Dissertations, Theses, and Masters Projects   Theses, Dissertations, & Master Projects

2001

Genetic Investigation of Interspecific and Intraspecific Relationships Within the Genus Rapana

Arminda L. Gensler
College of William and Mary - Virginia Institute of Marine Science

GENETIC INVESTIGATIONS OF INTERSPECIFIC AND INTRASPECIFIC
RELATIONSHIPS WITHIN THE GENUS *RAPANA*

_____

A Thesis
Presented to
The Faculty of the School of Marine Science
The College of William and Mary in Virginia

In Partial Fulfillment
Of the Requirements for the Degree of
Master of Science

_____

by
Arminda L. Gensler
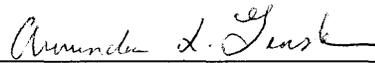2001

APPROVAL SHEET

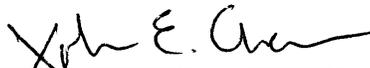This thesis is submitted in partial fulfillment of

The requirements for the degree of

Master of Science

_____
Arminda L. Gensler
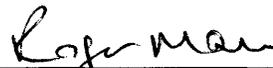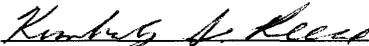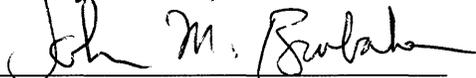
Approved, October 2001

_____
John E. Graves, Ph.D.
Co-Advisor

_____
Roger Mann, Ph.D.
Co-Advisor

_____
Kimberly S. Reece, Ph.D.

_____
John M. Brubaker, Ph.D.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The veined rapa whelk, *Rapana venosa*, is a large predatory neogastropod. While native to temperate Asian waters, *R. venosa* was found in the Black Sea in the 1940s, gradually spread through the Aegean, Adriatic, and Mediterranean Seas (1950s-present), and was recently identified in Chesapeake Bay, USA (1998) and Uruguay (1999). DNA sequencing of the 731 bp mitochondrial cytochrome *b* gene was used to investigate the genetic relationship of Chesapeake Bay *R. venosa* to a native Korean sample and to a sample from the introduced location along the Black Sea coast of Turkey. The cytochrome *b* locus from eight individuals from each location was sequenced. The haplotype diversity of the native Korean sample was significantly greater than that of either of the two samples from the introduced Chesapeake Bay or Turkish locations (h=1.0 vs. 0.0643 vs. 0.0, from, Korea, Chesapeake Bay, and Turkey, respectively). Mean nucleotide sequence divergence within populations decreased from 0.49% (SE=0.09) in Korea, to 0.14% (SE=0.05) in Chesapeake Bay, and to 0% (SE=0) in Turkey. Heterogeneity analysis of pooled haplotype frequencies revealed the three locations do not share a common gene pool (p=0.00); however, separate analysis of the Turkish and Chesapeake Bay samples failed to reject the hypothesis that the two locations comprise a single population (p=0.19). While the presence of a shared haplotype within the Turkish and Chesapeake Bay samples was consistent with the hypothesis that the source of the Chesapeake Bay *R. venosa* invasion was European, not Asian, the proximate origin of the Chesapeake Bay invasion could not be definitively determined,

The nuclear internal transcribed spacer (ITS) region was sequenced to examine taxonomic relationship of *R. venosa* relative to the putative congeneric species *R. bezoar*. ITS fragment sizes were approximately 1020 bp in *R. venosa* (n=31) and 1260 bp in *R. bezoar* (n=8). Maximum parsimony analysis indicated that individuals from the putative species *R. venosa* and *R. bezoar* represent two distinct reciprocally monophyletic clades with different evolutionary trajectories (bootstrap value = 100%). Between-group percent mean nucleotide sequence divergence was 6.1% (excluding gaps). These genetic data support the current classification of *R. venosa* and *R. bezoar* as distinct species.

GENETIC INVESTIGATIONS

OF

INTRASPECIFIC AND INTRASPECIFIC RELATIONSHIPS

WITHIN THE GENUS *RAPANA*

INTRODUCTION

Introduction to *Rapana venosa*

The gastropod *Rapana venosa* (Phylum: Mollusca, Class: Gastropoda, Order: Neogastropoda, Family: Thaididae, Genus: Rapana) was first described by Valenciennes in 1846, and was later independently described by Crosse in 1861. The Linnean name given by Crosse, *Rapana thomasiana*, is still used in scientific literature today; however, most researchers choose to use the original designation *R. venosa*. Common names for *R. venosa* include rapa whelk, veined rapa whelk, Thomas' rapa whelk, and Asian rapa whelk.

Rapana venosa is a large predatory neogastropod possessing a distinctive shell morphology. Mann and Harding (2000) describe this gastropod as having a "short spired, heavy shell, with a large inflated body whorl and a deep umbilicus. The columella is broad, smooth, and slightly concave [and] small elongate teeth are present along the edge of the large, ovate aperture". Both external and internal shell coloration are variable; the outer shell color ranges from grey to brownish orange, and the aperture and collumela can be shaded "orange, yellow, or off-white" (Mann and Harding 2000). Dark patterns of veination may be present internally. Shell length, as measured from the tip of the spire to the end of the siphonal canal, for the largest *R. venosa* documented in literature was 183 mm (Hwang et al. 1991), though a shell collection catalogue reports that a *R. venosa* of 212 mm was found in 1969 in Japan. *Rapana venosa* shell lengths in Chesapeake Bay have reached 172 mm.

The native range of *Rapana venosa* extends along temperate subtidal waters within the Sea of Japan, the Yellow Sea, the East China Sea, and the Gulf of Bohai. Adults live primarily on hard sand substrate, are often found at depths of 5-20 meters, and display diurnal burrowing with nocturnal emergence (Mann and Harding 2000). *Rapana venosa* are reported to tolerate both polluted and oxygen poor waters (Zolotarev 1996). Adult rapa whelks survive temperatures ranging from 4-27 °C, and have been found in salinities ranging from 16 to 32 ppt (Mann and Harding 2000). At the onset of colder winter water temperatures, mature whelks move from shallow nearshore habitats to deeper waters (Wu 1988, A. Occhipinti, University of Pavia, pers. comm.). In their native waters, rapa whelks feed primarily on large bivalve molluscs, and are consumed by octopus and humans.

Like many muricid gastropods, *R. venosa* females reproduce by laying large mats of egg capsules after internal fertilization. In Asian waters, the sexes are dioecious. Female *R. venosa* are capable of storing sperm, and extrusion of egg capsules can take place many days after mating has occurred. Each egg capsule can contain up to several hundred fertilized *R. venosa* eggs. The egg capsules are laid in groups known as egg mats or egg masses. A single egg mass can comprise just a few to several hundred egg capsules. Female *R. venosa* can lay two or three egg masses over the spawning season (Chung et al. 1993, J. Harding, VIMS, pers. comm.). Spawning periodicity varies with water temperature (Mann and Harding 2000). Records of spawning in Korea by Chung et al. (1993) indicate that egg masses are laid in April through July with temperatures of 13-26 °C. Reports from the Black Sea are variable. Chukhchin (1984) writes that spawning occurs from July to September (19-25 °C), while Sahin (1997) reports a

spawning from May to November in Turkey (11-19 °C, temperature data from Sevastapol).

After eggs hatch inside the egg capsule, the free swimming larvae congregate until the emergence pore at the top opens. Larval emergence from the egg capsule occurs up to 21 d after spawning, and the veliger larvae are planktonic. Most larval settlement occurs 14-28 days post emergence (Harding and Mann 1999, Chung et al. 1993), but observations indicated that they are capable of delaying metamorphosis for up to 100 d (Harding and Mann, unpublished data). Juvenile *R. venosa* preferentially settle on hard substrates and move to sandy bottomed areas upon reaching larger sizes (50 mm) or sexual maturity. Sexual maturity is believed to occur at age two (Harding and Mann, unpublished data).

The genus *Rapana*

Currently, three species are classified as belonging to the genus *Rapana: R. venosa, R. bezoar*, and *R. rapiformis*. In the South China Sea, the southern portion of *R. venosa's* range overlaps with the other two congeneric species, *R. bezoar* and *R. rapiformis*. There is some confusion regarding the taxonomic status of *R. venosa* and *R. bezoar*. In the current literature, the two are described as distinct species: *R. bezoar* is smaller and more tropical than *R. venosa*. However, some scientists believe that the two may represent conspecific (same species) populations exhibiting plastic differences in shell phenotype. The issue of taxonomy has not been resolved. Drapkin's (1963) description of the Black Sea invasion originally classified *R. venosa* as *R. bezoar*, and Kinzelbach (1986) reported *R. bezoar* was synonymous with *R. venosa*. The collection of

Rapana at the Smithsonian Institution in Washington, D.C. has been recently re-

examined; specimens originally identified as *R. bezoar* were recatalogued as *R. venosa*.

A small *R. venosa* from the Chesapeake Bay illustrated in Harding and Mann (1999)

resembles older descriptions of *R. bezoar*. Images of *R. venosa* and *R. bezoar* are shown

in Figure 1.

Some researchers believe that the differences between the taxa currently described

as *R. venosa* and *R. bezoar* may represent ecophenotypic or intraspecific variation.

Molluscs, especially gastropods, can show dramatic variation in phenotypic characters,

and different environmental stresses can cause variation in these phenotypes (Kirby et al.

1994). In the gastropod *Nucella lapillus*, environmental stresses produce a range of

phenotypic variation in traits such as reproductive strategy, growth rate, shell thickness,

shell shape, and shell color (Kirby et al. 1994).

Descriptive records of *R. venosa* and *R. bezoar* were difficult to obtain. Kira

(1961) describes *R. bezoar* as being smaller than *R. venosa,* and that in Japan *R. venosa* is

found in Hokkaido (one of the northern Japanese Islands), while *R. bezoar* exists in the

southern part of Honshuu. Morton (1994) reports that *R. bezoar* can be found on soft

subtidal bottoms in the waters south of Hong Kong. *R. bezoar* is also spread widely

through the warmer waters of the Indian and South Pacific Oceans (Mann and Harding

2000). The columella is wider on *R. bezoar* relative to the *R. venosa* (J. Harding, pers.

comm.). In color, *R. bezoar* is a faded version of *R. venosa* – the rich orange red of the

interior, as well as the warm brown outer shell coloration typical of most *R. venosa*

(Mann and Harding 2000) has been replaced by white, light brown, or cream shades (Kira

1961). Internal veination patterns are similar for both.

Figure 1: Images of *Rapana venosa* (A) and *R. bezoar* (B). The *R. venosa* photo, taken

by Dr. J. Harding (VIMS), represents an adult specimen (165 mm) recovered in

Chesapeake Bay. The *R. bezoar* image was obtained scanning a figure from the volume

Coloured Illustrations of the Shells of Japan (Kira, 1961). Kira describes *R. bezoar* as

being approximately half the size of *R. venosa*.
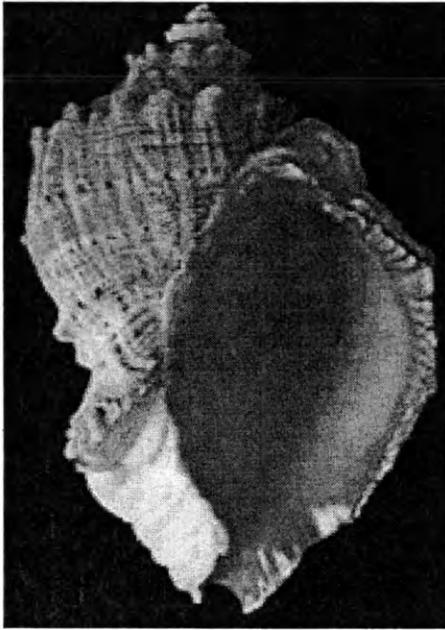
A
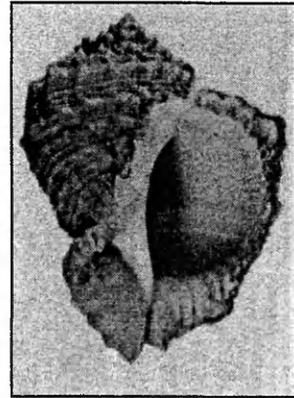
B

History of *R. venosa* invasions

Black Sea

The introduction of *R. venosa* into the Black Sea occurred in early 1940s, and was first

discovered in Novorosiysky Bay in 1946 (Konsoulova 1992). After establishment, *R.*

venosa spread along the Black Sea coast through the Crimean, Bulgarian, and Turkish

coastlines. After introduction, *R. venosa* "totaly [sic.] destroyed oyster banks...and

seriously reduced both oyster and mussel natural resources" (Konsoulova 1992).

While the destruction of large bivalve mollusc beds (composed largely of *Ostrea*

edulis, *Pecten ponticus*, and *Mytilus galloprovincialis*) was the most notable impact of *R.*

venosa introduction in the Black Sea, Drapkin (1963) and Zolotarev (1996) noted the

rapa whelk introduction restructured other ecological processes. Smaller soft sediment

molluscs such as *Venus gallina*, *Gouldia minima*, and *Pitar rudus* were also subject to

predation (Zolotarev 1996). The significant reduction of bivalve biomass resulted in a

lack of food for "bottom-feeding Black Sea fish, as well as of some plankton-eating fish,

which use bivalve larvae for food" (Drapkin, 1963). Hermit crab and sand eel

(*Ammodytes cicerellus*) populations were affected by the *R. venosa* introduction as well.

Hermit crab growth was previously limited by the smaller native Black Sea gastropod

shells (*Cerithium*, *Gibbula*, *Hinia*); when larger *R. venosa* shells were introduced, Black

Sea hermit crabs grew to larger sizes.

The mechanism for the Black Sea introduction is unknown, but Yonge and

Thompson (1976) suggest that the "impressive Japanese drill" *R. venosa* was introduced

via the transport of Japanese oysters into the Black Sea. They state that predatory

neogastropods "while still in egg capsules . . . [are] carried all over the world with the animals they consume".

Spread through the Mediterranean and other European waters

Kinzelbach (1986) reports that range expansion of the "conspicuous" non-native species *R. venosa* has been documented as proceeding from the Black Sea and Maramar Sea to the Adriatic and Tyrrhenian Seas. Koutsoubas and Voultsiadou-Koukoura (1990) describe its appearance in the Aegean Sea, and provide a map illustrating the full range expansion from the Black Sea into nearby waters. Koustoubas and Voultsiadou-Koukoura (1990) remark that the distribution of new reports of *R. venosa* in the Black Sea and Mediterranean region suggests that the species was transported by boats, though others (Mann and Harding 2000) state that the distribution could be explained by current transport of pelagic larvae. Since the mid-1990s, individuals have been collected in other European waters. *Rapana venosa* have been found off the coast of France and in the North Sea (P. Goulletquer, IFREMER DRV RA, France, pers. comm. 1999, McCarthy 1992).

Chesapeake Bay Invasion

Rapana venosa was first identified in Chesapeake Bay in August 1998. Two years later, over 1200 *R. venosa* from Chesapeake Bay were documented. The bulk of the animals have come from a region close to the port of Hampton Roads. Rapa whelks have been found in an area bounded by the Chesapeake Bay Bridge Tunnel, the mouth of the Rappahanock River, and the James River Bridge (Harding and Mann 1999).

Estimates by Harding and Mann suggest that the current records of 1200 *R. venosa* from

Chesapeake Bay may represent only a small percentage of the total adult population.

All live *R. venosa* at VIMS and most records of Chesapeake Bay *R. venosa*

originate from samples donated by or purchased from commercial fishermen. *R. venosa*

are bycatch in clam, blue crab, and conch fisheries. Thus, all current measurements of

rapa whelk densities, size distributions, and known catch locations may be biased due to

the nature of fisheries-dependent data collection.

In Chesapeake Bay, *R. venosa* experience temperature and salinity regimes

similar to their native habitat. Adult *R. venosa* from Chesapeake Bay have been found in

salinities ranging from 17-30 ppt, and the population has survived the lower Bay's annual

temperature fluctuations of 8 to 24 °C (Mann and Harding 2000). As in native Asian

waters, *R. venosa* prey upon bivalve molluscs. Studies at VIMS indicate that the rapa

whelk's prey item of choice in Chesapeake Bay is the hard clam, *Mercenaria*

mercenaria. This fact, as well as the rapa whelks' preference for the hard sand substrate

that represents important *M. mercenaria* habitat, suggests that the hard clam fishery may

be at risk. *Rapana venosa* in Chesapeake Bay exist in the same habitat as three native

whelk species: the channel whelk (*Busycotpypus canaliculatum*), knobbed whelk

(*Busycon carica*), and the lightning whelk (*Busycon contrarium*). However, while native

predators (i.e. sea turtles) can consume these thinner shelled native whelks, no known

Chesapeake Bay predators consume large *R. venosa*.

To date, all *R. venosa* taken from Chesapeake Bay are larger than 60 mm total

shell length (the biased size distribution results from the use of size selective fishing

gear). *Rapana venosa* from Chesapeake Bay are larger than those commonly found in

Korea and the Black Sea, a fact attributed to the intense fishing pressure in the latter two locations. The largest Chesapeake Bay *R. venosa* discovered to date is 172 mm TSL, corresponding to a probable age of 10-15 years (J. Harding, VIMS, pers. comm.), and indicates (if the population is a result of ballast water introduction of planktonic larvae) that *R. venosa* has been present in Chesapeake Bay for at least 10 years. If first reproduction is at age two, there may be as many as 7-14 year classes in Chesapeake Bay (J. Harding, VIMS, pers. comm.). Breeding populations of captured animals were established at VIMS in 1999 and 2000, and larval culture experiments have successfully raised *R. venosa* from egg to reproductively active adult snails.

The invasion of Chesapeake Bay by *R. venosa* was (and may be) likely mediated by ballast water transport of pelagic larvae. Ballast water discharge is now thought to account for many recent marine introductions (Ruiz et al. 1997, Carlton 1985, Williams et al. 1988). The port of Hampton Roads is one of the major commercial shipping centers and military ports on the east coast of North America, and ship traffic originating from Asian and Black Sea ports deballast outside Hampton Roads. A working hypothesis developed by R. Mann (VIMS) and G. Ruiz (Smithsonian Environmental Research Center) illustrates the theory of how ballast water introduction could bring *R. venosa* to Chesapeake Bay (Mann and Harding 2000). Ships from Hampton Roads export American coal to numerous ports along the Black Sea. Once the coal is offloaded, the empty vessels take on ballast water to stabilize the ship for the return voyage. This ballast water contains both seawater and living organisms. After reaching Hampton Roads, ships' ballast tanks are emptied, releasing both water and any surviving foreign organisms into Chesapeake Bay. Ruiz estimates that over 15 million tons of ballast water

are moved annually from ports containing reproducing *R. venosa* populations into

Hampton Roads (Harding and Mann 1999). Travel time for a commercial shipping

vessel from the Mediterranean and nearby waters (including the Black Sea) to the port of

Hampton Roads is 10 to 24 d (G. Ruiz, cited in Harding and Mann 1999). Thus, it is

realistic to assume that *R. venosa* larvae, which are usually planktonic for 14-28 d, and

can delay metamorphoses for up to 100 d (Harding and Mann, unpublished data), could

have been brought to the Chesapeake Bay via ballast water transport.

Biological invasions

       Biological invasions are defined by the National Research Council (NRC) (1996)

as "any species . . . that enters an ecosystem beyond its historic range". Invasions can be

further subdivided as either natural range expansions or as human mediated

introductions; it is the latter that has resulted in the global redistribution of many marine

and terrestrial species beyond that of their native habitats (Carlton 1999, Elton 1958).

       While human mediated marine biological invasions have received increased

scientific attention since the 1970s, mankind has been transporting nonindigeneous

species into new environments for at least 2000 years (Carlton 1999). Introductions are

"a great threat to the integrity of natural communities of plants and animals and to the

preservation of endangered species" and can result in human health risks as well as the

economic loss of millions of dollars (Carlton and Geller 1993). Introductions of such

flagship species as *Gymnodinium catenaturm* (a toxic dinoflagellete), *Dreissena*

polymorpha (zebra mussel), and *Mnemiopsis leidyi* (comb jelly) have recently brought

the bioinvasion issue to public and governmental attention (NRC 1996, Carlton and Geller 1993).

Human mediated introductions can be accomplished in a variety of ways. Historically, the most import vector for marine bioinvasions was through the accidental or intentional release of nonnative species (and their parasites) through aquaculture. Other means of introduction include release from the aquarium trade, transplantation of marine plants, migrations through canals, and vessel-facilitated transport. Prior to the 1880s, vessel transport of marine species occurred when organisms moved as part of an attached fouling community, such as boring species within the hull, or in the hard sand or rock ballast in ships (Carlton and Geller 1993, Carlton 1999). In modern ships, vessel stabilization is accomplished through the use of ballast water. Since the 1960s, vessel sizes (and thus ballast tanks) and rates of transoceanic shipping have increased, bringing more species into contact with novel environments. The NRC (1996) estimates that the shipping traffic of the world transports over 3,000 species in ballast water each day.

Ballast water is not the only ballast associated transport mechanism. Ballast tanks collect shallow layers of sediment. The mud-covered bottoms of these tanks can, even after the pumps complete deballasting of the vessel, retain several inches of water. Communities of small infaunal organisms have been observed living in this submerged region. These breeding communities could produce pelagic larvae that could be broadcast into full ballast tanks and be delivered into new colonial habitat upon deballasting. Also, sea chests (areas of the ship that hold ballast water before it is moved by the pumps into the ballast tanks) can store organisms during transoceanic journeys.

Ballast water transport is now believed to be the mechanism responsible for most modern marine introductions (Ruiz et al. 1997, Carlton 1985).

Not every organism picked up by a transport vector will become part of an established colonial population, and, as Williamson (1996) sums succinctly, "most invasions fail." The lack of invasion success is due to the many environmental filters that prevent the establishment of new species in new areas. Filters can be mechanical, biological, or ecological in origin. They exist during all steps of the invasion process. To illustrate this point, imagine a pool of propagules from a source population are pumped into the ballast tanks of a large vessel. Of the original population, only a small number are brought into the ship, and a fraction of the survivors will be crushed in the pumps or will find the ballast water inhospitable habitat. Time spent in ballast works as another filter to remove some of the survivors – some individuals will starve and others will metamorphose and will lack proper cues or substrate for settlement. After surviving the voyage, the organisms must survive the deballasting process and be able to tolerate the physical conditions of the novel environment. In addition, food and acceptable habitat must be present, and predators must be avoided. A final filter is capacity for reproduction after establishment. Invasions can fail because the numbers of introduced individuals are not sufficient to create a reproductively sustainable population.

How many successful invasions are there? The statistical "rule of tens" has been advanced to predict (1) the frequency of invasion success, and (2) the frequency of an invader having a negative impact economically, and thus becoming a "pest" species (Williamson, 1996). Williamson describes four stages a propagule must go through before establishment as a pest species:

Importation --> Introduction --> Establishment --> Pest.

Importation is defined as the chance of a species being brought into an area by a transport vector, introduction as the successful release of the propagule from the vector into the environment, and establishment as the founding of a reproductively successful population. Williamson (1996) assigns a 10% chance of a species passing from one transition to the next, but states the 10% figure is rough and that the actual figures usually run from 5 to 20%.

While Williamson's model primarily deals with the question of what fraction of invading species become established, it also provides a useful framework on which to consider propagule establishment rates in individual invasions. For example, in the context of the hypothesized ballast water transport of the pelagic larvae of *R. venosa*, the model illustrates how numbers of larvae could be reduced at each stage. Thus, the colonial habitat could be presented with only a small subset of the original larvae taken up in ballast. And, as these colonizing larvae may represent only a small percentage of the original population, one would expect a reduction in genetic diversity occur in the founding population.

Population size, genetic diversity, and invasion

Introduced populations are believed to be less genetically diverse than the original parent population (Barrett and Richardson 1986). Variation can be lost during colonization in two ways, and the interaction of these two forces results in reduction of genetic variation referred to as the founder effect. The new population contains only a subset of the alleles present in the parent population, and any alleles not included in the

founding propagules will be lost to the colonial population. After the new population is established, the subset of existing variation can be further reduced through genetic drift (bottleneck effect). While bottlenecks have been documented to result in severe reductions of genetic diversity when colonial populations are severely reduced relative to the native population and the new population has retains a small number of individuals over multiple generations (Nei et al. 1975, Avise 1994, Baker and Moeed 1987), other researchers debate the efficacy of these "bottlenecks" on the reduction of genetic diversity in all introductions. For example, Barrett and Richardson (1986) state, "[L]ong distance migration involving relatively few immigrants may have little effect on the amount of genetic variation if rapid population increase follows the initial establishment".

If bottlenecks reduce genetic diversity in some but not all populations, how important is genetic diversity to invasion success? Early work by Ehrlich (1986, 1989) suggested that potential successful invaders share certain characteristic attributes, one of which is a relatively large amount of genetic diversity (the reasoning being that large diversity measurements may indicate the presence of a pool of variation upon which natural selection can act, thus resulting in local adaptation). Barrett and Richardson (1986) disagree with Ehrlich, stating that there is no relationship between colonization efficacy and genetic variability. Duda (1989) goes on to suggest that the relationship between colonizers and genetic diversity may be a function of reproductive strategy and amount of genetic diversity of the source population.

Multiple studies have examined the genetic diversity among introduced marine populations using protein electrophoresis. Duda (1994) surveyed the Asian clam, *Potamocorbula amurensis*, an invasive species found within the San Francisco estuary.

The *P. amurensis* population in San Francisco Bay was first discovered in 1986; thus, the introduction is believed to be recent. Analysis of six allozyme loci revealed a large amount of genetic diversity (mean heterozygosity = 0.295, mean polymorphism = 0.75, average of 3.7 alleles per locus) is present in introduced *P. amurensis*. Genetic distances and F-statistics showed the population to be genetically homogeneous within the bay; although a contingency chi-squared test suggested the presence of population heterogeneity (Duda, 1994). Large genetic diversity in the introduced population can only result from the inoculation of genetically diverse larvae into the San Francisco Bay. As no study of the Asian source population was performed, Duda (1994) was unable to reject any of the several possible reasons for the presence of such genetically diverse larvae.

Levels of genetic diversity among introduced mollusc populations appears to be species specific (Duda 1989, see Table 1). These different levels of post-introduction diversity could have been caused by a variety of reasons. Causal factors could include that the colony population resulted from multiple introductions from different source populations each containing a low amount of diversity within the population, a single introduction from a genetically diverse population, or a single introduction from many genetically diverse populations.

While research examining the population genetics of introduced species in their colonial habitats has been performed, Barrett and Richardson (1986) note an overall "lack of information on the early stages of colonization," as well as a need for genetic data from the native and introduced populations. Such studies appear to be rare, especially for marine species, though work has been done by Marsden et al. (1996) and Reece et al.

Table 1: Allozyme studies reveal differing amounts of genetic diversity among introduced populations of molluscs (after Duda 1989).

Species Name	Study	Diversity (Heterozygosity)
Biomphalaria straminea	Woodruff et al. 1985	High
Crepidula onyx	Woodruff et al. 1986	High
Dreissena polymopha	Herbert et al. 1989	High
Macoma balthica	Meehan et al. 1989	High
Mytilus galloprovincialis	Grant and Cherry 1985	High
Potamocorbula amurensis	Duda 1989	High
Achatina fulica	Selander and Ochman 1983	Low
Cepea nemoralis	Selander and Ochman 1983	Low
Corbicula fluminea	Smith et al. 1979, McLeod 1986	Low
Littorina saxatilis	Knight et al. 1987	Low
Theba pisana	Johnson 1988	Low

(2001). Marsden et al. (1996) used allozyme analysis to survey populations of zebra

mussels in native European and introduced (North American) populations. High levels of

genetic diversity (average heterozygosity ranged from 27-43%) existed in both native and

colonized regions, suggesting either a large pool of individuals or multiple introductions.

No change in the amount of genetic variation in North American populations was

observed in the nine years since the original survey (Marsden et al. 1996). Reece et al.

(2001) surveyed the population genetics of the parasite *Perkinsus marinus* after its spread

into *Crassostrea virginica* populations along the northeastern American coast. Restriction

fragment length polymorphism (RFLP) analysis was used to survey eight loci from *P.*

marinus samples from the northeast, southeast, and gulf coast regions. Parasites were

shown to have "significant structure in the geographic distribution of genetic strains", and

the southeast coastal area contained more genetic variation than the Gulf of Mexico or

northeast regions. The recent introduction in the northeast region possesses less variation

than its probable source population.

Molecular techniques

"Invasions are fast, evolution is slow" (Williams 1996).

One must survey relatively rapidly evolving molecular markers to examine

genetic variation among introduced populations as invasions occur during an eyeblink of

evolutionary time. Instead of observing genetic divergence between a newly established

introduction and its parent population, one would rather expect to see shifts in allele

frequencies among locations. DNA sequencing, amplified fragment length

polymorphism (AFLP) fingerprinting, microsatellite loci, exon-primed, intron-crossing (EPIC) PCR, restriction fragment length polymorphism (RFLP) analysis, and even allozymes have all been used as molecular markers to investigate genetic diversity in introduced populations. Each of these techniques provides different levels of genetic resolution and different types of information; it is important to match the proper molecular techniques to the invasion situation at hand (Hillis et al. 1994). Several factors, including the level of genetic diversity revealed by the marker, how much of the genome one wishes to survey, which genome one will examine, and effort and cost per sample, should be considered.

Allozyme studies examine differences in protein mobility through a starch gel. If there has been a substitution of one or more amino acids within the protein, this charge or size difference can result in differences in protein mobilities through the gel. Allozyme studies survey multiple loci located throughout the nuclear genome. Allozyme electrophoresis is a powerful tool for examining population structure, mating systems, and heterozygosity. The technique is relatively non-expensive and can be used to screen large numbers of individuals. However, only phenotypic, rather than genotypic, information is collected, and the technique requires fresh or freshly frozen, non-degraded tissue (Hillis et al. 1994).

Amplified fragment length polymorphism (AFLP) fingerprinting allows for a broad scale screening of the nuclear genome. Large numbers of anonymous co-dominant markers are surveyed. This technique has proven useful for examining relationships between closely related organisms, including parentage studies, population genetics, and distance based phylogeny reconstruction.

Variable number tandem repeat (VNTR) loci are biparentally inherited nuclear markers. VNTR loci that are composed of 2-5 base pair repeated regions are known as microsatellites. Microsatellite loci have been very useful in population genetics work because they are highly variable regions (mutation rates of 10^{-2}/gamete/generation, Hillis et al. 1994). While microsatellite repeat regions would represent an ideal system for assessing genetic variation among the geographically disjunct populations of *R. venosa*, no *Rapana* specific loci have been developed to date. For the scope of this Masters' project, the cost and development time for generating a microsatellite library was prohibitive.

Exon-primed, intron-crossing (EPIC) Polymerase Chain Reaction (PCR) allows for the amplification of intron regions in a specific DNA locus. The PCR primers are anchored in highly conserved exons and are designed to amplify the intron region between these conserved regions. The nontranslated intron regions evolve more rapidly than the protein coding exon regions. Accumulated insertion/deletion events and base pair substitutions in these intron regions have proven useful for looking at population genetics and invasion studies. The markers generated by EPIC-PCR can be analyzed through either DNA sequencing of the amplified region or through restriction fragment length polymorphism (RFLP) analysis.

Restriction fragment length polymorphism (RFLP) analysis examines the presence or absence of restriction enzyme recognition sites in a certain DNA locus. In the past, whole molecule mitochondrial DNA was analyzed, but it is now common to amplify a specific DNA region using PCR. The amplified product is then digested with

restriction endonucleases. Base substitutions as well as insertion or deletion events can result in the loss or gain of a restriction site or in fragment size variations.

In this study, DNA sequencing was chosen to assess variation among native and introduced *R. venosa* populations because by detecting individual nucleotide differences in the locus sequenced it provides the highest level of resolution of the molecular techniques. The discovery of the maximum amount of variation possible is important, because one must look at differences in native and introduced locations based on the partitioning of allele frequencies. As the time since establishment is very short – with a maximum of 60 years/30 generations for the Black Sea invasion, and 15 years/7 generations for the Chesapeake Bay invasion (and that only 1-2 generations are available for genetic screening) -- one does not to expect to find sequence divergence due to mutation events between the introduced and native samples. Rather, DNA sequencing, through revealing a large number of allelic variants at the nucleotide level for individuals within a population, could allow one to discriminate populations based on shifts in allele frequencies among groups of individuals at different sampling locations.

Loci selected for sequencing

Two loci, the mitochondrial cytochrome *b* gene region and the nuclear internal transcribed spacer (ITS) region, were selected for sequencing. These loci were chosen as they both have yielded levels of variability appropriate for intraspecific studies in the past, and because they posses patterns of inheritance that reveal different levels of genetic information.

Bottlenecks, which occur when colonial populations are severely reduced relative to the founding populations and the numbers stay small over multiple generations, can result in a severe reduction in genetic diversity (Nei et al. 1975, Avise 1994, Baker and Moeed 1987). Avise, paraphrasing Wilson (1985), suggested mtDNA haplotypes might register founder effects more clearly than autosomal nuclear loci because of their expected fourfold lower effective population size. Nevertheless, issues such as kin-structured colonization may also significantly shape genetic variation in an introduced bottlenecked population. Wade et al. (1994) point out that in extreme bottleneck conditions: "colonization by a single randomly chosen but multiply inseminated female or by a group of related females (e.g., sisters) will preserve most of the nuclear variation but little of the cytoplasmic variation". Thus, nuclear DNA may reveal more allelic variation than the (usually) more rapidly mutating mt DNA. Thus, as different genetic information can be provided through the use of nuclear and mtDNA markers, both were incorporated into this project.

Cytochrome *b*, a protein coding mitochondrial gene, is part of the electron transport chain. It possesses both conserved and variable regions, and has been found to have sufficient variability for population analyses in both invertebrates (harpacticoid copepod *Microarthridion littorale*, Schizas et al. 1999) and some molluscan taxa (the protobranch bivalve *Deminucula proxima* and long finned squid *Loligo pealei*, Merritt et al. 1998). Mitochondrial genes are useful in population genetic studies as they have a smaller population size compared to nuclear genes. This reduction in relative population size results as mitochondrial DNA is inherited only through the maternal line, and each individual typically receives only one copy of a mitochondrial gene. Thus, mother and

offspring share identical haplotypes, allowing one to trace clonal lineages across time and space.

The nuclear locus ITS is part of the multi-copy ribosomal RNA gene cluster in eukaryotic organisms (Palumbi 1996). The ITS region amplified in this study is located between the nuclear large and small subunit rRNA genes; ITS possesses three sub-regions, one of which codes for the 5.8S structural rRNA gene, and two regions (ITS-1 and ITS-2) for which the function is not known. ITS-1 and ITS-2 are transcribed from DNA to RNA but are not functional rRNA products, nor is this RNA translated into a protein. Variation in ITS accumulates in these supposedly non-functional regions as they are relatively less constrained by selection, making ITS useful for intraspecific and interspecific genetic studies. Biparentally inherited nuclear gene regions, such as ITS, can reveal information about kin structured colonization events that would be lost if only mtDNA was studied. ITS has proven useful in analysis of population genetics in molluscs and other invertebrates. Schizas et al. (1999) surveyed ITS (along with cytochrome *b*) to examine population structure in a harpacticoid copepod, and Caporale et al. (1997) used a portion of this region to examine population structure in the soft-shelled clam (*Mya arenaria*) in the Northeastern United States. ITS has also proven useful in determining the taxonomic relationships between putative species of molluscs, including stagnicoline pulmonate snails (Remigio and Blair 1997), bivalves (Anderson and Adlard 1994; Kenchington et al. 2000), and the gastropod genus *Bulinus* (Stothard et al. 1996).

OBJECTIVES

The overall objective of this study is to investigate the genetic relationships of the

Chesapeake Bay population of *R. venosa* relative to other introduced and native

populations of and to evaluate the taxonomic relationship of *R. venosa* to *R. bezoar*. This

was accomplished through the study of the cytochrome *b* and internal transcribed spacer

(ITS) region of the ribosomal RNA gene complex. This work was designed to contribute

to the field of invasive species research, and it attempted to address the question of the

invasion dynamics of a marine introduction by examining the genetic relationships

between native and introduced populations. Potential source populations for the recent

Chesapeake Bay invasion were examined, and the genetic data analyzed to determine

whether the Chesapeake Bay *R. venosa* population can be described as the result of

multiple invasion events from different source populations, or as a single flourishing

introduction.

The examination of *R. venosa* and *R. bezoar* provides insight into alpha taxonomy

of the genus *Rapana*. Ultimately, this issue is important as it allows for an accurate

understanding of the physical tolerances for *R. venosa*. *R. venosa* is considered to be a

temperate species, while *R. bezoar* a more tropical species. Thus, if the two appear as

separate species both phenotypically and genotypically, we can be reasonably certain that

R. venosa has been correctly classified as a temperate species. Resolution of this issue

allows for proper risk assessment (e.g. how far south along the US east coast can *R.*

venosa invade?) and the development of effective management strategies.

MATERIALS AND METHODS

Sampling and DNA isolation

Rapana venosa were collected from three areas of the world: Korea, Turkey, and the United States (Figure 2). An additional single individual was obtained from the Atlantic coast of France. Samples are described by location and identification number (Tables 2 and 3).

Individuals from native Asian waters were purchased as live specimens from three city fish markets (Inch'on, Tongyeng, and Cheju-do) located on the South Korean peninsula. The *R. venosa* purchased at market were obtained from a commercial whelk pot fishery. Rapa whelk pots were typically located approximately 20 meters offshore on sandy muddy bottom habitat. The fish markets, while geographically separate (see Figure 3), each obtained rapa whelks from sources along southwestern coastline of Korea. Thus, while the fish markets are distant from one another, they do not represent localized collections. For this reason, all Korean fish market samples were pooled into a single Korean sample. Collections of Korean rapa whelks took place in summer 1999 by R. Green (VIMS) and Drs. Albert Choi (Cheju National University) and Jong-Geel Je (Korean Ocean Research and Development Institute).

Rapa whelks from the Black Sea were collected in Turkey at Sürmene-Çamburnu, a marine laboratory of Karadeniz Technical University located along the southeastern coast of the Black Sea, near Trabzon, Turkey (see Figure 4). *Rapana venosa* were

Figure 2: World map marking the three locations where samples of *Rapana venosa* were collected for population genetic analysis (the site of the single individual collected from South Brittany, France is not shown). The native range of *R. venosa* includes the waters of China, Japan, and Korea. Collection of native rapa whelks occurred in southwestern Korean waters (location A). Collection of non-native rapa whelks occurred along the southeastern coast of the Black Sea near Trabzon, Turkey (location B), and in the waters of the Chesapeake Bay, USA near Hampton Roads, VA (location C).

Physical Map of the World, April 2000
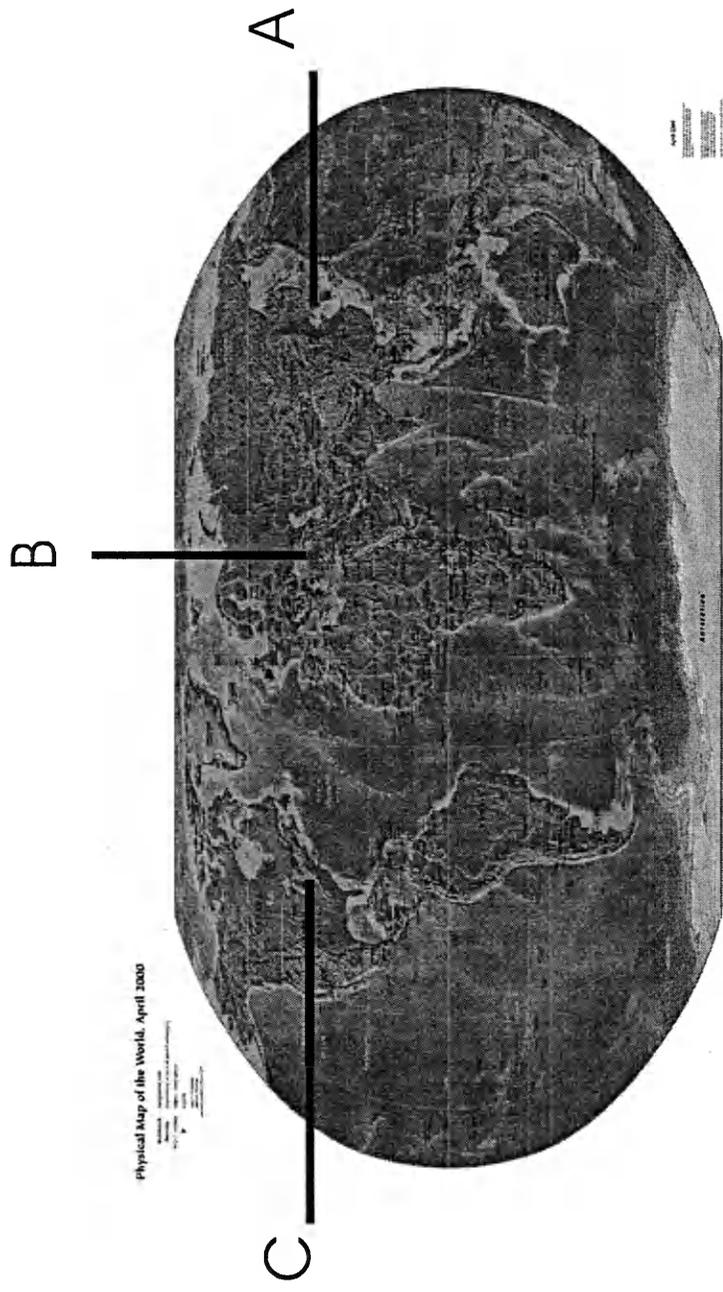
TABLE 2: Individual identification numbers, sample locations, and number of clones per individual sequenced for the cytochrome *b* analysis of *Rapana venosa* population structure. The following abbreviations are used: FM = Fish Market, HRBT-MMBT = Hampton Roads Bridge Tunnel to Monitor Merrimack Bridge Tunnel, and James River Ship = James River Shipyard.

Location	Sample ID	# Clones Sequenced	Location Notes
Korea	K108	3	FM: Inch'on
	K118	3	FM: Inch'on
	K250	3	FM: Tongyeng
	K255	1	FM: Tongyeng
	K260	3	FM: Tongyeng
	K261	1	FM: Tongyeng
	K263	2	FM: Tongyeng
	K277	1	FM: Tongyeng
Black Sea (Turkey)	T42	1	Sürmene-Çamburnu
	T47	3	Sürmene-Çamburnu
	T48	3	Sürmene-Çamburnu
	T49	3	Sürmene-Çamburnu
	T50	3	Sürmene-Çamburnu
	T53	1	Sürmene-Çamburnu
	T54	1	Sürmene-Çamburnu
	T56	1	Sürmene-Çamburnu
Chesapeake Bay (USA)	C41	3	Unknown
	C94	1	Ocean View
	C95	1	Ocean View
	C149	1	HRBT-MMBT
	C158	1	James River Ship
	C174	1	James River Ship
	C199	2	Ocean View
	C210	2	Ocean View
France	FRII	2	Bay of Quiberon

TABLE 3: Individual identification numbers, sample locations, and number of clones per individual sequenced for the ITS analysis of *Rapana venosa* population structure, and for the analysis of the taxonomic relationship between *R. venosa* and *R. bezoar*. The following abbreviations are used: FM = Fish Market, HRBT-MMBT = Hampton Roads Bridge Tunnel to Monitor Merrimack Bridge Tunnel, and James River Ship = James River Shipyard.

Location	Sample ID	# Clones sequenced	Location Notes
Korea	K1	1	FM: Cheju-do
	K65	2	FM: Inch'on
	K77	1	FM: Inch'on
	K91	1	FM: Inch'on
	K103	1	FM: Inch'on
	K108	1	FM: Inch'on
	K260	1	FM: Tongyeng
	K261	1	FM: Tongyeng
	K263	1	FM: Tongyeng
	K273	1	FM: Tongyeng
	K277	1	FM: Tongyeng
Black Sea (Turkey)	T42	1	Sürmene-Çamburnu
	T45	1	Sürmene-Çamburnu
	T46	1	Sürmene-Çamburnu
	T47	1	Sürmene-Çamburnu
	T48	1	Sürmene-Çamburnu
	K49	2	Sürmene-Çamburnu
	T50	1	Sürmene-Çamburnu
	T51	1	Sürmene-Çamburnu
	T53	1	Sürmene-Çamburnu
	T56	1	Sürmene-Çamburnu
Chesapeake Bay (USA)	C37	3	James River Ship
	C94	2	Ocean View
	C95	1	Ocean View
	C96	2	Ocean View
	C109	3	Ocean View
	C127	3	James River Ship
	C158	3	James River Ship
	C174	2	James River Ship
	C199	3	Ocean View
	C210	3	Ocean View
France	FRII	2	Bay of Quiberon

Figure 3: Map of South Korea showing locations of the three fish markets (A = Inch'on,

B = Tonyeng, and C =Cheju-do) where native *Rapana venosa* were obtained. While

individuals were obtained from all three fish markets, they were pooled into a single

Korean sample. *Rapana venosa* sold at these markets could have come from any location

along the South Korean coastline.

A

B

C

NORTH
KOREA

Demarcation
Line

Kangnŭng

⊗ SEOUL

Wŏnju

Ullŭng-do

Inch'ŏn

Pukp'yŏng-
dong

Suwŏn

Sea
of
Japan

Yellow
Sea

Taejŏn

P'ohang

Kunsan

Taegu

Chŏnju

Ulsan

Kwangju

Masan Chinhae

Mokp'o

Pusan

Yosu

Korea

Strait

Cheju-do

JAPAN
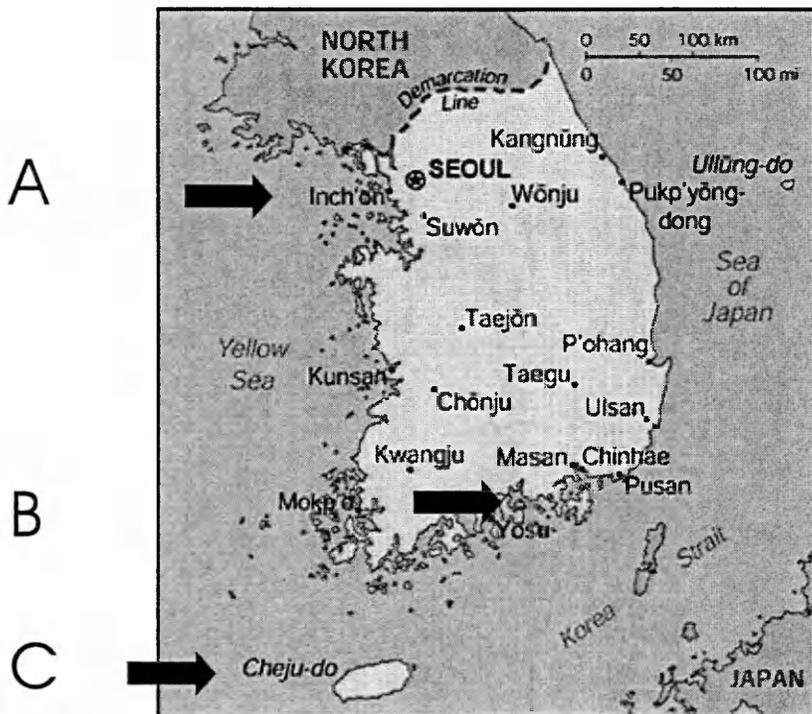
Figure 4: The black star shows the location of *Rapana venosa* collected from a small

embayment at Sürmene-Çamburnu, near the city of Trabzon, Turkey. Collection of

snails from the southeastern portion of the Black Sea occurred in October 1999.

collected using SCUBA in a small embayment on 28 October 1999. Rapa whelks were found on hard bottom at a depth of eight meters.

Samples of *R. venosa* from the Chesapeake Bay, USA were collected in 1999 and 2000 by the personnel of Virginia Institute of Marine Science and local fishermen. Chesapeake Bay *R. venosa* samples are part of a larger database (RAPTOR) generated by the Molluscan Ecology Working Group at VIMS, and can be cross referenced by ID number to RAPTOR samples (Tables 2 and 3).

The French *Rapana venosa* sample used in this study was obtained from the Bay of Quiberon, located in Southern Brittany, France. This individual was sent from Dr. Jean-Pierre Joly and Dr. P Goulletquer (IFREMER, Institut francais de recherche pour l'exploitation de la mer); no other information about this collection location is available.

Rapana bezoar individuals were obtained from Hong Kong. Twelve individuals were collected by Dr. Brian Morton (University of Hong Kong) from soft bottom with the Swire Insitute's offshore research trawl survey. Individual identification numbers for sampled *R. bezoar* are shown in Table 4.

The southern oyster drill, *Stramonita* (= *Thais*) *haemastoma*, was the outgroup species used in the phylogenetic comparison of *Rapana bezoar* and *Rapana venosa*. This thaidid mollusc is found in the Mediterranean, along the coast of west Africa, and, occurs as two 'subtypes' in the southeastern United States (Stickle, 1999). Kool (1993), who revised the phylogeny of the subfamily Rapininae (Neogastropoda: Murcidae) using morphological traits, places both *Rapana* and *Stramonita* in the subfamily Rapaninae (see Figure 5). A sample from the third member of the *Rapana* genus, *Rapana rapiformis*, was not obtained. Thus *Stramonita hemastoma* (the most closely related obtainable

TABLE 4: Sample location, individual identification code, and number of clones per individual for *Rapana bezoar* samples. These *R. bezoar* individuals were used in the analysis of the taxonomic relationship between *R. venosa* and *R. bezoar* using the ITS locus.

Location	Sample ID	# Clones Sequenced
Hong Kong	B1	1
	B3	1
	B4	1
	B5	1
	B7	1
	B9	1
	B10	1
	B11	1

Figure 5: Cladogram illustrating relationships within the subfamily Rapaninae. Note that the two genera used in this study, Stra (*Stramonita*) and Rap (*Rapana*), both marked with asterisks, are closely related (from Figure 30, Kool 1993).

Taxanomic groupings in the consensus cladogram are abbreviated as follows: Mur = *Muricanthus*; Hau = *Haustrum*; Nuc = *Nucella*; Tro = *Trochia*; For = *Forreria*; Aca = *Acanthina*; Cym = *Cymia*; Stra = *Stramonita*; Rap = *Rapana*; Con = *Concholepas*; Dic = *Dicathais*; Vex = *Vexilla*; Nas = *Nassa*; Pin = *Pinaxia*; Dru = *Drupa*; Plic = *Plicopurpura*; Mor = *Morula*; Cro = *Cronia*; Vas = *Vasula*; Tha = *Thais*; Pur = *Purpura*; Man = *Mancinella*; Neo = *Neorapana*; Trib = *Tribulus*.

PHYLOGENY OF RAPANINAE
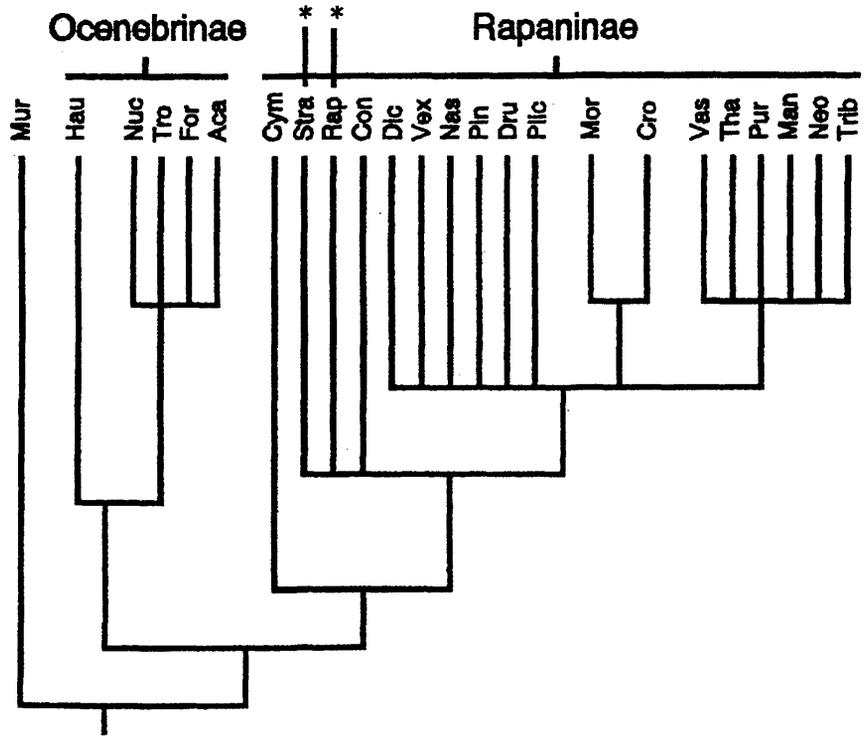
member of Rapininae) was used as the outgroup. The *S. haemastoma* was obtained from

Grande Isle, LA, by Dr. W. Stickle of Louisiana State University in June 2001.

For all DNA samples, approximately 1 cm³ of foot muscle was removed from a

living snail. All tissue samples were submerged in dimethylsulfoxide/disodium

ethylenediamine tetraacetate (DMSO/EDTA) preservation buffer, except for the French

R. venosa sample, which was sent in 95% ethanol. Samples were stored at room

temperature, and DNA was later extracted using the QIAmp DNeasy mini kit (Qiagen)

following the manufactures instructions. DNA isolates were resuspended in 400 µl of

AE buffer (Qiagen), and stored at 4 °C for up to nine months. After this time,

cytochrome *b* polymerase chain reactions (PCR) no longer produced amplified products.

PCR reactions

Working with cytochrome *b* was difficult. Spolsky et al. (1996) used cytochrome

b as a tool for population genetic studies in an Asian gastropod:

> Although cytochrome *b* sequencing has been used in numerous systematic
> studies . . . it was not a simple matter to apply these techniques to
> molluscs. Because of the very ancient branching of molluscs from the
> basal phylogeny, their cytochrome *b* sequences are divergent enough that
> the so-called "universal primers" for cytochrome *b* do not amplify this
> gene from molluscs . . . because of variable yield and purity of the product
> from different taxa, we cloned all amplified products prior to sequencing.

All of the problems listed above were encountered.

Cytochrome *b* primer sequences used in this project were from Collins et al.

(1996) who examined recent speciation events in *Nucella*, a genus of carnivorous

shallow-water gastropods. The sequences for the cytochrome *b* primers used were

aaaaagctttctaatctctcagtttgatgaa (145F) and

aaaaagctttgatcgaaaatagcataggcaa (155R).

The PCR protocol of Collins et al. (1996) was modified to amplify cytochrome *b* in *Rapana venosa*. The program used for the cytochrome *b* amplification reactions in the MJ Research Peltier Thermocyler (PTC-200) began with a denaturation at 94°C for 4 min, then 35 cycles of 94°C for 30 sec; 48°C, 1 min; 72°C, 3 minutes, and a final extension of 70°C for 5 minutes.

The internal transcribed spacer region primer sequences ITS-3 and ITS-5 were obtained from the literature (Goggin1994). The primer sequences used were originally designed to amplify ITS from the marine protist *Perkinsus*. The primers used to amplify ITS were as follows:

tatgcttaaattcagcgggt (ITS-3)

cgtaggtgaacctgcggaagg (ITS-5).

The program used to generate amplification products for ITS began with a denaturation at 94°C for 5 min, then 35 cycles of 94°C for 30 sec; 45°C, 1 min; 65°C, 3 minutes, and a final extension of 65°C for 7 minutes.

After a 'master mix' was created for a set of PCR reactions (Table 5), 18.8 μl was pipetted out into individual 0.2 ml strip tubes. DNA was then added. Reactions were very sensitive to DNA concentration, especially the cytochrome *b* amplifications. Often, test reactions had to be performed to determine the optimal concentration. Concentrations tested included 1:10, 1:20, and 1:50 DNA dillutions. Once the correct concentration for optimal amplification had been established, 0.25 μl of DNA from the 400 μl Qiagen isolate (or dilution) was added to an individual reaction.

Table 5: Concentrations and components for "cocktails" used for both ITS and cytochrome *b* Polymerase Chain Reaction (PCR). Components are from the GibcoBRL Life Technologies PCR Reagent System.

Ingredient	To make one 18.8 μl PCR cocktail...
PCR quality water	15.2 μl
10X PCR buffer + MgCl$_2$ (15 mM)	2.5 μl
Bovine Serum Albumin (1 mg/ml)	2.5 μl
dNTP (10mM)	0.5 μl
Forward Primer (100 pmol)	0.125 μl
Reverse Primer (100 pmol)	0.125 μl
Taq polymerase (5 U/μl)	0.125 μl

Because strong cytochrome *b* PCR amplification products were difficult to obtain initially, permutations of several parameters, including the concentrations of various components within the PCR reaction and the annealing temperature, were performed to note if the different parameters would result in increased reaction efficacy. Magnesium chloride concentrations were tested at 1.0, 1.5, and 2.5 mM concentrations. Bovine Serum Albumin (BSA) 1mg/ml was tested at 0, 6.25, 12.5, and 20 percent of the total reaction volume. Annealing temperatures for the cytochrome *b* reactions were tested using a temperature gradient thermocycler; stronger products resulted between 48° and 51° C. The lower temperature (48°) was used as the optimal annealing temperature in the remaining reactions

PCR reactions were tested for the presence of amplified product. Four microliters of the PCR amplification reaction were run out on a 1% agarose gel, and DNA was stained using ethidium bromide, and visualized under UV light. If product was present but lacking sufficient concentration for use in cloning, the remainder of the PCR reaction (approximately 10 µl) was run out on a 1% agarose gel and visualized using ethidium bromide. Then, the gel fragment containing the amplified product was excised. This gel fragment was dissolved in 200 µl PCR quality water and incubated at 65% for 15 minutes. One microliter of this mixture was used in an additional 20 µl PCR reaction. This usually resulted in a yield greater than the initial reaction.

Some cytochrome *b* amplifications required an additional step to clean the PCR product prior to its use in cloning and DNA sequencing. Reactions were purified using the GenElute PCR DNA Purification Kit (Sigma) when the amplified fragment was faint in comparison to bands of approximately 50-100 bp (the very small products may be

primer-dimer bands or excess deoxynucelotidetriphospates (dNTPs)). Column cleaning using the GenElute kit removed these low molecular weight fragments and resulted in successful ligation reactions. All protocols outlined in the GenElute kit were followed, except for the following modifications. First, to obtain DNA in concentrations sufficient for ligation, the volume of "unclean" PCR reaction run through the column was increased to at least 40 µl and the volume of water used during the elution was reduced to 20 µl. Second, PCR quality water was used instead of Tris/EDTA (TE) buffer in the final elution step. This eliminated the possibility that TE salts would interfere with later enzymatic reactions.

Cloning

Once a strong, clean amplified product was produced, it was cloned into a plasmid vector. Cloning was performed using the pCR2.1 vector, the Original TA Cloning Kit, and the protocols listed in the TA Cloning Kit manual, versions M - R (Invitrogen). Transformed bacterial colonies were screened for the presence of the PCR product. Colonies testing positive for the insert were grown up at 37 °C overnight (approximately 16 hours) in either sterile Terrific Broth (TB) or Luria-Bertani (LB) media with 6 µl of 50 mg/ml ampicillin (Sigma). Isolation of the transformed plasmid from the bacterial genomic DNA was performed using either the Perfectprep Plasmid Mini Kit (Eppendorf) or the QIAprep Spin Miniprep Kit (Qiagen). The resulting DNA, known as a mini-prep, was tested for the presence of the insert through a restriction digest with *Eco*RI (10.5 µl water, 1.5 µl React 3 buffer, 0.5 µl *Eco*RI enzyme, and 2.5 µl DNA) for three hours to overnight at 37 °C. Three microliters of stop solution were added to

each reaction, and fragments generated during the reaction were electrophoretically separated on a 2% agarose gel. Visualization of the DNA using UV light after staining with ethidium bromide revealed the presence of the proper sized fragment.

Sequencing

Mini-preps containing the vector and PCR insert of interest were quantified using a fluorometer prior to sequencing. Most sequencing reactions were performed with the Thermo Sequenase Primer Cycle Sequencing Kit (Amersham Pharmacia Biotech), although a small number of reactions were performed with the SequiTherm EXCEL II DNA Sequencing Kit-LC (Epicentre). Cytochrome *b* sequencing reactions were single direction and utilized either a 700 or 800 labeled T7 promoter primer. Sequencing of the ITS locus used the fluorescent dye-labeled primers M13 Forward and M13 Reverse. Bidirectional sequencing reactions incorporating the M13F 800 and M13R 700 primers were used for most ITS sequencing reactions.

Sequencing was performed on a LI-COR 4200L automated sequencer. DNA sequences were visualized by running 0.6 μl of each reaction on a 0.2 mm thick 66 cm long polyacryalmide gel (LI-COR). The polyacrylamide gel recipe used (40 ml of 3.7% KB Plus gel matrix (LI-COR), 267 μl 10% ammonium persulfate (Sigma), and 24 μl Temed (Sigma)). Electrophoresis conditions used were the defaults listed in the E-Seq v. 1.1 program for 0.2 mm/66 cm gels; the buffer used during electrophoresis was 0.8X TE KB Plus Buffer (LI-COR).

Sequences and SCF (v.2) files generated using E-Seq were imported into the computer program Sequencher4.1. Multiple alignments were performed using the

ClustalW algorithm (Thompson 1994) in MacVector 7.0 (Kaufman et al.1994) and

checked by eye. For the cytochrome *b* alignment, both the pairwise and multiple

alignment possessed an open gap penalty of seven and an extend gap penalty of one.

Delay divergence values were 40%, and transitions were unweighted. All values were

the same for the ITS alignment, except that the open gap penalty was five and the extend

gap penalty was two. These values were chosen as they provided the best combination of

the number of conserved residues and the lowest number of gaps inserted when a variety

of gap cost permutations were performed.

Nucleotide sequences obtained from these alignments were compared to

previously published sequences found in the National Center for Biotechnology

Information (NCBI) database. Comparisons were performed using the basic local

alignment search tool (BLAST) algorithm created by Altschul et al. (1990). These

comparisons confirmed the loci sequenced in this study were gastropod cytochrome *b* and

ITS.

Data Analysis – *R. venosa* population genetics

Phylogenetic and population genetic analyses were conducted using a variety of

genetic analysis software packages. Tajima's D and standard DNA polymorphism

statistics (number of polymorphic sites, number of haplotypes, haplotype diversity (*h*)

and mean nucleotide sequence diversity (π) were calculated using the computer program

DnaSequencePolymorphism (DnaSP) (Rozas and Rozas 1999). DnaSP excludes all

character sites containing missing data, ambiguities, and gaps from analysis. Translation

of the cytochrome *b* data set using the invertebrate mitochondrial DNA genetic code,

calculation of all pairwise percent sequence diversity and divergence values (and associated standard errors), and creation of the distance based neighbor joining tree were performed in MEGA2.1 (Kumar et al. 2001). Like DnaSP, all gaps were removed from analysis in MEGA2.1. The program Arlequin 2.000 (Schneider et al. 2000) was used to calculate ϕst, a measure of population subdivision looking at between-population nucleotide diversity, as well as to generate the minimum spanning network branch values. Heterogeneity of haplotype frequencies was tested using Monte Carlo analysis (Roff and Bentzen, 1989) within the Restriction Enzyme Analysis Package (REAP) (McElroy et al.1991).

Data Analysis – Differentiation of *Rapana venosa* and *R. bezoar*

To examine genetic differences between the putative species *R. venosa* and *R. bezoar*, internal transcribed spacer region (ITS) nucleotide sequences were obtained for eight *R. bezoar*. These sequences were obtained for comparison to ITS nucleotide sequences from each of the three (Korean, Turkish, and Chesapeake Bay) *R. venosa* locations. All PCR and sequencing reactions for ITS regions of *R. bezoar* were as described previously. From different sample locations, different numbers of clones per individual were sequenced. As ITS is multi-copy, several different allelic forms could be isolated from a single individual. Thus the term 'clone' as applied in this work does not necessarily denote identity among the isolates from a single individual, but instead refers to an individual allelic form amplified and isolated through the process of cloning. The mitochondrial cytochrome *b* locus described in the previous section was not used, as the

primer sequences did not amplify a product in any of the 12 Hong Kong *R. bezoar*, even after modifications of the reaction conditions and components.

For phylogenetic analysis of the ITS data set, pairwise distances (gaps excluded) were calculated using MEGA2.1 (Kumar et al. 2001). Creation of the maximum parsimony tree and bootstrap analysis was performed with the phylogenetic software package PAUP*4.0b8 (Swofford 2000). The tree was generated using a heuristic search procedure. The number of bootstrap replicates included was a thousand, and gaps were treated as informative characters. The species *Stromonita (=Thais) haemastoma* was used as an outgroup to root trees.

RESULTS

Genetic variation in *Rapana venosa*

A 731 bp segment of the cytochrome *b* gene was sequenced for twenty-five

Rapana venosa (aligned sequences are shown in Appendix A). The sequences comprised

twelve haplotypes (Table 6). Of the 731 total characters, 19 were polymorphic, and of

these, four were parsimony informative. There were 19 transitions, no transversions, and

one deletion. This deletion was disregarded in all analyses, as cytochrome *b* is a protein-

coding gene and a single deletion event would result in a frameshift mutation, disrupting

the protein. Thus, this event is likely a *Taq* error during PCR). The number of singleton

mutations found were higher than predicted under expectations of neutrality (Tajima's D

of -2.06955, $p < 0.05$).

Translation of the cytochrome *b* data set using the invertebrate mitochondrial

genetic code produces an amino acid sequence 243 residues in length. The protein

translation start site is at position two; the protein translation for all individuals is shown

in Appendix B. All 13 mutations found within the Korean sample were synonymous

(e.g. did not result in changes to the amino acid sequence). In contrast, only one

synonymous change (individual C174 at site 186) is found among the four Chesapeake

Bay variants, while the three remaining mutations found in the sample resulted in

nonsynonymous substitutions. In addition, the nonsynonymous changes resulted in

nonconservative amino acid substitutions.

Table 6: Aligned haplotypes and polymorphic sites for the *Rapana venosa* 731 bp cytochrome *b* data set. Polymorphism positions are given at the top of the figure. The most common haplotype (shared by 14 individuals from the Chesapeake Bay, Turkish, and French collection sites) is listed as haplotype number one; this sequence is shown in bold. Dots show identity to haplotype one, a dash represents a single base pair deletion. Haplotypes 2-13 are all unique haplotypes. Base pairs marked with an asterisk indicate the base pair change has resulted in a nonsynonymous substitution in the translated amino acid sequence.

ID	No.	7	79	112	211	262	313	326	334	367	370	403	439	523	537	586	590	685	696	700	712
	1.	**A**	**T**	**T**	**G**	**T**	**G**	**T**	**A**	**A**	**T**	**A**	**C**	**C**	**T**	**A**	**T**	**T**	**A**	**C**	**A**
K108	2.	.	C	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
K118	3.	.	C	.	A	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
K250	4.	.	C	.	A	.	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.
K255	5.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
K260	6.	G	C	C	.	.	.	.	.	G	.	.	.	.	.	.	.	C	.	.	G
K261	7.	.	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
K263	8.	.	C	.	A	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.
K277	9.	G	C	.	M	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	N
C94	10.	.	.	.	.	.	.	.	.	.	.	.	.	.	C*	.	.	.	.	.	.
C95	11.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	–	.
C174	12.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G*	.	.
C199	13.	.	.	.	.	.	.	C*	.	.	.	.	.	.	.	.	.	.	.	.	.

Diversity within R. venosa collections

Analysis of the level of DNA polymorphism within the three geographic samples revealed that the sampling sites possess different levels of diversity. The Korean location (possessing eight haplotypes and thirteen polymorphic sites) was more diverse than the Chesapeake Bay sample (consisting of four haplotypes and four polymorphic sites), and both locations were more diverse than the Turkish collection, which shared a single haplotype with the Chesapeake Bay sample and had no polymorphic sites (Table 7). The single sample from France possessed a haplotype identical to the most common haplotype found in Chesapeake Bay and Turkey.

Haplotype diversity (h), the probability of sampling different haplotypes when two individuals are drawn from a population, illustrates differences in diversities among collection sites. Haplotype diversity values ranged from 1.0 in the Korean location (all eight individuals exhibited unique haplotypes), to 0.0643 in the Chesapeake Bay location (four haplotypes present in eight individuals), to zero in the Turkish site (all eight samples were monotypic). Mean nucleotide sequence diversity (π), the average number of differences per site between sequences, also reflected that the different sampling locations had different levels of diversity. Mean nucleotide sequence diversity values decreased from 0.495% (SD = 0.09%) in Korea, to 0.138% (SD = 0.05%) in Chesapeake Bay, to zero in Turkey. Construction of 95% confidence intervals (p=0.05), showed that both haplotype diversity (h) and mean nucleotide sequence diversity (π) values for the three sampling locations were significantly different from one another.

Table 7: Diversity statistics for the twenty-five *Rapana venosa* individuals sequenced at the 731 bp cytochrome *b* locus. Number of samples sequenced per location (N), number of polymorphic sites, number of haplotypes observed, haplotype diversity (*h*), mean nucleotide sequence diversity (π) , and average number of nucleotide differences between different haplotypes (k) are shown for each sampling location. The abbreviation nc stands for 'not calculated'. Numbers in parentheses represent standard deviation values for *h* and π.

Sample Location	N	No. Polymorphic Sites	No. Haplotypes	*h*	π	k
Korea	8	13	8	1 (0.063)	0.00495 (0.00090)	3.6071
Turkey	8	0	1	0 (0)	0 (0)	0
Chesapeake Bay	8	4	4	0.643 (0.184)	0.00138 (0.00054)	1.000
France	1	0	1	nc	nc	nc

Divergence between sample locations

Genetic distances were calculated as group means between sampling locations, and are presented as mean nucleotide sequence divergence in Table 8. Due to the differences in within-sample diversity among samples, both mean nucleotide sequence divergence and net corrected percent differences were calculated. Corrected comparisons showed no pairwise distance differences between samples from Chesapeake Bay, Turkey, and France; however, samples from these three locations were all different from the Korean sample.

Population structure statistics

Population structure was investigated by pairwise ϕst values. For the pooled sample, a value of 0.386 was calculated, with 38.63% of the variance resulting from differences among samples and 61.27% of the variation resulting from within-population variation. Between-population comparisons (shown in Table 9) show that the only significant non-zero ϕst values were between Korea and Chesapeake Bay (ϕst = 0.47), and between Korea and Turkey (ϕst = 0.42). P values for these ϕst values remained significant after using the Bonnferroni method to correct for multiple testing (Rice 1989). ϕst values can range between zero and one. High ϕst values indicate strong population structuring, while values close to zero indicate high gene flow within the species.

Homogeneity of allele frequency distributions was assessed through Monte Carlo randomization and chi-square analysis (Roff and Bentzen 1989). Haplotype distributions pooled from the Korean, Turkish, and Chesapeake Bay samples were subjected to 10000 Monte Carlo randomizations and subsequent chi-squared analysis showed no randomized

Table 8: Mean nucleotide sequence divergence between samples for cytochrome b. Net between group means (corrected) are below the diagonal. Between group means (uncorrected) are above the diagonal. Standard errors are shown in parentheses. All distances calculated using uncorrected "p".

	Korea	Turkey	Chesapeake Bay
Korea	--	0.384 (0.00156)	0.454 (0.158)
Turkey	0.150 (0.141)	--	0.070 (0.035)
Chesapeake Bay	0.150 (0.141)	0 (0)	--

Table 9: ϕst values, a measure of population subdivision, were generated by the computer program Arlequin v. 2.000. Significant values are indicated by **; these values are still significant after using the Bonferroni method to control for multiple tests ($p \leq 0.000$).

	Korea	Turkey	Chesapeake Bay	Total Sample
Korea	0.00000			
Turkey	0.47465**	0.00000		
Chesapeake Bay	0.42041**	0.00000	0.00000	
Total sample				0.38631**

chi-squared values exceeded or equaled the observed value. This indicates that these samples do not share a common gene pool. In contrast, a second test looking only at homogeneity between the Turkey and Chesapeake Bay samples yielded 1968 of 10000 replicates where chi-squared values exceeded or equaled the observed value, indicating significant heterogeneity between the Turkish and Chesapeake Bay samples (p=0.197).

Minimum spanning haplotype network

Figure 6 shows a minimum spanning network that reveals the nucleotide changes among the different cytochrome *b* haplotypes. The haplotype network illustrates the diversity of the Korean haplotypes relative to that of the Chesapeake Bay based on the number of nucleotide changes separating haplotypes. This minimum spanning network reveals partitioning of haplotypes between the native Korean haplotypes and the French, Turkish, and Chesapeake Bay haplotypes.

Neighbor joining tree

The neighbor joining tree, an unrooted phylogram constructed using genetic distance data (Figure 7), shows the distances separating all haplotypes to be very small. Korean haplotypes are separated from those of the Turkish and Chesapeake Bay samples. The Korean haplotypes branch out over a greater area, indicating that the Korean haplotypes are more diverse than those found in the other areas. All Turkish individuals, the single French individual, and five Chesapeake Bay samples shared a common haplotype. The three other haplotypes in Chesapeake Bay are closely linked to this haplotype.

Figure 6: Minimum spanning haplotype network illustrating relationships among

cytochrome *b* haplotypes for 25 *Rapana venosa* individuals. Hatch marks identify

individual nucleotide substitutions between neighboring haplotypes. Text inside circles

shows the identification code for the individual possessing that haplotype. The large

circle represents the most common haplotype (shared by a total of fourteen individuals

from Turkey, Chesapeake Bay, and France). The small circles represent a single

haplotype that is possessed by only one individual. The identification codes are color

coded to indicate sample location.
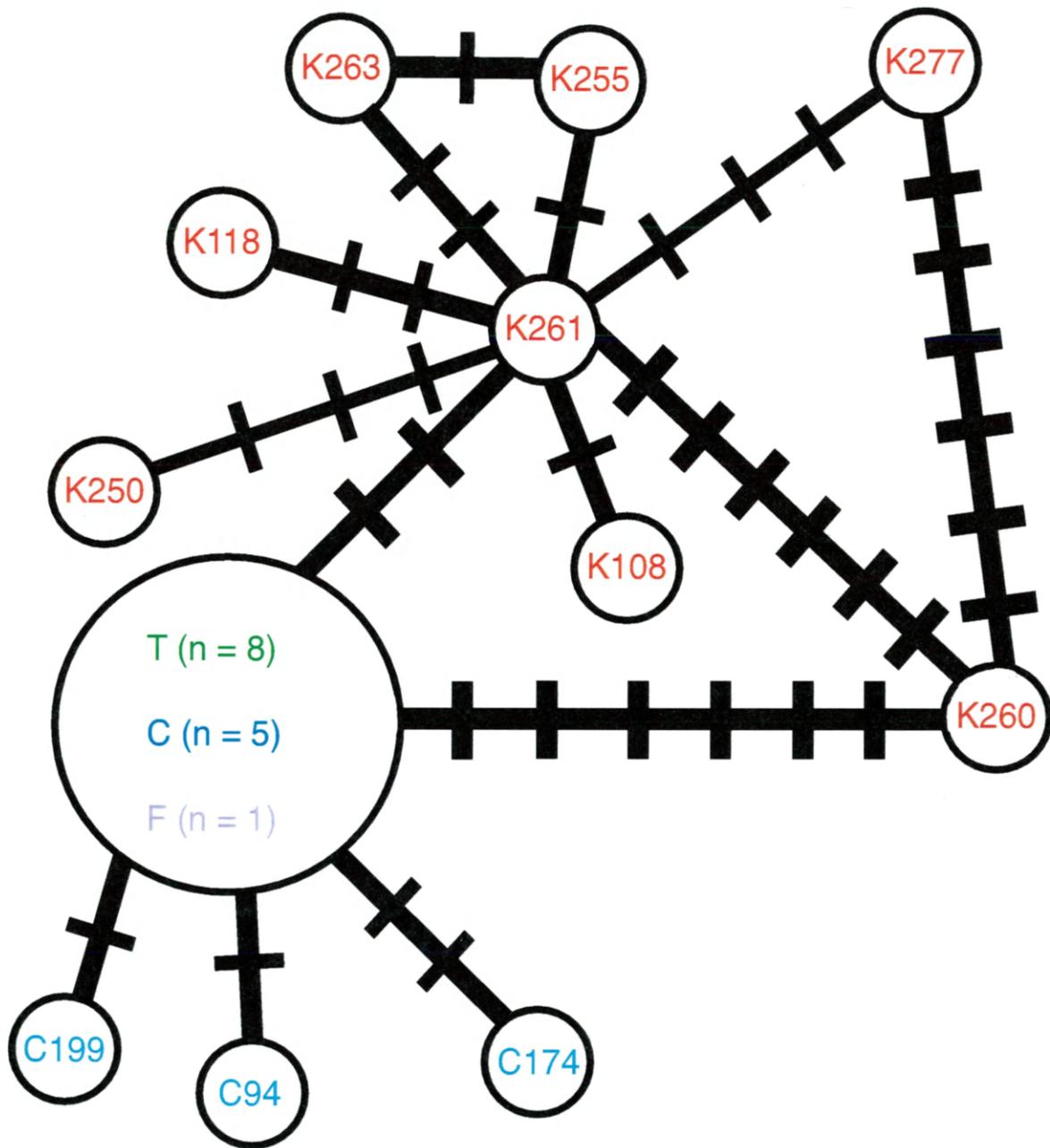
Figure 7: Neighbor-joining tree illustrating distance-based relationships among cytochrome *b* haplotypes for 25 *Rapana venosa* individuals. Korean individuals are shown in red, Chesapeake Bay individuals in blue, and Turkish individuals in green. The single French individual is shown in lilac.
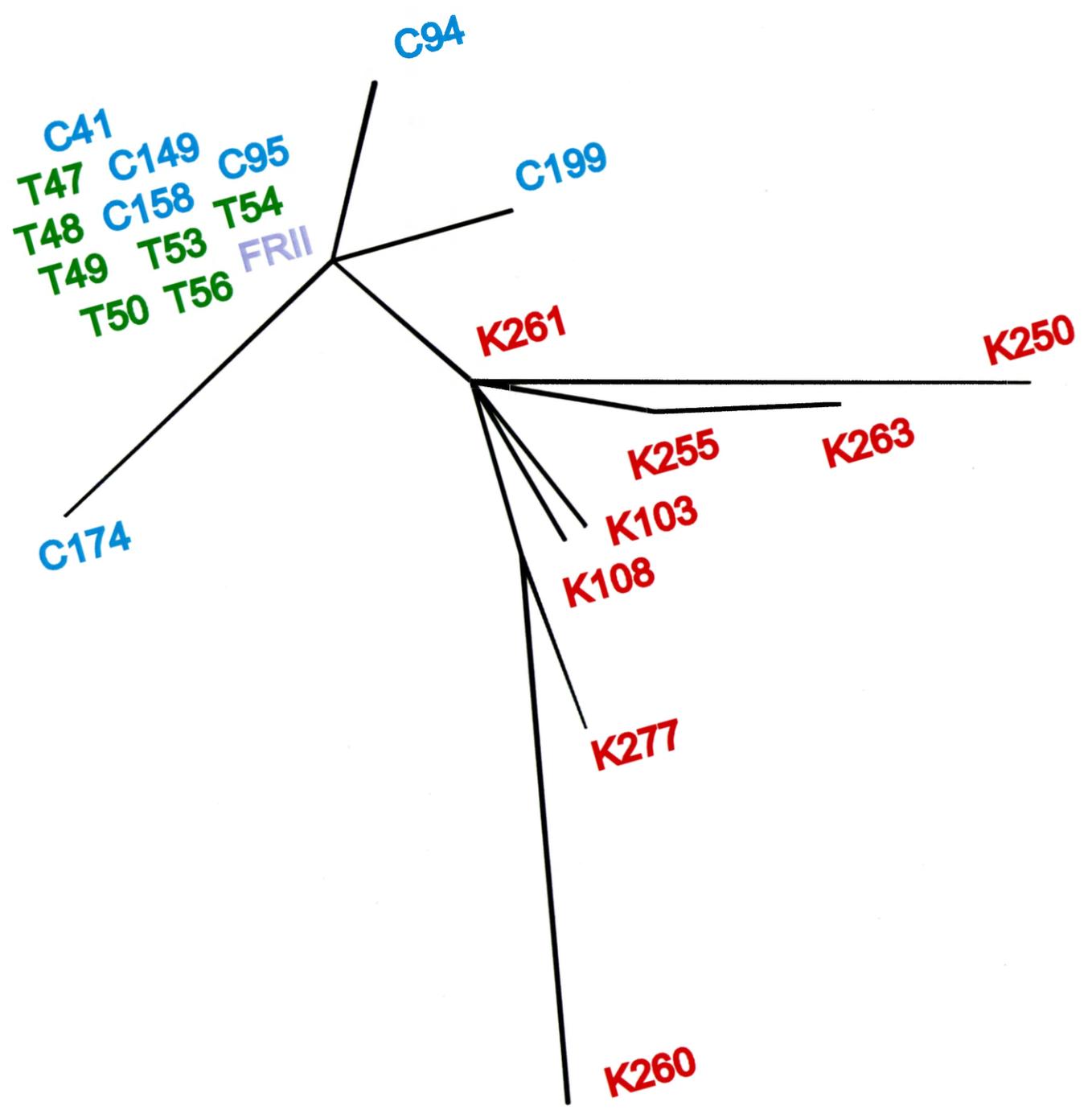
Genetic divergence between R. venosa and R. bezoar

The ITS region of the rRNA gene complex was sequenced for eight putative *R. bezoar* from Hong Kong and thirty-one *R. venosa* from Korea, Turkey, Chesapeake Bay, and France. Sequence information was also obtained for three clones of one individual of the outgroup species *Stromonita haemastoma*. The segment analyzed spanned the ITS-1, 5.8S, and ITS-2 regions. The size of the ITS fragment sequenced varied considerably between the three putative species due to insertion or deletion events within the nonstructural portion of the sequence. ITS fragments ranged in length from approximately 1020 bp in *R. venosa*, to 1260 bp in *R. bezoar*, and 1190 bp in *S. haemastoma*. Size differences in the ITS fragments were noted among individuals within species, and, in some cases, separate clones within individuals. Aligned ITS sequences are presented in Appendix C.

Between-species sequence divergence

Mean nucleotide sequence diversity within putative species and mean net nucleotide sequence divergence between putative species is displayed in Table 10. *R. venosa* samples showed relatively less within-species diversity (0.17%) compared to *R. bezoar* (0.42%) and *S. haemastoma* (0.24%). Within-species diversity values (0.17-0.42%) were an order of magnitude smaller than between species divergence values (6.1-14.2%). Corrected and uncorrected percent sequence divergence values reveal that there was significant genetic divergence between all comparisons of the three groups. The distance between *R. venosa* and *R. bezoar* (6.1%) was about half that for either of the

pairwise comparisons between the outgroup, *S. haemastoma*, and the putative congeneric

species *R. venosa* (13.9%) and *R. bezoar* (14.1%)

Maximum parsimony tree illustrating relationships among species

The phylogenetic relationship between *R. venosa* and *R. bezoar* was inferred

using a tree generated using maximum parsimony (see Figure 8). The tree, rooted using

sequences from the outgroup species *S. haemastoma*, revealed that individual ITS

sequences from *R. venosa* and *R. bezoar* segregate into separate clades. Bootstrap values

of 100% indicated strong statistical support for the separation of *R. venosa* from *R.*

bezoar, and separation of the *R. venosa* and *R. bezoar* clades from the outgroup.

Table 10: Matrix of percent sequence diversity and divergence within and between

putative species for the ITS gene region. Net between group means (corrected) are

located below the diagonal. Between group means (uncorrected) are above the diagonal.

Within location diversity is shown in bold along the diagonal. Standard error shown in

parenthesis. All distances calculated using uncorrected "p".

	R. venosa	*R. bezoar*	*S. haemastoma*
R. venosa	**0.168** (0.048)	6.420 (0.769)	14.096 (1.143)
R. bezoar	6.128 (0.729)	**0.417** (0.113)	14.480 (1.160)
S. haemastoma	13.890 (1.091)	14.150 (1.098)	**0.243** (0.137)

Figure 8: Phylogenetic tree generated using maximum parsimony showing the relationship among different ITS alleles. *Rapana venosa* are shown in purple (different sampling locations are shown in different shades of violet), *R. bezoar* are shown in pink, and the outgroup, *Stramonita haemastoma*, is shown in green. Identification numbers at the ends of the branches represent individual snails; numbers in parentheses following this ID code identify the individual clone. Individuals without a parenthetical identifier represent instances where only one clone was sequenced.

Numbers located on branches before nodes of this maximum parsimony tree are measures of percent bootstrap support. Bootstrap support is a statistical measure of the robustness of the generated tree. The tree was generated using a fast heuristic search; the number of bootstrap replicates used was 100, and gaps were classified as a fifth base.

T48(27)
T56
T53(17)
T53
T51
T49(5)
T50
T49(4)
T47(4)
T46(3)
T46
T45
T42
K277(26)
K273(28)
K263(18)
K261(7)
K260(H)
K108(17)
K103(11)
K091(7)
K077(23)
K065(7)
K065(3)
K01
FRII
C210(Q)
C210(O)
C210(M)
C199(V)
C199(S)
C174(W)
C158(F)
C158(E)
C158(C)
C127(E)
C127(D)
C127(C)
C109(S)
C109(R)
C109(O)
C096(L)
C096(K)
C096(H)
C095(21)
C095
C094
C037(B)
C037
C174(U)
B11
B10
B09
B05
B03
B01
B07
B04
Th1(1)
Th1(3)
Th1(2)

100
100
100
92
53
64
96
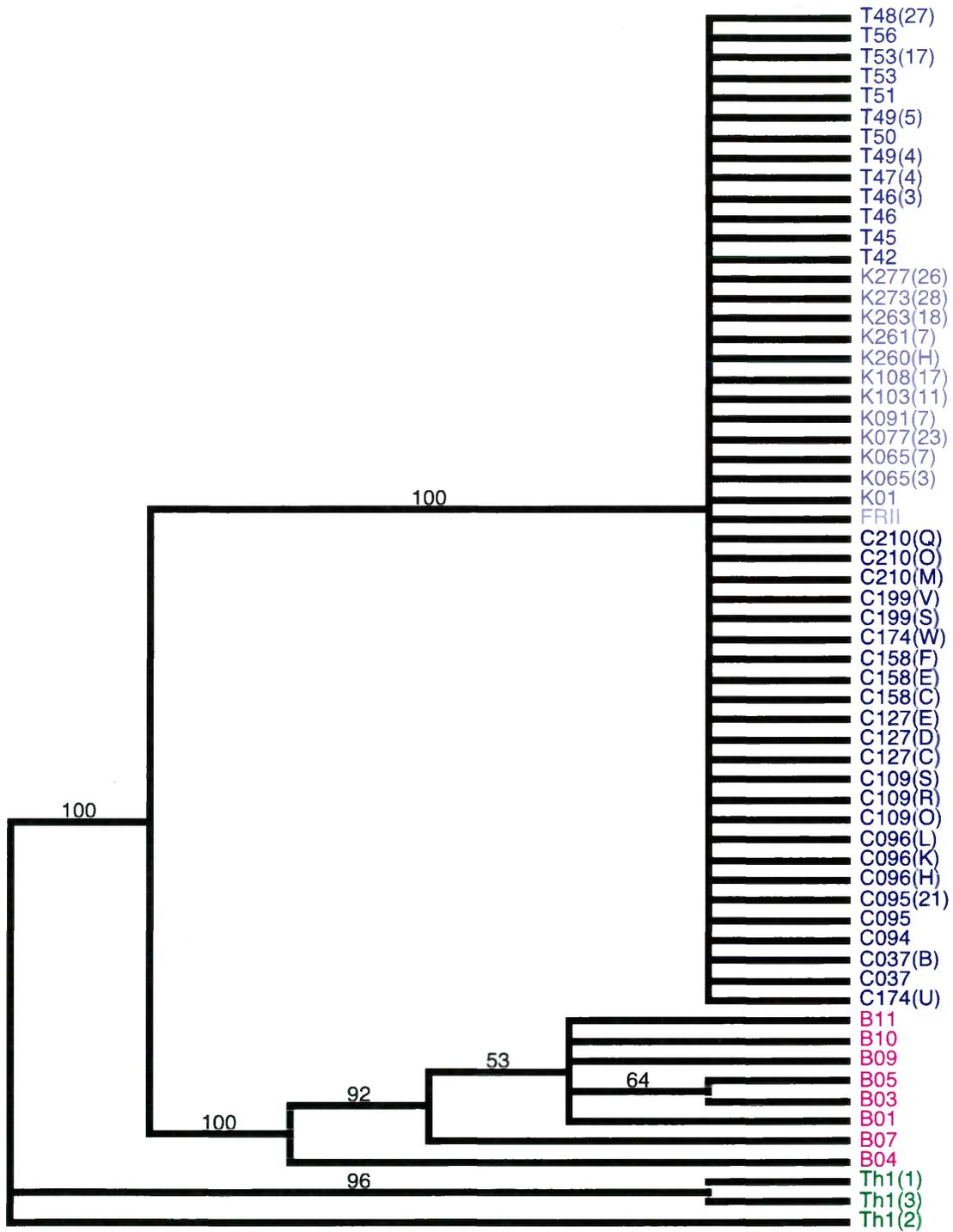
ITS and within-species diversity

The ITS gene region was not useful for analysis of diversity within *R. venosa*. Parsimony analysis failed to resolve any of the *R. venosa* ITS sequences into distinct clades (Figure 8). Within-sample diversity values were unable to be used to assess levels of diversity between separate collection areas. Some individuals possessed clones containing different haplotypes. To illustrate, a clone could have transitions, transversions, or insertion or deletions, that would differ from another clone from the same individual. However, this transition, transversion, or insertion deletion event could be shared by a clone from another individual. In some cases, diversity within clones from the same individual was greater than diversity between individuals, thus creating allelic phylogenies where individual clones would not cluster together. Table 11 shows that the range of diversity from clones from a single individual could be greater than two clones drawn from the same sampling location. Analysis with these data was further complicated by the fact that a disproportionate number of individuals in one sampling location (Chesapeake Bay) had up to three clones per individual sequences, while other areas (Korea, Turkey) had, on average, fewer than two clones per individual sequenced. Parsimony analysis (which, in PAUP*, treated gaps as a separate character) also failed to group together all clones within individuals. This lack of differentiation into statistically robust individual clades within *R. venosa* indicated the ITS gene region was not useful for population level analyses within *R. venosa*. The fact that multiple copies of ITS amplified using the primers from this study was a problem.

Table 11: Range of ITS diversity within individual clones and sample locations for *R.*

venosa. Values reported are percent sequence difference (uncorrected "p"); standard

error is shown in parentheses.

Pairwise comparison of . . .	Range of percent sequence differences
clones from a single individual	0.000 – 0.301 (0 – 0.166)
Korea (clones within a sample location)	0.000 – 0.601 (0 – 0.234)
Turkey (clones within a sample location)	0.000 – 0.938 (0 - 0.311)
Chesapeake Bay (clones within a sample location)	0.000 – 0.502 (0 - 0.132)

DISCUSSION

Several assumptions involving samples and molecular analysis must be examined before a detailed analysis of the results can be made. Interpretation of the results of this study is limited by small sample sizes. The number of individuals surveyed in each sample probably did not completely reflect the genetic composition of the area from which they originated. It is highly unlikely that all haplotypes in a location were represented by the samples. To illustrate, the eight individuals sequenced from Korea revealed eight unique haplotypes. If more of *R. venosa* were sequenced from this area, it is probable more haplotypes would be discovered.

The sampling regime, apart from the small sample sizes, also limits the interpretation of the results. Due to logistical constraints of sampling *R. venosa* from distant locations, I made use of what was available. Individual snails obtained from three sample sites (Korea, Turkey, and Chesapeake Bay) were gathered over very different geographic ranges. In Korea, *R. venosa* were bought at three different coastal fish markets, and thus the individuals could have originated from different points all along the southwestern Korean coastline. In the Chesapeake Bay, *R. venosa* were obtained from a variety of sites within the Bay. In contrast, the Turkish sample was gathered by SCUBA from a single small embayment. While all the individuals from a geographic region were treated in the analysis as single point samples, there is a correlation between size of an area sampled and the levels of genetic diversity discovered at each location. Geographic samples may not be representative of the entire region if small scale spatial structuring of

the population exists. Also, I did not sample for temporal differences within populations. To illustrate, only one sample was taken of the Korean sample (summer 1999) and Turkish sample (October 1999). Returning in subsequent years and sampling among a new collection of year-classes (or sampling among distinctly different size classes that would represent different year-classes during one collection period if sufficient individuals were available) would allow one to ascertain whether factors such as environmental differences or year class recruitment result in substantial changes in the haplotype frequencies found at the location. Such factors could bias interpretation of results and should be excluded when possible.

In the molecular analysis several assumptions are also made that could influence the interpretation of the data. First, is assumed the nucleotide sequences for the cytochrome *b* samples obtained exhibited fidelity to the true individual sequences, with the single exception of one individual whose sequence included a single basepair deletion. This haplotype was excluded from population level analysis, as this deletion was probably not real as the translated product would not be functional. A detailed study of the evolutionary history Northern hemisphere *Nucella* found no deletion events in the 718 bp region between individuals of different species. Thus, I believe the deletion found within the *R. venosa* data set was probably instead due to a *Taq* polymerase error. *Taq* error could have occurred during PCR when the polymerase skipped a nucleotide site, thereby failing to incorporate the correct dNTP into the DNA sequence. If done early in the PCR process, a single faulty base pair addition could be amplified and perpetuated exponentially in the reaction. *Taq* errors occur at very low rates (Lutton et al. 1992);

literature included within the Gibco/BRL *Taq* polymerase chain reaction kit states

incorporation error occur at approximately 5×10^{-4} errors per basepair.

All of the differences found within the 731 bp gene region of cytochrome *b* were

transitions; this is consistent with the theory that transitions should be more likely to

occur than transversions. This trend is found within other mollusc groups studied. For

example, Collins et al. (1996) analyzed a 718 bp region of cytochrome *b* from each of the

known Northern hemisphere species within the genus *Nucella*, and noted that of the

variable sites, 79.5% were transitions, 11.5% were transversions, and 12% were both. In

the *Rapana venosa* data set, the observed transitions (3 of 4 mutations found in the

Chesapeake Bay sample) caused three nonsynonymous substitutions in the three unique

Chesapeake Bay haplotypes (C94, C174, and C199), while the Korean sample possessed

only synonymous substitutions. This is surprising, as the Korean sample possessed a

great deal more diversity. In all cases, the chromatograms given by the sequencer were

inspected by eye to confirm the computer base calls for all sites containing mutations. If

visual examination revealed a base call to be in doubt, that base was labeled as an

ambiguous site and excluded from analysis. The nonsynonymous changes present in the

cytochrome *b* sample may have resulted from a *Taq* error during PCR when an incorrect

dNTP was incorporated into the DNA sequence. If done early in the PCR process, a

single faulty base pair addition could be amplified and perpetuated exponentially in the

reaction. However, errors due to mis-incorporation during PCR have been shown to

occur at very low rates (Lutton et al. 1992), and it is statistically unlikely that multiple

Taq errors occurred in this data set. Thus, I believe that the base changes found in the

cytochrome *b* data set may represent true differences.

Genetic variation in *R. venosa*

Cytochrome *b* sequence data from eight individuals taken from each of three sample

locations allowed the testing of three null hypotheses.

H_{o1}: R. venosa in the sample locations of Korea, Turkey, and Chesapeake Bay share a common gene pool.

The total sample heterogeneity analysis revealed the three sampling locations do not share a common gene pool (p=0.00). **Reject.**

This result was expected, as no shared haplotypes were present between the

diverse Korean sample and the two non-native (Turkey and Chesapeake Bay) samples.

This result was further illustrated by the neighbor-joining tree (Figure 7) constructed

using genetic distance information, and the minimum spanning network, which showed

the relationship between haplotypes (Figure 6). Both figures showed differentiation

between Korean and the Turkish and Chesapeake Bay samples. These mtDNA haplotype

results indicated that *R. venosa* sampled from Turkey or Chesapeake Bay were not direct

matrilineal descendents of the eight snails sampled in Korea.

The possibility that the *R. venosa* found in Turkey and Chesapeake Bay most

likely descended from haplotypes present, but unsampled, in southwestern South Korea

cannot be excluded. This possibility was supported as the relatively more diverse Korean

population had eight unique haplotypes obtained from the sequencing of eight

individuals. This strongly suggested that not all haplotypes were sampled. There is need

to increase the number of individuals surveyed for South Korea, and to expand the

geographic coverage of the sampling range to determine if population structure is present

in the native range. Historical research could identify which locations within the native range of *R. venosa* produced oysters for commercial export to in the Black Sea during the 1940s. Samples from these sites represent likely probable source populations for the resulting *R. venosa* invasion into the Black Sea.

H_{o2}: R. venosa in the introduced sampling locations of Turkey and Chesapeake Bay share a common gene pool.

Heterogeneity analysis of haplotype frequencies for the Turkish and Chesapeake Bay samples failed to reject the hypothesis that the two locations share a common gene pool (p=0.197). **Do Not Reject.**

This result was not surprising, considering the low haplotype diversity found within the samples, and that Turkey and Chesapeake Bay share the only non-unique haplotype found in the study. This haplotype was possessed by all eight of the Turkish individuals and five of the eight Chesapeake Bay individuals sampled and was closely related to the other three haplotypes found within Chesapeake Bay.

The genetic similarity of the Turkey and Chesapeake Bay samples, the two introduced locations, supports the hypothesis that the Black Sea/Western European region was a potential source for the more recent Chesapeake Bay introduction. Researchers who examined shipping routes and vessel transit times to and from possible source locations originally posited the source population for the Chesapeake Bay introduction was European and not Asian in origin (Mann and Harding 2000). Specifically, this hypothesis suggested the *R. venosa* population in Chesapeake Bay was the result of the introduction of planktonic larvae by coal ships deballasting in the port of Hampton Roads. These ships take on coal in Hampton Roads, sail to ports along the

Black Sea, offload the coal, and take on ballast water to stabilize the vessel for the return voyage.

This hypothesis cannot be accepted without qualifications for two reasons. First, the Chesapeake Bay, site of a secondary introduction from the Black Sea according to the above hypothesis, contains diversity not observed in the Turkish sample, the putative source location. This issue will be discussed at length in the next section. Second, the single French individual sequenced shared the same haplotype found in all of the individuals in the Turkish sample and the majority of those in the Chesapeake Bay sample. If the presence of the French haplotype is the result of *R. venosa* range expansion resulting from larval transport by currents or introductions via the ballast water of local ships (Kinzelbach 1986, Koutsoubas and Voultsiadou-Koukoura 1990, and Mann and Harding 2000), *R. venosa* possessing the common haplotype may be common to the waters of western Europe. If so, one cannot definitively ascribe the Chesapeake Bay invasion as originating solely from the Black Sea. Alternatively, the French sample could result from the introduction of the common haplotype from a non-European source (such as the Chesapeake Bay). For example, the movement of the blue crab, *Callinectes sapidus*, from the U.S. East Coast to the Black Sea has been documented (Makarov and Murina 1998). Increased sampling within the Mediterranean Sea, Adriatic Sea, Black Sea, and Atlantic coastal regions would improve our understanding of the relationships of Chesapeake Bay, Black Sea, and other European samples of *R. venosa*.

H_{o3}: Introduction of R. venosa in the Black Sea and Chesapeake Bay were not accompanied by a reduction in genetic variation.

Genetic variation (as measured by h and π) was reduced in the introduced samples (Black Sea and Chesapeake Bay) relative to the native Korean Sample. 95% confidence

intervals were constructed around values of h and π, and were found to be non-overlapping indicating the values are significantly different from one another. **Reject.**

All of the cytochrome *b* genetic diversity statistics were reduced in the two non-native samples relative to the native sample. The Korea sample (n=8) possessed thirteen polymorphic sites, eight haplotypes, a haplotype diversity of 1.0 (S.D. = 0.063), and a mean nucleotide sequence diversity of 0.495% (S.D. = 0.09%). These values were reduced in the introduced populations. Relative to the Korean sample, the Chesapeake Bay sample had one third the number of the polymorphic sites (four), half the number of haplotypes (four), a haplotype diversity value of 0.643 (S.D. = 0.184), and a mean nucleotide sequence diversity value of 0.138% (S.D. = 0.054%). The Turkish sample exhibited even less diversity; it was monomorphic. Ninety-five percent confidence intervals were constructed around the *h* and *π* values for three sample locations; these confidence intervals did not overlap and thus showed that there was a significant reduction in diversity from Korea to Chesapeake Bay to Turkey. Korea was more diverse than either of the introduced locations. Therefore, these results were consistent with the hypothesis that populations in an introduced location display lower levels genetic diversity relative to populations within the native location. As the Chesapeake Bay sample contains three haplotypes not found in the sample from Turkey, one cannot conclude that the Black Sea was the sole source for the Chesapeake Bay introduction.

Of the three locations sampled, the Korean individuals were the most diverse. Diversity may be high relative to the introduced populations because the Korean sample, from within the native range of *R. venosa,* was less likely to have gone through a bottleneck (a period of time where populations are reduced to relatively small number of individuals) than locations where non-native invasions have occurred. However,

diversity may also be high in the Korean sample relative to the introduced locations due differences in sampling; because individuals from Korea were purchased at fish markets it is likely that they were collected over a larger area than either the Turkish (single location collection) or Chesapeake Bay (two locations) samples.

Low genetic diversity within the Turkish sample may be explained by what we know of the historical record. While the mechanism of the initial Black Sea introduction is unknown, Younge and Thompson (1976) suggested it was accomplished when *R. venosa* (possibly as egg cases) were inadvertently included in shipments of Japanese oysters sent by train for aquaculture. If the introduced population was begun only from a small number of egg cases or a limited number of adult females, the number of source mitochondrial DNA lineages might well be very small. An extreme "bottleneck" could result in the low observed haplotype diversity. Additionally or alternatively, it is quite likely that point sampling in all locations did not adequately describe all the variation present. As sampling in Turkey was limited on the temporal and spatial scales (individuals were taken from a small embayment near Trabzon, Turkey), it is most probable the single collection may not comprise all the haplotypes present in the Black Sea. Also, fine scale genetic structure may be present, and could have resulted in the embayment having a population of *R. venosa* that are genetically distinct from other locations in the Black Sea. Unknown physical features within the region may create barriers to gene flow which would allow for the development of population structure, or the point sample may have, by chance alone, been composed of successful cohort from a single mtDNA line. Discriminating between these possible explanations would require

greater sampling coverage along the coastline Turkey and within the entire Black Sea region.

The present level of genetic diversity within *R. venosa* from Chesapeake Bay probably represents a subset of the diversity of the source population(s) (founder effect) and may have been further reduced by drift after the founding propagules were established. The observed levels of diversity are consistent with a reduction of variation after an introduction. That there are such large numbers of *R. venosa* in Chesapeake Bay so soon after the proposed date of the introductions suggests the numbers of individuals post-introduction did not stay small over several generations, making drift unlikely to be an important mechanism for reducing genetic variation after the introduction.

Can we detect a source population for the Chesapeake Bay introduction? Heterogeneity analysis of haplotype frequencies showed Turkey and Chesapeake Bay shared a common gene pool; however, the Chesapeake Bay sample contains three haplotypes that were not present in the Turkish sample. Thus, the additional diversity present in the Chesapeake Bay indicates one or more of the following options: a) the source population for the Chesapeake Bay invasion is not the Black Sea, but instead another location that has both the haplotype held in common between the Turkish sample and the Chesapeake Bay sample as well as the additional haplotypes found in Chesapeake Bay, b) the Chesapeake Bay population is the result of multiple introduction events from several source populations, at least one of which is the Turkish sample or that Chesapeake Bay sample has significant gene flow from a founder population that also exchanges genes with the Turkish sample, or c) the point sample from Turkey was non-representational of the diversity present in that location, and thus the source population

could be Turkey, and if the additional Chesapeake Bay haplotypes found are present in

the Black Sea and were not sampled. With these data, none of the possibilities may be

definitively discarded; thus, no exclusive source population assignment for the

Chesapeake Bay *R. venosa* invasion was possible.

Genetic differentiation between *R. venosa* and *R. bezoar*

Examination of the nuclear ITS rRNA region allowed the null hypothesis evaluating the
taxonomic classification of *R. venosa* and *R. bezoar* to be tested.

H_{o1}: *Sampled R. venosa and R. bezoar represent conspecific members of a single
population, and share a common gene pool.* **Reject.**

The maximum parsimony analysis of the sequence data from the ITS gene region

indicates that individuals assigned to the putative species *R. venosa* and *R. bezoar*

represent two different taxonomic groupings. High bootstrap values (100%) provided

strong statistical support for this result. Comparison of within species sequence ITS

diversity values to the between species sequence divergence values, supports this point as

well, as diversity values were an order of magnitude smaller than the between species

values. The between species sequence diversity (6.1%, not including gaps) was much

greater (more than an order of magnitude) than within group diversity values of 0.2% for

R. venosa individuals and 0.4% for *R. bezoar* individuals. The mean percent pairwise

sequence difference was approximately 14% when either *Rapana* species is compared to

the outgroup, which belongs to the same subfamily (Rapaninae) as *R. venosa* and *R.*

bezoar.

The cytochrome *b* primers used in this study did not amplify a gene product in *R.*

bezoar or the outgroup taxa. This result is probably due to sequence divergence within

the primer sites, and lends support to the hypothesis that *R. venosa* and *R. bezoar* are separate species that do not share a common gene pool. Spolsky et al. (1996) report that 'universal' metazoan cytochrome *b* primers do not amplify a gene product from molluscs, due to the "very ancient branching of molluscs from the basal phylogeny", that creating a universal mollusc cytochrome *b* primer has been difficult due to divergence within the phylum, and that their own species specific cytochrome *b* primers failed to amplify in all populations of the gastropod *Ocomelania* and the related species *Tricula*.

Thus, examination of the ITS data set by maximum parsimony analysis, pairwise distance comparisons of percent sequence differences within and between these putative species, and the failure of the Collins et al. (1996) cytochrome *b* primers to amplify outside the group described as *R. venosa* all indicated that there was substantial genetic distance between the two taxa. *R. venosa* and *R. bezoar* clearly do not share a common gene pool; thus, the classification of the two as separate species was supported by all genetic data gathered in this study.

General Discussion

Genetic variation within *Rapana venosa*

Cytochrome *b* variation was low, with mean percent pairwise differences within

sample locations ranging from 0% in Turkey to 0.14% in Chesapeake Bay, to 0.47% in

Korea. These values, especially those in the non-native samples were lower than

expected when compared to cytochrome *b* variation within some molluscs species. Work

by Spolsky et al. (1996) showed that cytochrome *b* diversity for the prosobranch

gastropod *Oncomelania hupensis* was 3.8 % between locations containing identical

subpopulations and 10-12% among locations containing unique subpopulations.

However, the study did not closely examine diversity within a subpopulation's location –

the cytochrome *b* sequences from two snails within the same location showed a 0.4%

difference (Spolsky et al. 1996).

Allelic variation within cytochrome *b* for some molluscs species was documented

by Merritt et al. (1998), but specific population studies utilizing these universal

molluscan cytochrome *b* primers have not yet been published. Merritt et al. reported that

within a 430 bp cytochrome *b* region sequenced from seven individuals of the bivalve

Deminucla proxima, there were nine nucleotide changes (all C to T transitions) that

resulted in seven haplotypes. Within the long-fin squid *Loglio pealei*, two nucleotide

changes generated three haplotypes in the ten individuals sequenced; however,

cytochrome *b* regions sequenced from other molluscs (spoon clam, *Yoldia limatula*, n=6;

veiled clam, *Solemya velum*, n=7) showed no variation.

Comparison of the cytochrome *b* region amplified using identical primers

(generated by Collins et al. 1996) for *R. venosa* and *Plicopurpura columellaris* and *P.*

patula was performed. The two *Plicopurpura* species are a geminate pair in the Pacific Ocean and belong to the same subfamily, Rapaninae, as *Rapana*. Alignments constructed between the two *Plicopurpura* species show substantial sequence variation. Of 718 bp sequenced, these two congeneric taxa share only 617 conserved sites. When the most common haplotype from the *R. venosa* population study was aligned alongside the *Plicopurpura* taxa, only 535 residues were conserved. Also, the *R. venosa* sequence was 731 bp long, compared to 718 bp for the two *Plicopurpura spp*. These substantial differences within the *Plicopurpura* genus, as well as between these species and *R. venosa*, which belong to the same subfamily, support the idea that cytochrome *b* is (somewhat) free to accumulate sequence differences, and individual mutations can still result in a functional cytochrome *b* gene. This lends further credence to the idea that the sequence variation found within *R. venosa* is real and not the result of *Taq* errors.

The internal transcribed spacer ITS rRNA gene region analysis

Between-species analysis

Maximum parsimony analysis of the ITS gene region and comparison of within- and between-species sequence pairwise comparison values revealed that *R. venosa* and *R. bezoar* are not conspecific populations exhibiting phenotypic plasticity along an environmental cline, but are instead correctly described by the scientific community as separate species within the genus *Rapana*.

The ITS divergence values found in this study are similar to those reported in the literature by researchers using ITS to test for species status among geographically distant but morphologically similar taxa. Anderson and Adlard (1994) used a portion of the ITS

locus to resolve the taxonomic status of the Sydney rock oyster, *Saccostrea*

commercialis, relative to that of the New Zealand rock oyster, *S. glomeratus.* The result,

that *S. commercialis* and *S. glomeratus* possessed 0.0% sequence divergence, supported a

synonymous grouping of the two putative species. Remigio and Blair (1997) used ITS to

study "problematic" species relationships within four pulmonate stagnicoline taxa. They

found genetic distances of 0.5 – 1.1% for ITS-1 and 0.2 – 0.2% for ITS-2 among the

three taxa *Stagnicola catascopium, S. emarginata,* and *S. elodes* (all within subgenus

Stagnicola s.str.) and concluded the taxa, currently described as separate species, may

represent a single species (or recently diverged sister species). A fourth stagnicoline

species, *S. caperata* (subgenus *Hinkleyia*), differed from these three taxa by 18.7 – 20.2%

at ITS-1 and 9.6 – 10.2% at ITS-2; *S. caperata* also possessed an ITS-1 sequence 44-45

bp longer and an ITS-2 region 10-13 bp shorter than the other taxa. These considerable

genetic differences between *S. caperata* and the other taxa led Remigio and Blair to

conclude *S. caperata* was not only not conspecific with the other taxa, but suggested

these data, coupled with differences in shell size and shape, prompted consideration of *S.*

caperata as a separate genus. Remigio and Blair also estimate nucleotide divergence

from restriction length polymorphism data from a study using ITS-1 to examine

taxonomic relationships within the gastropod genus *Bulinus* (Stothard et al. 1996).

Remigio and Blair inferred that different species groups within *Bulinus* possessed

nucleotide divergence values of 14.2% at ITS-1, while taxa belonging to the same species

group differed by 2.0% at ITS-1.

The whole locus ITS (ITS-1, 5.8S, ITS-2) divergence values reported in this

thesis for assessing the taxonomy of *R. venosa* and *R. bezoar* are consistent with the

estimates reported in the above studies, and support the following conclusions. First, *R. venosa* individuals form a single species (within-species diversity = 0.2%). Second, *R. bezoar* individuals form a single species (within-species diversity = 0.4%). Third, *R. venosa* and *R. bezoar* are congeneric, rather than conspecific (between-species divergence = 6.1%). Fourth, the *Rapana* genus (represented by *R. venosa* and *R. bezoar*) is different than that of the outgroup, *Stramonita hemastoma* (divergence between *R. venosa* and *R. bezoar* and *S. hemastoma* = 13.9 – 14.2%). Due these high percent sequence divergence values and consistent size differences at the ITS locus among *R. venosa, R. bezoar* and *S. hemastoma*, it is clear that these three taxa represent unique species. Based on the genetic data, these taxa each possess their own evolutionary trajectories and can be described as "an exclusive group . . . [where] members are all more closely related to each other than to any organism outside the group", and can thus be defined as separate species based on the Evolutionary Species Concept and the Genealogical Species concept, respectively (Harrison, 1998).

Designation of *R. venosa* as a distinct species from *R. bezoar*, while interesting as a question of alpha taxonomy alone, is also of practical import. Resolution of this relationship has implications for management of the introduced population in Chesapeake Bay. As the two are separate species, rather than a single conspecific population located along an environmental cline, one can use the present known distribution and environmental tolerances for R. venosa alone to predict which ports along the US coast are potentially at risk of invasion by the Chesapeake Bay *R. venosa* population. Specifically, as *R. venosa* is a more temperate species than the smaller, more tropical *R. bezoar*, several southeastern ports may possess thermal profiles that make them less

susceptible to colonization by *R. venosa*. Thus, the knowledge *R. venosa* and *R. bezoar* are separate species may assist in the management of *R. venosa* in Chesapeake Bay.

Within-species analysis

While the ITS rRNA gene region was informative in differentiating *R. venosa* and *R. bezoar* as separate species, it was not a useful genetic marker for assessing population structure within *R. venosa*. For intraspecific analyses, one to three clones of the ITS region were sequenced from each individual. There was a wide range of divergence between clones from the same individual, and sometimes this within-individual clonal diversity exceeded the diversity between individuals. For example, different clones from a single individual possessed transitions, transversions, or insertion/deletion events. These data resulted in allelic phylogenies where ITS clones from individuals did not cluster by individual or even by sample location. The presence of multiple haplotypes within individuals, which, when coupled with the failure to evenly sample the number of clones per individual among location, rendered the marker useless for studying population structure within *R. venosa* or assessing genetic diversity among sample locations. The decision to reduce the number of clones sequenced to one per individual was unfortunate in hindsight, but was made due to time and cost constraints before most of the sequencing was performed and prior to analysis of any of the ITS sequence data.

ITS is a multi-copy gene; hundreds of copies can be present in an individual genome, and intra-individual sequence variation is common (Caporale et al. 1997, Remigio and Blair 1997). Caporale et al. (1997) used direct sequencing of ITS-1 to study population structure of the soft shell clam (*Mya arenaria*) along coastal New England.

She found two variants of ITS-1 within some individuals; variants differed by a single point mutation and a three basepair insertion. Caporale et al. (1997) stated that "variation among copies of the same gene would be detected by direct sequencing only if each variant were common". They noted that sequencing would, in the presence of both copies, selectively amplify one variant to the complete exclusion of the other. They suggested that the variant with higher copy number would be the one preferentially amplified, and thus the direct "sequencing process may be presenting a biased representation of the distinctive variants in a genome and may produce misleading results owing to the potential influence of allele copy number". Remigio and Blair (1997) cloned ITS to examine the relationships among four stagnicoline snail taxa. They discovered clonal variants for two of the four taxa sequenced. These variants possessed substitutions at several sites, and had deletion or insertion events present. Similar insertions or deletions, and synonymous changes were found within the ITS clonal variants within a single individual in my study. Remigio and Blair (1997) state that a maximum parsimony analysis was not performed as there was a "lack of phylogenetically informative sites within the ITS data set.

ITS has proved informative for other populations genetic studies in molluscs and invertebrates. Caporale et al. (1997) performed allele specific PCR to amplify ITS-1 variants and found that soft shell clam (*Mya arenaria*) populations distributed in coastal New England were not significantly heterogeneous. Schizas et al. (1999) surveyed both cytochrome *b* and ITS-1 to examine populations of the harpacticoid copepod *Microarthridion littorale* in the southeastern and gulf coastal regions of the United States. Direct sequencing of one individual per location was performed, and 130 parsimoniously

informative sites within the 474 bp ITS-1 region were found. Clades based on ITS and cytochrome *b* data were concordant, with the exception of one monophyletic clade present in the cytochrome *b* data set that collapsed in the ITS-1 analysis. They attributed this lack of resolution to the reduced population size of the mitochondrial gene relative to the four-fold larger population size of the nuclear autosomal ITS gene region.

The mitochondrial cytochrome *b* locus in the analysis of *R. venosa* population structure also shows significant structuring relative to that of the nuclear autosomal ITS region. This pattern may be attributable to the same cause (i.e. the relative population sizes of mtDNA relative to nuclear DNA), but the pattern is undoubtedly exacerbated by the fact the Chesapeake Bay and Turkish *R. venosa* samples are the result of introductions. The presence of polymorphisms in the introduced location is due to transportation of ancestral polymorphism from the source population(s), and not the result of mutation, drift, and allelic fixation.

Further studies

Sampling – It would enhance the project if there were more intense sampling over the locations where *R. venosa* and *R. bezoar* are found.

First, for the examination of genetic relationships among native and non-native *R. venosa*, there should be increased coverage of the native range should. Sampling should include individuals from Japan and China, as well as additional sample locations on the opposite side of the Korean Peninsula. Also, to attempt to document invasion dynamics, including possible source locations for recent introductions, inclusion of additional non-

native samples is required. Adding sample locations from additional areas within the Black Sea could provide a better picture of diversity in the Black Sea invasion. This sampling, in combination with sampling *R. vensoa* in the Marmar, Adriatic, and Mediterranean Seas, and the Atlantic coast of France, could provide information about the spread of *R. venosa* in European waters. Sampling *R. venosa* in Uruguay could also be informative in tracking relationships among world-wide introductions.

Second, additional sampling would further support the evaluation of the alpha taxonomy of *Rapana*. Specifically, *R. bezoar* individuals should be sampled from at least one other geographic location. This would allow us to evaluate the degree of difference between sampling locations and between the two putative species for both *R. venosa* and *R. bezoar*. Though not included in this thesis, it would also be interesting to obtain samples of *R. rapiformis*, the third described species within the genus *Rapana*.

Creation of Rapana-specific cytochrome b primers – Obtaining strong amplification products of the cytochrome *b* gene region was difficult for *R. venosa* individuals and impossible for *R. bezoar* individuals. Analysis of conserved regions of the cytochrome *b* sequences for the 28 *R. venosa* individuals that have been sequenced could allow for the creation of several new *Rapana*-specific primer pairs. The new primer pairs would be helpful for two reasons. First, *Rapana*-specific primers could provide consistently strong cytochrome *b* amplifications for *R. venosa*, saving costly re-amplification attempts. Second, the ability to amplify cytochrome *b* in *R. bezoar* would strengthen the analysis of the taxonomic relationship of *R. bezoar* and *R. venosa*. Information from a second

molecular marker would provide additional evidence that could be used to evaluate the taxonomic relationship of the putative species *R. venosa* and *R. bezoar*.

RFLP analysis of cytochrome b – With an increase in the number of sample locations and the number of samples from each location analyzed, the cost of cloning and sequencing cytochrome *b* for a population genetics project would soon become prohibitive. Analysis of restriction enzyme sites within the 731 bp cytochrome *b* gene region for the twenty-five *R. venosa* sequenced revealed five restriction enzyme polymorphisms. The restriction enzymes *Hinf*I, *Taq*I, and *Nde*I each generate unique digestion patterns for one of the twenty-five *R. venosa* sequenced. The enzymes *Mbo*II and *Rsa*I cut at two of the four phylogenetically informative sites found to separate the individuals within the Korean sample from the Turkish and Chesapeake Bay samples. Digestion of cytochrome *b* using these enzymes would produce restriction fragment haplotypes that could be used to examine variation in a large number of samples.

Additional use of nuclear marker(s). The nuclear internal transcribed spacer (ITS) region was not, due to its diverse multicopy status, suitable for the population genetic study of *R. venosa*. Detailed assessment of the genetics of *R. venosa* invasions requires the use of biparentally inherited nuclear markers that possess a large amount of variation. The marker must also be able to be used with ease and with relatively low cost to allow screening of many individuals from different sample locations. A large scale nuclear diversity assessment would probably require the development of several microsatellite loci. While initial time and budget expenses would be high during the period of

microsatellite development, the cost of screening individuals after this period would be low. Data from large numbers of individuals from many sample locations to be screened relatively quickly with low cost. Possession of microsatellite data allow explicit testing for the presence of bottlenecks during invasion and allow for the examination of effect of invasion, mutation, and drift, on the nuclear genome of *R. venosa*.

CONCLUSIONS

Potential sources of introductions

The Turkish and Chesapeake Bay populations of *R. venosa* do not share a common gene pool with the Korean sample. This suggested that none of the introduced *R. venosa* in Turkey and Chesapeake Bay sampled in this study were direct matrilineal descendents of mtDNA lineages sampled in this study.

The two introduced locations, Turkey and Chesapeake Bay, share a common gene pool. However, as there are haplotypes present in the more diverse Chesapeake Bay location that were not discovered within the monomorphic Turkish population, the additional genetic variation may have originated from another location. Note, however, that the number of halplotypes discovered at the two locations could be an artifact of the small number of individuals surveyed in each location; additional sites and increased numbers of individuals sampled would assist in discriminating whether more diversity is present in the Turkey or Black Sea. Thus, no exclusive source population assignment for the Chesapeake Bay *R. venosa* invasion is possible.

Genetic diversity between native and non-native sample locations

The amount genetic diversity present in the both introduced sample locations of Chesapeake Bay and Turkey was reduced compared to the level of diversity present in the Korean sample, which was taken from within the native range of *R. venosa*. That the introduced locations possessed reduced diversity relative to the native population suggested that introduction events in *R. venosa* are associated with a loss of diversity.

This reduction in diversity is probably the result of the founder effect, and not genetic drift caused by low sample sizes after the invasion. Genetic drift is not believed to be important, as large numbers (2000+) of *R. venosa* were found in Chesapeake Bay within 10-20 years of the proposed date of the introduction. As this rapid population increase would occur over a relatively few number of generations, founder effect, rather than genetic drift, was likely to be an important mechanism for reducing genetic variation after the introduction.

Taxonomic classification of *R. venosa* and *R. bezoar*

The maximum parsimony analysis and associated statistical bootstrap support, as well as comparisons of within- and between- percent mean sequence diversity of the nuclear ITS rRNA region revealed that the putative *R. venosa* and *R. bezoar* individuals studied were different species. Thus, the current scientific classification of *R. venosa* and *R. bezoar* as separate congeneric species is correct.

Appendix A:
731 bp cytochrome *b* gene region

```
C041    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C095    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C094    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTNTCTTGC    [60]
C149    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C158    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C174    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C199    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
C210    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
FRII    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K108    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K118    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K250    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K255    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K260    TTTTGGGTCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K261    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K263    TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
K277    TTTTGGGTCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T42     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T47     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T48     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T49     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T50     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T53     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTNTCTTGC    [60]
T54     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
T56     TTTTGGATCTCTTTTAGGACTCTGTTTGGTAATTCAAATTGCTACTGGGCTGTTTCTTGC    [60]
```

```
C041    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C095    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C094    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C149    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C158    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C174    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C199    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
C210    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
FRII    AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K108    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K118    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K250    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K255    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K260    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATCAGGCGAGA    [120]
K261    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K263    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
K277    AATGCATTATACGGCTCACGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T42     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T47     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T48     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T49     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T50     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T53     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T54     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
T56     AATGCATTATACGGCTCATGTAGATCTAGCATTTAGTTCTGTAGTGCATATTAGGCGAGA    [120]
```

```
C041    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C095    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C094    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTNTCAT    [180]
C149    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C158    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C174    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C199    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
C210    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
FRII    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K108    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K118    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K250    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K255    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K260    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K261    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K263    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
K277    TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T42     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T47     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T48     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T49     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T50     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T53     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T54     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]
T56     TGTCACTTATGGTTGACTTCTTCGAGCACTTCATGCTAATGGAGCCTCTTGATTTTTCAT    [180]


C041    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C095    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C094    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C149    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C158    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C174    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C199    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
C210    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
FRII    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
K108    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K118    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K250    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K255    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K260    TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
K261    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K263    TTGTTTATATTTTCATATTGCTCGTGGTATATACTACGGATCATATCTTTATTTGCACGT    [240]
K277    TTGTTTATATTTTCATATTGCTCGTGGTATNTACTACGGATCATATCTTTATTTGCACGT    [240]
T42     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T47     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T48     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T49     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T50     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T53     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T54     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
T56     TTGTTTATATTTTCATATTGCTCGTGGTATGTACTACGGATCATATCTTTATTTGCACGT    [240]
```

```
C041   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C095   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C094   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C149   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C158   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C174   TTGANACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C199   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
C210   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
FRII   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K108   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K118   TTGAAACGTTGGAGTAATTCTCTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K250   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K255   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K260   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K261   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K263   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
K277   TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T42    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T47    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T48    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T49    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T50    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T53    TTGAAACGTTGGAGTAANTCTNNTATNNCNAATTATAGGAACAGCANTTTTAGGATATGT   [300]
T54    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]
T56    TTGAAACGTTGGAGTAATTCTTTTATTTCTAATTATAGGAACAGCATTTTTAGGATATGT   [300]


C041   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C095   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C094   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C149   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C158   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C174   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C199   TCTTCCATGAGGGCAAATATCTTTTCGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
C210   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
FRII   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTNATTACAAATTTACTCTCAGC   [360]
K108   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCGACTGTAATTACAAATTTACTCTCAGC   [360]
K118   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K250   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K255   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K260   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K261   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K263   TCTTCCATGAGGACAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
K277   TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T42    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T47    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T48    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T49    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T50    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T53    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T54    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
T56    TCTTCCATGAGGGCAAATATCTTTTTGAGGAGCAACTGTAATTACAAATTTACTCTCAGC   [360]
```

```
C041   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C095   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C094   ·AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA  [420]
C149   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C158   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C174   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C199   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
C210   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
FRII   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K108   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K118   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGGGGGTTTGCAGTTGATAA   [420]
K250   AGTTCCATACGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K255   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K260   AGTTCCGTATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K261   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K263   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
K277   AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T42    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T47    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T48    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T49    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T50    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T53    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T54    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]
T56    AGTTCCATATGTTGGTAAAATGTTAGTAGAATGAGTTTGAGGAGGGTTTGCAGTTGATAA   [420]


C041   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C095   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C094   TGCAACTCTTACACGATTCTTCGCTCTTCATTNTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C149   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C158   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C174   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C199   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
C210   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
FRII   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K108   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K118   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K250   TGCAACTCTTACACGATTTTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K255   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K260   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K261   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K263   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
K277   TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T42    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T47    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T48    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T49    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T50    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T53    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T54    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
T56    TGCAACTCTTACACGATTCTTCGCTCTTCATTTTCTTTTACCATTTGCTGTTGCAGGCTT   [480]
```

```
C041   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C095   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C094   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATCAAA   [540]
C149   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C158   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C174   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C199   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
C210   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
FRII   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K108   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K118   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K250   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K255   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAATAATCCATTAGGATTAAA   [540]
K260   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K261   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
K263   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAATAATCCATTAGGATTAAA   [540]
K277   AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T42    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T47    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T48    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T49    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T50    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T53    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T54    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]
T56    AGCAATCTTACATATGCTATTCCTTCATGAAACAGGCTCTAACAATCCATTAGGATTAAA   [540]


C041   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C095   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C094   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C149   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C158   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C174   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGGATTTAGTCGGTTT   [600]
C199   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
C210   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
FRII   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K108   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K118   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K250   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K255   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K260   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K261   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K263   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
K277   TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATCTTAGTCGGTTT   [600]
T42    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T47    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T48    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T49    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T50    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T53    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T54    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
T56    TAGAGATGGTGAAAAAGTTCCATTTCATTCCTACTACACTTTTAAAGATTTAGTCGGTTT   [600]
```

```
C041    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C095    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C094    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C149    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C158    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C174    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C199    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
C210    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
FRII    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K108    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K118    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K250    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K255    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K260    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K261    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K263    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
K277    TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T42     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T47     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T48     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T49     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T50     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T53     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T54     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]
T56     TTTAGTAGTTATAACACTTTTAACAATATTAGCTTTGTTTTCACCTCAATTATTAACAGA    [660]


C041    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
C095    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCC-AGTACACATTCAACCAGAGTG    [719]
C094    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
C149    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
C158    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
C174    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCGCCCCAGTACACATTCAACCAGAGTG    [720]
C199    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
C210    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
FRII    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K108    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K118    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K250    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K255    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K260    TCCTGAAAACTTTATTCCAGCTAACCCTCTTGTCACCCCAGTACACATTCAGCCAGAGTG    [720]
K261    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K263    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
K277    TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCANCCAGAGTG    [720]
T42     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T47     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T48     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T49     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T50     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T53     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T54     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
T56     TCCTGAAAACTTTATTCCAGCTAATCCTCTTGTCACCCCAGTACACATTCAACCAGAGTG    [720]
```

```
C041    ATNCTTTCTTT    [731]
C095    ATACTTTCTTT    [730]
C094    ATACTTTCTTT    [731]
C149    ATACTTTCTTT    [731]
C158    ATACTTTCTTT    [731]
C174    ATACTTTCTTT    [731]
C199    ATACTTTCTTT    [731]
C210    ATACTTTCTTT    [731]
FRII    ATACTTTCTTT    [731]
K108    ATACTTTCTTT    [731]
K118    ATACTTTCTTT    [731]
K250    ATATTTTCTTT    [731]
K255    ATACTTTCTTT    [731]
K260    ATACTTTCTTT    [731]
K261    ATACTTTCTTT    [731]
K263    ATACTTTCTTT    [731]
K277    ATACTTTCTTT    [731]
T42     ATACTTTCTTT    [731]
T47     ATACTTTCTTT    [731]
T48     ATACTTTCTTT    [731]
T49     ATACTTTCTTT    [731]
T50     ATACTTTCTTT    [731]
T53     ATACTTTCTTT    [731]
T54     ATACTTTCTTT    [731]
T56     ATACTTTCTTT    [731]
```

Appendix B: Amino acid translation of the 731 bp cytochrome *b* sequence for *R. venosa*. Translation start site was at position two. The invertebrate mitochondrial translation table from MEGA2.1 was used.

```
T47    FGSLLGLCLV IQIATGLLAM HYTAHVDLAF SSVVHISRDV TYGWLLRALH  [ 51]
T56    .......... .......... .......... .......... ..........  [ 51]
T54    .......... .......... .......... .......... ..........  [ 51]
T53    .......... .......... .......... .......... ..........  [ 51]
T50    .......... .......... .......... .......... ..........  [ 51]
T49    .......... .......... .......... .......... ..........  [ 51]
T48    .......... .......... .......... .......... ..........  [ 51]
T42    .......... .......... .......... .......... ..........  [ 51]
FRII   .......... .......... .......... .......... ..........  [ 51]
K277   .......... .......... .......... .......... ..........  [ 51]
K263   .......... .......... .......... .......... ..........  [ 51]
K261   .......... .......... .......... .......... ..........  [ 51]
K260   .......... .......... .......... .......... ..........  [ 51]
K255   .......... .......... .......... .......... ..........  [ 51]
K250   .......... .......... .......... .......... ..........  [ 51]
K118   .......... .......... .......... .......... ..........  [ 51]
K108   .......... .......... .......... .......... ..........  [ 51]
C210   .......... .......... .......... .......... ..........  [ 51]
C199   .......... .......... .......... .......... ..........  [ 51]
C174   .......... .......... .......... .......... ..........  [ 51]
C158   .......... .......... .......... .......... ..........  [ 51]
C149   .......... .......... .......... .......... ..........  [ 51]
C095   .......... .......... .......... .......... ..........  [ 51]
C094   .......... .......... .......... .......... ..........  [ 51]
C041   .......... .......... .......... .......... ..........  [ 51]


T47    ANGASWFICL YFHIARGYYG SYLYLHVWVG VIMGTALGYV LPWGQMSFWG  [110]
T56    .......... .......... .......... .......... ..........  [110]
T54    .......... .......... .......... .......... ..........  [110]
T53    .......... .......... .......... .......... ..........  [110]
T50    .......... .......... .......... .......... ..........  [110]
T49    .......... .......... .......... .......... ..........  [110]
T48    .......... .......... .......... .......... ..........  [110]
T42    .......... .......... .......... .......... ..........  [110]
FRII   .......... .......... .......... .......... ..........  [110]
K277   .......... .......... .......... .......... ..........  [110]
K263   .......... .......... .......... .......... ..........  [110]
K261   .......... .......... .......... .......... ..........  [110]
K260   .......... .......... .......... .......... ..........  [110]
K255   .......... .......... .......... .......... ..........  [110]
K250   .......... .......... .......... .......... ..........  [110]
K118   .......... .......... .......... .......... ..........  [110]
K108   .......... .......... .......... .......... ..........  [110]
C210   .......... .......... .......... .......... ..........  [110]
C199   .......... .......... .......... .......... ........R.  [110]
C174   .......... .......... .......... .......... ..........  [110]
C158   .......... .......... .......... .......... ..........  [110]
C149   .......... .......... .......... .......... ..........  [110]
C095   .......... .......... .......... .......... ..........  [110]
C094   .......... .......... .......... .......... ..........  [110]
C041   .......... .......... .......... .......... ..........  [110]
```

```
T47   ATITNLLSAV PYVGKMLVEW VWGGFAVDNA TLTRFFALHL LPFAVAGLAI   [162]
T56   .......... .......... .......... .......... ..........   [162]
T54   .......... .......... .......... .......... ..........   [162]
T53   .......... .......... .......... .......... ..........   [162]
T50   .......... .......... .......... .......... ..........   [162]
T49   .......... .......... .......... .......... ..........   [162]
T48   .......... .......... .......... .......... ..........   [162]
T42   .......... .......... .......... .......... ..........   [162]
FRII  .......... .......... .......... .......... ..........   [162]
K277  .......... .......... .......... .......... ..........   [162]
K263  .......... .......... .......... .......... ..........   [162]
K261  .......... .......... .......... .......... ..........   [162]
K260  .......... .......... .......... .......... ..........   [162]
K255  .......... .......... .......... .......... ..........   [162]
K250  .......... .......... .......... .......... ..........   [162]
K118  .......... .......... .......... .......... ..........   [162]
K108  .......... .......... .......... .......... ..........   [162]
C210  .......... .......... .......... .......... ..........   [162]
C199  .......... .......... .......... .......... ..........   [162]
C174  .......... .......... .......... .......... ..........   [162]
C158  .......... .......... .......... .......... ..........   [162]
C149  .......... .......... .......... .......... ..........   [162]
C095  .......... .......... .......... .......... ..........   [162]
C094  .......... .......... .......... .......... ..........   [162]
C041  .......... .......... .......... .......... ..........   [162]


T47   LHMLFLHETG SNNPLGLNSD GEKVPFHSYY TFKDLVGFLV VMTLLTMLAL   [212]
T56   .......... .......... .......... .......... ..........   [212]
T54   .......... .......... .......... .......... ..........   [212]
T53   .......... .......... .......... .......... ..........   [212]
T50   .......... .......... .......... .......... ..........   [212]
T49   .......... .......... .......... .......... ..........   [212]
T48   .......... .......... .......... .......... ..........   [212]
T42   .......... .......... .......... .......... ..........   [212]
FRII  .......... .......... .......... .......... ..........   [212]
K277  .......... .......... .......... .......... ..........   [212]
K263  .......... .......... .......... .......... ..........   [212]
K261  .......... .......... .......... .......... ..........   [212]
K260  .......... .......... .......... .......... ..........   [212]
K255  .......... .......... .......... .......... ..........   [212]
K250  .......... .......... .......... .......... ..........   [212]
K118  .......... .......... .......... .......... ..........   [212]
K108  .......... .......... .......... .......... ..........   [212]
C210  .......... .......... .......... .......... ..........   [212]
C199  .......... .......... .......... .......... ..........   [212]
C174  .......... .......... .......... .......... ..........   [212]
C158  .......... .......... .......... .......... ..........   [212]
C149  .......... .......... .......... .......... ..........   [212]
C095  .......... .......... .......... .......... ..........   [212]
C094  .......... ......S... .......... .......... ..........   [212]
C041  .......... .......... .......... .......... ..........   [212]
```

```
T47    FSPQLLTDPE  NFIPANPLVT  VHIPEWFL [243]
T56    ..........  ..........  ........ [243]
T54    ..........  ..........  ........ [243]
T53    ..........  ..........  ........ [243]
T50    ..........  ..........  ........ [243]
T49    ..........  ..........  ........ [243]
T48    ..........  ..........  ........ [243]
T42    ..........  ..........  ........ [243]
FRII   ..........  ..........  ........ [243]
K277   ..........  ..........  ........ [243]
K263   ..........  ..........  ........ [243]
K261   ..........  ..........  ........ [243]
K260   ..........  ..........  ........ [243]
K255   ..........  ..........  ........ [243]
K250   ..........  ..........  ........ [243]
K118   ..........  ..........  ........ [243]
K108   ..........  ..........  ........ [243]
C210   ..........  ..........  ........ [243]
C199   ..........  ..........  ........ [243]
C174   ..........  .........A  ........ [243]
C158   ..........  ..........  ........ [243]
C149   ..........  ..........  ........ [243]
C095   ..........  ..........  ........ [243]
C094   ..........  ..........  ........ [243]
C041   ..........  ..........  ........ [243]
```

Appendix C: Internal Transcribed Spacer Region DNA sequences

```
Th1(1)    ATCATTACCGGTGGTTACACAACCTTATCGTGTTGCCGTTGTTCTCCTCTTTTGT    [55]
Th1(2)    ATCATTACCGGTGGTTACACAACCTTATCGTGTTGCCGTTGTTCTCCTCTTTTGT    [55]
Th1(3)    ATCATTACCGGTGGTTACACAACCTTATCGTGTTGCCGTTGTTCTCCTCTTTTGT    [55]
T48(27)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T56(W)    ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T53(17)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T53       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T51       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T49(5)    ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T50       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T49(4)    ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T47(4)    ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T46(3)    ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T46       WTCATTACCGGT---TAC-CACCG----------------ACTC--------T    [25]
T45       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
T42       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K277(26)  ARGATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K273(28)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K263(18)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K261(7)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K260(H)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K108(17)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K103(11)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K091(7)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K077(23)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K065(7)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K065(3)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
K01       ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
FRII      ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C210(Q)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C210(O)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C210(M)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C199(V)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C199(S)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C174(W)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C158(F)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C158(E)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C158(C)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C127(E)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C127(D)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C127(C)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C109(S)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C109(R)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C109(O)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C096(L)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C096(K)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C096(H)   MTCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C095(21)  ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C095      ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C094      ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C037(B)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C037      ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
C174(U)   ATCATTACCGGT---TAC-CACCG-----------------ACTC--------T    [25]
B11       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B10       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B09       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B07       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B05       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B04       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B03       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
B01       ATCATTACCGGT---TAC-CACCGTTTTCTTTGCGGTTCAACACTCCCATCGTGT    [51]
```

```
Th1(1)     CTGTTCTCCCCTCCCC-----CCGCCGCGCGCTTTCAGGAGCCGCGGTAGGTAAG    [105]
Th1(2)     CTGTTCTCCCCTCCCCTTCCCCCGCCGCGCGCTTTCAGGAGCCGCGGTAGGTAAG    [110]
Th1(3)     CTGTTCTCCCCTCCCC-----CCGCCGCGCGCTTTCAGGAGCCGCGGTAGGTAAG    [105]
T48(27)    -CGT------CGTGCCC--------------------------------------    [35]
T56(W)     -CGT------CGTGCCC--------------------------------------    [35]
T53(17)    -CGT------CGTGCCC--------------------------------------    [35]
T53        -YGT------CGTGCCC--------------------------------------    [35]
T51        -CGT------CGTGCCC--------------------------------------    [35]
T49(5)     -CGT------CGTGCCC--------------------------------------    [35]
T50        -CGT------CGTGCCC--------------------------------------    [35]
T49(4)     -CGT------CGTGCCC--------------------------------------    [35]
T47(4)     -CGT------CGTGCCC--------------------------------------    [35]
T46(3)     -CGT------CGTGCCC--------------------------------------    [35]
T46        -MGT------CGTGCCC--------------------------------------    [35]
T45        -CGT------CGTGCCC--------------------------------------    [35]
T42        -CGT------CGTGCCC--------------------------------------    [35]
K277(26)   -CGT------CGTGCCC--------------------------------------    [35]
K273(28)   -CGT------CGTGCCC--------------------------------------    [35]
K263(18)   -CGT------CGTGCCC--------------------------------------    [35]
K261(7)    -CGT------CGTGCCC--------------------------------------    [35]
K260(H)    -CGT------CGTGCCC--------------------------------------    [35]
K108(17)   -CGT------CGTGCCC--------------------------------------    [35]
K103(11)   -CGT------CGTGCCC--------------------------------------    [35]
K091(7)    -CGT------CGTGCCC--------------------------------------    [35]
K077(23)   -CGT------CGTGCCC--------------------------------------    [35]
K065(7)    -CGT------CGTGCCC--------------------------------------    [35]
K065(3)    -CGT------CGTGCCC--------------------------------------    [35]
K01        -CGT------CGTGCCC--------------------------------------    [35]
FRII       -CGT------CGTGCCC--------------------------------------    [35]
C210(Q)    -CGT------CGTGCCC--------------------------------------    [35]
C210(O)    -CGT------CGTGCCC--------------------------------------    [35]
C210(M)    -CGT------CGTGCCC--------------------------------------    [35]
C199(V)    -CGT------CGTGCCC--------------------------------------    [35]
C199(S)    -CGT------CGTGCCC--------------------------------------    [35]
C174(W)    -CGT------CGTGCCC--------------------------------------    [35]
C158(F)    -CGT------CGTGCCC--------------------------------------    [35]
C158(E)    -CGT------CGTGCCC--------------------------------------    [35]
C158(C)    -CGT------CGTGCCC--------------------------------------    [35]
C127(E)    -CGT------CGTGCCC--------------------------------------    [35]
C127(D)    -CGT------CGTGCCC--------------------------------------    [35]
C127(C)    -CGT------CGTGCCC--------------------------------------    [35]
C109(S)    -CGT------CGTGCCC--------------------------------------    [35]
C109(R)    -CGT------CGTGCCC--------------------------------------    [35]
C109(O)    -CGT------CGTGCCC--------------------------------------    [35]
C096(L)    -CGT------CGTGCCC--------------------------------------    [35]
C096(K)    -CGT------CGTGCCC--------------------------------------    [35]
C096(H)    -CGT------CGTGCCC--------------------------------------    [35]
C095(21)   -CGT------CGTGCCC--------------------------------------    [35]
C095       -CGT------CGTGCCC--------------------------------------    [35]
C094       -CGT------CGTGCCC--------------------------------------    [35]
C037(B)    -CGT------CGTGCCC--------------------------------------    [35]
C037       -CGT------CGTGCCC--------------------------------------    [35]
C174(U)    -CGT------CGTGCCC--------------------------------------    [35]
B11        -TGT------CGTTGCA--------------------------------------    [61]
B10        -TGT------CGTTGCA--------------------------------------    [61]
B09        -TGT------CGTTGCA--------------------------------------    [61]
B07        -TGT------CGTTGCA--------------------------------------    [61]
B05        -TGT------CGTTGCA--------------------------------------    [61]
B04        -TGT------CGTTGCA--------------------------------------    [61]
B03        -TGT------CGTTGGG--------------------------------------    [61]
B01        -TGT------CGTTGCA--------------------------------------    [61]
```

```
Th1(1)     GAGAGAAGAA------CAATAATAGAGAGACAACAAAGGGAGGCTA---TTTATT    [151]
Th1(2)     GAGAGAAGAAGAAGAACAATAATAGAGAGACAACAAAGGGAGGCTA---TCTATT    [162]
Th1(3)     GAGAGAAGAA------CAATAATAGAGAGACAACAAAGGGAGGCTA---TTTATT    [151]
T48(27)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T56(W)     ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T53(17)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T53        ------AGAA--------ATG----------------GGGAGGTTTTYGTTCCTT    [60]
T51        ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T49(5)     ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T50        ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T49(4)     ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T47(4)     ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T46(3)     ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
T46        ------AGAA--------ATG----------------GGGAGGTTTTNGTTCCTT    [60]
T45        ------AGAA--------ATG----------------GGGAGGTTTTMGTTCCTT    [60]
T42        ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K277(26)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K273(28)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K263(18)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K261(7)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K260(H)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K108(17)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K103(11)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K091(7)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K077(23)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
K065(7)    ------AGAA--------ATG----------------GGGAGGTTTTMGTTCCTT    [60]
K065(3)    ------AGAA--------ATG----------------GGGAGGTTTTHGTTCCTT    [60]
K01        ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
FRII       ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C210(Q)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C210(O)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C210(M)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C199(V)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C199(S)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C174(W)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C158(F)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C158(E)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C158(C)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C127(E)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C127(D)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C127(C)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C109(S)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C109(R)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C109(O)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C096(L)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C096(K)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C096(H)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C095(21)   ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C095       ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C094       ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C037(B)    ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C037       ------AGAA--------ATG----------------GGGAGGTTTTCGTTCCTT    [60]
C174(U)    ------AGAA--------ATG----------------GGGAGGTTT-CGTTCCTT    [59]
B11        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B10        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B09        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B07        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B05        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B04        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B03        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
B01        ------AAAA--------AACAAA-------ACAAAAGGGAGGTTTTCGTTC--T    [93]
```

```
Th1(1)      GTCCTC--AAGGGTGGTCGGCGATGAGAGTCCCTCTCGCCGGCCCCGCCTCCCGC   [204]
Th1(2)      GTCCTC--AAGGGTGGTCGGCGATGAGAGTCCCTCTCGCCGGCCCCGCCTCCCGC   [215]
Th1(3)      GTCCTC--AAGGGTGGTCGGCGATGAGAGTCCCTCTCGCCGGCCCCGCCTCCCGC   [204]
T48(27)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T56(W)      GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T53(17)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T53         GTYCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T51         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T49(5)      GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T50         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T49(4)      GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T47(4)      GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T46(3)      GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T46         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T45         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
T42         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K277(26)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K273(28)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K263(18)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K261(7)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K260(H)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K108(17)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K103(11)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K091(7)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K077(23)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K065(7)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K065(3)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
K01         GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
FRII        GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C210(Q)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C210(O)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C210(M)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C199(V)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C199(S)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C174(W)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C158(F)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C158(E)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C158(C)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C127(E)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C127(D)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C127(C)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C109(S)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C109(R)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C109(O)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C096(L)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C096(K)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C096(H)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C095(21)    GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C095        GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C094        GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C037(B)     GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C037        GTCCTCGAAAGGG----------------------------TCG---------CCTCCCGC   [84]
C174(U)     GTCCTCGAAAGG-----------------------------TCG---------CCTCCCGC   [82]
B11         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCGTCGAGCCTCCCGC   [139]
B10         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
B09         GTCCTCGAAAGGGTTGTGGCTGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
B07         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
B05         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
B04         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGCCN--GAGCCTCCCGC   [137]
B03         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
B01         GTCCTCGAAAGGGTTGTGGCCGGCCA---------TCGGTCG-CGAGCCTCCCGC   [138]
```

```
Th1(1)      CCGAAGTGTGACTGTGGGGTACCTGTCTTGTCCGGGTTGTCGGGTCTATC-----   [254]
Th1(2)      CCGAAGTGTGACTGTGGGGTACCTGTCTTGTCCGGGTTGTCGGGTCTATC-----   [265]
Th1(3)      CCGAAGTGTGACTGTGGGGTACCTGTCTTGTCCGGGTTGTCGGGTCTATC-----   [254]
T48(27)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T56(W)      CCGAAGCGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T53(17)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T53         CCGAAGTGTGACTGTDGGGTACCTGTHCTGTCCGGGCTGTCGGGTCTYTT-----   [134]
T51         CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T49(5)      CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T50         CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T49(4)      CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T47(4)      CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T46(3)      CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
T46         CCGAAWGTGTGACTGTNGGGTACCTGTNCTGTCCGGGCTGTCGGGTCTCTT-----  [134]
T45         CCGGGAGTGTGACTGTDGGGTACCTGTHCTGTCCGGGCTGTCGGGTCTCTT-----  [134]
T42         CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K277(26)    CCGAAGTGTGACTGTGGGGTAC-TGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [133]
K273(28)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K263(18)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K261(7)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K260(H)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K108(17)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K103(11)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K091(7)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K077(23)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K065(7)     CCGAAGTGTGACTGTNGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K065(3)     CCGAAGTGTGACTGTDGGGTACCTGTYCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
K01         CCGAAGTGTGACTGTGGGGTAC-TGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [133]
FRII        CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C210(Q)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C210(O)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C210(M)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C199(V)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C199(S)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C174(W)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C158(F)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C158(E)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C158(C)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C127(E)     CCGAAGTGTGACTGTGGGGTAC-TGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [133]
C127(D)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C127(C)     CCGAAGTGTGACTGTGGGGTAC-TGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [133]
C109(S)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C109(R)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C109(O)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C096(L)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C096(K)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C096(H)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C095(21)    CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C095        CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C094        CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C037(B)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C037        CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [134]
C174(U)     CCGAAGTGTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTT-----   [132]
B11         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [194]
B10         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
B09         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
B07         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
B05         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
B04         CCGAAGTTTGACTGTGTGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCGCC   [192]
B03         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
B01         CCGAAGTTTGACTGTGGGGTACCTGTCCTGTCCGGGCTGTCGGGTCTCTCCCTCC   [193]
```

```
Th1(1)      --------------GCGC----------------TTTTGAA----CAAAGC--------        [271]
Th1(2)      --------------GCGC----------------TTTTGAA----CAAAGC--------        [282]
Th1(3)      --------------GCGC----------------TTTTGAA----CAAAGC--------        [271]
T48(27)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T56(W)      ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T53(17)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T53         ---------TYTTCCTC--------------TTCTYCT--TCCAAGGG-----AAG           [160]
T51         ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T49(5)      ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T50         ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T49(4)      ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T47(4)      ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
T46(3)      ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
T46         ---------TYTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
T45         ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
T42         ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K277(26)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [159]
K273(28)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K263(18)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
K261(7)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
K260(H)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
K108(17)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
K103(11)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K091(7)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
K077(23)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K065(7)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K065(3)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
K01         ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [159]
FRII        ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
C210(Q)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C210(O)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C210(M)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C199(V)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C199(S)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C174(W)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C158(F)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C158(E)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C158(C)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C127(E)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [159]
C127(D)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C127(C)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [159]
C109(S)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C109(R)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
C109(O)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C096(L)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C096(K)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C096(H)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C095(21)    ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C095        ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C094        ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [160]
C037(B)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
C037        ---------TCTTCCTC--------------TTCTCCT--TCCAAGGGG----AAG           [161]
C174(U)     ---------TCTTCCTC--------------TTCTCCT--TCCAAGGG-----AAG           [158]
B11         ACCACCTCCTCCTCCTC------TTCTTCTTCCTCCTCCTCCTCTGGAGGTGGAG            [243]
B10         ACCACCTCCTCCTCCTCCTCCTCTTCTTCTTCCTCCTCCTCCTCTGGAGGTGGAG            [248]
B09         ACCACCTCCTCCTCCTCCTCCTCTNCVVCVNCCTCCTCCTCCTCTGGAGGTGGAG            [248]
B07         ACC------TCCTCCTCCTCCTCTTCTNCTTCCTCCTCCTCCTCTGGAG-TGG--            [239]
B05         ACCACCTCCTCCTCCTCCTCCTCTTCTTCTTCCTCCTCCTCCTCTGGAGGTGGAG            [248]
B04         ACCTCCTCCTCCTCCTC------TTCTNNTTCCTCCTCCTCCTCTGGTGGTGGAG            [241]
B03         ACCACCTCCTCCTCCTCCTCCTCTNNNNNTTCCTCCTCCTCCTCTGGAGGTGGAG            [248]
B01         ACCACCTCCTCCTCCTCCTCCTCTTCTKCTTCCTCCTCCTCCTCTGGAG-TGGAG            [247]
```

```
Th1(1)     -----------------------------GCGG--------GAGGCCGTCGGCACT    [290]
Th1(2)     -----------------------------GCGG--------GAGGCCGTCGGCACT    [301]
Th1(3)     -----------------------------GCGG--------GAGGCCGTCGGCACT    [290]
T48(27)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T56(W)     --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T53(17)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T53        --------------------------G-----GTGG------CGGAGATCGTHGGCATT    [182]
T51        --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T49(5)     --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T50        --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T49(4)     --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T47(4)     --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
T46(3)     --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
T46        --------------------------G-----GTGG------CGGAGATCGTMGGCATT    [183]
T45        --------------------------G-----GTGG------CGGAGATCGTMGGCATT    [183]
T42        --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K277(26)   --------------------------G-----GTGG------GGG-GATCGTCGGCATT    [180]
K273(28)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K263(18)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
K261(7)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
K260(H)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
K108(17)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
K103(11)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K091(7)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
K077(23)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K065(7)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K065(3)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
K01        --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [181]
FRII       --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
C210(Q)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C210(O)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C210(M)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C199(V)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C199(S)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C174(W)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C158(F)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C158(E)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C158(C)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C127(E)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [181]
C127(D)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C127(C)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [181]
C109(S)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C109(R)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
C109(O)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C096(L)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C096(K)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C096(H)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C095(21)   --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C095       --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C094       --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [182]
C037(B)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
C037       --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [183]
C174(U)    --------------------------G-----GTGG------CGGAGATCGTCGGCATT    [180]
B11        GTGGTTGTGGTGGTGGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [298]
B10        GTGGTTGTGGTGGTGGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [303]
B09        GTGGTTGTGGTGGTGGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [303]
B07        ----TTGTGGTGGTGGT---TGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [287]
B05        GTGGTTGTGGTGGTGGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [303]
B04        GTGGTTG--------------GTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [281]
B03        GTGGTTGTGGTGGTAGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [303]
B01        GTGGTTGTGGTGGTGGTGGTTGGTGGTGGTGGTTTGGTGGGGGACCGTCGGCACT    [302]
```

```
Th1(1)      TCGGCAGGGTGAGCTCCGAGGTGACG-CAAGTGTT---------T-----GTT--    [328]
Th1(2)      TCGGCAGGGTGAGCTCCGAGGTGACG-CAAGTGTT---------T-----GTTGT    [341]
Th1(3)      TCGGCAGGGTGAGCTCCGAGGTGACG-CAAGTGTT---------T-----GTT--    [328]
T48(27)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T56(W)      TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T53(17)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T53        TYGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAANCCCACGCCGAAG-    [235]
T51        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T49(5)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T50        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
T49(4)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
T47(4)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
T46(3)     TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
T46        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [236]
T45        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [236]
T42        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
K277(26)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [233]
K273(28)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
K263(18)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
K261(7)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
K260(H)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
K108(17)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAASCCCACGCCGAAG-    [236]
K103(11)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
K091(7)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
K077(23)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
K065(7)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
K065(3)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
K01        TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [234]
FRII       TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
C210(Q)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C210(O)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C210(M)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C199(V)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C199(S)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C174(W)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C158(F)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C158(E)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C158(C)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C127(E)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [234]
C127(D)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C127(C)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [234]
C109(S)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C109(R)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
C109(O)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C096(L)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C096(K)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C096(H)    TCGGCGGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C095(21)   TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [236]
C095       TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C094       TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [235]
C037(B)    TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAGG    [237]
C037       TCGGCAGGGTGAGCTTGGAGGGGACGGCACGGCCG-CAAAAGCCCACGCCGAAG-    [236]
C174(U)    TCGGCAGGGTGAGCTTGGAGGGGACG-CACGGCCGGCAAAAGCCCACGCCGAAGG    [234]
B11        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAGG    [346]
B10        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAGG    [351]
B09        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAG-    [350]
B07        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAG-    [334]
B05        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAGG    [351]
B04        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAGG    [329]
B03        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAG-    [350]
B01        TCGGCAGGGTGAGCTTGGAGGG-ACGGCACGGCCG-CTAAAACCC-----GAAG-    [349]
```

```
Th1(1)      -TTGTTTTCGAAACAATGGGGAGCTCTCGCTCTCCCA-C--------------TT     [367]
Th1(2)      TTTGTTTTCGARACAATGGGGAGCTCTCGCTCTCCCA-C--------------TT     [381]
Th1(3)      -TTGTTTTCGAAACAATGGGGAGCTCTCGCTCTCCCA-C--------------TT     [367]
T48(27)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
T56(W)      TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTTCCTGTTC-----------TTTTT     [278]
T53(17)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
T53         TGTGTWG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [277]
T51         TGTGCGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
T49(5)      TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
T50         TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
T49(4)      TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [277]
T47(4)      TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
T46(3)      TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [280]
T46         TGTGTTG-TGAAGGGNGGGGAAGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
T45         TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
T42         TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
K277(26)    TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [275]
K273(28)    TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [278]
K263(18)    TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
K261(7)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGCTC-----------TTTTT     [280]
K260(H)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [280]
K108(17)    TGTGTGG-TGAAGGG-GGGGGASCTCTCGTTCCCTGTMC-----------TTTMH     [278]
K103(11)    TGTGTGGGTGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT.    [279]
K091(7)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
K077(23)    TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [279]
K065(7)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [277]
K065(3)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [277]
K01         TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC-----------TTTTT     [276]
FRII        TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC-----------TTTTT     [280]
C210(Q)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C210(O)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C210(M)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C199(V)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C199(S)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C174(W)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C158(F)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C158(E)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [279]
C158(C)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C127(E)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [276]
C127(D)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [277]
C127(C)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [276]
C109(S)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [279]
C109(R)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [280]
C109(O)     TGTGTGG-TGAAGGGGGGGGGGGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C096(L)     TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTYC----------TTTTT      [277]
C096(K)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C096(H)     TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C095(21)    TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [278]
C095        TGTGTGG-TGAAGGG-GGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [277]
C094        TGTGTGG-TGAAGGGGGGGG-AGCTCTCGTTCCCTGTYC----------TTTTT      [277]
C037(B)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT.     [280]
C037        TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [279]
C174(U)     TGTGTGG-TGAAGGGGGGGGGAGCTCTCGTTCCCTGTTC----------TTTTT      [277]
B11         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [390]
B10         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [395]
B09         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [394]
B07         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [378]
B05         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [395]
B04         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [373]
B03         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [394]
B01         TGTG------AAGGG--G---AGCTCTCGTTCCCCGATCCTTTTCAGTTGTTTTT     [393]
```

```
Th1(1)      GCTGGCTTCCCCTCCATGGGCGATGGTTTAGAGAGACGCCCGTCCGTT-------    [415]
Th1(2)      GCTGGCTTCCCCTCCATGGGCGATGGTTTAGAGAGACGCCCGTCCGTT-------    [429]
Th1(3)      GCTGGCTTCCCCTCCATGGGCGATGGTTTARAGAGACGCCCGTCCGTT-------    [415]
T48(27)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTHTCTC    [329]
T56(W)      GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
T53(17)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
T53         GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTYTCTC    [328]
T51         GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
T49(5)      GCGGCCT--CCCTCA--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
T50         GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
T49(4)      GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [326]
T47(4)      GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
T46(3)      GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [331]
T46         GCGGCCT--CCCTCC--GGGCGACGGTTTANNGKGACGCCCGTYCTCTCTCTCTC    [330]
T45         TCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [330]
T42         GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
K277(26)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [326]
K273(28)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
K263(18)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
K261(7)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [329]
K260(H)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [329]
K108(17)    MCRGCCT--CCCTCC--GGGCGACGGTTTAAANANACGCCCGTCCTCTCTCTC--    [327]
K103(11)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
K091(7)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
K077(23)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
K065(7)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [326]
K065(3)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [326]
K01         GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [325]
FRII        GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [331]
C210(Q)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C210(O)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C210(M)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C199(V)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C199(S)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
C174(W)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C158(F)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C158(E)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
C158(C)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C127(E)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [327]
C127(D)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [328]
C127(C)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [327]
C109(S)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [328]
C109(R)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [331]
C109(O)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [327]
C096(L)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [328]
C096(K)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
C096(H)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
C095(21)    GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [329]
C095        GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [328]
C094        GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [326]
C037(B)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [331]
C037        ACGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTCTC    [330]
C174(U)     GCGGCCT--CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTCTC--    [326]
B11         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [436]
B10         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [441]
B09         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [440]
B07         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [424]
B05         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [441]
B04         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [419]
B03         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [440]
B01         GCGGCC---CCCTCC--GGGCGACGGTTTAAAGAGACGCCCGTCCTCTCTC----    [439]
```

```
Th1(1)     ------------------TTTGGGCGGCCGCCCCCTGGCCTATTTATKCTTC--TT    [451]
Th1(2)     ------------------TTTGGGCGGCCGCCCC-TGGCCTATTTATTCTCC--TT    [464]
Th1(3)     ------------------TTTGGGCGGCCGCCCC-TGGCCTATTTATTCTTC--TT    [450]
T48(27)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
T56(W)     GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
T53(17)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
T53        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
T51        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
T49(5)     GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
T50        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
T49(4)     GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [374]
T47(4)     GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
T46(3)     GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [379]
T46        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTTTTTTTT-----    [378]
T45        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTNATTTTT-----    [378]
T42        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
K277(26)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCC-TGGC-TTTTCATTTTT-----    [373]
K273(28)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
K263(18)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
K261(7)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
K260(H)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
K108(17)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
K103(11)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
K091(7)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
K077(23)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
K065(7)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [374]
K065(3)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [374]
K01        GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [373]
FRII       GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [379]
C210(Q)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C210(O)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C210(M)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C199(V)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C199(S)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
C174(W)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C158(F)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TCTTCATTTTT-----    [375]
C158(E)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
C158(C)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C127(E)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C127(D)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
C127(C)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C109(S)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
C109(R)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [379]
C109(O)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [375]
C096(L)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [376]
C096(K)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
C096(H)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
C095(21)   GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [377]
C095       GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTWW-----    [376]
C094       GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [374]
C037(B)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [379]
C037       GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [378]
C174(U)    GCGCCCATCGT-GGCGCGTTGGGTGGCCGCCCCCTGGC-TTTTCATTTTT-----    [374]
B11        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [489]
B10        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [494]
B09        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [493]
B07        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [477]
B05        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [494]
B04        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [472]
B03        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [493]
B01        GCGCTC-TCGCCGGCGCGTAGGGTGGCCGCCCCCTGGC-TTTTTATATTTCCGTT    [492]
```

```
Th1(1)      TCCTATCGCAGCCTTTACACTTGAGAATTACGAATGTGCGAAAGGCG--CCGA-T    [503]
Th1(2)      TCTTATCGCAGCCTTTACACTTGAGAATKACGAATGTGCGARAGGCG--CCGA-T    [516]
Th1(3)      TCCTATCGCAGCCTTTACACTTGARAATTACGAATGTGCGAAAGGCG--CCGA-T    [502]
T48(27)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
T56(W)      TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
T53(17)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
T53        TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
T51        TCTTATCGCAGCCTTTTCATTTGAACATAACCAATGTGAGAAAGGTGGCCCGTCT    [432]
T49(5)      TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [430]
T50        TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
T49(4)      TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [429]
T47(4)      TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
T46(3)      TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [434]
T46        TCTTATCGCAGCCTTTTSNTTTGA-CATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
T45        TNTTATCGCAGCCTTTCCATTTGATCATAACCGATGTGAGANAGGTGGCCCGTCT    [433]
T42        TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
K277(26)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [428]
K273(28)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
K263(18)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
K261(7)     TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [432]
K260(H)     TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [432]
K108(17)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
K103(11)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
K091(7)     TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [431]
K077(23)    TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [431]
K065(7)     WCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [429]
K065(3)     TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGKAAGGTGGCCCGTCT    [429]
K01        TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [428]
FRII       TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [434]
C210(Q)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C210(O)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C210(M)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C199(V)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C199(S)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
C174(W)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C158(F)     TCTTATCTCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C158(E)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
C158(C)     TCTTATCGCAGCCTTTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [430]
C127(E)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C127(D)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
C127(C)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C109(S)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
C109(R)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [434]
C109(O)     TCTTATCGCAGCCTTTTCACTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [430]
C096(L)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
C096(K)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
C096(H)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
C095(21)    TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [432]
C095       WCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [431]
C094       YCTTATCGCAGCCTCTTCATTTGAACATAACTGATGTGAGAAAGGTGGCCCGTCT    [429]
C037(B)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [434]
C037       TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [433]
C174(U)     TCTTATCGCAGCCTTTTCATTTGAACATAACCGATGTGAGAAAGGTGGCCCGTCT    [429]
B11        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [543]
B10        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [548]
B09        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [547]
B07        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [531]
B05        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [548]
B04        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [526]
B03        TCTTATCGCAGCCTTTTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [547]
B01        TCTTATCGCAGCCTTCTCACTTGAACAAAACTGATGCGAGAAAGGTGGCC-GATT    [546]
```

```
Th1(1)     AAGGTCTC--GGTC-GTCGGCTCGCCCGCAAAAGCCTCTTGCGGGGCGGGGTCGC     [555]
Th1(2)     AAGGTCTC--GGTC-GTCGGCTCGCCCGCAAAAGCCTCTTGCGGGGCGGGGTCGC     [568]
Th1(3)     AAGGTCTC--GGTC-GTCGGCTCGCCCGCAAAAGCCTCTTGCGGGGCGGGGTCGC     [554]
T48(27)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
T56(W)     TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
T53(17)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
T53        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
T51        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
T49(5)     TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
T50        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
T49(4)     TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [483]
T47(4)     TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
T46(3)     TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [488]
T46        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
T45        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [487]
T42        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
K277(26)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [482]
K273(28)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
K263(18)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
K261(7)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
K260(H)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
K108(17)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
K103(11)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
K091(7)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
K077(23)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
K065(7)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [483]
K065(3)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [483]
K01        TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [482]
FRII       TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [488]
C210(Q)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C210(O)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C210(M)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C199(V)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C199(S)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
C174(W)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C158(F)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C158(E)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
C158(C)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C127(E)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C127(D)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
C127(C)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C109(S)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
C109(R)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [488]
C109(O)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [484]
C096(L)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
C096(K)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
C096(H)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
C095(21)   TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [486]
C095       TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [485]
C094       TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [483]
C037(B)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [488]
C037       TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [487]
C174(U)    TGGGTCTCCGGGAC-GCCGGCTCGCCCACAAAAGCCTCTTGTGGGTCGGGGTCGC     [483]
B11        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [598]
B10        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [603]
B09        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [602]
B07        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [586]
B05        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [603]
B04        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [581]
B03        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [602]
B01        TTGGTCTCGGGGTCTGCCGGCTCGCCCGCAAAAGCCTCTTGCGGGTCGGGGTCGC     [601]
```

```
Th1(1)      GCGGCCTGGGGCCGCCTTTCGCGCGCGCAAGCGAAACC----------CAAACTT      [600]
Th1(2)      GCGGCCTGGGGCCGCCTTTCGCGCGCGCAAGCGAAACC----------CAAACTT      [613]
Th1(3)      GCGGCCTGGGGCCGCCTTTCGCGCGCGCAAGCGAAACC----------CAAACTT      [599]
T48(27)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
T56(W)      GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
T53(17)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
T53         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
T51         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
T49(5)      GCGACCGGGACCTGCCTTTCGCACGC--AAGAAAAAACGT-------AAAAACTT      [530]
T50         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
T49(4)      GCGACCGGGACCTGCCTTTCGCACGC--AAGAAAAAACGT-------AAAAACTT      [529]
T47(4)      GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
T46(3)      GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [534]
T46         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGN-------NAAAACTT      [532]
T45         GCGACCGGGACCTGCCTTTCGCACGC--AAGAAAAAACGT-------AAAAACTT      [533]
T42         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
K277(26)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [528]
K273(28)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
K263(18)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
K261(7)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
K260(H)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
K108(17)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
K103(11)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
K091(7)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
K077(23)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
K065(7)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [529]
K065(3)     GCGACCGGGACCTGCCTTTCGCACGC--AAGHAAAAACGT-------AAAAACTT      [529]
K01         GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [528]
FRII        GCGACCGGGACCTGCCTTTCGCACGC--AAGAAAAAACGT-------AAAAACTT      [534]
C210(Q)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C210(O)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C210(M)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C199(V)     GCGACCGGGACCTGCCCTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C199(S)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
C174(W)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C158(F)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C158(E)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
C158(C)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C127(E)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C127(D)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
C127(C)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C109(S)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
C109(R)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [534]
C109(O)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [530]
C096(L)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [531]
C096(K)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
C096(H)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
C095(21)    GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [532]
C095        GCGACCGGGACCTGCCTTTCGCACGC--AAGCNAAAACGT-------AAAAACTT      [531]
C094        GCGACCGGGACCTGCCTTTCGCACGC--AAGAAAAAACGT-------AAAAACTT      [529]
C037(B)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [534]
C037        GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [533]
C174(U)     GCGACCGGGACCTGCCTTTCGCACGC--AAGCAAAAACGT-------AAAAACTT      [529]
B11         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GT-AAAAAAAAAAACTT      [649]
B10         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GT-AAAAAAAAAAACTT      [654]
B09         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GTTAAAAAAAAAAACTT      [654]
B07         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GTAAAAAAAAAAAACTT      [638]
B05         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GTAAAAAAAAAAAACTT      [655]
B04         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GT-AAAAAAAAAAACTT      [632]
B03         GTGGCCGGGACCTGCCTCTCGCACGC--AAGCAAAAA-GT-TAAAAAAAAAACTT      [653]
B01         GTGGCCGGGACCTGCCTTTCGCACGC--AAGCAAAAA-GTAAAAAAAAACAAACTT     [653]
```

```
Th1(1)      GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [655]
Th1(2)      GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [668]
Th1(3)      GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [654]
T48(27)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
T56(W)      GTGAGAACAACTT-AGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
T53(17)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
T53         GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
T51         GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
T49(5)      GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
T50         GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
T49(4)      GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [584]
T47(4)      GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
T46(3)      GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [589]
T46         GTGAGAACAACTTTAGGCGGAGGATCACTCGGCTCGTGCGTCGANGAAGAACGCA   [587]
T45         GRTAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [588]
T42         GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
K277(26)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [583]
K273(28)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
K263(18)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
K261(7)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
K260(H)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
K108(17)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
K103(11)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
K091(7)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
K077(23)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
K065(7)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [584]
K065(3)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [584]
K01         GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [583]
FRII        GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [589]
C210(Q)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C210(O)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C210(M)     GTGAGGACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C199(V)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C199(S)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
C174(W)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C158(F)     GTGAGAACAACTTTAGGCGGTGGATÇACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C158(E)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
C158(C)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C127(E)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C127(D)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
C127(C)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C109(S)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
C109(R)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [589]
C109(O)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [585]
C096(L)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
C096(K)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
C096(H)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
C095(21)    GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [587]
C095        GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [586]
C094        GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [584]
C037(B)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [589]
C037        GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [588]
C174(U)     GTGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [584]
B11         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [704]
B10         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [709]
B09         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [709]
B07         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [693]
B05         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [710]
B04         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [687]
B03         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [708]
B01         GAGAGAACAACTTTAGGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCA   [708]
```

```
Th1(1)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [710]
Th1(2)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [723]
Th1(3)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [709]
T48(27)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
T56(W)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
T53(17)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
T53        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
T51        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
T49(5)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
T50        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
T49(4)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [639]
T47(4)     GCCAGCTGCGTGAACTAATGTMAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
T46(3)     GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [644]
T46        GCCAGCNGCGNGAAC-AATGTAAATNGCAGGACAMATTGAACATCGACACATNGA   [641]
T45        GCCAGCTGCGTGAACWAATRTRAATTGCAGRACAAATTGAAMATCGACACNTTGA   [643]
T42        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
K277(26)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [638]
K273(28)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
K263(18)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
K261(7)    GCCAGCTGCGTGAACTAATGTRAATTGCAGGACATATTGAACATCGACACTTTGA   [642]
K260(H)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
K108(17)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
K103(11)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
K091(7)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
K077(23)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
K065(7)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [639]
K065(3)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [639]
K01        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [638]
FRII       GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [644]
C210(Q)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C210(O)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C210(M)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C199(V)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C199(S)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
C174(W)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C158(F)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C158(E)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
C158(C)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C127(E)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C127(D)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
C127(C)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C109(S)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
C109(R)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [644]
C109(O)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [640]
C096(L)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
C096(K)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
C096(H)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
C095(21)   GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [642]
C095       GCCAGCTGCGTGNACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [641]
C094       GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [639]
C037(B)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [644]
C037       GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [643]
C174(U)    GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [639]
B11        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [759]
B10        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [764]
B09        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [764]
B07        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [748]
B05        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [765]
B04        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTCGA   [742]
B03        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [763]
B01        GCCAGCTGCGTGAACTAATGTGAATTGCAGGACACATTGAACATCGACACTTTGA   [763]
```

```
Th1(1)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [765]
Th1(2)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [778]
Th1(3)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [764]
T48(27)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
T56(W)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
T53(17)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
T53        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
T51        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
T49(5)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
T50        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
T49(4)     ACGCATATTGCGGCCAAGGGTCCGCCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [694]
T47(4)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
T46(3)     ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [699]
T46        ANNAATATTGCGGCMAAGGGTCCGACCTTTGGCMACGCCCGTCTGAGGGTCGGCA     [696]
T45        AMGCATATTGCGGCCAAGGGTCCGTCCTTTGGCYACGCCCGTCTGAGGGTCGGCG     [698]
T42        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
K277(26)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [693]
K273(28)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
K263(18)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
K261(7)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
K260(H)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
K108(17)   ACGCNTATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
K103(11)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
K091(7)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
K077(23)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
K065(7)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [694]
K065(3)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [694]
K01        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [693]
FRII       ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [699]
C210(Q)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C210(O)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C210(M)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C199(V)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C199(S)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
C174(W)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C158(F)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C158(E)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
C158(C)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C127(E)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C127(D)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
C127(C)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C109(S)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
C109(R)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [699]
C109(O)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [695]
C096(L)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [696]
C096(K)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
C096(H)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
C095(21)   ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [697]
C095       AMGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTNAGGGTCGGCG     [696]
C094       ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [694]
C037(B)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [699]
C037       ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTKNNGGTCGGCG     [698]
C174(U)    ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [694]
B11        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [814]
B10        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [819]
B09        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [819]
B07        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [803]
B05        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [820]
B04        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [797]
B03        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [818]
B01        ACGCATATTGCGGCCAAGGGTCCGTCCTTTGGCCACGCCCGTCTGAGGGTCGGCG     [818]
```

```
Th1(1)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGTGCAGTGGGCTCTCGCAGCAG    [818]
Th1(2)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGTGCAGTGGGCTCTCGCAGCAG    [831]
Th1(3)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGTGCAGTGGGCTCTCGCAGCAG    [817]
T48(27)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
T56(W)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
T53(17)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
T53         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCWCCTTGGGHTCTCGCAGCCG    [749]
T51         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
T49(5)      AAGTTCTACCCATTCGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
T50         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
T49(4)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [747]
T47(4)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
T46(3)      AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [752]
T46         AAGTTCTACCCAT-CGCYGGAGGCTCTTT-CCGGCTCCTTGGGCTCTCGCAGCCG    [749]
T45         AAGTTCTACCCAT-CGCNGGAGGCTCTTT-CCGGCKCCTTGGGCTCTCGCAGCCG    [751]
T42         AAGTTCTACCCAT-CGCCGGAGKCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
K277(26)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [746]
K273(28)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
K263(18)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
K261(7)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
K260(H)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
K108(17)    AAGTTCTACCCAT-CGCCGGAGGKTKTKT-CCGGCGCCTTGGG-TCTKGCAGCCG    [747]
K103(11)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
K091(7)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
K077(23)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
K065(7)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [747]
K065(3)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [747]
K01         AAGTTCTACCCAT-CGCCGGAGGTYWTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [746]
FRII        AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [752]
C210(Q)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C210(O)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C210(M)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C199(V)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C199(S)     AAGTTCTACCCAT-CGCCGGAGCCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
C174(W)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C158(F)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C158(E)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
C158(C)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C127(E)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C127(D)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
C127(C)     AAGTTCTACCCAT-CGCCGGAG-CTCTTTTCCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C109(S)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
C109(R)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [752]
C109(O)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [748]
C096(L)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [749]
C096(K)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
C096(H)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
C095(21)    AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [750]
C095        AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCNCCTTGGGCTCTCGCAGCCG    [749]
C094        AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [747]
C037(B)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [752]
C037        AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [751]
C174(U)     AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCCTTGGGCTCTCGCAGCCG    [747]
B11         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [867]
B10         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [872]
B09         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [872]
B07         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [856]
B05         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [873]
B04         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [850]
B03         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [871]
B01         AAGTTCTACCCAT-CGCCGGAGGCTCTTT-CCGGCGCGTTGGGCTCTCGCAGCTG    [871]
```

```
Th1(1)      GGCGGGGCGAAGCCTTCGCGGYYTTTCCTGCC----CGGCTACGTGGCTCCAAGT    [869]
Th1(2)      GGCGGGGCGAAGCCTTCGCGGYYTTTCCTGCC----CGGCTACGTGGCTCCAAGT    [882]
Th1(3)      GGCGGGGCGAAGCCTTCGCGGTTTTTCCTGCC----CGGCTACGTGGCTCCAAGT    [868]
T48(27)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
T56(W)      GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
T53(17)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
T53        GGTAAAGGGACG--CTHTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
T51        GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
T49(5)      GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
T50        GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
T49(4)      GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
T47(4)      GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
T46(3)      GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [794]
T46        GGTAAAGGGACG--CTYTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
T45        GGTAAAGGGACG--CTYTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [793]
T42        GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
K277(26)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT·   [788]
K273(28)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
K263(18)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
K261(7)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
K260(H)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
K108(17)    GGKAAAGGGACG--CTGTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
K103(11)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
K091(7)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
K077(23)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
K065(7)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
K065(3)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
K01        GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [788]
FRII       GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [794]
C210(Q)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C210(O)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C210(M)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C199(V)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C199(S)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
C174(W)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C158(F)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C158(E)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
C158(C)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C127(E)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C127(D)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
C127(C)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C109(S)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
C109(R)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [794]
C109(O)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [790]
C096(L)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
C096(K)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
C096(H)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
C095(21)    GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [792]
C095       GGYAAAGGGACG--CTYTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [791]
C094       GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
C037(B)     GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [794]
C037       GGCAAAGGGACG--CTCTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [793]
C174(U)     GGCAAAGGGACG--CTYTC----TTTC--GC-----TCGCTACGTGGCTCCAAGT    [789]
B11        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [915]
B10        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [920]
B09        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [920]
B07        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [904]
B05        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [921]
B04        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCGAGT    [898]
B03        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [919]
B01        GGCAG-GCGAGGTTCTCTC----TTTC--GCCTAGCTCGCTTCGTGGCTCCAAGT    [919]
```

```
Th1(1)      ACAGACCTTCACCTCCCCCCTGCCACCACGAAAGTGAGGGTGGGGGGGGTGTGAC    [924]
Th1(2)      ACAGACCTTCACCTCCCCC-TGCCACCACGAAAGTGAGGGTGGGGGGGGGTGTGAC   [936]
Th1(3)      ACAGACCTTCACCTCCCCCCTGCCACCACGAAAGTGAGGGTGGGGGGGGTGTGAC    [923]
T48(27)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
T56(W)      TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [837]
T53(17)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
T53         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTRAC    [838]
T51         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
T49(5)      TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
T50         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [837]
T49(4)      TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [834]
T47(4)      TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
T46(3)      TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [840]
T46         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TWAAAAA-GGGTGAC    [837]
T45         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TWAAAAA-GGGTGAC    [839]
T42         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
K277(26)    TCAGACCTTCACCCTYGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [834]
K273(28)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
K263(18)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [837]
K261(7)     TCAGACCTTCACCCTCGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [838]
K260(H)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [837]
K108(17)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TGAAAA--GGGTGAC    [834]
K103(11)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
K091(7)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
K077(23)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
K065(7)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAA---GGGTGAC    [833]
K065(3)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [834]
K01         TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [833]
FRII        TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [839]
C210(Q)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C210(O)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C210(M)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C199(V)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C199(S)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [838]
C174(W)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [835]
C158(F)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
C158(E)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [836]
C158(C)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C127(E)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [837]
C127(D)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [838]
C127(C)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [837]
C109(S)     TCAGACCTTCACCCTTGGCGC--CT-CCAG---GTGCA--TAAAAAA-GGGTGAC    [837]
C109(R)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [840]
C109(O)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [836]
C096(L)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [838]
C096(K)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
C096(H)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
C095(21)    TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [839]
C095        TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAAAGGGTGAC    [838]
C094        TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAA--GGGTGAC    [834]
C037(B)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [840]
C037        TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [839]
C174(U)     TCAGACCTTCACCCTTGGCGC--CT-CCCG---GTGCA--TAAAAAA-GGGTGAC    [835]
B11         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [960]
B10         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [965]
B09         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [965]
B07         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [949]
B05         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [966]
B04         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [943]
B03         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [964]
B01         TCAGACCTTCACCACGGACACACCTGCCCG---GTG----TAAAC---TGGTGAC    [964]
```

```
Th1(1)     TTGCCGAT----------TCAAAGACCTCGCACTTTT------------------  [951]
Th1(2)     TTGCCGAT----------TCAAAGACCTCGCACTTTT------------------  [963]
Th1(3)     TTGCCGAT----------TCAAAGACCTCGCACTTTT------------------  [950]
T48(27)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
T56(W)     TTGCCGAT---------AAAAAAAAAA-G-AC-----------------------  [859]
T53(17)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
T53        TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [859]
T51        TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
T49(5)     TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
T50        TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [857]
T49(4)     TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [854]
T47(4)     TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [857]
T46(3)     TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [862]
T46        TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [859]
T45        TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
T42        TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
K277(26)   TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [855]
K273(28)   TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
K263(18)   TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [857]
K261(7)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [858]
K260(H)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [858]
K108(17)   TTGCCGAT---------AAAAAAAA----G-AC----------------------  [853]
K103(11)   TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
K091(7)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [857]
K077(23)   TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
K065(7)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [854]
K065(3)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [855]
K01        TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [854]
FRII       TTGCCGAT---------AAAAAAAA----G-AC----------------------  [858]
C210(Q)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
C210(O)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
C210(M)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
C199(V)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [857]
C199(S)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [858]
C174(W)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [855]
C158(F)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [857]
C158(E)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [856]
C158(C)    TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [858]
C127(E)    TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [857]
C127(D)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [859]
C127(C)    TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [858]
C109(S)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [858]
C109(R)    TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [862]
C109(O)    TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [858]
C096(L)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [859]
C096(K)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
C096(H)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [860]
C095(21)   TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [859]
C095       TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [858]
C094       TTGCCGAT---------AAAAAAAAAA---G-AC---------------------  [854]
C037(B)    TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [862]
C037       TTGCCGAT---------AAAAAAAAAAA-G-AC----------------------  [861]
C174(U)    TTGCCGAT---------AAAAAAAAAA--G-AC----------------------  [856]
B11        TTGCCGATTTCGGAGAGAAGAAAAAAAAAG-ACCCGCACTTTTCTTTTGTTCCCC  [1014]
B10        TTGCCGATTTCGGAGAGAAGAAAAAAAA-G-ACCCGCACTTTTCTTTTGTTCCCC  [1018]
B09        TTGCCGATTTCGGAGAGAAGAAAAAAAAAG-ACCCGCACTTTTCTTTTGTTCCCC  [1019]
B07        TTGCCGATTTCGGAGAGAAGAAAAAAAA-G-ACCCGCACTTTTCTTTTGTTCCCC  [1002]
B05        TTGCCGATTTCGGAGAGAAGAAAAAAAAAG-ACCCGCACTTTTCTTTTGTTCCCC  [1020]
B04        TTGCCGATTTCGGAGAGAAGAAAAAAAA-G-ACCCGCACTTTTCTTTTGCTCCCC  [996]
B03        TTGCCGATTTCGGAGAGAAGAAAAAAAAAG-ACCCGCACTTTTCTTTTGTTCCCC  [1018]
B01        TTGCCGATTTCGGAGAGAAGAAAAAAAAAG-ACCCGCACTTTTCTTTTGTTCCCC  [1018]
```

```
Th1(1)      CTTCTTTTTTCCCCCCTTGACCCTCTCGGAACAGAGAGGGGCAAAAAAGGTGTTG   [1006]
Th1(2)      CTTCTYTTTTCCCCCCTTGACCCTCTCGGAACAGAGAGGGGCAAAAAAGGTGTTG   [1018]
Th1(3)      CTTCTTTTTTCCCCCCTTGACCCTCTCGGAACAGAGAGGGGCAAAAAAGGTGTTG   [1005]
T48(27)     --------------GC--GCCGC----------------GT------------G   [870]
T56(W)      --------------GC--GCCGC----------------GT------------G   [869]
T53(17)     --------------GC--GCCGC----------------GT------------G   [870]
T53         ---------------C--GCCGC----------------GT------------G   [868]
T51         --------------GC--GCCGC----------------GT------------G   [870]
T49(5)      --------------GCGCGCCGC----------------GT------------G   [868]
T50         --------------GC--GCCGC----------------GT------------G   [867]
T49(4)      --------------GCGCGCCGC----------------GT------------G   [866]
T47(4)      --------------GC--GCCGC----------------GT------------G   [867]
T46(3)      --------------GC--GCCGC----------------GT------------G   [872]
T46         --------------GC--GCCGC----------------GT------------G   [869]
T45         --------------GC--GCCGC----------------GT------------G   [870]
T42         --------------GC--GCCGC----------------GT------------G   [870]
K277(26)    --------------GC--GCCGC----------------GT------------G   [865]
K273(28)    --------------GC--GCCGC----------------GT------------G   [870]
K263(18)    --------------GC--GCCGC----------------GT------------G   [867]
K261(7)     --------------GC--GCCGC----------------GT------------G   [868]
K260(H)     --------------GC--GCCGC----------------GT------------G   [868]
K108(17)    --------------GCGCGCCGC----------------GT------------G   [865]
K103(11)    --------------GCGCGCCGC----------------GT------------G   [868]
K091(7)     --------------GC--GCCGC----------------GT------------G   [867]
K077(23)    --------------GCGCGCCGC----------------GT------------G   [868]
K065(7)     --------------GC-CGCCGC----------------GT------------G   [865]
K065(3)     --------------GCGCGCCGC----------------GT------------G   [867]
K01         --------------GC--GCCGC----------------GT------------G   [864]
FRII        --------------GCGCGCCGC----------------GT------------G   [870]
C210(Q)     --------------GC--GCCGC----------------GT------------G   [866]
C210(O)     --------------GC--GCCGC----------------GT------------G   [866]
C210(M)     --------------GC--GCCGC----------------GT------------G   [866]
C199(V)     --------------GC--GCCGC----------------GT------------G   [867]
C199(S)     --------------GC--GCCGC----------------GT------------G   [868]
C174(W)     --------------GCGCGCCGC----------------GT------------G   [867]
C158(F)     --------------GC--GCCGC----------------GT------------G   [867]
C158(E)     --------------GCGCGCCGC----------------GT------------G   [868]
C158(C)     --------------GC--GCCGC----------------GT------------G   [868]
C127(E)     --------------GC--GCCGC----------------GT------------G   [867]
C127(D)     --------------GC--GCCGC----------------GT------------G   [869]
C127(C)     --------------GC--GCCGC----------------GT------------G   [868]
C109(S)     --------------GC--GCCGC----------------GT------------G   [868]
C109(R)     --------------GC--GCCGC----------------GT------------G   [872]
C109(O)     --------------GC--GCCGC----------------GT------------G   [868]
C096(L)     --------------GC--GCCGC----------------GT------------G   [869]
C096(K)     --------------GC--GCCGC----------------GT------------G   [870]
C096(H)     --------------GC--GCCGC----------------GT------------G   [870]
C095(21)    --------------GC--GCCGC----------------GT------------G   [869]
C095        --------------GC--GCCGC----------------GT------------G   [868]
C094        --------------GCGCGCCGC----------------GT------------G   [866]
C037(B)     --------------GC--GCCGC----------------GT------------G   [872]
C037        --------------GC--GCCGC----------------GT------------G   [871]
C174(U)     --------------GC--GCCGC----------------GT------------G   [866]
B11         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1069]
B10         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1073]
B09         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1074]
B07         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1057]
B05         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1075]
B04         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1051]
B03         CTACTTGTTCCCTTGCACACCACCCGACCCTACCTCTCGGTACAGAAAGGGATGG   [1073]
B01         CTACTTGTTCCCTTGCACACCACCCAACCCTACCTCTCGGTACAGAAAGGGATGG   [1073]
```

```
Th1(1)      TG-------------------------------------------CGTGTGTAGA    [1018]
Th1(2)      TG-------------------------------------------CGTGTGTAGA    [1030]
Th1(3)      TG-------------------------------------------CGTGTGTAGA    [1017]
T48(27)     T-------------------------------------------GCGTGTTGA     [880]
T56(W)      T-------------------------------------------GCGTGTTGA     [879]
T53(17)     T-------------------------------------------GCGTGTTGA     [880]
T53         T-------------------------------------------GCGTGTTGA     [878]
T51         T-------------------------------------------GCGTGTTGA     [880]
T49(5)      T-------------------------------------------GCGTGTTGA     [878]
T50         T-------------------------------------------GCGTGTTGA     [877]
T49(4)      T-------------------------------------------GCGTGTTGA     [876]
T47(4)      T-------------------------------------------GCGTGTTGA     [877]
T46(3)      T-------------------------------------------GCGTGTTGA     [882]
T46         T-------------------------------------------GCGTGTTGA     [879]
T45         T-------------------------------------------GCGTGTTGA     [880]
T42         T-------------------------------------------GCGTGTTGA     [880]
K277(26)    T-------------------------------------------GCGTGTTGA     [875]
K273(28)    T-------------------------------------------GCGTGTTGA     [880]
K263(18)    T-------------------------------------------GCGTGTTGA     [877]
K261(7)     T-------------------------------------------GCGTGTTGA     [878]
K260(H)     T-------------------------------------------GCGTGTTGA     [878]
K108(17)    T-------------------------------------------GCGTGTTGA     [875]
K103(11)    T-------------------------------------------GCGTGTTGA     [878]
K091(7)     T-------------------------------------------GCGTGTTGA     [877]
K077(23)    T-------------------------------------------GCGTGTTGA     [878]
K065(7)     T-------------------------------------------GCGTGTTGA     [875]
K065(3)     T-------------------------------------------GCGTGTTGA     [877]
K01         T-------------------------------------------GCGTGTTGA     [874]
FRII        T-------------------------------------------GCGTGTTGA     [880]
C210(Q)     T-------------------------------------------GCGTGTTGA     [876]
C210(O)     T-------------------------------------------GCGTGTTGA     [876]
C210(M)     T-------------------------------------------GCGTGTTGA     [876]
C199(V)     T-------------------------------------------GCGTGTTGA     [877]
C199(S)     T-------------------------------------------GCGTGTTGA     [878]
C174(W)     T-------------------------------------------GCGTGTTGA     [877]
C158(F)     T-------------------------------------------GCGTGTTGA     [877]
C158(E)     T-------------------------------------------GCGTGTTGA     [878]
C158(C)     T-------------------------------------------GCGTGTTGA     [878]
C127(E)     T-------------------------------------------GCGTRTTGA     [877]
C127(D)     T-------------------------------------------GCGTGTTGA     [879]
C127(C)     T-------------------------------------------GCGTGTTGA     [878]
C109(S)     T-------------------------------------------GCGTGTTGA     [878]
C109(R)     T-------------------------------------------GCGTGTTGA     [882]
C109(O)     T-------------------------------------------GCGTGTTGA     [878]
C096(L)     T-------------------------------------------GCGTGTTGA     [879]
C096(K)     T-------------------------------------------GCGTGTTGA     [880]
C096(H)     T-------------------------------------------GCGTGTTGA     [880]
C095(21)    T-------------------------------------------GCGTGTTGA     [879]
C095        T-------------------------------------------GCGTGTTGA     [878]
C094        T-------------------------------------------GCGTGTTGA     [876]
C037(B)     TT------------------------------------------GCGTGTTGA     [883]
C037        T-------------------------------------------GCGTGTTGA     [881]
C174(U)     T-------------------------------------------GCGTGTTGA     [876]
B11         TGAGGAGGGGGGGG-AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1123]
B10         TGAGGAGGGGGGGG-AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1127]
B09         TGAGGAGGGGGGGGGGAGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA   [1129]
B07         TGAGGAGGGGGGG--AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1110]
B05         TGAGGAGGGGGGG--AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1128]
B04         TGAGGAGGGGGGG--AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1104]
B03         TGAGGAGGGGGSGG-AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1127]
B01         TGAGGAGGGGGGG--AGGAGGAGGGAGGAAAAGGTAATAGAGGTGCGCGTGTCGA    [1126]
```

```
Th1(1)     GCCCCTTCCGTCCTACCCGC-GTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1072]
Th1(2)     GCCCCTTCCGTCCT-CCCGC-GTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1083]
Th1(3)     GCCCCTTCCGTCCTACCCGC-GTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1071]
T48(27)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
T56(W)     GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [933]
T53(17)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
T53        GCHCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
T51        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
T49(5)     GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
T50        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
T49(4)     GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
T47(4)     GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
T46(3)     GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [936]
T46        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [933]
T45        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
T42        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
K277(26)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [929]
K273(28)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
K263(18)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
K261(7)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
K260(H)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
K108(17)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [929]
K103(11)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
K091(7)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
K077(23)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
K065(7)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [929]
K065(3)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
K01        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [928]
FRII       GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
C210(Q)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
C210(O)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
C210(M)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
C199(V)    GCCC-TTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
C199(S)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C174(W)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
C158(F)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
C158(E)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C158(C)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C127(E)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [931]
C127(D)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [933]
C127(C)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C109(S)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C109(R)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [936]
C109(O)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C096(L)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [933]
C096(K)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
C096(H)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [934]
C095(21)   GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [933]
C095       GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [932]
C094       GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
C037(B)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [937]
C037       GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [935]
C174(U)    GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGACATTCCGCAT     [930]
B11        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1177]
B10        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGAACATGGCTTGGATATTCCGCAT     [1181]
B09        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1183]
B07        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1164]
B05        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1182]
B04        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1158]
B03        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1181]
B01        GCCCCTTCCGTCCT-CCTGCCGTGGAGCGAAGGACATGGCTTGGATATTCCGCAT     [1180]
```

```
Th1(1)      GCCTACGCGCCACCCGAAGGTGAGTACACACGCGCATCTTACCCGGGGCGGAGAA   [1127]
Th1(2)      GCCTACGCGCCACCCGAAGGTGAGTACACACGCGCATCTTACCCGGGGCGGAGAA   [1138]
Th1(3)      GCCTACGCGCCACCCGAAGGTGAGTACACACGCGCATCTTACCCGGGGCGGAGAA   [1126]
T48(27)     GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
T56(W)      GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [975]
T53(17)   ··GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
T53         GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGMACAC    [974]
T51         GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
T49(5)      GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
T50         GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
T49(4)      GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
T47(4)      GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
T46(3)      GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [977]
T46         GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [974]
T45         GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [975]
T42         GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
K277(26)    GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [970]
K273(28)    GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
K263(18)    GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
K261(7)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
K260(H)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
K108(17)    GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [971]
K103(11)    GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [974]
K091(7)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
K077(23)    GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
K065(7)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [970]
K065(3)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
K01         GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [969]
FRII        GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [975]
C210(Q)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
C210(O)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
C210(M)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
C199(V)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
C199(S)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
C174(W)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
C158(F)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
C158(E)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
C158(C)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
C127(E)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [972]
C127(D)     GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [975]
C127(C)     GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [974]
C109(S)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
C109(R)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [977]
C109(O)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [973]
C096(L)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [974]
C096(K)     GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
C096(H)     GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [976]
C095(21)    GCCTGCGCACCACCCGAAGG---G---------GATTTTTTTTGCGCGCGCACAC   [976]
C095        GCCTGCGCACCACCCGAAGG---G---------GATTTTTT-GCGCGCGCACAC    [974]
C094        GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
C037(B)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [978]
C037        GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [976]
C174(U)     GCCTGCGCACCACCCGAAGG---G---------GATTTTT--GCGCGCGCACAC    [971]
B11         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TTTTTT----   [1217]
B10         GCCTACGCACCACCGGAAGG---GT-----CTCGATTCTTTT---TTTTTT----   [1221]
B09         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TTTTTTT---   [1224]
B07         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TTTTT-----   [1203]
B05         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TTTTTTTCTC   [1226]
B04         GCCTACGCACCACCGGAAGG---GT-----CTCGATTCTTTT---TTTTTT----   [1198]
B03         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TCCCCTC-TC   [1224]
B01         GCCTACGCACCACCTGAAGG---GT-----CTCGATTCTTTT---TTTTTT----   [1220]
```

```
Th1(1)     TAAAAAGAAGAAGAAAAAGAGC-GTCTYT-----CTYATTTCCGACCTCAGATCG   [1176]
Th1(2)     TAAAAAGAAGAAGAAAAAGAGTTGTCTYT-----CTTATTTCCGACCTCAGATCG   [1188]
Th1(3)     TAAAAAGAAGAAGAAAAAGAGC-GTCTYT-----CTTATTTCCGACCTCAGATCG   [1175]
T48(27)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
T56(W)     --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1013]
T53(17)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
T53        --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1012]
T51        --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
T49(5)     --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1013]
T50        --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1011]
T49(4)     --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1011]
T47(4)     --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
T46(3)     --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1016]
T46        --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1013]
T45        --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1014]
T42        --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
K277(26)   --------------ACACGRGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1009]
K273(28)   --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
K263(18)   --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
K261(7)    --------------ACACGGGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
K260(H)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
K108(17)   --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1009]
K103(11)   --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1013]
K091(7)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
K077(23)   --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
K065(7)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1009]
K065(3)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
K01        --------------ACACGGGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1008]
FRII       --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1015]
C210(Q)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1010]
C210(O)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1010]
C210(M)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1010]
C199(V)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1010]
C199(S)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
C174(W)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
C158(F)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1011]
C158(E)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
C158(C)    --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1013]
C127(E)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1010]
C127(D)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1013]
C127(C)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1012]
C109(S)    --------------ACACGAGC-GTCTTTTTTT-CTYATTTCCGACCTCAGATCG   [1012]
C109(R)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1016]
C109(O)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1012]
C096(L)    --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1014]
C096(K)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
C096(H)    --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
C095(21)   --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1014]
C095       --------------ACACGAGC-GTCTTTTTT--CTCATTTCCGACCTCAGATCG   [1012]
C094       --------------ACACGAGC-GTCTTTTTTTTCTCATTTCCGACCTCAGATCG   [1011]
C037(B)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1017]
C037       --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1015]
C174(U)    --------------ACACGAGC-GTCTTTTTTT-CTCATTTCCGACCTCAGATCG   [1010]
B11        ---------------C--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1248]
B10        ---------------C--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1252]
B09        ---------------C--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1255]
B07        -----------------------TCTCTCT-CTCATTTCCGACCTCAGATCG   [1232]
B05        ---------------TC--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1258]
B04        ---------------C--------TCTCT-----CTCATTTCCGACCTCAGATCG   [1225]
B03        ---------------TC--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1256]
B01        ---------------C--------TCTCTCTCT-CTCATTTCCGACCTCAGATCG   [1251]
```

```
Th1(1)        GACGAGATT     [1185]
Th1(2)        GACGAGATT     [1197]
Th1(3)        GACGAGATT     [1184]
T48(27)       GACGAGATT     [1023]
T56(W)        GACGAGATT     [1022]
T53(17)       GACGAGATT     [1023]
T53           GACGAGATT     [1021]
T51           GACGAGATT     [1023]
T49(5)        GACGAGATT     [1022]
T50           GACGAGATT     [1020]
T49(4)        GACGAGATT     [1020]
T47(4)        GACGAGATT     [1020]
T46(3)        GACGAGATT     [1025]
T46           GACGAGATT     [1022]
T45           GACGAGATT     [1023]
T42           GACGAGATT     [1023]
K277(26)      GACGAGATT     [1018]
K273(28)      GACGAGATT     [1023]
K263(18)      GACGAGATT     [1020]
K261(7)       GACGAGATT     [1021]
K260(H)       GACGAGATT     [1021]
K108(17)      GACGAGATT     [1018]
K103(11)      GACGAGATT     [1022]
K091(7)       GACGAGATT     [1020]
K077(23)      GACGAGATT     [1021]
K065(7)       GACGAGATT     [1018]
K065(3)       GACGAGATT     [1020]
K01           GACGAGATT     [1017]
FRII          GACGAGATT     [1024]
C210(Q)       GACGAGATT     [1019]
C210(O)       GACGAGATT     [1019]
C210(M)       GACGAGATT     [1019]
C199(V)       GACGAGATT     [1019]
C199(S)       GACGAGATT     [1021]
C174(W)       GACGAGATT     [1020]
C158(F)       GACGAGATT     [1020]
C158(E)       GACGAGATT     [1021]
C158(C)       GACGAGATT     [1022]
C127(E)       GACGAGATT     [1019]
C127(D)       GACGAGATT     [1022]
C127(C)       GACGAGATT     [1021]
C109(S)       GACGAGATT     [1021]
C109(R)       GACGAGATT     [1025]
C109(O)       GACGAGATT     [1021]
C096(L)       GACGAGATT     [1023]
C096(K)       GACGAGATT     [1023]
C096(H)       GACGAGATT     [1023]
C095(21)      GACGAGATT     [1023]
C095          GACGAGATT     [1021]
C094          GACGAGATT     [1020]
C037(B)       GACGAGATT     [1026]
C037          GACGAGATT     [1024]
C174(U)       GACGAGATT     [1019]
B11           GACGAGATT     [1257]
B10           GACGAGATT     [1261]
B09           GACGAGATT     [1264]
B07           GACGAGATT     [1241]
B05           GACGAGATT     [1267]
B04           GACGAGATT     [1234]
B03           GACGAGATT     [1265]
B01           GACGAGATT     [1260]
```

LITERATURE CITED

Altschul, S.F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Anderson, T. J., and R. D. Adlard. 1994. Nucleotide sequence of a rDNA internal transcribed spacer supports synonymy of *Saccostrea commercialis* and *S. glomerata*. J. Molluscan Stud. 60: 196-197.

Avise, J. 1994. Molecular Markers, Natural History, and Evolution. Chapman and Hall.

Baker, A. J. and A. Moeed. 1987. Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). Evolution. 41(3): 525-535.

Barrett, S. C. H. and B. J. Richardson. 1986. Genetic attributes of invading species. in: ecology of biological invasions. Cambridge University Press. pp. 21-33.

Caporale, D. A., B. F. Beal, R. Roxby, and R. J. Van Beneden. 1997. Population structure of *Mya arenaria* along the New England coastline. Molecular Marine Biology and Biotechnology. 6(1): 33-39.

Carlton, J. T. 1985. Transoceanic and interoceanic dispersal of coastal marine organisms: the biology of ballast water transport. Oceanographic Marine Biology Review. 23:313-371.

Carlton, J. T. 1999. Molluscan invasions in marine and estuarine communities. Malacologica. 41(2).

Carlton, J. T. and J. B. Geller. 1993. Ecological roulette: the global transport of nonindigenous marine organisms. Science. 261: 78-82.

Collins, T. M., K. Frazer, A. R. Palmer, G. J. Vermeij, and W. M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution. 50(6): 2287-2304.

Chung, E. Y., S. Y. Kim, and Y. G. Kim. 1993. Reproductive ecology of the purple shell, *Rapana venosa* (Gastropod: Muricidae), with special reference to the reproductive cycle, deposition of egg capsules, and hatching of larvae. Korean Journal of Malacology. 9(2):1-15.

Chukchin, V. 1984. Ecology of the gastropod molluscs of the Black Sea. Academy of Sciences, USSR, Kiev Kaukova Dumka, 175 pp. In Russian.

Drapkin, E. 1963. Effect of *Rapana bezoar* Linne' (Mollusca, Muricidae) on the Black Sea fauna. SRR, Doklady Akademii Nauk 151(3): 700-703.

Duda, T. F. 1994. Genetic population structure of the recently introduced Asian clam, *Potamocorbula amurensis*, in San Francisco Bay. Marine Biology. 119:235-241.

Ehrlich, P. R. 1986. Which animal will invade?, in H. A. Mooney and J. A. Drake (eds) Ecology of Biological Invasions of North America and Hawaii. Ecological Studies 58, Springer-Verlag, New York, USA. pp. 79-95.

Ehrlich, P. R. 1989. Attributes of invaders and the invading process: vertebrates, in J. A. Drake, H. A. Mooney, F. di Castri, R. H. Groves, F. J. Kruger, M. Rejmanek, and M. Williamson (eds). Biological Invasions, a Global Perspective. SCOPE 37. John Wiley and Sons, Chichester, UK. pp. 315-28.

Elton, C. S. 1958. The Ecology of Invasions in Animals and Plants. Methuen, London, UK.

Goggin, C. L. 1994. Variation in the two internal transcribed spacers and 5.8S ribosomal RNA from five isolates of the marine parasite *Perkinsus* (Protista, Apicomplexa). Molecular and Biochemical Parasitology 65 (1) 1994. 179-182.

Grant, W. S. and M. I. Cherry. 1985. *Mytilus galloprovincialis* Lmk. in southern Africa. J. Exp. Mar. Biol. Ecol. 90: 179-190.

Harding, J. M. and R. Mann. 1999. Observations on the biology of the veined rapa whelk, *Rapana venosa* (Valencienne, 1846) in the Chesapeake Bay. Journal of Shellfish Research. 18(1): 9-17.

Harrison, R. G. 1998. "Linking Evolutionary Pattern and Process: the relevance of species concepts for the study of speciation." Chapter 2 *in* Endless Forms: Species and Speciation. Eds. D. J. Howard and S.H. Berlocher. Oxford, Oxford University Press. pp 19-31.

Herbert, P. D. N., B. W. Muncaster, G. L. Makie. 1989. Ecological and genetic studies of *Dreissena polymorpha* (Pallas): a new mollusc in the Great Lakes. Can. J. Fish. Aquat. Sci. 46: 1587-1591.

Hillis, D. M., C. Moritz, and B. K. Mable. 1996. Applications of Molecular Systematics, Chapter 12 *in* Molecular Systematics, second edition. ed. D. M. Hillis, C. Moritz, and B. K. Maple. Sinauer Associates, Inc.

Hwang, D. F., S. C. Lu, and S. S. Jeng. 1991. Occurrence of tetrodotoxin in the gastropods *Rapana rapiformis* and *R. venosa* venosa. Marine Biology 111(1): 65-69.

Johnson, M. S. 1988. Founder effects and geographic variation in the land snail *Theba pisana*. Heredity. Lond. 61: 133-142.

Kaufmann, J., S. Olson, and J. Panagrossi. 1994. "MacVector," v.7.0 Kodak Scientific Imaging Systems, New Haven.

Kenchinton, E. L., K. R. Freeman, S. P. Macquarrie, and S. M. C. Robinson. 2000. Use of DNA markers to detect differential larval settlement patterns of *Mytilus edulis* and *M. trossulus*. Journal of Shellfish Research. 19(1): 640.

Kinzelbach, R. 1986. New records of Thomas' rapa whelk, *Rapana thomisiana*, from the Black Sea and Marmara Sea. Zoology in the Middle East. 1: 122-124.

Kira, Tetsuaki. 1961. Coloured Illustrations of the Shells of Japan: enlarged and revised edition. Hoikusha. Osaka, Japan. 239 pp. In Japanese.

Kirby, R. R., B. L. Bayne, and R. J. Berry. 1994. Phenotypic variation along a cline in allozyme and karyotype frequencies, and its relationship with habitat, in the dog-whelk, *Nucella lapillus*, L. Biological Journal of the Linnean Society. 53:255-275.

Knight, A. J., R. N. Hughes, R. E. Ward. 1987. A striking example of the founder effect in the mollusc *Littorina saxatilis*. Biol. J. Linn. Soc. 32: 417-426.

Konsoulova, T. H. 1992. Mussel *Mytilus galloprovincialis* L a m. (Bilvalvia) natural resources along the northern Bulgarian Black sea coast in relation to *Rapana thomasiana* Grosse (Gastropoda) distribution. Trudove na Instituta po Okeanologiia. 1: 104-109.

Koutstoubas, D. and E. Voultsiadou-Koukoura E. 1990. the occurrence of *Rapana venosa* (Valenciennes, 1846)(Gastropoda, Thaididae) in the Aegean Sea. Boll. Malacologico, Milano26(10-12).

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Bioinformatics (in press).

Lutton, K., D. Walker, and D. Blair. 1992. Nuclear rDNA ITS sequence variation in the trematode genus *Echinostoma*: an aid to establishing relationships within the 37-collar spin group. Parasitology. 111: 609-615.

Makarov, Y. N., and V. V. Murina. 1998. Crabs the newcomers in the Black Sea. Kraby-vselentsy v Chernom more. In Russian. Priroda. 10: 39-42.

Mann, R. and Harding, J. M. 2000. Invasion of the North American Atlantic coast by a large predatory Asian mollusc. Biological Invasions. 2:7-22.

Marsden, J. E., A. P. Spidle, and B. May. 1996. Review of genetic studies of *Dreissena* spp. Am. Zool. 36(3): 259-270.

McCarthy, M. 1992. Giant whelks threaten British oysters. London Times. 16. August 1992.

McElroy, D., P. Moran, E. Bermingham and I. Kornfield. 1992. REAP: An integrated environment for the manipulation and phylogenetic analysis of restriction data. J. Hered. 83:157-158.

McLeod, M. J. 1986. Electrophoretic variation in North American *Corbicula*. Am. Malac. Bull. 2:125-135.

Meehan, B. W., J. T. Carlton, R. Wenne. 1989. Genetic affinities of the bivalve *Macoma balthica* from the Pacific coast of North America: evidence for recent introduction and historical distribution. Mar Biol. 102: 235-241.

Merritt, T. J. S., L. Shi, M.C. Chase, M. A. Rex, R. J. Etter, and J. Quattro. 1998. Universal cytochrome *b* primers facilitate intraspecific studies in molluscan taxa. Molecular Marine Biology and Biotechnology. 7(1): 7-11.

Morton, B. 1994. Prey preference and method of attack by *Rapana bezoar* (Gastropoda: Muricidae) from Hong Kong. *In*: Morton B (ed.) The Malacofauna and Southern China III. Hong Kong University Press. pp. 309-325.

National Research Council. 1996. Stemming the Tide: controlling introductions of Nonindigenous species by ships' ballast water. Committee on Ships' Ballast Operations, Marine Board, Commission on Engineering and Technical Systems, National Research Council. National Academy of Sciences. 141 pp.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution. 29:1-10.

Palumbi, S. R. 1996. Nucleic acids II: the polymerase chain reaction. *in* Molecular Systematics, second edition. ed. D. M. Hillis, C. Moritz, and B. K. Maple. Sinauer Associates, Inc.

Reece, K. S., D. Bushek, K. L. Hudson, and J. E. Graves. 2001. Geographic distribution of *Perkinsus marinus* genetic strains throughout the US Atlantic and Gulf Coasts. Marine Biology.

Remigio, E. A., and D. Blair. 1997. Relationships among problematic North American stagnicoline snails (Pulmonata: Lymnaeidae) reinvestigated using nuclear

ribosomal DNA internal transcribed spacer sequences. Can. J. Zool. 75: 1540-1545.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution. 43: 223-225.

Roff, D. A. and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. 6: 539-545.

Rozas, J. and R. Rozas. 1999. DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics. 15(2): 174-175.

Ruiz, G. M., J. T. Carlton, E. D. Grosholz, and A. H. Hines. 1997. Global invasions of marine and estuarine habitats by non-indigenous species: mechanisms, extent, and consequences. American Zoologist. 37(6): 621-632.

Sahin T. 1997. Investigations on some biological characteristics of the sea snail *Rapana venosa* (Valenciennes 1846) population in the Eastern Black Sea. Turkish Journal of Zoology. 21(4):461-466.

Schizas, N. V., G. T. Street, B. C. Coull, G. T. Chandler, and J. M. Quattro. 1999. Molecular population structure of the marine benthic copepod *Microarthridion littorale* along the southeastern and Gulf coasts of the USA. Marine Biology. 135:399-405.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Selander R. K. and H. Ochman. 1983. The genetic structure of populations as illustrated by molluscs. Isozymes. 10: 93-123.

Smith, M. H. J. Britton, P. Burke, R. K. Chesser, M. K. Smith, and J. Hagen. 1979. Genetic variability in *Corbicula*, an invading species. In: J. C. Britton (ed) Proceedings, First International *Corbicula* Symposium. Texas Christian University Research Foundation. Fort Worth, pp. 244-248.

Spolsky, C. M., G. M. Davis, and Z. Yi. 1996. Sequencing methodology and phylogenetic analysis: cytochrome b gene sequence reveals significant diversity in Chinese populations of *Oncomelania* (Gastropoda: Pomatiopsidae). Malacologia. 38(1-2): 213-221.

Stothard, J. R., S. Huges, and D. Rollinson. 1996. Variation within the internal transcribed spacer (ITS) of ribosomal DNA genes of intermediate snail hosts within the genus *Bulinus* (Gastropoda: Planorbidae). Acta Trop. 61: 19-29.

Swofford, D. L. 2000. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) v. 4. Sinauer Associates. Sunderland, Massachusetts.

Thompson, J. D., D. G. Higgins, and T.J. Gibson. 1994. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and matrix choice. Nucleic Acids Research. 22: 4673-80.

Wade, M. J., M. L. McKnight, and H. B. Shaffer. 1994. The effects of kin-structured colonization on nuclear and cytoplasmic genetic diversity. Evolution. 48(4):1114-1120.

Williams, R. J., F. B. Griffiths, E. J. Van der Wal, and J. Kelly. 1988. Cargo vessel ballast water as a vector for the transport of non-indigenous marine species. Estuar. Coastal Shelf Science. 26: 409-420.

Williamson, M. 1996. Biological Invasions. Chapman and Hall. 244 pp.

Wilson, A. C. 1985. The molecular basis of evolution. Sci. Amer. 253(4): 164-173.

Woodruff, D. S., L. L. McMeekin, M. Mulvey, and M. P. Carpenter. 1986. Population genetics of *Crepidula onyx*: variation in a California slipper snail recently established in China. Veliger. 29: 53-63.

Woodruff, D. S., M. Mulvey, M. W. Yipp. 1985. Population genetics of *Biomphalaria straminea* in Hong Kong. J. Hered. 76: 355-360.

Wu, Y. 1988. Distribution and shell height-weight relation of *Rapana venosa* Valenciennes in the Laizhou Bay. Marine Science/Haiyan Kexue. 6:39-40.

Yonge, C. M. and T. E. Thompson. 1976. Living Marine Molluscs. William Collins Sons & Co Ltd. Glasgow. 288 pp.

Zolotarev, V. 1996. The black sea ecosystem changes related to the introduction of new mollusc species. Ecology. 17(1-3): 227-236. Berlin d. Virtssenschafts-Verlag.

VITA

Arminda Lee Gensler

Born in Melbourne, Florida, on 17 May 1976. Graduated from Satellite High School, Satellite Beach, Florida, in 1994. Received a Bachelor of Arts degree in Biology and English from Williams College in 1998. Entered the master's program at the College of William and Mary School of Marine Science, in 1998.