
	Park Station, Canelands	29°38' 47.47"S	31° 2' 47.98"E	PS1	
Richmond		29°52' 27.70"S	30°16' 48.46"E	R1	DM8685
Gauteng Province					
	Hammanskraal	25°23' 26.68"S	28°18' 3.74"E	ARC 170	
				ARC 171	
Pretoria		25°44' 45.67"S	28°11' 13.63"E	UP02	
Limpopo Province					
	Nkomo – B	23°26' 25.84"S	30°45' 17.82"E	ARC 79	
				ARC 101	
Vietnam					
	Hung Yen Province, North Vietnam	20°38' 11.72"N	106° 3' 25.20"E	HP1*	
				HP2*	
Indonesia					
	Subang, West Java	06°34'5.05"S	107°45'35.39"E	S1**	

*Specimens donated by Dr. Nguyen Phu Tuan, National Institute of Plant Protection, Chem Tu Liem, Hanoi, Vietnam

** Specimens donated by Dr. M. P. Sudarmaji, Indonesian Center for Rice Research, Subang, West Java, Indonesia

2.2 Molecular analyses

2.2.1 DNA isolation, extraction and quantification

DNA was extracted from 25 mg of rat liver using a QIAGEN DNeasy® Tissue Extraction Kit according to manufacturer's instructions. Quantification of all extracted samples was conducted using a Nanodrop® spectrophotometer set at an absorbance wavelength of 260 nm. The absorbance of a 1µl DNA sample was measured with reference to a blank comprising 1µl AE buffer (QIAGEN). The integrity of the extracted DNA was assessed by electrophoresis (Figure 1) in 1.2% agarose gels containing 0.005 mg/µl ethidium bromide using a 0.5x TBE running buffer at 70 volts for ± $1^{1}/_{2}$ hours.

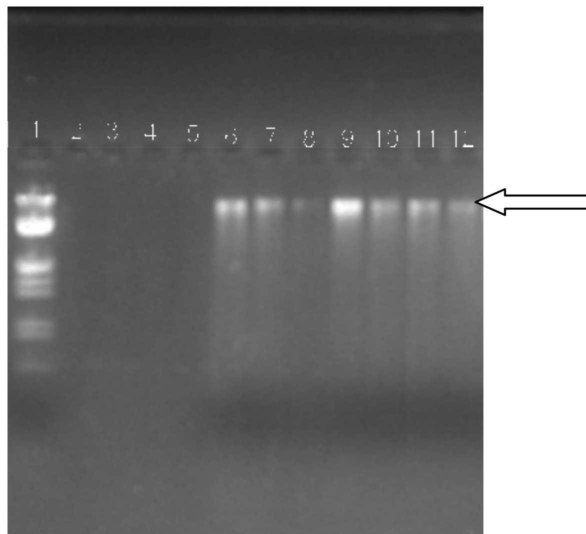

Figure 1: Agarose gel electropherogram used to assess the integrity of DNA isolated from representative samples of *Rattus* liver tissue (lanes 6-12). All samples in above electropherogram contain high molecular weight bands (indicated by the arrow). Lane 1 contains Molecular Weight Marker III (Roche Molecular Biochemicals).

2.2.2 Polymerase Chain Reaction (PCR) amplification

The cytochrome *b* gene was amplified as a single fragment using the forward primer L14724 and the reverse primer H15915-mus (Dr. A. D. Bastos pers. comm., 2006). The D-loop region was amplified using the forward primer P and reverse primer E (Wilkinson and Chapman, 1991). The primer sequences are presented in Tables 2 and 3, respectively.

Table 2: Primers used in PCR amplification of the mitochondrial cytochrome *b* region

Primer Name	Direction	Sequence 5' – 3'
L14724/ L24	Forward	TGACATGAAAAATCATCGTTG
H15915/ H15	Reverse	CATTTTTGGTTTACAAGAC

Table 3: Primers used in PCR amplification of the mitochondrial D-loop region

Primer Name	Direction	Sequence 5' – 3'
P	Forward	TCCTACCATCAGCACCCAAAGC
E	Reverse	CCTGAAGTAGGAACCAGATG

Amplification of the cytochrome b and D-loop regions was conducted by combining 32 µl of master mix (Tables 4 and 5) with 60 ng sample DNA in 18 µl sterile water to produce a reaction mixture with a total volume of 50 µl. Temperature profiles for these reactions are presented in Tables 6 and 7, respectively.

Table 4: PCR master mix composition for a single amplification of the cytochrome *b* gene

Component	Volume (µl)
Sterile Water	1.6
Buffer (10x)	5
$MgCl_2$ (25 mM)	8
Primer L14724 (6.25 µM)	8
Primer H15915-mus (6.25 µM)	8
dNTP's (10mM)	1
Taq (5 µ/µl)	0.4
Final Volume	**32**

Table 5: PCR master mix composition for a single amplification of the D-loop region

Component	**Volume (µl)**
Sterile Water	1.6
Buffer (10x)	5
$MgCl_2$ (25 mM)	8
Primer P (6.25 µM)	8
Primer E (6.25 µM)	8
dNTP's (10mM)	1
Taq (5 µ/µl)	0.4
Final Volume	**32**

Table 6: Thermal cycling procedure used for the Cytochrome *b* PCR

Stage	**Repeat**	**Temperature (ºC)**	**Time (minutes)**	**Purpose**
1	1	94	2	Denaturation of DNA
2	5	96	0:12	Denaturation of DNA
		49	0:25	Annealing of primers to DNA
		72	0:45	Extension of primer
3	10	96	0:12	Denaturation of DNA
		47	0:20	Annealing of primers to DNA
		72	0:45	Extension of primer
4	25	96	0:12	Denaturation of DNA
		45	0:10	Annealing of primers to DNA
		72	0:45	Extension of primer
5		72	7	Extension of primer
6		15	∞	Holds at this temperature after the reaction until DNA is removed from machine

Table 7: Thermal cycling procedure used for D-loop PCR

Stage	Repeat	Temperature (ºC)	Time (minutes)	Purpose
1	1	95	2	Denaturation of DNA
2	35	95	1	Denaturation of DNA
		55	1:30	Annealing of primers to DNA
		72	2	Extension of primer
3	1	72	7	Extension of primer
4	1	15	∞	Holds at this temperature after the reaction until DNA is removed from machine

The cytochrome *b* PCR products were separated by electrophoresis (Figure 2) in 1.2% agarose gels containing 0.005 µg/ml ethidium bromide for ± $1^1/_2$ hours in 0.5x TBE running buffer at 70 volts, while D-loop PCR products were separated by electrophoresis in a 1.2% agarose gel containing 0.005 µg /ml ethidium bromide for ± 24 hours in 0.5x TBE running buffer at 15 volts.

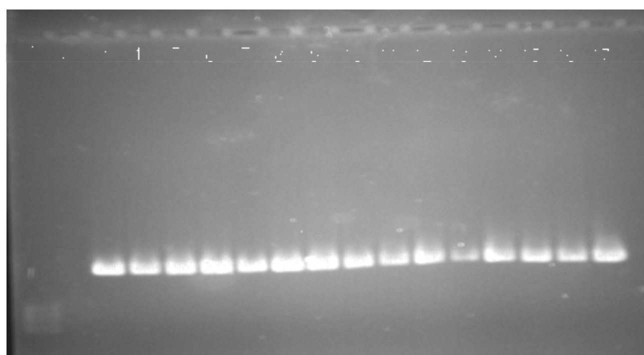

Figure 2: Agarose gel electropherogram containing cytochrome *b* PCR-amplification products comprising a single band (lanes 3 – 17) that indicated successful amplification reactions. Lane 1 contains Molecular Weight Marker III (Roche Molecular Biochemicals).The presence of smears and extra bands was attributed to the amplification of non-target DNA.

After electrophoresis the gels were transilluminated with an ultraviolet light source to locate the position of bands corresponding to the amplification products. Desired amplification products were identified by their mobility relative to a co-electrophoresed molecular weight marker (Fermentas). Appropriate bands were excised from the gel for purification with a QIAquick ® Gel Extraction Kit. Gel slices were stored in a -20 ºC freezer in appropriately labeled 1.5 ml

Eppendorfs until required.

Recovery of DNA from gel slices was conducted using a QIAGEN QIAquick® Gel Extraction Kit according to manufacturer's instructions. The recovered DNA was quantified using Nannodrop® spectrophotometry as described in Section 2.2.1 but with the use of buffer EB (QIAGEN) as a reference. Samples containing 50-80 ng of template DNA were suitable for use in sequencing reactions. In order to verify that only a single band had been purified, 1.2% agarose gels containing 0.005 µg/ml ethidium bromide were electrophoresed (Figure 3) in 0.5x TBE running buffer at 70 volts for ± $1^1/_2$ hours.

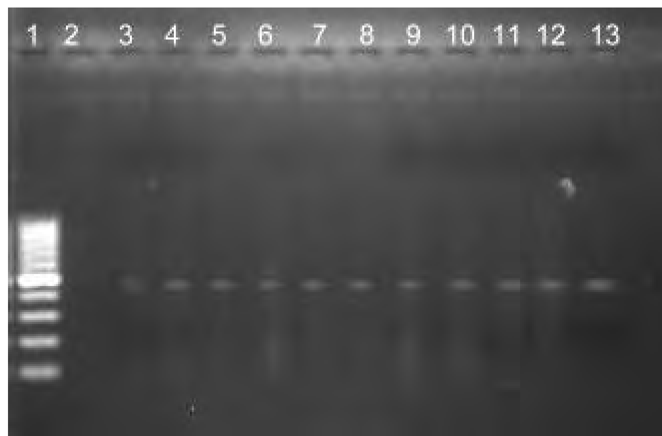

Figure 3: Agarose gel electropherogram containing extracted and purified mitochondrial DNA amplification products. Lane 1 contains a Molecular Weight Marker (Fermentas) while lanes 3–13 contain the extracted and purified D-loop PCR-amplification products.

2.2.3 DNA sequencing

DNA obtained from the gel extraction protocol was directly sequenced by technicians at Inqaba Biotech, Hatfield, Pretoria, South Africa. Both the cytochrome *b* and D-loop regions were sequenced in the forward and reverse directions using the primers used in the original PCR amplifications. Sequences were returned in the form of ABI data files.

2.2.4 Sequence editing and alignment

Electropherograms were imported into the BioEdit Sequence Alignment Editor (Hall, 1999) for editing. The forward and reverse electropherograms of each sample were compared and where discrepancies arose, appropriate changes to the sequence were made. Reference sequences for both the cytochrome b and D-loop regions of the mitochondrion of *R. norvegicus*, *R. rattus* and *R. tanezumi* were obtained from the NCBI GenBank (Tables 8 and 9) and aligned with the

sequences obtained in this study (Apendices 4 – 6). As certain key GenBank-derived cytochrome *b* reference sequences were too short in comparsion to study sequences generated, it was decided that two cytochrome *b* datasets would be analyzed viz. a cytochrome *b* dataset comprising sequences 1146 nucleotides long and a cytochrome *b* dataset comprising sequences 676 nucleotides long (Apendices 4 and 5).

Table 8: NCBI GenBank reference sequences included in the analysis of cytochrome *b* sequences

Sample Origin	Species	Sample Name	GenBank Accession No.
China	*R. norvegicus*	RnChina	AF295545
	Mus musculus domesticus	MmdChina	AF520636
Copenhagen, Denmark	*R. norvegicus*	RnCopen	AJ428514
Haiphong Port, Vietnam	*R. norvegicus*	RnHai	AB355902
Huahine, Society Islands, French Polynesia	*R. norvegicus*	RnHua	EF186461
Raiatea, Society Islands, French Polynesia	*R. norvegicus*	RnRai	EF186462
USA	*R. norvegicus*	RnUSA	DQ673916
Japan	*R. tanezumi*	RtJap	AB211042
	R. rattus	RrJap	AB033702
	R. argentiventer	RaJap	AB033701
	Diplothrix legata	DlJap	AB033696
Samoa	*R. Rattus*	RrSam	EF186475
Titirangi, New Zealand	*R. rattus*	RrTit	EF186470
Moitaka, Papua New Guinea	*R. rattus*	RrMoi	EF186471
Kuala Lumpur, Malaysia	*R. rattus*	RrKL	EF186413
Huahine, Society Islands, French Polynesia	*R. rattus*	RrHua	EF186469
Jakarta, Indonesia	*R. tanezumi*	RtJak	EF186491
Philippines	*R. tanezumi*	RtPhil	DQ191488
Northern Sulawesi, Indonesia	*R. tanezumi*	RtNS	EF186511
Yogyakarta, Indonesia	*R. tanezumi*	RtYog	EF186378
Amami Island, Japan	*R. tanezumi*	RtAI	EF186508
Philippines	*R. everetti*	RevPhil	DQ191485
Philippines	*R. praetor*	RpPhil	DQ191487
Philippines	*Limnomys bryophilus*	LbPhil	DQ191479
Philippines	*Tarsomys apoensis*	TaPhil	DQ191491

Table 9: NCBI GenBank reference sequences included in the analysis of D-loop sequences

Sample Origin	Species	Sample Name	GenBank Accession No.
Italy	*R. norvegicus*	RnItaly	X14848
Copenhagen, Denmark	*R. norvegicus*	RnCopen	AJ428514
Huahine, Society Islands, French Polynesia	*R. norvegicus*	RnHua	EF186346
Raiatea, Society Islands, French Polynesia	*R. norvegicus*	RnRai	EF186347
USA	*R. norvegicus*	RnUSA	AY172581
	R. rattus	RrUSA	RRU13750
Hauhine, Society Islands, French Polynesia	*R. rattus*	RrHua	EF186354
Kuala Lumpur, Malaysia	*R. rattus*	RrKL	EF186298
Madagascar	*R. rattus*	RrMad	DQ009791
Moitaka, Papua New Guinea	*R. rattus*	RrMoi	EF186356
Samoa	*R. rattus*	RrSam	EF186360
Titirangi, New Zealand	*R. rattus*	RrTit	EF186355
Northern Sulawesi, Indonesia	*R. tanezumi*	RtNS	EF186396
Hong Kong, China	*R. tanezumi*	RtHK	EF186440
Amami Island, Japan	*R. tanezumi*	RtAI	EF186393
Yogyakarta, Indonesia	*R. tanezumi*	RtYog	EF186378
Jakarta, Indonesia	*R. tanezumi*	RtJak	EF186376
New Zealand	*R. exulans*	RexNZ	AF104177
South Africa	*Mastomys coucha*	McSA	AY576896
Turkey	*Mus macedonicus*	MmTurk	AJ286327

2.2.5 Sequence data analyses

For both mitochondrial cytochrome *b* and D-loop sequences (Appendices 4 – 6), the ClustalW function contained within the BioEdit Sequence Alignment Editor version 7.0.4.1 (Hall, 1999) was used to produce a multiple alignment of sample and reference sequences which was further edited by eye. Once all sequences had been aligned, they were trimmed to the length of the shortest sequence. The aligned sequences were then input into ClustalX version 1.81 (Thompson *et al*, 1997), which was used to change file formats from Fasta (*.fas) used in BioEdit version 7.0.4.1 (Hall, 1999) to the Nexus file format (*.nxs) used in PAUP version 4.0 (Swofford, 2001). All the analyses were run in an Intel Celeron 1.6 GHZ Celeron M based personal computer with a Windows XP operating system.

Haplotype analyses were conducted using the computer programs DnaSP version 4.10.9 (Rozas *et al.*, 2003) and TCS version 1.6.0_02 (Clement *et al.*, 2000). Cytochrome *b* and D-loop sequence datasets were initially analyzed in DnaSP version 4.10.9 (Rozas *et al.*, 2003) to determine the number of haplotypes present and to produce a haplotype data file for further analysis. DnaSP version 4.10.9 (Rozas *et al.*, 2003) was further used to calculate haplotype and nucleotide diversity within the cytochrome *b* and D-loop sequence datasets. Once the number of haplotypes had been identified for the cytochrome *b* and D-loop datasets, the information was then input into TCS version 1.6.0_02 (Clement *et al.*, 2000) for the production of a haplotype network for each dataset.

Haplotype data files generated for the cytochrome *b* and D-loop sequence datasets were analyzed using the program MrModelTest version 2.2 (Nylander, 2004) to determine the most appropriate evolution models to be used in further analyses. Once suitable models were identified, the haplotype data files generated for the cytochrome *b* and D-loop sequence datasets were analyzed using PAUP version 4.0 (Swofford, 2001) and MrBayes version 3_0b4 (Ronquist and Huelsenbeck, 2003).

PAUP version 4.0 (Swofford, 2001) was utilized to generate a maximum parsimony tree from the haplotype data files obtained for the mitochondrial cytochrome *b* and D-loop sequence data sets, respectively. A distance matrix indicating the genetic distance values between haplotypes was also generated by PAUP version 4.0 (Swofford, 2001) via use of the genetic distance models recommended by MrModelTest version 2.2 (Nylander, 2004).

MrBayes version 3_0b4 (Ronquist and Huelsenbeck, 2003) was used to create phylogenetic

trees using Bayesian Inference under the model recommended by MrModelTest version 2.2 (Nylander, 2004). Phylogenetic trees created using the program MrBayes version 3_0b4 (Ronquist and Huelsenbeck, 2003) utilized 1 000 000 generations with burn-in values of 1000 and a sampling frequency of 100. Burn-in values were estimated by visual examination of probabilities in relation to number of generations during a preliminary run. The burn-in value was chosen based on the number of generations it took for the probabilities to plateau.

The reliability of nodes within phylogenetic trees created by PAUP version 4.0 (Swofford, 2001) was estimated by bootstrap re-sampling analysis (1000 iterations). In the case of the Bayesian analysis performed by MrBayes version 3_0b4 (Ronquist and Huelsenbeck, 2003), reliability of nodes was estimated as posterior probabilities.

2.3 Cytogenetic Analyses

The *Rattus* specimens listed in Section 2.1 (Table 1) were also used for cytogenetic analysis. Chromosome preparations for microscopic analysis were obtained either from direct harvesting of bone marrow or from fibroblast cultures established from ear biopsies.

2.3.1 Preparation of slides from bone marrow

The femur was removed from dead animals that were euthanized in accordance with the protocol approved by the Animal Ethics subcommittee of the University of KwaZulu-Natal Research Committee, i.e. by exposing these animals to an overdose of inhaled halothane. All excess tissue surrounding this bone was removed, after which an 18 gauge syringe needle was inserted through holes drilled into the upper and lower epiphyses. The bone marrow was removed from the femur by flushing it with 5 ml of a solution of DMEM 20 (Dulbecco's Modified Eagles' Medium supplemented with 20% v/v foetal bovine serum; Highveld Biological (Pty) Ltd., South Africa) containing 200 µl of 10 µg/ml colchicine (Sigma, Germany). Colcemid was included in this solution in order to inhibit the completion of the metaphase stage of mitosis within the bone marrow cells. The centrifuge tube containing the bone marrow was placed in an incubator at 37°C for 1 hour, after which it was centrifuged at 400 to 600 x g for five minutes using a MSE model Minor 35 centrifuge (MSE Scientific Instruments. Ltd., England).

The supernatant from this centrifugation step was decanted until at least 0.25 ml was left above

the centrifuged pellet. The pellet was then re-suspended in 5 ml of a 0.075 M hypotonic KCl solution and incubated for 20-27 minutes at 37°C. The solution was centrifuged at 400 to 600 g for five minutes and the supernatant decanted until at least 0.25 ml was left above the centrifuged pellet. The pellet was then re-suspended in 10 ml of Carnoy's fixative (absolute methanol:glacial acetic acid 3:1), added drop-wise. This suspension was centrifuged for five minutes, after which the supernatant was decanted until only 1 ml was left above the pellet. The pellet was then re-suspended in a further 10 ml of Carnoy's fixative, centrifuged again for five minutes, and the supernatant again decanted until only 1 ml was left above the pellet. The pellet was re-suspended in the remaining volume of supernatant and withdrawn with a Pasteur pipette, from which 2-3 drops were dropped onto a microscope slide from a height of 30 – 60 cm. The slides were heat dried by placing them into an incubator set at 65°C for 15-30 seconds, and then stained in a Copeland jar containing 3 % Giemsa (Merck (Pty) Ltd, South Africa), dissolved in deionised water, for five minutes. The slides were then rinsed in water and air-dried.

Slides were initially viewed with a Medilux-12 microscope (Kyowa Optical CO. Ltd, Japan) using the 20x objective to assess slide quality in terms of density of metaphase spreads per slide. Preparations with a sufficient number of spreads (> 5/slide) were then screened using a Vanox AHBS3 microscope (Olympus, Japan) to identify chromosome spreads suitable for recording. Such spreads were viewed with the oil immersion objective (100x) and captured digitally using a Snappy Snapshot image capture adapter (Play Incorporated, USA). These chromosome spreads were arranged into karyograms using the Open Source image manipulation program The Gimp version 2.2.12 (http://www.gimp.org) installed in a 1.6 GHZ Intel Celeron M based personal computer and a Windows XP operating system.

2.3.2 Skin fibroblast culture cells

Skin fibroblast cultures were established from a biopsy (approximately < 1 cm^2 in size) of disinfected (by wiping with 70% ethanol) ear or skin from the study organism. This biopsy was then placed in 70–100 % ethanol for approximately two minutes to disinfect the tissue. Thereafter, the biopsy was rinsed sequentially in three Petri dishes containing 2 ml of DMEM each. In the last Petri dish the tissue was minced into the smallest fragments possible using scissors and forceps. This suspension was transferred into a 15 ml Sarstedt sterile conical tube (Highveld Biological (Pty) Ltd., South Africa), centrifuged at 400 to 600 x g for five minutes and the supernatant discarded. Thereafter, 5 ml of a solution containing 0.25% w/v trypsin (Highveld Biological (Pty) Ltd., South Africa) and 0.02% EDTA in DMEM were added to the

recovered pellet of tissue fragments.

This suspension was then shaken with a Snijders vortex mixer (model 34524, Holland) for a total of 30 to 45 minutes at room temperature, centrifuged at 400 to 600 g for five minutes and the supernatant recovered. This supernatant was mixed in a sterile conical tube with 10 ml DMEM 20 supplemented with 200 unit/ml each of penicillin and streptomycin (Highveld Biological (Pty) Ltd., South Africa), centrifuged again for five minutes at the same speed and the supernatant discarded. The resulting cell-pellet was re-suspended in 5 ml DMEM 20 and transferred to a 25 cm^2 Cellstar tissue culture flask (Greiner Bio-One GmbH, Germany) which was then placed in a water-jacketed growth incubator (Forma Scientific, USA) supplied with a 5% CO_2 atmosphere at 37^oC. The culture flask was first inspected after 72 hours, using an inverted microscope (Olympus, Japan) at a 20x magnification, for evidence of cells, morphologically identifiable as fibroblasts, attached to the bottom. Thereafter, the culture was microscopically inspected daily for evidence of cellular growth, such as mitotic figures, and absence of microbial contamination. The growth medium (DMEM 20) was replaced every 48 to 72 hours.

Harvesting for the purpose of producing chromosome preparations was carried out shortly before the cells reached confluence. This time varied, according to specimen, from three to eight weeks.

Harvesting was initiated by adding 25 µl/ml of Colchicine (Sigma, Germany) to the culture flask, which was then incubated at 37°C for 45 minutes. The culture medium was then removed from the culture flask and the cells within the flask were rinsed with two changes each of 10 ml DMEM.

The DMEM was then replaced with 2 ml of 0.25% w/v trypsin (Highveld Biological (Pty) Ltd., South Africa); and the preparation incubated for three to five minutes at 37^oC and energetically tapped with the index finger to dislodge the cells from the bottom of the flask. This cell suspension was then transferred to a centrifuge tube containing 5 ml DMEM 20. The tube and contents were centrifuged for five minutes at 400 to 600 g, after which the supernatant was removed and replaced with 10 ml of a hypotonic 0.075 M KCl solution. After this step, the procedure was identical to the one described in Section 2.3.1 for the harvesting of bone marrow cells.

2.3.3 Banding techniques

2.3.3.1 G-Banding

G-Banding was performed one week after slide production. Slides were prepared for treatment by dividing the individual slides into three sections, as each division was exposed to the treatment at different times. Initial treatment of an individual slide was conducted by placing the slide into a Copeland jar containing 50 ml phosphate buffer (0.07 M Na_2HPO_4 combined 0.07 M KH_2PO_4 8:17; Merck, South Africa) pH 6.8, supplemented with 25 mg of trypsin powder (Sigma, Germany) and kept at a constant temperature of 37°C within a water bath. The first section of a slide being treated was placed into the above trypsin/phosphate buffer solution for 15 seconds, thereafter the slide was lowered so that the second section was immersed for 10 seconds, following which the final section was immersed for 5 seconds.

The slide was then removed from the trypsin/phosphate buffer solution and rinsed in a Copeland jar containing phosphate buffer supplemented with 20% v/v foetal bovine serum (Highveld Biological (Pty) Ltd., South Africa) for approximately two minutes. Thereafter, the slide was rinsed in a Copeland jar containing only phosphate buffer and stained in a Copeland jar containing a 3% Giemsa (Merck (Pty) Ltd, South Africa) solution (1.5 ml Giemsa stock solution in 50 ml phosphate buffer) for five minutes. Slides were then rinsed in water, air dried and inspected for banded chromosome spreads using the 20x objective of a Medilux-12 microscope (Kyowa Optical CO. Ltd, Japan) (Seabright, 1971).

After initial viewings, a Vanox AHBS3 microscope (Olympus, Japan) was used to identify suitable chromosome spreads. These were brought into focus using the oil immersion lens of the microscope and then digitally captured using the Snappy Snapshot image capture adapter (Play Incorporated, USA). The chromosome spreads were then arranged into karyograms using the Open Source image manipulation program The Gimp version 2.2.12 (http://www.gimp.org) installed in a 1.6 GHZ Intel Celeron M based personal computer and a Windows XP operating system.

2.3.3.2 C-Banding

C-Banding was performed after slide production. Slides were placed in a Copeland jar containing 0.2 N HCl (Merck (Pty) Ltd, South Africa) at room temperature for 20 minutes, rinsed in distilled water and air–dried. Thereafter, they were placed onto a slide rack over an open sink and flooded with a 0.1M BaOH solution (Merck (Pty) Ltd, South Africa) preheated to

a temperature of 65°C, for two to twelve minutes, after which they were thoroughly washed on each side with two litres of double deionised water, air dried and placed into a container lined with moistened tissue paper (Sumner *et al*., 1971).

Within the container, slides were flooded with 2x SSC buffer (0.3M NaCl added to 0.03M Sodium Citrate, pH 7.5) and placed in an incubator set at 65°C for two hours. They were then rinsed in water, air dried and stained in 3% Giemsa (Merck (Pty) Ltd, South Africa) solution (1.5 ml Giemsa stock solution in 50 ml phosphate buffer) for five minutes. Slides were then rinsed in water; air dried and inspected for banded chromosome spreads as described in Section 2.3.3.1.

CHAPTER 3

RESULTS

3.1 Molecular Analyses

3.1.1 Haplotype Analyses

Haplotype analysis using DnaSP version 4.10.9 (Rozas *et al.*, 2003) was conducted on both the 1146 nucleotide and 676 nucleotide cytochrome *b* datasets produced. For each dataset, it was observed that 11 haplotypes (six haplotypes belonging to field-identified *R. norvegicus* samples and five haplotypes belonging to field-identified *R. rattus*) (Table 10) were derived for the specimens used in this study. The 1146 nucleotide cytochrome *b* dataset contained 102 variable sites, whilst the 676 nucleotide cytochrome *b* dataset contained 89 varaiable sites. The mean genetic distance between the field-identified *R. norvegicus* and field-identified *R. rattus* haplotypes was 0.786 for both datasets.

The haplotype diversity of the 19 samples of field-identified *R. norvegicus* (forming 6 haplotypes) was 0.516 (with 6 variable sites) for both the 1146 nucleotide and 676 nucleotide cytochrome *b* datasets. The haplotype diversity of the 13 samples of field-identified *R. rattus* (forming 5 haplotypes) was 0.731 for both the 1146 nucleotide cytochrome *b* dataset (53 variable sites) and 676 nucleotide cytochrome *b* dataset (32 variable sites).

A consolidated statistical parsimony haplotype network was constructed in TCS version 1.6.0_02 (Clement *et al.*, 2000) (Figure 4) to illustrate mutational relationships between the *Rattus* cytochrome *b* haplotypes (Table 10).

Table 10: Mitochondrial cytochrome *b* haplotypes of South African field-identified *R. norvegicus* and *R. rattus* specimens obtained from the 1146 cytochrome *b* dataset (Haplotype No.[1]) and 676 cytochrome *b* dataset (Haplotype No.[2]).

Hapotype No.[1]	Hapotype No.[2]	No. Of Samples	Sample ID.	Species Field ID.	Origin
1	1	1	SA 75	*R. norvegicus*	Warwick Avenue, Durban CDB
2	2	15	SA 136 SA 141	*R. norvegicus*	Browns Road, Durban CBD
			SA 174 SA 175	*R. norvegicus*	View Street, Sydenham
			SA 213 SA 217	*R. norvegicus*	Cnr. Russell & St. George Street, Durban CBD
			SA 218 SA 219	*R. norvegicus*	Montclair Park, Montclair
			SA 232 SA 233	*R. norvegicus*	MaKhanyile CC1514, Cato Crest
			SA 234 SA 235	*R. norvegicus*	Cnr. Commercial & Soldiers Way, Durban CBD
			SA 252	*R. norvegicus*	145 Cato Maor Road, Cato Crest
			PS1	*R. norvegicus*	Park Station, Canelands, Verulam
			RnHua	*R. norvegicus*	Huahine, Society Islands, French Polynesia
3	3	2	SA 220 SA 221	*R. norvegicus*	Nzuza CC1504, Cato Crest
4	4	4	SA 253 SA 254	*R. tanezumi*	Pickford Estates, Shongweni
			ARC 79	*R. tanezumi*	Nkomo-B, Limpopo Province
			UPO2	*R. tanezumi*	Pretoria
5	5	10	SA 255 SA 256 SA 257 SA 258 SA 259	*R. rattus*	Axbridge Place, Widenham, Umkomaas
			ARC 170	*R. rattus*	Hammanskraal, Gauteng Province
			RrTit	*R. rattus*	Titirangi, New Zealand
			RrHua	*R. rattus*	Huahine, Society Islands, French Polynesia
			RrSam	*R. rattus*	Samoa
			RtJap	*R. tanezumi*	Japan
6	6	1	ARC 101	*R. rattus*	Nkomo-B, Limpopo Province
7	7	1	ARC 171	*R. rattus*	Hammanskraal, Gauteng Province
8	8	1	R 1	*R. tanezumi*	Richmond
9	9	1	HP1	*R. norvegicus*	Hung Yen Province, North Vietnam
10	10	1	HP2	*R. norvegicus*	Hung Yen Province, North Vietnam
11	11	1	S1	*R. norvegicus*	Subang, West Java, Indonesia
12	12	2	RnUSA RnRai	*R. norvegicus*	USA Raiatea, Society Islands, French Polynesia
13	13	1	RnCopen	*R. norvegicus*	Copenhagen, Denmark
14	14	1	RnChina	*R. norvegicus*	China
15	15	1	RrJap	*R. rattus*	Japan
	16	1	RtAI	*R. tanezumi*	Amami Islands, Japan

Hapotype No.[1]	Hapotype No.[2]	No. Of Samples	Sample ID.	Species Field ID.	Origin
	17	1	RtYog	*R. tanezumi*	Yogyakarta, Indonesia
	18	1	RtJak	*R. tanezumi*	Jakarta, Indonesia
16	19	1	RtPhil	*R. tanezumi*	Philippines
17	20	1	RaJap	*R. argentiventer*	Japan
18	21	1	RevPhil	*R. everetti*	Philippines
19	22	1	RexPhil	*R. exulans*	Philippines
20	23	1	RpPhil	*R. praetor*	Philippines
21	24	1	DlJap	*D. legata*	Japan
22	25	1	LbPhil	*L. bryophilus*	Philippines
23	26	1	TaPhil	*T. apoensis*	Philippines
24	27	1	MmdChina	*M. musculus domesticus*	China
	28	1	RnHai	*R. norvegicus*	Haiphong Port, Vietnam
	29	1	RrMoi	*R. rattus*	Moitaka, Papua New Guinea
	30	2	RrKL	*R. rattus*	Kuala Lumpur, Malaysia
			RtNS	*R. tanezumi*	Northern Sulawesi, Indonesia
25		1	RnRai	*R. norvegicus*	Raiatea, Society Islands, French Polynesia

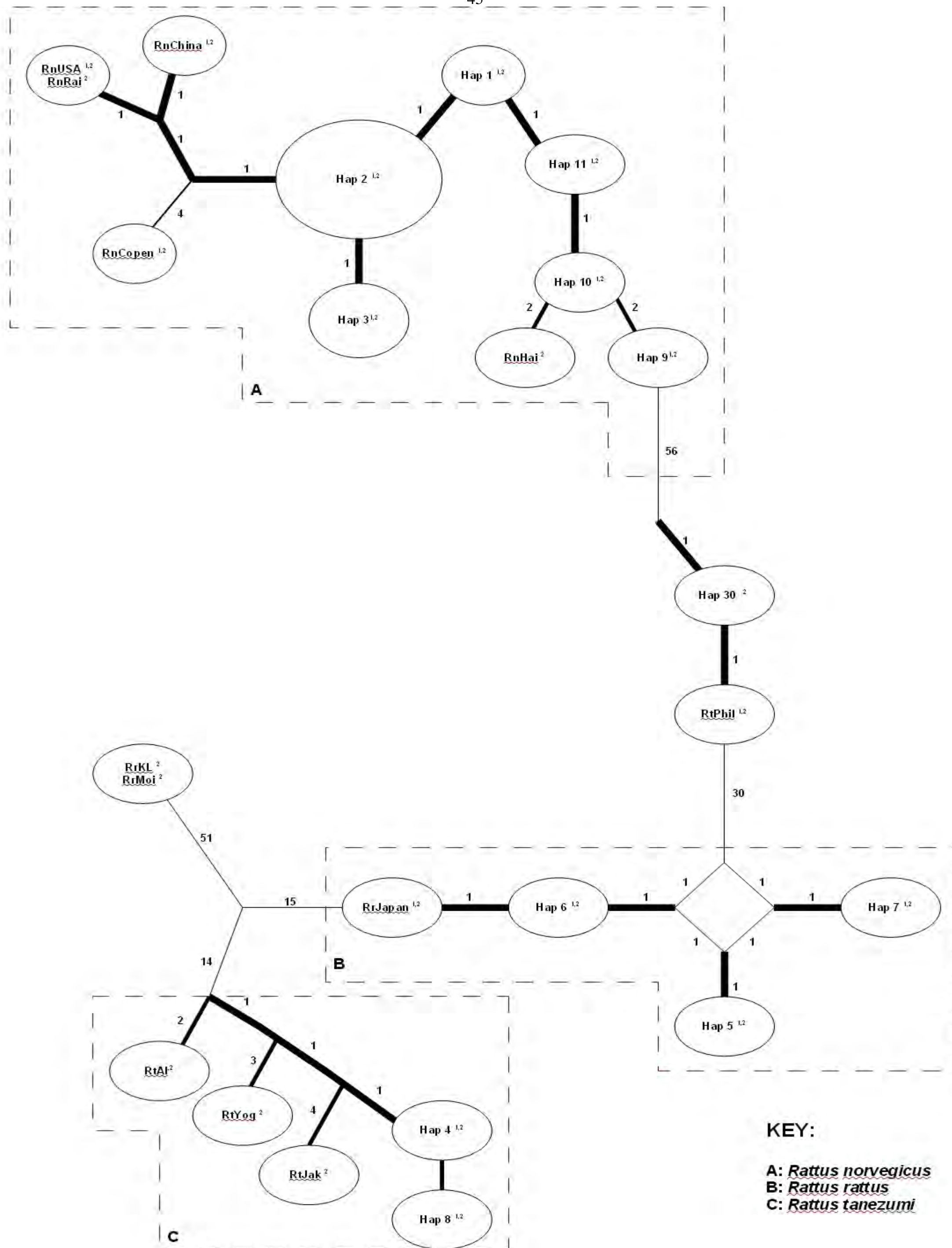

Figure 4: Consolidated statistical parsimony haplotype network showing the relationship between the haplotypes identified from the 1146 cytochrome *b* dataset ([1]) and 676 cytochrome *b* dataset ([2]) dataset for South African *Rattus* samples (Haps 1 to 11, Table 10) and those obtained from NCBI-GenBank. Connections between haplotypes demarcated by thick lines and less than 10 mutational steps indicate 95% parsimony connections, whilst connections demarcated by thin lines and greater than 10 mutational steps indicate < 95% parsimony

The consolidated cytochrome *b* haplotype network (Figure 4) contained three distinct sub-networks when TCS version 1.6.0_02 (Clement *et al.*, 2000) was set at a 95% parsimony limit; the six haplotypes (1, 2, 3, 9, 10 and 11) derived for the field-identified *R. norvegicus* specimens formed a *R. norvegicus* sub-network with NCBI-Genbank derived *R. norvegicus* haplotypes (RnUSA, RnRai, RnChina, RnCopen, RnHua), the five haplotypes derived for the field-identified *R. rattus* specimens (4, 5, 6, 7and 8) formed the remaining two sub-networks with haplotypes 4 and 8 forming a *R. tanezumi* sub-network with NCBI-Genbank derived *R. tanezumi* haplotypes (RtAI, RtYog and RtJak) and haplotypes 5, 6 and 7 forming an *R. rattus* sub-network with a NCBI-Genbank derived *R. rattus* haplotype (RrJapan). The program was then set to 100 steps to join all sub-networks into a single network (Figure 4).

The sub-network comprising the *R. norvegicus* haplotypes was observed to separate from the sub-networks containing the *R. rattus* and *R. tanezumi* haplotypes by a minimum of 88 and 117 mutational steps, respectively, whilst the *R. rattus* sub-network separated from the *R. tanezumi* sub-network by a minimum of 28 mutational steps. Within the *R. norvegicus* sub-network, all the haplotypes were separated from one another by a single mutation step with the exception of Hap 9, which was separated from Hap 10 by two mutational steps. Within the *R. tanezumi* sub-network haplotypes were separated from one another by a minimum of one to seven mutational steps, and within the *R. rattus* sub-network haplotypes were separated by between one and five mutational steps.

Haplotype analysis using DnaSP version 4.10.9 (Rozas *et al.*, 2003) was conducted on 320 nucleotides of the D-loop region of the mtDNA extracted from the South African field-identified *R. norvegicus* and *R. rattus* samples. This analysis revealed 7 haplotypes, four from *R. norvegicus* genotypes and samples and three from *R. rattus* genotypes (Table 11). 82 sites were variable over the 320 nucleotides of the D-loop region examined. The mean genetic distance between the field-identified *R. norvegicus* and field-identified *R. rattus* haplotypes was 0.752.

For 14 samples of *R. norvegicus* (forming four haplotypes) the haplotype diversity was 0.526 (with 12 variable sites), whilst the seven samples of field-identified *R. rattus* (forming three haplotypes) had a haplotype diversity of 0.810, with 27 variable sites over the 320 nucleotides of the D-loop region..

A statistical parsimony haplotype network constructed in TCS version 1.6.0_02 (Clement *et al.*, 2000) from 320 nucleotides of the mitochondrial D-loop region (see Table 11) is presented in

Figure 5.

Table 11: Mitochondrial D-loop haplotypes of South African field-identified *R. norvegicus* and *R. rattus* specimens and outgroups downloaded from the NCBI GenBank. Refer to Tables 1, 8, 9 and 10 for information regarding sample origin and species names.

Hapotype No.	No. Of Samples	Sample ID.	Species Field ID.	Origin
1	12	SA 141	*R. norvegicus*	Browns Road, Durban CBD
		SA 75	*R. norvegicus*	Warwick Avenue, Durban CDB
		SA 174 SA 175	*R. norvegicus*	View Street, Sydenham
		SA 221	*R. norvegicus*	Nzuza CC1504, Cato Crest
		SA 233	*R. norvegicus*	MaKhanyile CC1514, Cato Crest
		SA 234	*R. norvegicus*	Cnr. Commercial & Soldiers Way, Durban CBD
		SA 252	*R. norvegicus*	145 Cato Maor Road, Cato Crest
		PS1	*R. norvegicus*	Park Station, Canelands, Verulam
		S1	*R. norvegicus*	Subang, West Java, Indonesia
		RnUSA	*R. norvegicus*	USA
		RnHua	*R. norvegicus*	Huahine, Society Islands, French Polynesia
2	2	SA 213	*R. norvegicus*	Cnr. Russell & St. George Street, Durban CBD
		SA 218	*R. norvegicus*	Montclair Park, Montclair
3	3	SA 253 SA 254	*R. tanezumi*	Pickford Estates, Shongweni
		R 1	*R. tanezumi*	Richmond
4	1	SA 255	*R. rattus*	Axbridge Place, Widenham, Umkomaas
5	7	SA 256 SA 258 SA 259	*R. rattus*	Axbridge Place, Widenham, Umkomaas
		RrUSA	*R. rattus*	USA
		RrTit	*R. rattus*	Titirangi, New Zealand
		RrHua	*R. rattus*	Huahine, Society Islands, French Polynesia
		RrSam	*R. rattus*	Samoa
6	1	HP1	*R. norvegicus*	Hung Yen Province, North Vietnam
7	1	HP2	*R. norvegicus*	Hung Yen Province, North Vietnam
8	1	RnItaly	*R. norvegicus*	Italy
9	1	RnCopen	*R. norvegicus*	Copenhagen, Denmark
10	1	RnRai	*R. norvegicus*	Raiatea, Society Islands, French Polynesia
11	1	RrMad	*R. rattus*	Madagascar
12	1	RrMoi	*R. rattus*	Moitaka, Papua New Guinea
13	2	RrKL	*R. rattus*	Kuala Lumpur, Malaysia
		RtNS	*R. tanezumi*	Northern Sulawesi, Indonesia
14	1	RtYog	*R. tanezumi*	Yogyakarta, Indonesia
15	3	RtAI	*R. tanezumi*	Amami Islands, Japan
		RtHK	*R. tanezumi*	Hong Kong, China
		RtJak	*R. tanezumi*	Jakarta, Indonesia
16	1	RexNZ	*R. exulans*	New Zealand
17	1	RfAust	*R. fuscipes*	Australia
18	1	MmTurk	*M. macedonicus*	Turkey
19	1	McSA	*M. coucha*	South Africa

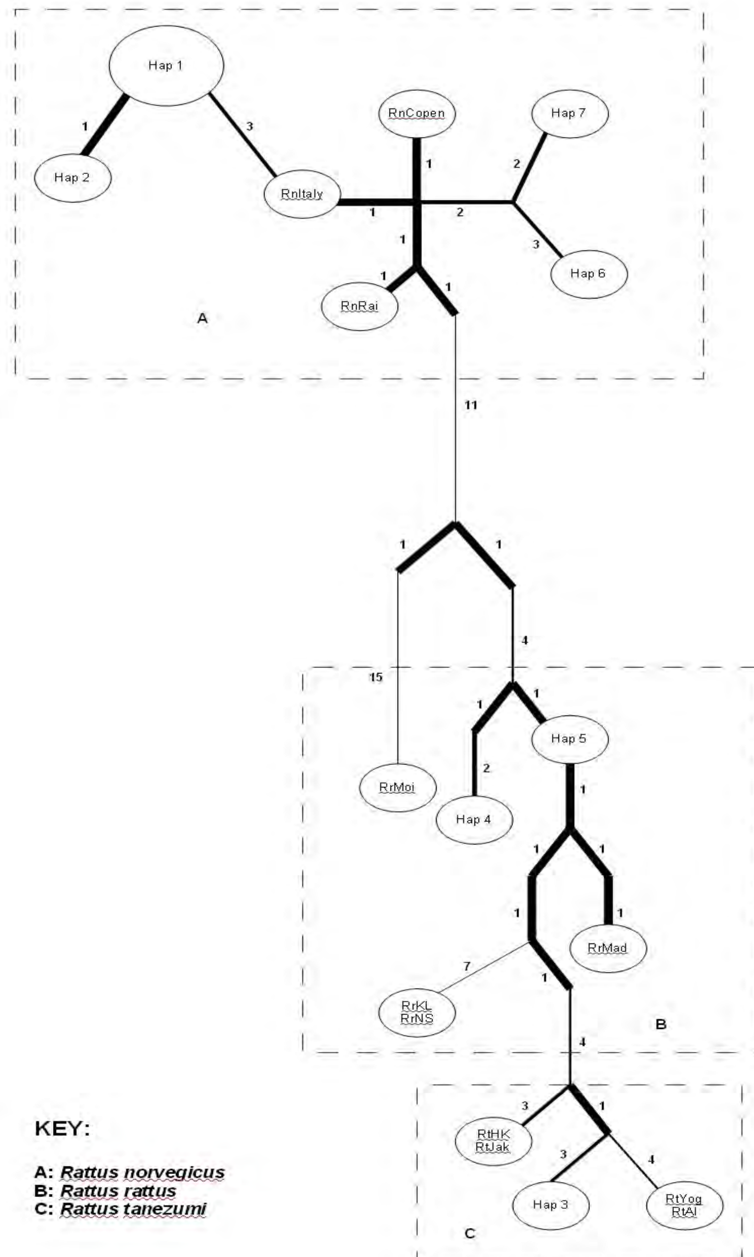

Figure 5: Statistical parsimony haplotype network showing the mutational relationships between *Rattus* mitochondrial D-loop haplotypes. Codes: Haps 1, 4, 6, 9, 10, 11 and 12 refers to haplotypes presented in Table 11. Connections between haplotypes demarcated by thick lines and less than 10 mutational steps indicate 95% parsimony connections, whilst connections demarcated by thin lines and greater than 10 mutational steps indicate < 95% parsimony connections.

The D-loop haplotype network (Figure 5), constructed by treating gaps as missing data in TCS version 1.6.0_02 (Clement *et al.*, 2000), produced three distinct sub-networks when set at a 95% parsimony limit; the four haplotypes (1, 2, 6 and 7) derived for the field-identified *R. norvegicus* specimens formed an *R. norvegicus* sub-network with NCBI-Genbank derived *R. norvegicus* haplotypes (RnItaly, RnRai, RnCopen), the three haplotypes derived for the field-identified *R. rattus* specimens (3, 4 and 5) formed the remaining two sub-networks with haplotypes 4 and 5 forming a *R. rattus* sub-network with NCBI-Genbank derived *R. rattus* haplotypes (RrMoi, RrMad, RrKl and RrNS) and haplotype 3 forming a *R. tanezumi* sub-network with NCBI-Genbank derived *R.tanezumi* haplotypes (RtHK, RtJak, RtYog and RtAI). The connection limit was then set to 50 steps to join all sub-networks into a single network (Figure 5). The sub-network comprising the *R. norvegicus* haplotypes (1, 2, 6 and 7) was separated from the *R. tanezumi* sub-network by a minimum of 26 mutaional steps, and from *R. rattus* sub-network by a minimum of 19 mutational steps. The *R. rattus* sub-network separated from the haplotypes of *R. tanezumi* sub-network by a minimum of 5 mutational steps. The haplotypes comprising the *R. norvegicus* sub-network, were separated from each other by between one and three mutational steps, whilst the haplotypes that comprised both the *R. rattus* and *R. tanezumi* sub-networks, were separated from each other by between one and seven mutational steps.

3.1.2 Genetic Distance Analyses

The GTR model was identified in both Modeltest version 3.7 (Posada and Crandall, 1998) and MrModelTest version 2.2 (Nylander, 2004) as the best-fitting both the 1146 and 676 nucleotide cytochrome *b* datasets. This model was used where appropriate in further analyses.

Table 12: Pairwise GTR genetic distances based on 1146 nucleotides of the mitochondrial cytochrome *b* region between *Rattus* study haplotype groups (Haps 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11), NCBI-GenBank derived haplotype groups (Haps 12, 13, 14, 15 , 16, 17, 18, 19, 20, 21, 22, 23, 25) and outgroups (Hap 24).

Haplotype Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	-																								
2	0.005	-																							
3	0.009	0.005	-																						
4	0.865	0.875	0.934	-																					
5	0.695	0.704	0.739	0.157	-																				
6	0.686	0.695	0.733	0.165	0.028	-																			
7	0.671	0.680	0.713	0.165	0.013	0.034	-																		
8	0.911	0.920	0.989	0.009	0.170	0.179	0.178	-																	
9	0.018	0.023	0.027	0.750	0.614	0.605	0.596	0.785	-																
10	0.009	0.013	0.018	0.822	0.663	0.655	0.642	0.864	0.009	-															
11	0.004	0.009	0.013	0.856	0.686	0.678	0.663	0.901	0.013	0.004	-														
12	0.018	0.013	0.018	0.916	0.731	0.722	0.707	0.967	0.027	0.018	0.013	-													
13	0.028	0.023	0.028	0.813	0.666	0.656	0.647	0.853	0.037	0.028	0.023	0.028	-												
14	0.018	0.014	0.018	0.896	0.697	0.688	0.674	0.945	0.028	0.018	0.014	0.009	0.030	-											
15	0.663	0.672	0.707	0.171	0.033	0.004	0.039	0.186	0.586	0.633	0.655	0.697	0.636	0.665	-										
16	0.578	0.586	0.610	0.221	0.200	0.188	0.195	0.234	0.519	0.554	0.571	0.612	0.565	0.583	0.195	-									
17	0.843	0.854	0.905	0.496	0.515	0.459	0.515	0.517	0.803	0.883	0.833	0.821	0.852	0.896	0.448	0.527	-								
18	0.896	0.872	0.910	0.528	0.652	0.627	0.678	0.551	0.859	0.859	0.884	0.980	0.935	0.904	0.612	0.523	0.837	-							
19	0.899	0.867	0.920	0.465	0.426	0.411	0.428	0.486	0.780	0.854	0.890	0.914	0.904	0.865	0.424	0.411	0.532	0.617	-						
20	0.920	0.935	0.985	0.654	0.654	0.651	0.683	0.688	0.877	0.947	0.902	0.915	1.233	0.863	0.635	0.734	0.916	6.713	0.528	-					
21	0.820	0.835	0.881	0.619	0.538	0.524	0.576	0.647	0.728	0.785	0.811	0.839	0.835	0.796	0.544	0.612	0.671	0.579	0.668	0.542	-				
22	0.671	0.659	0.676	0.575	0.713	0.648	0.749	0.603	0.646	0.646	0.660	0.635	0.963	0.598	0.671	0.588	0.800	0.589	0.464	0.574	0.718	-			
23	0.804	0.781	0.819	0.486	0.513	0.505	0.548	0.508	0.714	0.770	0.796	0.839	0.781	0.791	0.491	0.594	0.988	0.564	0.582	0.592	0.442	0.436	-		
24	1.390	1.407	1.529	1.609	3.356	6.713	6.713	1.737	1.398	1.398	1.390	1.521	1.599	1.402	6.713	1.358	2.455	6.713	6.713	6.713	1.066	1.377	1.152	-	
25	0.018	0.013	0.018	0.831	0.672	0.664	0.650	0.872	0.018	0.009	0.013	0.018	0.028	0.018	0.642	0.560	0.901	0.821	0.834	0.977	0.791	0.647	0.742	1.362	-

Table 13: Pairwise GTR genetic distances based on 676 nucleotides of the mitochondrial cytochrome *b* region between *Rattus* study haplotype groups (Haps 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11), NCBI-GenBank derived haplotype groups (Haps 12, 13, 14, 15 , 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30) and outgroups (Hap 27)

Haplotype Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	-																													
2	0.004	-																												
3	0.009	0.004	-																											
4	0.847	0.857	0.933	-																										
5	0.670	0.679	0.724	0.147	-																									
6	0.625	0.635	0.675	0.156	0.014	-																								
7	0.644	0.653	0.694	0.155	0.014	0.020	-																							
8	0.904	0.914	1.006	0.010	0.160	0.170	0.169	-																						
9	0.019	0.025	0.030	0.716	0.579	0.541	0.560	0.757	-																					
10	0.009	0.014	0.019	0.799	0.635	0.593	0.611	0.850	0.009	-																				
11	0.004	0.009	0.014	0.838	0.660	0.616	0.634	0.894	0.014	0.004	-																			
12	0.019	0.014	0.019	0.905	0.711	0.662	0.684	0.971	0.030	0.019	0.014	-																		
13	0.030	0.025	0.030	0.786	0.637	0.595	0.616	0.832	0.040	0.030	0.025	0.031	-																	
14	0.020	0.015	0.020	0.882	0.673	0.628	0.648	0.944	0.030	0.020	0.015	0.009	0.032	-																
15	0.602	0.611	0.649	0.163	0.019	0.004	0.025	0.178	0.522	0.571	0.593	0.637	0.575	0.604	-															
16	0.934	0.947	1.038	0.025	0.153	0.165	0.162	0.037	0.786	0.879	0.923	1.017	0.877	0.948	0.172	-														
17	0.980	0.992	1.097	0.025	0.153	0.163	0.162	0.037	0.818	0.920	0.970	1.066	0.907	0.991	0.170	0.031	-													
18	0.890	0.900	0.981	0.031	0.161	0.170	0.169	0.043	0.753	0.839	0.879	0.969	0.842	0.905	0.178	0.035	0.048	-												
19	0.546	0.554	0.580	0.236	0.202	0.196	0.195	0.250	0.483	0.521	0.538	0.586	0.530	0.544	0.204	0.258	0.260	0.269	-											
20	0.784	0.795	0.848	0.448	0.464	0.429	0.462	0.469	0.743	0.824	0.773	0.767	0.806	0.863	0.418	0.441	0.449	0.503	0.489	-										
21	0.895	0.870	0.909	0.472	0.568	0.560	0.594	0.494	0.855	0.855	0.880	4.718	4.718	1.082	0.544	0.508	0.500	0.529	0.480	0.743	-									
22	1.085	1.014	1.135	0.453	0.412	0.383	0.416	0.476	0.876	1.008	1.075	1.101	1.074	1.018	0.397	0.451	0.491	0.461	0.406	0.559	0.555	-								
23	0.994	1.016	1.098	0.673	0.672	0.640	0.713	0.712	0.940	1.046	0.974	0.997	4.718	0.927	0.622	0.665	0.734	0.742	0.803	0.909	4.718	0.521	-							
24	0.793	0.806	0.850	0.604	0.484	0.486	0.521	0.634	0.703	0.758	0.783	0.802	4.718	0.754	0.507	0.662	0.657	0.656	0.609	4.718	0.495	0.689	0.545	-						
25	0.794	0.779	0.800	0.585	0.725	0.668	0.774	0.618	0.755	0.755	0.771	0.704	4.718	0.645	0.696	0.575	0.636	0.592	0.624	0.801	0.560	0.496	0.568	4.718	-					
26	0.832	0.803	0.850	0.483	0.488	0.488	0.525	0.507	0.727	0.792	0.822	0.946	0.891	0.856	0.473	0.497	0.491	0.478	0.619	0.993	0.509	0.554	0.617	0.456	0.458	-				
27	1.248	1.258	1.372	1.328	1.735	2.359	1.971	1.414	1.251	1.251	1.239	1.352	1.426	1.252	4.718	1.546	1.583	1.427	1.195	1.810	1.714	1.674	4.718	1.002	1.175	0.989	-			
28	0.019	0.014	0.019	0.809	0.645	0.603	0.620	0.858	0.019	0.009	0.014	0.019	0.030	0.020	0.581	0.913	0.930	0.851	0.526	0.843	0.806	0.960	1.089	0.757	0.767	0.757	1.223	-		
29	0.779	0.749	0.789	0.547	0.629	0.598	0.612	0.578	0.74	0.809	0.765	0.817	0.816	0.763	0.579	0.534	0.555	0.551	0.827	0.638	0.747	0.549	0.407	0.839	0.735	0.613	4.718	0.786	-	
30	0.528	0.537	0.562	0.246	0.211	0.204	0.204	0.26	0.468	0.504	0.521	0.568	0.514	0.536	0.213	0.269	0.271	0.28	0.004	0.508	0.501	0.421	0.774	0.609	0.606	0.654	1.191	0.509	0.904	-

Pairwise comparison of the six *R. norvegicus* haplotypes identified in this study (1 to 3 and 9 to 11, Table 10) (Table 12 and 14) yielded genetic distance values ranging from 0.004 (0.4%) to 0.027 (2.7%) with a mean of 0.012 (1.2%). Pairwise comparison of these six *R. norvegicus* haplotypes to NCBI GenBank-derived *R. norvegicus* haplotypes (12, 13, 14, Table 11, 13, 14 and 28, Table 11) yielded similar values (minimum distance 0.013 (1.3%), maximum 0.037 (3.7%), mean 0.021 (2.1%)).

R. tanezumi haplotypes (4 and 8) (Figure 4), were separated by a genetic distance of 0.009 (0.9%) in the 1146 nucleotide cytchrome *b* dataset (Table 13) and by 0.010 (1%) in the 676 nucleotide cytchrome *b* dataset (Table 13), whereas *R. rattus* haplotypes (5, 6, 7) (Figure 4), were separated by 0.025 (2.5%) (Table 13) and 0.016 (1.6%) (Table 14). The mean genetic distance (combination of values from Table 12 and 14) between the *R. rattus* clade and *R. tanezumi* clade was 0.164 (16.9%).

Pairwise comparison of haplotypes 4 to 8 to the NCBI GenBank-derived *Rattus* haplotypes (haplotypes 12 to 19 and 25 in Table 10, haplotypes 12 to 23 and 28 to 30 in Table 10) revealed an apparent split within these *R. rattus* study haplotypes.

Haplotypes 4 and 8 were observed to be less distant from the *R. tanezumi* reference haplotypes (haplotypes 16 to 19 in Table 10) (mean genetic distance 0.084 (8.4%), range 0.025 – 0.250 (2.5% - 25%)) than from the *R. rattus* reference haplotype (15 in Table 10) (mean genetic distance 0.328 (32.8%). Similarly, haplotypes 5, 6 and 7 were more similar to the *R. rattus* reference haplotype (15, Table 10) (mean genetic distance of 0.025 (2.5%) and 0.016 (1.6%)) than they were to the *R. tanezumi* reference haplotypes (16, Table 10 and 16 to 19, Table 10).

Comparison of the six *R. norvegicus* haplotypes (1 to 3 and 9 to 11, Table 10) to the five *R. rattus* haplotypes identified in this study (4 to 8, Table 10) yielded genetic distance values ranging from a minimum of 0.596 (59.6%) to a maximum of 0.989 (98.9%) with a mean of 0.836 (83.6%).

Analysis of the D-loop dataset using Modeltest version 3.7 (Posada and Crandall, 1998), identified the HKY model as best fitting the data. The haplotype groups presented in Table 11 and outgroups were analyzed using this model to produce a HKY genetic distance matrix (Table 14).

Pairwise comparison of the four *R. norvegicus* haplotypes identified in this study (1, 2 6 and 7, Table 11) (Table 14) yielded genetic distance values ranging from 0.012 (1.2%) to 0.108 (10.8%) with a mean of 0.077 (7.7%). Pairwise comparison of these four *R. norvegicus* haplotypes to NCBI GenBank-derived *R. norvegicus* haplotypes (8, 9 and 10 Table 11) yielded similar values (minimum distance 0.038 (3.8%), maximum 0.092 (9.2%), mean 0.131 (13.1%)).

Pairwise comparison of the three field-identified *R. rattus* haplotypes identified in this study (3, 4 and 5, Table 11) (Table 14) yielded genetic distances between 0.052 (5.2%) and 0.185 (18.5%) with a mean of 0.158 (15.8%). These three haplotypes formed two clades in the D-loop haplotype network (Figure 5). *R. tanezumi* comprised haplotype 3 only (Figure 5), whilst *R. rattus* comprised haplotypes 4 and 5 (Figure 5), separated by a genetic distance of 0.052 (5.2%) (Table 14). The mean genetic distance between Clades 1 and 2 was 0.211 (21.1%).

Pairwise comparison of haplotypes 3, 4 and 5 to the NCBI GenBank-derived *Rattus* haplotypes (9, 10, 11, 12, 13, 14, 15, 16 and 17) (Table 11) showed that haplotype 3 was more similar to the *R. tanezumi* reference haplotypes (14 and 15, Table 11) (mean genetic distance 0.095 (9.5%)) than to the *R. rattus* reference haplotypes (11, 12 and 13, Table 11) (mean genetic distance of 0.291 (29.1%)).

Haplotypes 4 and 5 were more similar to the *R. rattus* reference haplotype (11, Table 11) (mean genetic distance 0.071 (7.1%)) than to the *R. tanezumi* reference haplotypes (14 and 15, Table 11) (mean genetic distance 0.237 (23.7%)).

Pairwise comparison of the four *R. norvegicus* (1 and 2, 6 and 7, Table 11) to the three *R. rattus* haplotypes identified in this study (3, 4 and 5, Table 11) yielded genetic distances between 0.308 (30.8%) and 0.468 (46.8%), with a mean of 0.363 (36.3%).

Table 14: Pairwise HKY genetic distances based on 320 nucleotides of the mitochondrial D-loop region between *Rattus* study haplotype groups (Haps 1, 2, 3, 4, 5, 6, 7), NCBI-GenBank derived haplotype groups (Haps 8, 9, 10, 11, 12, 13, 14, 15, 16, 17) and outgroups (Haps 18 and 19).

Haplotype Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-																		
2	0.012	-																	
3	0.350	0.375	-																
4	0.331	0.356	0.237	-															
5	0.308	0.330	0.185	0.052	-														
6	0.092	0.108	0.468	0.393	0.370	-													
7	0.079	0.094	0.398	0.353	0.328	0.064	-												
8	0.038	0.052	0.396	0.328	0.306	0.078	0.064	-											
9	0.067	0.082	0.398	0.329	0.307	0.078	0.064	0.025	-										
10	0.083	0.099	0.429	0.308	0.288	0.092	0.079	0.038	0.038	-									
11	0.330	0.355	0.209	0.102	0.040	0.447	0.405	0.375	0.376	0.354	-								
12	0.504	0.542	0.491	0.353	0.328	0.580	0.534	0.530	0.532	0.504	0.405	-							
13	0.289	0.311	0.174	0.221	0.144	0.396	0.318	0.355	0.356	0.388	0.164	0.415	-						
14	0.427	0.460	0.094	0.323	0.222	0.568	0.524	0.484	0.487	0.530	0.252	0.455	0.227	-					
15	0.407	0.440	0.095	0.220	0.183	0.475	0.434	0.462	0.464	0.511	0.212	0.503	0.235	0.117	-				
16	0.490	0.530	0.346	0.383	0.310	0.528	0.453	0.482	0.485	0.527	0.370	0.487	0.425	0.333	0.312	-			
17	0.475	0.504	0.601	0.512	0.481	0.550	0.501	0.447	0.447	0.475	0.511	0.309	0.588	0.636	0.704	0.596	-		
18	0.660	0.698	0.571	0.624	0.585	0.855	0.776	0.689	0.692	0.737	0.626	0.822	0.607	0.634	0.642	0.779	1.001	-	
19	0.459	0.483	0.604	0.525	0.465	0.593	0.534	0.532	0.533	0.567	0.494	0.750	0.489	0.682	0.614	0.815	0.902	0.630	-

3.1.3 Phylogenetic Analyses

Analyses conducted on the cytochrome *b* and D-loop haplotype datasets resulted in the production of two cladograms for each of the sequence datasets, viz. a maximum parsimony phylogenetic tree (Figures 6, 7 and 10) and a Bayesian Inference phylogenetic tree (Figures 8, 9 and 11).

Analysis of the cytochrome *b* and D-loop haplotype datasets in MrModelTest version 2.2 (Nylander, 2004) identified the GTR model as the best-fitting for the cytochrome *b* haplotype dataset and the HKY model as best-fitting the D-loop haplotype dataset.

Bayesian likelihood trees (Figures 8, 9 and 11) were created using MrBayes version 3_0b4 (Ronquist and Huelsenbeck, 2003) using the parameters set by MrModelTest version 2.2 (Nylander, 2004) for both the cytochrome *b* and D-loop haplotype datasets.

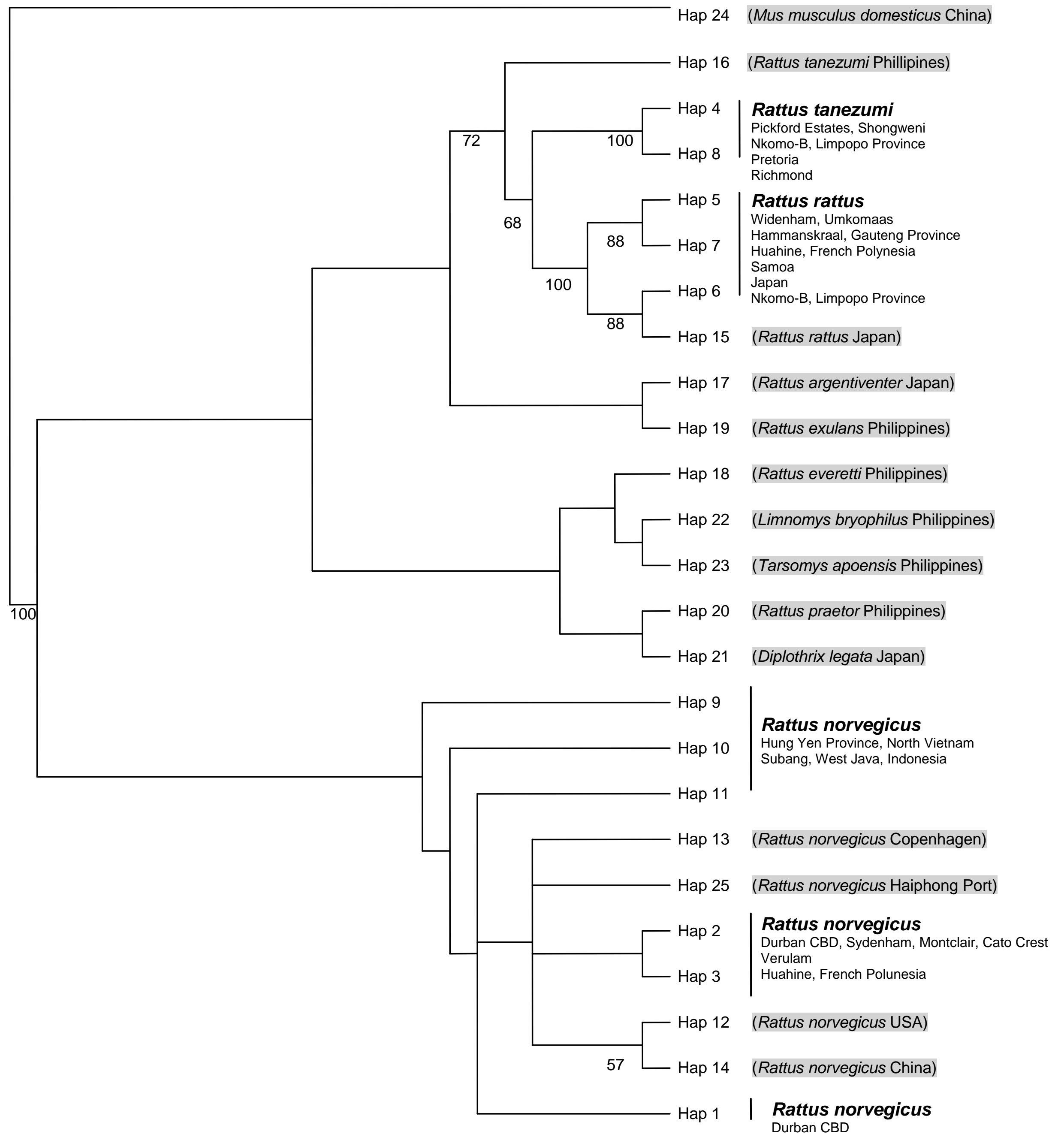

Figure 6: Maximum Parsimony tree showing evolutionary relationships between samples based on 1146 nucleotides of the mitochondrial cytochrome *b* sequence. Bootstrap values are indicated adjacent to the nodes. Tree rooted to the outgroup *Mus musculus domesticus (*Hap 24). Labels in bold indicate species identified in this study.

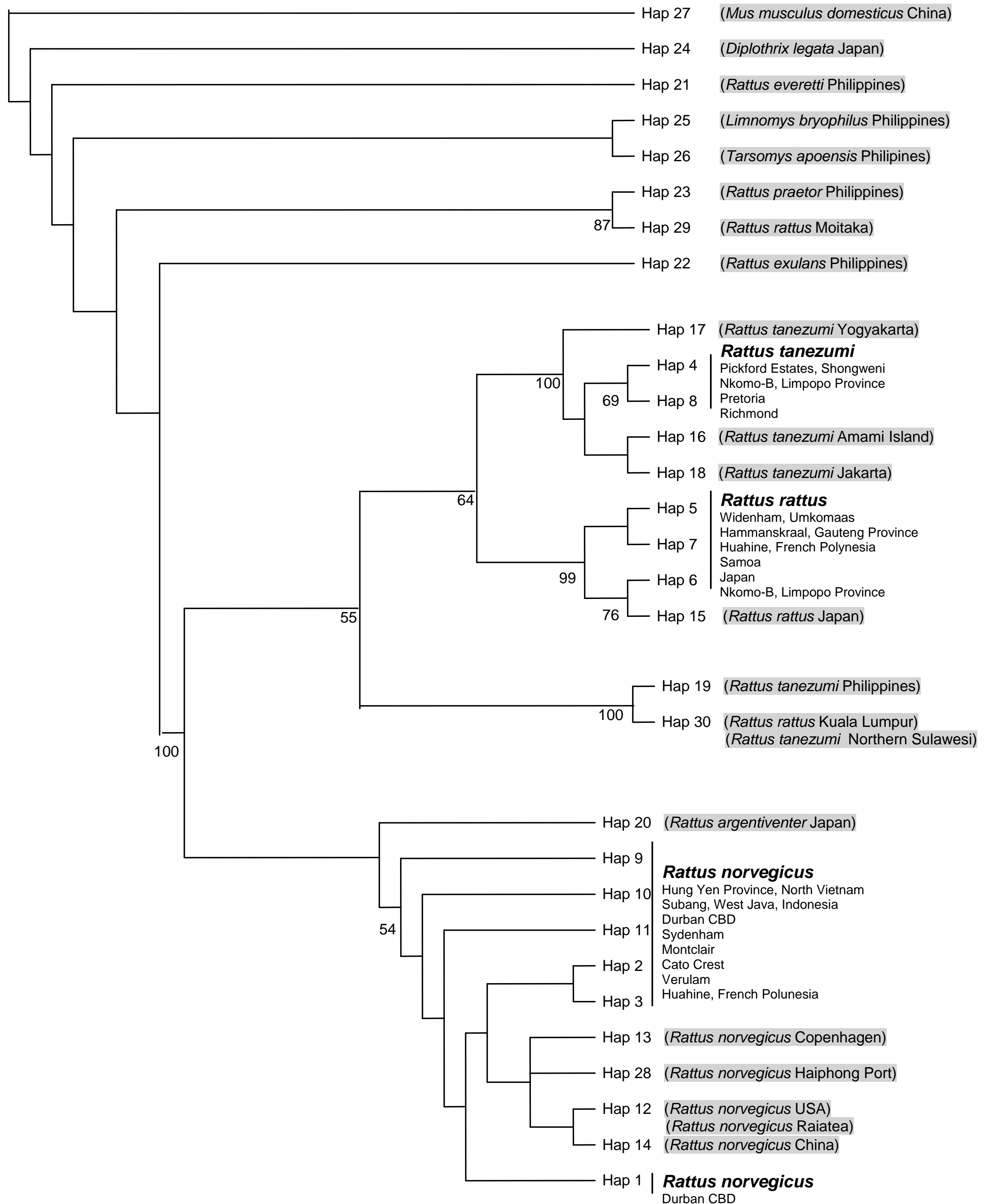

Figure 7: Maximum Parsimony tree showing evolutionary relationships between samples based on 676 nucleotides of the mitochondrial cytochrome *b* sequence. Bootstrap values are indicated adjacent to the nodes. Tree rooted to the outgroup *Mus musculus domesticus (*Hap 27). Labels in bold indicate species identified in this study.

Figure 8: Bayesian phylogenetic tree obtained from the analysis of 1146 nucleotide cytochrome *b* dataset using the GTR genetic distance model. Posterior probabilities are indicated adjacent to the nodes. Tree rooted to the outgroup *Mus musculus domesticus (*Hap 24). Labels in bold indicate species identified in this study.

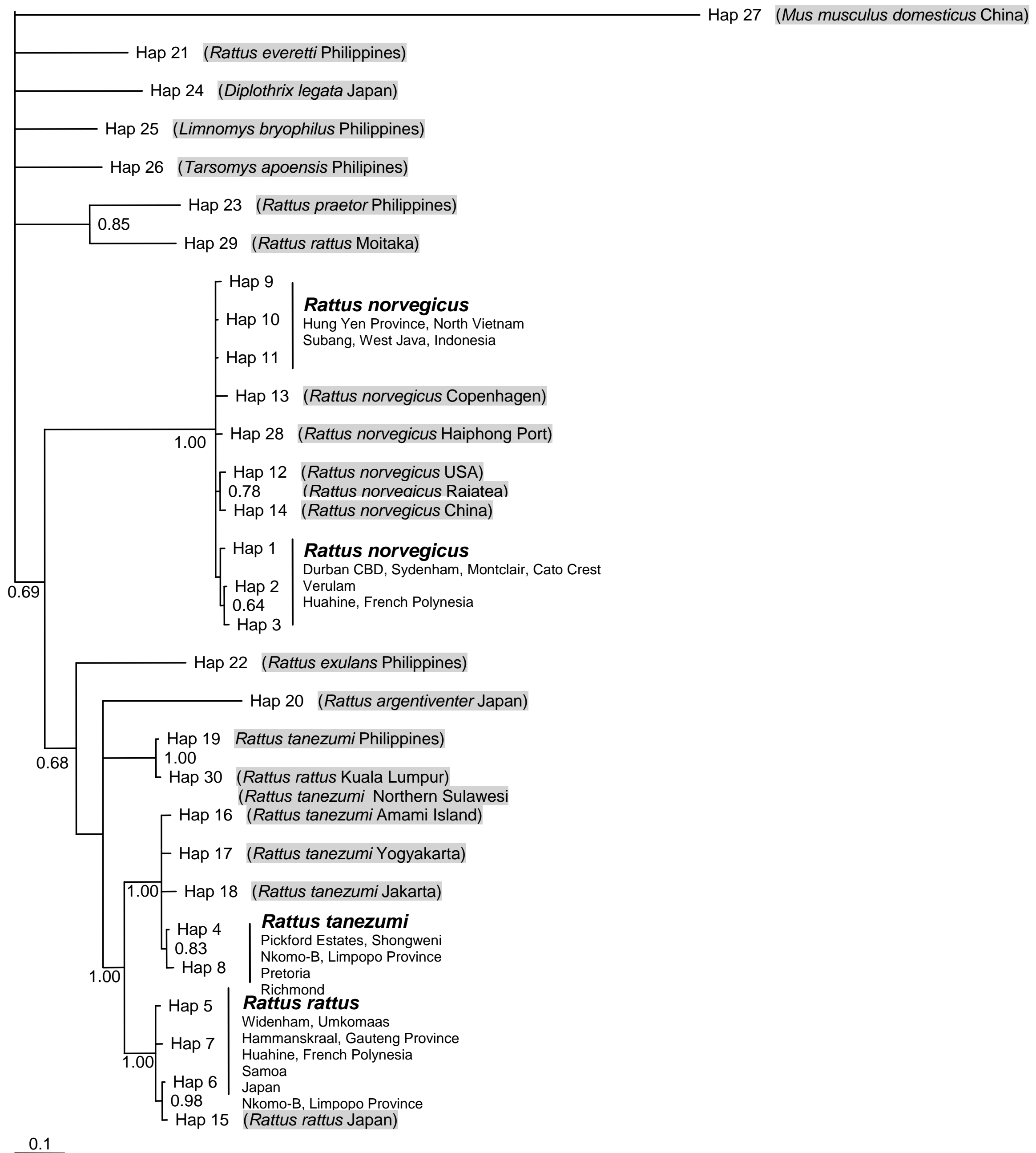

Figure 9: Bayesian phylogenetic tree obtained from the analysis of 676 nucleotide cytochrome *b* dataset using the GTR genetic distance model. Posterior probabilities are indicated adjacent to the nodes. Tree rooted to the outgroup *Mus musculus domesticus (*Hap 27). Labels in bold indicate species identified in this study.

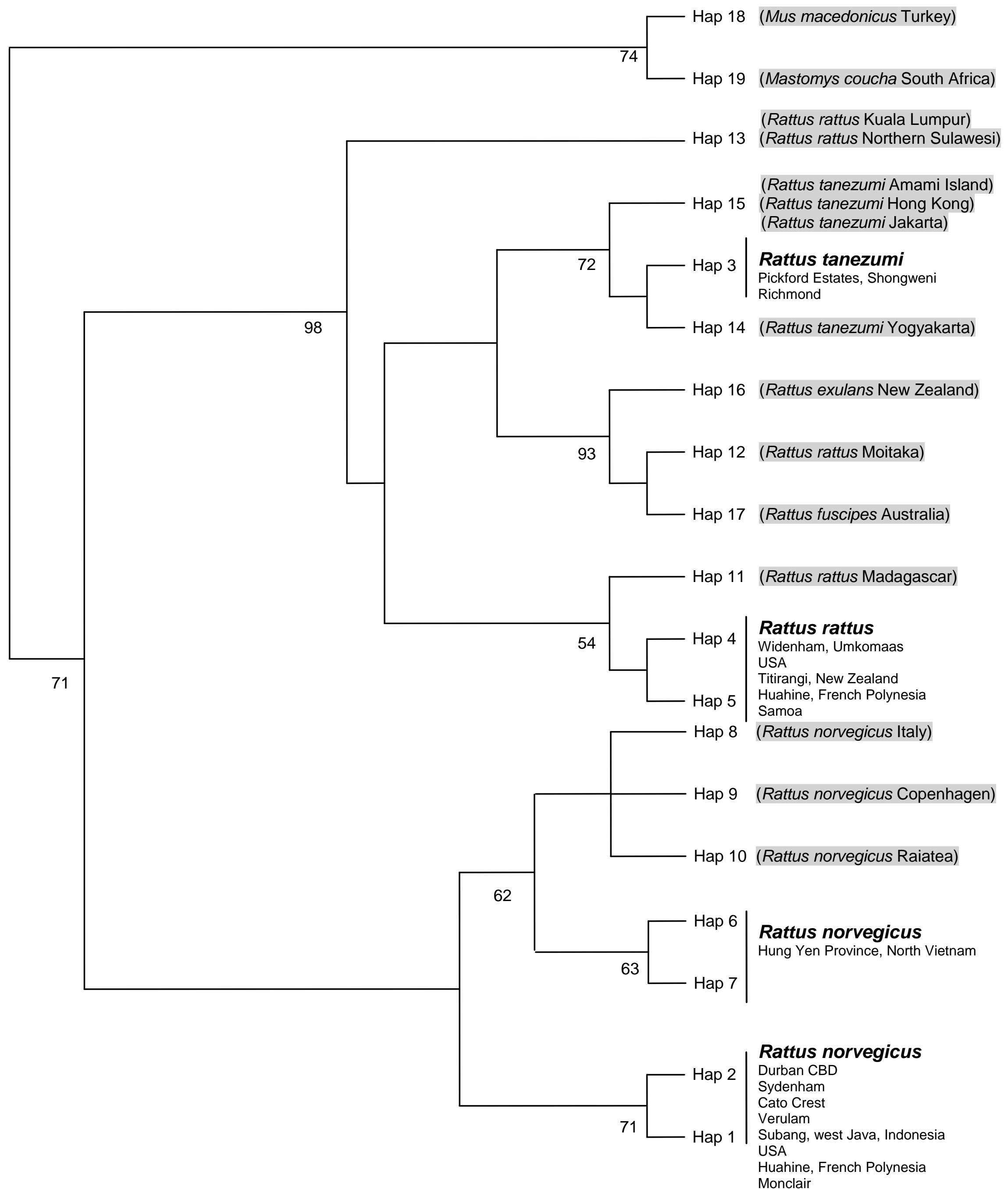

Figure 10: Maximum Parsimony tree showing evolutionary relationships between samples based on 320 nucleotides of the mitochondrial D-loop sequence. Bootstrap values are indicated adjacent to the nodes. Tree rooted to outgroups *Mus macedonicus* (Hap18) and *Mastomys coucha* (Hap 19). Labels in bold indicate species identified in this study.

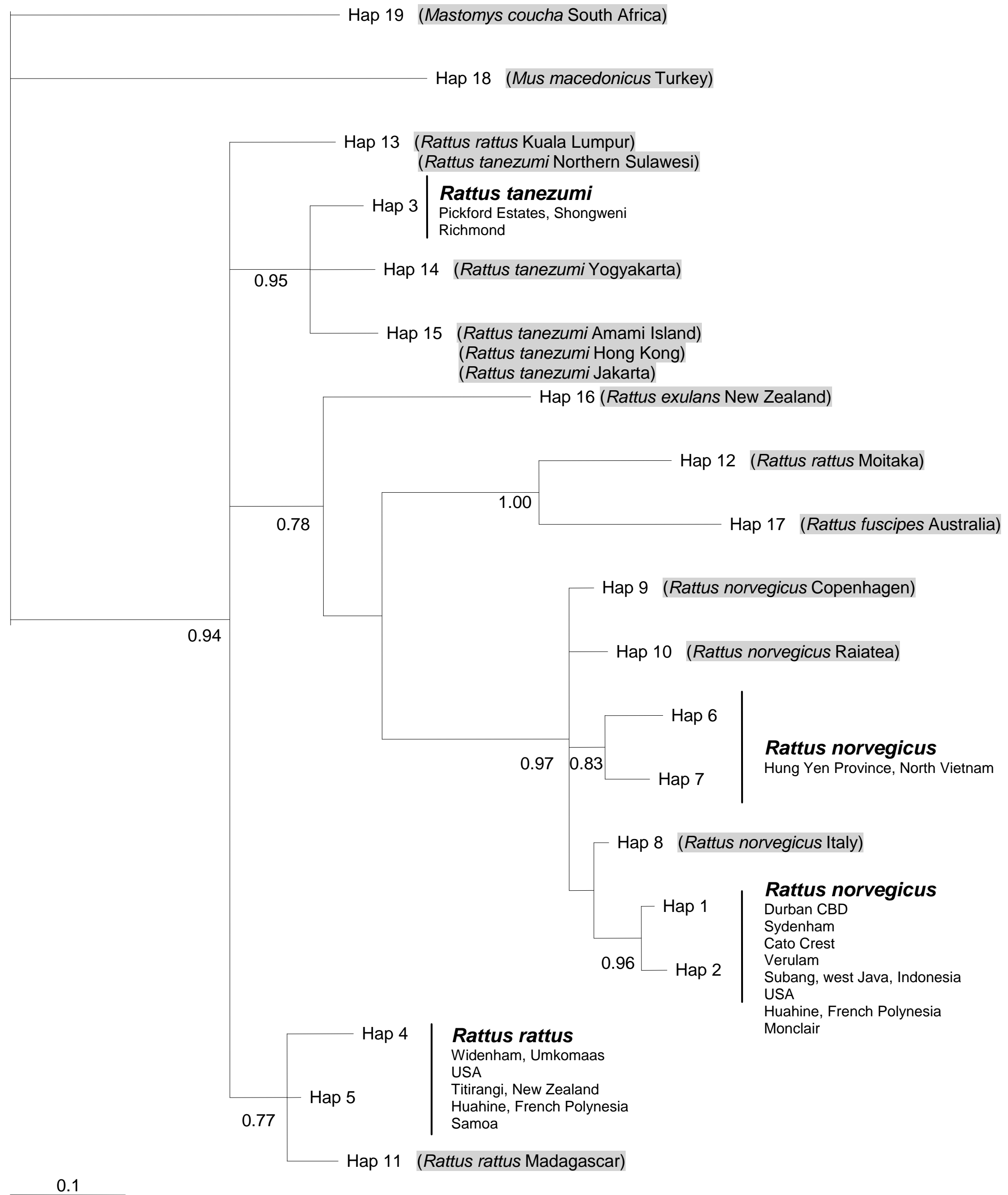

Figure 11: Bayesian phylogenetic tree obtained from the analysis of D-loop sequences (320 nucleotides long) using the HKY genetic distance model. Posterior probabilities are indicated adjacent to the nodes. Tree rooted to outgroups *Mus macedonicus* (Hap18) and *Mastomys coucha* (Hap 19). Labels in bold indicate species identified in this study.

Maximum parsimony and Bayesian analyses of the 1146 and 676 nucleotide cytochrome *b* haplotypes (Table 10) (Figures 6, 7, 8 and 9) were largely congruent. All *R. rattus*, *R. norvegicus* and *R. tanezumi* study haplotypes (1 to 11) formed a strongly-supported clade in the maximum parsimony analysis (bootstrap 100%) (Figures 6 and 7). This clade was essentially unsupported (posterior probability 0.69) in the Bayesian analysis of the shorter cytochrome *b* dataset (Figure 10), and not present in the 1146 nucleotide cytochrome *b* dataset (Figure 8).

All *R. norvegicus* haplotypes (1, 2, 3, 9, 10 and 11) formed a clade which was poorly-supported in the maximum parsimony analysis (<50% and 54%, Figures 7 and 8), but well-supported (posterior probability 0.99 and 1.00, Figures 8 and 9) in the Bayesian analysis.

Sister to this clade was a clade which included *R. tanezumi* and 2 study samples and other taxa (*R. everetti*, *R. argentiventer*, *R. exulans*, *R. praetor*, *D. legata*, *L. bryophilus* and *T. apoensis*), although there was no support for this wider grouping. The sub-clade containing *R. tanezumi* and two study samples was moderate- to well-supported (bootstrap 72% and 64%, Figures 6 and 7, posterior probability 0.69 and 1.00 (Figure 8 and 9, respectively). This sub-clade comprised two reciprocally-monophyletic *R. rattus* sub-clades. One of these sub-clades (support 100%, 69%, 1.00, 0.83; Figures 6, 7, 8, 9) contained *R. tanezumi* study haplotypes 4 and 8 as well as *R. tanezumi* haplotypes derived from Amami Islands in Japan and Yogyakarta and Jakarta in Indonesia (Figures 7 and 9). The other sub-clade contained *R. rattus* study haplotypes 5, 6 and 7 and *R. rattus* (Japan) haplotype 15 (support 100%, 99%, 1.00, 1.00; Figures 6, 7, 8, 9). There is good support (posterior probability 0.97, Figure 8) for the inclusion of a sub-clade comprising *R. argentiventer* and *R. exulans* in the clade containing the *R. rattus* study samples.

Analysis of the D-loop haplotypes (Table 11) (Figures 10 and 11) revealed, with respect to the outgroups *Mus macedonicus* and *Mastomys coucha* (haplotypes 18 and 19), all *R. rattus*, *R. norvegicus* and *R. tanezumi* study haplotypes (1 to 7) formed a clade which was moderately supported (maximum parsimony bootstrap 71%, posterior probability 0.94). Within this, *R. norvegicus* haplotypes (1, 2, 6 and 7) formed a sub-clade (<50% bootstrap support, posterior probability 0.97) which included *R. norvegicus* haplotype 8, 9 and 10 from Italy, Copenhagen and Raiatea, respectively. Study haplotypes 1 and 2 formed a moderately-supported (bootstrap 71%) sub-clade, and study haplotypes 6 and 7 a relatively poorly-supported sub-clade (bootstrap 63%, posterior probability 0.83).

In the maximum parsimony analysis, *R. rattus* study haplotypes (3, 4 and 5) were contained within a strongly-supported (98% bootstrap) clade which was sister to that containing the *R. norvegicus* haplotypes (Figure 10). In the Bayesian analysis, however, haplotypes 4 and 5, and haplotype 3 were part of separate sister-clades to the *R. norvegicus* clade. Both maximum parsimony (72%) and Bayesian analysis (0.95) provided good support for a clade comprising *R. tanezumi* (haplotype 3) and *R. tanezumi* haplotypes 14 (Yogyakarta) and 15 (Amami Island, Hong Kong and Jakarta).

3.2 Cytogenetic Analyses

For the construction of *Rattus* karyograms, the chromosomes were arranged according to the system used for *R. rattus* by Yosida (1985). Only karyograms obtained from uniformly-stained chromosomes are reported here. This is because the G- and C-banding methods did not yield clearly-defined banding patterns that could be confidently used for analysis.

Karyograms produced from chromosome spreads obtained from field-identified *R. rattus* specimens yielded three karyogram arrangements each characterized by a different diploid number (ranging from 38 to 42).

Arrangement 1 had a diploid number of 2n = 38 (Figure 12) and was characterized by nine pairs of acrocentric chromosomes, acrocentric M_1 and M_2 marker chromosomes, seven pairs of metacentric chromosomes and one pair of acrocentric X chromosomes. Specimens that had this arrangement were SA255, SA256, SA258, ARC101, ARC170 and ARC171.

Arrangement 2 had a diploid number of 2n = 40 (Figure 13) and was characterized by nine pairs of acrocentric chromosomes, acrocentric M_1 and M_2 marker chromosomes, eight pairs of metacentric chromosomes and one pair of acrocentric X chromosomes. Specimens that had this arrangement were SA257 and SA259.

Arrangement 3 had a diploid number of 2n = 42 (Figure 14) and was characterized by 13 acrocentric chromosome pairs, seven pairs of metacentric chromosomes and acrocentric X and Y chromosomes. No marker chromosomes were identified for this arrangement. Specimens that had this arrangement were SA253, SA254, UP02, ARC79 and R1.

Only a single karyogram could be produced from chromosome spreads obtained from field-identified *R. norvegicus* specimens. This arrangement was observed to have a diploid number of 2n = 42 (Figure 15) and was characterized by one pair of sub-telocentric chromosomes, 11 pairs of acrocentric chromosomes, eight pairs of metacentrics and a pair of acrocentric X chromosomes. All field-identified *R. norvegicus* specimens had this arrangement.

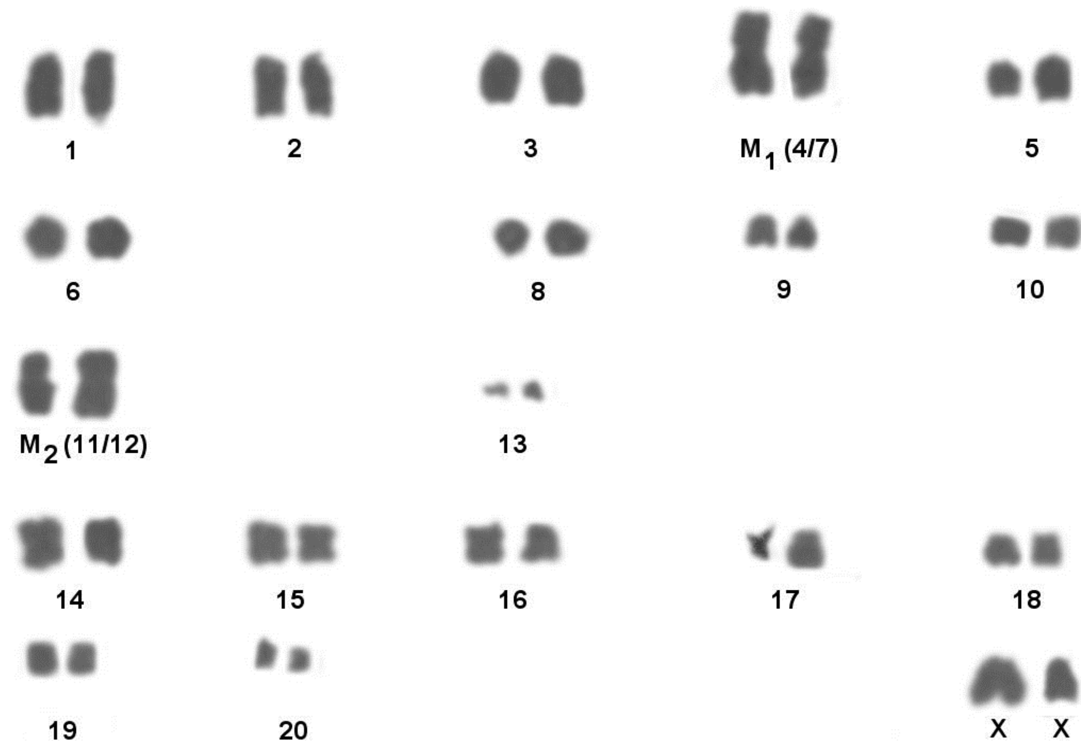

Figure 12: Karyogram of *R. rattus* Arrangement 1

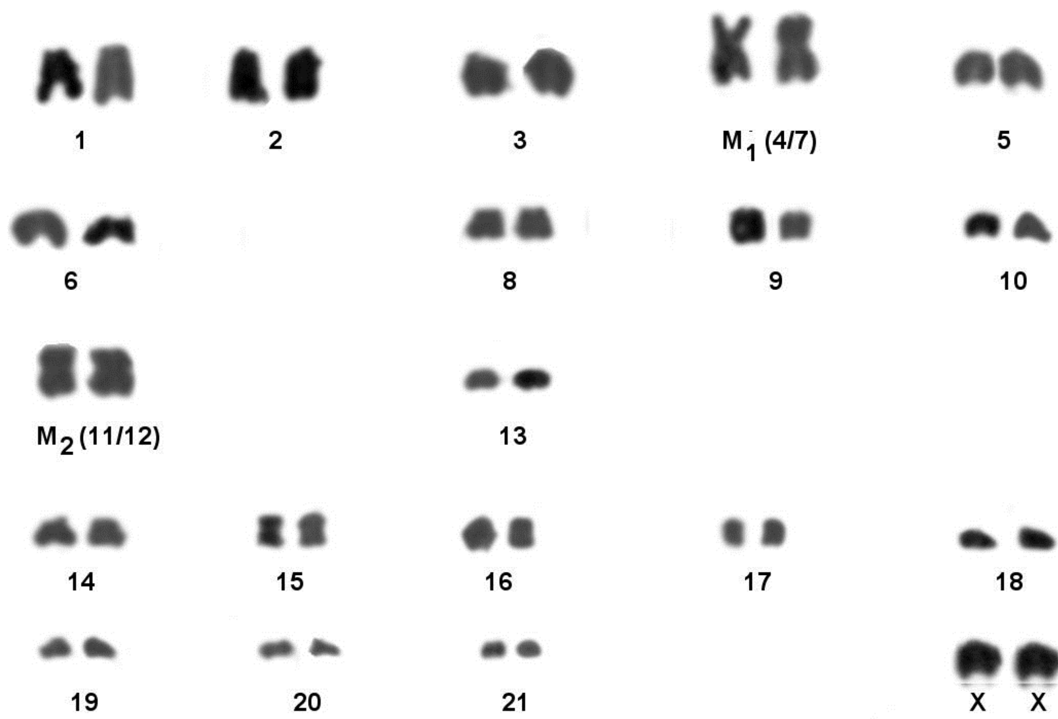

Figure 13: Karyogram of *R. rattus* Arrangement 2

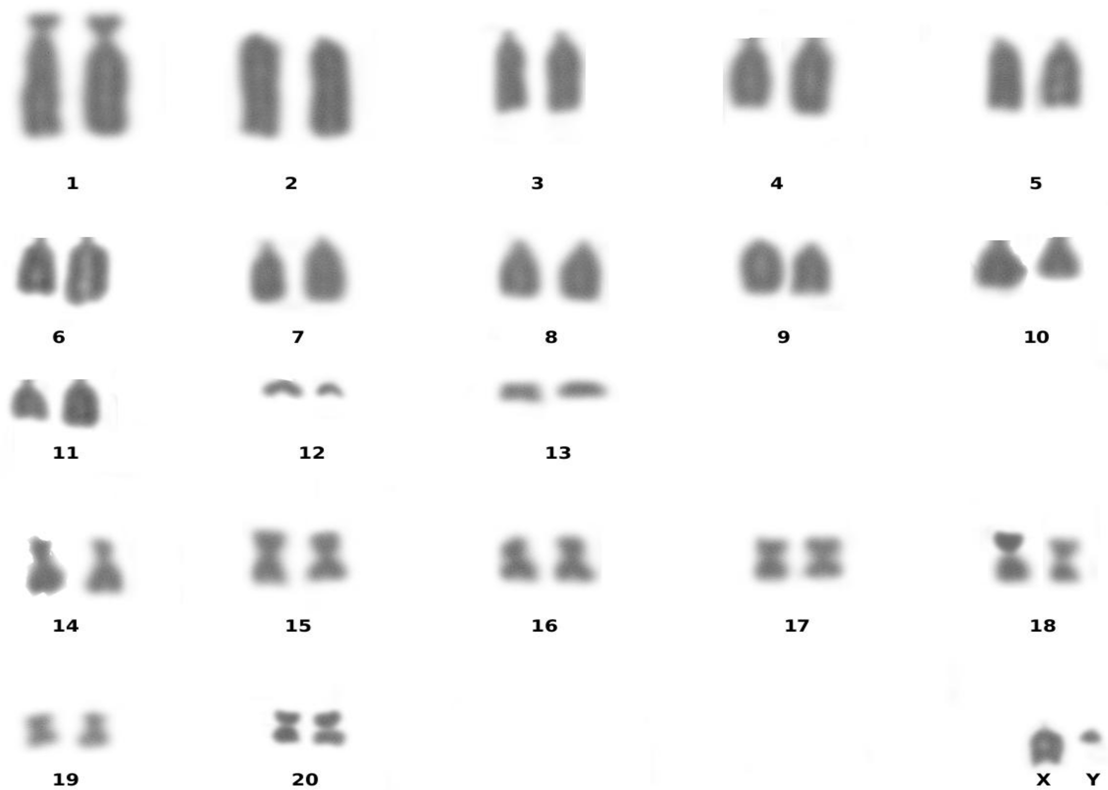

Figure 14: Karyogram of *R. rattus* Arrangement 3

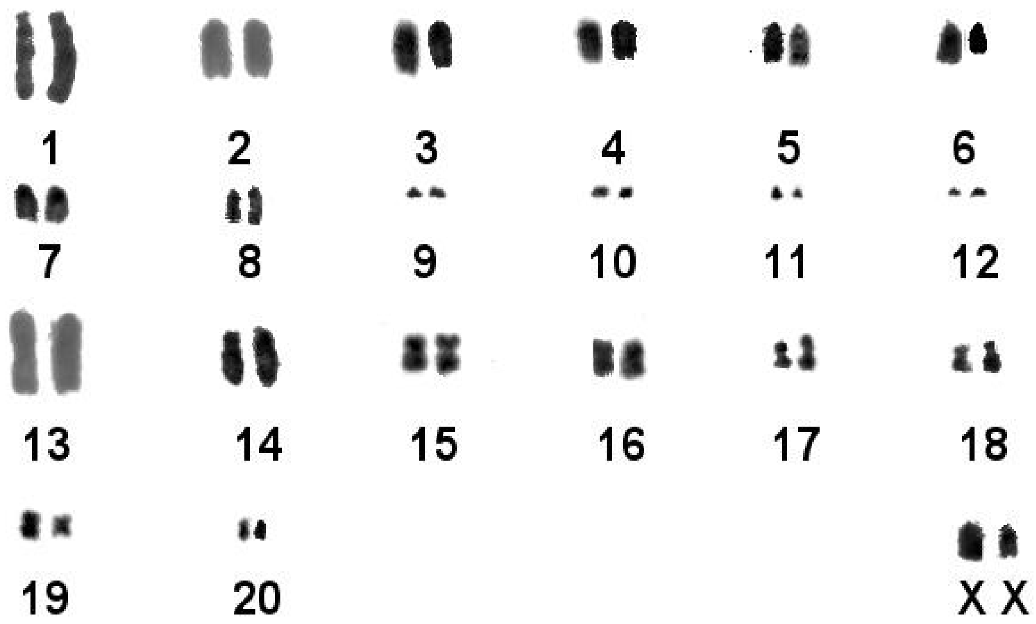

Figure 15: Karyogram of *R. norvegicus*

CHAPTER 4

DISCUSSION

Morphological identification of many *Rattus* species is generally considered to be difficult and in some cases unreliable, even if whole carcasses are available; this is mainly due to the lack of distinct morphological identification criteria (Robins *et al.*, 2007; Matisoo-Smith and Robins, 2009). Past morphological studies have identified the presence of only two invasive *Rattus* species viz. *R. norvegicus* and *R. rattus* within the southern African region (De Graaff, 1981; Skinner and Smithers, 1990). It was therefore expected that *Rattus* specimens collected for this study would either be *R. norvegicus* or *R. rattus*. Initial field identifications confirmed this assumption as specimens field-identified as either *R. norvegicus* or *R. rattus* conformed to morphological identification criteria published by De Graaff (1981), Skinner and Smithers (1990), Long (2003) and Musser and Carleton (2005).

However, molecular and cytogenetic analysis of these wild-caught specimens indicated that morphologically-based field identifications underestimated the number of *Rattus* species present in the sample. There were two genetically-distant monophyletic clades present within the sample set morphologically-identified as *R. rattus*, indicating the presence of a cryptic species.

4.1 Molecular Analyses

All analysis of the mitochondrial cytochrome *b* and D-loop datasets (genetic distance, maximum parsimony, statistical parsimony haplotype network and cytogenetics) were congruent in supporting *R. norvegicus* as a discrete, genetically distant *Rattus* species, and in finding cryptic genetic diversity within the clade of morphology-identified *R. rattus* samples. This led to the identification of *R. tanezumi* within the sample set, a first record of the presence of this species in KwaZulu-Natal.

4.1.1 *Rattus norvegicus*

Specimens collected from various points within the Durban Central Business District (CBD), Sydenham, Montclair, Cato Crest and Verulam (Tables 1 and 10) and initially field-identified as *R. norvegicus* showed close association with GeneBank-derived *R. norvegicus* reference sequences when analysed using Bayesian and maximum parsimony analysis of both the

cytochrome *b* and D-loop datasets, thus providing support for the identity of these specimens. The haplotype diversity amongst these *R. norvegicus* specimens was fairly low (0.516). This was further reflected in both the cytochrome *b* (Figure 4) and D-loop (Figure 5) haplotype networks: The six cytochrome *b* and four D-loop haplotypes obtained for these *R. norvegicus* specimens were observed to be linked to one another by 1-3 mutational steps. This was consistent with the low cytochrome *b* genetic distances between *R. norvegicus* haplotypes (mean 1.2%, range 0.4% to 2.7%) (Table 12 and 13), which fell within the range reported for intraspecific variation amongst rodents (mean 1.5%, range 0.0 to 4.7%) by Baker and Bradley (2006).

Rattus norvegicus specimens collected from the Durban CBD (with the exception of the specimen collected at Warwick Avenue), View Street in Sydenham, Monclair Park in Montclair, Makhanyile CC1514 and 145 Cato Manor Road in Cato Crest, Park Station in Verulam were observed to form a singular haplotype (2, Table 10 and 11) along with an *R. norvegicus* mitochondrial sequence obtained from GenBank that originated from Huahine, Society Islands, French Polynesia. *Rattus norvegicus s*pecimens collected from populations in View Street in the Durban CBD and from Nzuza CC1504 in Cato Crest which formed two separate haplotypes (1 and 3, respectively, Table 10 and 11), appear to be derived from the populations in haplotype 2 as comparisons between haplotypes 1, 2 and 3 yielded very low genetic distance values ranging from 0.004 (0.4%) to 0.009 (0.9%) and were only separated from haplotype 2 by one mutaional step (Figure 4). Although these three haplotypes formed a monophyletic clade with *R. norvegicus* samples retrieved from the NCBI Genbank and originating from Copenhagen, the USA and China (Figure 6) as well as French Polynesia (Raiatea) (Figure 7), genetic distance values indicate that these specimens are most closely related to the NCBI GenBank *R. norvegicus* samples that originated from China (between 1.4% and 1.8% GTR-corrected genetic distances, Table 12 and 13) and Raiatea in French Polynesia (between 1.3% and 1.8% GTR-corrected genetic distances, Table 12 and 13). This relationship between South African *R. norvegicus* specimens and the NCBI GenBank *R. norvegicus* samples from China and Raiatea was further supported when the D-loop sequences were analysed (Tables 11 and 14, Figures 5, 10 and 11). However, it is most likely that the *R. norvegicus* populations observed in KwaZulu-Natal are due to two separate introductions from French Polynesia as they share a haplotype with samples derived from this area.

4.1.2 *Rattus rattus*

The five cytochrome *b* and three D-loop haplotypes obtained for field-identified *R. rattus* specimens were seen to split into two monophyletic clades (*R. tanezumi* and *R. rattus*) (Figure 4 and 5) separated from each other by 13 to 31 mutational steps. The inter-clade genetic distance of 16.9%, is consistent with the inter-clade genetic distance reported for sister species (mean 9.55%, range 2.70% to 19.23%) by Baker and Bradley (2001) with haplotype diversity being fairly high (0.731).

4.1.2.1 *Rattus rattus*

Specimens collected from Widenham in Umkomaas, Hammanskraal in Gauteng Province and from Nkomo-B in Limpopo Province formed three distinct haplotypes. Specimens collected from Nkomo-B in Limpopo Province and one specimen (ARC 171) which was collected from Hammanskraal in Gauteng Province formed singlular haplotypes (6 and 7, Table 10 and 11), with specimens collected from Widenham in Umkomaas and the remaining specimen from Hammanskraal in Gauteng Province (ARC 170) forming a haplotype (5, Table 10 and 11) with GenBank-derived *R. rattus* sequences originating from Titirangi in New Zealand, Huahine in French Polynesia and Samoa. Interestingly, a reference sequence identified as *R.tanezumi* on the NCBI-Genbank was also seen to be part of this haplotype composition and it is highly likely that this sequence was mis-identified and probably originated from an *R. rattus* specimen and not from an *R. tanezumi* specimen. The close association of the specimens collected from Widenham in Umkomaas, Hammanskraal in Gauteng Province and from Nkomo-B in Limpopo Province with the GenBank *R. rattus* sequences provides support for the identification of these specimens as *Rattus rattus.*

There appear to be two lineages of *R. rattus* in South Africa possibly originating from two separate introductions of this species into the country (Figures 7, 8 and 9). The first lineage is comprised of populations found at Widenham in Umkomaas and Hammanskraal in Gauteng Province (haplotype 5, Table 10), with this lineage most likely originating from stock introductions into South Africa from Samoa, New Zealand or French Polynesia as specimens collected during this study share a haplotype with specimens originating from these locations. Haplotype 7 is considered to be derived from haplotype 5 as it separates from haplotype 5 by three mutational steps (Figure 4), has the lowest genetic distance (1.3% GTR-corrected genetic distances, Tables 12 and 13) when compared to haplotype 5, and forms strongly supported monophyletic clades with haplotype 5 in all cladograms produced (Figures Figures 6, 7, 8 and

9).

The second lineage is comprised of populations located at Nkomo-B in the Limpopo Province, with this lineage being considered to have been introduced into South Africa from Japan as specimens collected from this area were observed to form a monophyletic clade with the *R. rattus* sample retrieved from the NCBI Genbank originating in Japan (Figures 6, 7, 8 and 9) and was most similar to this Japanese sample as it consistently showed the lowest genetic distance (0.4% GTR-corrected genetic distances, Tables 12 and 13).

.

4.1.2.2 *Rattus tanezumi*

Specimens collected from Pickford Estates in Shongweni, Richmond and Pretoria as well as a specimen that was collected from Nkomo-B in Limpopo Province (ARC 79) were monophyletic with Genbank-derived *R. tanezumi* samples when analysed using both Bayesian and maximum parsimony analysis of the cytochrome *b* and D-loop datasets, thus providing support for the indentification of these specimens as *R. tanezumi*.

R. tanezumi study haplotypes (4 and 8, Table 10 and 3, Table 11) formed a monophyletic clade with *R. tanezumi* samples retrieved from the NCBI Genbank and originating from Japan (Amami Island), Indonesia (Jakarta and Yogyakarta) (Figures 7 and 9) and Hong Kong (Figures 10 and 11). The Amami Island and Yogyakarta samples consistently showed the lowest GTR-corrected cytochrome *b* genetic distance to both *R.tanezumi* study haplotypes (4 and 8) (between 2.5% and 3.7%, Table 13).

As both *R. tanezumi* study haplotypes formed a monophyletic clade (Figures 6 – 9), it appears that they may have originated from one stock introduction into South Africa possibly from either Amami Island or Yogyakarta or possibly from two introductions but with founder stocks not yet analysed.

4.2 Cytogenetic Analyses

R. rattus was initially characterized by Yosida in 1985 as a species that comprises four major chromosomal groups; Asian group, Oceanian, Ceylonese and Mauritian.

The Asian group is characterized by a diploid number of 2n = 42 and is considered by Yosida (1985) as being the basic *R. rattus* karyotype. This group has recently been elevated from the subspecies *R. rattus tanezumi* to the species *R. tanezumi.* This change in taxonomic status is primarily based on biochemical and cytogenetic differences between *R. rattus* and *R. tanezumi*, although, specimens that have this karyotype are still considered as Asian type black rats (Baverstock *et al.*, 1983; Chinnen *et al.,* 2005; Long, 2003; Hingston *et al.*, 2005, Robins *et al.*, 2007; Musser and Carleton, 2005). The Asian group is defined by a karyotype that has 13 acrocentric chromosome pairs, seven metacentric chromosome pairs and acrocentric X and Y (found only in the male karyotype) or a pair of acrocentric X chromosomes (found only in the female karyotype) (Yosida, 1985).

The Oceanian group is characterized by having a diploid number of 2n = 38 (Baverstock *et al.*, 1983; Yosida, 1985; Chinnen *et al.,* 2005; Hingston *et al.*, 2005, Robins *et al.*, 2007; Matisso-Smith and Robins, 2009; Musser and Carleton, 2005) and normally has nine acrocentric pairs, seven metacentric pairs, M_1 and M_2 marker chromosomes derived from the Robertsonian fusion of pairs 4 and 7 and 11 and 12, respectively and an acrocentric X and Y or pair of X chromosomes (Yosida, 1985).

The Ceylonese group is characterized by a diploid number of 2n = 40 and is considered as being an intermediate *R. rattus* karyotype (Baverstock *et al.*, 1983; Yosida, 1985; Chinnen *et al.,* 2005; Musser and Carleton, 2005). This karyotype is comprised of nine acrocentric pairs, eight metacentric pairs, one metacentric M_2 marker chromosome derived from the Robertsonian fusion of pairs 11 and 12 and an acrocentric X and Y or pair of X chromosomes (Baverstock *et al.*, 1983; Yosida, 1985).

The Mauritian group is characterized by a distinct 2n = 42 karyotype (Baverstock *et al.*, 1983; Chinnen *et al.,* 2005; Matisso-Smith and Robins, 2009), which is comprised of nine acrocentric pairs, M_1 and M_2 marker chromosomes derived from the Robertsonian fusion of pairs 4 and 7 and 11 and 12, five metacentric pairs and an acrocentric X and Y or pair of X chromosomes. The distinctive trait observable in this karyotype is that it also contains four small acrocentric pairs that are derived from the Robersonian fission of two metacentric pairs, namely metacentric pairs 14 and 18 (Yosida, 1985).

From the arrangements obtained in this study, three of the four *R. rattus* groups that were identified by Yosida in 1985 are present in KwaZulu-Natal. Arrangement 1 (Figure 12) and Arrangement 2 (Figure 13) conformed exactly to the description given by Yosida (1985) of the Oceanian group and Ceylonese group karyotypes, respectively. This indicates that specimens that had these karyotypes (*R. rattus*) are *R. rattus* as this is the species name given to all black rat populations with the Oceanian, Ceylonese or Mauritian group karyotype (Baverstock *et al.*, 1983; Chinnen *et al.,* 2005; Hingston *et al.*, 2005; Robins *et al.*, 2007; Musser and Carleton, 2005). Arrangement 3 (Figure 14) conformed to the description given by Yosida (1985) of the Asian group karyotype, thus indicating that specimens with this karyotype (*R. tanezumi*) should be referred to as *R. tanezumi* as this is the species name assigned to all black rat populations that present with the Asian group karyogram (Baverstock *et al.*, 1983; Chinnen *et al.,* 2005; Hingston *et al.*, 2005; Musser and Carleton, 2005; Robins *et al.*, 2007).

Rattus norvegicus is described as having a karyotype of 2n = 42 that is composed of two sub-telocentric pairs, nine acrocentric pairs, nine metacentric pairs and an acrocentric X and Y or pair of X chromosomes (Baverstock *et al.*, 1983; Yosida, 1985; Yiğit *et al.*, 1998). The arrangement seen in the karyogram obtained for *R. norvegicus* in this study (Figure 15) conforms to the description given by Yosida (1985) and Yiğit *et al.* (1998) as it has a diploid number of 42 and contains one pair of sub-telocentric chromosomes, 11 pairs of acrocentric chromosomes, eight pairs of metacentrics and a pair of acrocentric X chromosomes. However, due to the lack of karyological studies on this species (Yiğit *et al.*, 1998), it is uncertain whether polymorphisms in its karyotype do exist as in *R. rattus* (Yosida, 1985; Caldarini *et al.*, 1989; Chinnen *et al.,* 2005).

From the molecular and cytogentic data collected in this study it is evident that there are three invasive species of *Rattus* present in KwaZulu-Natal, viz. *R. norvegicus*, *R. rattus* and *R. tanezumi*. *R. tanezumi* is considered to be the latest invader to KwaZulu-Natal as specimens for this species are only now being identified. With regards to *R. rattus* and *R. norvegicus*, it is most likely that the two stock introductions of *R. rattus* occurred first followed by a single stock introduction of *R. norvegicus*, which along with its higher breeding rate and highly territorial behavior (De Graaff, 1981; Long 2003) has displaced *R. rattus* from the ports and some urban areas (Borchert *et al.*, 2007). However, further studies into the origin and dispersal of the various *R. rattus*, *R. norvegicus* and *R.tanezumi* populations present are still needed to fully understand the complexity of *Rattus* introductions to KwaZulu-Natal.

4.3 Conclusion

Mitochondrial cytochrome b and D-loop sequences and cytogenetic data obtained during this study were consistent with the presence of three distinct species of *Rattus* (*norvegicus*, *rattus* and *tanezumi*) within KwaZulu-Natal even though only two had initially been field-identified.

Study specimens identified as *Rattus norvegicus* yielded typical 2n = 42 karyograms, were monophyletic with *R. norvegicus* specimens from other continents, and were separated from other *Rattus* species by genetic distances large enough to be consistent with their allocation to a separate geneus or still-higher taxon. Cytochrome *b* phylogeny indicates that the *Rattus norvegicus* populations in KwaZulu-Natal are most like due to one original stock introduction.

Specimens morphologically-identified as *Rattus rattus* formed two clades, consistent with the existence of two species (*Rattus tanezumi* and *Rattus rattus*) on phylogenetic, genetic and karyological grounds. This represents the first reporting of *Rattus tanezumi* in South Africa. Cytochrome *b* phylogeny indicates that the *Rattus tanezumi* populations in KwaZulu-Natal are likely to have originated from one stock introduction and the *Rattus rattus* populations from two stock introductions.

This study has aided in expanding current knowledge of the number of invasive commensal *Rattus* species known to inhabit South Africa and allowed for an extension in distribution and genetic records of *Rattus norvegicus* (6 new records), *Rattus rattus* (6 new records) and *Rattus tanezumi* (3 new records) on the NCBI GenBank.

REFERENCES

Aboim, M. A., Menezes, M, Schlitt, T. and Rogers, A. D. (2005). Genetic structure and history of populations of the deep-sea fish *Helicolenus dactylopterus* (Delaroche, 1809) inferred from mtDNA sequence analysis. *Molecular Ecology* **14**, 1343 – 1354.

Agapow, P. M., Bininda – Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C. and Purvis, A. (2004). The impact of species concept on biodiversity studies. *The Quarterly Review of Biology* **79**, 161 – 179.

Aplin, K. P., Chesser, T. and ten Have, J. (2003). Evolutionary biology of the genus *Rattus*: profile of an archetypal rodent pest. *Rats, mice and people: rodent biology and management. ACIAR Monograph* **96**, 487 – 498.

Baker, R. J. and Bradley, R. D. (2001). A test of the genetic species concept: cytochrome *b* sequences and mammals. *Journal of Mammalogy* **82**, 960 – 973.

Baker, R. J. and Bradley, R. D. (2006). Speciation in mammals and the Genetic Speciation Concept. *Journal of Mammology* **87**, 643 – 662.

Ballard, J. W. O. and Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular ecology* **13**, 729 – 744.

Bastos, A. D. S., Chimimba, C. T., von Maltitz, E., Kirsten, F. And Belmain, S. (2005). Identification of rodent species that play a role in disease transmission to humans in South Africa. *Proceedings of the South African Society for Veterinary Epidemiology and Preventative Medicine*, 78 – 83.

Baverstock, P.R, Adams, M., Maxson, L.R., Yosida, T. H. (1983). Genetic differentiation among karyotypic forms of the black rat, *Rattus rattus*. *Genetics* **105**, 969 – 983.

Begon, M. (2003). Disease: Health effects on humans, population effects on rodents. *Rats, mice and people: rodent biology and management. ACIAR Monograph* **96**, 13 – 19.

Belmain, S. (2003). Ratzooman – A new project aimed at understanding emerging/re-emerging infectious diseases transmitted by rodents. *Rodent Research Newsletter* **16**, 5 – 6.

Berdoy, M., Webster, J. P., Macdonald, D.W. (2000). Fatal attraction in rats infected with *Toxoplasma gondii*. *Proceedings of the Royal Society, London* **267**, 1591 – 1594.

Bermingham, E. and Moritz, C. (1998). Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**, 367 – 369.

Borchert, J. N, Mach, J. J., Linder, T. J., Ogen-Odoi, A., Angualia, S. (2007). Invasive rats and bubonic plague in Northwest Uganda. *Managing Vertebrate Invasive Species:*

Proceedings of an International Symposium, 283 – 293.

Burnham, K. P. and Anderson, D. R. (2002). Model selection and multimodel inference: A practical information-theoretic approach. Springer, New York.

Caldarini, G, Capanna, E., Civitelli, M. V., Corti, M, Simonetta, A. (1989). Chromosomal evolution in the subgenus *Rattus* (Rodentia, Muridae): karyotype analysis of two species from the Indian subregion. *Mammalia* **53**, 77 – 84.

Chan, E. Y. (2005). Advance in sequencing technology. *Mutation Research* **573**, 13 – 40.

Chinen, A. A., Suzuki, H., Aplin, K. P., Tsuchiya, K., Suzuki, S. (2005). Preliminary genetic characterization of two lineages of black rats (*Rattus rattus* sensu lato) in Japan, with evidence for introgression at several localities. *Genes and Genetic Systems* **80**, 367 – 375.

Clement, M., Posada, D. and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657 – 1660.

Contrafatto, G., Meester, J. A., Bronner, G., Taylor, P. J., and Willan, K. (1992a). Genetic variation in the African rodent sub-family Otomyinae (Muridae): IV: Chromosome G-banding analysis of *Otomys irroratus* and *O. angoniensis*. *Israel Journal of Zoology* **38**, 277 – 291.

Contrafatto, G., Meester, J. A., Willan, K., Taylor, P. J., Roberts, M. A., and Baker, C. M. (1992c). Genetic variation in the African rodent subfamily Otomyinae (Muridae): II. Chromosomal changes in some populations of *Otomys irroratus*.. *Cytogenetics and Cell Genetics* **59,** 293 – 299.

De Graaff, G (1981). *The rodents of Southern Africa.* 225 – 228. Butterworth and Co. (SA) (PTY) LTD.

Ditchfield, A. D. (2000). The comparative phylogeography of Neotropical mammals: patterns of the intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. *Molecular Ecology* **9**, 1307 – 1318.

Duplantier, J. M., Catalan, J., Orth, A., Grolleau, B. and Britton-Davidian, J. (2003). Systematics of the black rat in Madagascar: consequences for the transmission and distribution of plague. *Biological Journal of the Linnean Society* **78**, 335 – 341.

Duplantier, J. M., Duchemin, J. B., Chanteau, S. and Carniel, E. (2005). From the recent lesson of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. *Veterinary Research* **36**, 437 – 453.

Engelbrecht, A., Dobigny, G. & Robinson, T.J. (2006). Further insights into the ancestral murine karyotype: the contribution of the Otomys – Mus comparison using chromosome painting. *Cytogenetic and Genome Research* **112**, 126 – 130.

Freeland, J. (2005). *Molecular Ecology*. John Wiley and Sons, Ltd.

Friedman, S. and De Ley, J. (1965). "Genetic Species" concept in *Xanthomonas*. *Journal of Bacteriology* **89**, 96 – 100.

Gharizadeh, B., Ohlin, A., Mölling, P., Bäckman, A., Aminia, B., Olcénc, P., Nyréna, P. (2003). Multiple group-specific sequencing primers for reliable and rapid DNA sequencing. *Molecular and Cellular Probes* **17**, 203 – 210.

Goodman, S. M. (1995). Rattus on Madagascar and the dilemma of protecting the endemic rodent fauna. *Conservation Biology* **9**, 450 – 453.

Gülbitti Onarici, S., and Sümer, S. (2003). Protein and DNA in Systematic Biology. *Turkish Journal of Biology* **27**, 47 – 55.

Hall, B. G. (2001). *Phylogenetic trees made easy: A how-to manual for molecular biologists.* Sinauer Associates, Inc.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symposium Series* **41,** 95 – 98.

Harrison, C. J. and Langdale, J. A. (2006). Techniques for molecular analysis: A step by step guide to phylogeny reconstruction. *The Plant Journal* **45**, 561 – 572.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F., Yoder, A. D. (2010). Phylogeography's past, present and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution.* **54,** 291 – 301.

Hingston, M., Goodman, S. M., Ganzhorn, J. U., Sommer, S. (2005). Reconstruction the colonization of southern Madagascar by introduced *Rattus rattus*. *Journal of Biogeography* **32**, 1549 – 1559

Huelsenbeck, J. P. and Bollback, J. P. (2001). Application of the likelihood function in phylogenetic analysis.In: Balding, D. J., Bishop, M. and Cannings, C. (Eds) *Handbook of Statistical* Genetics, pp 429 – 431. John Wiley and Sons, LTD.

Isabel, N., Beaulieu, J., Thériault, P., Bousquet, J. (1999). Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. *Molecular Ecology* **8**, 477 – 483.

Kerth, G., Mayer, F., König, B. (2000). Mitochondrial DNA (mtDNA) reveals that female Bechstein's bats live in closed societies. *Molecular Ecology* **9**, 793 – 800.

Koh, H. S., Yoo, M. H., Lee, B. G. and Park, J. G. (2004). Molecular DNA Systematic Analyses of East Asian Mammals: Sequence Variation of cytochrome *b g*ene and control region of mitochondrial DNA of common otter, *Lutra lutra lutra* L. (Mammalia, Carnivora) from Korea. *Korean Journal of Bioogical Science* **8**, 231 – 233.

Kumar, S.; Tamura, K. and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* **5,** 150 – 163.

Leclerc-Madlala, S., Janowski, M. (2004). Identification of perceptions and practices likely to be relevant to the transmission of plague, leptospirosis, and toxoplasmosis. *Social anthropological study for the Ratzooman Project – Prevention of Sanitary Risks Associated with Rodents at the Rural/Peri-Urban Interface.*

Long, J. L (2003). *Introduced Mammals of the World: Their history, distribution and influence*, pp 188 – 199. CABI and CSIRO Publishing.

Martin, Y., Gerlach, G., Schlötterer, C. and Meyer, A. (2000). Molecular phylogeny of European Muroid Rodents based on complete Cytochrome *b* sequences. *Molecular Phylogenetics and Evolution* **16**, 37 – 47.

Matisoo-Smith, E. and Robins, J. (2009). Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biological Invasions* **11**, 1521 – 1527. **DOI: 10.1007/s10530-008-9404-1**

Mathews, C. K; van Holde, K. E. and Ahern, K. G. (2000). *Biochemistry 3rd edition,* pp. 1074 – 1075. Benjamin/Cummings, an imprint of Addison Wesley Longman.

McNaught, M. K. and Owens I. P. F. (2002). Interspecific variation in plumage colour amnong birds: species recognition or light environment? *Journal of Evolutionary Biology* **15**, 505 – 514.

Meerburg, B. G., Singleton, G. R. and Kijlstra, A. (2009). Rodent-borne diseases and their risks for public health. *Critical Reviews in Microbiology* **35**, 221 – 270.

Mills, J. N., Yates, T. L., Ksiazek, T. G., Peters, C. J. and Childs, J. E. (1999). Long-term studies of Hantavirus reservoir populations in the South-western United States: Rationale, Potential, and Methods. *Emerging Infectious Diseases* **5** (1), 95 – 101.

Musser, G. and Carleton, M. D. (2005). Family muridea. In: *Mammal Species of the World: A Taxonomic and Geographic Reference 3rd Edition* (eds Wilson, D. E. and Reeder, D. M.), pp. 894 – 1531. The John Hopkins University Press.

Nylander, J. A. A. (2004). MrModeltest v2. Evolutionary Biology Centre, Uppsala University.

Posada, D. and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution.

Bioinformatics **14,** 817 – 818.

Posada, D. and Buckley, T. R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology.* **53**, 793 – 808

Rambau, R.V., Harrison, W.R., Elder, F.F.B. & Robinson, T.J. (1997). Chromosomes of Brants' whistling rat and genome conservation in the Otomyinae revealed by G-banding and fluorescence in situ hybridization. *Cytogenetics and Cell Genetics* **78**, 216 – 220.

Robins, J.H., Hingston, M., Matisoo-Smith, E. and Ross, H. A. (2007). Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* **DOI: 10.1111/j.147-8286.2007.01752.x**

Robinson, M., Catzeflis, F., Briolay, J and Mouchiroud, D. (1997). Molecular Phylogeny of Rodents, with Special Emphasis on Murids: Evidence from Nuclear Gene LCAT. *Molecular Phylogenetics and Evolution* **8**, 423 – 434.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572 – 1574.

Rozas, J., Sánchez-Delbarrio, J. C., Messeguer, X. and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496 – 2497.

Seabright, M. A. (1971). A rapid banding technique for human chromosomes. *Lancet* **2**, 971 – 972.

Singleton, G. (2003). CSIRO Sustainable ecosystems rodent research in SE Asia has major social, economic and environmental impacts. *Rodent Research Newsletter* **16**, 1 – 2.

Singleton, G. R., Hinds, L. A., Krebs, C. J. and Spratt, D. M. (2003). Rats, mice and people: an interwoven relationship. *Rats, mice and people: rodent biology and management. ACIAR Monograph* **96**, 9 – 10.

Skinner, J.D. & Smithers, R.H.N. (1990). *The mammals of the southern African subregion* (2nd edition). University of Pretoria, pp. 281-284.

Strimmer, K. and von Haeseler, A. (2003) *Nucleotide substitution models*. In: The phylogenetic handbook: A practical approach to DNA and protein phylogeny (eds Salemi, M. and Vandamme, A-M), pp 72 – 87. Cambridge University Press.

Sumner, A. T., Evans, H. J., Buckland, R. A. (1971). A new technique for distinguishing between human chromosomes. *Nature New Biol.* **232**, 31–32.

Swofford, D. L. (2001). PAUP *. Phylogenetic Analysis Using Parsimony (*and Other

Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taylor, P. J. (2000). Patterns of chromosomal variation in South African rodents. *Journal of Mammology* **81**, 317 – 331.

Taylor, P. J., Arntzen, L., Hayter, M. Iles, M., Frean, J., & Belmain, S. R. (2008). Understanding and managing sanitary risks due to rodent zoonoses in an African city: beyond the Boston Model. *Integrative Zoology* **3,** 38-50.

Taylor, P. J., Maree, S., Sandwyk, J., Baxter, R. & Rambau, R. V. (2009). When is a species not a species? Uncoupled phenotypic, karyotypic and genotypic divergence in two species of South African laminate-toothed rats (Murinae: Otomyini). *Journal of Zoology (London)* **277**, 317–332.

Templeton, A. R. (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* **10**, 779 – 791.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24,** 4876 – 4882.

Tollenaere, C., Brouat, C., Duplantier, J., Rahalison, L., Rahelinirina, S., Pascal, M., Moné, H., Mouahid, G., Leirs, H., Cosson, J. (2009). Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *Journal of Biogeography* **DOI: 10.1111/j.1365-2699.2009.02228.x**

Twigg, G. (1975). *The brown rat*. Latimer Trend and Company Ltd./David and Charles (Holdings) limited.

Vandamme, A. M. (2003). Basic concepts of molecular evolution. In: *The phylogenetic handbook: A practical approach to DNA and protein phylogeny* (eds Salemi, M. and Vandamme, A-M), pp 10 – 23. Cambridge University Press.

Wilkinson, G. S. and Chapman, A. M. (1991). Length and sequence variation in Evening bat D-loop mtDNA. *Genetics* **128**, 607 – 617.

Yang, D. Y. and Speller, C. F. (2006). Co-amplification of cytochrome and D-loop mtDNA fragments for the identification of degraded DNA samples. *Molecular Ecology Notes* **6**, 605 – 608.

Yiğit, N., Çolak, E., Sözen, M., Özkurt, S. (1998). The Taxonomy and Karyology of *Rattus norvegicus* (Berkenhout, 1769) and *Rattus rattus* (Linnaeus, 1758) (Rodentia: Muridae) in Turkey. *Turkish Journal of Zoology* **22**, 203 – 212.

Yosida, T. H. (1985). Chromosomal and biochemical evolution of the genus *Rattus. Acta Zool.*

Fennica **170**, 7 – 14.

Zhang, D-X, Hewitt, G. M. (2003). Nuclear DNA analyzes in genetic studies of populations: practice, problems and aspects. *Molecular Ecology* **12**, 563 – 584.

APPENDICES

APPENDIX 1: Preparation of 1.2% agarose gels

A 1% agarose gel that is produced by combining 1.2 g of agarose in 100 ml of 0.5x TBE. To this mixture 100 μl of 0.005 μg/ml Ethidium bromide is added and the gel allowed to set in a mould. Once set, samples are loaded into the wells present in the set gel and electrophoresed at 70V for +/- $1^{1}/_{2}$ hours.

APPENDIX 2: Preparation of DMEM

To prepare one litre of DMEM add 13.75 g/l DMEM powder to one litre of sterile double deionised water within a beaker placed on a magnetic stirrer set at medium speed. Once the DMEM powder has completely dissolved, check pH of solution and adjust to between 6.8 and 7.4 using sterile $NaHCO_3$. Once the pH has been adjusted, add 2 to 4 ml of antibiotics containing penicillin, streptomycin and fungizone. Solution is then to be filter sterilized using a Sartorius filtration system that has a pore size of 0.22 μ and a filter size of 47mm and is connected to a vacuum flask and pump. After filtration, filter sterilized DMEM can be stored in sterile Pyrex bottles.

APPENDIX 3: List of Published sequences

Sample Id.	Description	Accession Number
ARC 79	*Rattus tanezumi* SA Bastos bankit 784793	DQ439819
ARC 101	*Rattus rattus* SA Bastos bankit 796328	DQ439830
ARC170	*Rattus rattus* SA Bastos bankit 796332	DQ439833
ARC171	*Rattus rattus* SA Bastos bankit 796334	DQ439834
SA75	*Rattus norvegicus* Nair & Bastos	DQ439839
SA136	*Rattus norvegicus* Nair & Bastos	DQ439840
SA141	*Rattus norvegicus* Nair & Bastos	DQ439841
SA218	*Rattus norvegicus* Nair & Bastos	DQ439842
SA219	*Rattus norvegicus* Nair & Bastos	DQ439843
SA252	*Rattus norvegicus* Nair & Bastos	DQ439844
SA253	*Rattus tanezumi* Nair	DQ439849
SA254	*Rattus tanezumi* Nair	DQ439850
SA256	*Rattus rattus* Bastos & Nair	DQ439836
SA258	*Rattus rattus* Bastos & Nair	DQ439837
SA259	*Rattus rattus* Bastos & Nair	DQ439838

APPENDIX 4: 1146 nucleotide Cytochrome *b* consensus sequences

```
                  10        20        30        40        50        60        70
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75      ATGACAAACATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA136     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA141     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA174     ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAA-TCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA175     ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA213     ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA217     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA218     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA219     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA220     ATGACAAACATCCGAAAATCTCACCCCCTATCCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA221     ATGA-AAA-ATCCGAAAATCTCACCCCCTATCCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA232     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA233     ATGACAAA-ATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA234     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA235     A-GACAAA-ATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA252     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA253     ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
SA254     ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
SA255     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA256     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA257     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA258     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA259     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC79     ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
ARC101    ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC170    ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC171    ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
UP02      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
R1        ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
PS1       ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
HP1       -----AAACATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
HP2       ATGACAAA-ATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
S1        -----AAA-ATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnUSA     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnCopen   ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnChina   ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RrJap     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RtJap     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RtPhil    ATGACAAACATTCGAAAATCACACCCCCTACTCAAAATCATTAATCACTCCTTCATCGACCTTCCCGCCC
RaJap     ATGACAAACATTCGGAAATCACACCCCCTATTCAAAATTATTAACCACTCCTTCATCGACCTTCCTGCCC
RevPhil   ATGACAAACATCCGAAAATCTCACCCTCTACTCAAAATCATTAACCACTCTTTCATCGACCTTCCTGCCC
RexPhil   ATGACAAACATTCGAAAATCTCATCCTCTACTCAAAATTATTAATCACTCCTTTATTGACCTTCCCGCTC
RpPhil    ATGACAAACATTCGAAAATCCCACCCTCTGCTCAAAATTATTAATCATTCTTTCATCGACCTTCCCGCCC
DlJap     ATGGCAAACATTCGAAAATCCCACCCTCTAATCAAAATTATTAACCACTCTTTCATCGACCTCCCCGCCC
LbPhil    ATGACAAACATCCGAAAATCTCACCCCTTACTCAAAATTATTAACCACTCCTTCATCGACCTCCCCGCCC
TaPhil    ATGACAAACATCCGAAAATCTCACCCTCTGCTTAAAATTATTAACCACTCCTTCATCGACCTTCCCGCCC
MmdChina  ATGACAAACATACGAAAAACACACCCATTATTTAAAATTATTAACCACTCATTCATTGACCTACCTGCCC
RnHai     ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC

                  80        90       100       110       120       130       140
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA136     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA141     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA174     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA175     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA213     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA217     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA218     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA219     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA220     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA221     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA232     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA233     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA234     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA235     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA252     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA253     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA254     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA255     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA256     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA257     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA258     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA259     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC79     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC101    CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC170    CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC171    CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTC-ATAGTTCAAATTATCA
UP02      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
R1        CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
PS1       CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
HP1       CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
HP2       CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
S1        CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnUSA     CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnCopen   CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATTCTCA
RnChina   CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RrJap     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RtJap     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RtPhil    CATCCAACATCTCATCATGATGAAATTTTGGCTCTTTACTAGGAGTATGCCTC-ATAGTCCAAATTATCA
RaJap     CATCTAATATCTCCTCATGATGAAACTTTGGTTCCCTACTAGGAGTATGCCTC-ATGATTCAAATTGTTA
RevPhil   CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RexPhil   CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTATTAGGAGTATGCCTA-ATAATTCAAATCATTA
RpPhil    CATCCAACATCTCATCATGATGAAACTTCGGTTCTCTCCTAGGAGTATGCCTT-ATGATTCAAATCATCA
DlJap     CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
LbPhil    CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-GTAATTCAAATTATCA
TaPhil    CATCCAACATTTCATCATGATGAAACTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAATCCAAATTATCA
MmdChina  CATCCAACATTTCATCATGATGAAACTTTGGGTCCCTTCTAGGAGTCTGCCTA-ATAGTCCAAATCATTA
```

```
RnHai      CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA

                     150       160       170       180       190       200       210
SA75       CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA136      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA141      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA174      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA175      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA213      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA217      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA218      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA219      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA220      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA221      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA232      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA233      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA234      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA235      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA252      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA253      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
SA254      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
SA255      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA256      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA257      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA258      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA259      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC79      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
ARC101     CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC170     CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC171     CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
UP02       CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
R1         CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
PS1        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
HP1        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
HP2        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
S1         CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnUSA      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnCopen    CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnChina    CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RrJap      CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RtJap      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RtPhil     CAGGCCTATTCCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
RaJap      CAGGCCTGTTCCTAGCAATACACTACACATCCGATACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RevPhil    CAGGCTTATTCCTAGCAATACACTACACATCCGACACTATAACAGCATTTTCATCAGTTGCCCACATCTG
RexPhil    CAGGCCTGTTCCTAGCAATACACTACACATCCGACACTCTAACAGCATTCTCATCAGTTACCCACATCTG
RpPhil     CAGGCCTATTCCTAGCAATACACTATACATCCGACACTACGACAGCATTCTCATCAGTCACCCACATCTG
DlJap      CAGGTCTATTCCTAGCAATACACTACACATCTGACACTACTACAGCATTCTCATCAGTCGCCCACATCTG
LbPhil     CGGGCCTCTTCCTAGCAATACACTACACATCCGACACCATAACAGCATTCTCATCAGTCACCCATATCTG
TaPhil     CAGGTCTATTCCTAGCAATACACTACACATCCGATACAATAACAGCATTCTCATCAGTCACCCACATCTG
MmdChina   CAGGTCTTTTCTTAGCCATACACTACACATCAGATACAATAACAGCCTTTTCATCAGTAACACACATTTG
RnHai      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG

                     220       230       240       250       260       270       280
SA75       C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA136      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA141      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA174      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA175      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA213      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA217      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA218      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA219      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA220      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA221      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA232      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA233      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA234      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA235      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA252      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA253      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
SA254      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
SA255      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA256      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA257      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA258      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA259      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC79      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
ARC101     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC170     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC171     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
UP02       C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
R1         C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
PS1        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
HP1        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
HP2        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
S1         C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnUSA      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnCopen    C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACAGCCCAACGGCGCCTCAATATTTTTCATCTG
RnChina    C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RrJap      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RtJap      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RtPhil     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACACGCCAACGGAGCTTCAATATTCTTTATCTG
RaJap      C-CGAGACGTAAACTACGGCTGATTAATCCGATATTTACACGCCAACGGGGCCTCAATATTCTTCATCTG
RevPhil    C-CGAGATGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCCATATTCTTTATCTG
RexPhil    C-CGAGACGTAAACTACGGCTGACTAATTCGATACTTACATGCCAACGGAGCTTCAATATTCTTTATCTG
RpPhil     C-CGAGATGTAAACTACGGCTGACTAATTCGATATCTACATGCCAACGGAGCCTCTACATTCTTCATCTG
DlJap      C-CGAGATGTAAACTACGGCTGACTAATCCGATATATACATGCCAACGGAGCCTCAATATTTTTTATCTG
LbPhil     C-CGAGACGTAAATTACGGCTGACTAATTCGATACCTACATGCCAACGGAGCCTCAATATTTTTCATCTG
```

```
TaPhil     T-CGAGATGTAAACTACGGCTGACTAATTCGATACTTACATGCCAATGGAGCCTCAATATTTTTCATCTG
MmdChina   T-CGAGACGTAAATTACGGGTGACTAATCCGATATATACACGCAAACGGAGCCTCAATATTTTTTATTTG
RnHai      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG

                   290       300       310       320       330       340       350
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA136      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA141      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA174      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA175      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA213      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA217      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA218      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA219      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA220      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA221      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA232      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA233      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA234      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA235      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA252      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA253      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
SA254      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
SA255      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA256      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA257      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA258      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA259      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
ARC79      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
ARC101     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCCTAGAAACATGAAACATTGGA
ARC170     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
ARC171     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
UP02       CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
R1         CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
PS1        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
HP1        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
HP2        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
S1         CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnUSA      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnCopen    CCTATTCCTCCATGTGGGGCGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnChina    CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGA
RrJap      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCCTAGAAACATGAAACATTGGA
RtJap      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
RtPhil     CCTATTCCTTCATGTAGGGCGAGGGATATACTACGGATCCTATACTTTCCTAGAAACATGAAACATTGGA
RaJap      CTTATTCCTTCACGTAGGACGAGGAATATACTACGGATCCTACACTTTCCTAGAGACATGAAATATCGGG
RevPhil    CCTATTCCTACACGTAGGACGAGGGATATATTACGGATCCTACACCTTTCTAGAGACATGAAATATTGGA
RexPhil    CTTATTCCTCCACGTAGGACGAGGAATGTACTATGGATCCTACACCTTCCTAGAAACATGAAACATCGGA
RpPhil     CCTATACCTCCATGTAGGACGAGGAATGTATTACGGATCCTACACCTTCCTAGAAACATGAAACATCGGA
DlJap      CTTATTCCTTCACGTAGGACGAGGGATATACTATGGATCCTATACCTTCCTAGAAACATGAAACATTGGA
LbPhil     CTTATTCCTCCATGTAGGACGAGGAATGTACTACGGATCCTATACCTTTCTAGAGACATGAAATATTGGA
TaPhil     CTTATTCCTCCATGTAGGACGAGGGATGTACTACGGATCCTACACCTTTCTAGAAACATGAAATATTGGA
MmdChina   CTTATTCCTTCATGTCGGACGAGGCTTATATTATGGATCATATACATTTATAGAAACCTGAAACATTGGA
RnHai      CCTATTCCTCCATGTGGGACGAGGCCTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG

                   360       370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA136      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA141      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA174      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA175      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA213      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA217      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA218      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA219      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA220      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA221      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA232      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA233      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA234      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA235      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA252      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA253      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
SA254      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
SA255      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA256      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA257      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA258      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA259      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC79      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
ARC101     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC170     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC171     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
UP02       GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
R1         GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCAGGAGGGCAAATATCAT
PS1        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
HP1        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
HP2        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
S1         ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnUSA      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnCopen    ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnChina    ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RrJap      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RtJap      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RtPhil     GTAGTCCTACTATTTGCAGTCATAGCAACTGCATTCATAGGCTATGTACTCCCATGAGGACAAATATCAT
RaJap      GTTGTCCTTCTATTTGCAGTCATAGCAACCGCATTCATAGGCTATGTTCTTCCATGAGGACAAATATCAT
RevPhil    GTCGTCCTATTATTTACAGTCATAGCAACCGCATTCATAGGCTACGTGCTCCCATGAGGACAAATATCCT
RexPhil    ATCGTCCTACTATTTGCAGTCATAGCCACCGCATTCATAGGCTACGTACTTCCATGAGGACAAATATCAT
RpPhil     GTTATCCTGCTATTCGCAGTTATAGCAACCGCATTCATAGGCTACGTACTTCCATGAGGACAAATATCCT
```

```
DlJap     ATCATCCTACTATTCACAGTCATAGCAACCGCATTCATAGGCTATGTACTTCCATGAGGACAAATATCCT
LbPhil    GTCGTCCTGTTATTTGCAGTCATAGCAACCGCATTCATAGGCTATGTCCTTCCATGAGGACAGATATCCT
TaPhil    ATTATCCTACTATTTACAGTCATAGCAACCGCATTCATAGGCTACGTCCTTCCATGAGGACAAATATCCT
MmdChina  GTACTTCTACTGTTCGCAGTCATAGCCACAGCATTTATAGGTTACGTCCTTCCATGAGGACAAATATCAT
RnHai     ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT

                    430       440       450       460       470       480       490
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA136     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA141     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA174     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA175     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA213     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA217     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA218     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA219     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA220     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA221     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA232     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA233     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA234     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA235     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA252     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA253     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
SA254     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
SA255     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA256     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA257     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA258     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA259     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
ARC79     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
ARC101    TCTGAGGAGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
ARC170    TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
ARC171    TCTGGGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
UP02      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
R1        TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
PS1       TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
HP1       TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
HP2       TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
S1        TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnUSA     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnCopen   TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnChina   TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RrJap     TCTGAGGAGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RtJap     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RtPhil    TCTGAGGGGCCACAGTAATTACAAACCTATTA-TCAGCCATCCCCTATATTGGCACCACCCTAGTCGAAT
RaJap     TTTGAGGAGCCACCGTAATCACAAACCTATTA-TCAGCTATTCCCTATATTGGCACCACCCTAGTCGAAT
RevPhil   TCTGAGGGGCTACAGTAATTACAAACCTACTA-TCAGCTATCCCCTATATCGGAACAACCTTAGTCGAAT
RexPhil   TCTGAGGAGCCACAGTAATTACAAATCTACTA-TCAGCTATCCCCTATATCGGCACTACCCTAGTCGAAT
RpPhil    TCTGAGGCGCCACAGTTATCACAAATCTTCTA-TCAGCTATCCCTTATATCGGAACTACCCTAGTCGAAT
DlJap     TCTGAGGGGCTACAGTAATCACAAACCTGCTA-TCAGCCATCCCCTACATCGGAACTACATTAGTTGAAT
LbPhil    TCTGAGGAGCTACAGTAATTACAAACCTACTA-TCAGCTATCCCTTATATCGGAACTACCCTGGTCGAAT
TaPhil    TCTGAGGGGCTACAGTAATTACAAATCTACTA-TCAGCCATCCCTTACATCGGAACTACCCTAGTTGAAT
MmdChina  TCTGAGGTGCCACAGTTATTACAAACCTCCTA-TCAGCCATCCCATATATTGGAACAACCCTAGTCGAAT
RnHai     TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT

                    500       510       520       530       540       550       560
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA136     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA141     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA174     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA175     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA213     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA217     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA218     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA219     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA220     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA221     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA232     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA233     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA234     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA235     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA252     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA253     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
SA254     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
SA255     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA256     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA257     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA258     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA259     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC79     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
ARC101    GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC170    GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC171    GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
UP02      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
R1        GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
PS1       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
HP1       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC
HP2       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
S1        GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RnUSA     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTTATCCTCCC
RnCopen   GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RnChina   GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RrJap     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RtJap     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RtPhil    GAATCTGAGGAGGTTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC
RaJap     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTTCACTTTATCCTCCC
RevPhil   GGATCTGAGGGGGATTCTCAGTAGATAAAGCAACTTTAACACGATTTTTCGCATTTCACTTCATCCTCCC
```

```
RexPhil    GAATCTGAGGGGGATTCTCAGTAGACAAAGCAACCCTAACTCGTTTTTTCGCATTTCACTTCATTCTCCC
RpPhil     GAATTTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACTCGCTTTTTCGCCTTCCACTTCATTCTCCC
DlJap      GAATTTGAGGCGGTTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
LbPhil     GAATTTGAGGAGGCTTCTCAGTAGACAAAGCCACTCTAACACGCTTCTTCGCATTTCACTTCATTCTCCC
TaPhil     GAATCTGAGGTGGATTCTCAGTAGACAAAGCAACTCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
MmdChina   GAATTTGAGGAGGCTTCTCAGTAGACAAAGCCACCTTGACCCGATTCTTCGCTTTCCACTTCATCTTACC
RnHai      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC

                    570       580       590       600       610       620       630
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA136      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA141      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA174      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA175      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA213      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA217      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA218      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA219      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA220      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA221      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA232      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA233      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA234      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA235      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA252      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA253      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA254      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA255      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA256      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA257      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA258      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA259      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC79      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC101     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC170     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC171     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
UP02       ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
R1         ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
PS1        ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
HP1        ATTCATTATCGCCGCCCTTGCAATTGTACATCTCCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
HP2        ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
S1         ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnUSA      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnCopen    ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnChina    ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RrJap      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtJap      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtPhil     ATTCATCATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RaJap      ATTCATCATTGCCGCTCTTGCAATTGTACACCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RevPhil    ATTTATCATCGCCGCCCTCGCAATCGTCCACCTTCTTTTCCTCCACGAAACAGGGTCAAACAACCCCACA
RexPhil    ATTCATCATCGCCGCTCTTGCAATCGTACATCTCCTTTTCCTCCACGAAACAGGATCAAACAACCCCACA
RpPhil     ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTTTTCCTACACGAAACAGGATCAAACAACCCCACA
DlJap      ATTTATTATCGCCGCTCTTGCAATTGTACACCTCCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
LbPhil     ATTCATTATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCTACA
TaPhil     ATTTATTATTGCTGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCTACA
MmdChina   ATTTATTATCGCGGCCCTAGCAATCGTTCACCTCCTTTTCCTCCACGAAACAGGATCAAACAACCCAACA
RnHai      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA

                    640       650       660       670       680       690       700
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA136      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA141      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA174      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA175      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA213      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTG-AT
SA217      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA218      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA219      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA220      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTG-AT
SA221      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA232      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA233      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA234      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTG-AT
SA235      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA252      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
SA253      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTGGGAGTAT
SA254      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTGGGAGTAT
SA255      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
SA256      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
SA257      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
SA258      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
SA259      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
ARC79      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTGGGAGTAT
ARC101     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTGT
ARC170     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
ARC171     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
UP02       GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTGGGAGTAT
R1         GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTGGGAGTAT
PS1        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
HP1        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
HP2        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
S1         GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
RnUSA      GGATTAAACTCCAACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
RnCopen    GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
RnChina    GGATTAAACTCCAACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT
RrJap      GGACTAAACTCTGACGCAGACAAAATCCCATTCCATCCATACTACACAATTAAAGACTTACTTGGAGTGT
RtJap      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACACAATTAAAGACTTACTTGGAGTAT
RtPhil     GGGCTAAACTCTGACGCAGACAAAATTCCATTTCATCCATACTACACAATTAAAGACCTACTTGGAGTAT
```

```
RaJap      GGACTAAATTCCGACGCAGACAAAATCCCATTCCACCCATACTATACAATCAAAGACCTACTAGGGGTGT
RevPhil    GGGCTAAACTCCGACGCAGACAAAATCCCATTCCATCCATACTATACAATTAAAGACCTCCTAGGGGTAT
RexPhil    GGCCTAAATTCTGATGCAGATAAAATCCCATTTCATCCATACTACACAATTAAAGACCTACTAGGAGTAT
RpPhil     GGACTAGACTCCAACGCAGACAAAATCCCATTCCATCCATATTACACAATTAAAGACCTCCTAGGAGTAT
DlJap      GGACTAGATTCCAACGCAGACAAAATCCCATTTCACCCCTACTATACAATCAAAGATCTCCTAGGAGTAT
LbPhil     GGACTAAATTCCAATGCAGACAAAATCCCATTTCATCCATATTATACAATTAAAGACCTCCTAGGAGTAT
TaPhil     GGACTAAACTCTGACGCAGACAAAATCCCATTCCACCCATACTACACAATTAAAGACCTACTAGGAGTAT
MmdChina   GGATTAAACTCAGATGCAGATAAAATTCCATTTCACCCCTACTATACAATCAAAGATATCCTAGGTATCC
RnHai      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATACAATTAAAGACCTCCTAGGTGTAT

                    710       720       730       740       750       760       770
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA136      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA141      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA174      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA175      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA213      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA217      TTATATTACTATTATTCCTAATAACCCTAGTACT------------------------------------
SA218      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA219      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA220      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA221      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA232      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA233      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA234      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA235      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA252      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
SA253      TCATTCTACTCCTATTTCTAATAACTCTAGTATTATTTTTCCCAGACTTACTAGGAGACCCAGACAATTA
SA254      TCATTCTACTCCTATTTCTAATAACTCTAGTATTATTTTTCCCAGACTTACTAGGAGACCCAGACAATTA
SA255      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
SA256      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
SA257      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
SA258      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
SA259      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
ARC79      TCATTCTACTCCTATTTCTAATAACTCTAGTATTATTTTTCCCAGACTTACTAGGAGACCCAGACAATTA
ARC101     TCATACTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
ARC170     TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
ARC171     TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
UP02       TCATTCTACTCCTATTTCTAATAACTCTAGTATTATTTTTCCCAGACTTACTAGGAGACCCAGACAATTA
R1         TCATTCTACTCCTATTTCTAATAACTCTAGTATTATTTTTCCCAGACTTACTAGGAGACCCAGACAATTA
PS1        TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
HP1        TTATATTACTATTATTCCTAATAACCCTAG-ACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
HP2        TTATATTACTATTATTCCTAATAACCCTAG-ACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
S1         TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
RnUSA      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
RnCopen    TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
RnChina    TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
RrJap      TCATACTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
RtJap      TCATGTTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAACTA
RtPhil     TCATACTACTCCTATTTCTAATAACTCTAGTATTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA
RaJap      TCATATTACTTCTACTTTTAATAACCCTAGTATTATTCTTCCCAGACTTACTAGGAGACCCAGACAATTA
RevPhil    TCTTACTTCTACTATTCTCAATAATTCTAGTACTATTTTTTCCAGATCTATTAGGAGATCCAGACAACTA
RexPhil    TTATATTACTTCTATTTTTAATAACCCTAGTATTATTCTTCCCAGACTTACTAGGAGATCCAGACAATTA
RpPhil     TCCTGCTTCTTCTATTTCTAATAACCCTAGTACTGTTCTTCCCAGACCTACTAGGAGACCCAGACAACTA
DlJap      TCATACTACTCCTACTTCTAATAACCCTAGTACTATTCTTTCCAGACCTACTAGGAGACCCAGACAACTA
LbPhil     TTATACTACTCCTATTATTAATAACCCTAGTATTGTTCTTCCCAGATCTATTAGGAGACCCAGACAACTA
TaPhil     TCATCCTACTCCTACTACTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAACTA
MmdChina   TAATCATATTCTTAATTCTCATAACCCTAGTATTATTTTTCCCAGACATACTAGGAGACCCAGACAACTA
RnHai      TTATATTACTATTATTCCTAATAACCCTAGTACTATTCTTCCCAGACCTACTAGGAGACCCAGACAATTA

                    780       790       800       810       820       830       840
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA136      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA141      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA174      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA175      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA213      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA217      ----------------------------------------------------------------------
SA218      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA219      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA220      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA221      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA232      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA233      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA234      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA235      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA252      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
SA253      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATACTTCCTATTTGCCTACGCT
SA254      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATACTTCCTATTTGCCTACGCT
SA255      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
SA256      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
SA257      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
SA258      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
SA259      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
ARC79      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATACTTCCTATTTGCCTACGCT
ARC101     CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATATTTCCTATTTGCCTACGCT
ARC170     CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
ARC171     CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
UP02       CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATACTTCCTATTTGCCTACGCT
R1         CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGATACTTCCTATTTGCCTACGCT
PS1        TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
HP1        TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
HP2        TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
S1         TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
RnUSA      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
RnCopen    TACGCCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
RnChina    TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT
RrJap      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
```

```
RtJap      CACACCTGCTAACCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
RtPhil     CACACCTGCTAATCCACTAAATACCCCACCACATATTAAGCCAGAATGGTATTTCCTATTTGCCTACGCT
RaJap      CACACCTGCTAATCCCCTAAATACCCCACCACATATTAAACCAGAATGATATTTCCTATTTGCTTACGCT
RevPhil    CACACCTGCTAATCCCCTAAGCACTCCACCGCACATTAAACCAGAATGATATTTCCTCTTCGCTTACGCT
RexPhil    TATACCTGCCAACCCCCTAAATACCCCACCACATATCAAACCAGAATGATATTTTCTCTTCGCTTACGCC
RpPhil     CACACCCGCTAACCCCCTAAATACTCCACCACACATCAAGCCAGAATGGTATTTCCTCTTTGCCTACGCC
DlJap      CACACCCGCTAACCCCCTAAACACCCCGCCACACATTAAACCAGAATGGTACTTCCTCTTCGCTTACGCC
LbPhil     CACACCTGCTAACCCCCTAAACACCCCACCACACATCAAACCAGAATGATATTTCCTCTTTGCCTACGCT
TaPhil     TACACCCGCTAACCCCCTAAACACCCCTCCACACATCAAGCCAGAATGATATTTCCTCTTTGCCTACGCC
MmdChina   CATACCAGCTAATCCACTAAACACCCCACCCCATATTAAACCCGAATGATATTTCCTATTTGCATACGCC
RnHai      TACACCCGCTAACCCCCTCAACACCCCACCCCACATCAAACCAGAATGATATTTTCTCTTTGCCTACGCT

                     850       860       870       880       890       900       910
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA136      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA141      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA174      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA175      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA213      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA217      ----------------------------------------------------------------------
SA218      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA219      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA220      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA221      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA232      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA233      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA234      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA235      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA252      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
SA253      ATTCTACGCTCCATCCCTAACAAACTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA254      ATTCTACGCTCCATCCCTAACAAACTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA255      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA256      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA257      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA258      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
SA259      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
ARC79      ATTCTACGCTCCATCCCTAACAAACTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
ARC101     ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
ARC170     ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
ARC171     ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCCTAG
UP02       ATTCTACGCTCCATCCCTAACAAACTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
R1         ATTCTACGCTCCATCCCTAACAAACTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
PS1        ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
HP1        ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
HP2        ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
S1         ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
RnUSA      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
RnCopen    ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG
RnChina    ATTTTATGCTCCATTCCCAACAAACTAGGAGGAGGTCGTGAGGCCCTAATCTTATCAATCCTAATCCTTG
RrJap      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
RtJap      ATTCTACGCTCCATCCCTAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTCTAATCTTAG
RtPhil     ATTCTACGTTCCATCCCCAATAAACTAGGAGGGGTA---GTAGCCCTAGTCCTATCAATTCTAATCCTAG
RaJap      ATCCTACGCTCTATCCCCAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATTTTAATCCTAG
RevPhil    ATTCTACGCTCTATCCCCAACAAACTAGGGGGTGTA---GTAGCCCTAGTCCTATCAATTCTAATCCTAG
RexPhil    ATTTTACGCTCCATTCCCAACAAATTAGGAGGAGTA---GTAGCCTTAGTCCTATCAATCCTAATCTTAG
RpPhil     ATTTTACGCTCTATCCCCAACAAGCTAGGAGGAGTC---GTAGCCCTAGTCCTATCAATCCTTATCCTAG
DlJap      ATTCTACGCTCTATCCCCAACAAACTAGGAGGAGTC---GTAGCCCTAGTCCTATCAATTCTAATCCTAG
LbPhil     ATTCTACGCTCTATCCCCAACAAACTAGGAGGAGTA---GTGGCCCTAGTCCTATCAATTCTAATCCTAG
TaPhil     ATCCTACGCTCCATTCCAAACAAACTAGGAGGAGTA---GTAGCCCTAGTCATATCAATTCTAATTCTAG
MmdChina   ATTCTACGCTCAATCCCCAATAAACTAGGAGGTGTC---CTAGCCTTAATCTTATCTATCCTAATTTTAG
RnHai      ATTCTACGCTCCATTCCCAACAAACTAGGAGGAGTC---GTAGCCCTAATCTTATCAATCCTAATCTTAG

                     920       930       940       950       960       970       980
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA136      CCTTACTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA141      CCTTACTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA174      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA175      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA213      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA217      ----------------------------------------------------------------------
SA218      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA219      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA220      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA221      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA232      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA233      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA234      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA235      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA252      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
SA253      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAACTCTGTA
SA254      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAACTCTGTA
SA255      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
SA256      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
SA257      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
SA258      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
SA259      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
ARC79      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAACTCTGTA
ARC101     CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
ARC170     CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
ARC171     CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
UP02       CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAACTCTGTA
R1         CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAACTCTGTA
PS1        CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
HP1        CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
HP2        CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
S1         CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
RnUSA      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
RnCopen    CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
```

```
RnChina    CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTTTA
RrJap      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGCCTAACATTCCGCCCAATCACCCAAATTCTATA
RtJap      CCTTTCTACCATTCTTACATACCTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAATTCTATA
RtPhil     CCTTTCTACCATTCCTACATACTTCAAAACAACGCAGTCTAACATTCCGCCCAATCACCCAAGCCCTATA
RaJap      CTCTCCTACCATTTCTACATACTTCAAAACAACGCAGCCTAACATTCCGCCCTATTACCCAAATTCTATA
RevPhil    CCTTCCTACCATTCCTACATACTTCAAAACAACGCAGCCTAACCTTCCGTCCAATCACCCAAACACTATA
RexPhil    CCCTCCTACCATTCCTCCACACTTCAAAACAACGCAGCCTAACATTCCGCCCAATTACCCAAACCCTATA
RpPhil     CCCTCCTGCCATTCTTACACACCTCAAAACAACGCAGCCTAACCTTCCGCCCAATCACCCAAACCCTATA
DlJap      CCTTCCTACCATTCTTACACACCTCAAAACAACGCAGCCTAACATTCCGCCCAATCACTCAAACTCTATA
LbPhil     CCCTTCTTCCACTCTTACATACTTCAAAACAACGCAGCCTAACATTCCGCCCAATTACCCAAATCCTATA
TaPhil     CCCTCCTGCCATTCCTACATACCTCAAAACAACGCAGCCTAACATTCCGCCCAATCACCCAAATTCTATA
MmdChina   CCCTAATACCTTTCCTTCATACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTA
RnHai      CCTTCCTACCATTCCTGCATACTTCAAAACAACGCAGCTTAACATTCCGCCCAATCACCCAAATCCTCTA

                   990       1000      1010      1020      1030      1040      1050
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA136      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA141      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA174      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA175      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA213      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA217      ----------------------------------------------------------------------
SA218      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA219      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA220      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA221      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA232      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA233      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA234      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA235      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA252      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
SA253      CTGAATTCTAGTAGCCAACCTCTTCATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATT
SA254      CTGAATTCTAGTAGCCAACCTCTTCATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATT
SA255      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
SA256      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
SA257      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
SA258      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
SA259      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
ARC79      CTGAATTCTAGTAGCCAACCTCTTCATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATT
ARC101     CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
ARC170     CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGAGCCAACCAGTAGAACACCCATTTATT
ARC171     CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
UP02       CTGAATTCTAGTAGCCAACCTCTTCATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATT
R1         CTGAATTCTAGTAGCCAACCTCTTCATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATT
PS1        CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
HP1        CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
HP2        CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
S1         CTGAATCCTAGTAGCCAACCTCCTAGTCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
RnUSA      CTGAATCCTAGTAGCCAACCTCCTAGTCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
RnCopen    CTGAATCCTAGTAGCCAACCTCCTAGTCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
RnChina    CGGAATCCTGGTAGCCAACCTCCTTGTCTTAACAGGAATCGGAGGCCAACCAGTAGAACACCCATTTATC
RrJap      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
RtJap      CTGAATTCTAGTAGCCAACCTCTTCATTTTAACATGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
RtPhil     CTGAATCCTAGTAGCCAATCTCTTCATTTTAACATGAATCGGAGGACAGCCAGTAGAACATCCATTTATT
RaJap      TTGAATCTTAGTAGCTAATCTCCTCGTCTTAACATGAATCGGAGGCCAACCAGTAGAACACCCATTCATC
RevPhil    TTGAATACTAACAGCCAACCTACTCATCCTAACATGAATTGGAGGGCAACCAGTAGAACATCCATTTATC
RexPhil    CTGAATCCTAGTAGCTAACCTCTTCATCCTAACATGAATTGGAGGCCAACCAGTAGAACATCCTTTCATC
RpPhil     CTGAATCCTAGCAGCCAACCTCCTCATCTTAACATGAATTGGGGGCCAACCAGTAGAACACCCATTCATT
DlJap      CTGAATCCTAGTAGCCAACCTACTCATTTTAACATGAATTGGAGGACAACCAGTAGAACATCCATTCATC
LbPhil     TTGAATCCTAACGGCCAACCTACTTATTCTAACATGAATTGGGGGACAACCAGTAGAACACCCATTTATC
TaPhil     TTGAATCCTAACAGCCAACCTACTTATTCTAACATGAATTGGAGGCCAACCAGTAGAACACCCATTTATC
MmdChina   CTGAATCCTAGTAGCCAACCTACTTATCTTAACCTGAATTGGAGGCCAACCAGTAGAACACCCATTTATT
RnHai      CTGAATCCTAGTAGCCAACCTCCTAATCTTAACATGAATCGGAGCCCAACCAGTAGAACACCCATTTATC

                   1060      1070      1080      1090      1100      1110      1120
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA136      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA141      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA174      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA175      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA213      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA217      ----------------------------------------------------------------------
SA218      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA219      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA220      ATTATTGGCCAACTAGCCTCCATCAGCT-CTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA221      ATTATTGGCCAACTAGCCTCCATCAGCT-CTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA232      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA233      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA234      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA235      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA252      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
SA253      ATTATTGGCCAACTAGCATCTATTAGTTACTTCTCAATTATCCTTATTCTAATACCAATTTCCGGAATCA
SA254      ATTATTGGCCAACTAGCATCTATTAGTTACTTCTCAATTATCCTTATTCTAATACCAATTTCCGGAATCA
SA255      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
SA256      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
SA257      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
SA258      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
SA259      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
ARC79      ATTATTGGCCAACTAGCATCTATTAGTTACTTCTCAATTATCCTTATTCTAATACCAATTTCCGGAATCA
ARC101     ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTCTAATACCAATCTCCGGAATCA
ARC170     ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTTAATACCAATCTCCGGAATCA
ARC171     ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTCTAATACCAATCTCCGGAATCA
UP02       ATTATTGGCCAACTAGCATCTATTAGTTACTTCTCAATTATCCTTATTCTAATACCAATTTCCGGAATCA
R1         ATTATTGGCCAACTAGCATCTATTAGTTACTTCTCAATTATCCTTATTCTAATACCAATTTCCGGAATCA
PS1        ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
HP1        ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
HP2        ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
S1         ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
```

```
RnUSA      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
RnCopen    ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
RnChina    ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
RrJap      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATCCTAATACCAATCTCCGGAATCA
RtJap      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCAATTATCCTTATTTAATACCAATCTCCGGAATCA
RtPhil     ATCATTGGCCAGCTAGCATCCATCAGCTACTTCTCAATTATCCTTATTCTAATACCAATCTCCGGAATTA
RaJap      ATTATTGGCCAACTAGCATCCATCAGTTACTTCTCGATTATCCTCATTCTAATACCAATCTCTGGAATCA
RevPhil    ATTATTGGCCAACTAGCATCCATCAGCTATTTCTCAATCATCCTTATTCTAATACCAATCTCTGGAATTA
RexPhil    ATCATTGGCCAACTAGCATCTATAAGCTACTTCTCAATTATTCTTATCCTAATACCTATCTCTGGAATCA
RpPhil     ATCATTGGTCAACTAGCCTCTATCAGCTACTTCTCAATTATCCTTATCCTGATACCAATCTCCGGAATTA
DlJap      ATTATTGGCCAACTAGCATCCATGAGTTACTTCTCAATCATCCTTATCCTAATACCAATCTCTGGAATCA
LbPhil     ATCATCGGCCAACTAGCATCCATCAGCTATTTCTCAATTATCCTCATCCTTATACCAATCTCTGGAATCA
TaPhil     ATTATTGGTCAACTAGCATCCATCAGCTACTTCTCAATTATCCTCATCCTAATGCCCATCTCTGGAATTA
MmdChina   ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTCATACCAATCTCAGGAATTA
RnHai      ATTATTGGCCAACTAGCCTCCATCAGCTACTTTTCAATTATCCTCATTCTCATACCAATCTCTGGAATTG
```

```
                  1130      1140
           ....|....|....|....|....|.
SA75       TTGAAGACAAAATGTTAAAATGAAAT
SA136      TTGAAGACAAAATGTTAAAATGAAAT
SA141      TTGAAGACAAAATGTTAAAATGAAAT
SA174      TTGAAGACAAAATGTTAAAATGAAAT
SA175      TTGAAGACAAAATGTTAAAATGAAAT
SA213      TTGAAGACAAAATGTTAAAATGAAAT
SA217      --------------------------
SA218      TTGAAGACAAAATGTTAAAATGAAAT
SA219      TTGAAGACAAAATGTTAAAATGAAAT
SA220      TTGAAGACAAAATGTTAAAATGAAAT
SA221      TTGAAGACAAAATGTTAAAATGAAAT
SA232      TTGAAGACAAAATGTTAAAATGAAAT
SA233      TTGAAGACAAAATGTTAAAA-GAAAT
SA234      TTGAAGACAAAATGTTAAAATGAAAT
SA235      TTGAAGACAAAATGTTAAAATGAAAT
SA252      TTGAAGACAAAATGTTAAAATGAAAT
SA253      TTGAAGACAAAATACTAAAATGAAAC
SA254      TTGAAGACAAAATACTAAAATGAAAC
SA255      TTGAAGACAAAATACTAAAATGAAAC
SA256      TTGAAGACAAAATACTAAAATGAAAC
SA257      TTGAAGACAAAATACTAAAATGAAAC
SA258      TTGAAGACAAAATACTAAAATGAAAC
SA259      TTGAAGACAAAATACTAAAATGAAAC
ARC79      TTGAAGACAAAATACTAAAATGAAAC
ARC101     TTGAAGACAAAATACTAAAATGAAAC
ARC170     TTGAAGACAAAATACTAAAATGAAAC
ARC171     TTGAAGACAAAATACTAAAATGAAAC
UP02       TTGAAGACAAAATACTAAAATGAAAC
R1         TTGAAGACAAAATACTAAAATGAAAC
PS1        TTGAAGACAAAATGTTAAAATGAAAT
HP1        TTGAAGACAAAATGTTAAAATGAAAT
HP2        TTGAAGACAAAATGTTAAAATGAAAT
S1         TTGAAGACAAAATGTTAAAATGAAAT
RnUSA      TTGAAGACAAAATGTTAAAATGAAAT
RnCopen    TTGAAGACAAAATGTTAAAATGAAAT
RnChina    TTGAAGACAAAATGTTAAAATGAAAT
RrJap      TTGAAGACAAAATACTAAAATGAAAC
RtJap      TTGAAGACAAAATACTAAAATGAAAC
RtPhil     TCGAAGACAAAATACTGAAATGA---
RaJap      TTGAAGACAAAATACTAAAATGAAAC
RevPhil    TTGAAGACAAAATACTAAAATGA---
RexPhil    TTGAAGACAAAATACTAAAATGA---
RpPhil     TTGAAGACAAGATACTAAAATGA---
DlJap      TCGAAGACAAAATACTAAAATGAAAT
LbPhil     TTGAAGACAAAGTACTAAAATGA---
TaPhil     TCGAAGACAAAATACTAAAATGA---
MmdChina   TCGAAGACAAAATACTAAAATTATAT
RnHai      TTGAAGACAAAATGTTAAAATGAAAT
```

APPENDIX 5: 676 nucleotide Cytochrome *b* consensus sequences

```
                   10        20        30        40        50        60        70
SA75       ATGACAAACATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA136      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA141      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA174      ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAA-TCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA175      ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA213      ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA217      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA218      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA219      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA220      ATGACAAACATCCGAAAATCTCACCCCCTATCCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA221      ATGA-AAA-ATCCGAAAATCTCACCCCCTATCCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA232      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA233      ATGACAAA-ATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA234      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA235      A-GACAAA-ATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA252      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
SA253      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
SA254      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
SA255      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA256      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA257      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA258      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
SA259      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC79      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
ARC101     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC170     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
ARC171     ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
UP02       ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
R1         ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
PS1        ATGACAAACATCCGAAA-TCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
HP1        -----AAACATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
HP2        ATGACAAA-ATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
S1         -----AAA-ATCCGAAAATCACACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnUSA      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnCopen    ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnChina    ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RrJap      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RrTit      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RrHua      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RrSam      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RtAl       ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
RtYog      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
RtJak      ATGACAAACATTCGGAAATCACACCCCCTACTAAAAATCATTAACCACTCCTTCATTGATCTTCCAGCCC
RtJap      ATGACAAACATTCGAAAATCACACCCCCTAATCAAAATTATTAATCACTCCTTCATTGACCTTCCCGCCC
RtPhil     ATGACAAACATTCGAAAATCACACCCCCTACTCAAAATCATTAATCACTCCTTCATCGACCTTCCCGCCC
RaJap      ATGACAAACATTCGGAAATCACACCCCCTATTCAAAATTATTAACCACTCCTTCATCGACCTTCCTGCCC
RevPhil    ATGACAAACATCCGAAAATCTCACCCTCTACTCAAAATCATTAACCACTCTTTCATCGACCTTCCTGCCC
RexPhil    ATGACAAACATTCGAAAATCTCATCCTCTACTCAAAATTATTAATCACTCCTTTATTGACCTTCCCGCTC
RpPhil     ATGACAAACATTCGAAAATCCCACCCTCTGCTCAAAATTATTAATCATTCTTTCATCGACCTTCCCGCCC
DlJap      ATGGCAAACATTCGAAAATCCCACCCTCTAATCAAAATTATTAACCACTCTTTCATCGACCTCCCCGCCC
LbPhil     ATGACAAACATCCGAAAATCTCACCCCTTACTCAAAATTATTAACCACTCCTTCATCGACCTCCCCGCCC
TaPhil     ATGACAAACATCCGAAAATCTCACCCTCTGCTTAAAATTATTAACCACTCCTTCATCGACCTCCCCGCCC
MmdChina   ATGACAAACATACGAAAAACACACCCATTATTTAAAATTATTAACCACTCATTCATTGACCTACCTGCCC
RnRai      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnHua      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RnHai      ATGACAAACATCCGAAAATCTCACCCCCTATTCAAAATCATCAACCACTCCTTTATCGACCTCCCCGCCC
RrMoi      ATGACAAACATTCGAAAATCTCACCCTCTATTCAAAATTATCAACCACTCCTTCATTGACCTTCCTGCCC
RrKL       ATGACAAACATTCGAAAATCACACCCCCTACTCAAAATCATTAATCACTCCTTCATCGACCTTCCCGCCC
RtNS       ATGACAAACATTCGAAAATCACACCCCCTACTCAAAATCATTAATCACTCCTTCATCGACCTTCCCGCCC
```

```
                   80        90       100       110       120       130       140
SA75       CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA136      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA141      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA174      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA175      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA213      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA217      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA218      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA219      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA220      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA221      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA232      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA233      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA234      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA235      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA252      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
SA253      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA254      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA255      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA256      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA257      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA258      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
SA259      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC79      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC101     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC170     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
ARC171     CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTC-ATAGTTCAAATTATCA
UP02       CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
R1         CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
PS1        CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
HP1        CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
HP2        CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
S1         CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnUSA      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnCopen    CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATTCTCA
RnChina    CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
```

```
RrJap      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RrTit      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RrHua      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RrSam      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RtAl       CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAATTCAAATTATCA
RtYog      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RtJak      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAATTCAAATTATCA
RtJap      CATCCAACATCTCATCATGATGAAATTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RtPhil     CATCCAACATCTCATCATGATGAAATTTTGGCTCTTTACTAGGAGTATGCCTC-ATAGTCCAAATTATCA
RaJap      CATCTAATATCTCCTCATGATGAAACTTTGGTTCCCTACTAGGAGTATGCCTC-ATGATTCAAATTGTTA
RevPhil    CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
RexPhil    CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTATTAGGAGTATGCCTA-ATAATTCAAATCATTA
RpPhil     CATCCAACATCTCATCATGATGAAACTTCGGTTCTCTCCTAGGAGTATGCCTT-ATGATTCAAATCATCA
DlJap      CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-ATAGTTCAAATTATCA
LbPhil     CATCCAACATCTCATCATGATGAAACTTTGGCTCTCTCCTAGGAGTATGCCTT-GTAATTCAAATTATCA
TaPhil     CATCCAACATTTCATCATGATGAAACTTTGGCTCTCTTCTAGGAGTATGCCTT-ATAATCCAAATTATCA
MmdChina   CATCCAACATTTCATCATGATGAAACTTTGGGTCCCTTCTAGGAGTCTGCCTA-ATAGTCCAAATCATTA
RnRai      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnHua      CATCTAACATCTCATCATGATGAAACTTCGGTTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RnHai      CATCTAACATCTCATCATGATGAAACTTCGGCTCTCTACTAGGAGTATGCCTC-ATAGTACAAATCCTCA
RrMoi      CATCCAACATCTCATCATGATGAAACTTCGGTTCCCTCCTAGGAGTATGCCTC-ATAATTCAAATCATCA
RrKL       CATCCAACATCTCATCATGATGAAATTTTGGCTCTTTACTAGGAGTATGCCTC-ATAGTCCAAATTATCA
RtNS       CATCCAACATCTCATCATGATGAAATTTTGGCTCTTTACTAGGAGTATGCCTC-ATAGTCCAAATTATCA

                         150       160       170       180       190       200       210
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA136      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA141      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA174      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA175      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA213      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA217      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA218      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA219      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA220      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA221      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA232      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA233      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA234      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA235      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA252      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
SA253      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
SA254      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
SA255      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA256      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA257      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA258      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
SA259      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC79      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
ARC101     CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC170     CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
ARC171     CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
UP02       CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
R1         CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
PS1        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
HP1        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
HP2        CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
S1         CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnUSA      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnCopen    CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnChina    CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RrJap      CAGGCTTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RrTit      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RrHua      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RrSam      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RtAl       CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
RtYog      CAGGCCTATTCCTAGCAATACATTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
RtJak      CAGGCCTATTCCTAGCAATACATTACACATCCGACACCTTAACAGCATTCTCATCAGTTACCCACATCTG
RtJap      CAGGATTATTTCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RtPhil     CAGGCCTATTCCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
RaJap      CAGGCCTGTTCCTAGCAATACACTACACATCCGATACTTTAACAGCATTCTCATCAGTTACTCACATCTG
RevPhil    CAGGCTTATTCCTAGCAATACACTACACATCCGACACTATAACAGCATTTTCATCAGTTGCCCACATCTG
RexPhil    CAGGCCTGTTCCTAGCAATACACTACACATCCGACACTCTAACAGCATTCTCATCAGTTACCCACATCTG
RpPhil     CAGGCCTATTCCTAGCAATACACTATACATCCGACACTACGACAGCATTCTCATCAGTCACCCACATCTG
DlJap      CAGGTCTATTCCTAGCAATACACTACACATCTGACACTACTACAGCATTCTCATCAGTCGCCCACATCTG
LbPhil     CGGGCCTCTTCCTAGCAATACACTACACATCCGACACCATAACAGCATTCTCATCAGTCACCCATATCTG
TaPhil     CAGGTCTATTCCTAGCAATACACTACACATCCGATACAATAACAGCATTCTCATCAGTCACCCACATCTG
MmdChina   CAGGTCTTTTCTTAGCCATACACTACACATCAGATACAATAACAGCCTTTTCATCAGTAACACACATTTG
RnRai      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnHua      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RnHai      CAGGCTTATTCCTAGCAATACACTACACGTCTGATACCATAACAGCATTTTCATCAGTCACCCACATCTG
RrMoi      CAGGCCTATTCCTAGCAATACACTACACATCAGACACCACAACAGCATTCTCATCAGTCACCCATATCTG
RrKL       CAGGCCTATTCCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG
RtNS       CAGGCCTATTCCTAGCAATACACTACACATCCGACACTTTAACAGCATTCTCATCAGTTACCCACATCTG

                         220       230       240       250       260       270       280
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA136      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA141      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA174      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA175      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA213      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA217      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA218      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA219      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA220      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA221      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA232      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
```

```
SA233      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA234      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA235      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA252      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
SA253      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
SA254      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
SA255      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA256      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA257      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA258      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
SA259      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC79      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
ARC101     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC170     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
ARC171     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
UP02       C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
R1         C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
PS1        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
HP1        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
HP2        C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
S1         C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnUSA      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnCopen    C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACAGCCCAACGGCGCCTCAATATTTTTCATCTG
RnChina    C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RrJap      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RrTit      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RrHua      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RrSam      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RtAl       C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
RtYog      C-CGAGACGTAAACTATGGCTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
RtJak      C-CGAGACGTAAACTATGGGTGACTAATCCGATACTTACATGCCAATGGAGCCTCAATATTCTTTATCTG
RtJap      C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCAATATTCTTTATCTG
RtPhil     C-CGAGACGTAAACTACGGCTGACTAATCCGATACTTACACGCCAACGGAGCTTCAATATTCTTTATCTG
RaJap      C-CGAGACGTAAACTACGGCTGATTAATCCGATATTTACACGCCAACGGGGCCTCAATATTCTTCATCTG
RevPhil    C-CGAGATGTAAACTACGGCTGACTAATCCGATACTTACATGCCAACGGAGCCTCCATATTCTTTATCTG
RexPhil    C-CGAGACGTAAACTACGGCTGACTAATTCGATACTTACATGCCAACGGAGCTTCAATATTCTTTATCTG
RpPhil     C-CGAGATGTAAACTACGGCTGACTAATTCGATATCTACATGCCAACGGAGCCTCTACATTCTTCATCTG
DlJap      C-CGAGATGTAAACTACGGCTGACTAATCCGATATATACATGCCAACGGAGCCTCAATATTTTTTATCTG
LbPhil     C-CGAGACGTAAATTACGGCTGACTAATTCGATACCTACATGCCAACGGAGCCTCAATATTTTTCATCTG
TaPhil     T-CGAGATGTAAACTACGGCTGACTAATTCGATACTTACATGCCAATGGAGCCTCAATATTTTTCATCTG
MmdChina   T-CGAGACGTAAATTACGGGTGACTAATCCGATATATACACGCAAACGGAGCCTCAATATTTTTTATTTG
RnRai      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnHua      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RnHai      C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGCGCCTCAATATTTTTCATCTG
RrMoi      C-CGAGATGTAAATTACGGCTGACTAATTCGATACTTACATGCCAATGGAGCCTCTATATTCTTCATCTG
RrKL       C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGAGCTTCAATATTCTTTATCTG
RtNS       C-CGAGACGTAAACTACGGCTGACTAATCCGATACCTACACGCCAACGGAGCTTCAATATTCTTTATCTG
```

```
                     290       300       310       320       330       340       350
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA136      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA141      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA174      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA175      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA213      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA217      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA218      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA219      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA220      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA221      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA232      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA233      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA234      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA235      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA252      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
SA253      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
SA254      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
SA255      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA256      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA257      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA258      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
SA259      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
ARC79      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
ARC101     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCCTAGAAACATGAAACATTGGA
ARC170     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
ARC171     CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
UP02       CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
R1         CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGG
PS1        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
HP1        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
HP2        CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
S1         CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnUSA      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnCopen    CCTATTCCTCCATGTGGGGCGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnChina    CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGA
RrJap      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCCTAGAAACATGAAACATTGGA
RrTit      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
RrHua      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
RrSam      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
RtAl       CTTATTCCTCCATGTAGGCCGAGGTATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGA
RtYog      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGA
RtJak      TTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTTTTAGAAACATGAAACATTGGA
RtJap      CTTATTCCTCCATGTAGGCCGAGGGATATACTACGGATCCTACACCTTCTTAGAAACATGAAACATTGGA
RtPhil     CCTATTCCTTCATGTAGGGCGAGGGATATACTACGGATCCTATACTTTCCTAGAAACATGAAACATTGGA
RaJap      CTTATTCCTTCACGTAGGACGAGGAATATACTACGGATCCTACACTTTCCTAGAGACATGAAATATCGGG
RevPhil    CCTATTCCTACACGTAGGACGAGGGATATATTACGGATCCTACACCTTTCTAGAGACATGAAATATTGGA
RexPhil    CTTATTCCTCCACGTAGGACGAGGAATGTACTATGGATCCTACACCTTCCTAGAAACATGAAACATCGGA
RpPhil     CCTATACCTCCATGTAGGACGAGGAATGTATTACGGATCCTACACCTTCCTAGAAACATGAAACATCGGA
DlJap      CTTATTCCTTCACGTAGGACGAGGGATATACTATGGATCCTATACCTTCCTAGAAACATGAAACATTGGA
```

```
LbPhil     CTTATTCCTCCATGTAGGACGAGGAATGTACTACGGATCCTATACCTTTCTAGAGACATGAAATATTGGA
TaPhil     CTTATTCCTCCATGTAGGACGAGGGATGTACTACGGATCCTACACCTTTCTAGAAACATGAAATATTGGA
MmdChina   CTTATTCCTTCATGTCGGACGAGGCTTATATTATGGATCATATACATTTATAGAAACCTGAAACATTGGA
RnRai      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnHua      CCTATTCCTCCATGTGGGACGAGGACTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RnHai      CCTATTCCTCCATGTGGGACGAGGCCTATACTATGGATCCTACACTTTCCTAGAAACCTGAAACATTGGG
RrMoi      CCTATACCTCCACGTAGGACGAGGTATGTACTACGGATCCTACACCTTCCTAGAAACATGAAACATTGGA
RrKL       CCTATTCCTTCATGTAGGGCGAGGGATATACTACGGATCCTATACTTTCCTAGAAACATGAAACATTGGA
RtNS       CCTATTCCTTCATGTAGGGCGAGGGATATACTACGGATCCTATACTTTCCTAGAAACATGAAACATTGGA

                     360       370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA136      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA141      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA174      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA175      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA213      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA217      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA218      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA219      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA220      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA221      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA232      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA233      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA234      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA235      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA252      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
SA253      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
SA254      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
SA255      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA256      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA257      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA258      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
SA259      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC79      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
ARC101     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC170     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
ARC171     ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
UP02       GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
R1         GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCAGGAGGGCAAATATCAT
PS1        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
HP1        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
HP2        ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
S1         ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnUSA      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnCopen    ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnChina    ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RrJap      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RrTit      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RrHua      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RrSam      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RtAl       GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
RtYog      GTTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
RtJak      ATTGTCCTATTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTTCCATGAGGACAAATATCAT
RtJap      ATTATCCTACTATTTGCAGTCATAGCAACCGCATTCATAGGTTATGTACTCCCATGAGGACAAATATCAT
RtPhil     GTAGTCCTACTATTTGCAGTCATAGCAACTGCATTCATAGGCTATGTACTCCCATGAGGACAAATATCAT
RaJap      GTTGTCCTTCTATTTGCAGTCATAGCAACCGCATTCATAGGCTATGTTCTTCCATGAGGACAAATATCAT
RevPhil    GTCGTCCTATTATTTACAGTCATAGCAACCGCATTCATAGGCTACGTGCTCCCATGAGGACAAATATCCT
RexPhil    ATCGTCCTACTATTTGCAGTCATAGCCACCGCATTCATAGGCTACGTACTTCCATGAGGACAAATATCAT
RpPhil     GTTATCCTGCTATTCGCAGTTATAGCAACCGCATTCATAGGCTACGTACTTCCATGAGGACAAATATCCT
DlJap      ATCATCCTACTATTCACAGTCATAGCAACCGCATTCATAGGCTATGTACTTCCATGAGGACAAATATCCT
LbPhil     GTCGTCCTGTTATTTGCAGTCATAGCAACCGCATTCATAGGCTATGTCCTTCCATGAGGACAGATATCCT
TaPhil     ATTATCCTACTATTTACAGTCATAGCAACCGCATTCATAGGCTACGTCCTTCCATGAGGACAAATATCCT
MmdChina   GTACTTCTACTGTTCGCAGTCATAGCCACAGCATTTATAGGTTACGTCCTTCCATGAGGACAAATATCAT
RnRai      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnHua      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RnHai      ATCATTCTACTATTTGCAGTCATAGCAACTGCATTCATGGGCTATGTACTCCCATGAGGACAAATATCAT
RrMoi      GTTATCCTACTATTTGCAGTTATAGCAACCGCATTCATAGGATATGTACTTCCATGAGGACAAATGTCTT
RrKL       GTAGTCCTACTATTTGCAGTCATAGCAACTGCATTCATAGGCTATGTACTCCCATGAGGACAAATATCAT
RtNS       GTAGTCCTACTATTTGCAGTCATAGCAACTGCATTCATAGGCTATGTACTCCCATGAGGACAAATATCAT

                     430       440       450       460       470       480       490
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA136      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA141      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA174      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA175      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA213      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA217      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA218      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA219      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA220      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA221      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA232      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA233      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA234      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA235      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA252      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
SA253      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
SA254      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
SA255      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA256      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA257      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA258      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
SA259      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
ARC79      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
ARC101     TCTGAGGAGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
ARC170     TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
```

```
ARC171     TCTGGGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
UP02       TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
R1         TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
PS1        TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
HP1        TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
HP2        TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
S1         TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnUSA      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnCopen    TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnChina    TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RrJap      TCTGAGGAGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RrTit      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RrHua      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RrSam      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RtAl       TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATTCCCTATATTGGCACCACCCTAGTCGAAT
RtYog      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATTCCCTATATTGGCGCCACCCTAGTCGAAT
RtJak      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCTATCCCCTATATTGGCACCACCCTAGTCGAAT
RtJap      TCTGAGGGGCCACAGTAATCACAAACCTATTA-TCAGCCATTCCCTACATTGGCACCACTCTAGTCGAAT
RtPhil     TCTGAGGGGCCACAGTAATTACAAACCTATTA-TCAGCCATCCCCTATATTGGCACCACCCTAGTCGAAT
RaJap      TTTGAGGAGCCACCGTAATCACAAACCTATTA-TCAGCTATTCCCTATATTGGCACCACCCTAGTCGAAT
RevPhil    TCTGAGGGGCTACAGTAATTACAAACCTACTA-TCAGCTATCCCCTATATCGGAACAACCTTAGTCGAAT
RexPhil    TCTGAGGAGCCACAGTAATTACAAATCTACTA-TCAGCTATCCCCTATATCGGCACTACCCTAGTCGAAT
RpPhil     TCTGAGGCGCCACAGTTATCACAAATCTTCTA-TCAGCTATCCCTTATATCGGAACTACCCTAGTCGAAT
DlJap      TCTGAGGGGCTACAGTAATCACAAACCTGCTA-TCAGCCATCCCCTACATCGGAACTACATTAGTTGAAT
LbPhil     TCTGAGGAGCTACAGTAATTACAAACCTACTA-TCAGCTATCCCTTATATCGGAACTACCCTGGTCGAAT
TaPhil     TCTGAGGGGCTACAGTAATTACAAATCTACTA-TCAGCCATCCCTTACATCGGAACTACCCTAGTTGAAT
MmdChina   TCTGAGGTGCCACAGTTATTACAAACCTCCTA-TCAGCCATCCCATATATTGGAACAACCCTAGTCGAAT
RnRai      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnHua      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCTGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RnHai      TCTGAGGAGCTACAGTAATTACAAACCTATTA-TCAGCTATCCCTTACATTGGGACTACCCTAGTCGAAT
RrMoi      TCTGAGGTGCCACAGTAATTACAAACCTACTA-TCAGCTATCCCCTATGTCGGAACCACCCTAGTCGAAT
RrKL       TCTGAGGGGCCACAGTAATTACAAACCTATTA-TCAGCCATCCCCTATATTGGCACCACCCTAGTCGAAT
RtNS       TCTGAGGGGCCACAGTAATTACAAACCTATTA-TCAGCCATCCCCTATATTGGCACCACCCTAGTCGAAT

                        500       510       520       530       540       550       560
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA136      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA141      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA174      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA175      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA213      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA217      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA218      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA219      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA220      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA221      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA232      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA233      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA234      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA235      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA252      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
SA253      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
SA254      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
SA255      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA256      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA257      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA258      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
SA259      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC79      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
ARC101     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC170     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
ARC171     GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
UP02       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
R1         GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
PS1        GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
HP1        GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC
HP2        GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
S1         GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RnUSA      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTTATCCTCCC
RnCopen    GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RnChina    GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RrJap      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RrTit      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RrHua      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RrSam      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RtAl       GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
RtYog      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
RtJak      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTTGCATTCCACTTCATCCTCCC
RtJap      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
RtPhil     GAATCTGAGGAGGTTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC
RaJap      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTTCACTTTATCCTCCC
RevPhil    GGATCTGAGGGGGATTCTCAGTAGATAAAGCAACTTTAACACGATTTTTCGCATTTCACTTCATCCTCCC
RexPhil    GAATCTGAGGGGGATTCTCAGTAGACAAAGCAACCCTAACTCGTTTTTTCGCATTTCACTTCATTCTCCC
RpPhil     GAATTTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACTCGCTTTTTCGCCTTCCACTTCATTCTCCC
DlJap      GAATTTGAGGCGGTTTCTCAGTAGACAAAGCAACCCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
LbPhil     GAATTTGAGGAGGCTTCTCAGTAGACAAAGCCACTCTAACACGCTTCTTCGCATTTCACTTCATTCTCCC
TaPhil     GAATCTGAGGTGGATTCTCAGTAGACAAAGCAACTCTAACACGTTTTTTCGCATTCCACTTCATCCTCCC
MmdChina   GAATTTGAGGAGGCTTCTCAGTAGACAAAGCCACCTTGACCCGATTCTTCGCTTTCCACTTCATCTTACC
RnRai      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTTATCCTCCC
RnHua      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RnHai      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCCTAACACGCTTCTTCGCATTCCACTTCATCCTCCC
RrMoi      GAATCTGAGGAGGCTTCTCAGTAGACAAAGCAACCTTAACTCGTTTTTTCGCCTTCCACTTCATCCTCCC
RrKL       GAATCTGAGGAGGTTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC
RtNS       GAATCTGAGGAGGTTTCTCAGTAGACAAAGCAACCCTAACACGCTTTTTCGCATTCCACTTCATCCTCCC

                        570       580       590       600       610       620       630
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA75       ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA136      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
```

```
SA141      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA174      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA175      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA213      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA217      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA218      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA219      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA220      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA221      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA232      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA233      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA234      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA235      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA252      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
SA253      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA254      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA255      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA256      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA257      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA258      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
SA259      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC79      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC101     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC170     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
ARC171     ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
UP02       ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
R1         ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
PS1        ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
HP1        ATTCATTATCGCCGCCCTTGCAATTGTACATCTCCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
HP2        ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
S1         ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnUSA      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnCopen    ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnChina    ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RrJap      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RrTit      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RrHua      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RrSam      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtAl       ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtYog      ATTTATCATTGCCGCCCTTGCAATTGTGCATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtJak      ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtJap      ATTCATCATCGCCGCCCTTGCAATTGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtPhil     ATTCATCATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RaJap      ATTCATCATTGCCGCTCTTGCAATTGTACACCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RevPhil    ATTTATCATCGCCGCCCTCGCAATCGTCCACCTTCTTTTCCTCCACGAAACAGGGTCAAACAACCCCACA
RexPhil    ATTCATCATCGCCGCTCTTGCAATCGTACATCTCCTTTTCCTCCACGAAACAGGATCAAACAACCCCACA
RpPhil     ATTTATCATCGCCGCCCTTGCAATTGTACATCTCCTTTTCCTACACGAAACAGGATCAAACAACCCCACA
DlJap      ATTTATTATCGCCGCTCTTGCAATTGTACACCTCCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
LbPhil     ATTCATTATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCTACA
TaPhil     ATTTATTATTGCTGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCTACA
MmdChina   ATTTATTATCGCGGCCCTAGCAATCGTTCACCTCCTTTTCCTCCACGAAACAGGATCAAACAACCCAACA
RnRai      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnHua      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RnHai      ATTCATTATCGCCGCCCTTGCAATTGTACATCTTCTTTTCCTCCACGAAACAGGATCAAATAACCCCACA
RrMoi      ATTTATCATTGCCGCTCTTGCAATTGTACATCTCCTTTTTCTTCACGAAACAGGATCAAATAACCCCGCA
RrKL       ATTCATCATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
RtNS       ATTCATCATCGCCGCCCTTGCAATCGTACATCTCCTCTTCCTCCACGAAACAGGATCAAACAACCCCACA
```

```
                    640       650       660       670
           ....|....|....|....|....|....|....|....|.
SA75       GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA136      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA141      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA174      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA175      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA213      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA217      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA218      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA219      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA220      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA221      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA232      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA233      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA234      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA235      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA252      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
SA253      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA254      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA255      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA256      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA257      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA258      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
SA259      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
ARC79      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
ARC101     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
ARC170     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
ARC171     GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
UP02       GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
R1         GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
PS1        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
HP1        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
HP2        GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
S1         GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
RnUSA      GGATTAAACTCCAACGCAGACAAAATCCCATTCCATCCATATTATA
RnCopen    GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
RnChina    GGATTAAACTCCAACGCAGACAAAATCCCATTCCATCCATATTATA
RrJap      GGACTAAACTCTGACGCAGACAAAATCCCATTCCATCCATACTACA
RrTit      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
RrHua      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
RrSam      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
```

```
RtAl       GGACTAAACTCTGACGCAGACATAATCCCATTTCATCCATACTACA
RtYog      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
RtJak      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
RtJap      GGACTAAACTCTGACGCAGACAAAATCCCATTTCATCCATACTACA
RtPhil     GGGCTAAACTCTGACGCAGACAAAATTCCATTTCATCCATACTACA
RaJap      GGACTAAATTCCGACGCAGACAAAATCCCATTCCACCCATACTATA
RevPhil    GGGCTAAACTCCGACGCAGACAAAATCCCATTCCATCCATACTATA
RexPhil    GGCCTAAATTCTGATGCAGATAAAATCCCATTTCATCCATACTACA
RpPhil     GGACTAGACTCCAACGCAGACAAAATCCCATTCCATCCATATTACA
DlJap      GGACTAGATTCCAACGCAGACAAAATCCCATTTCACCCCTACTATA
LbPhil     GGACTAAATTCCAATGCAGACAAAATCCCATTTCATCCATATTATA
TaPhil     GGACTAAACTCTGACGCAGACAAAATCCCATTCCACCCATACTACA
MmdChina   GGATTAAACTCAGATGCAGATAAAATTCCATTTCACCCCTACTATA
RnRai      GGATTAAACTCCAACGCAGACAAAATCCCATTCCATCCATATTATA
RnHua      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
RnHai      GGATTAAACTCCGACGCAGACAAAATCCCATTCCATCCATATTATA
RrMoi      GGACTAAACTCCGACGCGGACAAAATCCCATTCCACCCATATTACA
RrKL       GGGCTAAACTCTGACGCAGACAAAATTCCATTTCATCCATACTACA
RtNS       GGGCTAAACTCTGACGCAGACAAAATTCCATTTCATCCATACTACA
```

APPENDIX 6: D-loop consensus sequences

```
                10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA141    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA75     CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA174    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA175    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA213    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA218    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA221    -----CATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA233    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA234    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA252    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
SA253    CAGTACATA-----AAATTATCTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
SA254    CAGTACATA-----AAATTATCTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
SA255    CAGTACATA-----AAATTATTTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
SA256    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
SA258    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
SA259    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
PS1      CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
R1       CAGTACATA-----AAATTATCTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
S1       CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
HP1      CAGTACATA-----AAATGATATAGGTACTTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
HP2      CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
RnItaly  CAGTACATA-----AAATGATAT-GGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
RnUSA    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
RnCopen  CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
RnRai    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATTGTACATTAAATTATTTTCCCC
RnHua    CAGTACATA-----AAATGATATAGGACATTAAAACATTTATGTATATCGTACATTAAATTATTTTCCCC
RrUSA    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RrTit    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RrHua    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RrSam    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RrMad    CAGTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RrMoi    CAGTACATA-----AAATGATATAGGACATTAAGACATTAATGTATATCGTGCATTAATTTATTTTCCCC
RrKL     CAGTACATA-----AAATTATCTAGAACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RtNS     CAGTACATA-----AAATTATCTAGAACATTAAAACATTTATGTATATCGTACATTAATTTATTTTCCCC
RtYog    CAGTACATA-----AAATTATCTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
RtAl     CAGTACATA-----AAATTATTTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
RtHK     CAGTACATA-----AAATTATTTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
RtJak    CAGTACATA-----AAATTATTTAGGACATTAAAACATTTATGTATATCGTACATTAATTTATTTCCCCC
RexNZ    -AGTACATA-----AAATGACATAGGACATTAAAACATTTATGTATATCGTACATTAATTTATCTACCCC
RfAust   CAGTACATA-----AAATGATATAGGACATTAATACATTAATGTATATCGTGCATTAATTTCTTTTCCCC
MmTurk   --GTACATA-----AATTTACATAGTACAATAGTACATTTATGTATATCGTACATTAAATTATTTCCCCC
McSA     C-GTACATA-----AAATTATCTAGTACATTAAAACATTTATGTATATCGTACATTAAATTATTATCCCC

                80        90       100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA141    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA75     A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA174    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA175    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA213    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA218    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA221    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA233    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA234    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA252    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
SA253    A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATGA-AC-TTAAATTTAACTAA--AA
SA254    A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATGA-AC-TTAAATTTAACTAA--AA
SA255    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAGCTAA--AA
SA256    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
SA258    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
SA259    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
PS1      A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
R1       A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATGA-AC-TTAAATTTAACTAA--AA
S1       A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
HP1      A---AGCATATAAGCA-TGTAATATATATCTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
HP2      A---AGCATATAAGCA-TGTAATATATATCTAATGATTTAGGACATCG-AT-TTAAACTCAACTAT--AA
RnItaly  A---AGCATATAAGCA-TGTAATATATATCTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
RnUSA    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
RnCopen  A---AGCATATAAGCA-TGTAATATATATCTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
RnRai    A---AGCATATAAGCA-TGTAATATATATCTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
RnHua    A---AGCATATAAGCA-TGTAATATATATTTAATGATTTAGGACATAC-AT-TTAAACTCAACTAT--AA
RrUSA    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
RrTit    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
RrHua    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
RrSam    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGACATAA-AT-TTAAACTCAACTAA--AA
RrMad    A---AGCATATAAGCA-TGTAATATATAATTAATGTATTAAGATATAA-AT-TTAAACTCAACTAA--AA
RrMoi    A---TGCATATAAGCAATGTAAATTTTAATTAATGTACTAAGACATAATACTTAAAA--CAACTTA--AA
RrKL     A---AGCATATAAGCA-TGTAATATA-AATTAATGTATTAAGACATCAAAT-TCAAATCTAACTAA--AA
RtNS     A---AGCATATAAGCA-TGTAATATA-AATTAATGTATTAAGACATCAAAT-TCAAATCTAACTAA--AA
RtYog    A---AGCATATAAGCA-CGTAATATA-AATTAATGTACTAAGACATAA-AT-TTAAATTTAACTAA--AA
RtAl     A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATAA-AT-TTAAGTTTAACTAA--AA
RtHK     A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATAA-AT-TTAAGTTTAACTAA--AA
RtJak    A---AGCATATAAGCA-CGTAATATA-AATTAATGTATTAAGACATAA-AT-TTAAGTTTAACTAA--AA
RexNZ    A---AGCATATAAGCA-TGTAATATA-AGTCAATGTATTAAGACATGATAT-TTAAACTCAACTAG--AA
RfAust   A---TGCATATAAGCA-TGTATATTCTAGTTAATGTATTAGGATATAACA--ATCAAATCAACT-T--GA
MmTurk   A---AGCATATAAGCA-CGTAAATTATA-TTAATGATATCACACATAAAAACTAATACTCAACATA--AC
McSA     A---AGCATATAAGCA-GGTATTATATA-TTAATGTAATAGGACATAAA--TTATATTTCCACATT--AA

               150       160       170       180       190       200       210
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA141    A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA75     A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA174    A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA175    A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA213    A--TTCATAACAATATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA218    A--TTCATAACAATATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA221    A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA233    A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
```

```
SA234     A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA252     A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
SA253     A--TCCAACCCAACAGGAATATTCT---TTTAA--ACATTAAAATAATGCTTTAAA--GACATATCTGTG
SA254     A--TCCAACCCAACAGGAATATTCT---TTTAA--ACATTAAAATAATGCTTTAAA--GACATATCTGTG
SA255     A--TTTAAACCAACATGAATATTCT---TTCGAT-ACATTAAGATAATGTTTTAAA--GACATATCTGTG
SA256     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
SA258     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
SA259     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
PS1       A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
R1        A--TCCAACCCAACAGGAATATTCT---TTTAA--ACATTAAAATAATGCTTTAAA--GACATATCTGTG
S1        A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
HP1       A--TTCATAATAACATGTCTATTCT---TA-AAT-ACATTAAGATAATGCTTATTA--GACATATCTGTG
HP2       A--TTCATAATAACATGTCTATTCT---TA-AAT-ACATTAAGATAATGCTTATTA--GACATATCTGTG
RnItaly   A--TTCACAACAACATGTCTATTCT---CA-AAT-ACATTAAGATAATGCTTATTA--GACATATCTGTG
RnUSA     A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
RnCopen   A--TTCACAACAACATGTCTACTCT---CA-AAT-ACATTAAGATAATGCTTATTA--GACATATCTGTG
RnRai     A--TTTACAACAACATGTCTATTCT---CA-AAT-ACATTAAGATAATGCTTATTA--GACATATCTGTG
RnHua     A--TTCATAACAACATGTCTATTCT---CG-AAT-ACATTAAAATAATGCTTATTA--GACATATCTGTG
RrUSA     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
RrTit     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
RrHua     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
RrSam     A--TTTAAACCAACATGAATATTCT---TTCAT--ACATTAAGATAATGTTTTAAA--GACATATCTGTG
RrMad     A--TTTAAGCCAACATGAATATTCT---TTCAT--ACATTAAAATAATGTTTTAAA--GACATATCTGTG
RrMoi     T--CCTATAACAACATGGATATTCT-----TAAATACATTAAGATAATGTTTTCCG--GACATATCTGTG
RrKL      A--CTCATATCAACATGAATATTCT---ACAAT--ACATTAAAATAATGCTTTATA--GACATATCTGTG
RtNS      A--CTCATATCAACATGAATATTCT---ACAAT--ACATTAAAATAATGCTTTATA--GACATATCTGTG
RtYog     A--TCCAATTCAACATGAATATTCT---TTCAT--ACATTAAAATAATGCTTTAAG--GACATACCTGTG
RtAI      A--TCCAACCTAACATGAATATTCT---TTCAT--ACATTAAAATAATGTTTTAAA--GACATATCTGTG
RtHK      A--TCCAACCTAACATGAATATTCT---TTCAT--ACATTAAAATAATGTTTTAAA--GACATATCTGTG
RtJak     A--TCCAACCTAACATGAATATTCT---TTCAT--ACATTAAAATAATGTTTTAAA--GACATATCTGTG
RexNZ     A--TCCACGATAACATGAATATTCT---CACAT--ACATTAATCTAATGTTTTACG--GACATACCTGTG
RfAust    A--TTCTCTACAACATGGATATTCT----TAAAT-ACATTAAGATAATGCTTTCCG--GACATATCTGTG
MmTurk    AT-ATCACA-CACCATGAATATTATA--CTTAAT-ACATTAAATTAATGCTTTAAA--GACATATCTGTG
McSA      AC-CTCAT--CAACATGTCTATTAT---TTCAAT-ACATTCAATTAATGTTCTTAA--GACATATCTGTG

                         220       230       240       250       260       270       280
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SA141     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA75      TTATTAGACATGCACCATTAA-GTCATAAA----------------------------------------
SA174     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA175     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA213     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA218     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA221     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA233     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA234     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
SA252     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATTCCTGTCCCCAATTGGT
SA253     TTATTAGACATACACCATTAAAGTCATGAATTCTTTCTCTTCCATATGACTATCCTTGACCTAAGTTGGT
SA254     TTATTAGACATACACCATTAAAGTCATGAATTCTTTCTCTTCCATATGACTATCCTTGACCTAAGTTGGT
SA255     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
SA256     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
SA258     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
SA259     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
PS1       TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
R1        TTATTAGACATACACCATTAAAGTCATGAATTCTTTCTCTTCCATATGACTATCCTTGACCTAAGTTGGT
S1        TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
HP1       TTATTAGACATACACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
HP2       TTATTAGACATACACCATTAA-GTCATAA-C-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RnItaly   TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RnUSA     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RnCopen   TTATTAGACATACACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RnRai     TTATTAGACATACACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RnHua     TTATTAGACATGCACCATTAA-GTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGTCCCCAATTGGT
RrUSA     TTATTAGACATACACCATTAAAGTCATAAATTCTTCCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
RrTit     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
RrHua     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
RrSam     TTATTAGACATACACCATTAAAGTCATAAATTCTTTCTCTTCCATATGACTATCCTCAACCTAAGTTGGT
RrMad     TTATTAGGCATACACCATTAAAGTCATAAATTCTTCCTCTTCCATATGACTATCCTCGACCTAAGTTGGT
RrMoi     TTATTAGACATACACCATAAA-GTCATAAAC-CTTCCTCTTCCATATGACTATCCCCGACCCCAATTGGT
RrKL      TTATTAGACATACACCATTAAAGTCATAAACTCCCTCTCTTCCATACGACTATCCTCGACCCAAGTTGGT
RtNS      TTATTAGACATACACCATTAAAGTCATAAACTCCCTCTCTTCCATACGACTATCCTCGACCCAAGTTGGT
RtYog     TTATTAGACATACACCATAAAAGTCACGAATTCTTTCTTTTCCATATGACTATCCTTGACCTAAGTTGGT
RtAI      TTATTAGACATACACCATTAAAGTCATGAATTCTTCCTTTTCCATATGACTATCCTTGACCTAAGTTGGT
RtHK      TTATTAGACATACACCATTAAAGTCATGAATTCTTCCTTTTCCATATGACTATCCTTGACCTAAGTTGGT
RtJak     TTATTAGACATACACCATTAAAGTCATGAATTCTTCCTTTTCCATATGACTATCCTTGACCTAAGTTGGT
RexNZ     TTATCAGACATACACCAT-AAAGTCATAAAC-CTTTCTCTTCCATATGACTATCCCTGACCCCAGTTGGT
RfAust    TTATTAGACATACACCATAAG-GTCATAAAC-TCTTCTTTTCCATATGACTATCCTCGACCCCAATTTGT
MmTurk    TTATCTGACATACACCATAAA-GTCATAAAC-CCTTCTCTTCCATATGACTATCCCCTTCCCCATTTGGT
McSA      TTATCTTACATACACCATTAACGTCATAAAC-CTTTCTCTTCCATATGACTATCCCCTTCCCCATTTGGT

                         290       300       310       320
          ....|....|....|....|....|....|....|....|
SA141     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA75      ----------------------------------------
SA174     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA175     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA213     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA218     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA221     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA233     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA234     CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
SA252     CTCTATTTCTACCATCCTCCGGGCA-CCTC----ACACAC
SA253     CTCTATTTCTACCATCCTCCGTGCAGACCA---CTCAC--
SA254     CTCTATTTCTACCATCCTCCGTGAAGATCA---CACAA--
SA255     CCCCATTTCTACCATCCTCCGTGAA-ATCA----ACAA--
SA256     CTCTATTTCTACCATCCTCCGTGAA-ATCA----ACAA--
SA258     CTCTATTTCTACCATCCTCCGTGAA-ATCA----ACAA--
SA259     CTCTATTTCTACCATCCTCCGTGAA-ATCA----ACAA--
PS1       CTCTATTTCTACCTACCTCCGTGAA-ACCA----GCAACA
R1        CTCTATTTCTACCATCCTCCGTGAA-ATCA----ACAA--
```