GENETIC DIVERSITY OF POPULATIONS OF THE MALESIAN MOSS, *ACANTHORRHYNCHIUM PAPILLATUM*, AS MEASURED BY MICROSATELLITE MARKERS AND *ITS*2 SEQUENCES

ALFREDO AMIEL P. LEONARDÍA
(M.Sc., Leiden University)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF BIOLOGICAL SCIENCES
THE NATIONAL UNIVERSITY OF SINGAPORE
2007

ACKNOWLEDGEMENTS

A PhD project –particularly one that involves both field and lab work– is never a solitary endeavor. Moreover, contributions to its completion are direct and indirect, physical and metaphysical, immediate, long-drawn, early, recent, last-minute. Though I labored to set down the names of all those I remember who helped, much time has passed since the project's inception that I may have inadvertently left out a person or two despite my best efforts. To them I sincerely apologize and to all I offer my deepest gratitude.

My sincere and deepest thanks to my supervisors, Associate Professor Benito C. Tan and Associate Professor Prakash P. Kumar. Thank you for letting me experiment on microsatellites to near-disastrous results. Thank you for your patience and continuous support. And thank you to the Tan Chin Kee Foundation, for funding my flight to Indonesia to start off my studies on mosses. Prof. Ben, it took me some time to figure this out, but I have not forgotten and I am truly grateful.

To Professor Lam Toong Jin, my sincerest thanks for teaching me to be (more) critical and for the suggestion to focus on one species.

To the National University of Singapore and the Department of Biological Sciences, for funding my research through my research scholarship and various research grants. This study would literally not be possible without your support.

And to Ee Ling, my loving wife, for never tiring nor failing to teach me how to use the pH meter. O flog my feeble brain!

And to those who welcomed me at the Cryptogam laboratory, I celebrate our great times and our friendship. To Boon Chuan, may you be the best muscologist of our generation. To Chang Ying, thank you for the (heated) discussions on

cladistics and for your friendship. To Farida, for teaching me how to be more respectful. To Meng-Shyan, for the early and fruitful discussions. To Su-Fern, for the seed game. To Mr. Chua, for Linux™ and more. To Serena and Hwee Boon, for the sushi. To Yao Hui, for painting with light the beauty of mosses. To Claudia and Kai Yang, for the jigsaw and company in Maxwell hill. To Woon Ping, for help with *Pogonatum*.

And to the geneticists at the Plant Morphogenesis laboratory, thank you for allowing some bench space to a moss guy. To Wang Yu, for the pBluescript. Ping Yu for genome walking. To Li Ang, for the discussions on ISSRs. To Mandar, for our (unfulfilled) vow and for the biscuits. To Dileep, for drinking water in the drawer and for (re)teaching me humility. Lian-Lian and Jaclyn and Xin Ying, how you all changed the lab and made it better. To Mrs. Ang, for your patience and help.

To Professor Haji Mohamed, my deepest gratitude for facilitating the collections in Malaysia as part of an existing research collaboration between your lab and Prof. Ben's. To the frighteningly impressive Yong Kien Thai, I look forward to seeing the wonderful places you and Boon Chuan will be bringing Southeast Asian bryology.

To Dr. Helena Korpelainen and her lab, for kindly hosting my short visit to her lab in the University of Helsinki.

To the National Parks Board, my sincere thanks for allowing us to do our collections in Singapore.

And to the 7th Floor mafia. Carol for commiserating. Serena for the crosslinker. Daryl for badminton. Teng Seah for badminton. Wee Kee for badminton.

To Goh Wee Kee. For being my MCB mother. Twice.

To Lars and Cecilia, it's been so long since we trudged around Maxwell Hill, thanks for the memories and the support in Stockholm.

And to my field mates, how we had great times together! Tzi Ming for taking me along on your trips and for sparking an interest in herps. To Su Lee for the same (minus the herps). Cheng Puay, for your boundless enthusiasm. Eunice for being such a sport. Bicky for demonstrating that Eunice is a sport. Jeremy for the same –and for being so darn calm all the time. Reuben for translating and for the maps. Norman for showing us what it's like to be a field pro. Matthew for Tioman and the RAM. Tommy, for Tioman and **the** crab. Heok Hui for being so impressive. Swee Hee for being so knowledgeable. Darren for being such a natural. Zeehan for being so effective. And Ngan Kee for being the mother to us all.

To Joelle, for planting the seed of microsatellites. See what you got me into?

To my Filipino barkada, for restoring the faith. Ming, for being. JC for the music. Arvin for the "Oh, S...!"

To Chye Fong and Bee Ling, for allowing me do fragment analysis. To Joan, Reena, Sally, Mrs. Chan, and Laurence, no graduate student could survive without you. Thank you for everything! To Ann Nee, thank you for DreamWeaver and all the great posters. To Mr. Soong, thank you for always processing our orders so promptly.

To Vaane and Kavita, Ee Ling has finally shown me the ice cream in Sunset Way. Here's to great times past and in the future.

To the makers of ZipLoc bags. Though your products be humble, they proved to be most indispensable.

To Donald Knuth, Leslie Lamport, Peter Borg and Richard Koch: A botanist wrote his thesis with LaTeX on a notebook with a glowing apple using a text editor iconificated (sic) by a wild strawberry. How strange is that?

To Ee Ling's family, for welcoming me and for badminton.

To my family. No words can capture or express. But you know you're all beyond that.

And again to Ee Ling. You taught me more than just how to use a pH meter. Your love and patience sustained me through the most trying times spent working on this project. But since you can read my mind then you already know that, hunh?

And finally, although this thesis project, in its mostly-quiet, sometimes-angry all-consuming relentlessness, has taken 16% of my life thus far, may I never forget what and Who really matter.

TABLE OF CONTENTS

SUMMARY

Mosses are important yet neglected inhabitants of tropical rain forests. In this study I investigated the genetic diversity of the Malesian moss species, *Acanthorrhynchium papillatum* (Harv.) Fleisch. using microsatellite markers and DNA sequences from the second internal transcribed spacer region, *ITS*2, of ribosomal DNA. Moss samples for analysis were collected from five sampling areas: three areas in Peninsular Malaysia, two in Singapore. These sampling areas differ from each other in habitat quality. Genetic diversity was assessed at several scales: within clumps of a moss, among moss clumps in one sampling area and among sampling areas.

Eight microsatellite markers were newly developed for *Acanthorrhynchium papillatum* for this project. Rigorous tests on these markers revealed that they were suitable for use in population genetic studies of the moss. These were the first markers reported for a tropical moss species.

The microsatellite markers revealed high levels of allelic and haplotypic diversity among clumps of *Acanthorrhynchium papillatum* in each of the sampling areas. Although a reduction in both allelic and haplotypic levels of diversity were detected in sampling areas considered to be more ecologically disturbed, diversity at both allelic and haplotypic levels in these areas was still generally high. High levels of diversity are hypothesized to stem from high mutation rates in the microsatellite markers used as evidenced by several observations in the data. Allelic diversity of *A. papillatum* was also found to be high in areas that were qualitatively thought to be more disturbed. Genotypic diversity was lower in these areas suggesting that vegetative reproduction is more important in these areas for maintaining the population numbers of this moss.

Genetic variation among clumps of *Acanthorrhynchium papillatum* was also seen in another marker, *ITS*2. Considerable levels of diversity among populations of *A. papillatum* were seen with this marker. Similar to the results using microsatellite markers, *ITS*2 haplotype diversity was lower in areas deemed to be more ecologically disturbed. Compared to the diversity levels in the microsatellite markers, however, diversity levels in *ITS*2 were lower, emphasizing the utility and importance of microsatellite markers in population genetic studies.

The microsatellite markers were also used to examine genetic diversity within clumps of *Acanthorrhynchium papillatum*. Most clumps studied had very low levels of diversity indicating that vegetative reproduction was more important within clumps than sexual reproduction. However, multilocus genotypes of samples within some of the clumps studied were not all alike, providing evidence again of high microsatellite mutation rates or of occasional sexuality within this species.

The results obtained provide baseline information on the genetic diversity of *Acanthorrhynchium papillatum* in Singapore and Malaysia and are hoped to form one of the first of a series of studies on bryophyte genetic diversity in Southeast Asia.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Chemicals and reagents

X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
rATP	adenosine triphosphate
BSA	bovine serum albumin
CTAB	cetyltrimethyl (hexadecyltrimethyl) ammonium bromide
DMSO	dimethyl sulfoxide
Na_2EDTA	disodium ethylene diamine tetraacetic acid
DTT	dithiothreitol
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
SDS	sodium dodecyl sulfate
SSC	standard saline citrate
TAE	Tris-acetate, EDTA
TBE	Tris-borate, EDTA
TE	Tris-EDTA

Molecular techniques

RAPD	random amplified polymorphic DNA
ISSR	inter-simple sequence repeats

Sampling areas

BTNR	Bukit Timah Nature Reserve
GBFR	Gunung Belumut Forest Reserve
KTP	Kota Tinggi Waterfalls Resort
MACR	MacRitchie Reservoir
SBRF	Sungei Bantang Recreational Forest

Statistics

A	alleles per locus
AMOVA	analysis of molecular variance
d.f.	degrees of freedom
F_{CT}	fixation index – between groups of populations
F_{SC}	fixation index – among populations within groups
F_{ST}	fixation index – among individuals among populations and among groups
P_M	match probability
h	Nei's gene diversity

D	Nei's genetic distance
D_A	Nei's net genetic distance
$\hat{h}$	Nei's unbiased gene diversity
PD	power of discrimination
P	probability
s.d.	sample standard deviation

Units and measurements

g	acceleration due to gravity
bp	base pairs
cm	centimeter
°C	degrees celsius
g	gram
ha	hectare
h	hour
kV	kilovolt
km	kilometer
m	meter
μ**A**	microampere
μ**g**	microgram

μL	microliter
μm	micrometer
μM	micromolar
mg	milligram
mL	milliliter
mm	millimeter
mM	millimolar
min	minute
M	molar
ng	nanogram
nM	nanomolar
OD	optical density
rpm	revolutions per minute
s	second
cm^2	square centimeter
U	unit
v/v	volume per volume
W	watt
w/v	weight per volume

Others

DNA	deoxyribonucleic acid
dNTP	deoxyribonucleic acid triphosphate
GPS	global positioning system
T_M	melting temperature
PCR	polymerase chain reaction
RNA	ribonucleic acid
rDNA	ribosomal DNA
ts.	transition
tv.	transversion

Chapter 1

GENERAL INTRODUCTION AND REVIEW OF LITERATURE

1.1 General Introduction

1.1.1 Background of the study

Mosses and other bryophytes are integral components of forests throughout the world. Although small and often unnoticed, they provide many significant ecological functions.

Bryophytes are among the first plants to colonize newly exposed surfaces (Hallingbäck & Hodgetts, 2000). They stabilize the soil crust and help in the accumulation of humus, paving the way for the growth of other plants. They regulate moisture in the forests, helping control erosion and flash floods. Their contribution to water regulation is particularly significant in highland environments where they can form a major proportion of the above-ground biomass.

Hallingbäck & Hodgetts (2000) state that "they are critical to the survival of a tremendous diversity of organisms". These include arthropods and other invertebrates that depend on bryophytes for habitat or for food. Bryophytes also provide seedbeds for several tree species (Glime & Saxena, 1991). Some bryophytes even provide substrates for the growth of Cyanobacteria and thus indirectly help in fixing atmospheric nitrogen (Glime & Saxena, 1991; Saxena & Harinder, 2004).

Despite the ubiquity and importance of mosses, little is known of the population biology of many moss species in the tropics. Most studies on bryophyte population biology were done on temperate species. Except for an early study

in the Philippines (de Vries *et al.*, 1983), studies on population genetic diversity of Malesian species are practically non-existent. Baseline information on population parameters such as degrees of genetic diversity, extent of clonal proliferation and characteristics of genetic structure have been established for almost no moss species in the Malesian region. The effects of anthropogenic environmental changes on these parameters are consequently also unknown. At the rate that natural habitats of mosses are disappearing in Southeast Asia, the probability of never finding answers to these questions is becoming all too inevitable.

In this project we established baseline information on the genetic diversity of the Malesian moss species, *Acanthorrhynchium papillatum* (Harv.) Fleisch. Genetic diversity metrics were examined at three spatial scales: within clumps of the moss, among clumps within the same population or sampling area and among different populations. Populations of *A. papillatum* from habitats with different degrees of disturbance were compared to infer the effects of habitat degradation on these metrics. The characteristics of genetic structure were also tested to detect genetically isolated populations.

At the outset of the project several considerations and limitations peculiar to working with mosses had to be addressed.

Many bryophytes are notoriously phenotypically plastic (Buryová & Shaw, 2005; Hedenäs, 2001; Mishler, 2001; Såstad *et al.*, 1999) and inferring genetic diversity from morpho-anatomical characters would have been both time-consuming and potentially futile. Genetic diversity therefore needed to be determined through the use of molecular markers.

Several classes of molecular markers were available for selection. However, the marker used had to be chosen carefully. The chosen marker had to be able to distinguish distinct individuals in a population in order to investigate the peculiar tendency of mosses and other bryophytes to readily, if not predominantly,

reproduce through clonal (asexual) means (Mishler, 2001). The marker had to be PCR-based to accommodate for limited amounts of starting material that can be extracted from the smaller individuals found in the field. The marker had to be species-specific to screen out contaminating DNA from symbiotic organisms that are difficult to remove from the samples. Finally, the marker ideally had to be amenable to high-throughput processing to facilitate the study of large numbers of samples. Microsatellite markers fit this bill entirely and were developed *de novo* for the species studied in this project.

Microsatellites are tandem repeats of short motifs (variously, 1–8 nucleotides) that have been found to be abundantly distributed in the genomes of species in which they have been searched (Bruford & Wayne, 1993). Repeat length of the same microsatellite locus may vary, often greatly, among individuals of the same species. They are heritable and are thus useful as molecular markers in the study of populations.

While microsatellites were the main markers employed, sequences from the second internal transcribed spacer (*ITS*2) of ribosomal DNA were also used as supplementary markers for this study. *ITS* sequences have been used in other studies on the population biology of mosses [e.g., Skotnicki *et al.* (2005); Chiang & Schaal (1999b); Shaw (2000)].

1.1.2 Objectives

The main objectives of this study were as follows:

1. To develop microsatellite markers for the moss, *Acanthorrhynchium papillatum*

2. To establish baseline information on genetic diversity between and among clumps of *Acanthorrhynchium papillatum* using microsatellite markers and *ITS*2 sequences

3. To detect and describe differences in genetic diversity between and among clumps of *Acanthorrhynchium papillatum* found in sampling areas of different degrees of disturbance using microsatellite markers and $ITS2$ sequences

4. To use microsatellite markers to describe the genetic diversity within clumps of *Acanthorrhynchium papillatum*

1.1.3 Scope and limitations

It was the goal of this project to draw generalizations on the genetic diversity of mosses in Malesia and how genetic diversity is affected by anthropogenic changes in their habitats. Constraints in time and logistics, however, limited several aspects of this study.

Only one species of moss, *Acanthorrhynchium papillatum*, was studied. Originally several species of mosses, each with different combinations of substrate preference, habit and sexuality, were to be studied. In fact, microsatellite libraries for four moss species, *Acanthorrhynchium papillatum*, *Pogonatum cirratum* ssp. *macrophyllum*, *Thuidium plumulosum* and *Thuidium cymbifolium*, were constructed in the labs at the Department of Biological Sciences at the National University of Singapore. In the end, however, and for various reasons, only studies on *A. papillatum* were completed.

Collections were done exclusively in Singapore and Peninsular Malaysia. While this obviously precludes the description of genetic diversity in other parts of Southeast Asia, data from samples collected, particularly summary statistics, should represent adequate approximations of the genetic diversity of *Acanthorrhynchium papillatum* in other areas.

Collections of *Acanthorrhynchium papillatum* were not randomly executed and sampling efforts were different for every collection site. The distribution of *A. papillatum* in the areas sampled was patchy; completely random sampling in these

areas would have been logistically prohibitive. Moreover, such a sampling strategy could have potentially reduced the samples collected to statistically cumbersome numbers. To compensate, large numbers of samples were collected per site. In fact, samples were taken from nearly all clumps of this moss encountered in the field. (Sampling, however, was conducted such that the moss clumps and surrounding areas received little impact and disturbance). The collection of large numbers of samples in each collection site also allowed for the better description of genetic diversity of this moss in smaller spatial scales.

Only microsatellite data and *ITS*2-sequences were used as markers. Although microsatellite markers have been shown to be ideal for this study, data from other, independent sources would have been welcome. In fact, *ITS*2-sequences, although inadequate for distinguishing individuals in a population, were used precisely to add more weight and data in describing the genetic diversity of *Acanthorrhynchium papillatum*. Other markers such as single-nucleotide polymorphisms (SNPs) and AFLP, may potentially enrich the findings but are left for future work on this species.

Despite these limitations, the project is the first of its kind in Southeast Asia. What its findings represent are not only significant contributions to the field of population studies on bryophytes, they are urgent contributions as well because of the rampant environmental degradation in the region. Moreover, the results on the diversity studies and the development of microsatellite libraries for *Acanthorrhynchium papillatum* pave the way for further studies on this and related species.

1.2 Review of Literature

1.2.1 Diversity of moss populations

Despite the grave and escalating reality of deforestation and the accompanying loss of species in Southeast Asia, much of the biology of many forest inhabitants

remains unstudied. For instance, although there are many accounts on the taxonomy and species-level diversity of Southeast Asian bryophytes, few studies have been made on the population-level diversity of bryophytes in our region. In fact, there has only been a single study, one that was conducted more than 20 years ago, on the population diversity of a moss species in the Philippines (de Vries *et al.*, 1983). Also, of the few studies that have been done to examine the differences in diversity between moss populations in natural and deforested areas, none has been done in Southeast Asia.

One could argue that the effects of deforestation on the diversity of moss populations could be inferred from the many studies conducted on the flowering plants, a group that is better studied –and is arguably more popular– than the bryophytes. But the many differences in the biology of these two plant groups (Mishler, 2001), particularly the small size of bryophytes and their occupation of microhabitats, make these inferences hard to accept. To understand how deforestation affects bryophytes, bryophytes have to be studied directly.

Notwithstanding the dearth of studies on the population diversity of bryophytes in Southeast Asia, many such studies have been done outside the region.

Early studies and isozyme analysis

The diversities of bryophyte populations were not at first examined directly, but were rather deduced from data on species-level diversity and implications of the dominant haploid life cycle of bryophytes [see historical summary of Cummins & Wyatt (1981)]. Bryophytes are thought to be an ancient group of land plants: fossils have been found that closely resemble modern day species (Anderson, 1963; Goffinet, 2000; Schofield, 2000). Ennos (1990) explains that this discovery led people to assume that bryophyte speciation peaked early in geologic history and that since then, speciation proceeded at a rate lower than that found in the vas-

cular plants. He continues, saying that it was hypothesized that the underlying cause of this low rate of speciation is a lack of genetic variation in bryophytes. Without adequate means of testing this assumption, and perhaps complicated by the phenotypic plasticity of bryophytes, this idea persisted until the availability and use of molecular biology techniques.

The first direct investigations on the genetic diversity of moss populations employed electrophoresis of isozymes, also called allozymes. The use of isozymes is one of the earliest molecular methods to be used for evolutionary studies. However, because of the simplicity of the technique and its relative low cost, it is still in common use today. The method is based on visualizing on a sieving gel the different molecular forms of enzymes that have the same catalytic activity (Gottlieb, 1971).

The earliest studies using isozymes on bryophytes were on liverworts. They were done by Maria Krzakowa and her colleagues. In one of her papers she describes the polymorphism of isozyme markers of populations of a liverwort, *Plagiochila asplenioides* (Krzakowa & Szweykowski, 1979). Over 400 shoots from five natural populations in Poland were collected, sustained in the lab, and examined for variability of the enzyme system of peroxidases. Their results showed high levels of intrapopulation variability for this enzyme comparable to those reported for many populations of the higher plants. While their results for levels of interpopulation variability indicated that *P. asplenioides* was divided into races in Poland, no comparisons with angiosperm data obtained in the same habitat were made. Their results, however, were the first to challenge the presumptions of the lack of genetic variability in bryophytes.

Among the first studies exploring the genetic variability of a moss species was on populations of *Atrichum angustatum* (Cummins & Wyatt, 1981). The authors collected 4 clumps from 15 populations of the moss from east Texas, U.S.A.,

and tested 10 enzyme systems on them. They found that 8 enzyme systems had varying degrees of staining activity of which 4 were scorable. Their results indicated significant levels of variability in these four enzyme systems, both among the different clumps of the same population and among different populations. Moreover, they found that, once again, levels of polymorphism were comparable to those of other plants and animals that had been studied using the same techniques. They further argued that because of the small sample sizes of their study, their results probably underestimated the actual genetic variability in their samples.

One of the earlier studies was, surprisingly, conducted in our region. Three populations of two species of moss, namely, *Racopilum spectabile* and *Racopilum cuspidigerum* were collected from the Philippines and examined for allozyme variability (de Vries *et al.*, 1983). Twenty-two enzyme systems were tested of which 8 were scorable. The study found that the levels of genetic variation in these species were similar to those found in the vascular plants, again contrary to earlier assumptions of the genetic homogeneity of bryophyte populations. All the samples were collected from natural populations, from intact montane regions in different parts of the country. No correlations between diversity and habitat type were sought nor therefore found. One interesting aspect of their research was that they collected samples that were at varying distances from each other, from a few decimeters to a maximum of 3 kilometers, allowing them to test for correlations between genetic distance and spatial distance. Exhaustive sampling at such a scale was not attempted, however, although they did find positive correlation between increasing genetic distance and increasing spatial distance at magnitudes of kilometers to hundreds of kilometers. Finally, few numbers of specimens, 7 to 20, were sampled from each population, limiting their ability to form strong conclusions about their study. The research, however, is a pioneering work on the population genetics of mosses in Southeast Asia. Unfortunately, apart from this

study, no other study on genetic diversity of bryophytes in our region has since been done.

Another isozyme study involved analysis of 1736 shoots of *Polytrichum commune* collected in the United States (Derda & Wyatt, 1999). Their study revealed lower mean levels of diversity within populations compared to other mosses. However, they found higher total genetic diversity in the species caused by the occurrence of several alleles with high frequencies in different geographical regions. Gene identities were therefore also high within these regions. It was found that region-specific genotypes and gene identities between regions also decreased with increasing geographical distance. However, genotypes known to originate in specific regions were found in distant populations indicating occasional long-distance dispersion and colonization. Still, differentiation among populations was high indicating that this dispersal was limited. They also found that the examination of a related species, *P. jensenii* revealed some similarity to *P. commune* but that overall differences between the two were high enough to consider them distinct species.

Many other studies were conducted using isozymes as molecular markers [see lists of Akiyama (1994) and Wyatt *et al.* (1989)], all of them showing similar results; that is, that levels of genetic variation in bryophytes were higher than originally supposed and were comparable to levels found in the vascular plants. As is common in pioneering work, many of these studies were exploratory in nature and were used mostly to refute the then prevailing idea that bryophytes are genetically depauperate organisms. Other studies went beyond just establishing levels of genetic variation in bryophytes: de Vries *et al.* (1983) used them to compare variation among and within moss species, as well as to establish correlations between genetic and spatial distance. Akiyama (1994) investigated the gene flow of an epiphytic moss, *Leucodon*, when most studies prior to his were on terrestrial

mosses. Wyatt (1992) studied the differences in the levels of genetic variability of species of *Plagiomnium* as a function of their distribution and abundance as affected by habitat quality.

The use of isozymes for studies of the population variability of mosses however, has its limitations. For example, plants collected in the field have to be grown and sustained in the lab to ensure that sufficient levels of enzymes are available for detection. This has at least two limiting effects: it increases the time between collection and analysis, consequently decreasing throughput. It also limits the plants that can be analyzed to those individuals that can be grown and sustained in the laboratory, possibly introducing bias to effective sampling. Neither the experimental procedure nor the conversion of allozyme data into genetic markers lends easily to automation. The technique also needs larger amounts of tissue than is needed for PCR-based techniques [see Selkirk *et al.* (1997)], amounts that might be difficult to obtain from smaller species or from degraded samples. Of greater significance is the observation that isozyme markers exhibit lower levels of variation compared to DNA-based markers (Cavalli-Sforza, 1998) and consequently have lower resolving power.

Later studies and PCR-based techniques

The advent of the molecular technique, polymerase chain reaction or PCR, opened new avenues in the study of bryophyte biology. Mosses, being generally smaller than angiosperms, yield comparatively little amounts of tissue and DNA per individual. Such small amounts of are difficult to use directly but are sufficient to serve as templates for PCR amplification.

Having already established the significant genetic variability of bryophyte populations, many of the PCR-based studies explored the population biology of bryophytes beyond just examining genetic variability. As in the case of studies using

isozymes however, most of these studies were conducted in temperate areas.

RAPDs One of the earliest PCR-based techniques used in population studies of mosses is the use of random amplified polymorphic DNA or RAPDs, a method independently developed by Welsh & McClelland (1990) and Williams *et al.* (1990). In this technique, primers of arbitrary sequence are used in low stringency PCR conditions to generate several amplicons from genomic DNA templates. These amplicons appear as bands when run through a gel and are useful as markers when either present or absent (polymorphic) among different samples.

Of the few studies on the population variability of tropical mosses, one used RAPDs and was conducted in Panama on species of the bark-growing genus of moss, *Octoblepharum* (Korpelainen & Salazar Allen, 1999). Six RAPD primers were used to detect genetic variation within three different species, *Octoblepharum albidum*, *O. cocuiense* and *O. pulvinatum*, and to measure genetic distances among these three species. Genetic distances within one species, *O. albidum*, collected from different areas were also measured. Thirty-four samples of *O. albidum*, 23 of *O. cocuiense* and 16 of *O. pulvinatum* were collected from different habitats in Panama and were analyzed for this study. Examination of within-colony variation of different gametophytes were also conducted in this study. Of the twenty RAPD primers screened, banding patterns from six that gave reproducible results were analyzed. About 20 polymorphic bands for each of the species were generated using these primers. Considerable genetic variations were found in each of the three species studied. The RAPD data, moreover, showed nearly equal distances among the species. Discriminant functions generated from the RAPD data correctly placed samples in their respective species, strengthening the evidence that genetic identities were well-established among these species. Finally, some polymorphism in the gametophytes of colonies of *O. albidum* and *O. pulvinatum* were found, while only clonal colonies were seen in the samples of *O. cocuiense* studied.

A series of studies using RAPDs were conducted on several moss species in Antarctica.

In one of the studies (Selkirk *et al.*, 1997), both isozymes and RAPDs were used on *Sarconeurum glaciale* to determine genetic variation within and among populations of this species. Sixty-six samples collected from various localities in Antarctica were tested with 5 isozyme systems and 5 RAPD primers. Levels of genetic variability were found to be higher for RAPD than for isozyme data. Data from RAPD also indicated clear separation of populations sampled from different areas, suggesting that the continent was populated by the moss through multiple colonization events.

A similar study was conducted on *Ceratodon purpureus* collected from a channel formed from a meltstream waterfall in Antarctica (Skotnicki *et al.*, 1998). The technique was used to discover variation within clumps and between clumps. The population of clumps in this waterfall channel were also compared to two other populations in Antarctica, one a few hundred meters away, another, 300 km away, and to a population in Sydney, Australia. Six shoots of the moss from each of three localities, the top, middle and bottom of the waterfall channel, were sampled to measure variation among clumps in the channel. A total of 46 shoots were compared to asses genetic distance and population structure among the populations in Antarctica and the one in Sydney. Forty-five RAPD primers generating 73 bands were used for analysis. Clear genetic variation was found within and between clumps of the meltstream samples. A neighbor-joining tree constructed from data from these samples indicated some clustering that followed the position of the samples in the channel. Another neighbor-joining tree was constructed from data from the different populations. Samples from the meltstream channel formed clear clusters that were distinct from the other population in Antarctica and in Sydney, themselves also forming clear clusters.

Examination of the genetic diversity of *Bryum argenteum* collected from a similar habitat in Antarctica was conducted by Skotnicki *et al.* (1999). A total of 39 samples of the moss were collected from three meltstream channels. Many more samples collected from other parts of Antarctica were included in the study. Four RAPD primers were used and generated 75 usable bands. Once again, samples from the meltstream channels showed some clustering according to their position in the channel: samples from the top clustered with other samples from the top, samples from the bottom with other samples from the bottom. Extensive genetic variability was observed in all of the populations sampled with high levels of within-population variation found. An interesting finding was that samples from geographic areas with highly sparse colonies exhibited the same level of variability among clumps as those from more lush colonies.

In the north-temperate regions, studies using RAPD markers were also conducted. Genetic differentiation among species of *Polytrichum* were examined by Zouhair *et al.* (2000). Ten RAPD primers were used on 30 colonies of *Polytrichum* representing 6 species. One hundred and sixty-six polymorphic markers were used in a cluster analysis to generate genetic distances and indices of similarity within and among the different species. While results of RAPD data were found to be largely consistent with results from morphological data, the high levels of genetic differences found using RAPD data highlighted its potential use in discriminating among closely related taxa.

The rare Norwegian peat moss, *Sphagnum troendelagicum*, was studied by Stenøien & Flatberg (2000) using RAPD markers. Because of the rarity of the species, only three (of only five known) populations were sampled. A total of 77 gametophytes were analyzed with 10 RAPD primers. While earlier isozyme data showed negligible genetic variability within and among populations of this moss, low to moderate levels of variation were found with RAPD analysis. Moreover,

both allelic and haplotypic levels of variation were found to be higher than expected for a species that is thought to reproduce asexually. These unusual results led the authors to hypothesize that the species had multiple origins and were formed through recurrent hybridization.

RAPD remains popular in many studies of population variability. Its simplicity, and cost-effectiveness favor its use in preliminary investigations where rapid results are desirable and where no or little sequence information is known for the species being studied. But RAPDs also have their limitations. For one, it is sensitive to contamination. Since RAPD markers are anonymous and not species-specific, exogenous DNA sources can produce artifactual data and lead to erroneous analysis and interpretation of results. This is an important consideration among mosses where their growth-habit and habitat preferences make it difficult to exclude exogenous DNA sources during sample preparation. Moreover, because of the low stringency of the PCR conditions that are characteristic of the technique, banding patterns are sensitive to the PCR setup (Hadrys *et al.*, 1992): they are affected by salt concentrations, thermal cycling conditions, type and brand of polymerase, and make of thermal cyclers. Small changes in any of these components can affect the reproducibility of the results. Moreover, the bands themselves can be difficult to score. Because some bands appear brighter than others, faint bands may be scored by some researchers and ignored by others, introducing bias to the results.

ISSRs and AFLP There are other PCR-based techniques that, like RAPDs, need little initial preparation or sequence information about the study organism. Unlike RAPDs, however, these techniques yield more consistent results and are less prone to bias.

One such technique is ISSRs, or inter-simple sequence repeats (Zietkiewicz *et al.*, 1994; Gupta *et al.*, 1994). The technique is similar in form and execution

to RAPDs. But instead of using short primers of arbitrary sequence, it uses longer primers composed of short tandem repeats, e.g., $(GA)_9$, $(GT)_9$, $(CTC)_6$. These primers bind to the complementary sequences in the genome of the study organisms. Where these sequences are sufficiently close to each other and oriented in the correct way, PCR ampicons are produced. The amplicons are visualized and analyzed in the same way as RAPD data. However, the higher annealing temperatures associated with the longer primers allow for more stringent PCR conditions. These help prevent mis-priming and the formation of artifactual bands.

One study using ISSRs on mosses was conducted by Hassel *et al.* (2005). In this study, the effects on the population genetics of range expansion of the species, *Pogonatum dentatum* was studied by comparing populations from a mountain area and from a recently colonized lowland area in Sweden. Four populations from each area were studied. From each population, 5 patches were sampled. From each patch, 5 shoots were selected for processing and DNA extraction. Four ISSR primers were used to study the populations. These generated 18 polymorphic loci allowing the recognition of 64 haplotypes out of the 194 total shoots studied. Similar levels of gene diversity were found in both populations although slightly fewer numbers of alleles per locus were seen in populations from the lowlands. The ISSR markers were able to detect recent bottleneck events in three of the four lowland populations. Moreover, the patterns of allelic diversity were found to suggest that loss of diversity through founder effects and genetic drift accompanied the expansion of the range of the species. The analyzed data also indicated that sexual recombination played a greater role over asexual reproduction in the lowland populations compared to the montane populations. Less genetic differentiation in the lowland populations also suggested more unrestricted gene flow in these populations.

Another method that is also similar to RAPDs is a technique called AFLP,

invented by Vos *et al.* (1995). This technique is more complicated in execution than either RAPDs or ISSRs but, like these two other techniques, requires no sequence information from the organism being studied. Although AFLP takes more time to execute, typically many more polymorphic bands are generated with this method than either RAPDs or ISSRs. This lends strength to the analysis and interpretation of the data. Moreover, the PCR conditions involved in AFLP analysis are also more stringent than those in RAPDs. Greater consistency in the data as well as a reduction in artifactual data are thus obtained.

The first application of AFLP on mosses was by Vanderpoorten & Tignon (2000). The technique was used to study the genetic variability among five populations of the moss, *Amblystegium tenax*, sampled in Belgium in areas of contrasting water chemistries. The technique was also applied to two populations of the closely related species, *A. fluviatile*, for comparison. Three pairs of AFLP primers were tried. Only one pair, Eco-`AGG`/Mse-`CAT` generated usable data. Thirty-five scorable fragments were found of which 30 were polymorphic. The results of the study indicated high levels of differentiation among the different populations, higher than between some of the populations and *A. fluviatile*. This led the authors to conclude and recommend a re-evaluation of the taxonomy of the genus and closely related members of the family Amblystegiaceae. They also found clear segregation of the populations collected from the classical area of distribution with those from the marginal populations, and that the segregation correlated with differences in the quality of the water in which these populations were found.

Another study on moss species using AFLP was done by Pfeiffer *et al.* (2006). In this study of the clonal diversity of the moss, *Rhytidium rugosum*, 21 samples from two closely located plots in Germany were subjected to AFLP analysis. Fourteen other samples from France, Russia, Canada and other parts of Germany were included in the analysis for reference. Using two primer combinations on all

these samples, a total of 144 bands were generated, of which 96.5% were polymorphic. Interestingly, of the closely situated plots, only 7.6% were polymorphic indicating the heavy extent of clonal propagation in these plots. Computed pairwise distances were also small for the non-clonal samples in the plots indicating low genetic differentiation. The findings helped confirm the clonal nature of the mosses from the German plots and shed light on the degrees of variability in these populations.

Both ISSRs and AFLP are more robust than RAPDs. However, since either technique uses anonymous, non-species specific markers, they, like RAPDs, also suffer from sensitivity to contamination from exogenous DNA sources.

Microsatellites Another PCR-based technique that has been used to study the population variability of organisms, including mosses, are microsatellite markers (Bruford & Wayne, 1993). Microsatellites, also called simple-sequence repeats (SSRs) or short tandem repeats (STRs), are tandem repeats of sequence motifs from 1 to generally 8 bases, e.g., $(GA)_9$ or $(TAGTCG)_5$. Microsatellite loci can be categorized according to the "purity" of their repeats. For example, Weber (1990) classifies repeat sequences that have no interruptions, e.g., $(GA)_8$ as "perfect"; repeat sequences that are interrupted somewhere along their run, e.g., $(GA)_8T(GA)_7$ are "imperfect"; and repeat sequences that have adjacent repeats of different motifs, e.g., $(GA)_7(GT)_8$ are "compound". Microsatellites, as has already been suggested in the discussion of ISSRs, are abundant and scattered throughout the genome of eukaryotic organisms. They are often length-polymorphic between different individuals, e.g., in one individual a microsatellite locus could be $(GA)_9$, in another individual, the same locus could be $(GA)_{12}$. Potentially more alleles in the form of number of repeat units can be generated per microsatellite marker than for any of the PCR-based techniques discussed earlier. (RAPDs, ISSRs and AFLP markers have a maximum of only two alleles per locus: the presence or the

absence of an amplicon).

The paper of van der Velde *et al.* (2000) is the first report of the development of microsatellite markers for a moss species. Fourteen microsatellite loci were developed for the moss, *Polytrichum formosum*, after screening a genomic library of 8900 recombinant bacterial clones and testing the candidate primers on 181 Dutch and Danish colonies of the moss. These resulting microsatellite markers were used in several subsequent studies.

The genetic structure of *Polytrichum formosum* was studied using microsatellite markers (van der Velde *et al.*, 2001b). The relative contribution of sexual versus asexual modes of reproduction on the amounts of genetic structure in the moss were tested using the same populations used to test for the variability of the microsatellite markers. Levels of microsatellite variability were found to be lower than for other plant species. However, genotypic diversity was found to be high, indicating that sexual reproduction contributed greatly to the levels of genetic structure within *P. formosum* populations. Congruent with their data using isozyme markers (van der Velde & Bijlsma, 2000), no significant genetic differentiation between populations in the Netherlands or Denmark were found. The authors hypothesized that these observations are probably best explained by effective spore dispersal throughout this geographical range.

Although the nature of microsatellite markers make them generally species-specific, microsatellite markers developed for one species can be tested to see if they are useful on other, closely related species. Several microsatellite markers developed for *Polytrichum formosum* were shown to be useful for several other species of *Polytrichum* (van der Velde & Bijlsma, 2003) and have been exploited for other studies.

One of the few studies to discover the effects of habitat change on genetic diversity of moss populations used the microsatellite markers developed for *Poly-*

trichum formosum. Two hundred and fifty-six cushions of *P. commune* from four populations in Northern Ireland peat bogs were sampled and tested using three microsatellite markers (Wilson & Provan, 2003). In contrast to the findings of van der Velde *et al.* (2001b), the authors of this report found that genetic diversity was higher than had been assumed for bryophytes. However, lower levels of genetic diversity were found in populations from completely cut bogs than in uncut peatlands. Moreover, more genetic structure was found in the populations from the cut bogs, suggesting that genetic drift was already affecting the fragmented populations.

Hybridization and asymmetric reproductive isolation between *Polytricum commune* and *P. uliginosum* were discovered using microsatellite markers originally developed for *P. formosum* (van der Velde & Bijlsma, 2004). Analysis of three microsatellite markers in sympatric sporophytes of the two species revealed interesting relationships between them. While the two species were previously thought to be reproductively isolated from each other, the results of the study found the relationship to be more complex. Reproductive isolation between female gametophytes of *P. commune* × male gametophytes of *P. uligionosum* was found to be complete and prezygotic (or early postzygotic). However, reproductive isolation of female gametophytes of *P. uliginosum* × male gametophytes of *P. commune* was determined to be clearly postzygotic as evidenced in large numbers of hybrid sporophytes among female *P. uliginosum*. The authors observed irregular development in these hybrids, indicating their inability to produce ripe spores. The authors also discussed the rare possibility of producing viable spores and the formation of allodiploids, and further discussed the importance of these processes in interspecific hybridization and bryophyte speciation.

The labor required to develop microsatellite markers, despite their power and utility, has evidently limited their widespread adoption in studies of population

genetics of mosses. Indeed, apart from the paper on the development of microsatellite markers for *Polytrichum formosum*, there have been only two other published reports of microsatellite markers for moss species. However, microsatellite markers are actively being developed for other moss species in the lab of Helena Korpelainen in Finland (pers. comm.). Primer sequences for these microsatellite markers have not yet been published at the time of this writing.

Sequence data Although direct DNA sequence information are more often used in phylogenetic studies at species or higher levels of classification, they have also been used in studies of genetic variation of mosses. Among the available DNA sequence markers that are useful in studies of genetic diversity are sequences of the internal transcribed spacers (*ITS*) of ribosomal DNA.

Ribosomal DNA encodes the ribosomal RNA subunits. They occur as tandemly repeated copies of ribosomal RNA genes of highly variable copy number and are coded at one or more chromosome locations (Capesius, 1997). Two internal transcribed spacer (*ITS*) regions are found among the different coding units (Figure 1.1). The first spacer, *ITS*1, is found between the 18S and the 5.8S ribosomal RNA coding regions, the second, *ITS*2, is found between the 5.8S and 28S ribosmal RNA coding regions.

Internal transcribed spacers evolve rapidly (White *et al.*, 1990) and their sequence composition as well as length in base pairs are highly polymorphic among species within a genus or among populations of a species (Capesius, 1997; Rogers & Bendich, 1987). Because the coding sequences for the small and large subunits of rRNA are highly conserved (Capesius, 1997), universal primers for the amplification of *ITS* units have made it possible to amplify and sequence these regions in different organisms (White *et al.*, 1990; Saar *et al.*, 2001). Both their polymorphic nature and easy amplification have made *ITS* regions useful as molecular markers in studies of population genetic variability.

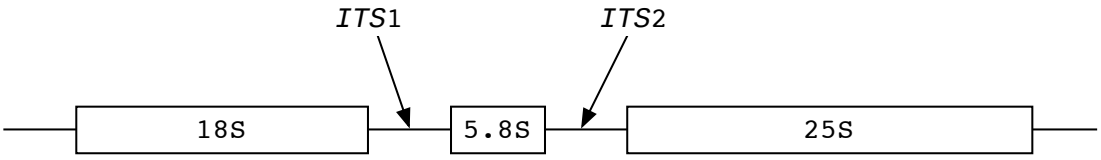


Figure 1.1: Location of internal transcribed spacer regions, $ITS1$ and $ITS2$, in ribosomal DNA. 18S, 5.8S and 25S are the ribosomal DNA coding regions.

Apart from being used in studies of population-level variability, *ITS* sequence data have also been used in taxonomic studies of angiosperms [e.g., Acevedo-Rosas *et al.* (2004); Baldwin (1992); Möller & Cronk (1997); Muellner *et al.* (2005)] and mosses [e.g., Stech (2004)]. They have also been used in studies of phylogeography and cryptic speciation of both groups of plants [e.g., Wittall *et al.* (2004); Chiang & Schaal (1999b); Shaw (2000)].

Chiang & Schaal (1999a) used DNA sequences from the *ITS*1 region to examine the genetic diversity of members of the family Hylocomiaceae. Eight species in six genera were sampled. Since both within-species and between-species levels of diversity were measured, two outgroup taxa were also included in the study for comparison using cladistic analysis. Specimens were collected from the United States and Canada. Their results indicated that genetic variation in the form of insertions/deletions or indels were a common finding in all of the taxa they studied. Most of these were single-base indels, many other indels were from 2–9 base pairs long with some indels longer than 10 base pairs. Nucleotide diversity was found to be lower in monotypic genera than in genera with more than one species. These differences in nucleotide diversity, however, were explained by the authors to be possible artifacts in classification and taxonomy: A genus with more than one species may in fact be composed of distinct genera. Lumping together distinct genera will have the effect of artificially increasing nucleotide diversity. The authors also found that nucleotide diversity was lower in widespread taxa than those with limited distribution.

Sequence information from *ITS* markers was also used to revisit the genetic diversity of Antarctic species of mosses (Skotnicki *et al.*, 2005). Both *ITS*1 and *ITS*2 regions of nine moss species from the Ross Sea region of Antarctica were sequenced and used to both infer genetic diversity of the species and evaluate their potential to be used in species identification. Sequences from over 30 specimens

from several regions were obtained; some of these specimens had not been positively identified using traditional methods. The results, however, indicated that these specimens belonged to 9 distinct species and helped place the ambiguously identified specimens into their proper taxa. Sequence variability in one of the taxa studied, *Bryum argenteum*, was low and showed that samples of *B. argenteum* from geographically close sites were similar. Specimens sampled from farther away exhibited greater genetic variation. Some of the authors of this research were involved in earlier studies of Antarctic mosses using RAPDs. They found that the *ITS* results concurred with their earlier RAPD data. Moreover, they reported sequence data as a simple technique that can be used with traditional methods for the identification of mosses.

Sequence data derived from a single locus, however, are products of the effects of evolution on only that locus. The evolutionary history of this locus may not be completely representative of the evolutionary history of the organism [see introduction of Zhang & Hewitt (2003)]. While this limitation may be overcome by using sequence data derived from several loci, such compensation may prove prohibitively expensive for projects dealing with large numbers of samples.

Summary of current state of research

There are many other reports on the genetic variability of mosses [e.g., Grundmann *et al.* (2007), see also part of the summary of Goffinet & Hax (2001)], some using a single class of marker, others using combinations. While they have not all been reviewed here, several insights are notable from the examples that have been presented:

All these examples show that while numerous studies have been done on the population variability of mosses, most were done on temperate species. Even areas like Antarctica that are harsher environments for study than the tropics

have had more native species studied than the tropics. Baseline data have already been generated in temperate areas and studies there have progressed further than just determining genetic variability of species. Only one study has been done in Southeast Asia, so long ago that the rapid changes in forest environments have probably rendered the results out of date. The number and type of studies on the genetic diversity of tropical species are badly trailing those on temperate species. This imbalance is only made worse by the fact that there are many tropical species of mosses and that the habitats where they are found are in greater risk of destruction. The results of any present studies on the genetic diversity of mosses in our area will not only contribute to our understanding of the population genetics of mosses but may also contribute to ways of conserving these potentially threatened organisms.

Although many of the past studies have reported more than just the genetic diversity of moss populations, few of these studies are explicit comparisons of diversity of populations of the same moss species between natural and degraded habitats. Akiyama (1994) found that one species of *Leucodon* that grew in disturbed habitats had lower gene diversities than other species that grew in more stable areas. However this comparison was made across species, not within populations of the same species. Hassel *et al.* (2005) compared populations of *Pogonatum dentatum* found in their natural environment and in recently colonized areas. Studies on the genetic diversity of relic populations that are left after habitat change have not been widely done.

Another item of note is that most of these reports used samples that were collected over broad geographic ranges. While genetic diversity has been detected and interesting results have been obtained in these studies, other information may have been literally skipped over and gone undetected. The small sizes of mosses imply that their population processes may operate at a smaller scale. Being

small also enables more individuals of a moss to cover the same surface area than individuals of higher plants. What this means to the researcher is that the scales at which sampling should be done is different for mosses as they are for the higher plants. Where one would sample in scales of kilometers for the higher plants, a moss researcher might need to sample in scales of tens of meters or meters to see the same patterns in mosses. This implies that potentially a lot of details on the diversity of mosses may be lost if sampling is not performed at an appropriately small scale.

1.2.2 Selection of molecular markers for this study

With the necessity to study the genetic diversity of tropical mosses now established, an appropriate molecular technique needed to be chosen for this study. A summary of the characteristics of each of these molecular techniques is presented in Table 1.1.

While each of these methods has proven its worth in studies of population variability, we see that all of them have their limitations, whether in preparation or execution of the technique or the interpretation or quality of the results. At the outset, isozymes were excluded as a choice to allow for higher-throughput processing, to take advantage of PCR-based techniques, and to allow unbiased sampling that may include individuals that are difficult to sustain in the lab. While it would be easiest to use RAPDs, the suspect rigor of the data obtained as well as the sensitivity to contamination of the method to exogenous DNA sources also excluded the technique from the choices. ISSRs and AFLP while known to produce data with greater rigor, are also sensitive to exogenous DNA contamination and were also excluded as choices. Moreover, at the time this project was started, AFLP was known to need large amounts of starting DNA that were difficult to obtain with smaller individuals of mosses.

Table 1.1: Summary of characteristics of different molecular techniques used in studies of population variability

Technique	Kind of variation	Advantages	Limitations
Isozymes	mobility polymorphism	inexpensive, easy execution	difficult to automate, less informative than PCR-based techniques
RAPDs	presence/absence	easy execution, no prior sequence information needed	prone to artifactual bands, sensitive to exogenous DNA
ISSRs	presence/absence	easy execution, no prior sequence information needed, more rigorous PCR-setup	sensitive to exogenous DNA
AFLP	presence/absence	many loci tested, no prior sequence information needed, more rigorous PCR-setup	sensitive to exogenous DNA
Microsatellites	length-polymorphism	easy execution, highly polymorphic	difficult development, generally species-specific
DNA sequencing	base changes, insertions/deletions	easy execution, informative	prone to linkage equilibrium from single locus tested

In line with the objectives of the study, and considering the strengths and limitations of available molecular techniques, microsatellite markers were chosen as the main marker for this project. The rigor of microsatellite data, ease of use of the markers in the laboratory plus the availability of new, easier techniques promised to offset the difficulty associated with their development.

To supplement data from microsatellite markers, direct sequence information from internal transcribed spacer regions were also included in this study.

1.2.3 Description of Acanthorrhynchium papillatum

Although ideally several species should be examined to gain a wider perspective of the genetic diversity of mosses in our area, time and resource limitations prevented this from being done within the course of a PhD project.

Since it is an objective of this project to examine the effects of habitat change on the diversity of moss species, a forest species that grows in both pristine and degraded environments needed to be chosen. It was also desirable to work on a species that could be reliably identified in the field to reduce the number of erroneously collected samples due to misidentification. It was essential that the chosen species be found in sufficiently large numbers in their habitats, firstly to enable the assessment of polymorphism of candidate microsatellite markers, and more importantly, to ensure proper sampling and statistical analysis of data on diversity.

Of the several species initially considered, *Acanthorrhynchium papillatum* was finally chosen. Early attempts to develop microsatellite markers for this species proved promising. Moreover, field trips revealed that they were sufficiently abundant in habitats of different degrees of degradation. (The field trips also revealed the dearth of the other species initially considered).

Taxonomy

Acanthorrhynchium Fleischer is a genus in the family Sematophyllaceae. It was originally described by Fleischer (1923) to include *Acanthodium* (a name that had already been used and needed to be replaced) and *Taxithelium* subgenera *Monostigma* and *Oligostigma* (O'Shea, 1997). It belongs to the subfamily Sematophylloideae in the same subgroup as *Acroporium*, *Rhaphidostichum*, *Sematophyllum* and *Trichosteleum*. The genus is found in the old world tropics, from the east African islands through Asia to western Oceania (O'Shea, 1997). Early studies put the number of species in this genus to 11, although Tan & Buck (1989) state that most of the species may in fact be synonymous with *Acanthorrhynchium papillatum*. In the recent study of Tan & Chang (2004) using *rbcL* sequences, *Mastopoma scabrifolium* was transferred to *Acanthorrhynchium* to form the combination *Acanthorrhynchium scabrifolium*, a second, recently established species of *Acanthorrhynchium* found in Malesia.

The type species of the genus is *Acanthorrhynchium papillatum* (Harv.) Fleisch., formed by the synonymy of *Hypnum papillatum*, *Trichosteleum atrocarpum* and *Trichosteleum enganano*ae Fleischer (1923). O'Shea (1997) revised the taxonomy of the African species of *Acanthorrhynchium*, synonymizing them all with *A. papillatum*, essentially extending the range of this once purely Asian species to Africa.

Description

Morphoanatomy *Acanthorrhynchium papillatum* has been described in detail in several publications (Bartram, 1939; Fleischer, 1923; O'Shea, 1997). It is distinctive in the field by a combination of characteristics. The plants are pleurocarpous, or creeping, in matted clumps when abundant. The main stem has complanate branches, the branches have complanate leaves. The leaves are ecostate, ovate-

lanceolate, often with acuminate tips and dentate to serrate margins. Alar cells are large, sometimes colored. Laminar cells are short with thick walls and are prominently unipapillose (Figures 1.2 and 1.3). Although few fertile specimens were found during the course of our study, sporophytes offer the additional distinctive features of having a gibbous capsule on a long seta.

Distribution With the synonymy of the African species of *Acanthorrhynchium*, *A. papillatum* is now distributed from the east African islands of Comoros, Madagascar and Seychelles, to Indochina, China, Malesia, Australia, New Guinea, Fiji and Samoa (O'Shea, 1997).

Local ecology *Acanthorrhynchium papillatum* is reported as being a corticolous moss on tree stumps, fallen branches or logs and leaf litter (O'Shea, 1997).

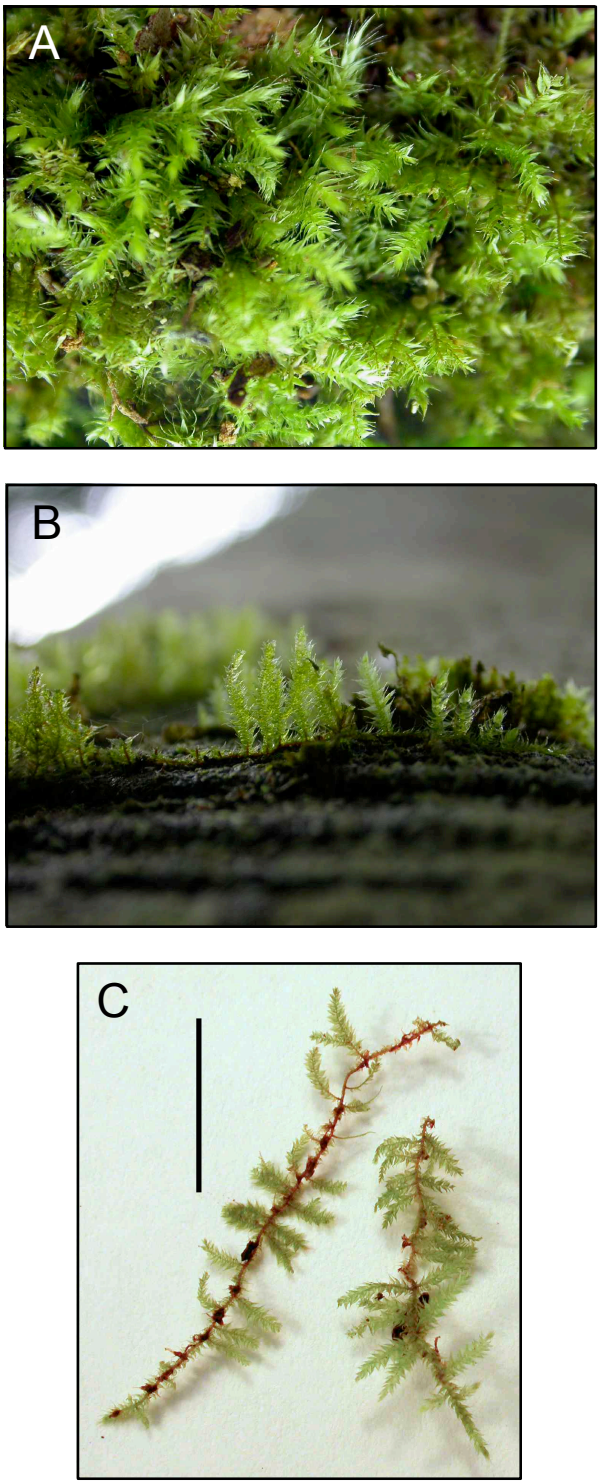


Figure 1.2: Habit and gross morphology of *Acanthorrhynchium papillatum*: (A) Matted clump (Photo by Lim Yao Hui). (B) Solitary ramet growing on a tree bark. (C) Ramets prepared in the lab (scale bar = 1 cm).

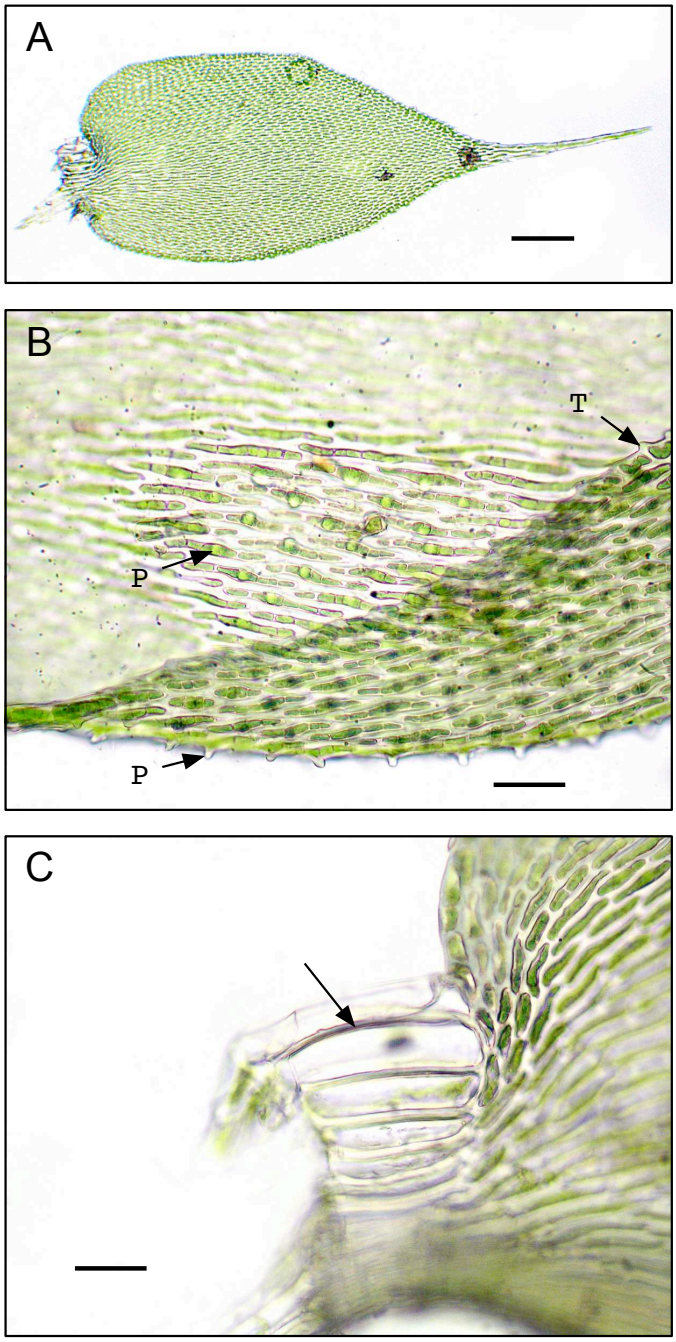


Figure 1.3: Leaf anatomy of *Acanthorrhynchium papillatum*: (A) Whole leaf fresh mount (scale bar = 100 μm). (B) Magnified view showing papillae (P – Papillae on leaf surface, T – tooth on leaf margin; scale bar = 10 μm). (C) Magnified view showing alar cells, indicated by arrow (scale bar = 10 μm). (Photos by Lim Yao Hui).

Chapter 2

DEVELOPMENT OF MICROSATELLITE MARKERS FOR *ACANTHORRHYNCHIUM PAPILLATUM*

2.1 Introduction

To be useful as markers, microsatellite loci have to be first isolated from the genome of the species for which they are being developed. Primers that flank and amplify microsatellite loci have to be designed and tested. Finally, the amplicons have to be tested for length-polymorphism by electrophoresis through a size-separting matrix. The difficulties associated with microsatellite marker development lie mostly in the process of isolating the microsatellite loci and their flanking regions.

There are several ways to isolate microsatellite loci from genomic DNA (Zane *et al.*, 2002; Selkoe & Toonen, 2006). Several of these methods were tried in the course of this project.

The least labor-intensive method is to use microsatellite markers that have already been developed for closely-related species. No microsatellite markers had been developed for tropical moss species prior to the inception of this project and so this method was not an option for this study.

The traditional, least complicated method to develop microsatellite markers from scratch is to screen genomic DNA libraries using labeled probes and to sequence the resulting "positives" for the presence of microsatellite loci and the sequences of their flanking regions (Zane *et al.*, 2002). Briefly, genomic DNA is first reduced into smaller fragments enzymatically or mechanically. After optionally selecting for fragments 100–1000 bp in size, the fragments are then ligated

into vectors and mobilized into competent cells of *E. coli*. The cells are plated and allowed to grow into colonies, producing a genomic DNA library. Probes of synthetic microsatellite oligonucleotides, end-labeled with either radioactive or non-radioactive (e.g., digoxigenine or biotin) tags are allowed to hybridize to blots of the library. The blots are then tested via appropriate radioactive, chemiluminescent or colorimetric methods to screen for the presence of microsatellite-bearing genomic DNA fragments. The "positive" fragments are then sequenced to confirm the presence of microsatellite loci and their flanking regions.

A variety of this method using radioactive-labeled probes was used by van der Velde *et al.* (2000) to isolate microsatellite loci from *Polytricum formosum*. A similar method, following the method of Estoup & Turgeon (1996) using digoxigenine-labeled probes was tested in the early stages of this project. The method proved unsuccessful as none of the clones sequenced were found to have microsatellite loci. The method was abandoned as more efficient methods became available.

Another protocol for the isolation of microsatellite loci is by PCR using microsatellite primers (similar to ISSRs) followed by genome walking (Siebert *et al.*, 1995). Briefly, genomic DNA is first digested by restriction endonucleases. Oligonucleotides of known sequence (linkers or adapters) are ligated to the ends of the fragments to provide priming sites for PCR. In separate parallel processes, amplicons generated by ISSR PCR are sequenced to find half of the flanking regions of microsatellite loci. Primers are designed to anneal to these halves. These primers are used in PCR setups with the linker-ligated genomic DNA fragments as templates and oligos complementary to the linkers used as reverse primers. If any amplicons are amplified by these reactions, they are sequenced to find the other half of the target microsatellite loci. This method was also briefly tried in this project. It was found to be promising but was slow and inefficient. It was also abandoned in favor of a technique that was ultimately used in this project.

Protocols based on screening enriched genomic libraries for microsatellite loci are currently the popular methods for microsatellite marker development. This is a variation of the first protocol described earlier. However, by using genomic libraries enriched for microsatellites, many more "positive" colonies are found even when far fewer colonies are screened. A variety of this technique, the protocol of Hamilton *et al.* (1999) [and its updated, more detailed version, Hamilton (2002)] was used in this project to isolate microsatellite loci from the moss species studied. The protocol is detailed in subsequent sections of this report.

The development of microsatellite markers was attempted throughout the course of this project for four species of mosses: *Acanthorrhynchium papillatum*, *Pogonatum cirratum* ssp. *macrophyllum*, *Thuidium plumulosum* and *Thuidium cymbifolium*. These mosses have different growth habits, habitat preferences and sexuality. It was an overly ambitious early objective of this project to develop microsatellite markers for each of these species and to use them in studies of population genetic variability. Although microsatellite markers were successfully developed for *A. papillatum* by the author of this report, development of microsatellite markers for *P. cirratum* ssp. *macrophyllum* were done mostly by honours students. Microsatellite loci were isolated for both *T. plumulosum* and *T. cymbifolium* but were never fully developed as markers. Studies on both *Thuidium* species were abandoned owing to the lack of samples (despite extensive searches in the field) needed to conduct tests on marker polymorphism and the time, effort and resources they took away from studies on *A. papillatum*.

The development of microsatellite markers for *Acanthorrhynchium papillatum* was published in a paper, Leonardía *et al.* (2006), in the journal, Molecular Ecology Notes. A copy of this paper is found in Appendix I.

2.2 Materials and Methods

Two microsatellite-enriched genomic DNA libraries, designated as APMS and APBMS, were constructed separately and independently for *Acanthorrhynchium papillatum*. The first library, APMS, yielded few usable markers necessitating the construction of a second. Some alterations to the protocol were made in the building of the second library, APBMS, to ensure that enough markers were generated for the genotyping aspect of the project. These differences are elaborated on in Section 2.2.11 on page 47. Other aspects of the protocol were identical for each library, differing only in the source of the moss samples.

2.2.1 Field collection and processing of moss samples

Large amounts of clean moss tissues were needed for the construction of the microsatellite-enriched genomic DNA libraries. For each library, one large clump (approximately 250 cm^2 in area) of *Acanthorrhynchium papillatum* with minimal dead and moribund tissues was selected in the field. The APMS library was constructed from a sample collected from Sungei Bantang Recreational Forest in Johore State, Peninsular Malaysia, while the APBMS library was constructed from a sample collected in Cameron Highlands, Pahang State, Peninsular Malaysia (Figure 2.1 and Table 2.1). Moss samples were collected by pulling and scraping the moss shoots carefully from their substrates, enveloping them in prepared paper packets and sealing them in large Ziploc® bags. Upon return to the laboratory, the samples were kept at 4 °C until further processing.

The moss samples were made as axenic as possible prior to DNA extraction to reduce the probability of contaminating the microsatellite-enriched genomic DNA libraries with foreign DNA. Each moss shoot was inspected under a dissecting microscope, both to confirm its identity and to clean it of foreign material. Debris, fragments of the substrate, dead or moribund tissue as well as any epiphytic fungal

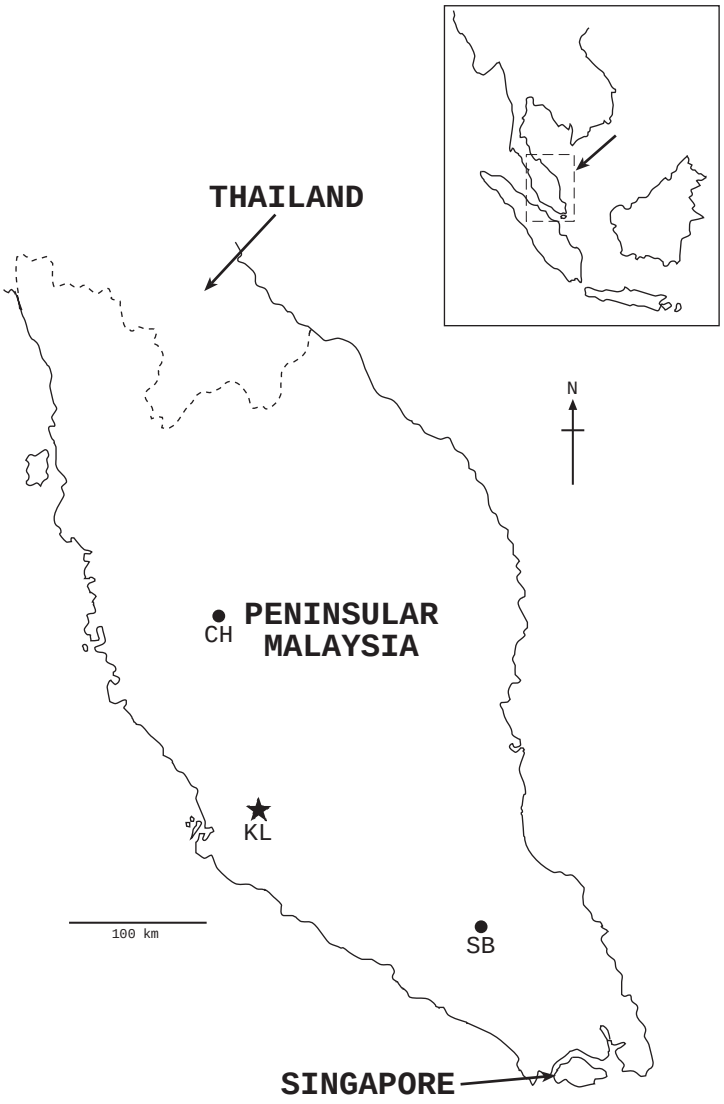


Figure 2.1: Collection localities of *Acanthorrhynchium papillatum* samples used in the construction of microsatellite-enriched genomic DNA libraries: `SB` – Sungei Bantang Recreational Forest, collection locality of sample used in constructing APMS; `CH` – Cameron Highlands, collection locality of sample used in constructing APBMS. `KL` – Kuala Lumpur is shown here for reference. A distance bar is indicated. Inset shows location of Peninsular Malaysia in relation to mainland Southeast Asia. (Maps courtesy of Reuben Clements).

Table 2.1: *Acanthorrhynchium papillatum* samples used in the construction of microsatellite-enriched DNA libraries. SBRF – Sungei Bantang Recreational Forest; CH – Cameron Highlands.

Library	Sample dry mass (mg)	Sampling area	Collection date
APMS	223	SBRF	22 April 2003
APBMS	831	CH	12 October 2003

or plant material were manually removed using fine forceps. The superficially clean tissues were then vigorously agitated in a dilute solution of Tween®20 (Sigma, USA) to further loosen any additional adhering material. The tissues were rinsed with Milli-Q® (Millipore, USA) water until they ran clear. They were then laid out on clean paper towels and allowed to air-dry until the tissues were brittle.

Early experiments with surface sterilization proved unsuccessful as the moss tissues themselves were bleached within 4–5 minutes of exposure to 8% commercial bleach solution. There would have been no way to chemically destroy contaminating DNA from exogenous sources using bleach without damaging genomic moss DNA since even a brief exposure to sodium hypochlorite would have compromised the downstream utility of the target DNA for use as templates in PCR (Prince & Andrus, 1992). The steps detailed in the preceding paragraph were used to compensate for this limitation and proved adequate, presumably since manual cleaning decreased the contribution of foreign DNA to minimal levels even as the total amount of DNA –both moss and foreign– was gradually reduced through the steps required to create the library.

2.2.2 Genomic DNA extraction

Genomic DNA for the construction of each library was extracted from pooled moss tissues prepared and cleaned as described in the preceding section.

A modified protocol of Doyle & Doyle (1990) was used to extract DNA from the moss samples. Approximately 500 mg of dry, brittle tissues were snap-frozen with liquid nitrogen in a chilled mortar. The tissues were ground with a chilled pestle to a fine powder, scraped from the mortar and pestle, and evenly distributed into 4, 6 or 8 microfuge tubes containing 500 μL preheated (60 °C) CTAB extraction buffer [2% CTAB (Sigma, USA) (w/v), 1.4 M NaCl (Sigma, USA), 20 mM Na_2EDTA (Sigma, USA) (pH 8.0), 100 mM Tris-HCl (J.T. Baker, USA and Merck, Germany)

(pH 8.0), freshly added 0.2% (v/v) 2-mercaptoethanol (Sigma, USA)]. Multiple tubes were used to increase the efficiency and yield of extraction. The tubes were incubated at 60 °C in a heat block for 30 min with occasional shaking. Following incubation, proteins were denatured and extracted by gentle mixing with an equal volume of 24:1 (v/v) chloroform:isoamyl alcohol (Merck, Germany). Phases were separated by centrifugation at 5000 *g*. The aqueous phases were pipetted out and transfered to clean microfuge tubes. DNA strands were precipitated from the aqueous phase by adding two-thirds volume of cold isopropanol. The resulting strands of DNA were pelleted by centrifugation at 5000 *g*. The alcoholic supernatant from each tube was discarded and the DNA pellets were washed with 1 mL cold 70% ethanol (Merck, Germany). Following the wash, the DNA pellets were recollected by centrifugation at maximum speed. The supernatant from each tube was pipetted out and the pellets were allowed to dry. The dry pellets were dissolved in 50 μL TE buffer [10 mM Tris-HCl (J.T. Baker, USA and Merck, Germany), 1 mM EDTA (Sigma, USA)]. DNA samples were pooled together.

DNA quality and yield were checked by electrophoresis through a 1.5% agarose gel in 1×TAE buffer [40 mM Tris -acetate (J.T. Baker, USA and Merck, Germany), 1 mM Na_2EDTA (Sigma, USA), pH 8] in the presence of ethidium bromide (Sigma, USA) and through OD readings at 260 nm and 280 nm using a DU® 600 spectrophotometer (Beckman Coulter, USA).

The construction of microsatellite-enriched genomic DNA libraries followed the protocol of Hamilton et al. (Hamilton, 2002; Hamilton *et al.*, 1999). A modification of this protocol, DeWoody (2002), was also used where indicated.

2.2.3 Genomic DNA digestion

Approximately 20 μg genomic DNA was digested in a 50 μL reaction volume with 10 U *Alu*I, 20 U *Hae*II, 10 U *Nhe*I, 1×Buffer 2 and 1×BSA (all enzymes and

supplementary reagents from New England Biolabs, NEB, USA). After approximately 3 h of incubation at 37 °C, the mixture was spiked with a further 5 U *Alu*I, 10 U *Hae*II and 5 U*Nhe*I and left to incubate overnight at 37 °C.

The enzymes were deactivated at the end of the overnight digestion by incubation at 65 °C in a heat block for 20 minutes.

Completeness of digestion was checked by running a 5 μL aliquot through a 1.5% agarose (Seakem, USA) gel in 1×TAE buffer in the presence of ethidium bromide (Sigma, USA).

2.2.4 Trimming and dephosphorylation of fragment ends

The digested genomic DNA fragments were made blunt-ended to allow for the ligation of the SNX-linkers that are part of Hamilton's protocol. This was done by mixing in 10 U mung bean nuclease (or MBN, from NEB, USA) to the deactivated digestion reaction, transfering the reaction mixture to a 0.2 mL, thin-walled reaction tube and incubating the mixture at 30 °C for 30 min with the aid of a PTC-100® (MJ Research, USA) thermocycler.

Following treatment with mung bean nuclease, the blunt-ended fragments were purified. Enzyme, buffer and the trimmed ends were removed by using a QIAQuick® Gel Extraction Kit (QIAGEN, USA) following the manufacturer's protocol for DNA cleanup of enzymatic reactions. The blunt-ended fragments were eluted with 53 μL of preheated (60 °C) elution buffer.

To help prevent the trimmed fragments from ligating to each other, the fragment ends were dephosphorylated with calf intestinal alkaline phosphatase (CIP). A 60 μL mixture of the blunt-ended fragments, 1×Buffer 2 (NEB, USA) and 10 U CIP (NEB, USA) was incubated at 37 °C for 2.5 h. Following incubation, the dephosphorylated fragments were purified using the same kit and protocol as for purification following MBN treatment. In this case, however, the dephosphory-

lated fragments were eluted with 30 μL of preheated (60 °C) elution buffer.

Recovery of the fragments following MBN and CIP treatment was checked by running a 1 μL aliquot through a 1.5% agarose gel in 1×TAE buffer in the presence of ethidium bromide (Sigma, USA).

2.2.5 Ligation of SNX linkers to MBN/CIP-treated fragments

SNX forward and SNX reverse oligos (Table 2.2) were ordered from MWG Biotech, Germany. Each oligo was dissolved in TE buffer to a final concentration of 100 μM. Equal volumes of forward and reverse oligos were combined, heated, and slowly cooled to create a 50 μM solution of SNX linkers.

SNX linkers were ligated to the MBN/CIP-treated genomic DNA fragments in the following reaction setup: A 30 μL reaction of 1×Buffer 2 (NEB, USA), 1×BSA (NEB, USA), 19.5 μM SNX linkers, 1 μM rATP (Sigma, USA), 20 U *Xmn*I (NEB, USA), 2000 U T4 DNA ligase (high-concentration, NEB, USA) and 10 μL MBN/CIP-treated genomic DNA fragments were combined together in a thin-walled, 0.2 mL reaction tube. The mixture was treated to 30 cycles of 30 min at 16 °C and 10 min at 37 °C using a PTC-100® (MJ Research, USA) thermocycler. The ligation-digestion reaction was terminated by a final incubation for 20 min at 65 °C using the same instrument.

Success of the ligation-digestion reaction was determined through PCR amplification using SNX forward as the single-species primer and the ligation-digestion reaction as template. A mixture of the components in the preceding paragraph minus the enzymes comprised the negative control template. Each template was used in the following PCR reaction setup: A 50 μL mixture of 0.8 μM SNX forward oligo, 0.8 mM dNTPs (Promega, USA), 1×ThermoPol buffer (NEB, USA), 0.6 U Vent_R® exo(−) polymerase (NEB, USA) and 10 μL template were mixed in a 0.2 mL, thin-walled reaction tube. The tubes were treated to the following

Table 2.2: Oligos used in the construction of microsatellite libraries. SNX forward and reverse oligos are from Hamilton (2002); Hamilton *et al.* (1999). Note that the SNX reverse oligo is phosphorylated at the 5' end. T3 and T7 are from Stratagene (2003).

Oligo name	Sequence (5'–3')
SNX forward	CTAAGGCCTTGCTAGCAGAAGC
SNX reverse	pGCTTCTGCTAGCAAGGCCTTAGAAAA
T3	AATTAACCCTCACTAAAGGG
T7	GTAATACGACTCACTATAGGGC

temperature profile in a PTC-100® (MJ Research, USA) thermocycler: initial denaturation for 5 min at 95 °C followed by 40 cycles of 45 s at 95 °C, 1 min at 62 °C, 1 min at 72 °C. The reaction mixture was held at 16 °C until recovery.

Results were checked by running aliquots of each reaction through a 1.5% agarose (Seakem, USA) gel in 1×TAE buffer in the presence of ethidium bromide (Sigma, USA).

2.2.6 Enriching for microsatellite-bearing fragments

Synthetic, biotinylated probes were mixed with the linker-ligated fragments to enrich the mixture for microsatellites. Hybridization reactions were carried out in several separate experiments for each library; some with all the probes together in one reaction, others with the probes in separate reactions. No distinct advantages in the results were found in either case but having the probes together in one reaction was, naturally, more convenient. The protocol detailed here is for the probes used together.

A 100μL mixture of 90 nM each of 3'-biotinylated oligos $\texttt{(GA)}_{15}$, $\texttt{(GT)}_{15}$, $\texttt{(CTC)}_{10}$, $\texttt{(CAGA)}_{8}$ (MWG Biotech, Germany), 6×SSC [0.9 M NaCl (Sigma, USA), 90 mM tri-sodium citrate (BDH, UK), pH 7.0], 0.1% SDS (Sigma, USA) and 14 μL of the ligation mixture were were combined in a 0.2 mL, thin-walled reaction tube. The mixture was placed in a PTC-100® thermal cycler and treated to the following temperature regime according to the protocol of DeWoody (2002): 10 min at 95 °C, 1 min at 75 °C, 10 min at 70 °C, slow ramp (−0.2 °C per 10 s) to 50 °C, 10 min at 50 °C, slow ramp (−0.5 °C per 10 s) to 40 °C, hold at 4 °C until recovery.

While the hybridization reaction proceeded, MagPrep® streptavidin beads (Novagen, Germany) were prepared to capture the hybrids. Twenty microliters of the bead suspension was aliquoted into a clean microfuge tube. The tube was

placed in a Magnetight^{TM} Multitube Rack (Novagen, Germany) to take the beads out of suspension. The supernatant was pipetted out. The beads were washed by adding 500 μL of binding and washing buffer [10 mM Tris-HCl (J.T. Baker, USA and Merck, Germany), pH 7.5, 1 mM Na_2EDTA (Sigma, USA), 1 M NaCl (Sigma, USA)], gently vortexing the tube to resuspend the beads, placing the tube in the rack, and pipetting out the supernatant. The beads were washed a total of four times with this procedure. After the final wash, the beads were resuspended in 250 μL of binding and washing buffer.

The resuspended beads were mixed together with the completed hybridization reaction. The reaction tube containing the mixture was taped to the axis of the rotisserie of a Hybaid Shake 'n' Stack hybridization oven (Thermo Electron Corporation, USA). The mixture was incubated for 2.5 h at 43 °C to allow the beads to capture the hybrids with the rotisserie set at 20 rpm to keep the beads from settling.

At the end of the capture period the tube was placed in the Magnetight^{TM} rack to separate the beads and the captured hybrids from the suspension. The supernatant was discarded.

The beads were then treated with the following washes: twice at room temperature with 500 μL 2×SSC, 0.1% SDS, twice at 45 °C with 500 μL 1×SSC, 0.1% SDS and twice at 60 °C or 65 °C (see Section 2.2.11) with 500 μL 1×SSC, 0.1% SDS. (SSC and SDS were made from the same components that were used to make the hybridization buffer). Each wash was conducted by resuspending the beads by gentle vortexing in the specified solution, incubating the beads for 5 min at the specified temperature, magnetizing the beads and finally discarding the supernatant.

Following the washes, the beads were resuspended in 80 μL of elution buffer [10 mM Tris-HCl (J.T. Baker, USA and Merck, Germany), 0.1 mM Na_2EDTA

(Sigma, USA)] and incubated for 10 minutes at 95 °C. The beads were then magnetized and the eluted, single-stranded, microsatellite-enriched DNA fragments were transferred to a clean microfuge tube.

2.2.7 PCR amplification of fragments enriched for microsatellites

The eluted, single-stranded, microsatellite-enriched DNA fragments were made double stranded and amplified following the same PCR reaction detailed earlier: A 50 μL mixture of 0.8 μM SNX forward oligo, 0.8 mM dNTPs (Promega, USA), 1×ThermoPol buffer (NEB, USA), 0.6 U Vent_R® exo(−) polymerase (NEB, USA) and 10 μL eluted fragments were mixed in a 0.2 mL, thin-walled reaction tube. The mixture was treated to the following temperature profile in a PTC-100® (MJ Research, USA) thermocycler: initial denaturation for 5 min at 95 °C followed by 40 cycles of 45 s at 95 °C, 1 min at 62 °C, 1 min at 72 °C. The reaction mixture was held at 16 °C until recovery.

The products of four such reactions were pooled together to increase the amount of material available for subsequent processing. The two hundred microliters of total reaction mixture were split into two tubes of 100 μL each. One milliliter of ice-cold, pure ethanol (Merck, Germany) and 44 μL of 3 M sodium acetate, pH 4.2 (Sigma, USA) were added to each tube to precipitate the products. The tubes were kept for about 15 min at −20 °C before the products were pelleted down by centrifugation at 14 000 *g* for 10 min. Each pellet was washed twice by resuspension in 500 μL 70% ethanol (Merck, Germany) and centrifugation at 14 000 *g* for 2 min, discarding the used wash solution each time. The pellets were air-dried at the end of washes until the ethanol completely evaporated. The products were then concentrated by dissolving each pellet in 43.5 μL of elution buffer [10 mM Tris-HCl (J.T. Baker, USA and Merck, Germany), 0.1 mM Na_2EDTA (Sigma, USA)].

2.2.8 Trimming SNX linker ends of the amplified fragments

The SNX linkers of the amplified, microsatellite-enriched fragments had to be trimmed to prepare them for insertion into the cloning vector.

A 50 μL reaction of 1×Buffer 2 (NEB, USA), 1×BSA (NEB, USA), 10 U *Nhe*I (NEB, USA) and 43.5 μL of the amplified, concentrated microsatellite-enriched fragments were combined in a fresh microfuge tube. The digestion mixture was incubated for about 3.5 h at 37 °C. The trimmed fragments were purified using a High-Pure PCR Purification Kit (Roche, Germany) following the manufacturer's protocol. The fragments were eluted with 50 μL elution buffer.

2.2.9 Insertion of trimmed fragments into cloning vector

The trimmed fragments were ligated to *Xba*I-digested, CIP-treated (both enzymes from NEB, USA) pBluescript® II SK+ Phagemid (Stratagene, USA) cloning vector for cloning and eventual sequencing.

A 20 μL reaction of 10 U *Nhe*I (NEB, USA), 400 U T4 DNA ligase (NEB, USA), 100 nM rATP (Sigma, USA), 1×Buffer 2 (NEB, USA), 1×BSA (NEB, USA) and equimolar ends of the trimmed fragments and the prepared vector were mixed together in a 0.2 mL, thin-walled reaction tube. The mixture was placed in a PTC-100® (MJ Research, USA) thermocycler and treated to 30 cycles of 30 min at 16 °C, 10 min at 37 °C. The enzymes were then deactivated by a 20 min incubation at 65 °C before being soaked at 16 °C until recovery.

2.2.10 Transformation of competent cells and culture of transformants

Competent cells of JM109 (Promega, USA) or DH5α (Life Technologies, USA) were transformed with the prepared plasmids using standard lab procedures. Briefly: 1–5 μL of the ligation-digestion mixture were gently mixed with 100 μL partially thawed competent cells and kept on ice for 30 min. The mixture

was then heat-shocked for 90 s at 42 °C. The cells were allowed to recover on ice for 4–5 min. Six hundred milliliters of lukewarm Luria-Bertani (LB) broth were slowly added to the cells which were then allowed to recover for 45 or 90 min (see Section 2.2.11) at 37 °C in an orbital shaker.

The recovered cells were gently pelleted down at 2000 *g* for 5 min and enough medium was removed to concentrate the cells to a volume of around 100 μL. Twelve microliters of a 100 μM solution of isopropyl-β-D-thiogalactopyranoside (IPTG, Promega, USA) and 20 μL of 122.5 mM solution of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal, Bio-Rad, USA) were mixed into the cell suspension. The mixture was then spread on LB plates with ampicillin (50 μM, Duchefa, the Netherlands) and left to incubate overnight at 37 °C.

Multiple cultures (about 5–7) were plated to increase the number of colonies to pick and screen.

2.2.11 Differences in protocol between APMS and APBMS libraries

Although the differences in the protocol used in the construction of the libraries, APMS and APBMS, are slight, the results suggest that they may be significant and worth presenting. Table 2.3 summarizes these differences.

2.2.12 Screening colonies for inserts

Well-isolated white and pale-blue colonies were picked using sterile pipette tips. Each colony was resuspended in 7 μL sterile, Milli-Q® (Millipore, USA) water. Aliquot of this suspension were used as templates in PCR reactions used to screen for inserts. Twenty microliter reactions of 200 nM each of T3 and T7 primers (Table 2.2, MWG Biotech, Germany), 200 μM dNTPs (Promega, USA), 1×F-501 buffer (FinnZymes, Finland), 0.8 U DyNAzyme™ II recombinant DNA polymerase (FinnZymes, Finland) and 2 μL of bacterial suspension were combined in

Table 2.3: Protocol differences between APMS and APBMS libraries.

Library	Temperature of final wash (°C)	Incubation time following heat shock (min)
APMS	65	90
APBMS	55	45

0.2 mL, thin-walled reaction tubes. The tubes were placed in a PTC-100® (MJ Research, USA) thermocycler and treated to the following temperature profile: initial denaturation for 2 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 52 °C, 1 min at 72 °C, followed by a final elongation and tailing step of 10 min at 72 °C and a soak at 16 °C until recovery.

Reaction products were checked for the presence of an insert by electrophoresis through a 1.5% agarose gel in 1×TAE buffer in the presence of ethidium bromide (Sigma, USA).

2.2.13 Culture and miniprep of colonies with inserts

Colonies with inserts (i.e., with PCR products longer than 150 bp) were cultured by inoculating the remaining 5 μL of bacterial suspension into 5 mL sterile LB broth with ampicillin (50 μM, Duchefa, the Netherlands) and incubating the culture overnight at 37 °C in an orbital shaker.

Plasmids from the overnight cultures were harvested using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega, USA) following the manufacturer's protocol without deviation.

2.2.14 Sequencing inserts

The inserts were sequenced using BigDye® Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems, USA).

Cycle sequencing reactions were set up according to the manufacturer's protocol and were performed on either a PTC-100® (MJ Research, USA) thermocycler, a GeneAmp® PCR system 2400 (Applied Biosystems, USA), a GeneAmp® PCR system 2700 (Applied Biosystems, USA), or a GeneAmp® PCR System 9600 (Applied Biosystems, USA) using the following thermal profile: Initial denaturation for 1 min at 96 °C, followed by 25 cycles of 10 s at 96 °C, 5 s at 50 °C, 4 min at

60 °C. Extension products were kept at 4 °C until recovery.

Extension products were purified according to the protocol of Applied Biosystems and were sequenced in a PRISM® 3100 Genetic Analyzer (Applied Biosystems, USA) equipped with a 50 cm capillary array running POP-6™ polymer, using default run and analysis modules. Bases were called using the instrument's default basecaller.

Each insert was sequenced from both T3 and T7 ends.

2.2.15 Contig assembly and identification of microsatellite-bearing inserts

Forward and reverse contigs of each insert were assembled manually. Linker and cloning vector sequences were also identified and excised manually from the sequences of the original genomic DNA fragments.

Inserts were screened for microsatellite sequences manually and with the program Tandem Repeats Finder (Benson, 1999).

Microsatellite-bearing sequences were compared with each other (to check for duplicate sequences) and with sequences in GenBank (Bilofsky & Burks, 1998) using BlastN (Altschul *et al.*, 1990) (to exclude the probability of having obtained contaminating sequences).

2.2.16 Primer design and testing

Primers flanking microsatellite sequences were designed with the aid of the programs GeneRunner (1994, version 3.00, Hastings Software) and Primer3 (Rozen & Skaletsky, 2000). Oligos were ordered from two local companies, Research Biolabs and 1st BASE Pte Ltd.

Testing for consistency of amplification

Primers pairs were first optimized and tested to amplify products consistently when given templates from different samples. (Details on the field sampling and extraction of genomic DNA from these samples are found in subsequent parts of this thesis).

Amplifications were carried out in 10 μL volumes of 1×F-501 buffer (FinnZymes, Finland), 250 μM dNTPs (Promega, USA), 1 μM each of forward and reverse primers, 0.4 U DyNAzyme^{TM} II recombinant DNA polymerase (FinnZymes, Finland) and approximately 10 ng genomic DNA as template using a PTC-100® (MJ Research, USA) thermocycler. The thermal profile used was: initial denaturation for 2 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at annealing temperature, 1 min at 72 °C, followed by a final elongation and tailing step of 10 min at 72 °C and a soak at 16 °C until recovery.

PCR products were checked by electrophoresis through 1.5% agarose gels in 1×TAE buffer in the presence of ethidium bromide (Sigma, USA).

Primer pairs that consistently amplified products when given templates from different moss samples were checked for length-polymorphism of products.

Testing for length-polymorphism of products

The length of polymorphisms of products from consistently amplifying primer pairs were tested by electrophoresis through either polyacrylamide gels or MetaPhor® High Resolution Agarose (Cambrex, USA).

Denaturing polyacrylamide gel electrophoresis of PCR products were performed using a Sequi-Gen GT Sequencing Cell (Bio-Rad, USA). Six percent polyacrylamide [19:1 acrylamide:bis-acrylamide, Bio-Rad, USA), 8 M urea (Sigma, USA) gels were cast 0.4 mm thin and run with 0.5×TBE [45 mM Tris-borate (J.T. Baker, USA and Bio-Rad, USA), 1 mM Na_2EDTA (Sigma, USA), pH 8]

buffer at constant power (60 W) and constant temperature (50 °C). All other electrophoresis parameters followed Bio-Rad's protocol. Following electrophoresis, the polyacrylamide gels were stained with silver following the protocol of Caetano-Anollés & Gresshoff (1994).

Three percent Metaphor® High Resolution Agarose (Cambrex, USA) gels were prepared and cast according to the manufacturer's protocol. Electrophoresis were performed in 0.5×TBE in the presence of ethidium bromide (Sigma, USA).

Length-polymorphic microsatellite markers were submitted to GenBank (Bilofsky & Burks, 1998).

Optimization of multiplex PCR reactions

Primer pairs exhibiting consistent amplification and length polymorphism of products were combined in multiplex PCR reactions to increase the efficiency of genotyping and to decrease errors in matching DNA samples and PCR amplicons.

Although multiplex PCR combinations of the primer pairs were done by trial and error, the combinations were governed by several considerations:

1. As many primer pairs as possible that would yield reliable and robust data were put together in the same multiplex reaction.

2. The Applied Biosystems DNA sequencer that was used was set up so that only 4 fluorescent dyes could be used per sample. Since a (ROX™) dye-labeled size standard had to be co-injected with the PCR products, only three dyes, 6-FAM™, HEX™ and NED™, could be used to label the primers.

3. Primers in the same multiplex reaction with amplicons of potentially overlapping size ranges had to be labeled with different flourescent dyes.

4. Primers that were labeled with the same dye had to be designed so that their amplicons did not overlap throughout their size range.

5. Finally, since the multiplex combinations had to be optimized while the primers were still unlabeled, primer pairs were combined so that amplicons in the same multiplex reaction did not overlap and could be easily discerned when run in a MetaPhor® gel.

After determining the final combination of primer pairs to be used in multiplex reactions, the concentrations of the primers needed to be optimized to more evenly amplify the different targets. Further optimization of PCR conditions was done following the recommendations of Henegariu *et al.* (1997).

Following successful optimizations of multiplex PCR combinations and conditions, half of each primer pair was dye-labeled with either 6-FAM™ (1st BASE, Singapore), HEX™ (1st BASE, Singapore) or NED™ (Applied Biosystems, USA) such that primers in the same multiplex reaction amplified products that were discernible from each other during subsequent fragment analysis runs on the Applied Biosystems DNA Sequencer.

2.2.17 Characterizing the microsatellite markers

Genotyping of a number of samples of *Acanthorrhynchium papillatum* was performed to characterize the developed markers. Details on the selection of the moss samples and the extraction of their genomic DNA are found in subsequent chapters.

Ten microliter reactions of 1×F-501 buffer (FinnZymes, Finland), 5% DMSO (Sigma, USA), 200 μM dNTPs (Promega, USA), 0.25 μM or 0.5 μM forward primer, 0.25 μM or 0.5 μM reverse primer, 0.4 U DyNAzyme™ II DNA polymerase (FinnZymes, Finland) and approximately 10 ng genomic DNA were combined in thin-walled, 0.2 mL tubes. Reactions were performed in a PTC-100®

(MJ Research, USA) thermocycler using the following thermal profile: Initial denaturation for 5 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 30 s at 58 °C, 2 min at 65 °C, and final extension for 30 min at 72 °C (the same annealing temperature was used for all primer pairs tested). Products were kept at 16 °C until recovery. (Details on these PCR conditions and specific primer combinations are discussed in the results).

Amplification products were purified using CleanSEQ® kits (Agencourt, USA) following the manufacturer's protocol.

Size separation and detection of the purified products were performed in an ABI PRISM® 377 DNA Sequencer (Applied Biosystems, USA) following the protocol of Applied Biosystems for fragment analysis applications. ROX™-labeled GeneScan™400HD (Applied Biosystems, USA) was used as the internal size standard.

Alleles were determined manually from fragment sizes called by the program GeneScan® version 3.1.2 (Applied Biosystems, USA). The number of alleles found for every locus was counted manually. Nei's gene diversity per locus, h, (also called heterozygosity) (Nei, 1987, 1973) was computed with the aid of the spreadsheet program Excel® (Microsoft, USA).

2.3 Results

2.3.1 Construction of microsatellite-enriched libraries

Similar results were obtained in the steps leading to the construction of both APMS and APBMS libraries. To avoid redundancy and in the interests of both brevity and clarity, only one gel electropherogram from each of these steps will be presented from either APMS or APBMS libraries where needed for illustrative purposes.

Genomic DNA extraction yielded 110 μg from the sample used for construction

of the APMS library (Figure 2.2A) and 400 μg from the sample used for the APBMS library. Extracted DNA was of sufficient quantity and purity for the construction of microsatellite-enriched libraries.

Overnight digestion of the genomic DNA resulted in a smear, the tail end of which lay beyond 10 000 basepairs in size despite the use of three endonucleases. Much of the smear, however, appeared evenly distributed from 100 to 2000 bp (Figure 2.2B).

Trimming with MBN and CIP should result in a barely-visible downward size-shift in the smear of digested genomic DNA (Figure 2.2C). Hence, the electropherogram resulting from this reaction is similar to that of the digested genomic DNA. Despite this similarity, however, gel-electrophoresis of the results of these processes was still performed to estimate the concentration of DNA in the solution to correctly set up the subsequent SNX-linker ligation.

Success of the ligation of SNX linkers to the treated genomic DNA fragments is demonstrated in Figure 2.3A. Expectedly, there were no amplicons present in the negative control. The amplified fragments in the experimental lane (still in a smear) extended to just past 1000 bp. Besides the copious primer dimers evident in both lanes, there was also a noticeable bias for smaller fragments in the experimental lane.

A similar electropherogram resulted after enriching the fragments for microsatellites and amplification of the enrichments via PCR (Figure 2.3B). Despite the use of identical reaction mixtures (except for the template) and the same PCR thermal profile, there are three things noticeably different from the earlier PCR check on ligation success: the absence of primer dimers, the absence of a bias for smaller fragments and faint but perceptible discrete bands.

Trimming the ends of the SNX linkers from the enriched fragments should once again produce a barely-visible downward shift in the size of the fragments. Gel

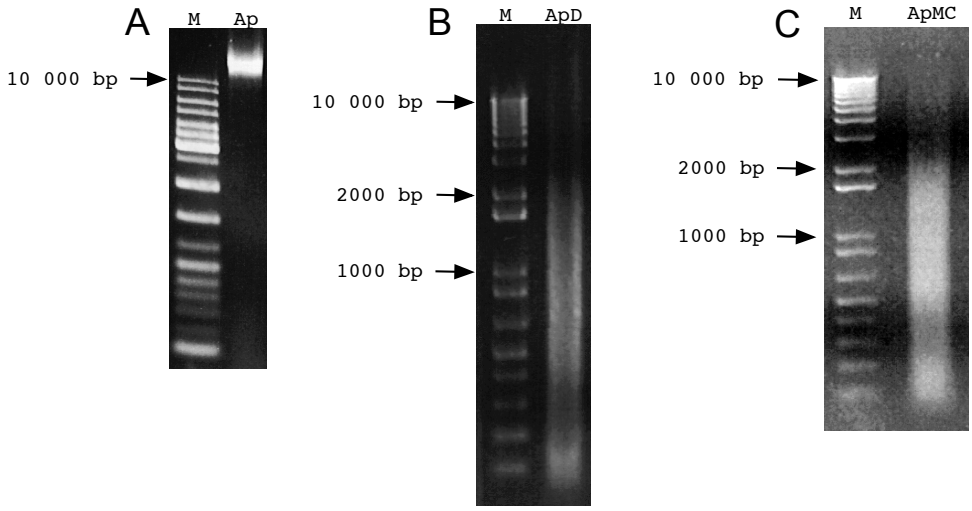


Figure 2.2: Genomic DNA of *Acanthorrhynchium papillatum* used for constructing a microsatellite-enriched library. (A) `Ap` – undigested genomic DNA. (B) `ApD` – digested genomic DNA. (C) `ApMC` – MBN/CIP trimmmed, digested genomic DNA fragments. `M` – ladder. All samples were resolved in 1.5% (w/v) agarose in 1×TAE buffer.

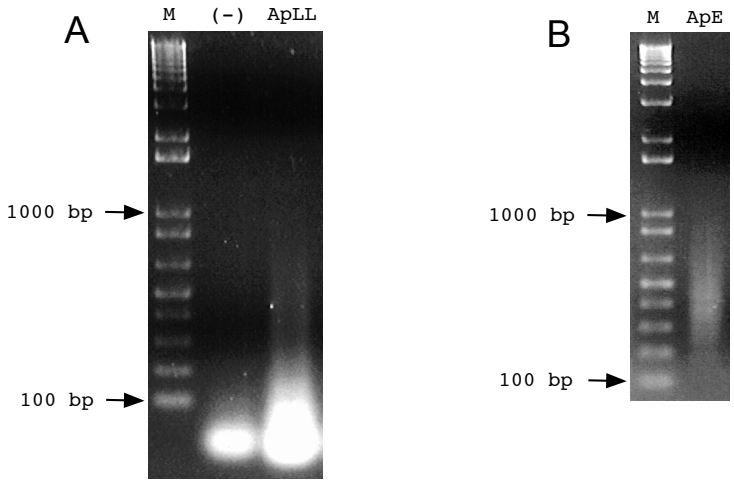


Figure 2.3: Linker-ligated and enriched genomic DNA fragments. (A) `ApLL` – PCR-amplified, SNX linker-ligated genomic DNA; `(-)` – negative control. (B) `ApE` – Genomic DNA fragments enriched for microsatellites. `M` – ladder. All samples were resolved in 1.5% (w/v) agarose in 1×TAE buffer.

electrophoresis of this step was not performed, however, to retain more material for ligation of these fragments to the cloning vector. Results of this subsequent ligation were likewise not checked through electrophoresis for the same reason. Instead, as is routine in the lab, successful ligation was inferred through successful mobilization of the produced plasmids into the competent cells.

Relatively few numbers of pure white colonies were obtained despite having multiple plates from both libraries. Initially, only white colonies were picked and screened for inserts. Pale blue colonies were eventually screened, however, and were also found to have inserts. No correlations between the color of the colony and the presence of inserts were sought and therefore found. Figure 2.4 shows one of several gel electropherograms used to screen for inserts.

2.3.2 Differences between the APMS and APBMS libraries

One hundred and sixty-nine plasmids in total were sequenced from both libraries. Despite the initial screening, one sequenced plasmid (from the APMS library) was found lacking an insert. Seventeen of the plasmids sequenced were essentially failed reactions with no discernible sequences. Proportionally more unique sequences were found in the APBMS library (93/95) compared to the APMS library (47/56). However, proportionally much more of the APMS library were found to bear microsatellites (44/47) compared to the APBMS library (22/93). All these microsatellite-bearing inserts had no significant homology with sequences in GenBank, indicating that the sequences had not been reported and were specific to *Acanthorrhynchium papillatum*. Of these inserts, proportionally more of the APBMS library had enough flanking regions for primer design (16/22) compared to the those found in the APMS library (13/44). Similar proportions of amplifying primers were found in both APMS (3/13) and APBMS (5/16) libraries. All of the amplicons were found to be length-polymorphic when tested across a range

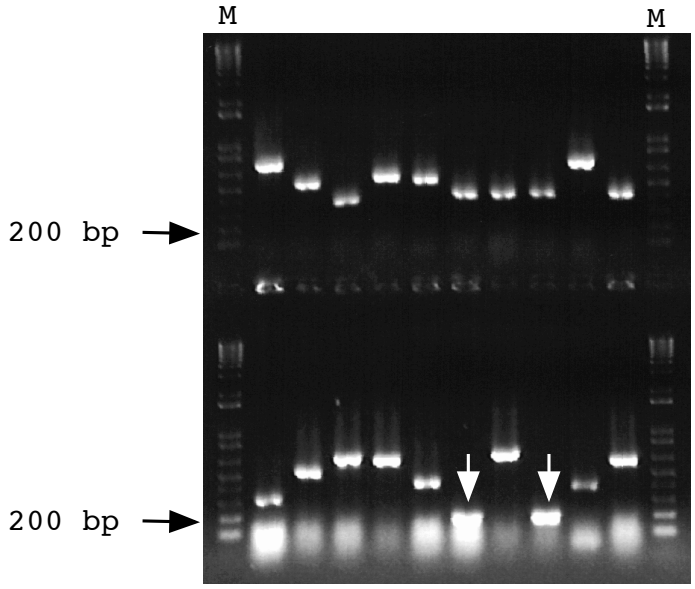


Figure 2.4: Screening of twenty samples for inserts. Ten samples were loaded in the top half of the gel, ten others were loaded in the bottom half. White arrows indicate samples without inserts. [1.5% (w/v) agarose in 1×TAE buffer]. M – ladder.

of samples (Figure 2.5). Table 2.4 summarizes these findings. Inserts or loci were named from the library and serial number of the colony bearing the genomic DNA fragment in which they were found. Appendices A–E on pages 220–253 list the sequences of the inserts as they are categorized (e.g., "dirty", replicated, without repeats, etc.).

2.3.3 Types of microsatellite loci isolated

Despite interrogating the genomic DNA of *Acanthorrhynchium papillatum* with four different microsatellite probes [to repeat: $\texttt{(GA)}_{15}$, $\texttt{(GT)}_{15}$, $\texttt{(CTC)}_{10}$, $\texttt{(CAGA)}_{8}$], nearly all of the microsatellite loci found were of $\texttt{(GA)}_n$ and $\texttt{(GT)}_n$ motifs. Only one locus, APBMS105, was found to have a $\texttt{(CTC)}_n$ motif. No loci bearing a $\texttt{(CAGA)}_n$ motif was found. Interestingly, two loci were found to bear motifs that were not probed or sought. APBMS56 was found with a few repeats of `(AGAGTG)`, while APBMS13 was found with a few repeats of `(AAAGAAAAGGA)`.

Following the nomenclature of Weber (1990), all three categories of microsatellites were found: "perfect" (e.g., APMS14), "imperfect" (e.g., APMS70, APBMS93) and "compound" (e.g., APMS65). Regardless of which category they belong, many of the loci found were long, greater than 20 repeats of the base motif.

Interestingly, the resulting 8 candidate microsatellite markers that amplified polymorphic products were primarily of $\texttt{(GA)}_n$ motifs. For simplicity, they were given the same name as their inserts or loci. Thus, the microsatellite markers developed for *Acanthorrhynchium papillatum* were named, in sequence, APMS4, APMS14, APMS28, APBMS3, APBMS14, APBMS23, APBMS61 and APBMS72.

2.3.4 Construction of multiplex PCR sets

On top of the considerations already stated that governed the PCR multiplexing of primer pairs (page 52), the characteristics of the candidate markers added a few

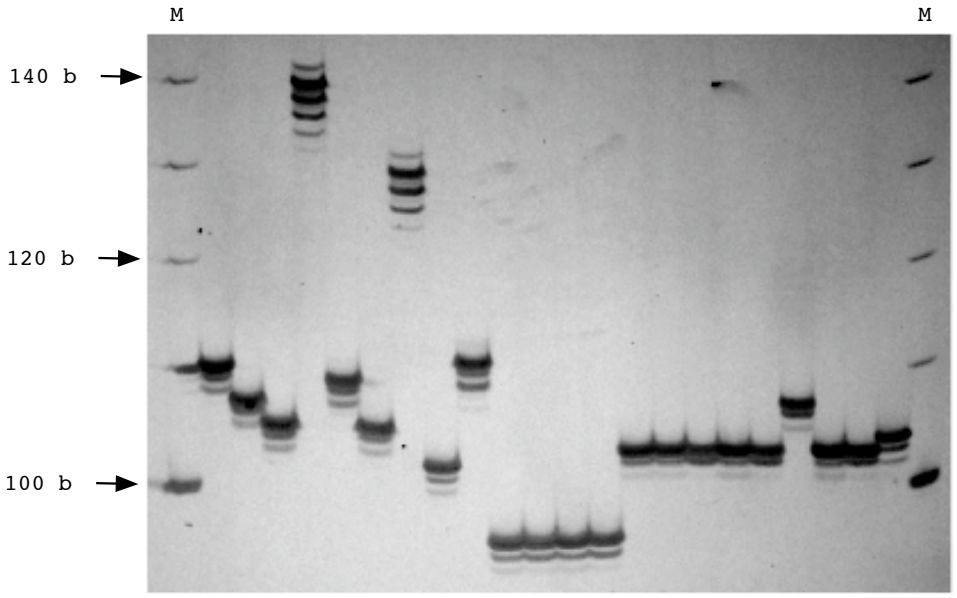


Figure 2.5: Electropherogram of PCR products of primers designed from the insert APMS28 tested on 22 samples of *Acanthorrhynchium papillatum*. Note stutter bands characteristic of microsatellite loci. (Silver-stained 6% denaturing polyacrylamide gel in 0.5×TBE buffer). `M` – ladder.

Table 2.4: Attrition of microsatellite-enriched inserts.

Library	Plasmids sequenced	(+) Inserts	Clean sequences	Unique	(+) Microsatellites	(+) Flanks	Amplifying primers	Polymorphic products
APMS	60	59	56	47	44	13	3	3
APBMS	109	109	95	93	22	16	5	5
Total	169	168	151	140	66	29	8	8

more. Some of the candidate primers were from inserts that were too short (for example, APBMS14) or had limited places for primer redesign. Moreover, preliminary analysis also revealed that some primers amplified products that varied widely in size. Both these limitations meant that the amplicons in a multiplex set had to be discernible mostly by being labeled with different dyes, not by having products that were in different size ranges. However, two primer pairs could be redesigned to amplify consistently longer fragments that didn't overlap in size with any of the other smaller amplicons. They could therefore be labeled with a dye that was also used in one of the smaller amplicons in a multiplex set.

Given all these considerations in combining primer pairs for multiplex PCR it was determined that at most only 4 primer pairs could be combined in the same reaction. At least 2 multiplex reactions were therefore needed to genotype samples with the 8 markers developed.

Notwithstanding these limitations, primers amplifying these 8 loci were amenable to different combinations of PCR multiplexing (see Figure 2.6 for an example). The two multiplex PCR sets finally chosen are described in Table 2.5.

2.3.5 Evaluation of markers through preliminary genotyping

The genotypes of ninety-eight *Acanthorrhynchium papillatum* samples from three populations were used to further characterize and evaluate the candidate markers. Screenshots of representations of the raw data are found in Figures 2.7 and 2.8.

Alleles and sample genotypes are presented in Appendix G. Table 2.6 summarizes the primer characteristics derived from the data.

Four to twenty-six alleles per locus (with an average of 11.5) were found from the genotypes of the 98 samples. Each locus showed high values of Nei's gene diversity (0.578 to 0.936) indicating both the large number of alleles found and the even spread of alleles in the samples tested. High values of gene diversity also

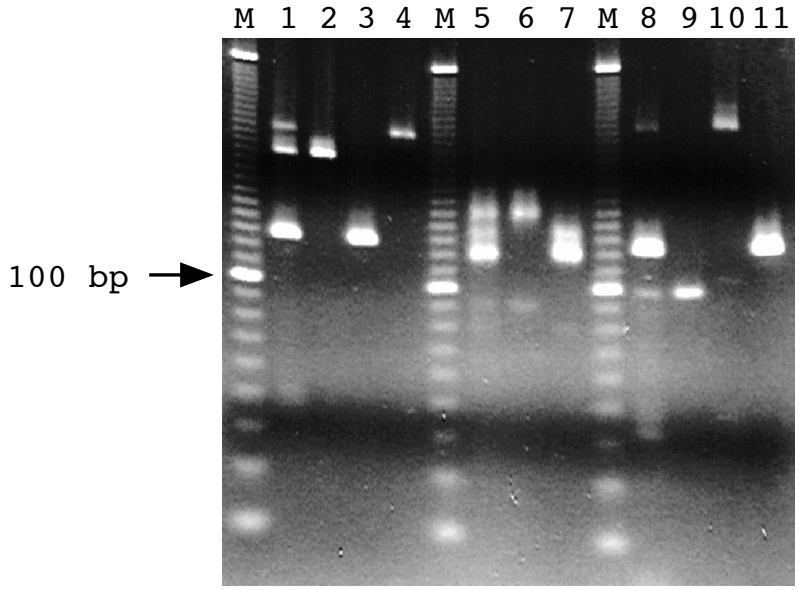
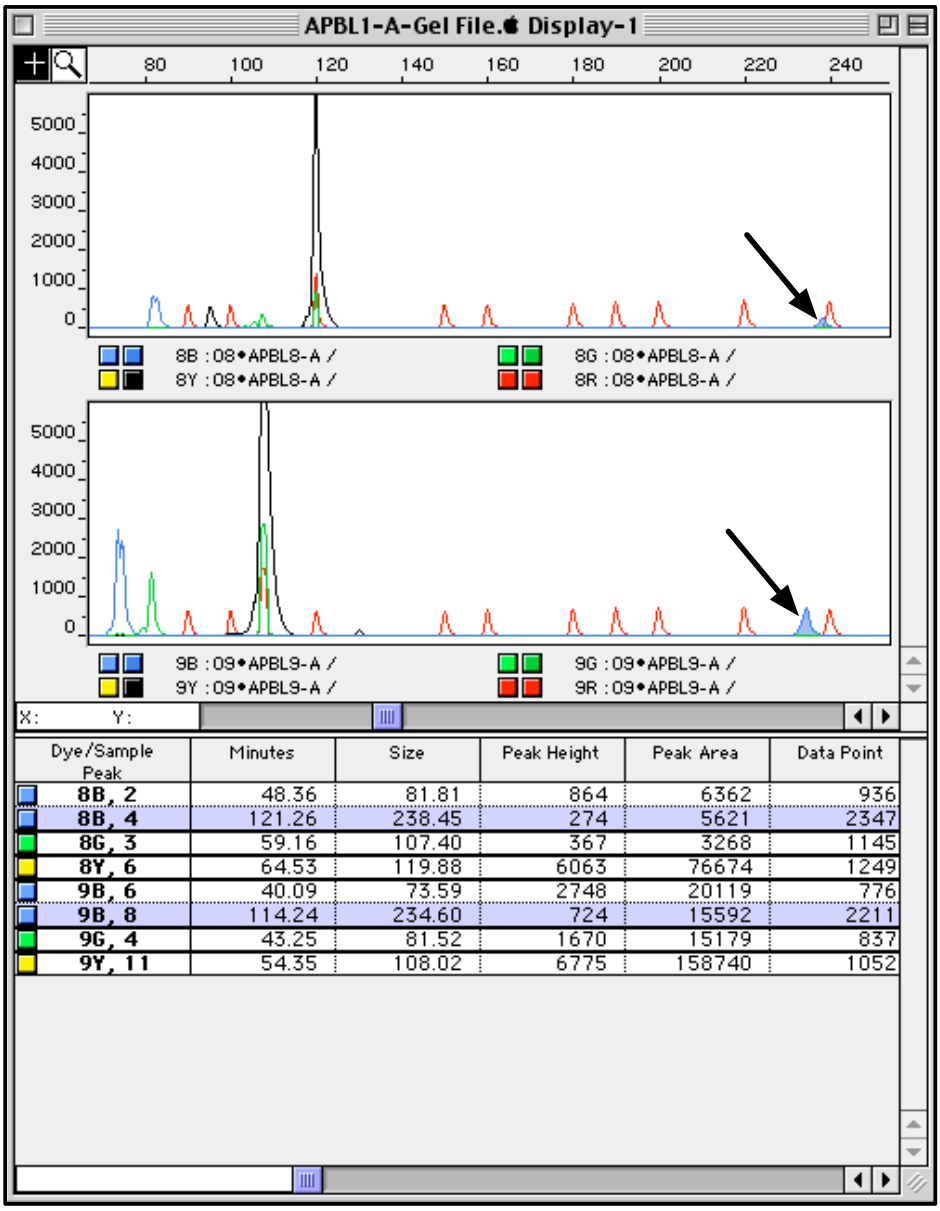


Figure 2.6: Example of multiplex PCR amplifying microsatellite loci in *Acanthorrhynchium papillatum*. M – marker; 1 – multiplex PCR of APMS4, APBMS23 and APBMS61; 2 – APMS4; 3 – APBMS23; 4 – APBMS61; 5 – multiplex PCR of APMS14, APMS28; 6 – APMS14; 7 – APMS28; 8 – multiplex PCR of APBMS3, APBMS14 and APBMS72; 9 – APBMS3; 10 – APBMS14; 11 – APBMS72.

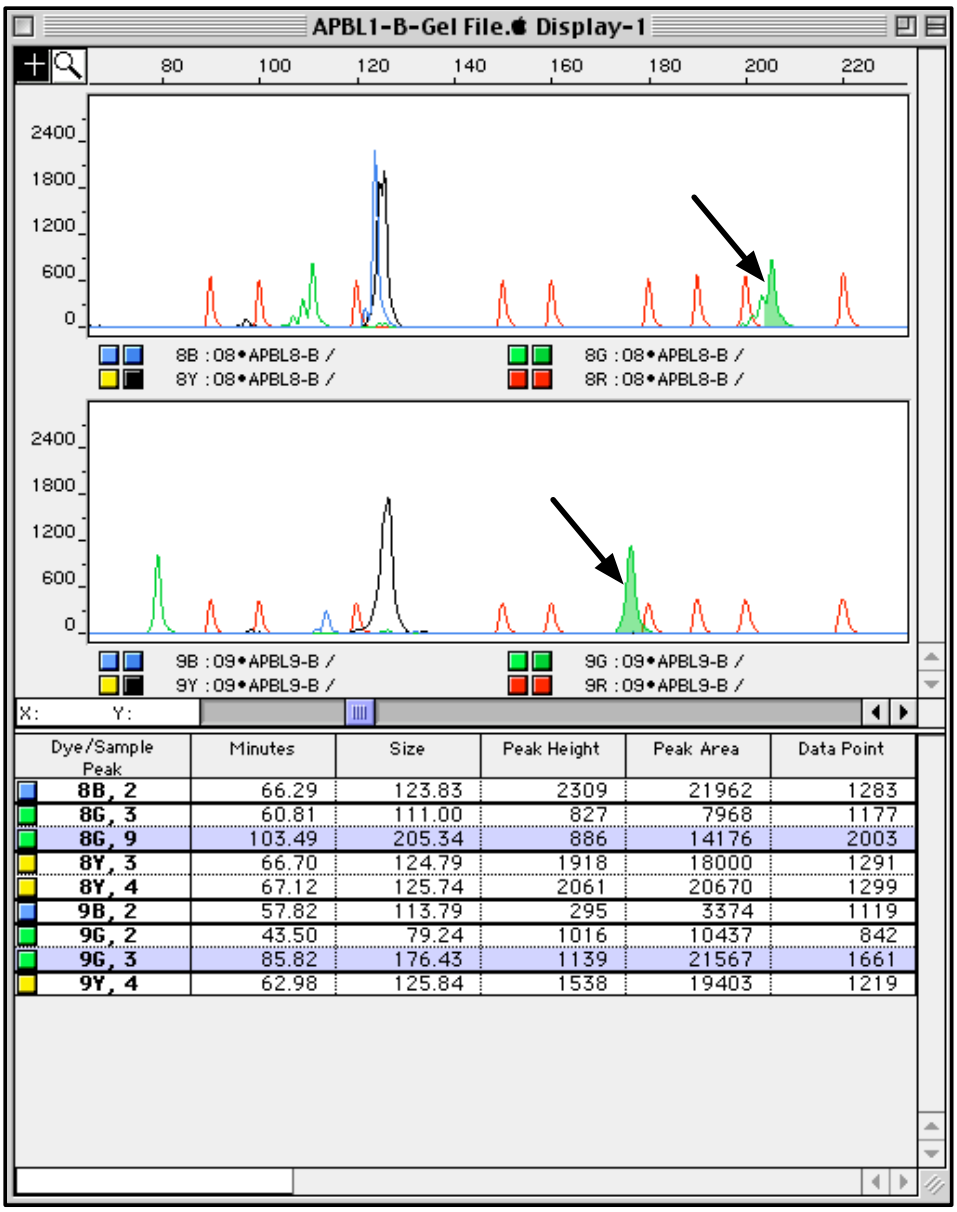
Table 2.5: Multiplex PCR sets for genotyping *Acanthorrhynchium papillatum* samples using microsatellite markers. Forward and reverse primers for the same locus were in equal concentrations (1×=0.5 μM). However, primer concentrations for different loci in the same multiplex PCR set were adjusted so that the amplicons more evenly fluoresce. Note that in each set, primers that are labeled with the same fluorescent dye amplify products in different size ranges.

Multiplex set	Locus	Fluorophore dye label	Relative primer concentration	Size range (bp)
A	APMS14	6-FAM™	1×	67–126
	APMS28	HEX™	1×	79–156
	APBMS23	NED™	0.5×	108–130
	APBMS61	6-FAM™	1×	233–239
B	APMS4	HEX™	1×	166–217
	APBMS3	6-FAM™	1×	92–164
	APBMS14	NED™	0.5×	94–158
	APBMS72	HEX™	1×	69–111



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
8B, 2	48.36	81.81	864	6362	936
8B, 4	121.26	238.45	274	5621	2347
8G, 3	59.16	107.40	367	3268	1145
8Y, 6	64.53	119.88	6063	76674	1249
9B, 6	40.09	73.59	2748	20119	776
9B, 8	114.24	234.60	724	15592	2211
9G, 4	43.25	81.52	1670	15179	837
9Y, 11	54.35	108.02	6775	158740	1052

Figure 2.7: Screen capture of electropherograms of multiplex PCR set A of two samples, APBL8 and APBL9. Dye-labeled DNA fragments are represented by colored peaks. Colors of the peaks indicate the dye label of the fragment: ROX^{TM}=red, 6-FAM^{TM}=blue, HEX^{TM}=green, NED^{TM}=black. The two electropherograms are oriented so that peaks of the same size (base pair length) are aligned vertically. Fragment sizes in base pairs are indicated on the combined x-axis at the top of the graphs. Y-axes indicate relative flourescence intensity. Red peaks are size standards (GeneScan^{TM}400HD). Blue peaks on the left of the graph are APMS14 loci. Green peaks are APMS28 loci. Black peaks are APBMS23 loci. Blue peaks indicated by arrows are APBMS61 loci (these loci also highlighted in the table). Table indicates various measurements, including fragment size in basepairs. Note differences in amplicon sizes as seen in the position of the peaks and as seen in the table.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
8B, 2	66.29	123.83	2309	21962	1283
8G, 3	60.81	111.00	827	7968	1177
8G, 9	103.49	205.34	886	14176	2003
8Y, 3	66.70	124.79	1918	18000	1291
8Y, 4	67.12	125.74	2061	20670	1299
9B, 2	57.82	113.79	295	3374	1119
9G, 2	43.50	79.24	1016	10437	842
9G, 3	85.82	176.43	1139	21567	1661
9Y, 4	62.98	125.84	1538	19403	1219

Figure 2.8: Screen capture of electropherograms of multiplex PCR set B of two samples, APBL8 and APBL9. Graph and table descriptions similar to Figure 2.7. Green peaks on the left of the graph are APBMS72 loci. Blue peaks are APBMS3 loci. Black peaks are APBMS14. Green peaks indicated by arrows are APMS4 loci (also highlighted in the table).

Table 2.6: Characteristics of *Acanthorrhynchium papillatum* microsatellites tested on 98 gametophytic samples from three populations. T_a = optimal annealing temperature; A = number of alleles; h = Nei's gene diversity.

Locus	Primer sequence (5'–3')	Repeat motif in clone	T_a	Size range (bp)	A	h	GenBank accession no.
APMS4	HEX™-CCTCCTACCAGCATTGAAAG ACCTGGGGAGGATTATAAATTCAG	$(GA)_{22}$	58	166–217	15	0.736	DQ092745
APMS14	6-FAM™-ATATTAACCTCGCTCGACG CAGCTCGAGATCCTTCTCTC	$(GA)_{46}$	58	67–126	11	0.852	DQ092744
APMS28	HEX™-CACTTTGCCGACTACAGTTG GTCCAGGTCTGATCACGTAAC	$(GA)_{23}$	58	79–156	26	0.936	DQ092746
APBMS3	GCTCTCTTTCGTTTGTTCAG 6-FAM™-CAATCGGCTTGCAATTATAC	$(GA)_{20}$	58	92–164	19	0.874	DQ092747
APBMS14	NED™-CGAAGCCGGTGGTAAGTCC CGGTCCTCTCTCCTCCAAAT	$(GA)_3(G)_4(GA)_9$	58	94–158	11	0.763	DQ092749
APBMS23	NED™-TCGAAATCGTCAGTACTTTCCAG ATTCACGAGATCCTCTGAGTTCG	$(GA)_4GT(GA)_5$	58	108–130	6	0.736	DQ092750
APBMS61	AACTGAAGTTCCTCCACTCTTTTC 6-FAM™-AGGTGCTTCTTAGAATCTTGTTGC	$(CA)_4(TC)_8CTT(C)_6TC(CT)_4$	58	233–239	4	0.578	DQ092751
APBMS72	HEX™-CAATCTCTTGCGTTTGCTTG AGAAGAGCGTCCACACTTGG	$(GA)_{30}$	58	69–111	11	0.602	DQ092748

suggest the utility of the markers for studies in genetic variation of populations. While Weber (1990) found that the longest uninterrupted, "perfect" repeats predicted greater informativeness of the polymorphism (or number of alleles) of loci, this is only partially supported by the results obtained.

2.4 Discussion

Although the major objective of this thesis was to investigate the population-level diversity of *Acanthorrhynchium papillatum*, a moss native to Southeast Asia, the process of the development of microsatellite markers for this moss also yielded interesting findings.

2.4.1 Library construction

The original protocol of Hamilton *et al.* (1999) called for the screening of bacterial colonies for the presence of microsatellites. This was accomplished through the onerous processes of colony lifts, membrane hybridization and subsequent chemiluminescent detection. These steps were followed in the earliest attempts in this project, only that colorimetric instead of chemiluminescent detection methods were used. However, having no positive and negative controls –this being the first microsatellite library constructed in the laboratory– made it difficult to optimize the hybridization and detection conditions. It also made interpreting the colorimetric results subject to ambiguity. What was particularly hard was deciding how well-developed the color reactions should be before the colony could be considered positively bearing microsatellites and therefore worth sequencing. Finally, while it was found that the colonies most positively reactive to the screening, i.e., the ones that gave the deepest colored spots in the colony lifts, did bear microsatellite loci, the loci were so long that the flanking sequences were too short for primer design. Therefore, the most "positive" microsatellite loci found by the screening,

although technically positive, were useless as microsatellite markers. Screening, it seemed, only hindered the development of microsatellite markers. Alternatives had to be sought.

These counter-productive findings led the researcher to try sequencing inserts from white colonies without first screening them for the presence of microsatellite repeats. Initial attempts were positive enough that the screening steps were finally abandoned. The final results show that even without colony screening, chances were better than 1/3 that a random insert-bearing colony would be found with a microsatellite repeat.

Differences between the two libraries

In fact, chances were closer to 3/4 when only the first attempt at a library, the construction of APMS, is considered. This is one of several differences observed between the two libraries, APMS and APBMS. This and others differences are discussed below.

Revisiting Table 2.4, one difference that bears discussion is the percentage of unique inserts in the two libraries. The large difference in the proportion of unique inserts in the APBMS (93/95=0.98) and APMS (47/56=0.84) libraries may be attributed to the different incubation times following the transformation of competent cells. In the construction of the APMS library the bacterial cells were incubated for 90 min as is routinely practiced in the laboratory. It is speculated that this extended incubation period not only allows transformed cells to recover from the stress induced by the heat-shock, but also allows them enough time to multiply. This is useful for researchers working on a single species of insert and when transformation efficiency is low: potentially more colonies may be obtained from fewer transformed cells. For researchers building a library however, it increases the chances of finding replicated inserts. In the construction of APBMS

the incubation period was reduced by half to 45 min and a larger proportion of unique inserts was observed.

The difference in temperature of the final stringent washes used in constructing the two libraries is thought to have contributed to two differences in the results.

More microsatellite loci, both in absolute and relative measures, were found in the APMS library compared to the APBMS library. It is hypothesized that only genomic DNA fragments that had strong complementarity to the probe sequences remained bound to the probes during the final stringent washes used in the construction of the APMS library. These are fragments with significant stretches of microsatellite repeats. Fragments that were only weakly bound to the probes, i.e., those without or with only short microsatellite repeats, melted away from the probes in the higher temperatures of the washes. On the other hand, the lower temperatures used in the washes of the APBMS library may have allowed fragments not bearing microsatellite repeats to remain bound to the probes. These fragments eventually made their way into inserts and plasmids and caused the reduction of the proportion of microsatellites found in the APBMS library. This, in turn, implied that, to obtain microsatellite loci from this library, many more inserts needed to be sequenced than would be necessary in a more enriched library.

Although efficiently isolating microsatellite loci using hotter stringent washes might seem desirable, they also had familiar counter-productive effects. Many of the microsatellite loci found in the APMS library had flanking sequences that were so short that primers could not be designed to amplify these loci. Many of these inserts had microsatellite repeats that abutted directly against the linker sequences and had no flanking sequences in one or both ends. The situation was expectedly reversed in the APBMS libary. Sequences suitable for primer design abounded at the ends of many inserts. However, having no or few microsatellite loci to amplify rendered many of these sequences useless.

Whether these differences are statistically significant was not tested. Nor were correlations between the seemingly slight differences in protocol (recall Table 2.3) and these differences checked. The proposed explanations are hypotheses that still need confirmation. The libraries were constructed to make tools for the study of the biology of *Acanthorrhynchium papillatum*. Time and resources had to be devoted to other aspects of the project and could not be spent sorting out these interesting matters that do not directly affect the main objectives of the project. However, it is hoped that these ideas may be useful for those who are keen on creating their own microsatellite-enriched libraries and may be following a similar protocol.

Keys to finding suitable microsatellite loci

Using proper parameters in seemingly trivial parts of the protocol are keys to efficiently obtaining microsatellite loci that can at least be tested to see if they can be amplified from genomic DNA. Incubating transformed cells for the right amount of time spells the difference between risking not finding enough colonies with inserts to sequence (not giving the cells enough time to recover from the stress of the heat shock) and wasting sequencing efforts on replicated inserts, as explained above. Performing stringent washes at the proper temperature determines whether long stretches of microsatellite loci with little or no flanking sequences are obtained, or short or no microsatellite loci are found.

Unfortunately these parameters may depend on other factors, like the type or transformation efficiency of the competent cells used or the nature of the species being studied. Moreover, whether ideal parameters have been correctly chosen cannot be known until many inserts from the library have already been sequenced. If it turns out that the parameters were not ideal, two choices present themselves. The first choice would be spending two or more weeks constructing another library

with slightly different parameters (and still not knowing if these new parameters are ideal). The second choice would be to sequence more inserts from the present library in the hope of finding unique and suitable microsatellite loci from among the redundant sequences or inserts with no or unusable microsatellite loci. Either choice offers no guarantees. What choice to make would depend on the careful assessment of time and material resources. The choices are not mutually exclusive, however, and both could be attempted to maximize the chances of finding suitable microsatellite loci.

Utility of short, cryptic microsatellite loci

Another observation was made that might be useful when developing microsatellite markers. While searching for microsatellite loci to amplify, focus was understandably on longer, unambiguous stretches of microsatellite repeats. As more and more of these loci failed to amplify despite repeated PCR attempts, attention was paid to shorter, more cryptic microsatellite loci in the hope that they would generate usable markers. The results of this refocus were that two loci, APBMS14 and APBMS23, earlier ignored for having shorter, cryptic repeats, turned out to be suitable markers possessing large amounts of polymorphism when tested across a range of samples. Having learned the lesson of not ignoring these kinds of inserts, more of them were tested. No other inserts, however, could be developed into markers.

2.4.2 Marker characterizations

Effects of "replicated" samples

As is typical in the development of microsatellite markers, difficulties do not end with the isolation of microsatellite loci. In fact, problems can arise even after these loci have been determined to be polymorphic by testing with a small number of

samples.

Because of time limitations in this project, experiments could not be conducted in a strictly ordered fashion. For example, sample collections could not be postponed until after all the microsatellite markers had been developed. Instead, samples were collected while the markers were being developed and nearly all the samples were available for testing by the time all 8 markers were found. This meant that the characterization of the markers didn't have to be performed using only a limited set of samples; all the samples available could be used instead. After genotyping all available samples and examining the data, it was apparent that the results were going to introduce complications in evaluating the usability of the markers.

Many of the *Acanthorrhynchium papillatum* samples were found to share exactly the same alleles for each of the 8 markers tested, in effect contributing "replicated" samples to the pool of all available samples. Some samples were found to differ from others in only one or two out of the 8 markers tested. While these findings could be of important biological significance (more on this in succeeding chapters), these results played havoc with computations of certain marker statistics.

Among the statistics that are usually reported when developing microsatellite markers is linkage disequilibrium (Lewontin & Kojima, 1960; Robbins, 1918). Linkage disequilibrium is a measure of the non-random association of alleles of pairs of loci. When one allele of a certain locus is found to occur with an allele of another locus more often than should occur if the loci were independent of each other, then the two loci are said to be in linkage disequilibrium. Since many samples of *Acanthorrhynchium papillatum* were found with exactly the same loci, then linkage disequilibrium will, by definition, be detected among these loci (and need not even be computed). Unfortunately, it cannot be determined if the linkage

disequilibrium truly reflects the relationship of the loci or if it is an artifactual product of the clonal nature of the samples of *A. papillatum*.

Nei's gene diversity is also affected by the nature of the *Acanthorrhynchium papillatum* samples. This statistic describes how evenly represented the different alleles are in a given locus. A marker is useful in genetic studies when its different alleles have more or less even relative frequencies with no alleles dominating the locus. In testing the markers of *A. papillatum*, alleles of all the "replicated" samples would appear with greater relative frequency than if the alleles were contributed by only one representative sample. Gene diversity for each of the loci would therefore appear smaller (and the utility of the marker somewhat diminished) if the "replicated" samples were included in the computations than if they were not. Again, the smaller values of gene diversity that would be obtained if the "replicated" samples were included in the computations could represent the true nature of the markers or could be products of the nature of the *A. papillatum* samples. As with the linkage disequilibrium computations, there was no way of knowing which was the true case.

Ultimately it was decided to retain the "replicated" samples in the computations for Nei's gene diversity. Despite the higher frequency of the alleles of the "replicated" samples, gene diversity for the loci remain high. (Linkage disequilibrium didn't need to be computed since it would inevitably show linkage among all pairs of loci. For completeness, however, linkage disequilibrium was computed using the software package, Genepop Raymond & Rousset (1995b). Results of the computations are found in Appendix H)

Other statistics

Papers on the development of microsatellite markers often also report three other statistics: deviations from Hardy-Weinberg equilibrium, PIC or polymorphism

information content (Botstein *et al.*, 1980) and observed heterozygosity. These statistics are appropriate only for diploid samples and do not apply to the present study of haploid plant material.

Null alleles

Null alleles were encountered while genotyping samples for marker evaluation. (A null allele is the absence of a PCR product at a locus. The length of the microsatellite region that is null at a locus is therefore indeterminate). At the start of the experimental analysis, samples with null alleles were excluded from computations of marker statistics. Admittedly, in retrospect, this should not have been done. However, had these samples been included in the computations marker statistics would only have been slightly different from those presented. Exact values of Nei's gene diversity would have been different, but these values would also change when all samples are finally evaluated. Results on linkage disequilibrium would have been the same. The biggest effect would be on the perception of the utility of these markers given the tendency of some loci to produce null alleles. The oversight of excluding samples with null alleles has been corrected, however, and these samples have been included in analyses that will be seen and discussed in later chapters.

[The original data set has been retained in this part of the thesis since it is this "trimmed" data set that was used in a paper reporting the microsatellite markers of *Acanthorrhynchium papillatum* (Leonardía *et al.*, 2006).]

Deviations from marker repeat units

Microsatellites are length-polymorphic because the polymerases that duplicate microsatellite loci are thought to "slip" and create copies that are one or more repeat units shorter or longer than the original template (Jarne & Lagoda, 1996;

Schlötterer, 2000). Microsatellite loci in different individuals, therefore, differ in length by multiples of these repeat units. For example, if a microsatellite locus is composed of dinucleotide repeats, then one individual may have an amplicon length of 125 base pairs while another individual may have an amplicon length of 127 base pairs.

A look at the genotypes of the samples (Appendix G) used to characterize the markers reveals that many of the amplicons do not follow this pattern of differing by multiples of the repeat units of their microsatellite loci. That is, while the repeat units of these microsatellite markers are mostly dinucleotide motifs, some amplicons do not differ from others by multiples of 2 base pairs.

This observation could be explained by the action of a different microsatellite mutation mechanism, the insertion or deletion of bases in or around the microsatellite loci, a combination of these two processes, or a yet-undescribed and undetermined cause. The examination of the sequences of these amplicons could shed light on what exactly is taking place. Unfortunately, sequences of some of these repeat regions have been difficult to obtain as evidenced in the development of the microsatellite-enriched libraries and as seen in later sections of this chapter. Notwithstanding these deviations from expected amplicon lengths, alleles obtained from the developed microsatellite markers should still prove useful in studying the genetic diversity of *Acanthorrhynchium papillatum*.

2.4.3 Long microsatellites

As mentioned earlier, several microsatellite loci in *Acanthorrhynchium papillatum* were found to be very long (e.g., APMS21 with 124 GA repeats, APMS31 with 95 GA repeats and APMS54 with 123 GT repeats – see Appendix D). Repeat lengths in these loci are longer than many of those reported in recent literature on microsatellite marker development [e.g., (Högberg *et al.*, 2006; Johnson *et al.*,

2005; Lee *et al.*, 2006; Mayor & Naciri, 2007; Nagai *et al.*, 2006; Schmidt *et al.*, 2006)].

Many of these loci are the ones found with short or no flanking sequences (Appendix D). While these long microsatellites could be artifacts of the library construction process, the long lengths of some amplicons generated with the developed markers, e.g., sample APBT31 using marker APBMS23, suggest that these long repeats may really be found in the genome of *Acanthorrhynchium papillatum*. Several attempts to sequence long amplicons generated by the developed markers were made to confirm this idea. But the difficulty of the template, characteristic of repeat regions, rendered these attempts unsuccessful.

Although the presence of long microsatellite loci in *Acanthorrhynchium papillatum* could not be positively ascertained, these unsuccessful attempts have no bearing on the utility of the markers or the objectives of this research project. However, their confirmed presence in *Acanthorrhynchium papillatum* can offer insights into the genome and biology of this moss that may be useful in future investigations.

2.4.4 Comparison with other microsatellite-marker studies

How does the project to develop microsatellite markers for *Acanthorrhynchium papillatum* compare with studies to develop microsatellite markers for other plants?

A paper by Squirrell *et al.* (2003) elaborates on the efforts needed to develop microsatellite markers for plant species by comparing various statistics from numerous reports. One interesting finding in the paper is that there is little difference in finding unique microsatellite loci between screened and unscreened enriched libraries. This statistic further justifies the decision in this project to bypass the screening step. The paper also reports that in order to obtain 10 working microsatellite markers, inserts from an average of 60 colonies from an enriched library

need to be sequenced to identify 38 unique microsatellites to be able to design 20 primer pairs. This implies a mean attrition of about 36% going from the total number of inserts sequenced to those bearing microsatellite markers, of 46% going from the total number of microsatellite-bearing inserts to those for which primers can be designed, and of 50% going from loci for which primers could be designed to those than can be used as microsatellite markers. The results of the present project obviously do not compare to these averages (in the order just presented, attrition of 61%, 56% and 0% respectively). However, while the mean attrition at these steps are 36%, 46% and 50% respectively, the range of these figures from which the means are computed is very broad: 0–90%, 0–79% and 0–89% respectively. Moreover, deviations from the average of the attrition figures for this project may be explained by the sub-optimal conditions used during library construction as has already been proposed.

The construction of microsatellite-enriched libraries and the development of microsatellite markers are still far from being straightforward, sure-fire processes. While statistics can be generated for various stages of these processes, large variations in these figures are found even in recent reports on microsatellite marker development. Table 2.7 summarizes figures from some of these papers.

2.5 Conclusions

Despite the difficulties involved in first finding microsatellite loci and then characterizing these loci in *Acanthorrhynchium papillatum*, 8 microsatellite markers were developed for this species, fulfilling the first objective of the thesis. The protocol used to develop these markers proved efficient in finding microsatellite loci. Unfortunately, most of the loci found could not be used as molecular markers. Characteristics of these 8 markers suggest their suitability in population genetics studies, studies that form the other objectives of this research project. Microsatel-

Table 2.7: Statistics of some recent efforts to develop microsatellite markers for plants. "Colonies" refer to the number of colonies whose inserts were sequenced; where screening or ISSR techniques were employed, only inserts from the "positive" colonies were sequenced. "Positives" refer to the number of colonies confirmed by sequencing to have microsatellite inserts. "Working" refers to the number of microsatellite markers reported for the organism.

Report	Genus	Colonies	Positives	Working
Jones *et al.* (2004)	*Luehea*	screened	16	8
Lee *et al.* (2006)	*Koompassia*	165	53	24
Mayor & Naciri (2007)	*Aster*	174	99	8
Provan & Wilson (2006)	*Sphagnum*	ISSR	17	9
Schmidt *et al.* (2006)	*Macadamia*	not reported	60	33

lites were also isolated from three other species of mosses: *Pogonatum cirratum* ssp. *macrophyllum*, *Thuidium plumulosum* and *Thuidium cymbifolium*. For various reasons, these could not be fully developed as markers. In the process of isolating microsatellite loci and developing microsatellite markers many observations were made on the construction of libraries enriched for microsatellite loci. These observations, the suggested explanations to these findings and the proposed compensations are hoped to be useful to other researchers attempting similar investigations. Other observations on the characteristics of the moss species, *A. papillatum*, are hoped to spur further interest and investigations on the biology and genetics of this under-studied group of plants.

Chapter 3

GENETIC DIVERSITY AMONG CLUMPS OF *ACANTHORRHYNCHIUM PAPILLATUM* AS MEASURED BY MICROSATELLITE MARKERS

3.1 Introduction

While it would have been more systematic to first study the diversity within clumps of *Acanthorrhynchium papillatum*, because the newly-developed microsatellite markers had to be tested on samples from different clumps, diversity between and among clumps was studied first.

Notwithstanding the order of execution and presentation of the experiments performed, this chapter represents the *raison d'être* of the thesis project. Many of the major objectives that were set at the beginning of the project are answered in this chapter. Found in this chapter are the results of the first study to be done on the genetic variation of a moss native to Southeast Asia in more than a quarter of a century.

In this chapter, the newly-developed microsatellites markers were more exhaustively tested. With the availability of more samples, amendments to the statistics presented in the earlier chapter were made. More importantly, using these microsatellite markers, baseline diversity statistics on populations of *Acanthorrhynchium papillatum* were established. Diversity was measured at two levels: at the marker (per locus) level and at the multi-locus genotype level (as measured through unique combinations of microsatellite alleles). Diversity metrics of *A. papillatum* populations from sampling areas of different degrees of disturbance

were also compared to assess the effects of anthropogenic change on the genetic diversity of the moss. Genetic distances among the different populations were computed and analyzed to help characterize patterns of relationships among populations of *Acanthorrhynchium papillatum*. Finally, population differentiation and the presence of genetic structure were investigated to identify populations that were genetically isolated from others.

3.2 Materials and Methods

Many of the methods used here are also applicable in other parts of the project. Where these methods are re-used are indicated in relevant sections of subsequent chapters.

3.2.1 Sampling areas

Samples of *Acanthorrhynchium papillatum* (Harv.) Fleisch. were collected from three areas in Johore State, Peninsular Malaysia: Sungei Bantang Recreational Forest, Gunung Belumut Forest Reserve and Kota Tinggi Waterfalls Resort, and two areas in Singapore: Bukit Timah Nature Reserve and MacRitchie Reservoir. For brevity, these sampling areas may be referred to as SBRF, GBFR, KTP, BTNR and MACR, respectively in the remainder of this thesis. These areas differ qualitatively in their degrees of disturbance and the conditions of their forests. The sampling areas in Malaysia are discontinuous with each other. Although the sampling sites in Singapore are only a few kilometers apart, they are separated by an expressway which may make them functionally discontinuous.

Trail maps of the sampling areas were unavailable; however, map coordinates (latitude, longitude and elevation) of sampling areas are provided. These were determined using a handheld global positioning system (GPS) device (Garmin™ eTrex®) and the software program Google™ Earth version 4.01569.0. Collections

were made within a 2 km radius from the stated coordinates.

Sungei Bantang Recreational Forest

Location and description Sungei Bantang Recreational Forest (2°21' N, 103°9' E, 30 m above sea level) lies about 100 km north-northwest of Singapore and about 8 km from the nearest town, Bekok (Figures 3.1 and 3.2). It is continuous with the southwest corner of the much larger (80 000 ha) Endau-Rompin National Park (Peh *et al.*, 2005). Sungei Bantang Recreational Forest is moderately developed with two carparks, a small office building, two public toilets and a few huts that dot a small recreation area at the entrance of the forest. Although it is frequented by visitors for its streams and picnic areas, much of its surrounding vegetation beyond the recreation areas remains intact and undisturbed. Moreover, no people were encountered beyond the recreation areas during any of the sampling trips (or other visits) to this site, suggesting that few people venture away from the picnic sites.

Sampling trails Samples of *Acanthorrhynchium papillatum* were collected from the trails that run beside the streams and into the vegetation of Sungei Bantang Recreational Forest. These trails do not have names. Some samples were collected from the recreation grounds as well.

Gunung Belumut Forest Reserve

Location and description *Acanthorrhynchium papillatum* was also sampled from Gunung Belumut Forest Reserve (2°3' N, 103°31' E, 200 m above sea level) which is approximately 70 km north of Singapore (Figures 3.1 and 3.2). Gunung Belumut Forest Reserve is approximately 30 000 ha in area (Peh *et al.*, 2005). Gunung Belumut Forest Reserve has also never been logged (Peh *et al.*, 2005) and its periphery is also popular as a recreation site. It is also more developed

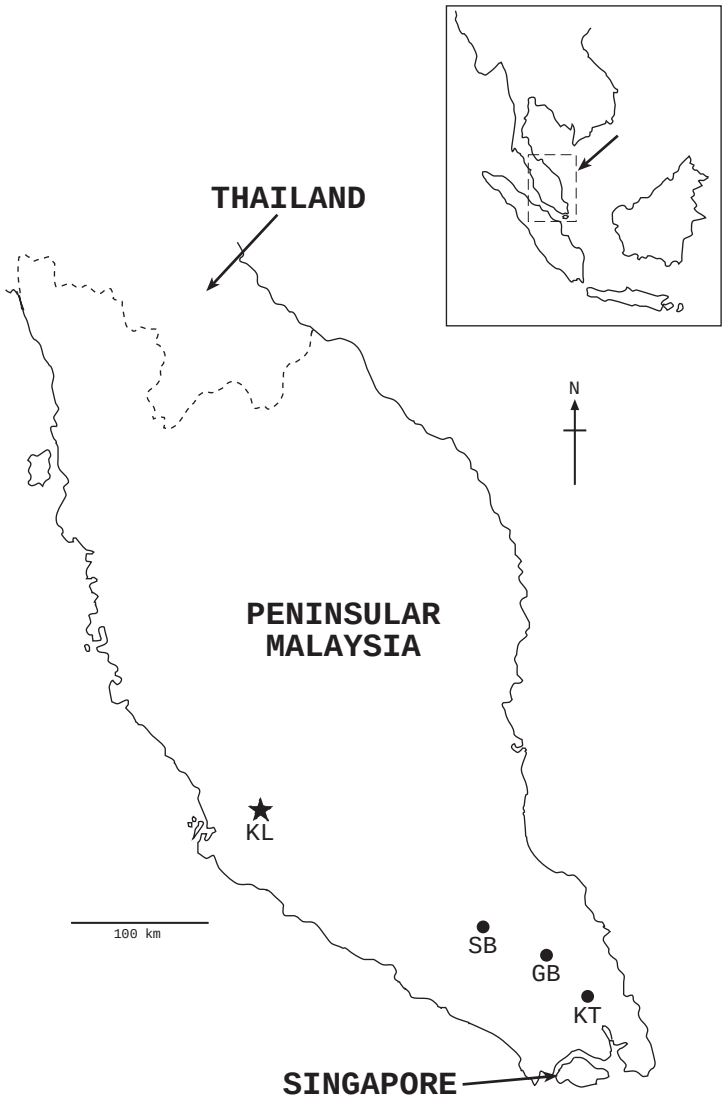


Figure 3.1: Location of sampling areas of *Acanthorrhynchium papillatum* in Malaysia: SB - Sungei Bantang Recreational Forest; GB - Gunung Belumut Forest Reserve; KT - Kota Tinggi Waterfalls Resort. Kuala Lumpur (KL) is shown for reference. A distance bar is indicated. Inset shows location of Peninsular Malaysia in relation to mainland Southeast Asia. (Maps courtesy of Reuben Clements).

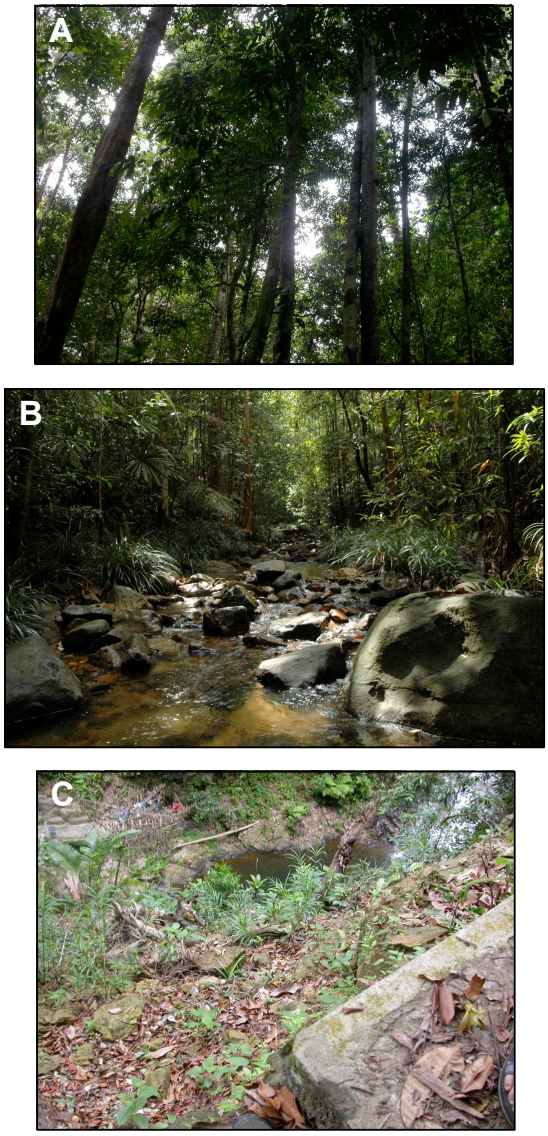


Figure 3.2: Habitat pictures of sampling areas of *Acanthorrhynchium papillatum* in Malaysia: (A) Sungei Bantang Recreational Forest. (B) Gunung Belumut Forest Reserve. (C) Kota Tinggi Waterfalls Resort.

with several chalets, administrative buildings, a small diner, huts and prepared campgrounds. The recreation areas are larger and visitors venture deeper into the surrounding forest.

Sampling trails *Acanthorrhynchium papillatum* samples were collected from the trails that run close to the tributaries of the main river at the base of the mountain. These trails, also, do not have names. Most of the samples were collected from beyond the more common picnic and recreation areas.

Kota Tinggi Waterfalls Resort

Location and description Samples of *Acanthorrhynchium papillatum* were also collected from Kota Tinggi Waterfalls Resort (1°49' N, 103°50' E, 100 m above sea level). Kota Tinggi Waterfalls Resort is approximately 50 km north-northeast of Singapore (Figures 3.1 and 3.2) and 16 km from the town of Kota Tinggi. The park has long been popular for its waterfalls that cascade from the slopes of Gunung Muntahak and Gunung Panti. Kota Tinggi Waterfalls Resort is now a well-developed resort. Concrete chalets with modern facilities, eateries and recreation halls have been built. The trails are mostly of concrete. Even the stream after the plunge pool of the main cataract is dammed in places to provide shallow swimming pools for visitors. The development of the resort and the popularity of the site contribute to making the surrounding vegetation of Kota Tinggi Waterfalls Resort the most disturbed of the three sampling areas in Malaysia.

Sampling trails Samples of *Acanthorrhynchium papillatum* were collected from the grounds of the recreation areas, the trail behind the chalets, the trail that runs along the waterfall leading to the summit of Gunung Muntahak and the trail that leads to Gunung Panti. These trails, also, do not have names.

Bukit Timah Nature Reserve

Location and description One of the sampling sites in Singapore is Bukit Timah Nature Reserve (1°21' N, 103°46' E, 100 m above sea level). Both Bukit Timah Nature Reserve and the other sampling area in Singapore, MacRitchie Reservoir are found near the geographic center of the island (Figures 3.3 and 3.4) and form part of the Central Catchment Nature Reserve. The closest margins of the two sites are only about 4 km apart but are separated by an expressway and a golf course. Although the forests of Bukit Timah Nature Reserve are mostly primary (Corlett, 1992), the trails that run through them are well-tended, some with wooden walkways and steel or concrete braces, and experience regular disturbances. Bukit Timah Nature Reserve is also a very popular recreation, research and teaching site.

Sampling trails Samples of *Acanthorrhynchium papillatum* were collected from the following trails in Bukit Timah Nature Reserve: South View Path, Cave Path, Tiup-tiup Path, Rock Path, Jungle Fall Path, Seraya Loop and North View Path.

MacRitchie Reservoir

Location and description Samples of *Acanthorrhynchium papillatum* were also collected from MacRitchie Reservoir (1°20' N, 103°49' E, 40 m above sea level). Although parts of MacRitchie Reservoir (Figures 3.3 and 3.4) remain primary forest (Corlett, 1992), the areas accessible to the public, including the only area where *A. papillatum* was found, are in secondary forests. MacRitchie Reservoir is likewise a very popular recreation site and receives many visitors in the form of joggers and nature hikers.

Sampling trails Trails around the entire circumference of MacRitchie Reservoir were explored. However, *Acanthorrhynchium papillatum* was only found and

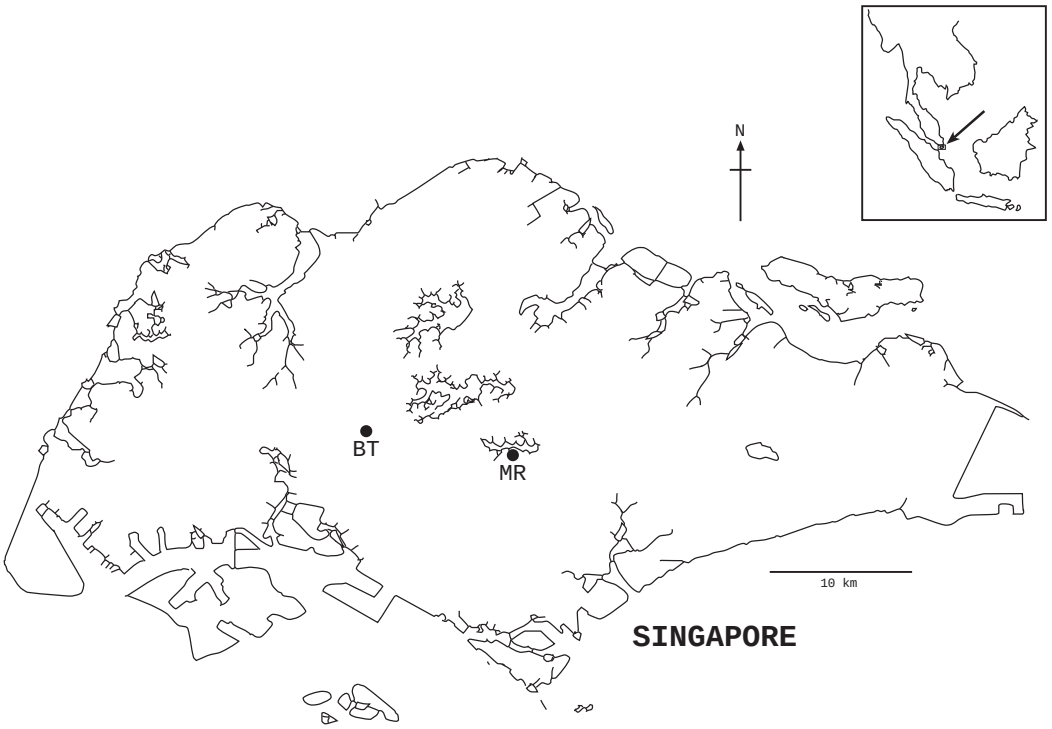


Figure 3.3: Location of sampling areas of *Acanthorrhynchium papillatum* in Singapore: BT - Bukit Timah Nature Reserve; MR - MacRitchie Reservoir. A distance bar is indicated. Inset shows location of Singapore in relation to mainland Southeast Asia. (Maps courtesy of Reuben Clements).

Figure 3.4: Habitat pictures of sampling areas of *Acanthorrhynchium papillatum* in Singapore: (A) Bukit Timah Nature Reserve. (B) MacRitchie Reservoir.

collected from Lornie Trail, located on the south side of the reservoir. Parts of this trail are exposed, with no tree cover, and are within 350 m of an adjacent highway.

Distances between sampling areas

Table 3.1 shows the approximate distances between the sampling areas of *Acanthorrhynchium papillatum*. Bukit Timah Nature Reserve and MacRitchie Reservoir are closest to each other being only about 6 km apart. MacRitchie Reservoir and Sungei Bantang Recreational Forest are most distant to each other being about 140 km apart. Sungei Bantang Recreational Forest and Gunung Belumut Forest Reserve together are almost the same distance away from the sampling areas in Singapore. Kota Tinggi Waterfalls Resort lies between the two sampling areas in Singapore and the two other sampling sites in Malaysia.

3.2.2 Collection dates

Table 3.2 details the collection dates of samples of *Acanthorrhynchium papillatum* used to determine diversity among clumps of the moss. When samples from the same locality were collected on several dates, they were collected in different trails, or different parts of the same trail.

3.2.3 Field selection and collection

Since no prior information on the clustering or distribution of *Acanthorrhynchium papillatum* was available, samples were selected and collected as they were found in the field. However, collections were done in such a way as to minimize impact on populations of the moss and on the surrounding vegetation. Samples were scraped, pulled or plucked carefully from their substrates and kept in paper packets. Multiple packets were kept together in sealed Ziploc® bags en route to the

Table 3.1: Approximate pairwise distances between sampling areas of *Acanthorrhynchium papillatum*. Figures are in kilometers.

	SBRF	GBFR	KTP	BTNR	MACR
SBRF	*				
GBFR	59.2	*			
KTP	101.5	92.1	*		
BTNR	131.9	114.3	53.5	*	
MACR	136.3	119.0	54.2	5.7	*

Table 3.2: Collection dates of *Acanthorrhynchium papillatum* samples that were used to determine variation among clumps of the moss.

Sampling area	Collection date(s)
SBRF	7 March 2004; 28–29 July 2004
GBFR	7 March 2004; 30–21 July 2004
KTP	2 July 2005
BTNR	7 April 2004; 18, 28 May 2004
MACR	16 March 2004; 17 April 2004; 31 May 2004

laboratory.

Samples were named according to their source locality using the following scheme: APBK → Sungei Bantang Recreational Forest, APBL → Gunung Belumut Forest Reserve, APKT → Kota Tinggi Waterfalls Resort, APBT → Bukit Timah Nature Reserve, APMR → MacRitchie Reservoir. Samples were numbered sequentially in the order of their collection. Thus, sample names were indicative of the relative distances of samples in the field.

3.2.4 Post-collection processing

Samples were kept moist in their bags at 4 °C prior to being processed for genomic DNA extraction.

A few shoots from each sample collected were examined under both stereo and compound microscopes to confirm the identity of the mosses collected in the field.

3.2.5 Lab sampling

One whole, continuous shoot of *Acanthorrhynchium papillatum* was selected from every sample collected. Shoots were processed for genomic DNA extraction and subsequent genotyping.

3.2.6 Genomic DNA extraction

Each selected shoot was allowed to air-dry on a bench top until tissues turned brittle (around 1 hour) and then inserted into a clean 2 mL microfuge tube into which 4 clean, 3 mm tungsten carbide beads had been placed.

Sample tubes were capped tightly, immersed in liquid nitrogen and allowed to come to thermal equilibrium. The samples were ground to a fine powder by vortexing the still deep-frozen tubes at maximum power for 30 s. The tubes were then re-immersed in liquid nitrogen. Cycles of vortexing and re-freezing were

repeated for a total of 3 times. This method was independently developed by this researcher and offers several advantages over the conventional mortar and pestle method of pulverizing samples. It is faster, allowing the grinding of at least 20 samples in 20 to 30 min, even faster for softer tissue. It is more efficient in the use of liquid nitrogen, using less than 800 mL for 20 samples. More importantly, it helps prevent cross-contamination from aerosols that form when samples are ground with a mortar and pestle. It is, however, less efficient than the use of a dedicated mixer mill. However, for labs that do not have access to a mixer mill, it offers an adequate method for the high-throughput grinding of plant samples prerequisite to the chemical extraction of genomic DNA.

Genomic DNA from the ground tissues was extracted using DNeasy Plant Mini Kits (QIAGEN, USA) following the manufacturer's protocol. One hundred microliters of Buffer AE were used to elute the DNA following the final washes of the filter membranes.

3.2.7 Fragment analysis with microsatellite markers

Fragment analysis using microsatellite markers to study the diversity of *Acanthorrhynchium papillatum* closely followed the scheme used in the development of these markers for this moss. Some alterations were made, however, to take advantage of technologies that more recently became available.

Multiplex PCR was carried out using a hot-start DNA polymerase to increase sensitivity and specificity of amplification and to minimize the formation of primer dimers. Ten microliter reactions of 1×F-522 buffer (FinnZymes, Finland), 5% DMSO (Sigma, USA), 200 μM dNTPs (Promega, USA), 0.25 μM or 0.5 μM forward primer, 0.25 μM or 0.5 μM reverse primer, 0.24 U Hot Start DyNAzyme^{TM} II DNA polymerase (FinnZymes, Finland) and approximately 10 ng genomic DNA were combined together in thin-walled, 0.2 mL tubes. Reactions were performed

in a PTC-100® (MJ Research, USA) thermocycler using the following thermal profile: Initial denaturation for 10 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 30 s at 54 °C, 2 min at 65 °C, and final extension for 30 min at 72 °C. Products were kept at 16 °C until recovery.

The multiplex PCR products were purified using CleanSEQ® kits (Agencourt, USA) following the manufacturer's protocol.

Fragment analysis runs of the purified products were performed in a PRISM® 3100 Genetic Analyzer (Applied Biosystems, USA) following the protocol of Applied Biosystems. ROX™-labeled GeneScan-400HD (Applied Biosystems, USA) was used as the internal size standard. Several changes in the run module parameters of the 3100 Genetic Analyzer were made to allow fragment analysis runs to be performed using 50 cm capillary arrays and POP-6™ polymer (Table 3.3).

3.2.8 Size-calling and allele-scoring

PCR fragment sizes were automatically calculated by GeneMapper® version 4.0 (Applied Biosystems, USA) by interpolation using the local Southern algorithm from a sizing curve generated from the internal size standards. The computed fragment sizes were in fractional base pair units and needed to be binned into discrete base pair units. Although the software was used to automatically accomplished this, the binning was reviewed and manually adjusted where needed.

Alleles for a microsatellite locus may be coded as the number of repeat units for that locus. This is easily derived by subtracting the sum of the length of the flanking regions of the microsatellite locus from the length of the PCR product and dividing the difference by the length of the repeat unit. Coding the alleles as the number of repeat units allows for data analysis using the stepwise mutation model (SMM) of Kimura & Ohta (1978). As will be seen later in the results, the lengths of the PCR products suggest that none of the microsatellite loci used

Table 3.3: Custom run module parameters that were optimized to allow for fragment analysis runs using Data Collection software version 1 of the PRISM® 3100 Genetic Analyzer (Applied Biosystems, USA).

Parameter	Value
Run Temperature	60 °C
Cap Fill Volume	184 steps
Current Tolerance	100 μA
Run Current	100 μA
Voltage Tolerance	0.6 kV
Pre Run Voltage	12.2 kV
Pre Run Time	180 s
Injection Voltage	2 kV
Injection Time	22 s
Run Voltage	12.2 kV
Number of Steps	10
Voltage Step Interval	60 s
Data Delay Time	1 s
Run Time	4800 s

in this study mutate in a strictly stepwise fashion. Because of this observation, alleles were instead coded as fragment sizes of PCR products in base pairs and data analysis methods were adjusted accordingly.

3.2.9 Data analysis

Alleles were entered into an EXCEL® (Microsoft, USA) spreadsheet after being exported from GENEMAPPER® Software v4.0 (Applied Biosystems, USA). The EXCEL MICROSATELLITE TOOLKIT version 3 (Park, 2001) was used to calculate various summary statistics and to format the data for use on other software.

Revised marker statistics

With the larger number of samples made available, it was possible to recalculate the marker statistics of each locus that were originally computed using the smaller data set used in the development of the microsatellite markers. Allele size ranges, the number of alleles and Nei's gene diversity for each locus were computed using the same methods used in the original calculations. In addition to the statistics originally computed, Nei's unbiased gene diversity for each locus (Nei, 1987) was also calculated. While the difference between the original and unbiased statistics are very small for sample sizes ≥ 50 (Nei, 1987), the unbiased gene diversity is a more appropriate measure and was thus used here. (Unbiased gene diversity statistics were also computed for the earlier data set for comparison). All these statistics help characterize the richness and evenness of the diversity of the microsatellite markers.

Total microsatellite diversity

In addition to figures for the revised marker statistics, other statistics that more fully describe the diversity of the microsatellite markers were also computed for

the entire data set. Allelic frequencies of each marker for the whole data set were computed using the Excel Microsatellite Toolkit version 3 of Park (2001). They were then used to determine other characteristics of the markers, such as the total and mean number of alleles (allelic richness) and the presence of null alleles (samples with no data at a particular locus) and rare alleles (define here as alleles found in $< 1\%$ of the population).

Microsatellite variability in each sampling area

Allelic frequencies of microsatellite markers in samples from each sampling area were computed, again using the Excel Microsatellite Toolkit version 3 (Park, 2001), to characterize microsatellite variability in each sampling area.

Additional statistics were derived for each sampling area from allelic frequencies. These statistics include the size ranges of the alleles, allelic richness in the form of the total and mean number of alleles, numbers of rare and private alleles (private alleles are those found in only one sampling locality) and Nei's unbiased gene diversities for each sampling area.

Size ranges of the alleles, and the numbers of private and rare alleles were determined visually from allelic frequencies. The total and mean number of alleles as well as Nei's unbiased gene diversities were computed using the Excel Microsatellite Toolkit version 3 (Park, 2001).

Since the total and mean number of alleles and the number of private alleles are influenced by sample size (El Mousadik & Petit, 1996; Kalinowski, 2004), these statistics were also computed using an alternative method that accounts for differences in the numbers of individuals collected from the sampling areas. A different program, HP-RARE version 1.0 (Kalinowski, 2005), was used to calculate allelic richness and private alleles to compensate for these differences in sample sizes. The program uses rarefaction, a method originally used for ecological studies on

species richness (Hurlbert, 1971) but recently adapted for use on allelic richness (El Mousadik & Petit, 1996; Kalinowski, 2004; Takahashi *et al.*, 2005).

The Friedman test followed (if necessary) by Dunn's multiple comparison test were used to compute statistical significance of differences in allelic richness and numbers of private alleles among the various sampling areas after rarefaction. The software package, InStat version 3 (GraphPad Software, San Diego California USA, www.graphpad.com), was used to calculate these tests of significance.

To help complete the description of the microsatellite variability among the sampling areas, the presence of null alleles were also determined. These were counted directly from the raw data and not from allelic frequencies.

Multi-locus genotypic diversity in each sampling area

Following the examination of diversity at the marker or genic level, genotypic diversity –combinations of alleles for all markers tested– at the level of individuals was also investigated.

The Excel Microsatellite Toolkit version 3 (Park, 2001) was used to determine the numbers of matching samples, those that shared the same alleles for all 8 markers.

From figures on matching samples, the numbers of unique multilocus genotypes for each sampling area were easily computed by subtracting the number of replicated samples in each sampling area from the total number of samples in each sampling area.

As a loose gauge of the genotypic richness of each sampling area, the theoretical maximum numbers of multilocus genotypes for each were computed using the power rule of probability theory.

To help distinguish different genets (individuals that are products of sexual reproduction) and recognize ramets (individuals that are products of vegetative

reproduction) within each area, two statistics, power of discrimination and the related statistic, match probability were computed. Power of discrimination is the chance that two different individuals will exhibit different multilocus genotypes. Match probability is the inverse: the probability that two different samples share the same multilocus genotype. Powers of discrimination, PD, were computed as exemplified by Gómez *et al.* (2001) and van der Velde *et al.* (2001a). Match probabilities, P_M, were computed simply as $P_M = 1 - PD$.

Genetic distances

Besides comparing degrees of variation of groups of samples from different sampling areas, genetic distances were also calculated to examine degrees of relatedness in two scales: among samples within the same sampling area, and among groups of samples from different sampling areas.

Within sampling areas Genetic distances of samples within each sampling area were calculated using allele sharing matrices (Bowcock *et al.*, 1994) as implemented by the Excel Microsatellite Toolkit version 3 (Park, 2001). Trees were constructed from the distance matrices with the Neighbor package of Phylip (Felsenstein, 2005) using the neighbor-joining method (Saitou & Nei, 1987). Trees were visualized using TreeView (Page, 1996).

Among sampling areas Pairwise genetic distances between the different sampling areas were computed using two measures of genetic distance, Nei's standard genetic distance, D (Nei, 1972), and net genetic distance, D_A (Nei *et al.*, 1983). Trees were constructed from matrices of these genetic distance measures using the neighbor-joining algorithm (Saitou & Nei, 1987). One-thousand bootstrap replicates over the loci were performed to assess the strength of each tree. The program Dispan (Ota, 1993) was used to perform the calculations. Trees were visualized

using TreeView (Page, 1996).

Correlations between geographic distances of the sampling areas and genetic distances (both Nei's standard and net genetic distances) were investigated using the Mantel test (Smouse *et al.*, 1986) performed using the software program, Arlequin version 3.11 (Excoffier *et al.*, 2005b).

Population differentiation

Exact tests of population differentiation were performed as described by Raymond & Rousset (1995a) using the online version of the software package, Genepop (Raymond & Rousset, 1995b). Sampling areas were defined as populations. A global test of differentiation of multi-locus genotype frequencies was performed using 100 000 000 Markov steps. This was done to compute the probability of accepting the null hypothesis that the genotypic distribution among all the populations was equal. As an added metric, an exact test of population differentiation on allelic distribution was calculated using similar parameters. In this test, the probability of accepting the null hypothesis, that the allelic distribution among all the populations was equal, was computed.

Population genetic structure

The presence of population genetic structure was investigated to determine if mosses in the different sampling areas were genetically isolated from mosses in other sampling areas. This was done in two ways.

Population genetic structure was inferred by calculating hierarchical fixation indices using analysis of molecular variance or AMOVA (Excoffier *et al.*, 1992; Michalakis & Excoffier, 1996). To perform this analysis, populations and groups of populations first needed to be defined. Sampling areas were defined as populations. Two groups of populations were defined for analysis, one comprised of

sampling areas in Singapore (BTNR and MACR), the other comprised of sampling areas in Malaysia (SBRF, GBFR and KTP). AMOVA was then used to detect population genetic differentiation among populations within groups, among individuals among populations and among groups, and finally between the two groups of populations. AMOVA was also used to determine the sources of variation in the data set. Calculations were performed using the software program, ARLEQUIN version 3.11 (Excoffier *et al.*, 2005b). Significance levels for overall fixation indices were computed through 1023 permutations (default number in the software package).

Population genetic structure was also inferred by calculating conventional fixation indices from pairwise differences among the different sampling areas. These were computed using the same software package, ARLEQUIN version 3.11 (Excoffier *et al.*, 2005b). *P*-values for these indices were also computed, again through 1023 permutations.

3.3 Results

3.3.1 Collected samples

A total of 279 samples were collected: 51 samples from Sungei Bantang Recreational Forest, 72 from Gunung Belumut Forest Reserve, 34 samples from Kota Tinggi Waterfalls Resort, 54 samples from Bukit Timah Nature Reserve and 68 samples from MacRitchie Reservoir. Collections were exhaustive in the trails explored: All specimens of *Acanthorrhynchium papillatum* encountered in the trails were sampled and collected. Interestingly, very few sporulating samples were encountered in the field. Efforts to isolate DNA from the few sporophytes found were attempted (to further enrich the data). However the sporophytes were so small that these attempts were unsuccessful as too much material was lost during the grinding process prerequisite to the chemical extraction of the nucleic acids. No

further attempts to study sporophytic material were made following these failed attempts.

Although a different instrument and software were used for this part of the project than what was used during the development of the microsatellite markers, electrophoretic data obtained were similar to those already shown in pages 66 and 67.

Tabulated PCR fragment sizes of the markers for all samples are presented in Appendix J.

3.3.2 Revised marker statistics

The revised marker statistics on the allele size ranges of the markers, the number of alleles, and Nei's unbiased gene diversity for each locus are presented in Table 3.4 (to aid comparison, the original marker statistics are also shown). These marker statistics, computed using all 279 samples collected from five populations, included samples with null alleles and those that had the same multi-locus genotypes, some of which were excluded in the original calculations.

Compared to the original marker statistics, the revised statistics showed larger ranges in amplicon sizes for each locus. These are expected results since more samples were used in the revised calculations. However, it is also evident that the upper bounds of the size ranges of some of the loci, namely, APMS4, APMS14, APMS28, APBMS3 and APBMS14 were slightly smaller (2–4 bp) than in the original set. These results may seem surprising: after all, the same samples that were used in the original calculations were also used in the revised calculations. Upper bounds should only increase and lower bounds should only decrease when more samples are added, as in the case of the revised calculations. However, it should be noted that two different instruments and software programs were used in the original (gel-based instrument and GeneScan® software) and revised (capillary

Table 3.4: Revised statistics of microsatellite markers tested on 279 gametophytic samples of *Acanthorrhynchium papillatum* from five populations. For clarity, only Nei's unbiased gene diversity statistics are shown. **O** – original data set; **R** – revised data set; A = number of alleles; $\hat{h}$ = Nei's unbiased gene diversity.

Locus	Size range (bp)		A		$\hat{h}$	
	O	**R**	**O**	**R**	**O**	**R**
APMS4	166–217	161–215	15	25	0.740	0.842
APMS14	67–126	59–122	11	26	0.856	0.874
APMS28	79–156	75–154	26	34	0.941	0.947
APBMS3	92–164	76–162	19	34	0.878	0.926
APBMS14	94–158	93–156	11	25	0.767	0.911
APBMS23	108–130	97–191	6	15	0.740	0.780
APBMS61	233–239	232–242	4	8	0.581	0.614
APBMS72	69–111	67–133	11	21	0.605	0.571

electrophoresis instrument and GeneMapper® software) data to determine the sizes of the fragments. These different instruments can cause the same-sized fragments to be called differently. It is therefore important that data to be used in the same analysis be collected and scored using the same experimental systems to ensure consistency in the results. In the experiments of this chapter, all the data were gathered using the same instrument (PRISM® 3100 Genetic Analyzer) and analyzed using the same software (GeneMapper®) to preclude any bias that may result from the use of different experimental conditions. (In higher stringency applications of fragment analysis, like in the case of human identification in forensics, allelic ladders are run alongside the samples for comparison to compensate for instrument variations. Since only a single instrument was used in the experiments of this chapter, allelic ladders did not need to be employed).

Greater numbers of alleles were also seen for all markers using the larger data set. These, in part, increased the already high levels of gene diversity computed from the original data set for almost all the loci. APBMS72 is notable as being the only locus with decreased levels of gene diversity in the revised calculations, this despite also having a larger number of alleles in the larger data set (21 in the revised versus 11 in the original). This is because many of these "new" alleles in the larger data set are represented by only one or few samples. A large proportion of the samples in the larger data set still possessed the dominant allele of 77 bp. This reduced the evenness of the distribution of alleles and consequently the gene diversity of this locus.

The general view that microsatellites with long repeats are more variable is not evident from the revised marker statistics. For example APMS14, which was found with 46 repeats of the GA motif when it was isolated, had fewer alleles (26) alleles in the data set than APBMS3 (34 alleles) which was found with only 20 repeats of a GA motif when it was isolated. The same is true for APBMS72. This

locus was found with 30 repeats of a GA motif when it was isolated but had fewer alleles (21) than APMS28 (34), which was found with 23 repeats of a GA motif when it was isolated.

3.3.3 Total microsatellite diversity

More statistics related to microsatellite diversity of the entire data set are presented in addition to those used when reporting microsatellite markers.

Overall allelic frequencies

Total allelic frequencies for each locus are depicted graphically in Figure 3.5.

A total of 188 alleles were generated from the 8 microsatellite markers (mean 23.5 ± 8.8 alleles per locus). Locus APBMS61 exhibited the fewest number of alleles (8) while loci APMS28 and APBMS3 showed the most number of alleles (34 each).

Most loci appeared unimodal, except for APBMS23 which had two prevalent alleles at 105 and 119 bp. While the allelic distribution of all the other loci had distinct peaks, APBMS28 exhibited a relatively flat distribution of alleles. Distributions were generally skewed to the right with more alleles found greater than the average than fewer. Gaps larger than 10 bp were found in distributions of APMS4, APMS28, APBMS3, APBMS14 and APBMS23. Outlying alleles were found in all but one locus (APBMS61) and were generally on the upper limits of the allele distribution of the loci. The largest range in upper and lower bounds of alleles is found in APBMS23 (94 bp) while the smallest is found in APBMS61 (10 bp). Allele ranges have a mean of 64 bp with a standard deviation of 24 bp.

While most of the markers were from loci with core sequences of dinucleotide repeats (exceptions are APBMS14 and APBMS61 that had complex core sequences that included both mononucleotide and dinucleotide repeats), it is clear from the

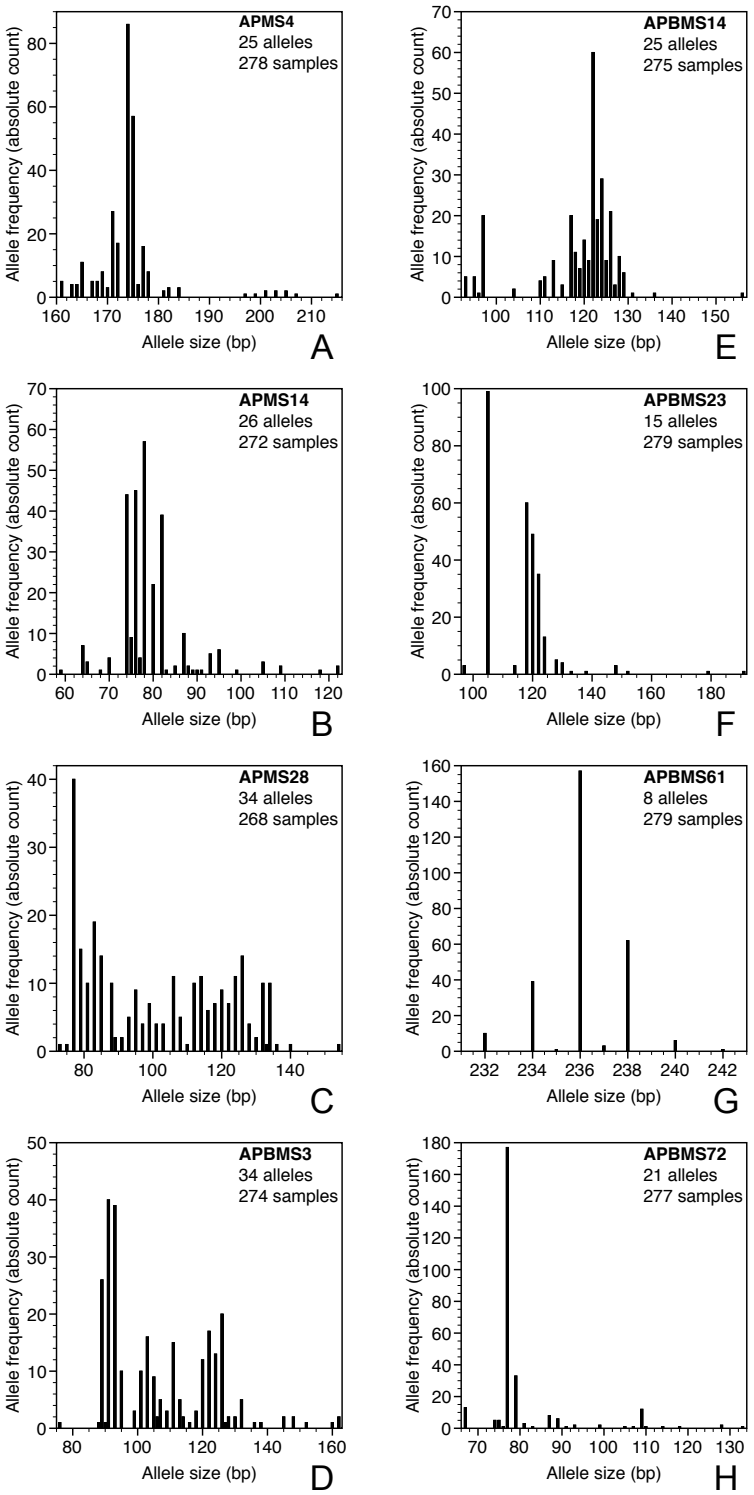


Figure 3.5: Allelic frequencies of the 8 microsatellite markers used. Allele frequencies are presented as absolute counts (number of samples having the allele). The number of alleles found for each locus and the number of samples used to construct each graph are also indicated (some samples were null at a locus and were excluded from the construction of the graph for that locus).

allelic distributions that none of the loci exhibit a strictly stepwise mutation pattern; that is, alleles for most loci are not strictly 2 bp apart. However APMS28, APBMS3 and APBMS61 do exhibit a nearly stepwise mutation pattern with deviations caused by only a few number of samples. Overall, the observations suggest that the stepwise mutation model may not be in effect among these loci and the subsequent statistical analysis must take this into consideration.

Null alleles

Figure 3.5 also indicates by inference the number of samples that were null at a locus. The total number of samples producing null alleles was 30 out of 279 (10.75%). No sample was null at more than 1 locus. The mean number of null alleles per locus was 3.75 with a standard deviation of 3.85. Null alleles were found in all but two loci, APBMS23 and APBMS61. Only one sample was null at APBMS4 while 11 samples were null at APMS28.

Rare alleles

Rare alleles of each locus, those that are found in less than 1% of the samples collected (1 or 2 of the 279 total number of samples), are shown in Table 3.5.

Rare alleles make up 68 out of the 188 (36.2%) total number of alleles (mean 8.5 ± 4.5 alleles per locus). Forty-seven alleles (25.4%) were found in only 1 sample each while 21 alleles (11.3%) were found in only 2 samples each. A majority of these rare alleles were found outlying the allele clusters. APBMS3 had the largest number of rare alleles (16) while APBMS61 had the fewest number of rare alleles (2).

Table 3.5: Numbers of rare alleles found in the 8 microsatellite markers used. Rare alleles are those that are found in less than 1% of the total number of samples. Given a total samples size of 279, rare alleles are thus those that were found in less than 3 samples each. Middle and right columns indicate the number of alleles represented by only 1 or 2 samples each.

	Sample representation	
Locus	1	2
APMS4	4	4
APMS14	8	3
APMS28	6	3
APBMS3	9	7
APBMS14	4	1
APBMS23	5	0
APBMS61	2	0
APBMS72	9	3
Total	47	21

3.3.4 Microsatellite diversity within sampling areas

Allelic frequencies

Allelic frequencies of the 8 markers for each sampling area are presented in Table 3.6.

While the table of allelic frequencies shows comprehensive data on the diversity of the loci and the sampling areas, clearer data are obtained when additional statistics are derived from this table.

Allele size ranges

The size ranges of the alleles found in each sampling location are presented in Table 3.7. The size ranges of the alleles are not usually used to measure genetic diversity among populations [although they are used in some statistics like Garza-Williamson's M index, first described by Garza & Williamson (2001) and modified by Excoffier *et al.* (2005a)]. They are presented here because they suggest the extent of the mutability of a locus and consequently, the genetic variation of the individuals in that sampling area. However they, like some other statistics related to diversity, are also influenced by sample size. The largest allele size ranges were found in APBMS3 in BTNR and in APBMS72 in SBRF, both with a range of 86 bp.

Allelic richness

Greater information may be obtained from statistics other than the size ranges of the alleles. Allelic richness statistics, as counted from the total and mean number of alleles per locus, are presented in Table 3.8.

The table shows that the samples from Gunung Belumut Forest Reserve exhibit the largest number of alleles: a total of 130 (mean of 16.2 ± 8.0 s.d. alleles per locus) while the sampling area with the fewest number of alleles is MacRitchie

Table 3.6: Allelic frequencies of all microsatellite markers used for each sampling area. Alleles are amplicon lengths in bp. Allele frequencies are presented as relative values.

Locus	Allele	SBRF	GBFR	KTP	BTNR	MACR
APMS4	161	0.080	0.014	0.000	0.000	0.000
	163	0.020	0.028	0.000	0.019	0.000
	164	0.000	0.028	0.029	0.000	0.015
	165	0.020	0.028	0.029	0.111	0.015
	167	0.000	0.014	0.000	0.074	0.000
	168	0.000	0.014	0.029	0.056	0.000
	169	0.060	0.042	0.000	0.037	0.000
	170	0.020	0.014	0.000	0.019	0.000
	171	0.040	0.056	0.059	0.130	0.176
	172	0.140	0.014	0.000	0.130	0.029
	174	0.440	0.306	0.559	0.185	0.191
	175	0.060	0.139	0.088	0.130	0.500
	176	0.020	0.000	0.059	0.000	0.015
	177	0.020	0.111	0.118	0.037	0.015
	178	0.040	0.028	0.029	0.000	0.044
	181	0.000	0.014	0.000	0.019	0.000
	182	0.000	0.042	0.000	0.000	0.000
	184	0.000	0.000	0.000	0.056	0.000
	197	0.000	0.014	0.000	0.000	0.000
	199	0.000	0.014	0.000	0.000	0.000
	201	0.020	0.014	0.000	0.000	0.000
	203	0.000	0.028	0.000	0.000	0.000
	205	0.020	0.014	0.000	0.000	0.000
	207	0.000	0.014	0.000	0.000	0.000
	215	0.000	0.014	0.000	0.000	0.000
APMS14	59	0.000	0.000	0.000	0.019	0.000
	64	0.020	0.045	0.000	0.056	0.000
	65	0.020	0.000	0.000	0.019	0.015
	68	0.000	0.015	0.000	0.000	0.000
	70	0.020	0.030	0.000	0.000	0.015
	74	0.118	0.182	0.091	0.222	0.162
	75	0.039	0.030	0.030	0.074	0.000
	76	0.039	0.076	0.212	0.074	0.397
	77	0.039	0.015	0.000	0.019	0.000
	78	0.294	0.364	0.121	0.167	0.044
	80	0.118	0.015	0.121	0.167	0.029
	82	0.098	0.091	0.030	0.130	0.294
	83	0.020	0.000	0.000	0.000	0.000
	85	0.000	0.015	0.030	0.000	0.000
	87	0.059	0.061	0.000	0.056	0.000
	88	0.020	0.015	0.000	0.000	0.000
	89	0.020	0.000	0.000	0.000	0.000
	90	0.000	0.000	0.030	0.000	0.000
	91	0.020	0.000	0.000	0.000	0.000
	93	0.000	0.000	0.152	0.000	0.000
	95	0.000	0.000	0.182	0.000	0.000
	99	0.000	0.015	0.000	0.000	0.000
	105	0.000	0.000	0.000	0.000	0.044
	109	0.020	0.015	0.000	0.000	0.000
	118	0.020	0.000	0.000	0.000	0.000
	122	0.020	0.015	0.000	0.000	0.000
APMS28	75	0.000	0.000	0.000	0.019	0.000
	77	0.275	0.097	0.161	0.057	0.180
	79	0.020	0.042	0.032	0.151	0.033
	81	0.078	0.014	0.097	0.038	0.000
	83	0.078	0.069	0.000	0.151	0.033
	85	0.039	0.014	0.032	0.038	0.131
	88	0.039	0.014	0.000	0.038	0.082
	89	0.020	0.014	0.000	0.000	0.000
	91	0.020	0.014	0.000	0.000	0.000
	93	0.000	0.069	0.000	0.000	0.000
	95	0.000	0.042	0.000	0.113	0.000
	97	0.020	0.028	0.000	0.019	0.000
	99	0.000	0.083	0.000	0.019	0.000
	101	0.020	0.042	0.000	0.000	0.000
	103	0.059	0.014	0.000	0.000	0.000
	106	0.039	0.028	0.000	0.019	0.098
	108	0.059	0.014	0.000	0.019	0.000
	110	0.000	0.000	0.000	0.019	0.000
	112	0.000	0.028	0.129	0.075	0.000
	114	0.020	0.069	0.129	0.019	0.000
	116	0.020	0.028	0.000	0.038	0.016
	118	0.000	0.014	0.194	0.000	0.000
	120	0.020	0.056	0.097	0.019	0.000
	122	0.020	0.056	0.000	0.019	0.016
	124	0.078	0.028	0.065	0.000	0.049
	126	0.039	0.056	0.000	0.000	0.131
	128	0.020	0.014	0.065	0.000	0.000
	130	0.000	0.028	0.000	0.000	0.000
	132	0.000	0.000	0.000	0.113	0.066
	133	0.000	0.014	0.000	0.000	0.000
	134	0.020	0.014	0.000	0.000	0.131
	136	0.000	0.000	0.000	0.000	0.016
	140	0.000	0.000	0.000	0.000	0.016
	154	0.000	0.000	0.000	0.019	0.000
APBMS3	76	0.000	0.000	0.000	0.019	0.000
	88	0.000	0.014	0.000	0.000	0.000
	89	0.122	0.086	0.059	0.226	0.000
	90	0.020	0.000	0.000	0.000	0.000
	91	0.102	0.229	0.118	0.075	0.162
	93	0.184	0.043	0.059	0.075	0.309
	95	0.082	0.014	0.000	0.038	0.044
	99	0.000	0.029	0.000	0.019	0.000
	101	0.020	0.014	0.000	0.038	0.088
	103	0.143	0.086	0.088	0.000	0.000
	105	0.041	0.029	0.118	0.019	0.000
	106	0.000	0.000	0.059	0.000	0.000
	107	0.000	0.014	0.029	0.038	0.015
	109	0.020	0.000	0.000	0.000	0.029
	111	0.000	0.043	0.000	0.000	0.176
	113	0.082	0.000	0.000	0.019	0.000
	114	0.000	0.014	0.000	0.019	0.000
	116	0.000	0.014	0.000	0.000	0.000
	118	0.041	0.000	0.000	0.000	0.015
	120	0.020	0.086	0.088	0.019	0.015
	122	0.000	0.100	0.000	0.170	0.015
	124	0.020	0.029	0.118	0.113	0.000
	126	0.041	0.086	0.088	0.000	0.132
	127	0.000	0.000	0.000	0.019	0.000
	128	0.000	0.029	0.000	0.000	0.000
	130	0.000	0.014	0.029	0.000	0.000
	132	0.000	0.014	0.118	0.000	0.000
	136	0.000	0.014	0.000	0.000	0.000
	138	0.020	0.000	0.000	0.000	0.000
	145	0.000	0.000	0.029	0.019	0.000
	148	0.000	0.000	0.000	0.038	0.000
	152	0.020	0.000	0.000	0.000	0.000
	160	0.020	0.000	0.000	0.000	0.000
	162	0.000	0.000	0.000	0.038	0.000
APBMS14	93	0.000	0.069	0.000	0.000	0.000
	95	0.039	0.014	0.029	0.020	0.000
	96	0.020	0.000	0.000	0.000	0.000
	97	0.000	0.014	0.382	0.120	0.000
	104	0.039	0.000	0.000	0.000	0.000
	110	0.000	0.028	0.000	0.040	0.000
	111	0.020	0.042	0.000	0.020	0.000
	113	0.020	0.014	0.029	0.120	0.000
	115	0.020	0.000	0.029	0.000	0.015
	117	0.020	0.056	0.029	0.000	0.206
	118	0.039	0.014	0.118	0.020	0.044
	119	0.000	0.028	0.000	0.060	0.029
	120	0.000	0.056	0.029	0.180	0.000
	121	0.020	0.014	0.029	0.100	0.015
	122	0.059	0.194	0.147	0.060	0.515
	123	0.235	0.028	0.000	0.040	0.044
	124	0.255	0.181	0.059	0.000	0.015
	125	0.020	0.056	0.000	0.040	0.029
	126	0.118	0.139	0.059	0.020	0.029
	127	0.020	0.014	0.000	0.020	0.000
	128	0.000	0.014	0.000	0.120	0.044
	129	0.039	0.028	0.029	0.000	0.015
	131	0.000	0.000	0.029	0.000	0.000
	136	0.020	0.000	0.000	0.000	0.000
	156	0.000	0.000	0.000	0.020	0.000
APBMS23	97	0.000	0.042	0.000	0.000	0.000
	105	0.510	0.389	0.735	0.259	0.088
	114	0.020	0.000	0.059	0.000	0.000
	118	0.098	0.153	0.088	0.130	0.500
	120	0.020	0.181	0.088	0.204	0.309
	122	0.039	0.097	0.029	0.333	0.103
	124	0.020	0.139	0.000	0.037	0.000
	128	0.078	0.000	0.000	0.019	0.000
	130	0.078	0.000	0.000	0.000	0.000
	133	0.020	0.000	0.000	0.000	0.000
	138	0.000	0.000	0.000	0.019	0.000
	148	0.059	0.000	0.000	0.000	0.000
	152	0.020	0.000	0.000	0.000	0.000
	179	0.020	0.000	0.000	0.000	0.000
	191	0.020	0.000	0.000	0.000	0.000
APBMS61	232	0.098	0.069	0.000	0.000	0.000
	234	0.353	0.111	0.176	0.111	0.015
	235	0.000	0.014	0.000	0.000	0.000
	236	0.549	0.486	0.765	0.611	0.515
	237	0.000	0.042	0.000	0.000	0.000
	238	0.000	0.222	0.029	0.241	0.471
	240	0.000	0.056	0.000	0.037	0.000
	242	0.000	0.000	0.029	0.000	0.000
APBMS72	67	0.000	0.000	0.382	0.000	0.000
	74	0.000	0.000	0.000	0.000	0.074
	75	0.000	0.014	0.000	0.037	0.029
	76	0.000	0.000	0.000	0.019	0.000
	77	0.633	0.694	0.529	0.889	0.441
	79	0.041	0.014	0.059	0.000	0.412
	81	0.000	0.014	0.000	0.037	0.000
	83	0.000	0.014	0.000	0.000	0.000
	87	0.020	0.097	0.000	0.000	0.000
	89	0.020	0.028	0.000	0.000	0.044
	91	0.000	0.014	0.000	0.000	0.000
	93	0.020	0.014	0.000	0.000	0.000
	99	0.020	0.000	0.000	0.019	0.000
	105	0.020	0.000	0.000	0.000	0.000
	107	0.020	0.000	0.000	0.000	0.000
	109	0.102	0.097	0.000	0.000	0.000
	110	0.020	0.000	0.000	0.000	0.000
	114	0.020	0.000	0.000	0.000	0.000
	118	0.000	0.000	0.029	0.000	0.000
	128	0.041	0.000	0.000	0.000	0.000
	133	0.020	0.000	0.000	0.000	0.000

Table 3.7: Size ranges of alleles for every locus and for each sampling area. Figures are in bp.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	161–205	161–215	164–178	163–184	164–178
APMS14	64–122	64–122	74–95	59–87	65–105
APMS28	77–134	77–134	77–128	75–154	77–140
APBMS3	89–160	88–136	89–145	76–162	91–126
APBMS14	95–136	93–129	95–131	95–156	115–129
APBMS23	105–191	97–124	105–122	105–138	105–122
APBMS61	232–236	232–240	234–242	234–240	234–238
APBMS72	77–133	75–109	67–118	75–99	74–89

Table 3.8: Number of alleles per locus for each sampling area. Figures are direct counts taken from allelic frequencies and do not compensate for the differences in sample sizes from each sampling area. Sample sizes are indicated in parentheses. Total, means and standard deviations (s.d.) for each sampling area are also shown.

	Sampling area (sample size)				
Marker	SBRF (51)	GBFR (72)	KTP (34)	BTNR (54)	MACR (68)
APMS4	14	23	9	13	9
APMS14	18	16	10	11	8
APMS28	21	28	10	20	14
APBMS3	17	21	13	18	11
APBMS14	17	19	13	16	12
APBMS23	13	6	5	7	4
APBMS61	3	7	4	4	3
APBMS72	13	10	4	5	5
Total	116	130	68	94	66
Mean	14.5	16.2	8.5	11.7	8.2
s.d.	5.4	8.0	3.7	6.0	4.0

Reservoir with a total of 66 (mean of 8.2 ± 4.0 s.d. alleles per locus). Interestingly, Kota Tinggi Waterfalls Resort has a marginally greater number of alleles (68, mean of 8.5 ± 3.7 s.d. alleles per locus) than MacRitchie Reservoir despite having half the number of samples (34/68).

To account for differences in sample size in each sampling area the technique of rarefaction was performed on the data. A base sampling size equal to the smallest sampling size among the different sampling areas was chosen. Although the sampling area with the fewest number of samples was Kota Tinggi Waterfalls Resort with 34 samples, the base sampling size was set at 31 individuals. This is because 3 samples from Kota Tinggi Waterfalls Resort were null at a locus. The results of the rarefaction analysis are presented in Table 3.9.

The rarefacted data show that, when a base sampling size of 31 individuals is taken, Sungei Bantang Recreational Forest and Gunung Belumut Forest Reserve now show very similar levels of allelic richness with a total of approximately 89 alleles for all 8 marker loci (mean for SBRF is 11.2 ± 3.9 s.d., mean for GBFR is 11.2 ± 4.8 s.d.). These two sampling areas exhibit the most number of alleles among those studied. MacRitchie Reservoir still has the fewest number of alleles after rarefaction with a total of 51.8 alleles for all 8 marker loci (mean 6.5 ± 2.9 s.d.).

The results of the Friedman test on the rarefacted data indicate extremely significant differences in allelic richness of the sampling localities ($P \approx 0.0007$, estimated from the χ^2 distribution). The post Dunn's multiple comparisons test indicate significant differences in allelic richness between Sungei Bantang Recreational Forest and MacRitchie Reservoir ($P < 0.05$), Gunung Belumut Forest Reserve and Kota Tinggi Waterfalls Resort ($P < 0.05$) and highly significant differences between Gunung Belumut Forest Reserve and MacRitchie Reservoir ($P < 0.001$).

Table 3.9: Number of alleles per locus for each sampling area after rarefaction. Base sampling size was set at 31 individuals. Total, means and standard deviations (s.d.) for each sampling area are also shown.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	10.9	14.1	8.6	11.2	6.4
APMS14	13.6	10.9	9.8	9.5	6.3
APMS28	15.9	18.8	10.0	15.1	11.3
APBMS3	13.6	14.3	12.7	13.8	8.3
APBMS14	12.8	13.2	12.3	13.2	8.5
APBMS23	10.0	5.8	4.9	6.0	4.0
APBMS61	3.0	6.1	3.8	3.8	2.5
APBMS72	9.4	6.2	3.9	3.8	4.5
Total	89.3	89.4	66.0	76.4	51.8
Mean	11.2	11.2	8.2	9.5	6.5
s.d.	3.9	4.8	3.6	4.5	2.9

Private alleles

The number of private alleles, defined as alleles present exclusively in one sampling area, for each locus and for every sampling area are presented in Table 3.10. Totals, means and standard deviations for each sampling area are also shown.

The data show a similar trend to that seen in allelic richness. Sungei Bantang Recreational Forest exhibited the most number of private alleles in all 8 marker loci with a total of 23 (mean 5.1 ± 2.6 s.d.). Gunung Belumut Forest Reserve closely follows with a total of 21 private alleles (mean 4.7 ± 1.7 s.d.). The sampling area with the fewest private alleles was MacRitchie Reservoir with a total of only 4 (mean 0.9 ± 0.8 s.d.).

Data on private alleles were also rarefacted in the same process used for allelic richness. The same base sampling size of 31 individuals was used. The results of the calculations are presented in Table 3.11.

Much of the same trend is seen in the rarefacted data on private alleles as in the raw counts. When a base sampling size of 31 individuals is set, Sungei Bantang Recreational Forest still exhibited the most number of private alleles in all 8 marker loci with a total of 22.5 (mean 2.8 ± 1.7 s.d.). Gunung Belumut Forest Reserve closely follows with a total of 16.7 private alleles (mean 2.1 ± 1.1 s.d.). The sampling area with the fewest private alleles was still MacRitchie Reservoir after rarefaction with a total of 4.0 (mean 0.5 ± 0.6 s.d.).

The results of the Friedman test on the rarefacted data on private alleles indicate very significant differences in the numbers of private alleles of the sampling localities ($P \approx 0.0021$, estimated from the χ^2 distribution). The post Dunn's multiple comparisons test indicate highly significant differences in the numbers of private alleles between Sungei Bantang Recreational Forest and MacRitchie Reservoir ($P < 0.01$), and Gunung Belumut Forest Reserve and MacRitchie Reservoir ($P < 0.01$).

Table 3.10: Private alleles per locus and per sampling area. Figures are direct counts taken from allelic frequencies and do not compensate for the differences in sample sizes of each sampling area. Totals, means and standard deviations (s.d.) of private alleles for each sampling area are also shown.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	0	6	0	1	0
APMS14	4	2	3	1	1
APMS28	0	3	0	3	2
APBMS3	4	4	1	4	0
APBMS14	3	1	1	1	0
APBMS23	6	1	0	1	0
APBMS61	0	2	1	0	0
APBMS72	6	2	2	1	1
Total	23	21	8	12	4
Mean	5.1	4.7	1.8	2.7	0.9
s.d.	2.6	1.7	1.1	1.3	0.8

Table 3.11: Private alleles per sampling area after rarefaction. Base sampling size was set at 31 individuals. Totals, means and standard deviations (s.d.) of private alleles for each sampling area are also shown.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	1.5	4.0	0.4	2.1	0.0
APMS14	3.8	1.7	3.4	0.8	1.0
APMS28	1.7	3.4	0.8	2.2	1.3
APBMS3	3.6	2.5	2.5	3.4	0.5
APBMS14	2.3	1.3	1.2	1.3	0.0
APBMS23	4.8	0.9	0.4	0.6	0.0
APBMS61	0.1	1.4	0.9	0.1	0.0
APBMS72	4.7	1.5	1.9	1.4	1.1
Total	22.5	16.7	11.6	11.8	4.0
Mean	2.8	2.1	1.4	1.5	0.5
s.d.	1.7	1.1	1.1	1.1	0.6

Gene diversities

Nei's unbiased gene diversities per locus and for each sampling area are presented in Table 3.12.

Gene diversities when all marker loci are considered ranged from 0.210 to 0.961(APBMS72 at BTNR and APBMS28 at GBFR, respectively). Mean gene diversities of each sampling area were relatively uniform and ranged from 0.707 (in MacRitchie Reservoir) to 0.807 (in Gunung Belumut Forest Reserve). A Friedman test applied to the gene diversities indicated no significant difference among the different sampling localities ($P \approx 0.1209$, estimated from the χ^2 distribution).

Details on null alleles

Details on the distribution of samples with null alleles in each sampling area are presented in Table 3.13.

Sungei Bantang Recreational Forest had the most number of samples with null alleles (8) while samples from Kota Tinggi Waterfalls Resort had the fewest (4). Interestingly, the number of null alleles are positively correlated to the number of samples taken from each area. Samples may be null at a microsatellite marker for different reasons (e.g., any number of mutations at primer binding sites) and it is difficult to objectively compare null alleles as measures of diversity. Further observations on null alleles are shown in succeeding sections.

The data therefore show that high levels of genic (allelic) diversity are found in *Acanthorrhynchium papillatum*. The data also show significant differences in allelic richness and private alleles among the sampling areas. Particularly, that samples in MacRitchie Reservoir have fewer numbers of alleles than in other sampling areas. That there are no differences in Nei's gene diversity among any of the sampling areas indicate that the evenness of the distribution of the alleles are the same for these areas, despite the reduction in the number of alleles.

Table 3.12: Gene diversities of each sampling area per microsatellite locus. Mean gene diversities of each sampling area are also shown.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	0.783	0.874	0.676	0.904	0.689
APMS14	0.882	0.824	0.884	0.876	0.735
APMS28	0.907	0.961	0.903	0.928	0.904
APBMS3	0.918	0.913	0.936	0.903	0.831
APBMS14	0.870	0.903	0.829	0.918	0.694
APBMS23	0.725	0.773	0.453	0.776	0.646
APBMS61	0.576	0.702	0.394	0.565	0.521
APBMS72	0.594	0.504	0.586	0.210	0.637
Mean	0.782	0.807	0.708	0.760	0.707

Table 3.13: The distribution of samples with null alleles for each marker and for every sampling area. Totals and means per sampling area are also presented.

	Sampling area				
Marker	SBRF	GBFR	KTP	BTNR	MACR
APMS4	1	0	0	0	0
APMS14	0	6	1	0	0
APMS28	0	0	3	1	7
APBMS3	2	2	0	1	0
APBMS14	0	0	0	4	0
APBMS23	0	0	0	0	0
APBMS61	0	0	0	0	0
APBMS72	2	0	0	0	0
Total	5	8	4	6	7
Mean	0.6	1.0	0.5	0.8	0.9

3.3.5 Multi-locus genotypic diversity within sampling areas

Numbers of matching samples

The data reveal that there were samples that had the same alleles for all 8 markers. Details on these matching samples are presented in Table 3.14. (Where samples are null at the same locus, they are still considered matches when all their other alleles correspond).

Samples that shared the same alleles for all 8 markers were limited to the same sampling area, i.e., no sample matched another sample outside its own sampling area. Because samples were named according to their order of collection in the field, sample names were indicative of relative distances. The table therefore also shows that many of the matching samples were those that were close to or even adjacent to each other.

At least one pair of matching samples was found in every sampling area. Five pairs of samples matched in Sungei Bantang Recreational Forest while only two pairs of samples matched in Gunung Belumut Forest Reserve. Samples that were matched by three or more others were found in Kota Tinggi Waterfalls Resort and Bukit Timah Nature Reserve and were most often seen in MacRitchie Reservoir. Two pairs of samples and a set of three other samples matched in Kota Tinggi Waterfalls Resort. Matching samples were more abundant, both in numbers of samples that were matched by others and numbers of samples that had a match, in samples collected from Singapore localities. The sample that was matched by the most number of other samples was found in MacRitchie Reservoir (matched by 10 others).

APBL25 and APBL26 were both null at the same locus, APMS14, but matched at all others. APBT20, APBT21, APBT22 were all null at APBMS14 but matched at all others. Finally, seven samples, APMR34, APMR35, APMR36, APMR37, APMR38, APMR39 and APMR52 were all null at APMS28 but matched at all

Table 3.14: Matching multi-locus genotypes. Samples in the same row share the same alleles at all 8 markers."Reference samples" are designations for the first sample found with the multi-locus genotype and are labeled as such for convenience.

Reference samples	Matching samples
APBK01	APBK02
APBK13	APBK14
APBK26	APBK29
APBK41	APBK43
APBK46	APBK48
APBL09	APBL10
APBL25	APBL26
APKT09	APKT10
APKT14	APKT15, APKT17
APKT25	APKT28
APBT01	APBT02, APBT03, APBT05, APBT12
APBT07	APBT08, APBT10, APBT11
APBT15	APBT16
APBT20	APBT21, APBT22
APBT29	APBT30
APBT32	APBT34
APBT36	APBT37
APBT44	APBT45, APBT46
APMR01	APMR02
APMR03	APMR04, APMR05, APMR06, APMR07
APMR11	APMR12, APMR13
APMR15	APMR17
APMR21	APMR22, APMR28, APMR29, APMR31, APMR56
APMR23	APMR58, APMR59
APMR30	APMR32, APMR40, APMR41, APMR42, APMR43, APMR44, APMR45, APMR46, APMR49, APMR51
APMR34	APMR35, APMR36 APMR37, APMR38, APMR39, APMR52
APMR54	APMR55, APMR62
APMR60	APMR61, APMR63, APMR64, APMR65, APMR66, APMR67, APMR68

others.

More on matching samples are discussed below.

Multilocus genotypes

The numbers of multilocus genotypes, unique combinations of alleles for all the markers used, were counted by subtracting the number of replicated samples from the total number of samples in a sampling area. (The number of replicated samples was easily determined from Table 3.14). Data on the numbers of multilocus genotypes for every sampling area and in total are shown in Table 3.15. The total number of unique multilocus genotypes was derived from the sum of the number of unique multilocus genotypes from each of the sampling areas. This was possible since no sample was identical to any other outside its own sampling area.

Table 3.15 shows that there were replicated multi-locus genotypes in each of the sampling areas. However, almost all samples from Gunung Belumut Forest Reserve, 70 of the 72 samples collected (97.2%), were distinct from each other. In the other extreme, MacRitchie Reservoir stands out in having fewer than half of its 68 samples having unique multilocus genotypes. Genotypic richness overall indicates that about 76% of the samples have unique multilocus genotypes and that the remainder are replicates.

The discovery of matching samples in the sampling areas and the finding that not all 279 samples have unique multi-locus genotypes are not surprising given the ability of mosses to reproduce vegetatively. Two individuals of mosses that are products of the same vegetative reproduction (ramets) are expected to be clonal and will match at all loci, barring autosomal mutation at that locus. However, given the data, with what degree of confidence can we claim that two samples with matching multi-locus genotypes were indeed clones of each other and products of vegetative reproduction? To answer this question, a few more statistics needed to

Table 3.15: Unique multilocus genotypes for each sampling area and in total. "No." is the number of unique multilocus genotypes, "% total" is the relative number compared to the total number of samples.

Sampling area	No.	% of total
SBRF	46	90.2
GBFR	70	97.2
KTP	30	88.2
BTNR	39	72.2
MACR	27	39.7
Total	212	75.6

be computed.

Theoretical maximum number of multilocus genotypes

Although the raw numbers of alleles per locus may not be as ideal for measuring allelic richness as the rarefacted figures, they are still useful in other ways. The theoretical maximum number of multilocus genotypes for each sampling area may be computed from the data in Table 3.8 using the power rule of probability theory. The results of these calculations are found in Table 3.16.

Gunung Belumut Forest Reserve shows the largest figure with a theoretical maximum greater than 1 billion multilocus genotypes, that is, a potential of greater than 1 billion multilocus genotypes may be found in this sampling area. Although MacRitchie Reservoir exhibits the smallest figure, it is still fairly high at almost 8 million multilocus genotypes. While these figures may never be realized in natural populations of the moss, these figures give an indication of the extent of genotypic diversity possible in each of the sampling areas. With numbers in these orders of magnitude, the chances of finding two non-clonal samples in the field that possess the same multi-locus genotypes purely by chance is very small. Probabilities are more formally calculated using the following statistics, power of discrimination and match probabilities. These statistics are described below.

Powers of discrimination and match probabilities

Powers of discrimination, PD, and match probabilities, P_M, for each sampling area are presented in Table 3.17.

Powers of discrimination were very high for each of the sampling areas. Assuming independent segregation of the loci used, chances were better than 0.9999 that two unrelated samples will have different multilocus genotypes in any of the sampling areas. Phrasing this in terms of match probabilities, chances are less

Table 3.16: The theoretical maximum number of genotypes computed using the power rule of probability theory

Sampling area	Max No. genotypes
SBRF	7.7540×10^8
GBFR	1.7267×10^9
KTP	1.2168×10^7
BTNR	1.1532×10^8
MACR	7.9834×10^6

Table 3.17: Powers of discrimination, PD, and match probabilities, P_M, as computed for each sampling area.

Sampling area	PD	P_M
SBRF	0.999997244	2.7561×10^{-6}
GBFR	0.999999450	5.5011×10^{-7}
KTP	0.999983502	1.6498×10^{-5}
BTNR	0.999998645	1.3549×10^{-6}
MACR	0.999962923	3.7077×10^{-5}

than 1 in 100 000 that two unrelated individuals will share multilocus genotypes in any of the sampling localities. Match probabilities are least among samples from Gunung Belumut Forest Reserve ($< 1/55\ 000\ 000$) and greatest (but still very small) among samples from MacRitchie Reservoir ($< 1/370\ 000$).

These statistics strongly suggest that when samples share the same multilocus genotype, they are actually clones of each other. This is because the alternative possibility, that the match is explained by chance, is very remote.

Overall, the data suggest that vegetative or clonal reproduction is present in all sampling areas. However, it is also clear from the data that this phenomenon is found in greater degrees in sampling areas in Singapore: MacRitchie Reservoir and Bukit Timah Nature Reserve.

3.3.6 Genetic distances

Within sampling areas

Trees showing allele-sharing distances among samples within the same sampling area are shown in Figures 3.6 to 3.10. These trees help reveal patterns of distributions of genotypes within each sampling area.

Again, since the samples were named in order of their collection, sample names suggest relative sample positions and provide a reference in examining the topologies of the trees.

While some pairs of proximate samples from Sungei Bantang Recreational Forest and Gunung Belumut Forest Reserve clustered together (e.g., APBK53 and APBK54; APBL31 and APBL32, APBL35 and APBL36), no extensive clustering of samples is exhibited in trees from these areas. In these sampling areas, the spatially more proximate samples weren't always in adjacent or proximate branches in their respective trees. Also, samples that were spatially distant in the field also clustered together in their respective trees. The data therefore suggest that

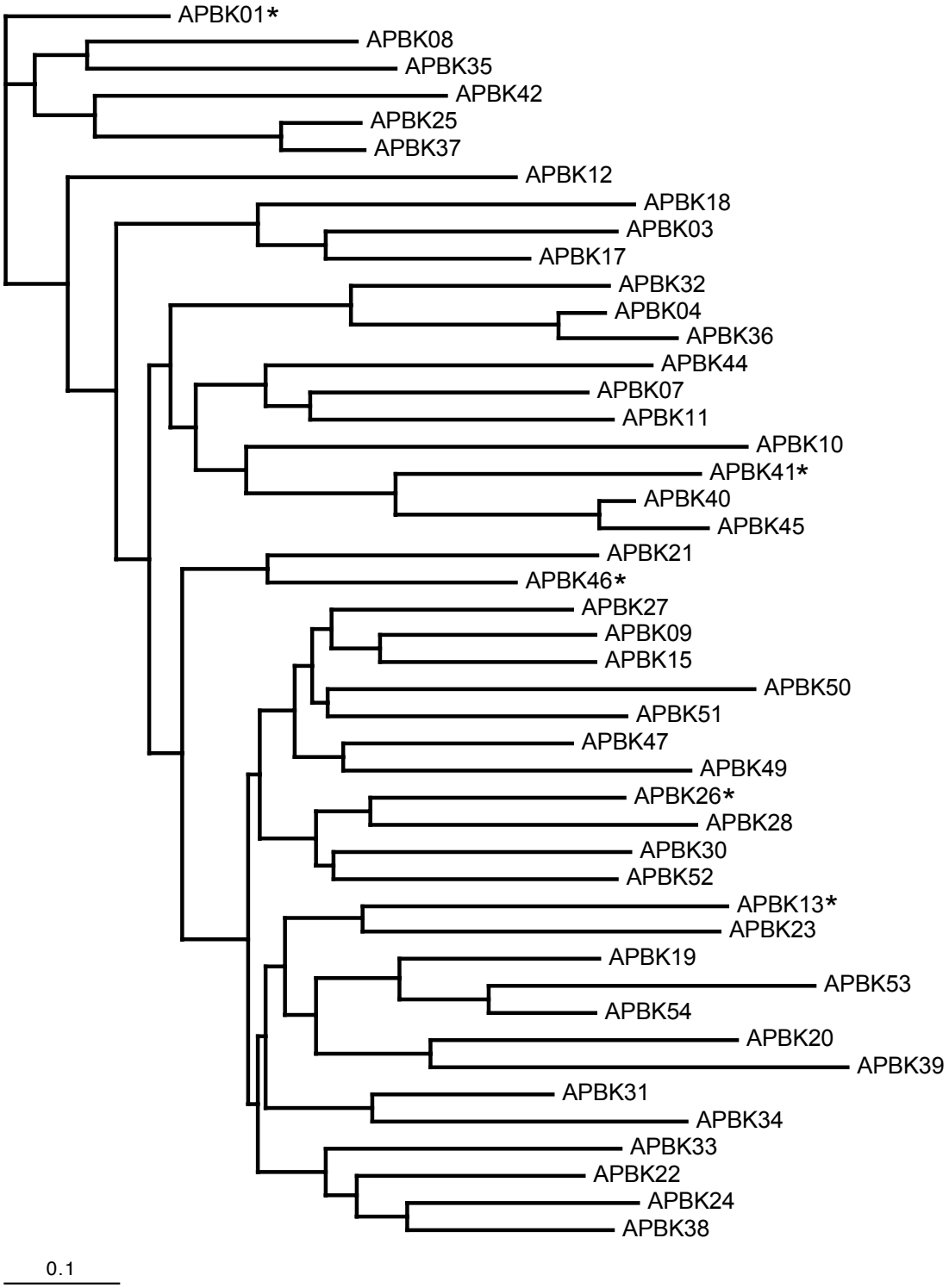


Figure 3.6: Neighbor-joining tree constructed from allele sharing distances of samples from Sungei Bantang Recreational Forest. Samples with asterisks had genotypes that matched other samples as shown in Table 3.14. Only reference samples are included in the tree. Scale bar indicates branch lengths. This tree is unrooted.

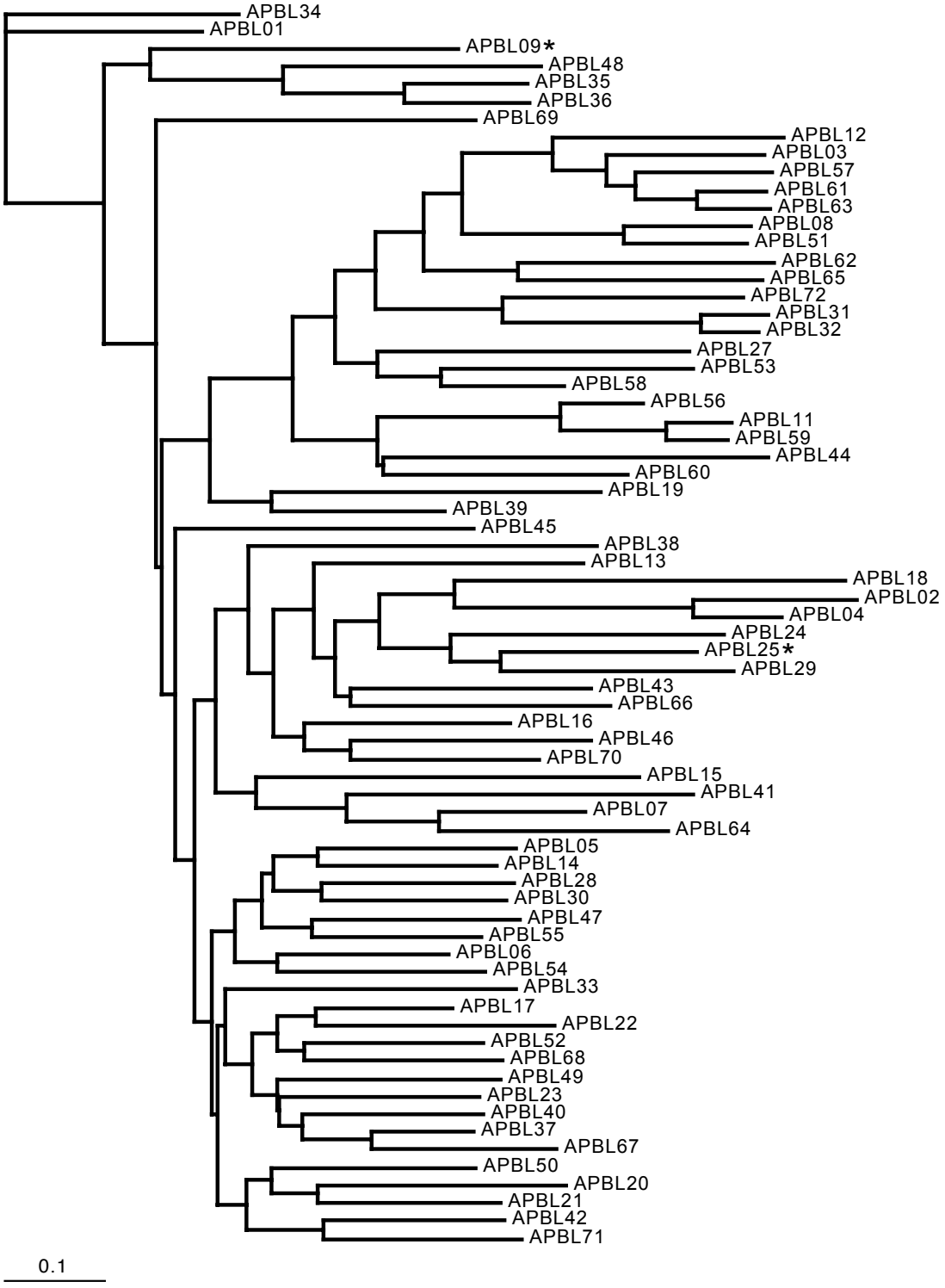


Figure 3.7: Neighbor-joining tree constructed from allele sharing distances of samples from Gunung Belumut Forest Reserve. Samples with asterisks had genotypes that matched other samples as shown in Table 3.14. Only reference samples are included in the tree. Scale bar indicates branch lengths. This tree is unrooted.

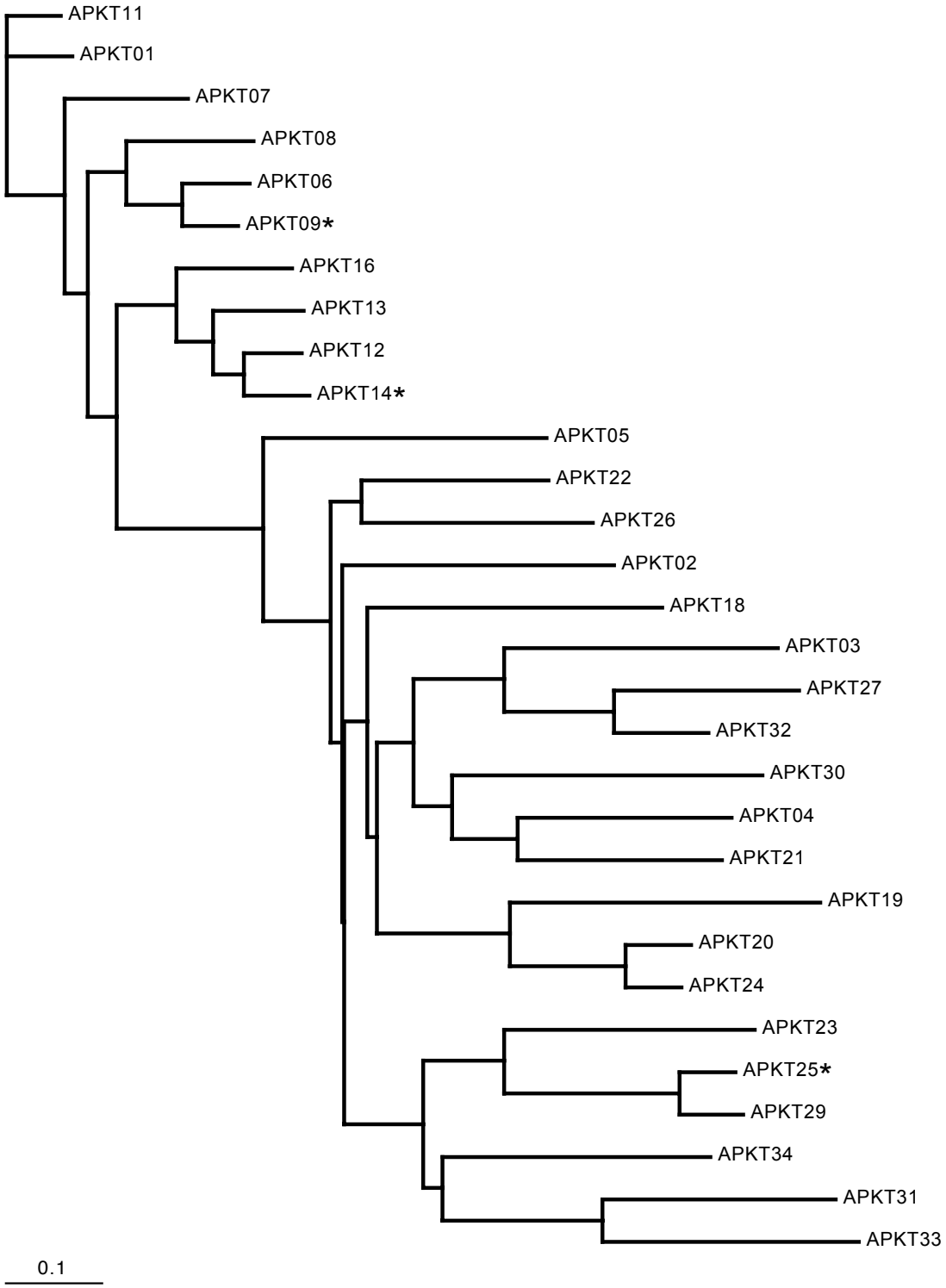


Figure 3.8: Neighbor-joining tree constructed from allele sharing distances of samples from Kota Tinggi Waterfalls Resort. Samples with asterisks had genotypes that matched other samples as shown in Table 3.14. Only reference samples are included in the tree. Scale bar indicates branch lengths. This tree is unrooted.

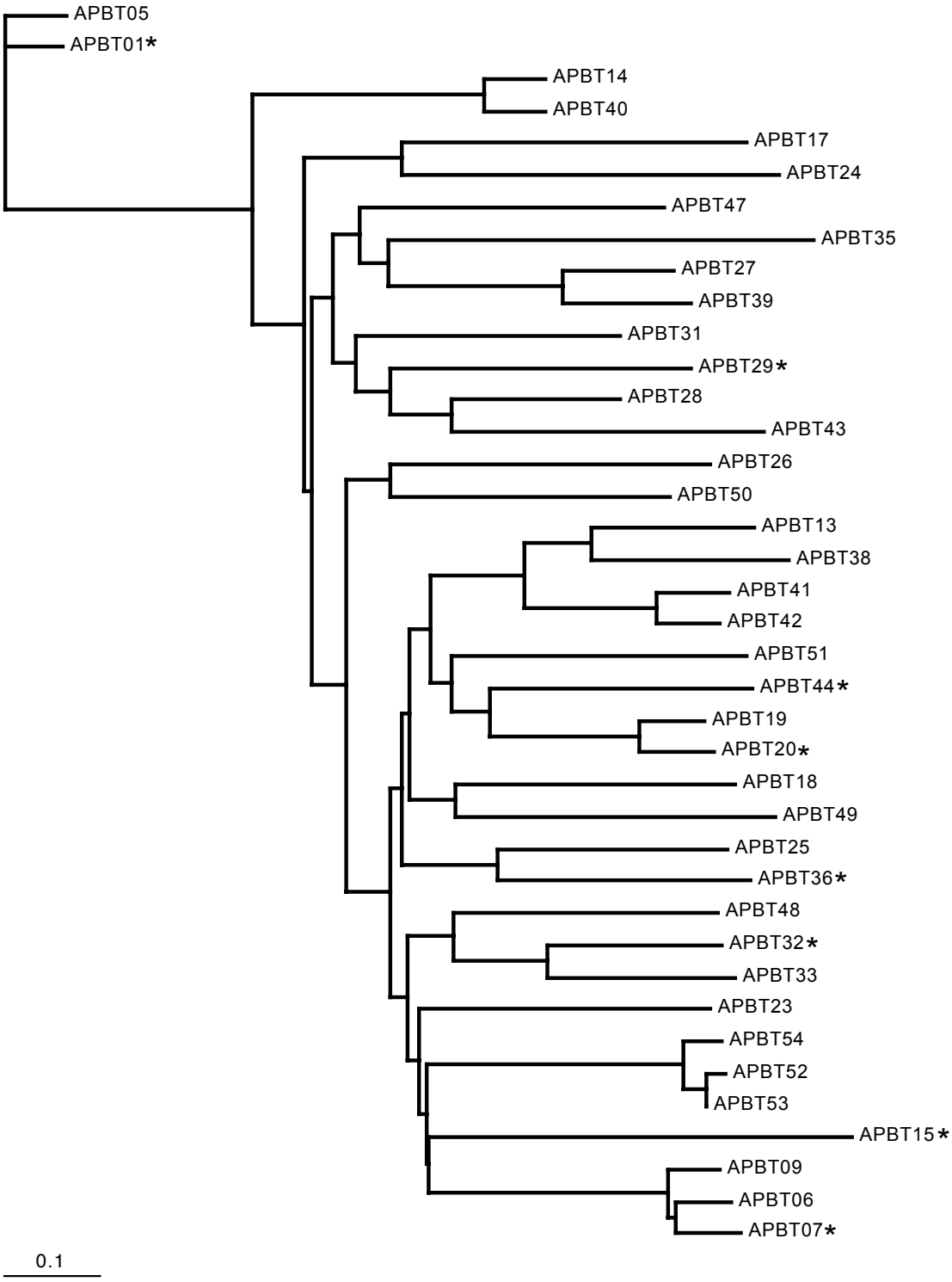


Figure 3.9: Neighbor-joining tree constructed from allele sharing distances of samples from Bukit Timah Nature Reserve. Samples with asterisks had genotypes that matched other samples as shown in Table 3.14. Only reference samples are included in the tree. Scale bar indicates branch lengths. This tree is unrooted.

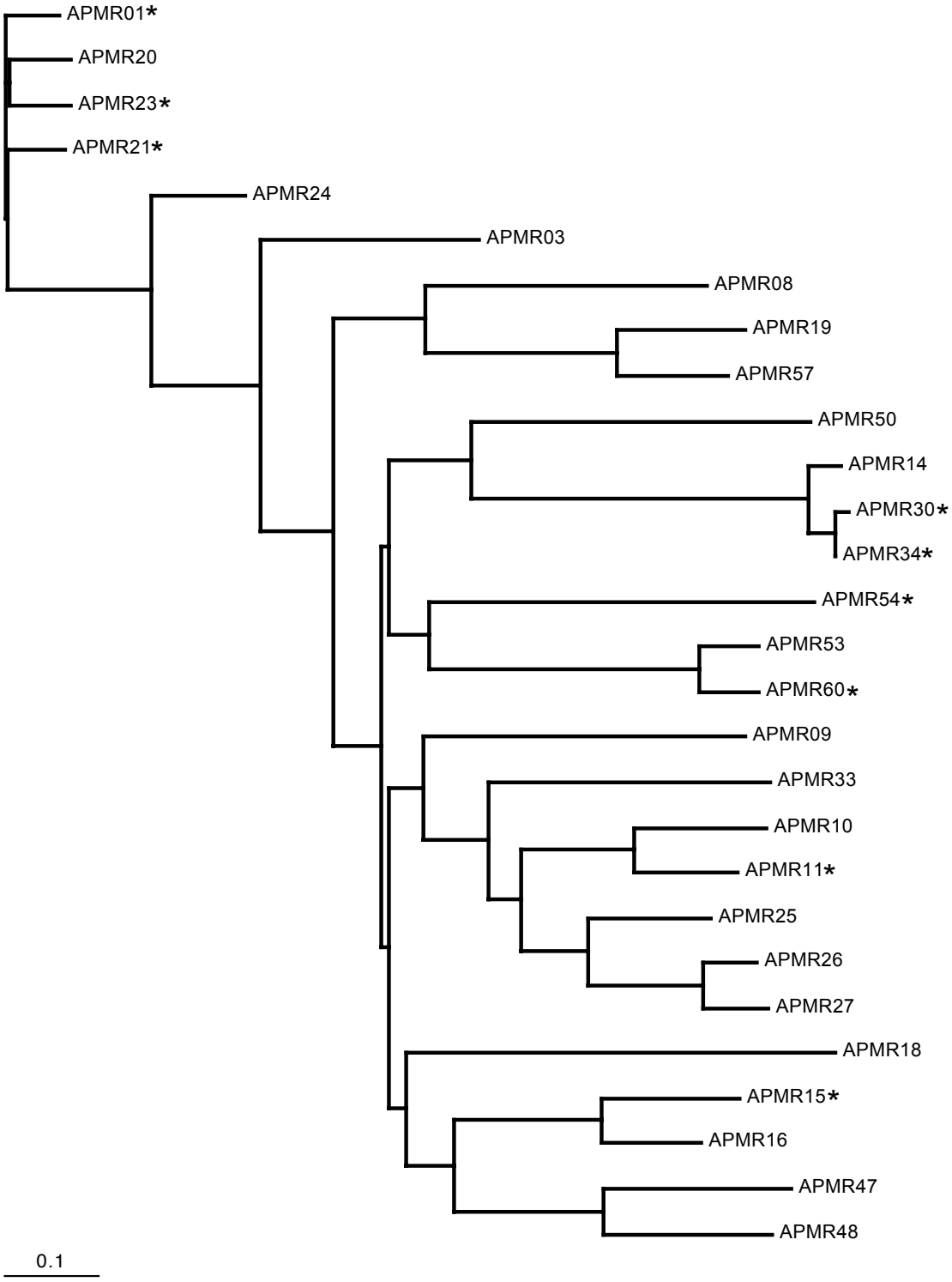


Figure 3.10: Neighbor-joining tree constructed from allele sharing distances of samples from MacRitchie Reservoir. Samples with asterisks had genotypes that matched other samples as shown in Table 3.14. Only reference samples are included in the tree. Scale bar indicates branch lengths. This tree is unrooted.

the spatial distribution of samples within these sampling areas do not follow any obvious genetic structuring.

Some clustering of several proximate samples in Kota Tinggi Waterfalls Resort is seen. Most of the samples between APKT01 and APKT17 were found together in the same cluster. Two proximate samples, APKT31 and APKT33 were also found clustering together. Many of the other samples were intermixed and do not show any other obvious clustering. The smaller number of samples collected from Kota Tinggi Waterfalls Resort, however, makes it difficult to compare these different trees. Overall, however, no obvious spatial clustering is seen.

Clustering of several proximate samples was also observed in the trees of Bukit Timah Nature Reserve and MacRitchie Reservoir. Samples from APBT06 to APBT11 all clustered together (APBT07, APBT10 and APBT11 were matching samples). Samples from APBT19 to APBT22 all clustered together (APBT20, APBT21 and APBT22 were matching samples).

In MacRitchie Reservoir, APMR01, APMR20, APMR21, APMR23 and all their respective matching samples were found clustered together. Also found clustered together were APMR25, APMR26 and APMR27.

Overall, the trees constructed from the genetic distances of samples within the same sampling area indicate that some clustering of proximate samples can be found, indicating that genetically similar individuals can be found close to each other spatially. However, spatially distant individuals can also be found clustered together in the trees of genetic distance. The lack of correlation between spatial distance and genetic distance using microsatellite markers therefore suggest that this marker is not very useful in showing dispersal patterns of this moss within the same population.

Among sampling areas

Trees of pairwise genetic distances using the neighbor-joining algorithm are shown in Figure 3.11.

Neighbor-joining trees were constructed from both Nei's standard genetic distance, D, and net genetic distance, D_A, show similar tree topologies. The trees show the Malaysian sampling areas clustering together and the Singaporean sampling areas clustering together. However, support for the clades of the Malaysian sampling areas are strong in the tree constructed from D but is weaker in the tree constructed from D_A. As seen in either tree, MacRitchie Reservoir seems to be genetically most distant from the rest as seen by the relatively long branch length that separates this sampling area from the others.

The results of the Mantel test did not show significant correlation between geographic distances of the sampling areas and genetic distances among the sampling areas. The correlation coefficient when the standard genetic distance was investigated was -0.179 while the correlation coefficient when the net genetic distance was investigated was -0.325.

3.3.7 Population differentiation

The global test of population differentiation on multi-locus genotype frequencies indicated that there was significant differentiation among populations (Non-differentiation $P = 0$). Likewise, the results of the exact test on all pairs of populations indicated differentiation among the populations. For each pairwise comparison, the probability that the observation was as extreme as or more extreme than the data under the null hypothesis is $P = 0$. These are expected results given that no samples from any of the different populations shared the same multi-locus genotype (all multi-locus genotypes of a population were private to a population).

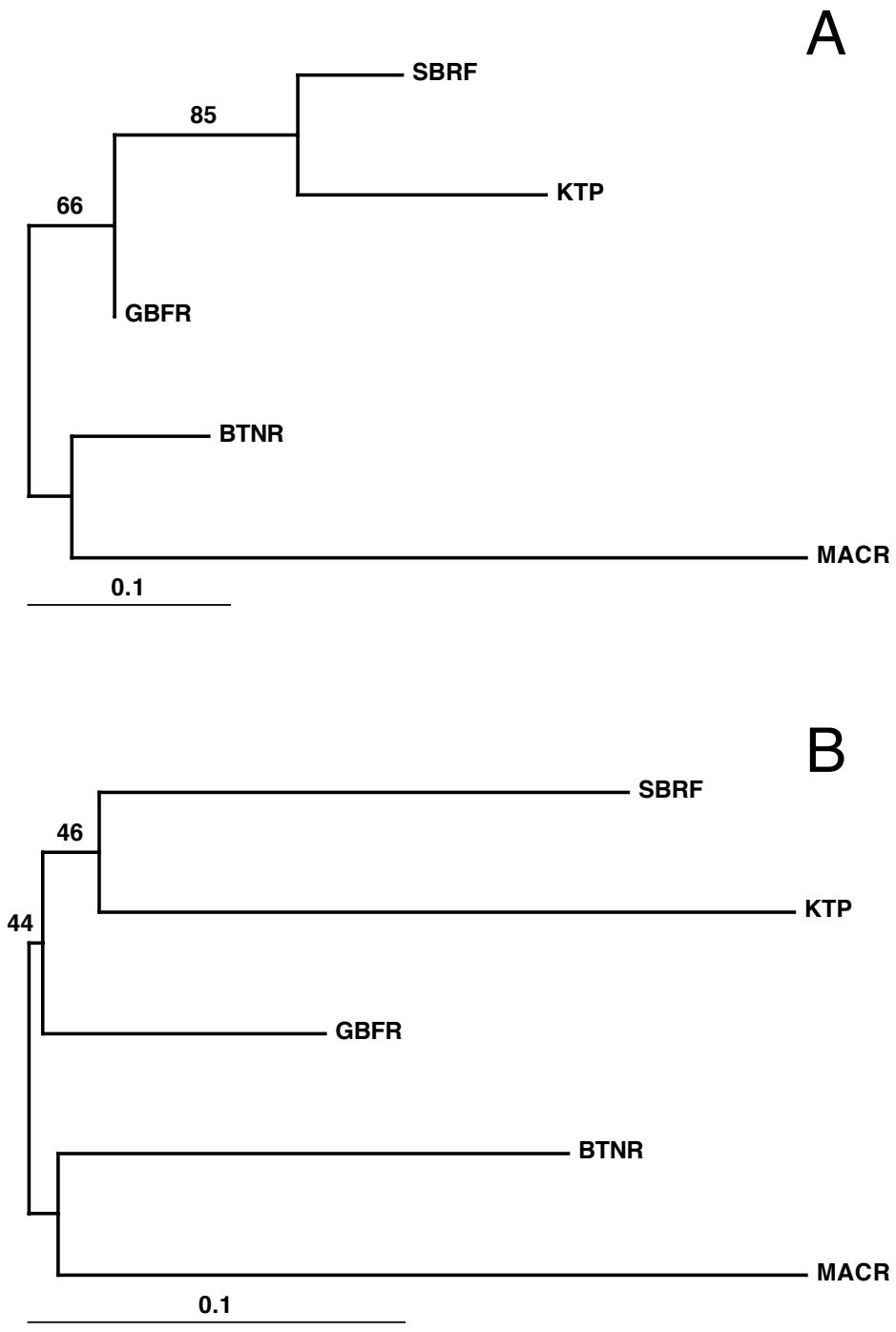


Figure 3.11: Neighbor-joining trees constructed from genetic distances among sampling areas. Tree A is generated from Nei's standard genetic distance, D. Tree B is generated from Nei's net genetic distance, D_A. Bootstrap percentages are shown in adjacent branches where applicable. Scale bars indicate branch lengths. This is an unrooted tree.

Similar results were obtained with the exact tests of population differentiation on allelic distributions. The probability of accepting the null hypothesis that there was no differentiation among the populations was $P < 0.001$ for the global test and each pairwise test. The significant presence of rare and private alleles probably contribute to these results.

The results therefore strongly indicate that both the genotypic and allelic distributions were different across populations.

3.3.8 Population genetic structure

The results of AMOVA at the haplotype level are shown in Table 3.18.

The results show that almost 91% of the variation in the data can be explained by variation within populations (sampling areas). Only about 7% can be explained by the variation between or among populations within groups and only 2% can be explained by variation among groups (groups were defined as Singapore and Malaysia).

AMOVA results were also used to compute fixation indices. The computations suggest that there is moderate genetic differentiation among populations within groups, $F_{SC} = 0.06978$, and that this statistic is significantly different from zero ($P = 0$). There is also moderate genetic differentiation among individuals among populations and among groups, $F_{ST} = 0.08866$, and that this statistic is also significantly different from zero ($P = 0$). Finally, there is little genetic differentiation between the two groups of populations, $F_{CT} = 0.02030$ but that this measurement may not be significantly different from zero ($P \approx 0.1$). The data from AMOVA therefore suggest that the moss populations in the different sampling areas are moderately genetically isolated from each other.

Population pairwise F_{ST} values computed using microsatellite markers are shown in Table 3.19. All F_{ST} values are significant ($P = 0$).

Table 3.18: Results of analysis of molecular variance, AMOVA, for microsatellite data. d.f.=degrees of freedom.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	25.069	0.06543	2.03
Among populations within groups	3	44.117	0.22032	6.84
Within populations	274	804.767	2.93711	91.13
Total	278	873.953	3.22285	

Table 3.19: Pairwise F_{ST}s between populations using microsatellite markers.

	SBRF	GBFR	KTP	BTNR	MACR
SBRF	0.00000				
GBFR	0.02346	0.00000			
KTP	0.05360	0.05971	0.00000		
BTNR	0.05797	0.02782	0.08876	0.00000	
MACR	0.13763	0.09185	0.16236	0.11508	0.00000

The results suggest that there is some genetic differentiation between SBRF and GBFR and between GBFR and BTNR, and that there is moderate genetic differentiation between most other pairs of populations. The results suggest that MACR is moderately to greatly genetically differentiated from all the other populations.

3.4 Discussion

The results just presented describe the genetic diversity of *Acanthorrhynchium papillatum* as measured by microsatellite markers, helping fulfill the second and third objectives of the thesis. Substantial levels of genetic variation were detected at several scales of measurement. Hardly significant differences in gene-level variation were observed in populations that were considered to be more disturbed. Larger decreases in multi-locus genotypic variation were observed (indicating a greater tendency to clonal reproduction) for these populations. Beyond the numbers and statistics that can be generated from any data set, however, are interesting observations and insights into the biology of the moss that might have implications on this species' chances of survival in their increasingly disturbed natural habitats.

3.4.1 Allelic diversity explains much of the diversity in Acanthorrhynchium papillatum

The most salient finding in this chapter is that overall, genetic diversity is high in *Acanthorrhynchium papillatum*. The results contest the early notion that moss populations have low levels of variation. They also strengthen the findings of more recent studies that moss populations possess levels of variation that match or rival those of higher plants. Moreover, the extensive numbers of alleles and genotypes found suggest that genetic diversity in this species of moss may be higher

than in the other few, temperate, mosses studied using microsatellite markers [*Polytrichum*: see van der Velde & Bijlsma (2003, 2000); van der Velde *et al.* (2001b, 2000); *Sphagnum*: Provan & Wilson (2006); Wilson & Provan (2003)].

It is evident that at the root of these high levels of genetic diversity are the large numbers of alleles found in each locus. Numbers of alleles (allelic richness) for 7 out of the 8 marker loci studied were in double-digits. Even the least allele-rich locus, APBMS72, with 8 alleles was nearly so. These numbers were obtained after the examination of fewer than 300 samples in 5 sampling areas; if more samples were obtained and studied, potentially more alleles could also have been found.

With such large numbers of alleles in fairly even distributions, alleles may be combined to potentially create large numbers of unique multi-locus genotypes. Indeed, both actual and theoretical figures indicate the presence and potential of many unique multi-locus genotypes in the overall data set.

Therefore high levels of allelic diversity in *Acanthorrhynchium papillatum* have effects on both allelic and genotypic levels of variation. Moreover, they also have effects on other measures of genetic diversity.

The large numbers of alleles also affected the topologies of trees constructed to reveal patterns of relationships of samples within the same sampling area. These trees were constructed from allele-sharing matrices of samples within the same area. The results do indeed show some clustering of samples, particularly those from more disturbed populations in MacRitchie Reservoir and Kota Tinggi Waterfalls Resort. The overall findings, however, do not reveal distinct patterns or clustering and suggest that the distributions of alleles and genotypes within the same sampling area were generally even. The lack of discernible patterns in the results may be the result of the presence of the large number of alleles in each population. With more alleles available, the number of shared alleles between samples are reduced and patterns of relationships among samples in the same area become

less evident.

Another apparent effect of the large numbers of alleles is seen in the results on the two tests on population genetic differentiation, at the allelic level and at the level of individuals as measured through multi-locus genotypes. The results of the exact tests suggested significant differentiation among the populations. Indeed, the test need not have been performed at the level of multi-locus genotypes to arrive at this conclusion. Since no genotypes are shared among populations (owing largely to the large numbers of alleles available) then the result of the test on genotypic differentiation would predictably show that the populations were differentiated. The extensive presence of both rare and private alleles help explain the differentiation that is seen among sampling areas at the allelic level.

3.4.2 Microsatellite loci of Acanthorrhynchium papillatum *experience high rates of mutation*

Given the pervasive and broad effects of the large numbers of alleles present in the data set, a pressing question now arises: what is the source of these large numbers of alleles and high levels of allelic diversity? Several observations in the data suggest that high rates of mutation are responsible for these findings.

The substantial numbers of rare and private alleles in the data set suggest that there are high rates of mutation in the microsatellite loci examined. There is, however, an alternative explanation for the presence of rare and private alleles. They could be rare and private alleles only because of insufficient sampling. That is, they possibly could also be found in adjacent or surrounding areas but that samples were not collected in these areas. If more samples in more sampling areas were collected and examined, then these present rare and private alleles may be found in larger numbers also and would not be rare and private anymore. However, other observations corroborate the hypothesis of high rates of mutation.

Most studies using microsatellite markers show data with amplicon sizes that are different from each other by multiples of the length of their motif, i.e., dinucleotide microsatellite markers should have amplicons that are different by multiples of 2 bp, trinucleotide markers should have amplicons that are different by multiples of 3 bp, etc. The data from the present study clearly show that mutation of the microsatellite markers did not proceed in a stepwise fashion: most of the markers had core sequences of dinucleotide repeats but the data show many amplicons that were a single base apart. Mutations that can cause changes in the repeat units of a microsatellite marker, e.g., the deletion of a single "A" from a core sequence of "GA" repeats can explain the presence of amplicons that do not fit into the stepwise mutation model.

Despite very low match probabilities in each of the sampling areas, several samples were observed that matched at 7 out of 8 loci. If these samples were products of vegetative reproduction, i.e., if they were the same ramets, then they should match at all loci. If instead they were different individuals or from different genets, then how could they match at 7 out of 8 loci when the chances of this happening by chance are extremely low? One possible explanation to these observations is that the samples were indeed vegetatively reproduced but that the locus where the samples do not match mutated only after the samples had split. Mutation levels in samples in the populations should be high enough for this to be detected in several instances in the data set.

Finally, the substantial number of null alleles found in the data may also be explained by high rates of mutation. Microsatellite loci and their surrounding sequences are known to be areas of higher levels of mutation. This fact, after all, is the basis of using inter-simple sequence repeats (ISSRs) as molecular markers. Mutations in the core sequence can generate different detectable microsatellite alleles but mutations in the flanking sequence can generate null alleles if primers

for the locus can no longer bind as a result of the mutation.

High rates of mutation in *Acanthorrhynchium papillatum* may have important implications. In the face of restrictions in either genetic recombination (sexual reproduction) or gene flow (migration), genetic variation can be maintained in an isolated population through somatic mutation. This source of variation is enhanced by the ability of the moss to reproduce vegetatively, producing more individuals in the absence of sexual reproduction that can themselves undergo somatic mutation. This is an interesting finding that may have consequences in the survivability of isolated populations of this moss.

3.4.3 Differences in diversity among sampling areas and population genetic structure

Despite the confounding observations that may be caused by the large numbers of alleles present in the data set, the results still suggest that degrees of diversity, both at the gene and genotypic levels, were not the same for all the sampling areas studied.

At the gene level, significantly fewer total numbers of alleles and private alleles were found in MacRitchie Reservoir, an area that was adjudged to be more disturbed than the rest. This suggests a founder event where only a subset of the total number of alleles was available in the area. Interestingly, Nei's gene diversity is apparently unaffected in these areas as seen in the test comparing this statistic across the different populations. This is perhaps because even in disturbed areas, large (but reduced) numbers of alleles were maintained in similarly even patterns of distribution. Here again the effects of having large numbers of alleles are seen. The results further imply that mutation rates are similar in populations of different degrees of degradation.

At the genotypic level the differences were starker. While matching samples

(and therefore, the incidence of vegetative reproduction) were found in each population, the numbers of matching samples in MacRitchie Reservoir were found in such a degree that chances were better than 50% that a sample taken at random from this sampling area would be a clone of another. Interestingly again, large numbers of clumps were still found in this area and so it seems that, in MacRitchie Reservoir, clonal reproduction is important in maintaining the numbers of individuals of this moss.

The data therefore suggest that, potentially, despite isolation and the restriction of gene flow suggested by the fixation indices obtained, populations of *Acanthorrhynchium papillatum* can still reproduce clonally and produce adequate numbers of individuals. These individuals may all experience somatic mutation, increasing the levels of variation within the population that may all be recombined when opportunities for sexual reproduction arise.

3.4.4 Limitations

As has been shown, the large numbers of alleles found in *Acanthorrhynchium papillatum* may help populations of the species survive isolation and restrictions of gene flow. Ironically this and a few other characteristics of the species create limitations in the analysis of data.

Microsatellite data are often analyzed with the assumption that loci mutate according to the stepwise mutation model. Statistics assuming this model, e.g., R_{ST}, the analog of F_{ST}, could provide additional information on the diversity and relationships of samples and populations of the species studied. In the case of the present data, all the loci studied had alleles that do not fit into a strictly stepwise mutation model, presumably because of high rates of mutation. Statistics assuming the stepwise mutation model therefore could not be computed and potentially telling information was lost.

The tendency of *Acanthorrhynchium papillatum* to undergo vegetative reproduction also presents problems in computations. As earlier alluded in the chapter on the development of microsatellite markers, the presence of matching samples creates problems in computations of linkage disequilibrium, a test on whether the loci being studied are linked or not. Even when matching samples are excluded from the data set used to check for linkage disequilibrium, significant levels of linkage are still seen between pairs of loci. Presumably, these significant levels of linkage may be due to samples that match at 6 or 7 out of 8 loci, i.e., clonal samples that had mutated after they had separated. Too many of these in various combinations are present in the data that excluding them would have been both impractical and subject to too much bias. Nevertheless, some indication of the independence of pairs of loci is seen when matching samples are excluded suggesting that the loci are not linked.

3.5 Conclusions

In this chapter, the genetic diversity of *Acanthorrhynchium papillatum* was characterized using microsatellite markers, helping complete the second and third objectives of the thesis.

One small but significant finding in this chapter was not on the genetic diversity of *Acanthorrhynchium papillatum* but was rather an observation made on the methodology. The type of instrument and other experimental conditions used may affect the size-calling of labeled PCR fragments like microsatellite markers. It is therefore important that the same type of instrument and the same experimental conditions be used throughout the course of the same study.

The results obtained indicate high levels of genetic diversity at all levels explored and for all populations studied. Differences in allelic richness and genotypic diversity were detected among the populations, however, suggesting some slight

evidence that diversity is still affected by habitat quality. Higher levels of diversity were seen in areas that were considered less degraded. Genotypic diversity, in particular, was lower in areas that were considered more disturbed. In these areas, vegetative reproduction seemed to be more important as more than half of the samples collected was a clone of another.

At the root of these high levels of diversity is extensive allelic diversity: large numbers of alleles that were fairly evenly distributed in each of the sampling areas examined. High levels of allelic richness were attributed to high mutation rates as supported by several observations.

The suspected high rates of mutation are thought to have interesting implications on the biology and survivability of the moss species. For example, it is proposed that the isolation of a population of this moss poses no immediate threat to its survivability. The mosses can revert to vegetative reproduction, increasing the number of individuals in the population, and then allow its inherent high rates of mutation to restore variability in the population.

To complete the picture on the population variability of *Acanthorrhynchium papillatum*, diversity needed to be examined at one more level. All the data that have been shown so far involve single samples from each clump of moss. Since *A. papillatum* also forms clumps composed of many individual branches of moss, diversity at this level –within clumps of the moss– needed to be examined. The results of this study are presented in later parts of this thesis.

However, before the study on the diversity within clumps of *Acanthorrhynchium papillatum* is shown, one other study is presented first. Diversity levels between and among clumps of *A. papillatum* were determined using another marker, sequence information from the second internal transcribed spacer, *ITS*2. This was done to provide additional information on the diversity of the moss. The description and results of this study are found in the next chapter.

Chapter 4

GENETIC DIVERSITY AMONG CLUMPS OF *ACANTHORRHYNCHIUM PAPILLATUM* AS MEASURED BY VARIATION IN *ITS2* SEQUENCES

4.1 Introduction

In addition to data from microsatellite markers, DNA sequence data were also analyzed in order to provide an independent and supplementary source of information on the genetic diversity of *Acanthorrhynchium papillatum*. A suitable marker had to be found and so candidate markers from literature were evaluated. These included the chloroplast genes, *rbcL* (Chiang & Schaal, 1998; Chang, 2003) and *rpl*16 (Assmussen, 1999), the chloroplast intergenic spacer, *trnL-trnF* (Taberlet *et al.*, 1991), and both nuclear internal transcribed spacers, *ITS* (Sun *et al.*, 1994; White *et al.*, 1990).

Despite repeated attempts, only the internal transcribed spacer sequences proved viable for sequencing and subsequent analysis. *rbcL* was too long and required at least four sets of primers to sequence completely. Preliminary results also showed very little variation in this gene among test samples. These early results suggested that the efforts and costs required to sequence *rbcL* were not worth the data that could be generated. In the case of *trnL-trnF*, the sequences were too short and so, again the benefits of sequencing this locus were thought to not outweigh the costs. Finally, *rpl*16 could not be successfully amplified and had to be abandoned early in the project. *ITS* sequences could easily be amplified and sequenced but possessed attributes that needed careful consideration.

The primers for amplification of *ITS* were derived from fungal species (White *et al.*, 1990). To reduce the probability of amplifying and sequencing any potential fungal contaminants in the moss DNA samples, new primers were designed and used. Primers for amplification of the second internal transcribed spacer region, *ITS*2, were designed from sequences found in public repositories.

The characteristics of DNA sequence data necessitated different analyses procedures and the computation of different indices of diversity from those used on microsatellite data. Thus, direct comparisons of the same indices of diversity computed from microsatellite data and from DNA sequence data were limited to statistics common to both DNA and microsatellite markers, e.g., analysis of molecular variance. However, as will be seen, *ITS*2 still provides a valuable source of information on the genetic diversity of *Acanthorrhynchium papillatum*.

4.2 Materials and Methods

4.2.1 Design of primers for ITS*2 amplification and sequencing*

New primers needed to be designed to preclude amplification and sequencing of potential fungal DNA contamination in the moss samples. Primers for amplifying the *ITS*2 region of *Acanthorrhynchium papillatum* were designed from consensus sequences of the 5.8S rRNA and 25S rRNA genes of three moss species, *Bryum capillare*, *Bucegia romanica* and *Hookeria lucens*. Sequences from these moss species were sourced from GenBank (Bilofsky & Burks, 1998) accessions (Table 4.1). Sequence alignment to identify consensus regions was performed using ClustalW version 1.83 (Thompson *et al.*, 1994) using default parameters. Primers were designed manually (Table 4.2) but were checked for suitability for use in PCR using Primer3 (Rozen & Skaletsky, 2000).

Table 4.1: GenBank accessions used in designing primers to amplify the *ITS*2 region of *Acanthorrhynchium papillatum*.

GenBank accession number	Species	Included loci
AJ252136	*Bryum capillare*	18S rRNA gene (partial), internal transcribed spacer 1 (*ITS*1, 5.8S rRNA gene, internal transcribed spacer 2 (*ITS*2), 25S rRNA gene
AJ251929	*Bucegia romanica*	18S rRNA gene, internal transcribed spacer 1 (*ITS*1), 5.8S rRNA gene, internal transcribed spacer 2 (*ITS*2), 25S rRNA gene
AJ252137	*Hookeria lucens*	18S rRNA gene (partial), internal transcribed spacer 1 (*ITS*1), 5.8S rRNA gene, internal transcribed spacer 2 (*ITS*2), 25S rRNA gene (partial)

Table 4.2: Primers designed to amplify the *ITS*2 region of *Acanthorrhynchium papillatum*. T_M (melting temperature) values were computed using the online tool of FinnZymes, Finland at `http://finnzymes.com/Java/tm_determination.htm` using default parameters.

Oligo name	Sequence (5'–3')	T_M
ITS forward	`ACTCTCAGCAACGGATATCTTGG`	65.6 °C
ITS reverse	`ATATGCTTAAACTCAGCGGGTAGTC`	64.5 °C

4.2.2 PCR amplification

The same DNA samples used to analyze variation among clumps of *Acanthorrhynchium papillatum* with microsatellite markers were also used for DNA sequence analysis.

The *ITS*2 regions of samples of *Acanthorrhynchium papillatum* were sequenced by first isolating and amplifying this region through PCR. Hot-start PCR was performed to increase the sensitivity and specificity of PCR amplifications and to minimize the formation of primer dimers. Ten microliter reactions of 1×F-522 buffer (FinnZymes, Finland), 5% DMSO (Sigma, USA), 200 μM dNTPs (Promega, USA), 0.5 μM each of *ITS* forward and reverse primers, 0.24 U DyNAzyme™ II Hot Start DNA polymerase (FinnZymes, Finland) and approximately 20 ng genomic DNA were combined in thin-walled, 0.2 mL tubes. Reactions were performed in a PTC-100® (MJ Research, USA) thermocycler using the following thermal profile: Initial denaturation for 10 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 30 s at 62 °C, 40 s at 72 °C, and final extension for 10 min at 72 °C. Products were kept at 16 °C until recovery.

PCR products were purified using an AMPure® kit (Agencourt, USA) following the manufacturer's protocol. The purified products were used as templates for subsequent cycle sequencing reactions.

4.2.3 Sequencing

The purified PCR products were sequenced using BigDye® Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems, USA) from both forward and reverse ends using the same primers used for PCR amplification.

Sequencing reactions were set up according to the manufacturer's protocol but were scaled to a quarter volume of the full-strength reactions. Cycle sequencing reactions were performed on a PTC-100® (MJ Research, USA) thermocycler

using the following thermal cycling profile: Initial denaturation for 1 min at 96 °C, followed by 25 cycles of 10 s at 96 °C, 5 s at 50 °C, 4 min at 60 °C. Extension products were kept at 4 °C until recovery.

Extension products were purified using CleanSEQ® kits (Agencourt, USA) following the manufacturer's protocol and eluted in Milli-Q® (Millipore, USA) water.

Purified extension products were sequenced in an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, USA) using a 50 cm capillary array and POP-6™ polymer with default run and analysis modules.

4.2.4 Basecalling, contig assembly and alignment

Basecalling was performed using the KB basecaller version 1.1 of the software package Sequencing Analysis version 5.2 (Applied Biosystems, USA). Forward and reverse sequences were assembled using Phrap (Phil Green, `http://www.phrap.org`). Assembled contigs were aligned with ClustalW version 1.83 (Thompson *et al.*, 1994) using default parameters. The alignment was visually checked and, where necessary, edited with the help of MacClade version 4.05 (Dwayne R. Maddison and Wayne P. Maddison, `http://www.macclade.org`).

4.2.5 Data analysis

Aligned *ITS*2 sequences were used for analysis.

To measure genetic diversity, nucleotide and haplotype diversity indices were calculated using Arlequin version 3.11 (Excoffier *et al.*, 2005b). The software was also used to determine the number and distribution of the different *ITS*2 haplotypes in the data set.

To measure the relatedness of the *ITS*2 haplotypes, genetic distances were computed. The genetic distance between each *ITS*2 haplotype was computed from

the absolute number of differences between each haplotype. Genetic distances were calculated with the aid of PAUP* version 4.0b10 (Swofford, 2002). A neighbor-joining tree (Saitou & Nei, 1987) was constructed from the genetic distances using the same program. Trees were visualized using TreeView (Page, 1996).

To test if mosses from the different sampling areas belonged to the same population, exact tests of population differentiation (Raymond & Rousset, 1995a) were performed at the global level and by pairwise comparisons of haplotype frequencies in each sampling area. Both global and pairwise tests were performed using 1 000 000 Markov steps.

Population genetic structure was also inferred by computing fixation indices using analysis of molecular variance, AMOVA (Excoffier *et al.*, 1992; Michalakis & Excoffier, 1996), and through conventional fixation indices by pairwise comparisons of each sampling area. For AMOVA, groups and populations were defined in the same way as in the microsatellite analysis of genetic variation among clumps of *Acanthorrhynchium papillatum*: one group formed by the Malaysian sampling areas (SBRF, GBFR, KTP) and one group formed by the Singaporean sampling areas (BTNR and MACR). AMOVA was also used to determine the partitioning of variation, that is, to determine how much of the variation observed is due to variation among groups of populations, among the populations and within the populations.

All tests of population genetic differentiation were performed using Arlequin version 3.11 (Excoffier *et al.*, 2005b).

4.3 Results

4.3.1 Sequencing difficulties

Of the 279 samples of *Acanthorrhynchium papillatum* available for analysis, only 193 yielded sequences with unambiguous base calls throughout the length of their

*ITS*2 region. The breakdown of the distribution of the numbers of usable samples is shown in Table 4.3.

Samples that could not be included in the analysis were those that yielded sequences with ambiguous basecalls from both forward and reverse ends (or 5.8S and 25S ends, respectively). Repeated attempts to sequence these samples did not improve the quality of the sequences. These results are indicative of the presence of more than one template in each of these samples. The literature suggests that these multiple templates may be from intragenomic (intra-individual) variation found among the different repeat units of rDNA (Ruggiero & Procaccini, 2004). Clone-mediated sequencing could potentially have revealed more information on this source of variation. However, too many samples possessed intragenomic variation that clone-mediated sequencing could not be performed given the time and resources limitations.

From Table 4.3 it can be seen that a fairly even percentage of samples of *Acanthorrhynchium papillatum* from each sampling area were made unavailable for analysis because of unambiguous basecalls. Approximately 30% of the total number of samples from each sampling locality could not be used. The figures suggest that no apparent correlation exists between the number of unusable sequences and the habitat quality of the source samples. The majority of sequences from each of the sampling areas were of high-enough quality to be used for analysis of genetic diversity of populations of *A. papillatum*.

4.3.2 Characteristics of ITS*2 in* Acanthorrhynchium papillatum

High-quality sequences of *ITS*2 used for analysis of genetic diversity among clumps of *Acanthorrhynchium papillatum* are shown in Appendix K. Because of the number of samples involved, the sequences are presented in FASTA (Pearson, 1990) format for better readability. Gaps to accommodate sequence alignment are indi-

Table 4.3: Numbers of samples of *Acanthorrhynchium papillatum* with unambiguous basecalls throughout the length of their *ITS*2 regions. Percentages of the total number of samples available for each sampling area are also shown.

Sampling area	No. usable	% of total
SBRF	33	64.7
GBFR	51	70.8
KTP	23	67.6
BTNR	33	61.1
MACR	53	77.9
Total	193	69.1

cated in the sequences.

The length of the *ITS*2 region in *Acanthorrhynchium papillatum* ranged from 325 bp (APBL4 and APBL9) to 369 bp (APBL15 and APBL37) with an average of 341.5 bp. Because of the many insertions/deletions found in the *ITS*2 region of *A. papillatum*, the length of the aligned sequences stretched to 381 bp. *ITS*2 in *A. papillatum* is GC-rich with an average of 72.3% GC-composition.

Table 4.4 shows the summary statistics related to the sequences of *ITS*2 from samples of *Acanthorrhynchium papillatum*.

While large differences in the numbers of usable sequences from each sampling area prevent direct comparisons from being made among these localities, the summary statistics still provide information about the overall characteristics of the variation of *ITS*2 in *Acanthorrhynchium papillatum*. A lot of the variations in *ITS*2 in *A. papillatum* was in the form of insertions/deletions or indels. There were about 6–7 common indels (again, see Appendix K), 4–12 bases in size. Several other, smaller indels were also found. There were more than 30 sites of base substitutions, either transitions or transversions. However, the majority of the bases in the *ITS*2 region of *A. papillatum* were conserved or nearly conserved in all the samples analyzed.

From the data, it is evident that the *ITS*2 region in *Acanthorrhynchium papillatum* is very variable. Both base substitutions and indels were common in samples from all areas. Samples from the more pristine localities in Malaysia (SBRF and GBFR) exhibited more substitution and indel sites than samples from areas considered to be more degraded (KTP, BTNR, MACR). There are more polymorphic sites in the more pristine areas as well. Again, however, these observations may just be an effect of the different numbers of samples available from the sampling areas studied.

Table 4.4: Summary statistics of nucleotide-level variation of the *ITS*2 region in *Acanthorrhynchium papillatum*. ts. – transition, tv. – transversion, s.d. = standard deviation.

Statistics	SBRF	GBFR	KTP	BTNR	MACR	Mean	s.d.
No. of transitions	15	21	5	10	7	11.6	6.5
No. of transversions	22	28	5	14	10	15.8	9.2
No. of substitutions	37	49	10	24	17	27.4	15.7
No. of indels	77	82	19	68	68	62.8	25.2
No. of ts. sites	15	21	5	10	7	11.6	6.1
No. of tv. sites	22	28	5	14	10	15.8	9.2
No. of subst. sites	37	49	10	24	17	27.4	14.3
No. private subst. sites	6	9	2	2	1	4.0	3.4
No. of indel sites	77	82	19	68	68	62.8	25.2
No. of polymorphic sites	102	109	29	89	84	82.6	28.2

4.3.3 Nucleotide diversity

One measure of genetic diversity is nucleotide diversity, a per-site or per-base comparison of the sequences being compared. In calculating nucleotide diversity, the entire length of the sequences can be compared or calculations can be limited to only the polymorphic sites within these sequences. Nucleotide diversity of the polymorphic sites of samples from each sampling area were calculated and the results are presented in Table 4.5.

The calculated nucleotide diversity values for *ITS*2 sequences for all the sampling areas were uniformly high. While the population at MacRitchie Reservoir had the lowest nucleotide diversity at 0.2102 and the population at Kota Tinggi Waterfalls Resort had the highest at 0.3863, standard deviation values for these measures suggest that these nucleotide diversity values are hardly different from each other. Nucleotide diversity values are lower than any of the values of Nei's gene diversity (an analog of nucleotide diversity) computed from microsatellite data. The low levels of nucleotide diversity are explained by the fewer number of possible alleles (limited to the bases, A, C, T, G) in any of the sites of the *ITS*2 sequences. Nucleotide diversity is also affected by the number of samples available for analysis and direct comparisons between sampling areas with different numbers of samples must be made with caution.

4.3.4 Haplotype diversity

More interesting results are observed when whole sequences, *ITS*2 haplotypes, are compared and analyzed (as against a per-base comparison of the sequences). Of the 193 samples of *A. papillatum* analyzed, 65 different *ITS*2 haplotypes were found. The diversity of these haplotypes is shown in Table 4.6.

Of the 65 haplotypes, 34 were unique to one individual sample and therefore found in only one sampling area. Thirty-one haplotypes were found in more than

Table 4.5: Nucleotide diversity in *ITS*2 sequences of *Acanthorrhynchium papillatum*. s.d. = standard deviation.

Sampling area	Nucleotide diversity	s.d.
SBRF	0.2392	0.1513
GBFR	0.3229	0.1710
KTP	0.3863	0.1296
BTNR	0.3002	0.1820
MACR	0.2102	0.1356
Total	0.2918	0.1539

Table 4.6: Haplotype diversity of *ITS*2 in *Acanthorrhynchium papillatum*. Samples in the same row have the same *ITS*2 sequences and are different from those in other rows. "Reference samples" are artificial designations for representative samples of a haplotype and are labeled as such only for convenience.

Reference sample	Matching samples
APBK1	APBK2, APBK3, APBK17
APBK4	APBK7, APBK11, APBK35, APBK36, APBK41, APBK42, APBK43, APBK44
APBK8	APBK12
APBK9	APBK18
APBK10	APBK25, APBK39, APBK40, APBK45, APBL4, APBL18
APBK20	
APBK26	APBK53, APBL9, APBL10, APBL35, APBL50, APBT17, APBT25, APBT49, APMR60, APMR61, APMR63, APMR64, APMR65, APMR66
APBK31	
APBK32	
APBK33	
APBK37	APBL62
APBK38	
APBK47	
APBK50	
APBK52	
APBL1	APBL45
APBL2	
APBL3	APBL8
APBL6	APBL14, APBL33
APBL11	APBL56, APBL59
APBL12	
APBL15	APBL37
APBL17	APBL21, APMR3, APMR4, APMR5, APMR6, APMR7, APMR24
APBL20	
APBL22	APKT25, APKT28, APKT29, APBT1, APBT4, APBT50, APMR18
APBL26	
APBL28	
APBL30	
APBL31	APBL32, APBL61, APBL72
APBL34	
APBL39	

continued on next page

continued from previous page

Reference sample	Matching samples
APBL41	
APBL42	
APBL43	
APBL44	
APBL46	
APBL47	
APBL51	APBL53, APBL57, APBL58
APBL52	APBL64
APBL65	
APBL66	
APBL68	
APBL69	
APBL71	APBT6, APBT7, APBT8, APBT9, APBT10, APBT11, APBT23, APBT36, APBT37, APBT53, APMR9, APMR62
APKT1	APKT6, APKT7, APKT8, APKT9, APKT11, APKT12, APKT13, APKT14, APKT15, APKT16, APKT17
APKT2	
APKT3	
APKT4	APKT22, APKT23, APBT18, APMR19
APKT20	APKT24
APKT32	
APBT15	APBT16
APBT19	APBT20, APBT21, APBT22
APBT26	APMR27
APBT29	APBT30
APBT32	
APBT35	
APBT38	
APBT44	APBT45, APBT46
APBT51	
APMR1	APMR2, APMR20, APMR21, APMR22, APMR23, APMR28, APMR29, APMR31, APMR58, APMR59
APMR10	APMR11, APMR12, APMR13
APMR14	APMR30, APMR32, APMR34, APMR35, APMR36, APMR37, APMR38, APMR39, APMR40, APMR41, APMR42, APMR43, APMR44, APMR45, APMR46
APMR16	
APMR25	APMR26
APMR47	APMR48

one sample. Of these, 23 haplotypes were found within only one sampling area (private haplotypes) and 8 were found in more than one sampling area. The 57 private (out of 65) haplotypes suggest the presence of haplotypic differentiation among the different sampling areas, a hypothesis that will be more formally tested later. However, the observation that 8 haplotypes were found in more than one sampling area (up to 4 out of 5 sampling areas in the case of the haplotype represented by APBK26) suggests the potential for dispersal of this moss to at least as far as the sampling areas are distant from each other (approximately 100 km). The distribution of unique and replicated haplotypes is summarized in Table 4.7.

It is worth noting that samples deemed to be clonal copies of each other by microsatellite analysis also had matching *ITS*2 haplotypes (or had *ITS*2 loci that could not be successfully sequenced). No conflicts between matching multi-locus microsatellite genotypes and matching *ITS*2 haplotypes were therefore observed. However, more matching *ITS*2 haplotypes are seen than matching multi-locus microsatellite genotypes. This observation corroborates the idea that microsatellite markers have a greater ability to distinguish among samples.

Gene diversities at the haplotype level (analogs of nucleotide diversity) were also calculated. Gene diversity at the haplotype level measures the number and evenness of distribution of haplotypes. Gene diversities for each sampling area are shown in Table 4.8.

Calculated values for gene diversity at the haplotype level were high for all the sampling areas examined. This was probably due to the substantial numbers of private haplotypes found in each area. Diversity was lowest (but still high) in Kota Tinggi Waterfalls Resort at 0.7115 ± 0.0195 s.d. and highest in Gunung Belumut Forest Reserve at 0.9675 ± 0.0086 s.d. Again, caution must be taken when interpreting these numbers since gene diversity measures are affected by the

Table 4.7: *ITS*2 haplotype distribution and gene diversities at the haplotype level among samples of *Acanthorrhynchium papillatum*. U – unique haplotypes, RP – replicated haplotypes found only within the same sampling area, RG – replicated haplotypes found in multiple sampling areas.

Sampling area	Total	U	RP	RG
SBRF	15	8	4	3
GBFR	31	18	8	6
KTP	7	3	2	2
BTNR	13	4	5	4
MACR	11	1	5	5

Table 4.8: *ITS*2 gene diversities, h, at the haplotype level for each sampling area. s.d. = standard deviation.

Sampling area	$h \pm$ s.d.
SBRF	0.8958 ± 0.0360
GBFR	0.9675 ± 0.0086
KTP	0.7115 ± 0.0195
BTNR	0.8826 ± 0.0401
MACR	0.8447 ± 0.0299

number of samples used in the analysis.

4.3.5 Genetic distance

The neighbor-joining tree constructed from the genetic distance matrix of all observed *ITS*2 haplotypes is shown in Figure 4.1.

The many terminal branches found in the tree in Figure 4.1 illustrate that the levels of *ITS*2 variation in *Acanthorrhynchium papillatum* were high (many haplotypes). The considerable lengths of these branches illustrate that the degrees of variation were also high (the differences between haplotypes were not due to only 1 or 2 mutations). The tree topology also does not exhibit distinct clustering of neighboring samples, or even samples from the same sampling area. Instead, haplotypes from different sampling areas were commonly found clustering together in the tree. These interesting results are discussed in the next section.

4.3.6 Population differentiation and genetic structure

The exact test of population differentiation performed at the global level suggested differentiation of haplotypes among the populations (probability of non-differentiation, $P = 0$). Exact tests of pairwise comparisons of populations yielded the same results (all pairwise tests, probability of non-differentiation $P = 0$). These results are probably explained by the overwhelming number (57/65) of private haplotypes found in the data set.

The results of the analysis of molecular variance, AMOVA, at the haplotypic level are presented in Table 4.9. The results reveal that about 81% of the variation is found within populations, about 14% is found among populations within groups and that about 5% of the variation in the data set is found among groups of populations.

Fixation indices were also computed from the results of AMOVA. The compu-

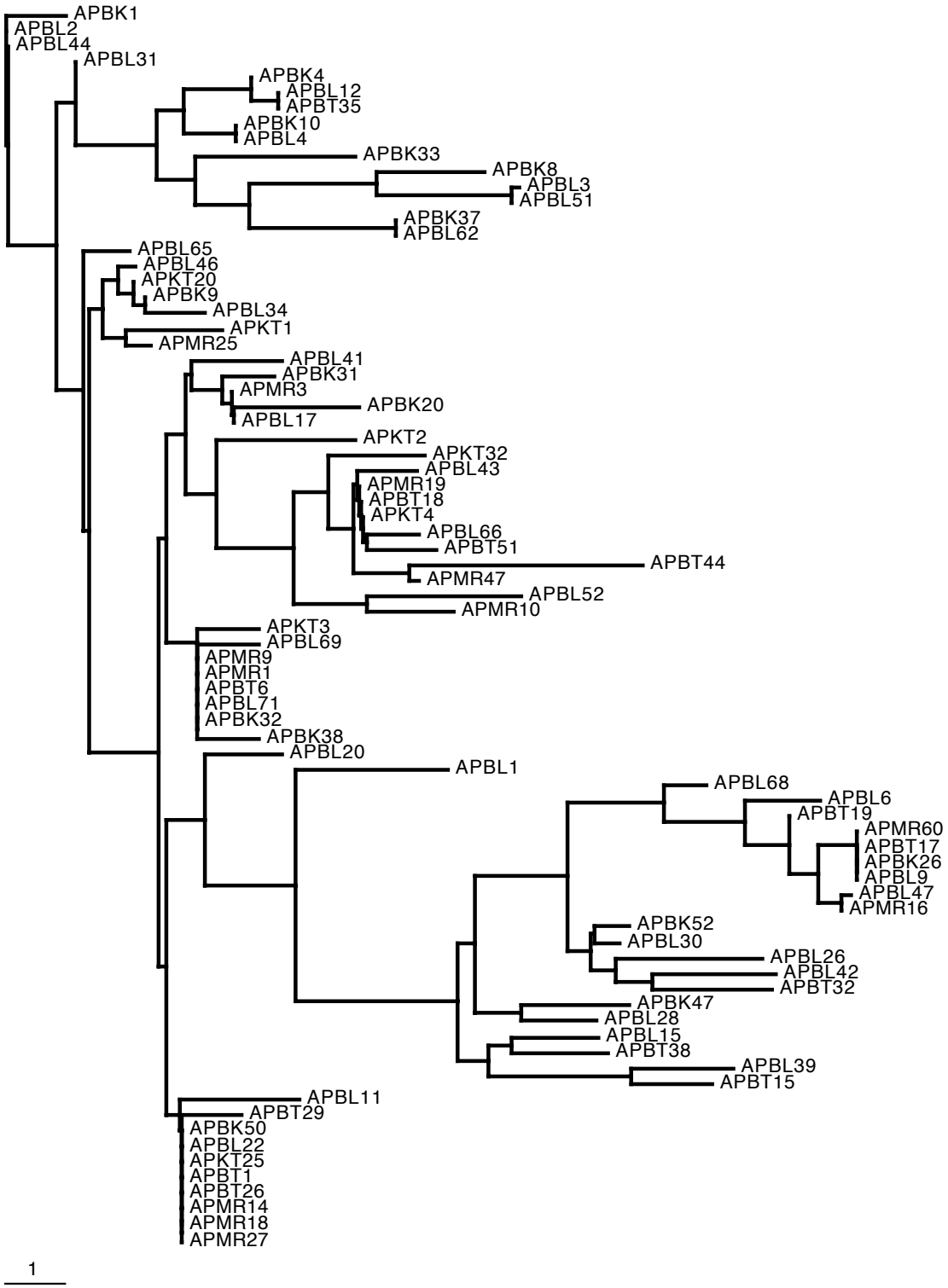


Figure 4.1: Neighbor-joining tree constructed from genetic distances of *ITS*2 haplotypes. Scale bar indicates branch lengths. This tree is unrooted.

Table 4.9: Results of analysis of molecular variance, AMOVA, from *ITS*2 sequence data averaged over 131 loci. d.f.=degrees of freedom.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	167.025	0.69461	4.66
Among populations within groups	3	261.146	2.07644	13.94
Within populations	188	2279.990	12.12760	81.40
Total	192	2708.161	14.89866	

tations suggest that there is moderate genetic differentiation among populations within groups, $F_{SC} = 0.14619$, and that this statistic is significantly different from zero ($P = 0$). There is also great genetic differentiation among individuals among populations and among groups, $F_{ST} = 0.18599$, and that this statistic is also significantly different from zero ($P = 0$). Finally, there is little genetic differentiation between the two groups of populations, $F_{CT} = 0.04662$ but that this measurement may not be significantly different from zero ($P \approx 0.19$). The data from AMOVA therefore suggest that the moss samples are greatly differentiated from each other and that populations in the different sampling areas are moderately genetically isolated from each other.

Population pairwise F_{ST} values computed from $ITS2$ sequences are shown in Table 4.10. All F_{ST} values are significant ($P = 1.0$).

The results of the pairwise F_{ST} comparison echo the results of the AMOVA. They generally suggest that there is moderate to great genetic differentiation among the populations.

4.4 Discussion

The results of the experiments in this chapter help complete the second and third objectives of the thesis. They provide baseline information on the genetic diversity among clumps of *Acanthorrhynchium papillatum* as measured by $ITS2$ sequences. High levels of $ITS2$ sequence variation, at both nucleotide and haplotype levels, were found in this moss. The results also suggested lower levels of genetic diversity among clumps of *A. papillatum* in sampling areas with greater degrees of disturbance. The results, while not unexpected, provide valuable empirical data on the genetic diversity of this species. Moreover, the results generally echo the findings of experiments using microsatellite markers. These independent sources of evidence (microsatellite markers and $ITS2$ sequences) strengthen the claim that

Table 4.10: Pairwise F_{ST}s between populations using $ITS2$ sequences.

	SBRF	GBFR	KTP	BTNR	MACR
SBRF	0.00000				
GBFR	0.08903	0.00000			
KTP	0.22767	0.19939	0.00000		
BTNR	0.23916	0.03490	0.30366	0.00000	
MACR	0.28162	0.13324	0.30399	0.12325	0.00000

although high levels of genetic diversity are found in *A. papillatum*, lower levels are found in areas that are more disturbed.

There is more to the data on *ITS*2 sequences of *Acanthorrhynchium papillatum* than just diversity, however. The data reveal other aspects of the biology of *A. papillatum*.

4.4.1 High rates of mutation are seen in the ITS*2 region of* Acanthorrhynchium papillatum

One of the findings of the study using microsatellite markers is that the mutation rates of microsatellites in *Acanthorrhynchium papillatum* were found to be high. What do the data on the *ITS*2 sequences reveal about the mutation rate of this locus? Are high mutation rates characteristic solely of microsatellite loci of the species? Or are high mutation rates also seen in other loci? The data suggest that high rates of mutation are also found in *ITS*2 in *A. papillatum*. This hypothesis is supported by a number of observations listed below.

The discovery of a large number of private *ITS*2 haplotypes, haplotypes that are found in only one sampling area, suggest that there are high mutation rates of this locus in *Acanthorrhynchium papillatum*. Had the large numbers of haplotypes been caused by other sources of variation, like gene flow from adjacent populations, then the haplotypes would not be private. One could argue that perhaps adjacent populations were missed in the sampling, and that if more extensive sampling had been done, then haplotypes that were thought to be private may have been found in other sampling areas. This, in effect, would negate their "private" status (similar arguments to these are found in the discussion of diversity of *A. papillatum* using microsatellite markers). While this is a valid argument, in the light of the other evidence in the data, the alternative of high mutation rates is favored.

The different copies of *ITS*2 found in many samples of *Acanthorrhynchium*

papillatum, as evidenced by these samples having *ITS*2 sequences with ambiguous basecalls, suggest that high rates of mutation are seen in this locus. This is because the many copies of rDNA and its associated spacers (including *ITS*2) within individuals of a species should have homogenous sequences caused by the action of concerted evolution (Dover, 1986; Ruggiero & Procaccini, 2004; Smith, 1976). The presence of *A. papillatum* samples with different *ITS*2 sequences (intragenomic or intra-individual variability) suggests that the mutation rates in *ITS*2 loci of *A. papillatum* are faster than the homogenizing effects of concerted evolution.

While it can be argued that intragenomic variability of *ITS*2 may be due to some source of variation apart from mutation, other findings of the thesis project corroborate the hypothesis that the variability is caused by high mutation rates inherent in *Acanthorrhynchium papillatum*. If intragenomic variability of *ITS*2 was influenced by external processes such as sexual reproduction, then different levels of intragenomic variability should be seen in the different sampling areas. This is because the moss populations in the different sampling areas have been shown to rely on sexual reproduction in different degrees. However, the actual findings show that the proportions of samples with intragenomic variability are the same for all sampling areas (around 30%). This suggests that intragenomic variability is not influenced by the degree of sexual reproduction found in a population. The findings therefore strengthen the assertion that intragenomic variability is due to mutation, as against other sources of variation.

The data show that some *ITS*2 haplotypes are shared among the different sampling areas. This observation suggests that the populations from these sampling areas, at one time or another, shared genes. Why then are there so many private *ITS*2 haplotypes in sampling areas that can share genetic material? The simplest answer is that high mutation rates in *ITS*2 create different haplotypes from those that were once shared.

4.4.2 ITS*2 tree topology*

The most noticeable finding in the topology of the tree constructed from genetic distances of *ITS*2 haplotypes is that there is, at best, only weak clustering of moss samples that are spatially adjacent in the field. This suggests that the dispersal of the moss is relatively even within these sampling areas. Similar results were seen in the trees constructed from allele-sharing matrices of the microsatellite markers. Consequently, a similar hypothesis is proposed for this data set, that is, that high rates of mutation in the *ITS*2 locus of *Acanthorrhynchium papillatum* obscure patterns of relationships of samples and contribute to the tree topology obtained.

4.4.3 Limitations

As a measure of genetic diversity of *Acanthorrhynchium papillatum*, the *ITS*2 locus has been shown to be a suitable marker to indicate the extent of variation within and among different populations of the moss. However, if more information is needed from the marker, then some limitations in *ITS*2 can pose problems.

The multiple copies of *ITS*2 can prevent the marker from being successfully sequenced without resorting to cloning techniques. The presence of multiple copies also impede the reconstruction of relationships of individuals of *Acanthorrhynchium papillatum* (which copy of *ITS*2, if several different copies exist in an individual, should be used when constructing a tree?).

The data also show that the levels of diversity in *ITS*2 are lower than those using microsatellite markers. *ITS*2, in particular, cannot be used to identify individuals of *Acanthorrhynchium papillatum* the way microsatellite markers can. The results therefore underscore the utility and importance of microsatellite markers in population genetics studies. Nevertheless, as has been shown in this chapter, *ITS*2 still provides a valuable molecular marker in measuring the extent of genetic diversity in *A. papillatum* and in showing how genetic diversity is different

between and among sampling areas of different habitat quality.

4.5 Conclusions

The second internal transcribed spacer, *ITS*2, was used to describe genetic variation in *Acanthorrhynchium papillatum*. High levels of variation were seen at both nucleotide and haplotype levels using this marker. These results join those obtained using microsatellite markers in again helping to disprove the early notions that the genetic diversity of mosses are low. The results also suggested that levels of diversity, particularly at the haplotype level, were reduced in populations from areas of lower habitat quality.

*ITS*2 diversity is suggested to arise from high rates of mutation in this locus. This hypothesis is supported by several observations and is seen in characterisitics of the data.

Lower levels of diversity were seen in *ITS*2 loci compared to microsatellite markers. These results underscore the utility and power of microsatellite markers. However, *ITS*2 has been shown to be useful marker in describing the genetic variability of *Acanthorrhynchium papillatum*.

Chapter 5

PRELIMINARY STUDY OF GENETIC DIVERSITY WITHIN CLUMPS OF *ACANTHORRHYNCHIUM PAPILLATUM* AS MEASURED BY MICROSATELLITE MARKERS

5.1 Introduction

In the earliest experiments to develop microsatellite markers for *Acanthorrhynchium papillatum*, candidate markers were occasionally tested on samples belonging to the same clump of moss. The trials were not exhaustive since the understandable priority at that time was to finish developing the markers. Moreover, it was felt that to better gauge variability and thus, utility of the markers, they needed to be tested on samples collected from spatially separated clumps, not on samples from the same clump. However, the few early experiments did show that samples coming from the same clump had the same multi-locus genotype and therefore could be clonal in nature. This was a consistent observation (of the few trials done) and seemed to be irrespective of the size of the clump and its source.

In the process of using the markers to determine the genetic variability among clumps of *Acanthorrhynchium papillatum*, a better understanding of the properties and utility of the markers was also gained. This includes understanding how the markers have different expected heterozygosities and powers of discrimination (or match probabilities) in each of the different sampling localities. Equipped with these statistics, it was now possible to return to the early experiments and try to conduct more meaningful experiments to determine genetic variability within

clumps of *A. papillatum*.

Unfortunately, several limitations prevented the exhaustive study of within-clump variation of *Acanthorrhynchium papillatum*. These limitations stemmed both from inherent characteristics of the species being studied and from the finite time and material resources available in the project.

Acanthorrhynchium papillatum forms amorphous clumps of widely-varying size and numerical composition of moss individuals. Some of these clumps are composed of only a few strands of the moss. Other clumps seen in the field covered much of the surface of large fallen logs, several square meters in area. Clumps of even 200 cm^2 in area can be composed of over a hundred individuals.

It would therefore be hard to compare the genetic variability within clumps when the clumps are highly variable in size and numerical composition of individual plants. To obtain adequate samples representative of even moderately-sized clumps would be logistically forbidding. Multiplying these numbers by the number of replicates needed to create a suitable profile of the within-clump diversity of an area only exacerbates this problem (and this is before even trying to compare statistics among different sampling localities).

Exact counts are made even harder to determine since *Acanthorrhynchium papillatum* is a creeping moss: teasing apart the tightly-matted clumps cause breakage of strands no matter how carefully the attempt to unravel them is made. This consequently muddles the accurate counting of the number of individuals and their genotypes.

The creeping nature of *Acanthorrhynchium papillatum* also presents difficulties in studying the orientation of individuals of different genotypes, or genets, within clumps. The strands grow in every direction, overlap and form tight mats. The initial spatial orientation of the genets is obscured as the strands grow and cannot be compensated for by sampling. (If sampling in a grid is used on an intact clump

it is possible that a sample taken from one grid-square actually originated from another many squares away).

Although an exhaustive study of the within-clump diversity of *Acanthorrhynchium papillatum* could not be made within the timeline and resources of this project, a preliminary investigation was performed that was nonetheless more thorough than the earliest experiments. The levels of sampling that could be done were not adequate to perform meaningful statistical comparisons at any level. Nevertheless, the results here presented help complete the picture on the genetic diversity of *A. papillatum* and are hoped to serve as seeds in future investigations on the within-clump diversity of this and other species.

5.2 Materials and Methods

Many of the materials and methods used to study the within-clump diversity of *Acanthorrhynchium papillatum* are the same as those described in earlier chapters. Sampling areas, field selection/collection, and post-collection processing pertinent to studies of within-clump diversity are described in Sections 3.2.1, 3.2.3, and 3.2.4, respectively. The molecular lab processes of DNA extraction, fragment analysis runs, and the analysis steps of size-calling and allele-scoring have also all been detailed in an earlier chapter.

Details on the methodology unique to this part of the thesis project follow:

5.2.1 Collection dates

Table 5.1 shows the collection dates of *Acanthorrhynchium papillatum* for samples used to determine variation within clumps of this moss.

Table 5.1: Collection dates of *Acanthorrhynchium papillatum* for studies on variation within clumps of the moss.

Sampling area	Collection date
Sungei Bantang Recreational Forest	19 March 2006
Gunung Belumut Forest Reserve	18 March 2006
Kota Tinggi Waterfalls Resort	2 July 2005
Bukit Timah Nature Reserve	25 May 2006
MacRitchie Reservoir	25 May 2006

5.2.2 Sampling within clumps to determine variation

Sampling within clumps was done in the laboratory on clumps brought in from the field.

Field collections from Sungei Bantang Recreational Forest, Gunung Belumut Forest Reserve, Bukit Timah Nature Reserve and MacRitchie Reservoir were done nearly two years prior to the time this part of the project was started. New collections from these areas needed to be made to ensure that the integrity of the DNA in the samples was sufficient for microsatellite analysis. About 15 clumps approximately 100 cm^2 in area from each of these areas were collected in a manner similar to that described in Section 3.2.3. Since the original collections from Kota Tinggi Waterfalls Resort were relatively fresher, it was decided that the samples used for studying variation among clumps could also be used for studying within-clump variation. This was done with some risk that the integrity of the DNA would be insufficient for microsatellite analysis.

To test for the extent of variation within clumps of *Acanthorrhynchium papillatum*, 8 clumps were randomly chosen from those available from each of the 5 sampling localities. Four shoots from each of these clumps (160 strands and 40 clumps total) were randomly selected to form a set of samples. This was done by either teasing apart all the strands from a clump and randomly selecting 4 strands, or by laying a grid on the intact clump, randomly selecting 4 grid-squares, and taking the strands directly under the squares. (The risk of breaking strands is present in either method). To aid in later discussion, this sampling scheme and the generated data will be designated "main".

To test if more genotypes could be revealed from increased sampling within a single clump, an additional clump was randomly selected from each of the sampling areas for testing. From this clump 20 strands (5 clumps and 100 strands total) were randomly selected for testing with the microsatellite markers using the techniques

just detailed. Again, to aid in later discussion, this sampling scheme and its generated data are designated "supplementary".

5.2.3 Sample naming

Samples were named in similar fashion to those used in the study of variation among clumps of *Acanthorrhynchium papillatum*. The new clumps from Sungei Bantang Recreational Forest, Gunung Belumut Forest Reserve, Bukit Timah Nature Reserve and MacRitchie Reservoir were sequentially numbered starting from 100 to easily differentiate them from those used for studying variation among clumps, e.g., APBK101, APBL113, etc. Samples coming from the same clump were suffixed with letters in sequence, e.g., APBK111A, APBK111B, etc. The 'A' samples of samples from Kota Tinggi Waterfalls Resort clumps are also the original samples used for the analysis of among-clump variation.

5.2.4 Analysis

The numbers of multi-locus genotypes per clump were determined manually by examination of the results.

Where applicable, pairwise genetic distances of samples within a clump were calculated in order to estimate the similarity of the said samples. Genetic distances were calculated in two ways. The first way was by counting the number of different alleles between the two samples using the formula

$$\hat{d}_{xy} = \sum_{i=1}^{L} \delta_{xy}(i),$$

where $\hat{d}_{xy}$ is the genetic distance between samples x and y, L is the number of loci and $\delta_{xy}(i)$ is the Kronecker function, equal to 1 if the alleles of samples x and y are identical for the ith locus and equal to 0 otherwise (Excoffier *et al.*, 2006).

The second way was by computing the sum of the squared number of repeat differences between the two samples (Slatkin, 1995) using the formula

$$\hat{d}_{xy} = \sum_{i=1}^{L} (a_{xi} - a_{yi})^2,$$

where $\hat{d}_{xy}$ is the genetic distance between samples x and y, L is the number of loci and a_{xi} and a_{yi} are the number of repeats of the microsatellite for the ith locus for samples x and y respectively. It is obvious from the formula and the description that this measure of genetic distance assumes that the mutation of microsatellite loci in *Acanthorrhynchium papillatum* proceeds in a stepwise manner. This assumption, as already shown in previous chapters, is tenuous at best. It is used here, however, just as an additional measure to try to further characterize the data.

Genetic distances using either formula were computed with the aid of the spreadsheet program EXCEL® (Microsoft, USA).

5.3 Results

Alleles of the microsatellite markers for the different samples used to study within-clump diversity of *Acanthorrhynchium papillatum* are presented in Appendices L and M. Observations concerning the data are presented below.

5.3.1 Null alleles

An inordinate number of null alleles were found in both main and supplementary samples. Most of these are from the marker APMS28 although many samples also generated null alleles for the markers APMS14 and APMS4. (The rest of the markers were null in only one or two samples). The presence of these null alleles creates uncertainties in the computation of the within-clump genetic di-

versity of *Acanthorrhynchium papillatum*, particularly in determining the extent of vegetative reproduction within clumps of this moss. For reasons that will be discussed later, null alleles were excluded from these computations. That is, multi-locus genotypes and pairwise distances were determined by only considering the contributions of the non-null alleles.

5.3.2 Multi-locus genotypes

Table 5.2 summarizes the findings on the multi-locus genotypes of the main samples.

The results clearly show that the majority of the clumps are composed of samples of single multi-locus genotypes. This was observed from at least half of the clumps from each of the sampling localities to as many as 7 of the 8 clumps in the localities in Singapore. Clumps with 2 or 3 multi-locus genotypes were also found but were always fewer than those with a single multi-locus genotype for any of the sampling localities. Only one clump, from Sungei Bantang Recreational Forest, with 4 different multi-locus genotypes was found. Examination of the raw data shows that samples from Kota Tinggi Waterfalls Resort that were different from others within the same clump were only different at one locus. This is in contrast to what was found in the other localities. In these localities, samples having dissimilar multi-locus genotypes were different at two or more loci.

Data from the supplementary samples show that the single clumps picked from Gunung Belumut Forest Reserve, Kota Tinggi Waterfalls Resort, Bukit Timah Nature Reserve and MacRitchie Reservoir are all composed of single multi-locus genotypes. That is, all 20 samples taken from the single clump from each of these localities exhibit the same alleles for all eight markers tested on them. In contrast, 2 multi-locus genotypes were found in the single clump picked from Sungei Bantang Recreational Forest.

Table 5.2: Multi-locus genotypes of 4 samples taken from each of 8 clumps of *Acanthorrhynchium papillatum*. Header row indicates number of multi-locus genotypes. Entries indicate number of clumps (out of 8 tested) that have these multi-locus genotypes.

	Number of multi-locus genotypes			
	4	3	2	1
Sungei Bantang Recreational Forest	1	0	2	5
Gunung Belumut Forest Reserve	0	0	3	5
Kota Tinggi Waterfalls Resort	0	2	2	4
Bukit Timah Nature Reserve	0	0	1	7
MacRitchie Reservoir	0	0	1	7

5.3.3 Pairwise genetic distances

Using the Kronecker function

Pairwise genetic distances of the main samples employing the Kronecker function as described by Excoffier *et al.* (2006) are shown in Appendix N. Examination of the data reveals that this method is helpful in rapidly showing which samples within a clump share a common multi-locus genotype. This method can also be used to show the number of loci at which two samples are different.

The maximum pairwise genetic distance using this method corresponds to the number of loci at which two samples can be compared. If data are present for both samples at all 8 markers, then the maximum difference between these samples is 8. If a sample is null at one marker, then only 7 markers can be used to compare the two samples and the maximum difference is 7. By summing the pairwise differences of samples within a clump the degree of variability with respect to the number of loci at which the samples are different can be quantified. The effects of the exclusion of null alleles can also be estimated by computing the sum of the maximum possible pairwise differences (e.g., if data from all loci are present, then this sum is equal to 8×6=48). A summary of these sums are presented in Table 5.3.

Pairwise genetic distances restate the observations already made from the number of multi-locus genotypes but also add more information. Each of the four samples from APBK101 are different from the rest at 3–5 loci. Moreover, the sum total of the pairwise differences for this clump are more than half the maximum possible value. Two samples from APBK111 share the same multi-locus genotype and are different at two loci from the two other samples (which in turn also share the same multi-locus genotype). Almost all the other clumps that show more than one multi-locus genotype are represented by a single sample that differs from the rest, albeit at two or more loci. Again, the data from pairwise genetic distances show

Table 5.3: Summary of pairwise genetic distances of samples within clumps of *Acanthorrhynchium papillatum*. "Total" is the sum of pairwise genetic distances [computed as described by Excoffier *et al.* (2006)] of samples within a clump. "Max" is the largest possible total given the number of usable markers. Computations are further explained in the text.

Clump name	Total	Max
APBK101	24	45
APBK102	0	30
APBK104	0	30
APBK107	0	48
APBK108	0	36
APBK111	8	48
APBK113	15	48
APBK115	0	48
APBL102	0	48
APBL103	15	48
APBL105	0	48
APBL106	0	42
APBL107	0	48
APBL108	0	48
APBL109	12	36
APBL110	12	42
APKT05	0	48
APKT10	3	48
APKT12	3	48
APKT15	0	48
APKT17	0	48
APKT18	0	42
APKT19	3	36
APKT28	5	48
APBT102	21	48
APBT103	0	48
APBT104	0	48
APBT105	0	48
APBT107	0	48
APBT108	0	36
APBT109	0	48
APBT110	0	48
APMR103	0	48
APMR105	0	48
APMR106	0	48
APMR108	0	36
APMR109	0	48
APMR111	0	48
APMR113	0	42
APMR115	15	48

how the multi-locus genotypes of samples from Kota Tinggi Waterfalls Resort, if different from the rest of the samples from the same clump, are only different at a single locus. This translates to low sums of pairwise genetic distances, much lower than the maximum possible totals for the clumps in consideration.

Because the supplementary data show that only the clump from Sungei Bantang Recreational Forest shows more than one multi-locus genotype, pairwise genetic distances from only this clump are shown in Appendix O. In this clump, APBK112, 16 samples were found possessing one multi-locus genotype while 4 shared another. These two genotypes were different at 2 loci, APMS28 and APBMS14. The total sum of pairwise genetic differences for this clump is 102 while the maximum possible is 1520.

Using sums of squared-differences

The method of Slatkin (1995) for computing pairwise genetic differences between samples is also helpful in showing which samples within a clump share a common multi-locus genotype. However, the utility of the method comes to fore in suggesting the degree of difference between alleles of the same loci of different multi-locus genotypes. This is done by recalling that the stepwise mutation model of microsatellite evolution suggests that larger differences in the number of repeats of alleles of the same loci imply greater genetic distance between these alleles. By summing these squared differences for all usable loci, the total becomes a measure of the overall genetic distance between two samples when the stepwise mutation model of microsatellite evolution is assumed.

Earlier data for the microsatellite markers used suggest that while most of the markers are based on dinucleotide repeats (and therefore should have repeat units that are 2 bases apart), many samples exhibited alleles that did not fall neatly in multiples of 2 bases away from other alleles. As has been discussed before,

this suggests that the mutation of these microsatellite markers may not be strictly stepwise. In this regard, all the markers were assumed to have repeat units 1 base apart. That is, the difference in the number of repeat units between two samples for any marker corresponds to the base-pair difference in the fragment lengths of the samples for that marker.

Pairwise genetic differences using the method of Slatkin (1995) are found in Appendix P. Large values were obtained for most of the pairs of samples that were different. Although this can once again restate that the samples were different at multiple loci, these values are better explained by the large base-pair differences in fragment sizes at these several loci. While most of the base-pair differences were found to be large, smaller values are seen for one clump, APKT10. In this clump, only one sample differs from the rest and at only one locus and for only a few repeat units.

Total sums of these pairwise genetic differences are presented in Table 5.4. Since it is difficult to establish upper bounds of the fragment lengths for any of the markers, maximum possible sums cannot be computed nor shown.

Concerning the supplementary sampling, data are again shown from only one clump, APBK112. Pairwise genetic distances using the method of Slatkin (1995) are presented in Appendix Q. The low values restate the observation that the two multi-locus genotypes are different at only two loci and that at either locus, the alleles are different by only 2 or 3 repeat units from the rest.

5.4 Discussion

The results of the experiments in this chapter are straightforward: *Acanthorrhynchium papillatum* clumps are mostly composed of a single-multilocus genotype regardless of the source of the clump. Vegetative reproduction seems to be more important than sexual reproduction within clumps of *A. papillatum*. These de-

Table 5.4: Summary of pairwise genetic distances of samples within clumps of *Acanthorrhynchium papillatum* computed using the method of Slatkin (1995). "Total" is the sum of pairwise genetic distances of samples within a clump. Computations are further explained in the text.

Clump name	Total
APBK101	3710
APBK102	0
APBK104	0
APBK107	0
APBK108	0
APBK111	4916
APBK113	4593
APBK115	0
APBL102	0
APBL103	4065
APBL105	0
APBL106	0
APBL107	0
APBL108	0
APBL109	3012
APBL110	1779
APKT05	0
APKT10	48
APKT12	4563
APKT15	0
APKT17	0
APKT18	0
APKT19	386
APKT28	5763
APBT102	816
APBT103	0
APBT104	0
APBT105	0
APBT107	0
APBT108	0
APBT109	0
APBT110	0
APMR103	0
APMR105	0
APMR106	0
APMR108	0
APMR109	0
APMR111	0
APMR113	0
APMR115	4665

ductions, however are weakened by the presence of null alleles that were found in the data. Although null alleles introduced some complexity to the analysis, they nevertheless strengthened ideas on the biology of the moss that were first seen in earlier chapters. Another interesting finding is that a small amount of genetic variation was found within clumps of the moss. Null alleles and the genetic variation found within clumps of *A. papillatum* are discussed below.

5.4.1 Null alleles

The large number of null alleles found in the data presented problems in determining the within-clump diversity of *Acanthorrhynchium papillatum*, that is, clonality or the extent of vegetative reproduction was difficult to ascertain. Pairwise genetic distances were also hard to compare between samples that had a full complement of alleles and those that were null in one or more alleles.

Although multi-locus genotypes can be determined from mere examination of the data, the determination of the exact number of multi-locus genotypes found in a group of samples requires that alleles be present for all markers being considered. This is because two samples can share the same alleles for 7 out of 8 markers, differ in only one marker, and thus be of different multi-locus genotypes. If one or both samples are null at one marker, it would be impossible to say for certain if the two samples share the same multi-locus genotype or if they do not, precisely because there is no information at the locus of the null allele.

However, knowing the ability of *Acanthorrhynchium papillatum* to reproduce asexually, it would be reasonable to assume that two samples from the same clump that share the same alleles for most of the markers but are null at one are products of vegetative reproduction. Indeed, this assumption is further bolstered by the low combined matching probability of the markers: unrelated samples have very low probabilities of having the same alleles at multiple markers (see again Table 3.17)

and thus the samples are most probably clones of each other.

Being products of vegetative reproduction would imply, however, that the samples should share the same alleles for all markers, including those that have been empirically determined to be null. Null alleles do not necessarily conflict with this assumption. After all, null alleles could be products of the same mutation in the binding site of one of the primers and are therefore still the same allele that the samples share. These would be reasonable assumptions that allow for the simplest explanations. And yet the empirical data show that it is possible for two samples to share the same alleles for 7 loci and yet be different at another. This is seen in the clump APKT10. In this clump, the "A" sample has an allele of 126 at locus APBMS3 while all other samples have an allele of 130. This is also seen in the clump APKT12. In this clump, the "B" sample has an allele of 93 at locus APBMS3 while all the others have an allele of 132. The simplest explanation, and one that has already been implicated in earlier chapters of this thesis, is that these differences in alleles may be explained by somatic mutation.

Pairwise genetic distances are also influenced by the presence of null alleles. Since these distances are computed from sums of locus-to-locus comparisons of two samples, the presence of null alleles prevent the full comparison of two samples and potentially reduce the total sum (by preventing two loci from being compared) and computed genetic distance between the two. Obviously therefore, genetic distance between two samples with null alleles can be higher than actually computed. Hence, direct comparisons of genetic distance between sets of pairs of samples would be inappropriate when null alleles are present. Assertions that there is greater genetic distance between one pair than another cannot be made when one or both pairs have null alleles because the pair with null alleles could be more distant than the other. This is true regardless of the method used to compute genetic distance [Excoffier *et al.* (2006) or Slatkin (1995)].

How then do we compensate for the presence of null alleles?

Interestingly, as far as multi-locus genotypes are concerned, null alleles behave somewhat like hypothetical markers that have not been used on the data: samples may share similar or different alleles at these markers but they do not affect data that are already present. Loci with null alleles can therefore be disregarded and the number of multi-locus genotypes in a set of samples determined from data that are available. No assumptions need to be made about the alleles. Excluding loci with null alleles can be done without bias when looking only at samples within the same set. The presence of null alleles, however, has to be taken into account when comparing among sample sets that have differing numbers of usable loci.

The exclusion of loci with null alleles can also be done when computing genetic distances between pairs of samples. Once again, the presence of the null alleles have to be highlighted when comparing different groups of samples. Although not conventionally done, one way this can be achieved is by computing the maximum possible genetic distance between samples [again, this cannot be computed when genetic distances using the method of Slatkin (1995) is employed as reasoned in the results chapter]. Samples with null alleles will have lower values since the null alleles cannot be compared. These values can serve to normalize the computed genetic distances so that better comparisons can be made of groups of samples with null alleles.

Overall, the presence of null alleles may complicate, but does not preclude analysis of the data. However, their presence demands that caution must be observed particularly when comparisons are made between different groups of samples. This implies that the presence of null alleles, in addition to the relatively small numbers of samples available in this experiment, limit the analysis to only qualitative or, at best, semi-quantitative methods.

5.4.2 Extent of vegetative reproduction

One of the main objectives for conducting experiments in this part of the thesis was to determine the extent of asexual or vegetative reproduction within clumps of *Acanthorrhynchium papillatum*. The use of the microsatellite markers developed earlier in this project helped in this regard. Their high combined power of discrimination suggests that samples with the same or similar multi-locus genotypes are derived from products of asexual reproduction.

Examination of the multi-locus genotypes of samples of *Acanthorrhynchium papillatum* taken from the same clump suggests heavy reliance on asexual reproduction within clumps of this species. This is supported by findings from both main and supplementary data. Nearly 3/4 of the main clumps surveyed were composed of samples with a single multi-locus genotype. Moreover, nearly all of the main clumps surveyed (38/40) had at least two samples with a single multi-locus genotype. Supplementary data show that more extensive sampling within a clump may not necessarily reveal more genotypes. In the supplementary data, clumps from 4/5 sampling localities were found to be composed of samples of single multi-locus genotypes. The single locality, Sungei Bantang Recreational Forest, with a clump having more than one multi-locus genotype, had only two, with one clearly dominating in number. The data also show that *A. papillatum* seems to rely greatly on asexual reproduction regardless of the source and, by extension, the quality of the habitat of the clump.

The data show, however, that the clumps are not purely composed of moss strands with a single multi-locus genotype. A sample different from the rest of the samples within the same clump was found in at least one clump from each of the sampling localities. Most of these samples were different at several loci and were genetically distant from the rest of the samples in the clump. This suggests that these clumps were "seeded" at least more than once by samples of different

origins. Had these samples been different at only one locus then the variation within the clumps could be explained in another way as suggested below.

Several clumps from Kota Tinggi Waterfalls Resort, had samples that were different at only one locus. Since the high combined power of discrimination (or low matching probability) of the microsatellite markers diminish the probability that unrelated samples share the same alleles for the seven loci, it is suggested that these samples are, in fact, related and are products of clonal reproduction. It is further suggested that the differing alleles are products of somatic mutation, not of sexual reproduction. Recalling that vegetative growth of mosses proceed from a single meristematic cell, somatic mutations in this cell are easily propagated in the younger tissues of the moss.

If somatic mutation is assumed to be the reason for deviations from complete correspondence of all loci between two samples, then it is intriguing to note that the differing alleles in these samples are not adjacent to the main alleles (see again discussion on the clumps APKT10 and APKT12). This implies that the assumption of stepwise mutation model of microsatellite evolution for this group of samples may be false and that the markers are evolving in a different way. However, more extensive testing needs to be done to confirm this and all the other preceding suggestions.

The suggestions on somatic mutations also have implications on the treatment of null alleles so pervasively found in the samples. If asexual reproduction can give rise to two samples that share the same alleles in all markers except one (the one where somatic mutation has taken place), then conversely, this implies that two samples need not have the same alleles at all loci for them to be considered products of asexual reproduction. This means that samples having null alleles at a single locus can still be considered products of asexual reproduction. The true alleles could be different or the same and it would not change this evaluation.

There would be no need to assume the true alleles before deducing that the samples are products of asexual reproduction.

The suggestions above may be true for samples that are null at only one locus. Many of the samples, however, are null at several loci. Caution must still be taken and confirmatory steps must be done in future experiments if adequate comparisons are to be asserted for these samples.

5.4.3 Presence of multiple genets in a clump

The presence of multiple genets in a clump, whether from somatic mutations of products of asexual reproduction or from sources exogenous to the clump, directly contributes in maintaining variation in the population.

5.5 Conclusions

This chapter fulfills the fourth and last objective of the thesis in addition to completing the picture on the genetic diversity of *Acanthorrhynchium papillatum*.

Notwithstanding the limitations of the sampling and the problems associated with null alleles, the results on the experiments to characterize the genetic diversity within clumps of *Acanthorrhynchium papillatum* still provide interesting insights into the biology of the moss.

Asexual reproduction seems to be more important than sexual reproduction within clumps of *Acanthorrhynchium papillatum*. This was evident while utilizing two modes of sampling and three measures of genetic variation. Asexual reproduction seems to be the prevalent mode within a clump regardless of its source or the quality of its habitat.

Clumps of *Acanthorrhynchium papillatum* are not always solely composed of a single genet, however. Individuals of different multi-locus genotypes were detected within clumps from each of the sampling localities. Individuals of different

multi-locus genotypes may have originated exogenously or from somatic mutations in mosses that originally were clonally produced. The number of loci where the samples are different can offer clues on the origins of the differing genets: when individuals are different at only one locus then there is statistical reason to conclude that the origins are from somatic mutation. Differences in multiple loci could either be from somatic mutations or from exogenous sources and further confirmatory experiments need to be done. Nevertheless, the presence of different genotypes helps contribute and maintain genetic variation in the population.

Further experiments would be desirable to elucidate the true origins of the different multi-locus genotypes in clumps of *Acanthorrhynchium papillatum*. Moreover, better, more extensive sampling can perhaps be made and tests to see if sample number, marker number, clump number are sufficient to draw quantitative conclusions on the within-clump diversity of the species and how they compare among the different localities.

Chapter 6

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

This thesis project is the first report on the population biology of a moss indigenous to Southeast Asia in almost 25 years. Four main objectives were set at the beginning of the study, each of them was fulfilled at the end.

The first objective was to develop microsatellite markers for a moss that was native to the Southeast Asian tropics. Eight microsatellite loci were isolated, characterized and evaluated to be suitable markers for studies on the genetic diversity of *Acanthorrhynchium papillatum*. The markers were developed only after several attempts and considerable optimization of laboratory techniques. Microsatellite libraries were also constructed for three other species of mosses, *Pogonatum cirratum* ssp. *macrophyllum*, *Thuidium plumulosum* and *Thuidium cymbifolium*. Microsatellite markers were developed for *P. cirratum* ssp. *macrophyllum* by honours students. Markers could not be fully developed for *T. plumulosum* and *T. cymbifolium* because of limitations in time and other resources. A welcome offshoot in the development of microsatellite markers, however, is that the lessons learned in the process will make it easier to develop these markers for other species in the future.

The second objective was to obtain baseline information on the genetic diversity among clumps of *Acanthorrhynchium papillatum* using both the newly-developed microsatellite markers and *ITS*2 sequences. The third objective was to determine if genetic diversity as measured by these markers was affected by the quality of the habitat in which the moss is found. To address these objectives,

samples of the moss were collected from habitats of different degrees of degradation and examined using both markers.

Using the microsatellite markers, high levels of allelic and genotypic diversity in *Acanthorrhynchium papillatum* were discovered in each area sampled. The results obtained oppose the early notions that mosses have low levels of genetic diversity. Such high levels of diversity were attributed to high mutation rates seen in the microsatellite markers used. The comparison of genetic diversity among clumps from different sampling areas revealed that areas that were qualitatively determined to be of reduced habitat quality had slightly lower levels of allelic diversity. Genotypic diversity in these areas was markedly reduced suggesting that, in habitats of reduced quality, vegetative reproduction was the more important mode of reproduction in *A. papillatum*. Somatic mutation in individuals that are products of vegetative reproduction is hypothesized to allow the species to maintain genetic variation when other sources of variation are restricted.

A small but important finding in the study using microsatellite markers is that experimental conditions, including the type of instrument used, can affect the data. While fragments of the same length are consistently sized using the same instrument type, they may be sized differently when another instrument type is used. Relative differences of different fragment lengths within one type of instrument, however, remain the same. This allows similar statistical data to be obtained when a different instrument type is used on the same data set. Within one study, however, it is imperative that the same experimental conditions be used.

Another molecular marker, sequence information from *ITS*2, was also used to study the genetic diversity among clumps of *Acanthorrhynchium papillatum* to supplement the data obtained using microsatellite markers. Considerable levels of nucleotide and haplotypic variation were detected in this locus between and among

the different samples studied, again belying the early assumptions that mosses are genetically depauperate. Comparisons of *ITS*2 measures of variation to their analogs in microsatellite markers reveal that higher levels of variation is seen with microsatellite markers than with *ITS*2 sequences, underscoring the importance and utility of using microsatellite markers in population genetic studies.

As measured by both markers, significant genetic differentiation and moderate to high levels of genetic structure are seen among the populations studied. These results suggests that the *Acanthorrhynchium papillatum* populations from the different sampling areas are genetically isolated despite the short distances between and among these areas. However, the high rates of mutation particularly evident in the microsatellite markers might mitigate any tendency towards allele fixation by the creation of new alleles by way of mutation (instead of recombination through sexual reproduction or through gene flow from neighboring areas).

The fourth and final objective was to describe the genetic variation within clumps of *Acanthorrhynchium papillatum* using microsatellite markers. Branches of this moss were selected from clumps sampled from the same areas and habitats visited earlier and studied using microsatellite markers. Very low levels of diversity were found within clumps of this moss regardless of the habitat quality of the source of the clump. The results therefore show that vegetative reproduction is the more important mode used by *A. papillatum* within clumps. Despite this finding, low levels of allelic diversity were also found, corroborating the suggestion that microsatellite mutation rates in this moss are high.

The high rates of mutation seen in both microsatellite loci and *ITS*2 sequences of *Acanthorrhynchium papillatum* have interesting implications in the survival of this species. In the absence of sexual reproduction or exchange of genetic material through migration, genetic variation can be maintained in an isolated population through somatic mutation. This source of variation, on top of the ability of the

moss to reproduce vegetatively, allows the moss to produce more individuals that can themselves undergo somatic mutation. Thus, even an isolated population of this moss may survive or even thrive as long as there are places for it to grow. This finding, however, is not a license to disturb or destroy the remaining habitats of this moss but rather is a testament to the ability of a species to survive in the face of habitat destruction.

While the results obtained from this study answered the questions that were set at the beginning of the project, improvements to many aspects of the project are possible. No thesis project, however carefully planned and executed, is perfect. This one, in particular, could have benefited in several ways:

Despite many attempts at isolating microsatellite loci and evaluating them as markers, only 8 were successfully developed. If more microsatellite markers had been found and used then they could have strengthened the data and lent more rigor to the findings.

The microsatellite markers could have been better evaluated (particularly regarding linkage disequilibrium) if diploid (sporophytic) specimens of *Acanthorrhynchium papillatum* had been available for collection and analysis. Unfortunately, few sporophytes were encountered in the field. Moreover, DNA extraction attempts on these few specimens were unsuccessful. The collection of enough sporophytic material and the development of a method for extraction of DNA from these samples would help more fully characterize the microsatellite markers developed for this species.

Overall, significant insights into the genetic diversity of the moss, *Acanthorrhynchium papillatum*, were revealed by this thesis project. However, it should not be forgotten that *Acanthorrhynchium papillatum* was studied as a representative of pleuocarpous or creeping mosses in Southeast Asia. Thousands of other pleurocarpous moss species under threat from habitat loss remain unstudied. The

results from this study, while useful, only suggest what could be happening to the population biology of other species; the only way to really know would be to study these species directly. With the development of new techniques and the inevitable improvement of technologies in the life sciences, the study of the genetic diversity of mosses is now easier than when the project was first started. It is hoped that the present project is only one of the first of many future enquiries into the genetic diversity and population biology of mosses in Southeast Asia.

REFERENCES

Acevedo-Rosas, R., Cameron, K., Sosa, V. & Pell, S. (2004). A molecular phylogenetic study of *Graptopetalum* (Crassulaceae) based on *ETS*, *ITS*, *rpl*16, and *trnL-F* nucleotide sequences. *American Journal of Botany* **91**, 1099–1104.

Akiyama, H. (1994). Allozyme variability within and among populations of the epiphytic moss *Leocodon* (Leucodontaceae: Musci). *American Journal of Botany* **81**, 1280–1287.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Anderson, L. E. (1963). Modern species concepts: Mosses. *The Bryologist* **66**, 107–119.

Assmussen, C. B. (1999). Towards a chloroplast DNA phylogeny of the Tribe Geonomeae (Palmae). In: *Evolution, Variation and Classification of Palms* (Henderson, A. & Borschsenius, F., eds.), vol. 83 of *Memoirs of the New York Botanical Gardens*.

Baldwin, B. G. (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. *Molecular Phylogenetics and Evolution* **1**, 3–16.

Bartram, E. B. (1939). Mosses of the Philippines. *Philippine Journal of Science* **68**, 1–437.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–580.

Bilofsky, H. & Burks, C. (1998). The GenBank genetic sequence data bank. *Nucleic Acids Research* **16**, 1861–1864.

Botstein, D., White, R., Skolnick, M. & Davis, R. W. (1980). Construction of genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–331.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

Bruford, M. W. & Wayne, R. K. (1993). Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development* **3**, 939–943.

Buryová, B. & Shaw, A. J. (2005). Phenotypic plasticity in *Philonotis fontana* (Bryopsida: Bartramiaceae). *Journal of Bryology* **27**, 13–22.

Caetano-Anollés, G. & Gresshoff, P. M. (1994). Staining nucleic acids with silver: An alternative to radioisotopic and fluorescent labeling. *Promega Notes Magazine* **45**, 13–20.

Capesius, I. (1997). Analysis of the ribosomal RNA gene repeat from the moss *Funaria hygrometrica*. *Plant Molecular Biology* **33**, 559–564.

Cavalli-Sforza, L. L. (1998). The DNA revolution in population genetics. *Trends in Genetics* **14**, 60–65.

Chang, Y. (2003). *The systematics of Sematophyllaceae (Musci) of pensisular Malaysia, Philippines and Singapore based on* rbcL *gene sequences and morphological study.* Master's thesis, National University of Singapore.

Chiang, T.-Y. & Schaal, B. A. (1998). Universal primers for amplification and sequencing a noncoding spacer between the *atpB* and *rbcL* genes of chloroplast DNA. *Botanical Bulletin of Academica Sinica* **39**, 245–250.

Chiang, T.-Y. & Schaal, B. A. (1999a). Genetic diversity of internal transcribed spacer 1 region of nuclear ribosomal DNA in moss family Hylocomiaceae. In: *Biodiversity and allelopathy: From organisms to ecosystems in the Pacific* (Chou, C. H., Waller, G. R. & Reinhardt, C., eds.). Academia Sinica, pp. 47–62.

Chiang, T.-Y. & Schaal, B. A. (1999b). Phylogeography of North American populations of the moss species *Hyloco*mium splendens based on the nucleotide sequence of internal transcribed spacer 2 of nuclear ribosomal dna. *Molecular Ecology* **8**, 1037–1042.

Corlett, R. T. (1992). The ecological transformation of Singapore, 1819–1990. *Journal of Biogeography* **19**, 411–420.

Cummins, H. & Wyatt, R. (1981). Genetic variability in natural populations of the moss *Atrichum angustatum*. *The Bryologist* **84**, 30–38.

de Vries, A., van Zanten, B. O. & van Dijk, H. (1983). Genetic variability within and between populations of two species of *Racopilum* (Racopilaceae, Bryopsida). *Lindbergia* **9**, 73–80.

Derda, G. S. & Wyatt, R. (1999). Levels of genetic variation and its partitioning in the wide-ranging moss *Polytrichum commune*. *Systematic Botany* **24**, 512–528.

DeWoody, J. A. (2002). *DeWoody's microsatellite cloning protocol (Spring 2002) based in part on the SNX-linker protocol by Hamilton et al. 1999 (*Biotechniques *27:500) and on Travis Glenn's "Dynabeads2002" protocol*. URL `http://www.agriculture.purdue.edu/fnr/html/faculty/DeWoody/DeWoodyweb/pdfs/msatclngprtcl.pdf`.

Dover, G. A. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics* **2**, 159–165.

Doyle, J. J. & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15.

El Mousadik, A. & Petit, R. J. (1996). High level of genetic differentitation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832–839.

Ennos, R. A. (1990). Population genetics of bryophytes. *Trends in Ecology and Evolution* **5**, 38–39.

Estoup, A. & Turgeon, J. (1996). *Microsatellite markers: isolation with non-radioactive probes and amplification*, 12/1996 ed. URL `http://www.inapg.inra.fr/dsa/microsat/microev1.zip`.

Excoffier, L., Estoup, A. & Cornuet, J.-M. (2005a). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**, 1727–1738.

Excoffier, L., Laval, G. & Schneider, S. (2005b). Arlequin version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Excoffier, L., Laval, G. & Schneider, S. (2006). Arlequin *version 3.1 user manual*. Computational and Molecular Population Genetics Lab (CMPG),

Institute of Zoology, University of Berne, Baltzerstrasse 6, 3012 Bern, Switzerland. URL `http://cmpg.unibe.ch/software/arlequin3`.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, U.S.A.

Fleischer, M. (1923). *Die Musci der Flora von Buitenzorg*, vol. IV. E. J. Brill, Leiden.

Garza, J. C. & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305–318.

Glime, J. M. & Saxena, D. (1991). *Uses of bryophytes*. New Delhi: Today & Tomorrow's Printers & Publishers.

Goffinet, B. (2000). Origin and phylogenetic relationships of bryophytes. In: *Bryophyte Biology* (Shaw, A. J. & Goffinet, B., eds.), chap. 4. Cambridge University Press, pp. 124–149.

Goffinet, B. & Hax, N. P. (2001). Bibliography of 'molecular systematic' studies of Bryophytes. I. 1985–2000. *Cryptogamie, Bryologie* **22**, 149–155.

Gómez, A., Pintos, B., Aguiriano, E., Manzanera, J. A. & Bueno, M. A. (2001). SSR markers for *Quercus suber* tree identification and embryo analysis. *The Journal of Heredity* **92**, 292–295.

Gottlieb, L. D. (1971). Gel electrophoresis: New approach to the study of evolution. *BioScience* **21**, 939–944.

Grundmann, M., Ansell, S. W., Russel, S. J., Koch, M. A. & Vogel, J. C. (2007). Genetic structure of the widespread and common Mediterranean bryophyte *Pleurochaete squarrosa* (Brid.) Lindb. (Pottiaceae) – evidence from nuclear and plastidic DNA sequence variation and allozymes. *Molecular Ecology* **16**, 709–722.

Gupta, M., Chyi, Y. S., Romero-Severson, J. & Owen, J. L. (1994). Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theoretical and Applied Genetics* **89**, 998–1006.

Hadrys, H., Balick, M. & Schierwater, B. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* **1**, 55–63.

Hallingbäck, T. & Hodgetts, N. (eds.) (2000). *Mosses, liverworts, and hornworts. Status survey and conservation action plan for bryophytes.* IUCN/SSC Bryophyte Specialist Group.

Hamilton, M. B. (2002). *Microsatellite-enhanced genomic library construction, streptavidin-biotin bench version, SNX linker*, 1.1 ed. URL `http://bioserver.georgetown.edu/faculty/hamilton/SNXmicro11.pdf`.

Hamilton, M. B., Pincus, E. L., Di Fiore, A. & Fleischer, R. C. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* **27**, 500–507.

Hassel, K., Såstad, S. M., Gunnarson, U. & Söderström, L. (2005). Genetic variation and structure in the expanding moss *Pogonatum dentatum* (Polytrichaceae) in its area of origin and in a recently colonized area. *American Journal of Botany* **92**, 1684–1690.

Hedenäs, L. (2001). Environmental factors potentially affecting character states in pleurocarpous mosses. *The Bryologist* **104**, 72–91.

Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H. & Vogt, P. H. (1997). Multiplex PCR: critical parameters and step-by-step protocol. *BioTechniques* **23**, 504–511.

Högberg, N., Svegården, I. B. & Kauserud, H. (2006). Isolation and characterization of 15 polymorphic microsatellite markers for the devastating dry rot fungus, *Serpula lacrymans*. *Molecular Ecology Notes* **6**, 1022–1024.

Hurlbert, S. H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**, 577–586.

Jarne, P. & Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424–429.

Johnson, P. C. D., Fowlie, M. K. & Amos, W. (2005). Isolation of microsatellite loci from the common buzzard, *Buteo buteo* (Aves: Accipitridae). *Molecular Ecology Notes* **5**, 208–211.

Jones, F. A., Poelchau, F., Bouck, A. C. & Hubbell, S. P. (2004). Eight microsatellite markers for the neotropical tree *Luehea seemannii* (Tiliaceae). *Molecular Ecology Notes* **4**, 5–8.

Kalinowski, S. T. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 530–543.

Kalinowski, S. T. (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187–189.

Kimura, M. & Ohta, T. (1978). Stepwise mutation model and distribution of

allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 2868–2872.

Korpelainen, H. & Salazar Allen, N. (1999). Genetic variation in three species of epiphytic *Octoblepharum* (Leucobryaceae). *Nova Hedwigia* **68**, 281–290.

Krzakowa, M. & Szweykowski, J. (1979). Isozyme polymorphism in natural populations of a liverwort, *Plagiochila asplenioides*. *Genetics* **93**, 711–719.

Lee, C. T., Lee, S. L., Faridah, Z., Siraj, S. S., Ng, K. K. S., Norlia, B. & Mat-Isa, M. N. (2006). Isolation and characterization of microsatellite markers in *Koompassia malaccensis* (Leguminosae), an important tropical timber species. *Molecular Ecology Notes* **6**, 1198–1201.

Leonardía, A. A. P., Kumar, P. P. & Tan, B. C. (2006). Development of microsatellite markers for the tropical moss, *Acanthorrhynchium papillatum*. *Molecular Ecology Notes* **6**, 396–398.

Lewontin, R. C. & Kojima, K.-i. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.

Mayor, R. & Naciri, Y. (2007). Identification and characterization of eight microsatellite loci in *Aster amellus* L. (Asteraceae). *Molecular Ecology Notes* **7**, 233–235.

Michalakis, Y. & Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061–1064.

Mishler, B. D. (2001). The biology of bryophytes–bryophytes aren't just small tracheophytes. *American Journal of Botany* **88**, 2129–2131.

Möller, M. & Cronk, Q. C. B. (1997). Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (*ITS*) sequences. *American Journal of Botany* **84**, 956–965.

Muellner, A. N., Samuel, R., Chase, M. W., Pannell, C. M. & Greger, H. (2005). *Aglaia* (Meliaceae): An evaluation of taxonomic concepts based on DNA data and secondary metabolites. *American Journal of Botany* **92**, 534–543.

Nagai, S., McCauley, L., Yasuda, N., Erdner, D. L., Kulis, D. M., Matsuyama, Y., Itakura, S. & Anderson, D. M. (2006). Development of microsatellite markers in the toxic dinoflagellate *Alexandrium minutum* (Dinophyceae). *Molecular Ecology Notes* **6**, 756–758.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist* **106**, 283–292.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321–3323.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Nei, M., Tajima, F. & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153–170.

O'Shea, B. J. (1997). A revision of *Acanthorrhynchium* (Sematophyllaceae) in Africa. *Tropical Bryology* **13**, 125–130.

Ota, T. (1993). DISPAN: Genetic distance and phylogenetic analysis. URL `http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.htm`.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Park, S. D. E. (2001). *Trypanotolerance in West African cattle and the population genetic effects of selection.* Ph.D. thesis, University of Dublin.

Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* **183**, 63–98.

Peh, K. S.-H., de Jong, J., Sodhi, N. S., Lim, S. L.-H. & Yap, C. A.-M. (2005). Lowland rainforest avifauna and human disturbance: persistence of primary forest birds in selectively logged forests and mixed-rural habitats of southern peninsular Malaysia. *Biological Conservation* **123**, 489–505.

Pfeiffer, T., Fritz, S., Stech, M. & Frey, W. (2006). Vegetative reproduction and clonal diversity in *Rhytidium rugosum* (Rhytidiaceae, Bryopsida) inferred by morpho-anatomical and molecular analyses. *Journal of Plant Research* **119**, 125–135.

Prince, A. M. & Andrus, L. (1992). PCR: how to kill unwanted DNA. *BioTechniques* **12**, 358–360.

Provan, J. & Wilson, P. J. (2006). Development of microsatellites for the peat moss *Sphagnum capillifolium* using ISSR cloning. *Molecular Ecology Notes* **7**, 254–256.

Raymond, M. & Rousset, F. (1995a). An exact test for population differentiation. *Evolution* **49**, 1280–1283.

Raymond, M. & Rousset, F. (1995b). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.

Robbins, R. B. (1918). Some applications of mathematics to breeding problems III. *Genetics* **3**, 375–389.

Rogers, S. O. & Bendich, A. J. (1987). Ribosomal RNA genes in plants: Variability in copy number and in the intergenic spacer. *Plant Molecular Biology* **9**, 509–520.

Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. & Misener, S., eds.). Totowa, NJ: Humana Press, pp. 365–386.

Ruggiero, M. V. & Procaccini, G. (2004). The rDNA *ITS* region in the Lessepsian marine angiosperm *Halophila stipulacea* (Forssk.) Ascjers. (Hydrocharitaceae): Intragenomic variability and putative pseudogenic sequences. *Journal of Molecular Evolution* **58**, 115–121.

Saar, D. E., Polans, N. O., Sørensen, P. D. & Duvall, M. R. (2001). Angiosperm DNA contamination by endophytic fungi: Detection and methods of avoidance. *Plant Molecular Biology Reporter* **19**, 249–260.

Saitou, N. & Nei, M. (1987). The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **44**, 406–425.

Såstad, S. M., Pedersen, B. & Digre, K. (1999). Habitat-specific genetic effects on growth rate and morphology across pH and water-level gradients within a population of the moss *Sphagnum angustifolium* (Sphagnaceae). *American Journal of Botany* **86**, 1687–1698.

Saxena, D. & Harinder (2004). Uses of bryophytes. *Resonance* **9**, 56–65.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365–371.

Schmidt, A. L., Scott, L. & Lowe, A. J. (2006). Isolation and characterization of microsatellite loci from *Macadamia*. *Molecular Ecology Notes* **6**, 1060–1063.

Schofield, W. B. (2000). Classification and number of bryophyte species. In: *Mosses, Liverworts, and Hornworts. Status Survey and Conservation Action Plan for Bryophytes* (Hallingbäck, T. & Hodgetts, N., eds.), chap. 2. IUCN, Gland, Switzerland and Cambridge, U.K.: IUCN/SSC Bryophyte Specialist Group, pp. 2–5.

Selkirk, P. M., Skotnicki, M. L., Adam, K. D., Connett, M. B., Dale, T., Joe, T. W. & Armstrong, J. (1997). Genetic variation in Antarctic populations of the moss *Sarconeurum glaciale*. *Polar Biology* **18**, 344–350.

Selkoe, K. A. & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615–629.

Shaw, A. J. (2000). Molecular phylogeography and cryptic speciation in the mosses, *Mielichhoferia elongata* and *M. mielichhoferiana* (Bryaceae). *Molecular Ecology* **9**, 595–608.

Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A. & Lukyanov, S. A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* **23**, 1087–1088.

Skotnicki, M. L., MacKenzie, A. M., Clements, M. A. & Selkirk, P. M. (2005). DNA sequencing and genetic diversity fo the 18S–26S nuclear

ribosomal internal transcribed spacers (*ITS*) in nine Antarctic moss species. *Antarctic Science* **17**, 377–384.

Skotnicki, M. L., Selkirk, P. M. & Beard, C. (1998). RAPD profiling of genetic diversity in two populations of the moss *Ceratodon purpureus* in Victoria Land, Antarctica. *Polar Biology* **19**, 172–176.

Skotnicki, M. L., Selkirk, P. M. & Ninham, J. A. (1999). RAPD analysis of genetic variation and dispersal of the moss *Bryum argenteum* in Ross Island, Victoria Land, Antarctica. *Polar Biology* **21**, 417–422.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191**, 528–535.

Smouse, P. E., Long, J. C. & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**, 627–632.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M. & Powell, W. (2003). How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* **12**, 1339–1348.

Stech, M. (2004). Supraspecific circumscription and classification of *Campylopus* (Dicranaceae, Bryopsida) based on inferences from sequence data. *Systematic Botany* **29**, 817–824.

Stenøien, H. K. & Flatberg, K. I. (2000). Genetic variability in the rare norwegian peat moss *Sphagnum troendelagicum*. *The Bryologist* **103**, 794–801.

Stratagene (2003). *pBluescript® II phagemid vectors instruction manual*, revision #083001m ed. URL `http://www.stratagene.com/manuals/212205.pdf`.

Sun, Y., Skinner, D. Z., Liang, G. H. & Hulbert, S. H. (1994). Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **89**, 26–32.

Swofford, D. L. (2002). PAUP*. *Phylogenetic analysis using parsimony (*and other methods). Version 4.* Sinauer Associates, Sunderland, Massachusetts.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105–1109.

Takahashi, T., Tani, N., Taira, H. & Tsumura, Y. (2005). Microsatellite markers reveal high allelic variation in natural populations of *Cryptomeria japonica* near refugial areas of the last glacial period. *Journal of Plant Research* **118**, 83–90.

Tan, B. C. & Buck, W. R. (1989). A synoptic review of Philippine Sematophyllaceae with emphasis on Clastobryoideae and Heterophylloideae (Musci). *Journal of the Hattori Botanical Laboratory* **66**, 307–320.

Tan, B. C. & Chang, Y. (2004). Molecular contribution to the systematic position of *Mastopoma scabrifolium* (Broth. *in* Moell.) B.C. Tan & Tran Ninh (Sematophyllaceae, Bryopsida). *Cryptogamie, Bryologie* **25**, 301–308.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

van der Velde, M. & Bijlsma, R. (2000). Amount and structure of intra- and interspecific genetic variation in the moss genus *Polytrichum*. *Heredity* **85**, 328–337.

van der Velde, M. & Bijlsma, R. (2003). Phylogeography of five *Polytrichum* species within Europe. *Biological Journal of the Linnean Society* **78**, 203–213.

van der Velde, M. & Bijlsma, R. (2004). Hybridization and asymmetric reproductive isolation between the closely related bryophyte taxa *Polytrichum commune* and *P. uliginosum*. *Molecular Ecology* **13**, 1447–1454.

van der Velde, M., During, H. J., van de Zande, L. & Bijlsma, R. (2001a). The reproductive biology of *Polytrichum formosum*: clonal structure and paternity revealed by microsatellites. *Molecular Ecology* **10**, 2423–2434.

van der Velde, M., van de Zande, L. & Bijlsma, R. (2001b). Genetic structure of *Polytrichum formosum* in relation to the breeding system as revealed by microsatellites. *Journal of Evolutionary Biology* **14**, 288–295.

van der Velde, M., van der Strate, H. J., van de Zande, L. & Bijlsma, R. (2000). Isolation and characterization of microsatellites in the moss species *Polytrichum formosum*. *Molecular Ecology* **9**, 1678–1680.

Vanderpoorten, A. & Tignon, M. (2000). Amplified fragments length polymorphism between populations of *Amblystegium tenax* exposed to contrasting water chemistries. *Journal of Bryology* **22**, 257–262.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Weber, J. L. (1990). Informativeness of human (dC-dA)$_n$ · (dG-dT)$_n$ polymorphisms. *Genomics* **7**, 524–530.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**, 7213–7218.

White, T. J., Bruns, T., Lee, S. & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols: a guide to methods and applications* (Innis, M., Gelfand, D. H., Sninsky, J. J. & White, T. J., eds.), chap. 38. Academic Press, Inc., San Diego, California, USA, pp. 315–322.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531–6535.

Wilson, P. J. & Provan, J. (2003). Effect of habitat fragmentation on levels and patterns of genetic diversity in natural populations of the peat moss *Polytrichum commune*. *Proceedings of the Royal Society of London B* **270**, 881–886.

Wittall, J. B., Hellquist, C. B., Schneider, E. L. & Hodges, S. A. (2004). Cryptic species in an endangered pondweed community (*Potamogeton*, Potamogetonaceae) revealed by AFLP markers. *American Journal of Botany* **91**, 2022–2029.

Wyatt, R. (1992). Conservation of rare and endangered bryophytes: Input from population genetics. *Biological Conservation* **59**, 99–107.

Wyatt, R., Odrzykoski, I. J. & Stoneburner, A. (1989). High levels of genetic variability in the haploid moss *Plagiomnium ciliare*. *Evolution* **43**, 1085–1096.

Zane, L., Bargelloni, L. & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**, 1–16.

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

Zietkiewicz, E., Rafalski, A. & Labuda, D. (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**, 176–183.

Zouhair, R., Corradini, P., Defontaine, A. & Hallet, J.-N. (2000). RAPD markers for genetic differentiation of species within *Polytrichum* (Polytrichaceae, Musci): A preliminary survey. *Taxon* **49**, 217–229.

Appendix A

"DIRTY" SEQUENCES

The following are the FASTA-formatted, "dirty" sequences from both APMS and APBMS libraries. Sequences are dirty when too many uncertain base calls prevent the accurate identification and trimming off of linker and cloning vector sequences. (In two cases, the insert is a chimera of two microsatellite-enriched fragments and is also classified as dirty). Where forward and reverse contigs of a given plasmid were impossible to assemble (given the uncertainty of many of the base calls) only one sequence, that from the primer end indicated, is presented.

>APMS13
GATTNTTTANAAGTTCCTGNTNGGGTTTTCGNCCCCTTNTNAANTTGAGGGGTCGATTTTTTTGGAT
TGNTCGTTCAGGGGGGGNGGAGNCNTATGGAAAAACGCCAGCAACGNGGNCNTTTTTACGGTTTCCT
NGNCTTTTTGCTGGCNTTTTGCTCACATGTTNTTTCNTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG
AGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAAT
GCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTA
GCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTG
AGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTC
ACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGCAAGGCCTTAGCACACAC
ACACACACACACACACACACACGCACACACACATACACTCACCCTGGTATCCTCCGTCAATGGTCCA
CTAATACCTTCAGCTTCTGCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTA
TCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGC
GCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCT
TGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAA
CAGTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAA
TTTTTGTTAAATCAGCTCATTTTTTTAACCAATAGGCCGAAATCGGCCAAAATCCCTTATAATCAAA
AGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAANGAGTCCCCTATAAAGAACGGG
GGACTCCCACGTCCAAGGGGNGAAAAAACNGCNCTATNAAGGGGGGATGGGCCCACCTACGTGNAAC
NATCCCCCCTTAATCAAGTTTTTTTGGGGGGCNGGAGGGGGCCNTAAAAGNNNCTAAANCNGGNAAC
CCCTTAAAGGGNGNNCCCCCCGAATTTNTAAAANTTTTGACGGGNGAAAAAAACCCGGNCAANCCTG
GNCAGGAAAA

>APMS19 --T3 end
GCCNNNNGNNCANNGTTTGTANCTCNNNCCNNGNNGGANGNNNNNAAAGGGANNGNNNGGCNCTTNG
GNCCNNCTCGCNANACTGAAANNATNCNCNGNNNNNNCTNTGCCCNTCGCNCTNNANNNANNATNGN
CGAGCNNNCTTANATTATTNNCTTTTCCCCCGANNAANATTNTGNNTANGACCANAGCACTGCNNAN
GCAAGAATCNNTCNCNTGTATNTCTNNNNNCCCGCCCCGCTNNTNAACNTCNCNNCNNNNNAAACNA
NNACGCCTGNNNGNCTTNGGTNTNTATANAANNNGAAAGATNTTNNANGAACAGANTTGTTNNANGA

NNGNGGGTGNGTNNNNGNTNANAGTANNNNCTNNCACTCCTNCCTCTTTGNCCNNNCACTNTTTGNA
ACCNNANNTNTAGNNATCGANGCNNNNANCGNNNGACNATTNNNTGNATNCAGNCTCTCTTNACNNN
TNNNATGCCTNTCNCNNCCANCNAACCTTANGTNNNNGCTCGTNTNNGTNCNTNGGCNNANTNCNGG
AGANNNNNAGACANGNAGCAAGNTNTNATNANANNTCTATTNNANGGGNNANGNGAGNAGANNCNGT
GGTGATACTTCCTCTNTTNAAGNNGTATNATAAAGNANGGANNGNNNANNGGNGNCAGGGNGANGAC
NNNTTCACCGNTNAANGNNGNTGNANCTNTNNCTTCTCTTCNNCNTTTCATCTCNATCTNTCTNGNG
NNTGTGGCTGNCNNNNNTNNNGNCNTCGCNAGNTATGTANANGNGTTNTNAGCTANTACATANAANN
GTNTATGAGNCTNAANANTNCTCGANTTNGTNNNCAGNNNTCGANNCTTNNGNNCNNCGTNNGNCGA
NNNTNNCNAANCCNATTCACNTCNNANCTCTNNGNNNNGAGGNNNGCTNTNATGTAANGNTGNGNNG
GTGGCNTCCGANGCTGANNCGTNCGTNCGNNNTGGCGCCGAGCNNAGAGNCTCGCAACNGGNCGNAG
ANNTCTACTNCCACNNGNNNACNTCNGNNCAANACNNANTNGNGANCAGTGTGCNAGTACTGNANTG
AGCANTCCANNNNNNCANTATATANCTATNATNTNNACNTNNTGTNNAGNGANNCNTATCANTGTTA
NGNACNGGTNNNTACNNCAGANCNTGTATATGTCNGTNNCTNNGNNATTNTCCCNCTNNNNANCACT
AGTNNCTCCCNCTGNTNCTNTGCNNTTNANTCNNTCCTNANTNNTCNTATNTCNCNTNTTNNNNGNT
NTNAGGNTNCNGNAGTNANGGCNTNNNANNNNNGNNTNTNNNTG

>APMS49 --T3 end
CCCCGGGGGGNAAANNANGANTTNNNNNCCTTAGGAGGAGNAAAAAGGGNGGANGCCCTTTTTTCCC
CCNATCNGNNGNAGAAAAAAANNTNGTNGANANGATNCNGNANNNACGCACANGGGNTCTACNGCCN
CNGGNGGANGATTGANNNCCNNCGAGNGGACGACTNTNNGCTNNCNNCGGANTACANATANANNTCN
AGAATCAGNGCANAGNAANTNGCTANGNNCNCNGNCNCCCNCNNGNTNGTCNANNCNTNNNCNTGGC
NANNNAANNGNNGNTATTACTANNACNATAGNTNTNNGNGNCAANANNCNNANNCNCANNNAGGATC
NGCGGGNGGTCNTNNNCNTGCTGNNGCNNGGAGNGNGNNTGGNCNNNCNGGCGGTANGNGAGGCGCT
ACNCNNAATNATAANCNTANTANNANAANNNNGANTANNAANNNANTGNNNGGNAGGCNTCANNTAG
NCNCNATTNNNTNANATGNNCGGNGCNNNNNATTGTTATCTANTTNANATANAAANGNANGNNGTNT
NNACTNAATNANNCCCGNNTCNNNNGCCTANNTNANAAGCGCCTNANANAGCGNNGGNNGAAGGNNA
GTNNNNNNCNCTACNCNTTNTNGACTCACGCNACNCTGGATNANATNNTGNTNNNTNGNGNCGGAAG
AANANANGACNNATGNTNNTANTTNCGCANNCAACANNANNNGAAGGGCGGTAGNNNANCNGTAGNN
TGNATNNNGNTAGNNCGTTNNNNCGGGGNGGCGGTAGGGNGTNNGNNAGNCAGNNGNGGNNAGAGNN
TGNCAGNNGAGNTNCNAGNGGAGTNNCGNGTGNNNTNACGCNATNANNNAAATAANTNATGCGNCNG
NGNNTANNTGTCNNTGGATCGAGAATGTTTTANANNNNAATTATATNNNNGTGCNNTNTNAGCTNNN
AAGAGGTGANTGNNNNGGATCATNCANTNCANGAAGGGNCNNNGTNANNNAGNCNNATNGANNTGTA
GCGGNAGNNGNNNNNGCATAGNGNNAGCAGCGNGNNCNCNNTCTNGTNNGCGNAGTCGGANNNCTAC
NAAGNGNAGACACNCNTNANNNANGTACGAAAGACNGAAANACNNTGCNAANGNACTGACNNNCNNA
CAGNANGANAAATGNNTCTTCTCNNNNCTNNCCTCTNNNCGCNCGCNCCTNGGCNCATNANTACNAT
CNAAAGCNAAGTANAGAGAANATNAANNGNNTNNNNNGNNNATNAGAGCCTNNCNACNTNCNNACCT
GACT

>APBMS9 --T7 end
NGNTNAGGNNGANAGGCNCGGGCTTTTTCCTAACACGGGGGAACAAAANAAANAGGGNGATGCNGAA
NGNTNGNCTNGTTATTGTCCANNCCCCCCAGGNNTCCGCCGCTGANNNGGGTNCGTATNGAGNGGGA
NGACNGCNGGCGNNCNANCTNTNNNNNNNTTNNAGNNNNNNGNTAAGNNNCNNNNNAAANAAAANAA
NCANNCNCGTCTTGNGANATNNNGAGTGGNNNNGCTNANCGCNATNCTCCCANGANGNNGCCGAGAA
CNANCCCCNATTTTTTTCNNGCNNGCCANNGCGNCNGGAGNACAAGANNANNCGCNCNCCAATCNNA
ANAGNNNGAAAGTCNTTTNNANNNNNCGCNGNCGCNCNNGNGNNAANTAANGNNNNNAGNNCGCGGC

GGGGNNNNNNNNGTCTNTNNNNTTTNTGGTCNNCNNTNGCNGNNACCNNNGNGNGGNANAGGGNNGA
NNGNGNNNNTNGGGATGNGACANNTACNTNAGTAGANGNNTNTNNANNGNNNCANCTANNCNCCAGC
CTNGNTGNACCNNGCNACNANTNCACCCGCACNANGNNGGCCNNNTANCCCNCTCCCANTACTCNCN
ACACCNGCNNACANCNNCNCGCNTCNNCNNCCNNGCCCGGCNNNNCTNANAGCNNCCCCNNANNCNN
TCACNCNNTCCTCGNNTNGAGGGGGNTNGNTNANAAANNGNGCANNANTNNGANNNGNANNNNTGAC
NNNTGCNGCNCNNNNCNCCGNNNCCGNNTNCCCNAGGNNNGNNGANGCNCANNNCNGGGNGCGNNNC
GCNNGGNNGNCNNGNNNCNNNNNCAGNGNGGGNNANACGCGCCGCNNCCTACCCCACCNNCCCTCTN
AACCNNACCCNGCGGGNTCNAGACGNGGANNGGGGCCTNNAGGNNNNNAGGAGGAGNNAGNNNNNCG
CGGCGNNCCCNGNCGCNCTGNNGGCCCGNTCNGCAGNCGCAGNNANNANAGAATANTATCANANNNC
ATNGAATCNANCNTCGCNCGCCTNAGCNANGTATGTTCAGNNCATTTNTAGNTNAGTNNTNTNTNTN
NANCTAANANANCNATAAGAGCAAGTAGGNATGNAACAAANGAATNANNNGNTNTANCAGNACANNA
ANNNTCCANTGTGCANCNNAGCNCGTAGAGNCNGANCCNANCGAAG

>APBMS26 --T7 end
GGNAGACNCGNGCNTANNNNNGNNANGNACNGNCNNTNNGNGTNGTATTTNGGNNNAGAGTNANGNT
ANACTCAGNAGCNCGGNGCTACANCNANATNNNNNNCNNCNTNANNTNTNGGANCGNNATTTGANTN
ANNNNCANNNTNCCNAANTCNTGNTNGTCNGNNNNNATNCGTCTANNNNTNCNNNCNGNATAGTNAN
TANAANNTANNNTCGNGNNNGTCNNNNATGNAGGNTNNNTNNACNNCAGAGTANCNAGTNGNNCGNA
NNCAGNNTGAGACTGCGCACNNGANGCGGANTAAATGCNNNGGNNGGGTTCANNATGNCCCANNGTN
NNNCTCNCNTCGCTNACCNNNACGNGCNGCTNGGNCCNCNNNATCGGNGNGNNAGACCGNGNNNTGA
NNANNNNTANNNATNNTGNNNCTNANTNTGNANNNATTNTCGGGNTNNCNTNACGTCGCNAAANNCG
TNGTNTGCCAAGAGNNNTNNNCCANCATGACCNTTNTNNTCNCCNCNNNNAGNCTTNTCNNCNCNNN
NNNCGCNCNGNCNNCCTNNTCNNCCTNCTNCGTTNCCTCTNCGNACTNGATNNTNTCCTCNTNNNCC
NCNCGNNTCTNCCNGNCCTTNNNNTNTNTNNNNGNTTGATNTTGNTTNNTGTNGANTANTGNNNNTG
ANNGNNTCTCNATTTNNNNNGNNNATNCAAGNNAGNNNAAGGGANAANAATAAANACGAATAAAAAA
CAGNGTNNNAGANAAGNAAACGNANGNACNATNNCCNNTTGNGTGGNNGNNNNNNNANTNCACGNNN
GGANANTTGNCCNANNAAAAAGNNNAGNTNNGGCNTGGNNGGNGGANNNNNANACCCANNNGNAAAN
NAGGTGTGNNCNCGGNGNGNGTNNNCNTCNGGGTNNCGNNANGCAAGNGNNCGNCCNCCNNNANAGN
AAGAGCGNNNNNTNNNAACNAGCAANGNNTGATNTNNNNGAGCNNNGNNACCNNCACNNCNCANNGN
CGNNNNNGGNNATNGTNGGAGANNACGATGTNCGGCNGTTGTTANTNACNAACCACNAGACNCNAAN
NGTGGTGAGGCGNCCNNGGTCNCNNACAAACANNANATNNNNNCCCTACNNNATNNGTNNNACTNAG
GAGNGATATAGNGCNNAGNTNAGCAAGCNACAACGACNNNGCAACNACANNTNANNNGGAACACNAN
GAANTTNNCCNGTANCNNANGCACGCACATCNGCNCTNATGNGANATGANNCGNTNCCCGTCNCNNC
NCGNCCTCCGNGTCGGTNNNNCGGCGNGNNNNGANNGAGNTGGCNNACANNGCCACANNCNCNTTTT
GNGTTAGTGCGCNTNGANTGGGCTGATNNTAGATTACACGNGAGANAACTNANATAATAGATNNATG
ATAANNNACTNGNCNCANNTACNTACACTNAAAACANNATTACANCTCCCNNGGGCNGNNNTNGCNG
NNNACACANACTCNTNTNCCCNTGNACTNTGATGNCANNAGGNTNCGCGTNNGCNGCGANNNGGNNA
TAAGNNACNANNAAATNGAATGNTGGACNNN

>APBMS27 --T7 end
TCTCCCCGTTCNTCCCNCCCTCCCGNCNCTTCNTCNTNCCGCTCGNCNNNNANANCCNNNNCNNCNC
CGNNCCGANNANNACCCNACGANAANGCANGCCGNNGACTANNCTNNNGNANCNNNNCCGANGNGNG
NNANGNNNNTNCNATCCANNCATCTCNNNTCCCCGNANCNNCCTNCNANTTNNGGNCNTNCNNNNAN
TANTTNCCNCCNNCCCCGNNAACTCAGGNANNNCCGGNCNNGGCGTCNCGACCNGCACCCGNATCNN
NAANNGATNCCNNCANNNNNANAANANTNTNNTTTTNNTTACNTANCTNNTTTNTTTTNNTNGNTGN

CNNTNTNNNTNTTNNNCAGNNNNNNTNGNNNCNCNGNGCNGACCGCNACCCCNCCNNCCCCGCCTGC
CCANNCCNNNCTGTCTGTCNTNNTTCNGANNANANCNCATCATCCCTCNTNNTATNTACCCANTTCA
ANTNTTTGGNNNCNNNNTCNTNNGGGANNGCNNNGNNNTNGGNGTNGGATANNGNANCGNTATAGNN
NNTGNNNNTNTGANNNTGTNGCNNCCTCGATGNCCGGANNACNTANANNANNNNNNGNTGCNAGTGG
TAGTCGAATNTCTGGGNTAANNCGNAGCNTANCNANACTCANTANATNNGCNNTGANCGNTNCCCNG
GGNNGTGCANCNCNCCNCCNCNNNGNCCTNGTCGTNNGNCCCTNNTACNNNGACCGCTGNNCTANGC
NGACATGNCTNNANGCGANNTGCNNNNATANNGGNAAANNACTNGCTNCNCCGTTCNCGNTCTNNCT
CCNTNNTGTCCNNTCTANNCTGCNNCNGNCCNNCGCTCNGGNNNGNCNNTACTCCNTCNGCCTNGGC
CNCNCNNATCGNTTCTNCTCCATCTNTCGNCNGNCGCNNCCNGNTTNNNNNNATAGNACGNNGTNTG
CATNCCTNAATAAGCAANANCNNACANANAANNTNCANNAATCAGNCCGNCAACATCANTNANATNA
TNCNANATNNCNNNNANANTCNNNCNTNCCACGATNNGNACNNACCNGNTCNGCNNNNCGTCCNGAG
GTCTCTCNNCAANTGGANANCTGANNNCGTNTNCGNCNCTANTCCGCGCGACTCNNCNGCCTNTCTT
ACNGACTNACNTNACCTCNNNANTNNACNNCNCANNNNCANACANGGATCTGNGCNNNACANNGCNN
GNNANGGNAATCNANGNNNNNATAAACTNCTANCGTNCNGTGNNCGGNGCCNCACTCCTCNGCTCNN
NTCNNNCGCNNGNNTNNNTANNNNNACNNCTNGNCGNNNTGTNNGNGGCCGCNTGNGTCANGAGNGC
ACGAANNNNACATNGGACNCNNNTNGANNTCGANCCCANGANCNCTGAANAGCNNCTGTCANCANNN
NANCNGNCTATAGNGCGNAGCGCGNNNAGCNNCCCCGCCCNNNNCNCTCGANTTNCCNCCCACNCAN
GNNACNNGNNNCGNCACNCGCCAGGGGNNNTGCCCCAGCGCNANCNCTCCNGAGCTCNCACANGNNN
CGACACNNNGGNNNNTNGCCCAC

>APBMS28 --T7 end
CCCNNNNGTNNNNNCAACCTTGCNTNNGGNNTTTCNTGNNGTNNNGGGGGTTTCNTNNCNTNNCNCN
TCCANANNNCCTCCCCNCTNNTNCCTNTNANCCAAGTGNCAAGAGANANNAGGGAGAATTCNNNNNC
ANGNNNCTCCATTCTTCNGGNANTCNCCTCTATTGTNTCCCCNACTNTNCTNAGNAGTCTATTTNGA
GGAAANTGGANANNTGNCNTNAGANNGACNAGTGAANGGNAATGNCANNGGTTNGCNTNGGNNTCCN
CNCTGTTNGGGNTTTTTATCGNNNTGCCCGACNTGANNACTTCNTCTGNANGAGCNTCTGCTACAGN
GGNCGCCNNCGCGGGGGAGCTCCAGCTTTTGTNTNCTTNACCGNGGGTTAANTGGGCGCNTTGGNGT
NANNNTGGNGNNAGANNNNTCCNGTGTGAAACNGTTATGCAGCTCACNTTTTACNCACNATNATANG
ANCNNGAAGNATNAAAGNGTCCANNCTGGCGTGCCCCNANGNNTACNGCCCNNCNNCANNCAAAANG
GGNAGGNNCNNNAATGCCNNNNACTNCGNNANGGGNNANCCCNNCANCGACANNCNCCTTTNTGAAN
CTGTCCNNACTTANCCGGNANNCGCCGTCCTNNCANTTNNTNNCCNNCTTCTNGNTNNNTCGTTCNN
TTTATTNTTCTGTACATCGAATNNNTTTNNATTNTNGTCTTAACNTCGAANNCTNTCNNNTCNCAGG
GGGGTCGATTCTCTNTTATTNAGCANTATTNTAAGTNANNAAANGGCNACATATAACCCNGNTTACA
TAAANNGCCCGACNAAACNGCNGNNTCCCTGTNNNAAGNCNGNCNNNGTNGTNNNTTTCCCCTTACN
GNNCCTCCCCACTNANGGGGTNTNCGATTCATCNNTGTNNNGNNNCCAGCATAGNCTGCTCANGNCG
GGCNNTATCCNNNNNCTNNCNGNATTCCCTNTGNNNCGNCCCCCGTTNCTNTNANNCTTCTANNCNA
TTTCCCGNTATCNGTNNNCCTCTCNNTAGGNAGANANNTNTNTTNTGNANTACNNANNNTNNTATNN
ATTACNGTATCNCCTNNNCNNNTGNGTACNNNNCNCCCNNTTTTNCCTNTNNNTTTNAANANAATNN
CTNTGTGCTNCNNANNTNNNTTTTGCNCNGTNTNTANT

>APBMS33 --T7 end. Chimera
GCTTCAGGAACTCGTCCGCGGAGATCTTGCCCTGCATGACTCCGGTGGTGCCGCAGGCGAGCACGAC
CTTGACATCGGTGATGCCGATGAAGCCGAGGATGGCCTTGAGGTAGGGGGTCTCCTGGTTGTAGCCC
TCGATCGGTGAGCCCGGGGCATATTCGCCCCCGCTCGCGACGAGCAGCGTGGCCGTCTTGCCGGTGG
CCAGGCCGTGGTAGCCGGTCTGATCGACCTTGAAGGTGATATTGACGCGCACGATGTGGTCGATCCA

TGCCTTGACGGCGGCGGGAATCGCAAAGTTGTACATCGGGGTGCCGAGAACGATCTCATCGGACTGG
AGCAGGCTTCTGCTAGCAGAAGCACTTCCATGACACATCTCCAGAGTTGCTCAGGCTTGCCGTGAAC
CACGGCAAGGTGCGCGCCGAATGGGCGCGAGAAGGTTAGTACTTCGCCGCGTAGGGCAGGAGCGCGA
TCGCGCGGGCCTGCTTGATGGCGCGCGTCAG

>APBMS35 --T7 end
GGCNCNTGATTCNNNGNCGAAGNCCTNCTNNCCNNNGTNTNNNCGGNCCNNNNCTTCNCCNTCTAGT
NTNCCGANNANTCCCNNCTAGCGCTNTNCCCNNCCCNGNNGGNTGTNNTNTTTTNGGNGNNNGNGNN
NNGAAAAGGGGNAAATNAACCGTTNCANNCTCCTNTTGTNNGNNGGNGNACNCNNNCNNNCGGGNTA
CGGNTNGGNNCNGNTTTNCCTTNCACNCCNNNNCNNNNGANNCCCTCNTNGCTCNANCCCCNNANCN
NNGCCGCCCATCATANNTNNTNTCTCNGTGGCNNTNTNNGNCNNNGCCTTTCCCCCCCCNTGCTTCN
CTCGNNNNGTNNGGNNNAATCCNNANTNNANTGANTACGNNNNGGTNNGNNNTNNAGNCTCCCATCT
NGANCGNAACGGNNANNNCCNGNGGGTNNNNTNTNANATNTCNTTNTGANNNNCTANNNGGANNNGN
NCNTGNCTGNGGCGNNNNGNNNCNNNNTCTACTNTGCNNTNNAANCNNACNNGNNNGACGTGCNCCA
CNNGGNNATNNCTNCNCCANAANTNTNNNGTANNCACTNAATCANTCCNTATANNANATNATNCCNC
TNNCAGACCTNGTTNNNGANTNTTACNCATGTGTGTTTNNNNTNNNNNNTCTNANCNNNNNCTAGNT
TTATANNCTNNCNCTTNNANNNCATNNTCNNCCNTCCGTNNTNCTNTNNAGCATCCTATTNGTTNNT
GNCNNANNGTGTTCNTCNNTNCCNACGNTTNTTNTTNNCNNNTANACAGAGANATANGCNNNANNCN
CNNNGCTNCANNTNTANAGGNNANTCNTNTNCNNNANGNNGNGCGNNCNNNNANTNNTAGNNNTNTN
CTNNNNTNATTNANNNCGANCNANGATTNNAGCNANNGGCAGNNNGNCACTGNCTCNCCCNAACANN
NNNNNNNNANGNANGANNTNNNGGANCNCCCNNNTTNTGGAAANNNGNCGNNGCTCTNGCTTNNNCN
TCACCGNNTCAGAGNGNTNCTTANNTCNCNNNNTGNCNTGCNNGGTNCTTTGTTTNGTNNNACCTNN
NAGTGGNTTCANNNNCTCGCCTCNCCNNCNCTNNCTCNCNAGCTANCCACGCATCCGACACCTNCAN
CCNCCNNGCGCNTCCCTCANGCGNNNGCNCCCCTNNTCTCNNTCCNNCCCNNACACCTACANCCGCC
NNNGCGTCNAANACCCNNCNNTANCNANNNCNGATCNNTNGANNNGNGNTNTCNCANGCNANN

>APBMS38 --T7 end
CNNNGGGNGNNGGNGGGTCGNCNCGANCNNGANNCNNNNNCNNNAGGNTNNCANCTTCTTAATTNCC
CCACNATNNCNTGTGTNTTNTNTNCTCNCTTNNNNNNGGGNGNGCGTANANAGTANCCNANAAGAAN
GNTGNGGGNNNNGNGTNNGNCCNNNNGNNNCTANCNTGTAAATGAANANNNANTNAAANNTTNNTTN
TNATNTGTTCTCTCTCTCNANTTCNGNTTNAANTAGNTNCCGCGGNAGGTNNANACACTNNTANTAG
CGNNNTNCACNNANNNTTNGNAAGANGGCNACNCANNNCCGCNGNTCNGAGCCNCCNNGCTGNGTNN
NNGNAGCNAAGTGGCAGNAGGNGANNNNCTGCNGCCGNTCNANTNNNAGGNNCANNNNTANTTTNNT
NTANANNNATNCTNCGNNAATACGNNGCGGTGNATNGAGGGCGNTNGANNTATNNTNCTNCNAAACA
NTNNTNNCATNNNCTNTGNCNANNNNNTANNAGTNTCCNGTNCNGNNNGCNTCTCTNCNNTCATCNC
ANNNNCCNGTCCNNNGGTTTTGCTNGCNCCTTNAGGGNNGGGGCNTCNGNAGCTGCNACTTGGCATT
NNTGTNNNAAANANTACTNNNNTCCNACAAGNNGGGTNGANATAGNNTAANNTNCATTNAAAANATN
TGCNNNCCNANNTNGNGTAGNCNGGNNNGNNNTTNNANNGGNNNAGGNNNNCGNNNGCNGTCATNAN
AGAAGCTGAANNANNANANNANANNAGGNNNTGCANTTACNNACCANTTGNNNANNTTNAANNANNT
NCNGTNAANCNANGNGGATGTAACNNNNNANNTATNTCTCNNTNNNANNNTGTNNNTGGGNNCGNAA
GGTGTCTNNNNGTTNTNGANTNCGNNNNNGNANANTNANNGTCNNNNTCTNNNGCTGTATCCCNNCT
NANTNNANNCNTCANNNGGNGTNTTGNCNTCTTTNATNTNNNNGNTNTATTTTTGTTTNNAGTTTGN
GGGTGTNTNGGTGTGTNGCNGTATGNNNNTAGTCNCNACAACCTCAATNTCTNNNCCTTGTNCATCC
TTATTTCTNTNTCTNNNTTCTCCTCCTCNCTCNCATCTNGNCNNCNNNTCGCGNTCTCNTCNNNCCN

CNTNNNNCNNNCNNCNNTNTNAANACAGTAACAATACNAANNCGGNNGATCNCGTNTCTANNCTAAT
TCAGACATGGNNCTTNNTANTNTTNGAANNTTNTTCTCGNNTNNNNTCTNTGNNNACNTCTATNATG
ANTNNCAAACNAAANCTCNNANCCTCGTNNNNANTCTAATACNTAGACAANNATANTAGNGCNGANN
CNTNANGGACNNANACNNNNTN

>APBMS43 --T7 end
GTCTANGGTATCNACTAAGNTTGATATCTGAATTCCTGCAGCCCGGGGGATCCACTANTTCTAGCAA
GGCCTTGCTAGAGCGGNCGCCACCGCGGTGGAGCTCCNGCTTTTGTTCCCTTTAGTGAGGGTTAATT
GCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC
ACAANATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGCCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTANTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGC
NTGCGNTCGGTCGTTCGGNTGCNGCGAGCGGTTATCANCTCACTCAAAGGCGGGTAATACGGTNTAT
CCANAGAATCAGGNTGANAACGCANGANANANCATGNNGAGNAAAAGGCCAGCCAAAAGGCCNGGCA
ACCNTNAANAAGGGCCGCNGTTGNTGGCGNTTTNTCCATTAGGATCANCCCCCCCTGANGANGCATC
NCAAAANNTCGACGCGTCNAANTCAAGAGGNNGTCGANACCCCGACNAGGACTATAAAGANACCATG
GCNTNTTNCCCCTNNCAAGNCTCCCTCCGTNCGGCNTTNCTGGTNNCGACCCCTGGACCGCTTAACC
GGNNNCCTGTTCCGNCCTTTTACCACNTTCNGGNAAANGTGANNNNGCTTTTNNCNCANCGCCANGC
TGCANAGTAANTCNATATCNNNGGCNNGCTACGTTCCNCTTCCTANCNTGGGGNNNTTGANNCCNCA
NNNCCTCCNCATTNNNGNCCNGNNGGCTNGGGCNNNATCNAGGTAANTNTGTCTCTTGCGNTCAACC
GNNAAATAACNTATTNTNNCACNTNGAGAGTNACCNNNGGATAANGNTCCCAACANNGNAAGGCNNC
CGGANCCCCCGNNCTTGAGNNGGGCCNTANCNATNTNNNNAAAAANATNNTTTGTANTTCNNCCCGG
GACAGCANCACNGAANAGAGTGGCGTCNCCNGAAAAACCACNGNNGGGTGNTTNTTTNCACANNTTC
CNTAAAAGNNTANNCTNTTTANTTNAGGGNNGNNGGGAGNCCNCANGNGTGGNCNANATACNCNCNC

>APBMS45 --T7 end
NACNGANGACNGNTATCGATAANGCCNNGATATCNAATTCCTGCNTNCCNGGNGATCCACTAGTTCT
AGCNGAAGCAATTCACATNNCACAGANTTTNNNANACNCACGCGAGAGAGAGNGAGAGAGAGAGAGN
GAGAGAGAGACNGACAGCGACNNNGAGAGACAGAGAGAGACNATNAGAAATGAGAGAGGANCAGCGC
GGTNGCGCAAAGGAGCAACAACNCTCTTTCAAATCATTTGAACNGCCNTCCGCTAGAGCGGCCGCCA
CCGCGGNGGAGCTNCAGCTTTTGTTCCCTTTANTGAAGGTTAATTGCGCGCTTGGCGCAATCATGGN
CATAGCTGTTTCCTGTGNGANATTGTTATNCGCTCANATTCCACACAACATACGAGCCGGAAGCATA
AAGTGTNAAGCCTGGGGTGCCCTAANGAGGTGNGCTAACTCNNATTAATTGCGNNTGCGNTCACNTG
NCNCGNTTTTCNCAGANCGNGANAAAACTTGNTCGTGCNCAGCTGCCANTTAATTGNAATTGNNNCN
ANGNAGCGGGNNNGAGNCGGGTGTNGNCGCTAATTGCCGCCGCNCTTTCCGCTANNCNCNCCNNAAT
NAACTCNCATGAGNNTCGNNNCNCTACNTCTNCGNNNNCNGNCGGNCATCTNTNTCCCGTCNAANGG
AGGGTNATNACNTGNTTTTCCCCGGNAGNNNTGGTNTATAACGNNCNNGNAAACANACTNTTGTNGN
CNNACAGNCCNGCCNAANGGGCCAGGGAACNNNTACAAAGGGTNGGNTGTATTTGNGGNTTTTTTCT
NNTNGNNANNCGNCCCCACCGNNACGACNTNCCNAANATCGTCGTNTCAAATGNAANGNNGTGNANA
CCCGNACGGGAATATTAAGNATNNNAGCGGGNTTCNCCCNGGTAGTCACTNCTGCCNNCTCCCNGTT
TTCGNCCCNAGCGATTAANTGNANTNANNCTNNNTTTATNCCTTNNGGNGAANGCANCATTTCTNTA
TNTTCCTTGTTNCNNTATCTNTATTCGNGNTCCTTNCNTCNCATNCGGGGTGGGNNANNAACCCCNC
NTNNACATNNCCTCNCAGNCTNTANNNNTATTTCGANGNNNGACCNNGTAAGAATTCCCCTANCCTN
NGTNNGNANACANANCNACACTCAGCGCTNGCCTCTCTTTNGGGGNNNCACCCTCCANATANATCGN
ATGNTGNNNG

>APBMS59 --T7
GCGAGNGGGNGNGAGGNNANCGGNNCTNNAGNGTGAANCTAGNNAAAANNAGACTNNNTGNTATGGA
AGNNNNGNGTGANNGNANACACCNTGCACTNANNTCTGTGCNTTGGGNANGCNNTCGCNNTTNNNTT
AANANTACCTAATTGTGNGTNNNNNNNNNNNNNGNAGANNGACNNNGAATNANGNGTNTTGTGGNCC
NTGCGTNTNNNTNNNAATTNNGCTCNACCTNNTNNGTNGTNTNGAANCAAATAACNTNANNTNNGNN
NTNNACTCTNNTCNANGCNNACNNCTACNNCCGGTCNNCTNNNCAGNNTCCGANANTNATACNCNNN
ACNNGNNGTGTNGTNNTNTGTGNATTATTTGNCNANATACACNGNCCTNNTGGTNGAACAAATNAAN
NNTNNTCNCNTCTCANNNTNNGCGNGTGNNNTTTTANGTNTNCTTNCNNTTGTTATNNCAATAATAN
NTAANGAAACNNANCNNAACTNNCATATTGNGNNATAACNGGNGATNANATANNANGNNNGANNNCN
AGGTNCTTNANCAGNGGNGCNNGCGTNNNGCANNNGTGNNCGNCANNNANNCCCAGCNTCTNNGTNC
GTTTCGGNNAGNTACNATGNGCGNNGAACNCANTNTANANANANNNNANTANNNTCTAANNNNNNNAA
CAGGNNGANNGNANGNCGGNTGNNNNGACNNNNGNANGNNTANCGNGCANNTCNTCANTNNNCNNTN
NNAGGNTCGNCANCACNGATGNNNNTNATNTAANCCNTNNCCTNANNATNNCTNTNNNTNTCTGTGG
ACANGGNNNANACTANNNNNTCAGNNATGCAATNNNTGNCTNCGCCCTGGNNTNCNTGTCCNCNCTN
TCTGTNATTCTGTAGGGTNCTCTNGCTATGNNGNNNGCNNCGNGNNTGGGNNCGAGNTGNNCTNTNN
ACNNNGNGCNNNCGTCTGCTNGNGTNNTNNNCNNAGNNTGTNNNAGCTGANCTCNNGNTNTCGGNCA
TCGCNTNNACGTANCTTCGNCNCNNATNCNGNNANCTTTNNNGCGANTNTCCGNATCTNNNGNNAGN
GTCNNNCGCNNGNCNATGNNGGAGNGTCGCNNNTGGCGTNCNCGGAGNNCANNAGNANNGANNNTGN
NNATGCNTNNTCNCGGCGNNNGGGCTTTCNNGCNTNCNNNNNNGTCTCNTAGNGCNNCANNNNNNCC
ANNNGNGNCCANTCNANCNNNNGNACANANCTNACCATACNNNACNCNCACNNNNCNNGNNGNCGCN
CCTCNCNCANNTCNTNNNGCNCGNNNGAANGANNACNNNTAACCNNAATNTNTANAGTGGANCAGNA
ANNATACNANAGAGANACCGANTTTTNANATNNNGNTCNCANTGATGNNTTNANGTNNTGNAGTNNN
TNNCNNNNTGTANNTNNGTGCNNGNTCTAGTNTTNTTNAAANNAGACNAGTCGNTANNTNAGGAGGN
NANNNNNANAGNNACNGTNGTANNGTANTAACGTCGAGTNCANAGNGAGATTNGNANATNGNGNNNT
GNACNTGNCNNGNACTNCNATNNNTTACTNATNNCTCTNNATACACNTGGTTTNTTGTCACGCTNCT
NAATCANACCACNTGTTANTNNNNTNNTGCNTGTACGCNTNATATATNNTCNNNCTGTCNTNNTGNT
CCNCNNTCNNCTNCNNNNCNNGNGNGTNGGGAGNGGGCGCTCGNGGAGNTTNNTNCTNANACCNGCN
GCTNNAGATGTTNNCNCTCNNGCCGNGGNTCNNNNCTCNNGNGNNTTNTCGGCCCNTTTTCACACCN
CGCNNTCNANNTNNGN

>APBMS67 --T3 end. Chimera
CTAAATACTTCTTGTAGTATTGGCCCCTTTATATTTTATAGAGTGTGGTCTATGAAATGAATAGGTT
CAGTTCTTTGACCTCTCTTGACATGCATGCTTAATTGAGGGTTTTCCTATCACAAATTATGTATCCT
TACATTTAGTTTTGGAAATAAAGATATAGAGAGAGAGAGAGAAAGAGAGAGAGAGACTTCACTTGTC
AAAGCAACATCTTGTTCACATCAATATGTTCCATGCTATAGGCCCTTGCAATGCCCAATTTGAAATG
CTTCTGCTAGCAGAAGCGTTTCTTTTTTCCTTAATTTTTTTTATTCTTCAAAATGATTAATCTGCAG
GCAATCTCTTGCGTTTGCTTGATCTTCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCCGTGTGNGGGCCAAGTGTGGANNGCTCNGGGGTGCGNGCTAAATCATAGTGTG
TGTCTCGCGCGNGGCCCGCGCCNCGCCCCNTCANTTCTGTGNNTGAGCTGTGTAGNATCTCAGNGGN
GNTGNTGTTGNGANGCGTTTGNTCATCNGCGNGAGCGAGNGCGTTNGNGCNGGAGCGCCGNGTGCTG
TGCGCTAGCTGCGGGNGNAACTATATGGANCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGA
TACCGNCGACCTCGAGGGGGGGNCCCGGGTACCAATTCGCCCTATAGNGNGNCGNATTACGCGCGCN
CCCTGGCCGNCGTTTTACAACGCCGNGACTGGGAAAACCCTGGCGGTTCCCCACTTAATCGCCTNGN
AGNNCATCCCCTTTCNCCAGNNGGGGNAANGCGAAAAGCCCNCNCGATCGCCCTNCCNANAGTGGCC

CNCCNGAAGGGNNAAGGAAANTGNAGCGNNAATTTNNTAAATCNCGNAAAATTTGNTAANCCCCNNT
TTTTACCAANGNCNAANCGNAAAACCCTTAAAAAAAAAAACCCAAAGGGNGNGGNNNCCNTTGANNA
GCCCTTNAAAGNGGCCCCCCAGGGNANCNNTNGGGGGGGCCCNNGNACCCCAAAGTTTGGGGGGGCN
NAANAANGNCNNGGNCCCTNGNGGGAACGNNCCNAAAAAAA

>APBMS76 --T7 end
AGTNCTCCGGAAAAAAAAANNTTGNTNTCGAATTCCTGCNGCCCGNGNGATCCACTAGNTCNAGCAA
GGCCTTGNTAGAGCGGCCGCCACCGCGGTGGAGCTCCNGCTTTTGTTCCCTTTAGTGAGGGTTAATT
GCGCGCTTGGNGTAATCATGGTNATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC
ACAACATACGAGCCGGAAGCATAANGTGTAAAGCCTGGGGTGCCTAANGAGTGAGCTAACTCACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGNAAACCTGTCGTGNCAGCTGCNTTAATGAATC
GGNCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGNGCTCTTCCGCTTNCTCGCTCANTGACTNGN
TGCGCTCGNTCGTTCGGNTGCGGCGAGCGGTATCAGCATCACTCAAAGGCGGNTAATGCGGCNNATC
CGCNGNAATCCGGGGNNTAACNCAAGAAAATAACATGTGTAGCANNANGNTCANCNNAANGTCCNNG
TAACCNNTAAANAGGGCCNCGGTNGCCNGGCGTTATTNTCCATAANGCTNCNNNCCNCCTGANAANN
NACANAAAAAATCNNNACATCAAANAAAAANGNGGGAGNANACCCGACNGGATTANNAAGGANANCA
GGCGTNTCCCNCCATGGAANCCTNNCCTCAAGGCGCTNTTCNTGTTCACGATTCCTGCCNNNTTCAC
NAGANANACNGGTCCGCCCTNCNCTCCTTCNNGNANNCGGNCNCNNTTCCNATAANTNNCNGNCGNN
ANANNNAGANCCGAATGAANNNGCNNANCCCTANNCGGNGCTNGTGTTNANAACACCCNNTTTNCNC
GANNGNNNATCTTNANNGNAAANACCTNTTGNNNCNCNTCCNGNNNANAGNNNCNTNNCTCCNGNNG
NANCGATGGNAACAAGTNTAGNAGNCCTAGNATNATNGCGNAGNCNTAAAANATTTCAGNNCNCNCN
AACAGNANANAATNNAAAAGNNNANGNGATNAGGCTNGGNNNNTCGNTCTACTCCATAAAAGNNCGG
NANATNTTCAGNACANAACACCNNCTCGCACGGNCTCTGT

>APBMS79 --T7 end
NTNATNNNGTTNTNGTNTNNANNNNACNNANGNNNTNTNNNCNCNCNCCNCNGGGGGCCNNNTACAA
CNNACGTNCCATTCNCNGCNNTNNATCTACCNNCTATTNTNTGNCNTNATGNGGNTTGNAAGAANGG
NGTTNNCNTTNNTTCTTTNAAANAANTNNTCNNGNNNANGNTNNTATNNATCTANNTAATCNCCCGN
NGGANANNNNGGAANCNCCCTCNTCTACTTANAGNNTTAANTNNNNCNGATNNNGNNTGGNTNNNTT
NTGCANNTANTNTCATNTANNTTTNNNNNNNNNNTTTNNATGCATNTNNATGNGANNNTCNGNNGNA
CNGNTGGANANNTGNGACNCANGNGTNTNAGGANGTNNNGGGAGTNGTTNGTTNGNGNTNNCNNTCN
ANAANACCNGAGGGANTCCGTCNCGNAGCCGGCNCGGCCNNTNGANGNNGCNNNNNNCCNNGNNNNA
NNCACCNNNNNNANTNNNAGNATNAACNAGNCTNGNNTGGTGCCNNGTCANCGNNTTCCANNNNGGN
NNNCANGGNNNGNAAANGGCNANGANCNTCCNNCTTTCNNNNNGGTCGNNCNNNGNAANNAGNNANA
AANNAGGAANNNNCTNGNAAGNNCNNTNNGNCNNGAAGNTTTNTCTCTNTTNNTTNTNTATAANTAT
ATTATATNCANNTTTANNTANTNNNATANNNCTTCNTNTGCNTTTTTTTTTTNNNATNNTCCCNCCN
NNNCCGGNGGCGAACCANANTCGAGGANAGANAGAANNGGAGNNANGNGAANATNANGANGNNGNTG
NCTNAAANTTTNNNCTTTTCTCTNNNANNGNNNGNACNGNTANATANTTCNTCGCCCCNTCNCNCNA
NTANNNNNNNNGCTNNNGAANAAGNGNNTGNNNGAGGTANNNAANTGGCNGGATNTNNNTNANNCCT
CTATTNNTACCNNCTCCNNNNNNTNNCCCNCCNNCCNNNNCCGNNANTCTGGNTNCNNCNACTNTNN
NNNCNGNNNCNNGCANGTTNCTNCNNTNTCNNCATGTCANCANNCCNNGCGCCNNTGATNNTNAANT
NCNNCCTNTNTNNCNANCCTTCCTNCNNANNCNTNGNNTNNNTTNNTAGCGNNCTCNNNNTNCAGNC
ACTCNTNTNCNNCTNATCNNTCCNTNCN

>APBMS80 --T7 end

CCNNCNNNCACNAAGGTAANCCTNNNNTNNCCNNANCCTNNNNNGCCCGNGCNNGGGCCNANGCCCA
NAGNAATTCNNANNANNNAANGNTTGNCGNNTNCNAACNNCCCTCCCNNGGCNNCNAGAGANAANGG
ANTTNANGGNNGCCTCCNTNNNGCTGNNTNNGNGCCNNTATTACTNNCNACNNNNNCACNTNANCCN
NGCCANCNCGAANACCANNNTAGCAANNACNCGAATGNNNNNANCNNGCNANTTCCCNAAGNAAGNA
TATNTANANTTNATNNNNTTNNTATTNAANNNATNTCNGNNGCNTNCTGGGGNGCAGNAGANANNGC
CCANCGCGTNGGNTNGTANCGNACNGCTTTATANGANTCCCNNTTNAGAACAAGAANNNNCNNGAGN
NGACTNGGNNGNNANANAAGGNNNCNATANATNGNNNNCNNNGGNTAGAAAACNNNTTANCAGNGNC
TNTNCNTNTNTNAGNNTCTTGTTGNNCCTGGNTGNNATGNNTGNTGNNNNNGNNCTTNGNNNNGTTC
NANANGNTNNGNNCNCTCCCTCACCTNNCTCTTTNGNTNTNTNCNGNTTNANNTGNCNNNCTTTNNC
AANCNCNGCGNANCACNCGNNGTCNANATNTCNTNNNNNNGNNACTNCCCANACNCNGNNGNNGTAA
NGGNNNNNANNNNNATTAGNGNCGNNCTTNCGNTCTTNNTAGCNTNACCNNNNNNCCGNGCCNNNNN
NCNNNNNGGANNNTNNNNNNNGNNNNNACNNTNTTCACACGGTNGTNTAACNNGNANCTTNATNATC
TNTNACCGNNNNACTCACNCNACANNNNNGNGNANNNANNNGGGAGNTNGTGNGCNTGGNTTCNGNN
NNCNNANGNGGNANATGGTGNTTNANGANNTNNCGNANGNCNCNAANNNNTATANATATNTTATNGA
TAAATNNNAATNNCNNTANCNTNGTNNGTNTCTNTACACNNGNNNCAGNTNTTCCATCNANANNTCN
NGNNAGCNNCNCNNATCGAGNNCTNNNCTTNNNNTNNGTNNTNNNTNCNNTNNCAANTNATNCTGNA
TTNNCTCTTNANNGCTTGNGANCNTNNATCAAGNNNANNNTTTGNTAANTNNNGTTNNNTNAGTNAT
TCGANCCNCCCCNNNNNNNNNNNTNNNNNNGAAAANCTNNNNNNNNNNAACNNNNNNNNNNCNCCCN
CNCCNAACGTTATANANANNNATGTATGTNTTNTNTATNNNNGTNTTTA

Appendix B

REPLICATED SEQUENCES

The following are sequences of inserts that were found more than once in either the APMS or APBMS library. They are FASTA-formatted and the inserts of which they are duplicates are indicated. If a sequence is derived from only one primer, the primer used to generate the sequence is indicated. Otherwise, the presented sequence is a consensus of the assembled forward and reverse contigs. Where possible, the sequences are trimmed of both linker and cloning vector sequences.

>APMS25 --Duplicate of APMS23
CTGCACACCTGCTGTGTCTTCTTTGGTATGTTGCTTGTGTCTGGAGGCAGTCGAGCAGTCATTCTCT
CTTTCCCCCTTTCTCCCTCTGTCATACACGCACACACACTCTCTCTCTGTCTCTCTCTCCCACACAC
ACACACACACACACACACACACACACACACACACACGCATGTGCATCCGCACCCGCTTTAGCTGGTTGAGT
GTAAGTATCGGTTTTTGCGTTCCAATGTGTGTTGCGGAGGTGATCCCATTTTTGCTTCCA

>APMS26 --Duplicate of APMS24
CTGCAGCTGCAGTATGGGATGAAGATGATGACGGTGATGCTGTGGTCACACACACACACACACACAC
ACACACACACACACACACACACACACACACACACACAAACACACACACACACACGCACACACACACACACG
CACACACACACACACACACACACACACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTTTCTCTCTCTCTGCTTTTCCACTTTCACACTCTCCTTCCACTTTTACCACTT
TCTTAATCGTGTGCCCTTTCAGTGCACTTGNTGCCTC

>APMS32 --Duplicate of APMS2
ATTCCATCGTGACTGGTGGGAGAGAGAGAAAAATGTTAGCATCTTCACATGTACAACCCATTTGCAG
TCAGCGAGTAAAGTTTTTCCATTTTGTGCACACGCACACACATACATACACACACACACACACACAC
ACACACACACACACACACACACACACACACGCACGCTCACACACACGTATATTCGAAATAACCTTTATA
GATCCTGTTGCCGTTGCGTCAAGAATGCAGTCGGCGATGTCCTTGAGATCCTGGAAAGACGATCTCA
TCACATGTAAG

>APMS36 --Duplicate of APMS35
AAGCGGAAGCCGCTGCGAGGTCTGTGTATCCCAGTCGTGTGACTGATGTGTGTGAGGGATCGTGTGG
TATGCCATCACCTGTTGTGGCGGTTGTATGTTGGGTTGTGTGTGTGTGTGTGTGTGTTCAGGGAT
GAGGCAGAGTGTGAGCGTCGTCGCGCCGCTGCTGCTGCGGTGCAACAGCCGGACGCGAAGGCTGCGT
TGGCGGCCCTTGCGGCTGACAACAGACGCCGTGCAG

>APMS39 --Duplicate of APMS35
CTGCACGGCGTCTGTTGTCAGCCGCAAGGGCCGCCAACGCAGCCTTCGCGTCCGGCTGTTGCACCGC
AGCAGCAGCGGCGCGACGACGCTCACACTCTGCCTCATCCCTGAACACACACACATACACACACACA
CACAACCCAACACACAACCGCCACAACAGGTGATGGCATACCACACGATCCCTCACACACATCAGTC
ACACGACTGGGATACACAGACCTCGCAGCGGCTTCCGCTT

>APMS41 --Duplicate of APMS14
GCAGCAATATTAACCTCGCTCGACGAGTAACAAATATATACATAAGTTAAAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGAGAGAGAAGGATCTCGAGCTGAGGGCTGGACGGATTCGAAGTGAATTTTTCTGTTGTCCCAGGCC
TTTCAG

>APMS52 --Duplicate of APMS47
TGGAAGCAAAAATGGGATCACCTCCGCAACACACATTGGAACGCAAAAACCGATACTTACACTCAAC
CAGCTAAAGCGGGTGCGGATGCACATGCGTGTGTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGGGAGAGAGAGACAGAGAGAGAGTGTGTGTGCGTGTGTGACAGAGGGAGAAAGGGGGAAAGAGGGA
ATGACTGCTCGACTGCCTCCAGAAACAAGCAACATACCAAAGAAGACACAGCAGGTGTGCAGAGGTG
TGTAG

>APMS58 --T3 end. Duplicate of APMS47
GGANTGTTGCTTGTTTCTGGAGGCAGTCGAGCAGTCATTCTCTCTTTCCCCCTTTCTCCCTCTGTCA
CACACGCACACACACTCTCTCTCTGTCTCTCTCTCCCACACACACACACACACACACACACACACAC
ACACACACGCATGTGCATCCGCACCCGCTTTAGCTGGTTGAGTGTAAGTATCGGTTTTTGCGTTCCA
ATGTGTGTTGTGGAGGTGATCCCATTTTT

>APMS62 --Duplicate of APMS14
TGCAGCAATATTAACCTCGCTCGACGAGTAACAAATATATACATAAGTTAAAGAGAGAGAGAGAGAG
AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAAGGATCTCGAGCTGAGGGCTGGACGGATTCGAAGTGAATTTTTCTGTTGTCCCAGGC
CTTTCAG

>APBMS75 --T3 end. Duplicate of APBMS29
CTCAGTTTGCTGTCGTGCAGTAGAAACAGGTGGGGATGTTGATTGATGGGTGGAATTTTTCAGAGGT
ACTGCTTGTTACACGAGTGATTCACAAGGTCATCTGGTGGTTTTGAATGCTGTGGGCTGTCTTACAG
ATTCCAGAAAAACCGAATACTTCGCTCATTTCCATTCAAATTAAACACGCTGGGGACTCGGGTTCTG
TGGAAACTCGGGTTCAATAGAGTTAGTTGGTGTGGTTCAGTTTTTAGTGTTGTTGTATGAGGGAGTG
GCATGGCAAGAACTGCTTCTGCTAGCAGAAGCGTTTCGGTAAAGGTCCCTTGCTTCCAAGATGGCTA
GACTTCCTACTTTGGGACAGTCCTTTTCCACCCCTTGTGCTGCCTTCCTTGGCACCCCTGGTGCTGC
CGTCTTTGTCTGCTTCGGTGCTCCCATCCCTAGTGCCCTTGGTACTTCCTTCCTTGTCTCCCTTAAG
AATTCCATGCTGACTCGACCTAATATTGCCATCTTTAGTAAATCCGTCATCCCCCTGGGAGTATAGC
CCATTTCCGAAATTCCCATTCTCAAACTGCCCACCGGAGAAATCTGCAGAAAATTCTCAAACTCTTA
AAACCATTTGAACGGGAAATCTGCACATACATCAACGCAGCAAGTGCACATAATAACATTGGAAGAC
TTGATTTGAATAAGTAGCATACCAAAATTCCTAAAATGATGGATGATCAAGGACTTCCCACACACGA
ACCACACGGGTCCTGGCCAGCACTGGCGANATATCGTCCGTCAGGACTAAATTNCATGGNCCNAATT
GCCCCTTGATGAGNCTCTGCTAAANTAGNGGATCCCCGGGCTGCAGAATCNANATCANCTTACGATC
CGNCGACCCNAGGGGGGCCGGTACCATNCCCCTNANGAGCGNATACCGGCCCTGGCGCGTTNNACGC
NGNNGGAAACTGGGTACCATNANCNCTGAAAANCCCTTCCANGGGAAANAANNCCCCCACCCCCCAA
NTGCNCNGAGGAAGANTNANNTAATTGNAACCGNAATTGAANNCTTTTACAGGCAANGAACNTAAAA
AAACNA

>APBMS104 --Duplicate of APBMS102
CTTCTTCAGCCTCCCGTTGCTTCGCAGCATCTAGAAAGCGTGCTTTCTTCTCCACGTTCATTTCATC
TTCACTATCCTCATCACTCTCATCTTCCTCGTCGCTGCCATCATCCTCCTCCATTTCGATTTCACCA
TCTCCCTCGTCGTCTTCGTCTTCCTCTCCCTCCTCCTCCTCATCTTCCTCTCCCTCCTCCTCCTCAT
CTTCGTCCTCGTCGCTTTCAATCAACTTCGCCTTCAACCACTTCTTATTGTCGTCAGTGAAACCCTT
GTTCTTGGTTGACACTTTTGTCTCCAGGAATTCATCACTCTCCTCCTTCTTGGCCTTACGCTGGACA
ACATTTTGCAATGACAGGGTTTAGGGTTTAAGGATTGCCACATACAGAACCAATGAAAACACAATAC
ATAAGAACTAGCCTAAATTGCAGGACTATCTTGTCCAAAAATTCAGACACTACAGTGAAAACCCTAA
AATTGCAACTAGTCCAGAAGCGCATCCCCATATCTGAAAT

Appendix C

SEQUENCES WITHOUT MICROSATELLITES

The following are sequences from both APMS and APBMS libraries that were found to be without microsatellite sequences. They are in FASTA format and, where possible, trimmed of both linker and cloning vector sequences. If a sequence is derived from only one primer, the primer used to generate the sequence is indicated.

>APMS57 --T3 end only
GGCGGCCTCTCTAGCAAGGCCTTAGCTCATTTTTTTNACCAATAGGCCGAAATCGGCAAAATCCCTT
ATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATT
AAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAA
CCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGA
GCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAA
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCG
CTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT

>APMS61
CACTGCGAGGTCTGTGTATCCCAGTCGTGTGACTGATGTGTGTGAGGGATCGTGTGGTATGCCATCA
CCTGTTGTGGCGGTTGTGTGTTGGGTTGTGTGTGTGTGTATGTGTGTGTGTTCAGGGATGAGGCAGA
GTGTGAGCGTCGTCGCGCCGCTGCTGCTGCGGTGCAACAGCCGGACGCGAAGGCTGCGTTGGCGGCC
CTTGCGGCTGACAACAGACGCCGTGCAG

>APMS68
CGCTGCGAGGTCTGTGTATCCCAGTCGTGTGACTGATGTGTGTGAGGGATCGTGTGGTATGCCATCA
CCTGTTGTGGCGGTTGTATGTTGGGTTGTGTGTGTGTGTGTATGTGTGTGTGTTCAGGGATGAGGCA
GAGTGTGAGCGTCGTCGCGCCGCTGCTGCTGCGGTGCAACAGCCGGACGCGAAGGCTGCGTTGGCGG
CCCTTGCGGCTGACAACAGACGCCGTGCAG

>APBMS1
CTCGGCGGCACGCGCAGCCGCACGTCGAGGCCCTTGGCATTGACCGTCTTGACGTCCCACGTCCACC
GCGCGCGCCCATCCGACGACGTGCCGGCGACCTGCGCGAACCCGGTCATGCTGCGGACACGACCGGA
CGACACGCCGTCAGTGGACAAGGCTGGCCGGGATCGTCTTGGCGGTGTTCAGATCGTAGGTCTTGTT
GCGAAGCGGATCGAGCGGGGTGTTCTCCGACGCATCGAAGA

>APBMS5
GATTCTCTCGTTAGCGGAGGGATTTGAATGAGAGGGCTTAAGGTTGTGATGCTTTTATCGATGTTTC
TGTTTATATTGTGGAGTCAGGATGCTCGCGAAGCGCAGGAGTGTTGATCCAAGGTCGTTCTGTGTGC
GGAGAAGACTTACCCACGACGCAGGGCCGAAGTGATTCGCAAGAAG

>APBMS7
CTGGATCAGCGACGGCCGAGTCCGCCACCGCGAACAGCCGCGGGTCATCTCCGGCGGGGTCGATGTA
GAGGGACTCTTCCAACGCGACGTGGTACTTCGGGCGGACGCGCGGAACCAGTGCCGACCGCAGCGCG
ACCAGGAACGAGGTGTGGAAGTCCCGCCAGACGTGGGACTGCTCCAGGTACGGGTTCATACCGGGGA
ACGGCGACGGCATCTTGGGCCTCCTGCAGCCATTGTACCCTCACACGCCCTTCGCCGCGTCGCGGAC
CTCGCCCTTGGGCAGGTTCGGCGTCGCTTTCAGGAGGCGCATCATGTTCGCCTTGTACGCCTCCACT
CCCGGTTGACCGTAGGGGTTCATGCCGCCCAACCGGCCCTCCACCACCGTCGCCAGCATCAGCATCT
GGAGCAGTTGGCCGATGGTGTGCTCGGAGATCATCGGCAGCACGATGTCGGCCGTCGGCCGGGCGAC
GTCCGCGTAGGCCTCGTTCGCGCCGCGGAACGCCGCGTCCAGCAGGTCCGGTATCCCCTTGCGGCTG
AACT

>APBMS10
GATTCGCGACCGCGCTCGCAAGCGCGTGGAAGAACTGCACCAGCGCGGCGTGCCGAAAGCGTATCTT
TATGGCGATACCGCGACGGATACGTATTCTGAAATGCATTCCTTTTATCTGCTGGTCGATGAGCCGT
CGGTGTACGGTCTGCCGAATACTCCATTCAACCCTTG

>APBMS11
GTTTCGAAATTAGGGCAAGCCGTGCAAAACACGCTGACCTTACATTTTTCCGCACACAATGCCCACG
ACATCCAGCGAATAACTCGCGAACGAGATAGCCTCTCGCACAAGTGCAGTGCGCTCTAGGTAGAATT
CGACAAGCTGCGGGAACAGACCTCTGTCTCGGTCCAACGTCAGTCGAACATGATTGCAGAAATGGGA
GACCAACTTCGAAAGCTGCACGAAGCCCTCGCC

>APBMS13
CTGCGGATGATGATCAAGGAGAAGCGATCTAAAGTTGAAAGAGAAAAAGAAAAGGAGAAAGAAAAAG
AGAAAGAAAAGGAGAGGGAGAAGGAACGGGAGAAAGAACGTGAAAGAGAGCGAGAAAGGGAGAGAGA
GAGAGAGAGCGTCAAGAACGTGAGAAGGAGAGGGAGCGAGAAAAAGAGCGGGTGGAACGCGAGAGAG
ATCGAGAGAGAGAGAAGGAAAAGGAAAGAGAGAAAGAAAGAGAGAGAGACCGATCTGAGCGAGAGAA
GGAGCGCGATGAGAAAATCAAAGTCAAAGTAGAAGAGCCTCGAGGAGTGCCAAGTGGAGGTAAGTGT
TTGCTAATCACCTTGCGAACTGTGCCTCGGTGTTGGTAGTTCGATGTTTTGAGTCGCACAATTCAAT
CAGGTGTGTTCTGGTTGTAAGAAGAGCCGGCCGTTTTCTCTTGGTAATTCTTTTTCTAGAACTTCGC
GGCATTGATTTTGAGCGGTATTTGTGCCAATACCCGCGGGTGCAATTTGTTCATGCTTTTTGTTTTT
CCTAAGTGATTATCGGGCCGGAACAACTTATAGTATTCCGGAAAG

>APBMS15
CTCAGAAGACCAAGAATGCTCTGGATCAATGGGCAAAGACCAACAAGGTCACTGGCTGGAATTCCCG
AATCATAGGGAAGGAGGGACCAACCAATGTCCACTAAAGTATTGATTGAAGACGGAGCCGTGATGAC
GCGCTTCAAGCAAAAGAAGTTGCTGGAGAACTATCGAGACAAAGGCATCGTGATGTTGCCGGAAGGT
CTCTATGTGAAATCGATTCAAACACTTCCGGACTCCGAGCAAACGACCGTGATCGTGATTGGAAGGC
TCGACCAATGAACAGGGAACTGATAATAAATCTGCCATCACGTGAAATCGACCTGAGCCAATCTCCT
AGTACTCCTGATCTGTATGGACGCGCCGAAACCGGCAGGGCCGATGCCCATATCGAGAAATTTCGCA
GGACTTTGGAAGAGGTCATTCACGAGTTGACCTCAGCGACTCAGACCGCAATGGACGGTCAAACGAG
CCATTTCAACGAGCGCATCAAACTATTCTCGGAAAC

>APBMS18

CTCGACTGGAGAGACTTTTTGCATCACAACTCTCCATATTTGCCAACATACAGTCGTCCTCGTCTCA
TGAGTTCGCCAGAGTTGGAGTCTAACTCTTCTCATTCAGATGACGAACAAGCTCAAAATTTGGGCAT
CAATGAGTTTGGATCCAACGACGACCTCGTCCGCAATATGGGCAATGATCCAGATTCAGAACCACTC
AGAGTTCAAGAAGGAAGTCCCCCTGTAGATGCTGATCCACCGAGGGAAGA

>APBMS20
GGTTCCGGCCCGAGGCGCCGCTGTTCCTGGCGGCGTACGGGGCGCTCCAGGGTTCCGACGCGATCGT
CCACTTCACGCTCGACGGCGCCGAATGGTCGGTGAAGCCGGGATACTTCATGCAGCCCTGGACGCTG
ATGTCGCCCGCGATGATGGGGCAGTTTCCCGCCGCGGCGCGGATCTATCGCCAGGGCCTGATCGCCG
AGGGCGATCTCCTGGTCGACCTCGACCTGAAG

>APBMS21
CTTCCCATCAGCACCGGCCGGATGATGCCGAGCACCAGCATCAGCCAGATGGTCGACAGAAACGCCC
AGGCGGTGTGCGTGCCGAGGCCGAGCGCGCGGGCGCGCCGGTACATGCGCAGCCACCAGAGCAGCAC
GGATGAGGTCAAGAAGAATCCGGCCATCAGCCACCACCCGCCGTCCTGCAGGCGCGGAAAATGCAGG
CCGTTCGACGGGGCCGGCGGCTCGAGCGCGAGCCACGGCAATTGGCGCACGAACTGAATTGGGTTCC
AGTGCACCGAGGCGAACA

>APBMS22
CTCCTGGACGTCGAGCAGGCGGCAGGGCACCTTGTTGATGGCGTACTCCGAGTCGCCGGAGCGGAAC
AGGGTCCGAGTTACGGTTACCTCGGAGAACTCGATGGGCAGCACACCCGAGGAGTTGTCGATGGTAA
GGGACACCTCGGCCCGGCCGAGGGCAGCCCGTTTCGGAGTCCCTGCGAAGATGACGTCCTCCATCTT
GGCCGAGCGCAGCGTCCTCGGCCCCTGCGCGCCGAGGACCCACGCCACCGCGTCGACGATGTTCGAC
TTCCCGCTGCCGTTCGGTCCGACCACGACCGTCACGCCCGGCTCGAACTCGAGGGTGGTGGTGTCCG
CGAACGACTTGAAAC

>APBMS24
TCTTCTCGCGCGCTTCTTTCAATTCTCTCAACCCGACCTCTCGATCATCCACCTCACATTTTCTTTT
CTCTCTCTTATGTCAATCGTGATGCGAACGCCATCCAGTTCTTCAATTTCCAAAGCGAGTCATATCA
CTCTCTGTGTTTGTTTTCCTACTGAATTGCAACCTCTGTCCATTGCATCAAAAATTTGTATCGGAGA
TGACACTGTAACCACTCATGAATTCCGTACAGTTGTTTTCGCAATGCTCATCGACAGTTATAAATGA
GTTTGCAGTTGTCTATCGTTCCCTTTCGTCTCCAATGCCGACGCCAAATCAAAACTGCTACTGTAGA
GTTGGGGCAGCACAGAATTCGAGGGTTCGTCTCCACAGCCCAG

>APBMS25
CTATCTTTATATAAGCCAACGCGTTACCAGGTTGGCGTGCAGCGTCATGCGTGCGCGATGGCCGGCC
GAAAGCGATCCAATCCTTACTTTTCATCGCTTTTAGCGAAATTGGCCCTTGACCTTTGGCGGTCACA
CAAGTATCTACGCGCACTCTGTTAGATCGGCTGTGGCGTAAGCCAGACGCGGTCTGCAAGCACAAAC
CATCGTTTCGCCGGCACGCCTCATCTTTCATGGGGACTTGGCGGGACGTTGAAGGCCCACGCGGGTA
ACGCTGCGGGGTCTTTGTTTTTGCGGAAATCGCGTTCTTCCCCAAGGTCGGGGAGGGGCTATGGTCT
TTGAAAT

>APBMS29
GTTTCGGTAAAGGTCCCTTGCTTCCAAGATGGCTAGACTTCCTACTTTGGGACAGTCCTTTTCCACC
CCTTGTGCTGCCTTCCTTGGCACCCCTGGTGCTGCCGTCTTTGTCTGCTTCGGTGCTCCCATCCCTA

GTGCCCTTGGTACTTCCTTCCTTGTCTCCCTTAAGAATTCCATGCTGACTCGACCTAATATTGCCAT
CTTTAGTAAATCCGTCATCCCCCTGGGAGTATAGCCCATTTCCGAAATTCCCATTCTCAAACTGCCC
ACCGGAGAAATCTGCAGAAAATTCTCAAACTCTTAAAACCATTTGAACGGGAAATCTGCACATACAT
CAACGCAGCAAGTGCACATAATAACATTGGAAGACTTGATTTGAATAAGTAGCATACCAAATTCCGA
AAATGATGGATGATCAAGGACTTCCCACACACGAACCACACGGTCCTGGCCAGCACTGGCGAGATAT
CGTCCGTCAGGACTAAATTTCATGGTCCAAATTGCCCCTTGATGAG

>APBMS30
CTTTGAAGTGAAGACTAGCAAACTATATTGTAAGTATCCGTTCCGTATACGCAGTGAGGATGACCGC
CCACGCTTGTTTCAATTTAATGATGGTACCCAAAGTAACCGCCTAGAGGGAAATCGCATTCAACAAC
CTTGCTCCTATGTGAATTGCCCCAAACGATGTGGGGTATGAAAACTGTTGCAAGTAGTGTGAAATGA
GTGTGTGTCGAGTACTCCCTGGAAAT

>APBMS31
ACTTCGCAGCATACCTTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTA
ATATCTGATAGGGGGGTCATCGATCCCCATGATATTAATCTAATTTTTCTTGGGAGCGGGTTCACCC
GGCA

>APBMS34
ACTTCTCAGGACAGGGAAGACATTTCTGAGGCCTGTAATTCGCACACTCATACGTGCCACTTGGTCC
CAAAAAAGAAGACTCAAAGCCATGTGAACAAGATCTACCTCATCGTCCCATTTTACTTTTAGAAACA
GCCTATAATACGACAGTGACTGTAATTCTAACATGGAAAGGAAAGGAAAGAGGACGTAGACAAACGG
CTGTCCTGCTGCTTCCCAGCATCTCTGATGACCATTTTTGCGCTGAAACCACACGATAAAAATTTGC
GCTAAACACCCACATGAACATACTGAGTGAAACTTGAGTACATGCTCTTATAGTCAGGAGAGCAGTA
TACTTCATTTCAACAATAAATGTCTCCGAAGAAGACGTTGTGGCACTGACCTTGTGCGACTCCTCCA
CCGCGTTCACCACACCTTGATTCGCCCAAGACTGGACAGTCTCCTCTGCAAAGGAG

>APBMS37
CTCCAACATCTTCACCTTCTGCATCAGTATTTTCAACACAGACTCTCCAGACGGTCGGCCTGATAAA
TGCAGCCAACCTTCCTGTGACGGACCCCCAGTGACTCCCTTGGCAGGTTCTTCTTTACCTCTGGCCT
CAGAACCAACGCTACTTTCACGTTCGTCCTTCACTGGCTTCTCAAGAGGTTCCGACAAGAAATCAAC
GGTGGCCTCATCTTCAACTGATTCGTTCACTTTCACAGGGTCAGTTGCACCTTTCTCAGGGACACTT
TCATTGGATCTGGCACCCGTAACGGACGATGTTCCATTCGGAATAAAACGACGACCACCTTTCCCCG
GATCGATGTCATCTCCGACAATCCAATCCTCCAAAAGATGCTCAATAGCATCAACTCCATGAATTTG
TAACACACTCAGTGTGCAATAAAATTCGGAACCATAGTGCGAAAGCAAGCGCAGTTTCAAAGTCCGT
ACCCACTTGTGCTCTGGAAGTGTGAAATTCTGAACATGACGAACA

>APBMS39
TCTTCACTTTGACATTCAGAGCACTGGGCAGAAATCACATTGCGTCAACATCCGCAGGGACCATCGC
AATGCTTTGTTTTAATTAAACAGTCGGATTCCCCTTGTCCGTACCAGTTCTGAGTCGGTTGTTCACT
GCCTAGGGAGAGCCGGGGTCGCCCCCGACGTTCTGCGTCCGTCGCCCGGTCGGCACGCGGGGGTCCG
CCCTCGCCGCGCGAGCAG

>APBMS40
TCTTCAGGTTGGATCTTGTACACGCGCGGCGAGTACGTGTGGATGACGCGAGGAGCGTACGCGTGGA

CGACGCGCGGTGGCTTCGCACACCCGCCGCTCGACGCCTTGTGAAGATGCAAACCAGTTGACGACTT
CAGCGTACTAGTTGGTTTGCTCGCGTTTACGTGCGCTTCAATTGCCCCCATTCCCTCGAAATGGCAC
GCCCGAACAGGTTGCGTGAATGGAAAAAGAATCCCTCGACAGATCTCGCTCCTGACCAACAAGATTA
AGTTCTTATCGAAGAAGTCAGTGGCGCGCTTCAATGTAAAGGCTCGACCAAAAAATAGCGTAATCGC
CACTTAAGAGGCTCACCACCACCACCTGAATCTCAAAAATCATTGAGGCTCAACCAAGGTCATGCGC
TACCAGGAATGGTCTCAGTCCATAGATTTTCGACGAGTCGAG

>APBMS41
TATTCGACGGCGTTTGCTCCAAAAACATTTTCACCACATCGTCCCGGCCCCACTGGGCCGCCAAACT
CAAAGGGCTGCTCCCATCAAAATTGGTGCCATCAACCACGTCATGATTTTGGAGCATTGCCCGCAAG
ATAATTTGCGTGGCTGGCACATGCCCCCATGTCGCGCTGAAATGCAAAGGCGTGTATCCCTTGGCGT
CTTTGAGCGTGGGATCTGCCGAATTCGTCAGCAGCAGCAGCACCATATCCAGGTTGGTTTCAATCGC
GGCAATATGCAGGGCCGTCTCCCCGTCACGATTGGCAGCGTTTACATCAGCGTCTTGCGCCAGCAGG
ACTTTGGCGATTTCGGTATTCCCGTTTTCTGCCGCCAACTGCAACGGCGTCGTGCCCCGGTAAGTCT
TCATGTTGGCATTG

>APBMS42
CATTGCGCAGGATCTTGTTATTCCGGAAGATCGAAATCTCGCCGCCAATTTCCTGCACCAGGCCGTC
GCCATCGGAGTTCTCCGTCATGGTATTGCCGTCCACGACCTCGCTTGTGCAGTCCTCATCAAAGTGG
ATGGCGGAGCCTGTATTGTTGCGGATGAGATTGCCCCGGATCGTCACACCGCTGGTTGAGCCGATCT
TGATGCCGCCCGAGCCGAATGTAGGATCGAAGTGGGCATAGTCGTTGTGTTCGATCGTGTTTCCCTG
AATCAGAATTCCGGACATCGTTGACTGCGTTGCGGGATTTTTTTCCACGCCGATACCGCCCCCGATA
CCGGTCTCACCGTTATCGTGAACGTAGTTGTTGAGAATATGCATACGGTAGTTGATCCGTATGCCGA
AACCGTGATTCAGCCGTACATCACAGTGTTCGACCGTCCAATTCGTTTCCTGCGTCAGCGCATCGGC
TCCCTGCGTCTCCCCGATCGCGCCGACCGGATACATGTCGGCGAATTTCTCGATGGTTAGATATTTG
ATGGTGACGTTATTGGCGGAGCCGCCGAAGC

>APBMS44
TCTTCGGCGGGAAGGAGTTTCGCGGTGGAACCGACGCGAGTGAGGTCAAGTGGACGGAGCGACGAGG
CATAATCTGCTTCATTACTGTCGCTGGTTTTTTGAACGGACCGGGATTTCAGAAGATGCGGTCGGAT
CTGCGGCAGGACGCGGACGAGATCTTTGTCGTGGACTGCTCGCCAGAGGGTCATCAACCACCGGTCG
CAACGCGCATCTTCCAGGCTGTGCAGCAGCCCGTATGCATCGTTCTCGCACTTCGCCGTTCTTCCTC
GAAAGAGGGTGGTTCCGCGATCGTGCGGTTCCGCAGCCTCTCGGAAGGCTCGCGAGAGGCTAAGTTT
CAAGAG

>APBMS46
GTTTCAACGCAAGTACGTGCTCGACGTTCTCGAGCGCAACAACGGCAACCGAACGCAAACGGCGCGT
GACTTGGGGGTCGACCCGCGAACCATCTTCCGTTACCTGGAGCGCGAAGCGAACCCACCGCCACCCG
CCGGCTCCCGCGCGCTCGACGAGTAGGGCGAGCCCCTCTCATCTTTCCCGCGGGTGAGTTGCCGCGG
AGACATGCGGGCACGCGATGTGCTGAGAGGCTTCTGCTAGCAGAAGCCTTGAACATGGCCTCGATGG
CCGTGCGTTCGCTGGGTCCTTGCACGAGAAAGCGGGAGCCGGGCAGATCAGCAGCCCGGCGAATCGG
CCGCCGCGGCAGCGACATCAGCGCGGAGGGGGATATGGGAAACAAAGACGCGATAATGACAAAGTCA
TTGCCGCGCAGGCAGGCATCGAGAAGCGGCAGCCACTCGGCTTCGCCGATGTGGGCCTGGCGGGCGG
CCACCTCGACTACCGGTTGCGGCGCATTGGGGCCACCGGGCCACACCTTCAGTGGAACC

>APBMS48
GCTTCAAGGTCGTGCCCGGCAAGGAGGCAACGTTCGAGCATGTCTGGACCAGCCGCGACAGCCATCT
GCGCGACGTGCCGGGTTTTGTCGAGTTTCACCTGCTGCGCGGTCCGCAGCGCGACGATCACGTGCTG
TATTCCAGCCACACGATCTGGGCTGACCGGCCGGCCTTCGAGGCCTGGACGACGTCCGAGGCGTTCC
GCGCCGCCCATCGCGATGCCGGCCGTATCCGCGCGGAACATGGCCCGCTGTATATCGCGCCGCCGGA
GTTCGAGGGCTTCGAGGTCGTCCAGACGGTCAAGTGACGCAGGCGGCGCGTCAGTCGCAG

>APBMS50
TTTTCCCTTCTTCTCCTCGTGAATTTCAGCCAAGCAGGCAGTCAGTCAGGCTGCCAGTTATTCCCGG
TTTCCCTTCTCCCAATGCACATAAACATACACACGTGTTCTCCTAATTAGGGTAAATGTAATAAAAA
AGGACATCACTCTTTCGCAAACATCTCTCTCCCAACCCTGTCTCCTTTCCCCTCTCTTCAACAAAAT
CAAACCGACAG

>APBMS53
CTGGTCAACTCCTCGCCTCCTCGGCTGTCAGAGAACTGGCGACGATCGACGCGCCAGCCACTAAACC
CCTACTCCTCATCAGCATTTGTAACCTCCCAATGTGTTTTGCCTCGTTTGCAACTCGATGAACTGGG
TTCGCAATGAGCATCGCACCAAGGGCAGGCAACCTCCCAGCGCGCTGAATCGGCCTTCGCGACAACG
GCTTCGCAATGCCGACTACAGAACCATCATATTTTCTTTCCGTCGGGGCCGAACTGCTTCTGCTAGC
AGAAGCTGTTCTCACCGAGTAAGCCGGCGATCGCGCTCGGCGAAACCCTCAACCCGCCCTGGGCCCC
GCGGCCTGGATCGCGGGATCGATGCCTTCGAACTTCCGGAAGTTCTCGACGAACCGGACGGCCAGGT
GACGTGCCTGTTCGTCGTAGGCGAGCGGGTCGGGCCAGGCGGCTCGGGGGTCGAGGATCTCGGGCGG
GACCCCGGGGCAGGTCTTGGGGAATTCGAGGCCGAAGATCGGGAGGTTGGCGTAATCGACGTGGTCG
AGCGAGCCGTTCATGACGGCGTCGACCAGGGCCCGGGTGTGGTTCAGGCTGATTCGCCTGCCCTGGC
CGGTCCCGCCCCCCGTCCAGCCGGTATTGAGCAGCCAGGCCCTGGCTTGGTGGCGTCGGAN

>APBMS54
AATTCAACTTCGCGGCTACTGTAAACAACACAGGTGTGCTTTGTACCCCAAAGTGCGAATAGTGCCG
TTGACACGCCACAACCGGTTTCGAGTGTGCGCGCGCCATTGGTGACATGGCGCGCCATGACGGCGTA
ACAGTCGGTAGCAGTCGCCCAAACGCGCTGCGCACCGGCCGCCTCGCCGTGGACACCGGGCGGATTG
GCGAGTACGCGTGCGATCGCGTCCAGACGCTCGCGATTCTCGTTGCCGGGTGCGATTTCGGTGGCGG
CCATCGACATCTCTCTCCTAGTTTTTTAGATCGTGTTGACGAGCTCGCGCAGTGAGCTCGGGTCGTA
GCGAGGCGCGTCGATCATCGCCCCGTTGGCGGGATCGCTGGAGAAGATCGCCGCGAGCCCCTCGACC
GGAACGTCAGCCGGCGCAACCGCGACGAGTTCGGGGCCGCCCTCACCGGCAAGTCCCGCGCGCGAGA
CGACCTCGATCAGCGGTGCGGTCTCCTGCGGGTCGGTGCGGATCACAAGGCGTACGTTGTCGCAAGG
CTCACACTCACTTGACCATGCCTGCAACAGGCTTTCATCGGCGATCAG

>APBMS55
GGTTCGTCTGTTCCCCCAAGTTGGATGCGTCGTCGATCGATTTTGAAAATGTGTAGCCGGCCAGCAC
CGTCGCGCGCTGTCCAACCGCATAACGAAAGTTGGCTTCCAAGGCATTATAGTTTGAGTTACCGATC
GTTTTCTGTGCTGTCATCGAGCCGTAATTGGGCCCCAAGCCGACACGCGTTCCCCGAATAGTTTGTC
CGGAGGCCGTCGTATAGGCGGCATCCTCGCCAAAAGGACCGCAGGTCGCGCTTCCAG

>APBMS57
ATTTCGTCGCCGCGCGCGGTCCCTCGTCTGTGGAGACCGTGAGGCTCAGGTCCTGCCACCGGTCACC
GTAAGTATCACCGATCTCGGTCAGCATGATTAGGCCGCGATCTGTGGTGACCAGCGTGTCACCGGTT

AGGCAGCCGGTCGGCGCGAGCAGGGAGGCTTGCGCGTTGCGGTAACCGTTCTTCTCGCCCAGGGTCA
GGCAGTCCGACCAGGCCTTCGCGGCGGCGGCCAGGATCTCAGCGTCCATGCCGCCGAGCGTGCGGAT
CTCGTGGCAGGCGGCGGCGTGCTTGGCCATCACCCGGTTGTGCGCGGACTTGTTCCGCTCGTAGCCG
TCGTAGGGGCCCACGACCGCGGCCAG

>APBMS60
CTGACTTTTGGTATCACCATTAGTTCCCTCCCATTCGTTCCATCTCTTGAACCACGCCTTTGCTCCT
CTCGGCTTCAATCCTTTCATCTTTACTACGCGTTCAAAAACAGCCCGAACTATCTCCTGGTCGCCTT
GTTGAATCTCCAGATCTAAGAG

>APBMS62
GCTTCGGCTTCATCACCCCGGACGAGGGCAGTCGCGACCTGTTCGTCCACCACACCTCGATCACCTC
TGACGGCTACCGCTCACTGGCGGAGGGCTCCAAGGTCTCCTACGAGGAGGAGGCCGGTCCGAAGGGT
CCGAAGGCGATCAACGTCACCAAG

>APBMS63
GCTTCCGTCGGGACCACCGCCACCAAGTTCAGCGAGGAAATTCTTGTATTCATCATCCATCTTATTA
CCCTGTGCAGACCCTTTTAAGGGACAATCGATTGTTGGATGACCACCATCTCCACAAATGCGACACG
AAACATCGCTTTTGAACGTGGAGTGTCTAGCCGGACAAGCATACTGCCTGTGACCAGGTTCGCCACA
AAGACGGCAATATTCGTCATCTCGTATAGTTCCGTTCAAGGCCGCAAG

>APBMS64
CTGCGCAGGAGTCGCGGATGGTATACGCACCTGCGCCGGCCACCGGTGTCCGGCCGGCATCGTGACC
GACGGCCGCACCGCAAACAGCGCGTTGCCAAGCCAGAAAGCGCAGACCCGGCAGATGCCGGGGTCTG
GCCGACTGCGATCATGTCTGCAAATCAACCTGCTGACATGGCCGGGCGAGCGGCCAGGCTGAGGAAA
GGCGCACCGCAATGAGCGGCAAGAACCGCTTCTGCTAGCAGAAGCACTTCCCCTCCAACATCCAGGG
CCTCCCCACCTGGTACGAGATCCGGGTGACGCGCGAGGCCTATGCCGCGCGCGCTGGCCGGGTCGAC
ATCATGGTGGCGATGAACGCCGAGACCTACGCGAAGGACGTGCGCGAGGTGACCTCGGGCGGATACC
TGTTCTACGACTCCACCTGGCCGCGCAGCGCGCTCCTGCAGCGCGGCGACATCAACGTCATCGGCGT
GCCGCTCGCGCAG

>APBMS65
AGTTCAAGCCGGGCAAAGCAACATGATCCCCTGAGCATCCTGTTCTGCGGCGCAGATGAATTCAGCA
TCTACTCGCTGCGCGCCTTGCATGAATATCAACACGTGAGACCAGACAAAATAGCCAGCATCGATGT
GTTGTGTAGGCCGGATAAGCGAGTCGGTCGAGGGTTGAAGCAG

>APBMS66
GCTTCCGCCTGGTGGTGGCCGACCGCGGCGTCGGGCGCAGTGGATCGCGCCAGGGTGATGGCGCGCG
CCAGGGCTTTGGCACGCGCATGCTGAGGGCGGTGGTCGAGCAGATCTCTGGAAGTCTGGAGGACGGC
GACAACAAGCCGGGCTTGCGCGTGTTCGTCACCGCGCCCATCGTCCGGCCATAAGCCTGACGCACGC
CGGCCTCGCGCGCCGACGCACGCCTTCGGGGGCCGCCATGCGTTCCGACGCCCGCCTGACGCTGCGC
GAACGCACCGCCGAATCTCATCGCCAACTCGATGCCGACCTCGCCCGCTTCGACCTCGCCGACCGGA
ACGACTATCGCGACTTCCTGCTGGCGCACGCCCACGCCGCCGCGCCCGTTGAGGCCGCGATCGCCAA
G

>APBMS69
ACTTCTCTCCGCACCCGACAGTGAGATCCTGAATCCCCGTGGCCCTCAAGGCTCCGTAACAGACCAG
TCTCATAGATCCGAAAGTAAGCCATACGAAAATGTCGCATGGGGTCACGTCGTGGAGTGCACAAAAG
AATGCCACTCACAGTTCAACCGCGATCACTAATACACTTCGTCGAAAACAGACACCATGCCACACCC
GAGAGTCAAAATCTGCAGAATGGCTTCTGCTAGCAGAAGCGCTTCTCAAACCAAGCCTTGCCAGCAA
CAGCAAAAACACTGTAACGGGTAACCACACCCGCTACAAGAGGAACGCCCAGAAACAGAAGAACACA
CTTCGACACGGCCCACACACCCACATACACACTACACCCTCCTAAAACCTTCACATAAAACAACGCC
AGCGGGAGGTACAGGACCACTTGCAGAGTGGAATTCATCGCCACCAGAACAGCACACAG

>APBMS70
AGAGCACGGCTCACCTCGTGCACAGCATCAAGCGACCTCCTTTGCTGCTGACAGCCATGGTCGACAC
TACACTGGCTCAGGAGGCGTTGGTTCGCCACGTGCATCAGGGCATCACCCGAGGAGCCATGGTGTGC
AACTCGTTCGTTCCCACGGAAGAAGTGCATAACCAGCCAACAAAGAGGCGAAGGACCAAAGACCAGC
AGCAGCACCGACTCAATAGCAGAGGTGATTACGAGTTTCGGAACA

>APBMS73
CTCCCGCACTCCTAACGCACCACGACCTAAATTCCCTCTCTCTTCGTCGTGGCACTGCATCCCCAAA
GAACTCTCCAAAATTTCTAGGCAAAACAGCCACAGTAATCCTCGGTATCACTCTGCTCTCCCTCCTA
CTTGACGAGAATCCTACAATTCGCTCATAAGACTCAGGCCACAAGTTCCTGATGTGAAAC

>APBMS74 --T3 end
TCTCGGTCCCTTNTCCCCGCCATGATCGGGCCTATAGTATTACCTGGACCCCTGTCAAGCAGCCCGC
CGTCGACGCGAGGATGAACAATGCCTGGGCCCGGTCGGATTTCGACAGTTTCGCGAGGAAATCAGCG
GACGCGTTGCTGTCGGCGTTCCGGGAAGCCTTAAAATCGCAAGACCTCCGCGCCGTCTCGTCGATGT
CATTATACACCCCAGACGGCGATACCCGCCTTGACATGCGCGGCTGTATCGGTTGGCATGATCACAT
GGTAGCCGCGATGCCTCACGATCGGGTGGTGTCCTTCGGCAACGACCACATCCTGAATCGGGCGTGG
TGGGGAGAACCGCTTCTGCTAGCAGAAGCAATTCACTGTGAGAAGCAACAGTCTTTACGATCACAAT
CAGAAACCGAAAACTTCGTTCTCATAACTGCAAATTTTCCAAATTCACTCCACCAAAAACCTCACAG
CAATTATCAACTAGGCCATAGCCCCTCTTATAGCCCCTACAAACCCTCAAACCACACCGTCGCGTCT
TCACAACTCCCGTCCAATCGTCACACCATACTAGGCAGATTTCATACTGCCAATGCAAAAAATCATT
TCCATGCTCGCCACTTTCAACTCAAACCTTCCCAGACACAAAAGCAAAGTTCAAAAG

>APBMS77
TATTCCCACCTAAGGAAAAACATGAAAAAGGCGGACCAGCCCTAAAATGAGCAGTTTCTGAGCCTCT
TAGCCTACACCCCTAGTGATGATGAAAGGAAGTGTCGAAGATCAGAGACGATCCTCTCGACGTGTGT
CCAAGCAAACTCAGAGAGGGACTTATCCTGCACCTCGTTTCATTTCCTCTCTGACTATGGCGATTTA
GGGTTTCAAATCACACCATCGCCATCTCCCAGCACACTCTGTCTAACAAACCTTGCAGAACAAAGAA
CGACCATGGCGATGGCAGGCTTCTGCTAGCAGAAGCGGTTCACCGGTCGAGCGTTGGGGCAACCCGG
TTTGAAGCCCTCGATCCGCGTCGGGGAGACGGGCGTGCGAGAGGTCGCTGCGTATTTGCTGGACCAC
GAGCAGTTCGCCAGGGTTCCCGTGACGGCTCTGGTGAACGCGACGCACTCCGTTTTCCATGTGAATG
CGGACTACGCCAGCGCAGGCCAAGCCCAGAACGCGAAGATTGCTTAGTTTCAGCAGTTCGTGAGACA
CGACTACGATGCCAACGATCACGGCACCTCGCAGTTTCCCGTTTCCAACGTCCACCGCATCGCCATC
CTGGACGTTCGCCTGTTCAACACCGACCGCCACGCAGGCAACATCCCGGTGAAGAAGATGGACGTGG
AGTCCGGGAG

>APBMS78
ACTTCCAAGACGGTCACGATAAAGAACTCCCAAACGAGACAAGAACCTGCTCGTACTGATCAATGGT
GTACTCGTATCCCCCACTTCGATTCACTCGAGCGGGATCGTCCACACTCGGACTTCACTTCCGTAGA
TTTACGTGAACGACCACAAAACATCTCAAACCAAAGATCAGACCATGCAATTGCTTTCTCATGGTCA
GTAGAGGAAACCCAAATGCAAACGTCACAGATGCACGCTTTTCGGCTCGTCAAACATAGACCAGCAG
TACCGTTGAGAGCGTTTCTGATGGTCCCGTTTGGACTCTTCCACATCTCCTTAAGTCCAAATTCCTT
CACGCGGGCTTCATCTGTAGACGACAAAAAACACATTCATTTGCTGACAGAAAAGTAGCAACCGAAG

>APBMS82
CTTCAGCCCGTAGGTCCCGCAGCCATCCGAACAACCGGGGCTTTAGCCCATGAGGGCAACGCTGTCA
ACCTGCAAGACCTGCTCGACGAAAAATACCGGGGCATCCGTCCCGCCCCCGGCTACCCAGCCTGCCC
AGACCACACCGAGAAAGGCATCCTCTGGCATCTGCTGGACGTGGAGAAAAACGCCGGCATGCGCCTC
ACTGAATCCTTCGCCATGTGGCCTGGCTCCAGCGTAAGCGGCCTTTACTTTGCCCATCCGGATGCGC
GCTACTTCACCTTGGGCAAGATCGAGCGGAATCAGGTAGCGGATTATGCGATCCGTAAAGAGATGAG
CGAGGCGGAGGTAGAACGGTGGCTTGGGCCTAATTTGCACTATGAGCCGGCAGGATAACACCGTCGG
AAAC

>APBMS83
ACTTCGAACCACTTGGAATCCGCCAAGGACGCGGCCTGTGGGGCGGGAAGCGGTTCCGGAGCGCGGA
ACGCATCCGGAACCGCCACCTTGGGCCGGACATAATTGGGTCCAACTGTGCAACCAACTAAGGCGGC
CGCAGTAAGGGCTACCGCAATCGATCGATAATTCATACTAGTCTCCCTCCGCCATGGCCGGCGACAG
GCCCGACGCGTGATTCTTCGCGCCGGATAG

>APBMS85
CTCGACTACACCGAGAAGTGATTAGCGCGGCATCACCAAGATCGGTGTCTCGCTAACCTGCAGCCTC
ACCTTGCCGTCCGCAACCGACTCCTCCGAGCCTGTCCCAGTCACCGAAGATCCAGCCAGACTCACCA
CCGTCGCCGACCTCCCGCTCATGGCAAGCGAGTACTCCACGTGCTTGTCCTCATCGGTCGGACACCA
AACGGCATAAGCCACCCTGGAACC

>APBMS87
TCTTCAGTTCGGTGCCCATCGAAATCTTGGCTTCTCCACTCTCAACTTGTCACGTAGAAAGATGATG
TCGGAGGGAAACTTTCCAACGCGGAGAAGGAAGGTTACCATGCGTTGATCTGCATCACAGATCTTTG
CCAATTCGAAGCCGTCCAGGTTCAACACGCCACACACTGCAAGCGTTTCGTCACTTTCCTTGCTCGT
AGTTCGCCTGCAAAGCACATGATGCACATCCCTAAAGGGAGTCAAAGCCAGAGAATCCTGAAAGGGC
TCCAAGTAGATCACATAGAAG

>APBMS88
TCTTCAAAATGACCCCTACGCAAAACTGCTTAAAAAAAAAACCAAAAGTCCTCTTGCGCCGTGAACAC
ATGGTGTTTGCAGTTTCCTGCCTTCAGTGGCACTTTCGAGATCCTCCTTGCGAAGTTGCTCAAACGC
GCGGCGGTTGTACACGCGATGCACCATAACATCCGCCAGCATGGTGAGCACGTACCCCAGTATGGTG
AACATCGCCGCGAACGAGTACACCTTCGATGGCATCAGGTGCTTGAACGACACGTTGGAATCGAACA
GGAAGTGCATCAATGCGGTGGTGAGGAATACCCCCCCTGCGAATTGGGTGCCCAGTATCAGGAACGA
G

>APBMS89

CTCAACCAGATCGCCCATCTTCAAAAACTTGGAATGAGTGCCACTGCATTCTGGATCCCTGGCAGTG
GAAGTTGAGAGGTGCGCGAAGGCGTGCTGCTCATCTTCCTCGTCTTCCTGGTGGCGCGAAAATGACT
CGACGTTGGGGAGTCGGGTGAGATTATTGGGGGTGTCACGAGCGTCGAGTTCGTCATCGTACTTGAA
TT

>APBMS90
CTTCAGCCCGTAGGTCCCGCAGCCATCCGAACAACCGGGGCTTTAGCCCATGAGGGCAACGCTGTCA
ACCTGCAAGACCTGCTCGACGAAAAATACCGGGGCATCCGTCCCGCCCCCGGCTACCCAGCCTGCCC
AGACCACACCGAGAAAGGCATCCTCTGGCATCTGCTGGACGTGGAGAAAAACGCCGGCATGCGCCTC
ACTGAATCCTTCGCCATGTGGCCTGGCTCCAGCGTAAGCGGCCTTTACTTTGCCCATCCGGATGCGC
GCTACTTCACCTTGGGCAAGATCGAGCGGAATCAGGTAGCGGATTATGCGATCCGTAAAGAGATGAG
CGAGGCGGAGGTAGAACGGTGGCTTGGGCCTAATTTGCACTATGAGCCGGCAGGATAACACCGTCGG
AAAC

>APBMS92
GCTTCAACGGTGGTGACCCGTACTTCGCGGGTAACTCCGCCGCGACGGCCGTTGTCAACACGAACCT
GTGTACCGCGGTGGCGTTCCTGGTCTGGGTTGCCTGGGACTACATCACGGGCCGCAAGCCAAGCCTG
ATCGGCGGCGTCAACGGTCTGATCACGGGACTCGTGGCGATCACACCGGCCGCCGGTTACGTCAACG
GCTTCGGTGCGATTGCCCTCGGCGTGATCGCGTGCACGATCGTGTACTTCTCGCTGAACTACCTGAG
CAGTCTGCCTCCGTTCCGGAACGTGGACGACACGCTCGGCGTGATCTACACGCACGGCATCGCCGGA
TTTGCCGGCGGGTTGCTGACCGGGATCTTCGCGGACCCGGGCATCACCGAGTACTACACGAAGGTGA
AGGGTGCAGTCGTGCCGGTTGGTGGTGCTGCCGGTCTCGTGTACACCGGCAG

>APBMS94
GGCGTTTTCGCTTCAGGGCGTCGTCCGCATGCTCACTCCAACGCGCACCTGGACATTGCCGCCGTCC
CGGGCGGTCTGGCTTCCGTCTTATGTGGACCACGAGTTACACGCAGTCGGGAAGTCGGAAATCTGCA
ATGTGTACATCGAACCCGAGGTGTTCCCGTGGTCCTGGACGGAGCCCGCCATTATCGCCATCACACC
GTTGCTGCGGGAACTGGCGTTGCTGGTGAGCGGAGGCGGTGATGGATACCGCGCGGAGAGCCGCGCT
GCGCGGGCCGCGTCCTTGTTGCTGGAAGTCCTGGCCGATCAGCCGTCACTGCCCGACAGTGGCATCC
CGCTTCCACGCGACGAGCGATTGCTGGAAATCTGCGGACATATGCTCAATCATCCGGCAAGCAACCA
GTCGCTTGATGTCTGGGGCGAGAAATTTGGCACGACCGGAAGGACGCTCGCGAACC

>APBMS95
TCTTCTCCTGCTTTTCAGGCGGGATACCGATGCCCGTGTCGCTCACCTCGAAACGCACGCGTACCTC
TTCCGAACCAGAAGTCAGAGGACAGAGGTCAGAGATCAGAAATGCAGAGGCCACTCCTTTATCTTCT
GACTTCTGACTTCTGACTTCTGAGTCGGCGACTTCCAGTCGCACAACGACTTCGCCCTCTTCGGTGA
ACTTGATGGCATTGCCGACCAGGTTGAGCAACACCTGACGCAACCGGCCGGCATCTCCTACGAGGGT
GTCGGGCACGTCCGGTTGAACCTGGCTGACCAGTTCCAGCCCCTTCTTGTGGGCCCGAATGGCCAGC
ACACGAAGGGTGTCGCTCAGAG

>APBMS97
ACTTCACCCCGGCTACCGCCGCCATCAATCAGGCCCCCGGCCCGGCCATGTTCAGCCAGTTGTCATC
CGGAGCCACGTTACCTTACGACATGGTCTATGTGACCATGAACGACGGCACCGTGGTCGCGCTGGCG
GCCGAAAATTCGTTGCTCAACCCCCCACCGGCGGCGTTCAACACGCAG

>APBMS98
CTCGCCTAGAAGATGAAGCTGTACCAGCATTCCTCCCTCCTTGACCTCTTCCACTTCCACGTCCAAC
TCCATATTGACCACTCAAAGGTTGACTAGGTGGTGGAGGATTTCTATTGACAATAGGAACAGTAGGT
ACTGTATTTACTAGGACCTCAACATTTCCCACTGAGCCTAAGCCTTCTTGATGAGGTCTACGTATAA
CGGTGGAACTCATCGTTGTTGTATTGTTTTGGGAAGAATTCACCAAAGTATTTTGACGCTGTCCTAC
AGTAGGATTCGGTACTACTACGTGTCGTGGTGGATGAACCACATCATAATGGACATTGTCATCATTA
ATTTGCTCAGCACCAATAACCCGTCCACCTTGTACTGGGGTTGCATTTCTTGTCACAATGTTTGGCC
TTCCAATATTTGCTGAATTTGTACGAAATTGCTCTTGGCGCCTTGCATTATTATCCAAGGGAAGAAC
T

>APBMS99
AATTCGGTCTGGTGAGTGGTTGTCTCTTGTTGAAGAATGCTGTCTGTTTCTGTTAAAATTGACAAGA
TTCAGTTTAATTTTTGGAGACTGGTGGGTCTATTTTAGTTTCTTAAGGAAGAGCGATTTGGCAATTG
GAAAATGGTCAATTAAATGAATCTTGTCTTAATCACGGTTGCTTTGGTGTCGGATGGGCTTCCAG

>APBMS100
GGTGGAANTGGGTCGACGGCGGTTCGTATGATGGCGGGTATCGCCGCGACCATGGCCGGCCGGACAA
CTTCGGCAACCTGATGAGCGGTTTCAAGCGCAAGGGCAAGGGCGGCGGCAGCAATGGCGGCCAGCAG
CGCAAGGAGCGCGGGTTCGTGGGGCCGATGGACCACAG

>APBMS101
CTGATGGTCGACGGCGCGGACGACCAGCCGAGGGTGAAGGCCCAGGCGCCGGCGTGGGTCCGCCGGT
ACGCCCCGTTCCTCGGCCGGTTCTGGGGCGAAGGGTTGACGGTCGTCGATCGTCCGGGCTTCAACGG
GCCAATCCTCGACTGGGCCAAGCGCGACCCCGACGGATTCCTCAAGGCCGCGCGGGGGCTCGCCGCG
AGCCATGGGAACGACGGGGAGACGGCCGACGAGCGTCAGCTCCTGCGGGCCATGAACC

>APBMS102
CTTCTTCAGCCTCCCGTTGCTTCGCAGCATCTAGAAAGCGTGCTTTCTTCTCCACGTTCATTTCATC
TTCACTATCCTCATCACTCTCATCTTCCTCGTCGCTGCCATCATCCTCCTCCATTTCGATTTCACCA
TCTCCCTCGTCGTCTTCGTCTTCCTCTCCCTCCTCCTCCTCATCTTCCTCTCCCTCCTCCTCCTCAT
CTTCGTCCTCGTCGCTTTCAATCAACTTCGCCTTCAACCACTTCTTATTGTCGTCAGTGAAACCCTT
GTTCTTGGTTGACACTTTTGTCTCCAGGAATTCATCACTCTCCTCCTTCTTGGCCTTACGCTGGACA
ACATTTTGCAATGACAGGGTTTAGGGTTTAAGGATTGCCACATACAGAACCAATGAAAACACAATAC
ATAAGAACTAGCCTAAATTGCAGGACTATCTTGTCCAAAAATTCAGACACTACAGTGAAAACCCTAA
AATTGCAACTAGTCCAGAAGCGCATCCCCATATCTGAAAT

>APBMS103 --T3 end
GTGCTTTCTTCTCCACGTTCATTTCATCTTCACTATCCTCATCACTCTCATCTTCCTCGTCGCTGCC
ATCATCCTCCTCCATTTCGATTTCACCATCTCCCTCGTCGTCTTCGTCTTCCTCTCCCTCCTCCTCC
TCATCTTCCTCTCCCTCCTCCTCCTCATCTTCGTCCTCGTCGCTTTCAATCAACTTCGCCTTCAACC
ACTTCTTATTGTCGTCAGTGAAACCCTTGTTCTTGGTTGACACTTTTGTCTCCAGGAATTCATCACT
CTCCTCCTTCTTGGCCTTACGCTGGACAACATTTTGCAATGACAGGGTTTAGGGTTTAAGGATTGCC
ACATACAGAACCAATGAAAACACAATACATAAGAACTAGCCTAAATTGCAGGACTATCTTGTCCAAA
AATTCAGACACTACAGTGAAAACCCTAAAATTGCAACTAGTCCAGAAGCGCATCCCCATATCTGAAA
T

>APBMS106
CTTTTCTTGCGATTGCCCGGCGATTCGATCCTGCGAAAGCACATCTCCTGCTAAAAATGGCCGCCCG
CACTCGACGCCGCAAGTCTGAAATACCTCGTCCGAGCCGGGGCTTCTGACATTCGACGCTACAATCA
CTTGACCTTCGCGCAGATCCGGCTGTAAGGCTCCGCACAGGCCGACGCTAACGACCGCTTCCAAAGC
CGACGATTGAAGATCGGCTGCGGAGATGGCACGAATAGCGATTTCTAGGGCGTGAGCGGCGAGTTTC
GGACCAGTTCCGTTGGCCGCCAGCAGGATTCGACGACCCTCCAGAATC

>APBMS107
CTTCAACAATTTGATAGCGTTGACTTAATGCGCCTGGCAGTTTGATCAACGTGGCGGGCGATGGGGT
AGCGATAAGCGGCCGGCCGCAGTATAGACATTCTGTGCTGCCAGCAGGATTAGATTGCGCGCAATCA
GCGTAAGCGCATAGCAGCGTAGCACTTGCTGTTGCTGGGACGATTTCGGCTACCACAGATTTTCCTT
CATCCGCCGAAATCAGATCGACTCCGGCCAGCATGGTGCCGCAGGTGCAGCGCAATGTGTTCGGATT
GTTTTTCGTGCCGCAACTCGGGCACAGCCGGAAAT

>APBMS108
TCTTCACTTTGACATTCAGAGCACTGGGCAGAAATCACATTGCGTCAACATCCGCAGGGACCATCGC
AATGCTTTGTTTTAATTAAACAGTCGGATTCCCCTTGTCCGTACCAGTTCTGAGTCGGTTGTTCACT
GCCTAGGGAGAGCCGGGGTCGCCCCCGACGTTCTGCGTCCGTCGCCCGGTCGGCACGCGGGGGTCCG
CCCTCGCCGCGCGAGCAG

>APBMS109
GGTTCCTTTGGAGAAGCCACTGCCGCCGGATCAGCGCGAGGTATTGGGTATCGGCGCGGCGGAGTCG
GTGCGGTACCGGCGAGTGCGGCTGATCTGCGGTTCGGTCGTATTGTCGGAGGCGGACAACTGGTATG
TACCGGCACGCCTGACTTCCGAGATGAACGAGCAG

>APBMS110
ATTTCATTAGCAATGGGATGGTAGGATAGTCATTCCTCCGAAGATCCATGCTCCATCAACCAAACAA
TGCATCAATGTCGTCATGTGTATGTCCCACCACCATAAATGACACTTTGATTTCGTTGAAGATTTGC
TTGGCTACTAGCAATGACCAGAAGCACATGATGTGGCGGCTTTTGTTGTCCTTCCAACAATTGTCCA
ATTGCACATGCAAGATAGGGGGCAACTTCCGAGATACTTCGCCCACCTCTGGCCGGCACTTCATCGC
TTGCTTGCACTCATCACTTCCCACTAAAAGTGCCTTCGACAAAGGCGACAGTCTGGGTCCTTGAAAG
AGCTGCGACGGATTACGGTACATTGGTTCTTGCTCCAGGTCTCTCAAAAGCTGACTTATGGACCCTA
TGGTCTGGTTACAATCAGCAGGAAACTTATCCACAATGAAATGAGCATATTTTCCATCACCGTAACC
ATGAGCCAAGACCCCCGAAATGGATAGTGGTAATTTCGCAAGCCCCTCCTTCACTTTTGTCTTCTTG
GCAAAGCAAGGGGACGCAGTCTTCGAGTGATCCATCTTATCATGCATGATAGACAGCACCTTCTCTG
AATAGTGGTGCGATAGAG

>APBMS111
GCTTCTTGAAGATGGCCAGGCTCCCGACCCATCCACCCTCCGTCGTGTGGAGAACGCCGGTCCGAGG
CGCGTCGAGCAGGACGCCCTGGTTGAAATGAGGAATGCCGCCGATGTGGGTGATGGGAAACCTGCCC
GCCTGCTCCAGCCACGGTTTGACAGTATCGGCCATTTCGCACCCTCCTGGATTTATCTTTGCAGCAA
TTTCTCGTCGATCGAAGATCAAGAG

>APBMS112 --T3 end

CTCGAAGAGGCGGGTTCTACCGCAAATACGACGACGTCTACACCTTCGCCAACCGTCGCCACGTCCA
CCAGTACCACATCGACGGTAGATGCGGCGCAGAGGGATGTAGAGTCCTTCACCCTACCTGTACATGA
CGGCTCGAAGGAGAAGCGAGTGGGCCCAGTGGATGATTCTTTGGCGGTTTCGTCCACTCAAACCGCC
TTCAGT

>APBMS113
CTGCCCCTTGAAGGAAATCTGTGCGTAGTCGCCGTCGGCGAGCGGTCGCGACTCTTCAACCGGCTCC
ATTGTCGAGCGCGAATCACGCACCCGCTCCAGTTCGGCGGCAACCTCTTCCTCGGTGACTGTTGCCT
CCGGCCGCTGGACTTTGATCTCCTGATACCCGGTGATGTCAATCTCTGGCACTACTTCAAAGACAGC
CTTGAAGGTCAGCGGTTCGCCATCGACCAGGTGCAGCTCGGTCACCTGTGGCTGCGAGACAGGCTTC
AGATTGCTCTCGGCGATGGCCGCGCGAAAGTGTTCCGGCAGGACGGACTCGACCACATCTTGGCGAA
TC

>APBMS114
ACTTCGAAAGCGGTGATCAACCACTATAATGTCCAGCCCACTTACGACATCTACGCGAATGTAGAAG
GGCGCGACCTGGGAGGCGTGGCGCGCGACATCCAAGCCGTTCTCGACTCCTGGACATCTAAACTGCC
GCACGGCACCACCCTGAAACTGCGCGGCCAAATCGAGAGCATGAACGAATCGTTCATCGGTCTCGGC
GGCGGGTTGATCGTCGCGATTGTGCTTGTCTACTTCTTGCTGGTAGTCAACTACCAGTCTTGGGTCG
ATCCATTTGTCATCATCTGCGCTTTGCCAGGAG

>APBMS115
TCTTCACTTTGACATTCAGAGCACTGGGCAGAAATCACATTGCGTCAACATCCGCAGGGACCATCGC
AATGCTTTGTTTTAATTAAACAGTCGGATTCCCCTTGTCCGTACCAGTTCTGAGTCGGTTGTTCACT
GCCTAGGGAGAGCCGGGGTCGCCCCCGACGTTCTGCGTCCGTCGCCCGGTCGGCACGCGGGGGTCCG
CCCTCGCCGCGCGAGCAG

>APBMS116
ACTTCCTGTGAGAGCTATCTGAGTTTGCGTATGTTTCTGTATGGTCTACAATGAGCCGTAACAATAC
CTCAATCATTGTTAATGTATTGTTCAGAGGAGTACCTAGTCCTTTGCATATCCTAGAGAGTGAGGAC
CGTTACATAATGAAGGACACTGACAGATGGGGCAACCTAGTTAACTACAAGCTCTCTTGGGAAAAGA
ACCCTTTGTTAGTGAAGTCTTTGCAAACGTGCTTTGATTGTCCTTCTAAAAGGTTTACTCCCTACAA
TACCCTTATTGTAGACACCGATCGCACCAGGTATTTCACGAACCTTCCAGAAAATGTTGTTTTCTTC
AAGTTGTGGACTGTCAACGAAGATGGAGCCAAGGACTTAGTCCTCATGAATTTTTTTCTACCTTGGT
TGTAGCGGTTGTATGCCGCTAGAGATAGGGGTATTTGGAATTTTCATTGGGACGATAGCATTGGGGA
ATGCCACTTGTATGAAGAGTCTGACCGATGGTTAGCGAACGATCTAGATAAGGCTCTAATAGTTTCA
GATAACTTCTATGATGAATGGGTTGATAGATATCCCTATGAGAAGTGGGTTTCTAATCTAAAG

>APBMS117
ATTTCAGGCGAGGTTCGGATTCATAGTGTCGAGTGAAAACATTCTCTGATGTTGGAGAAGACAGATC
ACAGAGTTGAAGTTTGTAGCAGCAGGATTCTGGAGTTTCTGACAAGTTTTTGGCAAGTTTGTCGCGA
TGGGGCGTAACCGATACTCGAAGAAAGGCGCATCATCGCAGCCTCGGGCGACGTTCTGGGAAGACGA
GAAGGATCAGCAGAGTTTACCATCGATCACGACGCCTCCGCCTGGCGAGGCGAGCGAGCGCGGTGCC
CCGAAGCGACTCGCCGTTCACCTCGCGATGTGGGATTTCGGACAATGCGATTCCAAGCGATGCACGG
GTCGGAAG

Appendix D

SEQUENCES WITHOUT FLANKS

The following are sequences from both APMS and APBMS libraries that possess microsatellite sequences but short or non-existent flanking regions, making it impossible to design primers to amplify the loci. They are in FASTA format and are trimmed of linker and cloning vector sequences. Where a sequence is derived from only one primer, the primer used to generate the sequence is indicated.

>APMS8
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACAT
ACACACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACA
GAGACAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS10
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACATACACAC
ACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAGAGACA
GAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS15
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGATANAGATACACACACATACACATATACAC
ANACANAGANANAGANANAGANANAGATANAGANATAGAGACAGAGANANAGACANAGACANAGANA
NAGANANAGAGACANANANNNNCNCACCTTCTGNTNTTCTNNTNGNATTCNNGGNTCTGCCGGGNTN
CNATATTTCGATTTTCNGTTTNTNNATNCTCTCGNGNTNGANNGGGGNCCCGNTCCCCNTTTCGNNC
TNTNNNNTGNNNNNTTTCNCTGGCTCTCTGNTTTGNCNNTTTACGACTGNGTGANTGNTNNAACNTC
TNNNNTTTCCTNNCTTTGTCGCCTTTCCCCCTTTCCCCNTTTNGCCTGNTGGCGTAGNGGCCCACAG
NGCTCGACCTATCNCACTTNTNCACAGTTTGAATGGCTGAATGGNATTTGGANANTGTATNTNTTGA
TANNTTGNTCAATTAATTNTTGAATNTTTGGTCTNATTTTTTNATTNTTTTANCCNATATGCNGAAA
ATTCCCTTAATTCCTTTATAAATTAAACCCAATATACCGGTTTAGNGTTNGTTCCGTTNTTNCCANC
TTGGGANCCCTTTTTCNANNTATTTGNGAATCCTGGGCTNCAANGGGGAAAAANNCGNNTTNTCNGG
NTTNTTNNGGCCNAATTGCCNCAATCCNT

>APMS16
CTCAGCCACATCAGTGCCGCCTCCAACATTCTCCTCGTCCTCTTCCTTCTCTTCCTCCTCGTCATTT
TGCTCTTCCACCATTTGATGCAAAGCCTCCTTGATGGGATCGTTCTCAGGGTCATTCTTGGTATACC
ACAGATTGTCCCCACCAGCAGCGGCAATGACCTCCTTCCCTGCCGCAATCTGCTCTTTCGTCAAAGG
AACAAACTCCTCCTCTTCTTCCTCCTCCTCCTCCTCCTCCTCGAATT

>APMS17
CTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TTTCTTTCTNTGNTTTNCNTGNTNATGTNATCTCGNCNTNNNTCGGANCTTGACAATGGNAATGTNT
GTNANNNNNNTGNCTAGGTTAAGACGCTCCACCNN

>APMS20
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGANAGAGAGAGAGAGAGAGAGAGAGAGAGATANAGATACACACATACAC
ATACACACANACAGAGACAGAGACAGAGACAGAGATAGAGATAGAGANANAGAGACAGAGACAGAGA
CAGAGACAGAGACAGAGANANACANACCNNCNCC

>APMS21
CTGCTGCAGAGATGATCGCGCTCTTTGTCGTCTGTTGCGCACAACGAAGAGCATGTGACTCCATCCT
TCTCTCGCATACAACTGCACAGTACTCGTTCACACACACACACACACACACACACACACACACACAC
ACACACACACACACACACACACACACAACACACACACACACACACACACACACACACACACACACAC
ACACACACACAACACACACACACACACACACACACACACACACACACACACACACACACACACACAA
CAACAACAACAACAACAACAACAACACACACAACAACACACACACAACAACACACACAC

>APMS22 --T3 end
AGAGACACAGAGACACAGAGACATAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACANAGAGAGAGAGAGAGAGAGAGAGAGAGATACANATACACACATACACATACAC
ANANACAGACANAGACANAGACANAGANATAGATANAGACAGANAGACANAGANANAGACANAGACA
NAGACANAGANAGANANACACCNNCCNCCNTTCTNTTTGNTNTAGNNNATGCNTCCNCCTGGCTGNA
TNNNTTCTNTATCTTNTTTATCCCTTCCNNCTNCCTNGGGGGGNNGGCCNNNACNTCNTTCTATCTN
NAGTGNGNTTTNTTNCGCTCGCTNNCTGNCCNTTTTTTTNNNNCTGCCTGGCTGNACNNTNCCTGTN
NTTACCTTANTTNNTTTGCTTGNCATCCNTCTTTCTTTCGCTGGCTGGNTNANTAACAGNNANGCCC
NNATCGNTCTTCCTTACNGNTGCTTGCGCTGNCTGGNTGGTGGATGTTGATTGNGNNCNTNTTTNTT
TTNTTTACATNNTCGATTTNTGTTTGNTCANTTCANTTTTTTTTTTAANCNGTCGGACTNNGTCGGN
TNCCTNCNTTATCAATCGAATNAATNGACCTAGGTTTGGNTGANTGTTNCTTTCTTGTTTGNANCAT
GCACTCTTTNATTAANGGNGNGNNTCCNTCNTNCANNNGGAGGGA

>APMS27 --T3 end
CTGCAGCTGCTCCTGAGCCTCCAGTGAAGGAGAGACGAAAGTCAGGGTGCAAGCACTATGATCGCGA
CTGTCTTTTGAAGGTGTGTGCTCTCTTTCTCTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCNCNCNCTCTCNCTCACACNCANANC
NNNNNATTAAAANCNANAGNGNNCNNCGANANNNNGNNNATNTAANNNNNGNCNNCANNNNNNNANG
GATANCGAGANANNNCGGGCNCGNGNGTNAANCNNACTANAANNGTGAAAANAGAAAAAAANGANAA
NATNAANAANGAAANNNNANAGAAANCNCNNTNNNTNGNAATNGNNTNCAAGNTNANNNGAGGGAAN
GNGTNGNTGGNANNGNNAAGNNNATAGANAAAAAAAANNANAAGGNNAANANGAAAAANNGTAANAC
NNANACAAAAGNNAANGAAGANATAGGAAGNNGANGACGTATNNAAAAAAAANGNNAANNCAGAANN
AAANGAGAAAANNCNGAAANANNNGANAACANAAAGNNGANNAANTCAGNNANAGANNNNACNGTAN
CNAGNCTNTNNACACANCGTGNGGAAAGGGATGNGNTGCNNGNGACNAAAAAAANNNNAAAANAACT
NAAAAANANNAAANANATAACAANNAATAAATAAATCNNTANNCGAANGNATAANAANNACAGTTGA
GAATATAGGATAANGANNGNNNAGACCTANNANCAGNANNAGNAN

>APMS30
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACAT
ACACACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACA
GAGACAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS31
GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGAGAG
AGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACAGAGAGAGAAAGAGAGCACACACC
TTCAAAAGACAGTCGCGATCATAGTGCTTGCACCCTGACTTTCGTCTCTCCTTCNCTGGAGGCTCAG
GAGCAGCTGCAG

>APMS34 --T3 end
TCACACACACACACACACACTCACACACTCACACACTCACACACTCACACACTCACACACACACA
CACACACACACACACACACACACACACACACACACACACACACATGCACACACACATGCACACACATGCAC
ACACACATGCACACACACACATGCACACACACACACACACACACACGCACACACACACACACGCACG
CACACACACACACACACACACATGCACACTCACTCACTCACTCACTTGTACCACAATCACACTCAAACT
TACCATCGAACCTGGCTGGCTGGTCAAC

>APMS37 --T3 end
TGGAGCTCCCGCGGTGGCGGCNNNTCTAGCNGAGGCCTGCAGCTGCAGTATGGGATGAAGATGATGA
CGGTGATGCTGTGGTCACGCACACACACACACACACACACACACACACACACACACACACACACACACA
CACACACACACAGACACACACANACACTTACTTACTTACANNCACTCACACACACACATACATACAC
ACACTCACTCACACACACACACACACTCTCTCTCTTTCTTANANACACACACACACACACACACACA
CACACACATACATACACACACACGCGCACACACATTTNTATACACATATNCACATATACACACACAC
ACACACACACACACACACACACACACACACACACACACACACACACACACACACACGCACACGCACACACA
CGCACACGCACACACACACACACACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTNTNTGTGTGTNTGTGTGTGTGTGCTTTCNCGCTCTCGCNCNCGTTTTANACT
NTATCNNGTNNGTGCGNGTGTGGNGTGTNNGCGGGCTNGCTGNCGCGANTAGCGNNTGGAGTTGTGG
GGGNTGNCGGGGGTNCGGAGTNNTTGGATNTTNAGGTNNNGGTGGANNCCGGGGAGGGGGGGGGGGG
GGGGGCGCGGGNTNGCCNNTTTNGNGGAGNTGGTTNGGNNGGGNNTTTTGGTGGGGGGNNGTGTGGN
NTGGTNTTTTTTNTGATTTGGGGGGNGANGTGGTGNAATTTTGGTNTGTTTNTCNGGGTT

>APMS38
AGAGAAGAGAGAGAGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAGAGAGAGAGAGAGA
GAGAAAGACACACACACACACACACACACACACACACACACACACACAGAAATACACATTCTCCCTCATCG
ACATGATGTAACTGCGCACCGCACGCGTGAGCAACCGTTACGATCGCACAAATGTTTTTTAAACACA
CCACACACACACACACACACACTCACTCACCGCCCTTGATCAA

>APMS40
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACAT
ACACACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACA
GAGACAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS43
AGCAGCAAACACATCACACACACACACACACACACTCACACACTCACACACTCACACACTCACACAC
TCACACACACACACACACACACACACACACACACACACACACACACACACACACACACACACATGCACACACACA
TGCACACACATGCACACACACATGCACACACACATGCACACACACACACACACACACACACACGCACAC
ACACACACACGCACGCACACACACACACACACATGCACACTCACTCACTCACTCACTTGTACCACAATC
ACACTCAAACTTACCATCGAACCTGGCTGGCTGGTCAAC

>APMS45
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGT
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACATACAC
ACACACAGAGACAGGGACAGAGACAGAGACAAAGATAGAGACAGAGAGACAGAGACAGAGACAGAGA
CAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS48 --T3 end
CTGGACTCCCGCGGTGGCGGCNCNTCTAGCAGAAACTGCAGCAACGGTNNCCTTTCATATGNAAAGC
CCTGAAATCTGCCTGTTTTTCCGGACGAATAGGGGCTTGATCACCAACCATTACAGACGTATTCGGT
TGCATTCTGTGTGTGGTGTGTGTGTGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCCCCNANNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNANNNNNNNNNNANNNNANNNNANNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNANNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNANNNN

>APMS50 --T3 end
CTGCTCGTCGGGTTGGCAGATGTTGTGGCCGTTCACTAAGGAAACGTTTGCGACTGGTGCGGCTGCC
AAGTGCTGGGCGCACGAGTTGCCGGTGAGCGAATTTGTTTGGTGATGTTCTGTTAGTCTCTCTCTCT
CGCTCTCTCTCTCTCTCTCTCTCTCGTTCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCNCTCCCNCTCNCGCGCGCGANATATATATANNTCNCACACACACACTACNGNTAG
AGAGGGGGGANNCNACNCACNGCANGATANAGAGAGAGAGNNNACNCACNCNCTCGANCTCTAGNGN
NAGACCGGTANNNNAGANAAAAAACAGTGAGNCNAAANANANAANCNCNNNGGACACNNNNCNCANA
NCNCGATNGACNGGAAAAAAAAANNNAGAGACNCACATANATCGCNCTTNAAANACATACGCACTNT
GTGNGGGGGGNGAAAAAANAAGAAANCGCCCNCATANNATNNAANATTCCCAAAAAAAGCTNCNNNN
AAAATGCCNCAATAGTCANTANNAAAGGGGAGAAGATATTNGGAAAAAACCCCCATTAAANAAAAAA
AAAAAAANAGCANAATTTNNNNNAGGGANNNGCCCGAAANNGGNAAAAAANNCCCNTAAATAAAAAA
GAAAAGAGACCCCCACCNNAGNGGNNNNANAAAAACNACCTTTGGNAGGGGGNGCNCCCTATAAAAA
ANCCCACCCCAATNTTTNAANNNTGGGAAAANGGNNTNTCANGACNAAACACNACNCNAAN

>APMS51
AGAGCGAGAGACAGGCAGAAAAACTGACGGAAAGCCATACCGAAAAATGAACTGAAAGACAGACAGA
AAGCCGGACGAAATGACTAAGAGTATAGATGCGGCATGAGACACATTGACACACATGAATGGAAGAA

GACTGAGTGAGCCGTGGAACGCACACACGCGCCGCCCCCACCACAACCCCCAATTCACACACACTTG
CACACACACACACACACATACACATACACATATACACACACACACACACACACACACTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTTTCTCTCTCTCTCTCAGTGTGTGTTTCATGT

>APMS53 --T3 end
CTGCTGCTGTTGTTGTTGTTGATGATGATGATGATAGTGTTGTTGATGACAATTTTGTGGTTTGTTG
TGGTTTGATCTCTCTCTCTCTCAGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCNCTCTCTCTCTCTCTCTCTATATCTATN
TNTATNTCTCTCTCTCTACNCAGNCTACACNCACACNCACACACACACACACACACACACNCACNCACN
CNCNCGNNNNCNCNCACACACACACACACGCANCNCACACGCNTCAATATACACACNACNCNACACA
NACANANANAGANAAAAAAACANANAGNCAAANAGANAGANNNNNAACAANNANANATGNNGGNACA
NAGAGAATACGACANNNGNACANACACACACACACANANACACNCACACGNGACANACANAAANGAN
TGTGCNTGNNANAGAGAGCNAGCGANGAGNNGAAAAAATGAAGCAGAGAGNGAGAGTGGAGNNAGNG
CGCANNNNAGGGGGAGCNNGTGNNNAGAAANAGNGCAGCNCNAGGNGNGNGGCNGNAANGNNGANAN
AAAGNGNNATGNANAANCNACAACTCNAGGGGGGGGAANGAGGAAAAAANANAAACNAAANGAGNNG
CGGNAAAAAGGAGGCNCNCACGGGGAGNGNGGNNTNAACAAAGAGAGAGGAGAGGGNAAAAANNNTG
GGGGNAAANANAAAAGTGAAGGNGCACGNGGNGGACAAACNCAAGATTAANAGGGANNGAGAGGAAA
TAGGNGGAAAAAAGACG

>APMS54
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACATAC
ACACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAGA
GACAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS55 --T3 end
CTGCTGCTGTTGTTGTTGTTGATGATGATGATGATAGTGTTGTTGATGACAATTTTGTGGTTTGTTG
TGGTTTGATCTCTCTCTCTCTCAGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCNCTCTNNANCNNGCN
CNCNCNCGCNCNCNCTCNCNCNNNNNCNCNANAGNNAANGNGNNNGGNNAAGCGAGNGNGCGCNCNNN
GANANAGAGCNAGCNANANNNCNCNCNCNCNNGNGNGCGCNAACGANCNNANACNCCCNCACACANN
AANCACAACAANAAAAGAAGAGNNGAAGNGAGNGAGAAAGAGANAGNGAGCGNAAAAAAAANCGNAA
AAAANNGAAAAAAAAAACACGAGAAAANAAGNGAGANANANANANAGGNNNNNNGNNAAAAANNGAN
AAAAAAAAAAAAANAGACNANANANAGAAACAAANAGNNAGANNCNGNGAGAAAAAANAGAAGNGNNA
ANAAAGAAAGNAANATCANAAANAAAAAGAAANAAAAAAAANGNAGAGANAAGAGANAAACCNANAAA
AAAANAANNAAANAAGGGAAAAAAGNNAGANNAANAAGANAAAANANAAAACAAANNNAGAANAAAA
NAAGANAANCNNANAGAAANANAAAGAANNAAAAGGNG

>APMS59 --T3 end
CTGCTGCTGCTGTNGATGTGTCGGTCTCGCTGGGTTCGGTCTGAGGGGAAAGAAAAGAGAGAGTGAG
ATGAATGTAATGCTTGCTAATTGTGATTGAAGATCTGTTTGCGAGCGTGTGTCTACCTTTCTTTCTC
CCTCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTNTCTCNNTCNCTCTC
TNNTCCTGNNGNAGCNANTNNGGNTGATTNNCNANTCNNCCCTNNTNCNNCNNGNCGNTGACCNAAN
GANANNNANNNCNANNTNNTGNCGNNGGGTNGGNNCNGACAGNACNNANNANATCANNANAAAGNNN

ANNAATAAATATNTATTNNTNAGANNTCNNTCNCGATNCCCNNNCCNANTNAAACNNAGANNAAACA
AAAANCTNANNNNNCATNNAGAAANNCGAANCAAAACAGGNNNAAANAANAANAANCTANNANAANT
AGANANNNANAANTCNNACACCNNNAGNANAAAANAANAGNGNGGCCNNNNNACNGCNANGNNCNNG
CNGCNTANACAACCGCNGNNCCCCNNANTNGCAGCANCGNTNNNGCGATNTAAGATGNANGNTTNCN
NCGTCCNANNNNGNNNNCANNANNACTACAACNTAATAGNTAANNNNNNNNGAANATNNATGTGCAC
AATTAAANNNNTATNAACCCGAGAGCCTGCGNCTNGCNCCAGACAGACAANCCCANCCNGCNACGCN
CNTNGACANNANNCNCGNGNCAATANCGGGGTNNGTNTACCTNNNNCTTNNNTANNNACGNNTGGAA
GNTANANNACANNTNCANANCNACCNCANGAATGNGNNACNANNNNNNANCNNGNAGGTGTNNNNAC
NTAAANTCAANNATNNCTNAANCNACCAACANCACANNAGNTNNNTAAGTNTGTNNCTTCANACTGN
GNGA

>APMS63
AGAGACACAGAGACNCCGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACACATACAC
ACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAGAGA
CAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS64
CTGCTGCAGAGATGATCGCGCTCTTTGTCGTCTGTTGCGCACAACGAAGAGCATGTGACTCCATCCT
TCTCTCGCATACAACTGCACAGTACTCGTTCACACACACACACACACACACACACACACACACACAC
ACACACACACACACACACACACACACACACAACACACACACACACACACACACACACACACACACAC
ACACACACACACACACAACACACACACACACACACACACACACACACACACACACACACAACAACAACA
ACAACAACAACAACAACAACAACAACACACACAACAACACACACACGACAACACACACAC

>APMS65
AGAGAAGAGAGAGAGAGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAAAGACACACACACACACACACACACACACACACACACACACACACACACACACACACACAGAAATACACAT
TCTCCCTCATCGACATGATGTAACTGCGCACCGCACGCGTGAGCAACCGTTACGATCGCACAAATGT
TTTTTAAACACACCACACACACACACACACACACTCNCTCACCGCCCTTGATCAA

>APMS67 --T3 end
GGCTNNTTCTAGGGGGAGNCTGCAAGCGTGCAGTGTGGGATGAAGATGATGACGGTGATGCTGTGGT
CACGCACACACACACACACACACACACANACNCACACACACANNCACNCACACACACACACACACANAA
ACACACACTTNCTTNCAAACACTCACTCACACACACACATACACACACACACTCACACACACACACA
CACACTCTCTCTCTTAAAAACNNNCACACACACACNCNCACACACACACACANANNCACNCACACAC
GCACACACACATTCATACACACATNCCCACNTCCACACACANACACACACACNCACNCNCANNCNCG
CACACNNNCNCAAACANACACANNGTNNNNGCAGACNCACACACANACNCACACNCNCNCNCANACC
TCCTCTTCTCTCTCTCTCTCTNATCTCNTCTCTCTCNNTNATNNGGGNGGCTNCTNCNNCTTCTGGA
NTNNTGTGTNNGGCTNCTGNTGAANGGNGGANACNNTGTNANANGATTATTNGGTNTTGNNTNTNGC
TNNTCTGNTTNTNCCNCNCTCAGCGGGCGNNNCTTNNGTCGNNNTNNAGAGNNTNGGNGTNGGNNNA
GANGCNGGCGNTCTNGANANNNGTNCNNCCCANNNAGNGCGACNNTCNNACCTTNGNNTATGNNGCC
NGCGNGGNGCNNNNNGANCGGAGNGTNCGNGTTNTNTGNGATNGGACNNTGTNNGANNNNGTGNNTT
GNTAAGCANCTNGNACTNCCANNTGAAANAAAANTTNTCTGAGNGGTGNTGTGGGTNNGAGGTNGAC
TTNGNNNATGGNGGTNANNTNANCGCGNCNGNNNTCGNGGNNTGNGACNTGTCCTCCANGATGNANG
TTNTCCNTNTNNNGNGTATNCNANGNACACGNNTGTANCNATNNNGAGTNGGGGGGGTTGTTNGGNG

NGTGANTGTGG

>APMS69
AGAGACACAGAGACTCCGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAATACACACACACACATACAC
ACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAGAGA
CAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS71
AGAGACACAGAGANNNNCCAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAG
AGACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACATA
CACACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAG
AGACAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APMS73
AGAGACACAGAGACACAGAGACAGAGACACAGAGACAGAGACTGAGAGCGAGACAGAGAGAGACAGA
GACAGAGACAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACACACACACATACAC
ACACACAGAGACAGAGACAGAGACAGAGACAGAGATAGAGACAGAGAGACAGAGACAGAGACAGAGA
CAGAGACAGAGACAGAGAGACACACACCGCCGCA

>APBMS6
CTGCTCCGGGTACAAGTAGCCAGACAGATGAATGGATAGAAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAGAGAGAGAGAGAGAAGAGAGCGGAAGAGG
AAGTGAAAGAGGGAGCACAAAGAGAAGGAAAAGAAAGAAAGAAGAAACAAAGAAAGCACAAGGAAAA
ATGAACC

>APBMS16
CTAGATCAATGCGATGGCCAAGGGGGAGGAACCGAGGAATGTGGACTGTGATGGAGAGGGGAAATTC
TGCTGCTCAAGGAAGCCAAAAGGGACGGACGGAAGAAGAAGAAGAGGACAGAGGAATCTGGGAAAGC
AGCGATGGGACTGTGGGAGTGGGCGAGATATTTTAGACTTTAGAGATACTGAGGGAGCGACGGCAGC
CAATGAGGCCACTCGGGAGTCTTCCTTTTGCTCGCCCGAGCGGGGCTTCGGAATTTCCGAAGCGGAA
GGGAACT

>APBMS17
TGTTCGCTTGTTTGCTTGTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCNCTCNCCCNTTGATGTNTNTGTNCTGNTACCNTGATGATGACGCAGTGAGCGGG
CTAAGCTGACACCCCANGTCAGCGTCNCAGTCNACACGAGCAG

>APBMS19
ACTTCAAGGTCTCAAAGGGCAAGCCTTGCTGTGTTTGCCTTCTTCTCTCCTCGTCTTCTTCGATACC
CCTGACTCTGTGTTCTTCTTTCGCCTCGCTCGCCGCTCTTCTCATGGGCCGCCCCTTGCCGTCCGAA
TTGGTGGATCTGGTAACTGTGTGCCTCCCTTCTTCGTGGTTTCGCTTCACAAATTCTTTGCAAATTT
CTTTCAATGGCGCATAATTGAGAGTGCTCACGCCCAAAATGATGCTAAAACGAGAATGTAGGGCCGC
GGACTGATCCATTTCTGCCCTGTCTTGATCGTGAG

>APBMS93
CTAGAGCAGAAACACACACAGACACCACACTCAGACACACACAGACACAGACACACACACACACAGA
CACACACAGACACCCACCCACACACACACACACACACAGACACACACCCACACACACACACACACAC
ACACACAAGACAAAAAGACAACGCCGACCTGAAAGTATTTGGGCAAGGTTTTGAGGTCGTTGGCCTT
ATAGAATC

>APBMS96
CTTTTTCTGCTCGGCAATCGGTGATCTCGTGGCGACCCGATTACCGAATTTGAGTGTCAGTCGTGGT
ATTACCAAGTGGATCGAAATAAATCGTGTCAACCTCTGCGAGCAGGAGAGACCTCGAGCGATCACAG
CGATTCTGCGCGATCGTCAAGTAGGATCGATTATTCCGATCAAGGCCTTTGGATAATATGAGAAATC
TGTCGCATTCTGATAGTTATCGCCAATAATGTACGATGGTGGGACATTGCCAGTTGTCCGTTGTCCG
TTGTCCGTTGTCCGTTGTCCGTTGTCCGTTGTCCGTTGTCCGTTGTCCGTTGTCATGGTCAGTCGTC
AGTCGACCGTTGCTTGTCGTCGTCGGATCGAGCCTGCATCGAGAAACGTCACGAACGAAGCCACTCT
CTTGTCCGTTGTCCGTTGTCAGTTGTTAGTTGTCCATTGTTGGTTGCCTATCTCCGAGTCTGCACCG
AGAAATGCGCCGAACGAAGCCATCTATGAGAGGGCCTTTGTCAAGCGAGGCGGTGCAGATTTTGTGT
TTTGG

Appendix E

SEQUENCES WITH FAILING PRIMERS

The following are sequences from both APMS and APBMS libraries that possess microsatellite sequences, enough flanking regions to design (and redesign) primers, but whose designed primers fail to amplify. The sequences are in FASTA format and are trimmed of linker and cloning vector sequences where possible. Where a sequence is derived from only one primer, the primer used to generate the sequence is indicated.

>APMS2
CTTACATGTGATGAGATCGTCTTTCCAGGATCTCAAGGACATCGCCGACTGCATTCTTGACGCAACG
GCAACAGGATCTATAAAGGTTATTTCGAATATACGTGTGTGTGAGCGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTATGTATGTGTGTGCGTGTGCACAAAATGGAAAAACTTTACTCGC
TGACTGCAAATGGGTTGTACATGTGAAGATGCTAACATTTTTCTCTCTCTCCCACCAGTCACGATGG
AAT

>APMS11
TTGGTATGCATGAGATGAACTGTTACCTTTCTTGAGTAATCACCTTGACATTGACTTTGAGGGGATG
TGACTGTTGCCTCTTCAAAAGCACTGCACTGGAAATGTACATGTAAGAAAAAAAAAGTACAGTGTAGA
GAGAGAGAGTGTGTGTGTGTGTGTGTGTGCGCGCGTTTTCCCCTTCTTATAATTTCAATGTTTGCCT
TGAAAACCAAATGTTGGACATGTACACAG

>APMS18
AAGCGGAAGCCGCTGCGAGGTCTGTGTATCCCAGTCGTGTGACTGATGTGTGTGGGATCGTGTGGTA
TGCCATCACCTGTTGTGGCGGTTGTGTGTTGGGTTGTGTGTGTGTGTGTATGTGTGTGTGTTCAGGG
ATGAGGCAGAGTGTGAGCGTCGTCGCGCCGCTGCTGCTGCGGTGCAACAGCCGGACGCGAAGGCTGC
GTTGGCGGCCCTTGCGGCTGACAACAGACGCCGTGCAG

>APMS23
CTACACACCTCTGCACACCTGCTGTGTCTCCTTTGGTATGTTGCTTGTTTCTGGGGGCAGTCGAGCA
GTCATTCTCTATTTCCCCCTTTCTCCCTCTGTCACACACGCACACACACTCTCTCTCTGTCTCTCTC
TCCCACACACACACACACACACACACACACACACACACACACACACACACACACGCATGTGCATCCGCACCCG
CTTTAGCTGGTTGAGTGTAAGTATCGGTTTTTGCGTTCCAATGTGTGTTGCGGAGGTGATCCCATTT
TTGCTTCCA

>APMS24
CTGCAGCTGCAGTATGGGATGAAGATGATGACGGTGATGCTGTGGTCACGCACACACACACACACAC
ACACACACACACACACACACGCACACACATAGACCACACACACACACACACAGACCACGCACACACA
CACACACACACACACACACACACACACACACACACACACACATCACACACACACATCACACGCTCACCTGCTG
GAGCAACCGCTCACGATCCTGCTGCAG

>APMS33 --T3 end
GANTGTGACTCCATCCTTCTCTCGCATACAACTGCACAGTACTCGTTCACACACACACACACACACA
CACACACGCACACACACACACACACACACACACACACCTGCACACACACACGCACACACACACGCAC
ACACACGCATACACACACACACACACATACATACACACACACACACATACACACACACACACACACA
CACACACACACACACAGACAGACACACACTCAGCACACATACAATCCAATCTAAAACGGCCGACAAA
AGCAACCGTATTGCTGCTGCTGCCACTGCTGCCACCATGTCGGTCGCCGATGTCGGGCAGTCTGTGC
AGCGGCCGGGCTCGCGTCTGAGCGGCAAG

>APMS35
CTGCACGGCGTCTGTTGTCAGCCGCAAGGGCCGCCAACGCAGCCTTCGCGTCCGGCTGTTGCACCGC
AGCAGCAGTGGCGCGACGACGCTCACACTCTGCCTCATCCCTGAACACACACACATACACACACACA
CACAACCCAACACACAACCGTCACAACAGGTGATGGCATACCACACGATCCCTCACACACATCAGTC
ACACGACTGGGATACACAGACCTCGCAGCGGCTTCCGCTT

>APMS46
AGAGCGAGAGACAGGCAGAAAAATTGACGGAAAGCCATACCGAAAAATGAACTGAAAGACAGACAGA
AAGCCGGACGAAATGACTAAGAGTATAGATGCGGCATGAGACACATTGACACACATGAATGGAAGAA
GACTGAGTGAGCCGTGGAACGCACACACGCGCCGCCCCCACCACAACCCCCAATTCACACACACTTG
CACACACACATACACATACACATACACATATACACACACACACACACACACACACACTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTTTCTCTCTCTCTCTCAGTGTGTGTTTCATGT

>APMS47
TGGAAGCAAAAATGGGATCACCTCCGCAACACACATTGGAACGCAAAAACCGATACTTACACTCAAC
CAGCTAAAGCGGGTGCGGATGCACATGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGGGAGAGAGAGACAGAGAGAGAGTGTGTGTGCGTGTGTGACAGAGGGAGAAAGGGGGAAAGAGAGA
ATGACTGCTCGACTGCCTCCAGAAACAAGCAACATACCAAANAAGACACAGCAGGTGTGCAGAGGTG
TGTAG

>APMS70
AGCAACAAGCAGCAGCCGCAGCAGCAGCAGCGGCCGCAGCAGCAGCAACAGTCGTAGCAGCGAGAGT
AAACCTATAACACTCGCAAGTGCAAGGATAAGCAGAATCAAAAGGTAAAAACACACGATGCGGTCCG
ATGGAATAATGCACATCGAGAGTTGAAGACGCGTGTGCTAGTCTGCCATCAATTGAAACTCGCACAC
TCACACACACACACACACACACCACACACGCACACGCACACACACACACACACACACACACACACAC
ACACACACACACACACACACACAACACACACACACAGACAGACATACATACATTTACTTACTCACTC
GCTTACCCACTCTCACACTCACTCAATAACTCACTTGCTCACTCACACCTTTCCACTCATACTTACG
CACACTTCACTTTCTGACCACAAACAGGATTTGACATGCCCGTCTTCGAAGGTGGCCCGCACAAGGT
CTGGGGCCTGACAGCCTTTGTGCTATCAAGTCTGCTCCACACTCTGGCTGACCCATCCGTGAAAAG

>APBMS2
AGTTCGAGGATTTTACCGGGTATCTCTTCTTCTATTTACTTACTTTCTTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCGCGCTCA
TTCTCGCAGGGCCTTGGGACAAATCNGGTCCGAACGGGGCCGACCTCCGCNANCACNCTCACTTCCT
ATGAGGCNTCTGCTAGAGCGGNCGCCACCGCGGGGGAGNTCCAGNTTTTGNTCCCTTTAGTGAGGGT
NAATTGCGCGCNTGGGGNAANCATGGTCATAGNTGTTTCCTGTGNGAAATTGTTATCCGCTCACAAT

TCCACACAACATACGAGNCGGAAGCATAAAGNGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTNGGGAAACCTGTNGTGNCAGNTGCATTAAT
GAATCGGCCAACGCGCGNGCAGAGGCGGTNTGNGTATTGGGCGCTCTTCC

>APBMS8
CTAATTGAACATAAATCCGAGTCTTTCATTCAAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGAGAGAGAGAGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAGCGTTATAGTCT
GGGAGCAGAGGCTTAGGGAGAAGAGGAGAGTAAAATAAAAGGGTTTTTCCACGTTAATTTATTTCCA
GGTTTCGAAATTCCTGGGTGTGTTTTTTGTCGACGCTAGTTAAAGATACACCTGAAGA

>APBMS4
TCTTCATAACCAATTCACTGAGAGAACCGCGACATCCCTTTAAAAACAAAAATACAACAAAAACAAA
AACATATTCTCATCTACAGTGAGTGATCTCGATTTCAAATCTTTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTGAGACATATATGTGTGCGTGTGTATGCAGATTTTTACAAGAAAAGA
GATTTAACAAAGAAAAAAAACAAAAAAACAAAATAAGAGTAGTGACCTCAGCAAAGATGGTTTGTTT
GGCTCTGAGTTCAACTCCACGAGAGAAAGTGTCACGAATTGTTCCATCCTTGCCAATGCCCACATCA
G

>APBMS32
CTATGGAGATTGGGAAAACCCAAGGGGGAGGGAGGGATGGAGGGAGGGAGAGAGAGAGAGAGAGA
GAGGCCGAAATCACCCGCACGAAAGCAAATTTTTTTTGCAATGGTACAGCGAAAAAAAAGAGGTTTT
CTTTCTCTCCTCTGTCTGGCGCGGCCAGACCAACCGCTGGCGGAACTTAATTAACTACTGTATGGGA
GAAGACCGAGGCAGACCTTCTTCTTCTGGCGCGCGCGCACTAAGTCATCTACGGCATGTATGTTGCG
TAGGCCCGACGACATTATCATCCATAGCATAATGATTTTCTCATATCTCATATCTAAAATCGAGCAC
GAAAT

>APBMS47
TTTTCGACATCCCTCGTATTCCCTCGACACATCCCTCGCCATTCCAAGACCATGGCCTACTTTTCGC
TTCGAGATTCTCAGGAAGGCCGTAGTTTGAAAAACTATTACCCAGATGAGAGATGTGTAAATGTGAA
CTGACAACTACACGAGCCTGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCCTCTCTCTC
TCTCTCTCTGCATCTATCTCCCACTCTCTCATGTCCGACGGCTGGTAGCGTCAAGAGAGGAATCCTG
TTTGGTAGCATGGCAGTAGGCATAGAGAGGTTCTATTTGTAATTGTCTCATTTCCAG

>APBMS51
ATTTCAGAGTCAAATGTNCTGTGTGTGCGCCGCCCAAATACCTATAATAAGCAAAATGGGAAGCAAA
AGGGGGGGGGGGAAGCAGAAGAGATATAATTAAGTAAATGGCCGTCTTATCTTGGGGTTAGGGGCAA
CGATGATTAAGGATACAACACCCGTTTTTTTTACCATAGTCAGAGTGATNGCAGCAATTGGAAGACC
GCNTTACATAATGTTCGACTTTTAGGGTCCATACTATTATTTCTCCTCNAGGAACTCNCGAATCTGN
ATTGGTNGGTNTCCTCTCTCCCNNAGGNGAGGGAGGGGAGAAAGGGGAAAAGATTGACTACCTAAAC
TTCCACGGTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATCCAC
ACTTTATAGTCATTTTTGGCNTTTCGGCTGTGACTCCCCCCCTTTTTTTGGTTTAAGCCATGGTNG

>APBMS52
TTTTCGTTTGCGTGTCGTGTTTCGTTATGGGTCTGTTTCTTATGTGTATTGCTACTCACTCTTGCCA
ACCTAACTGAACTAGCACTGCTATTTATAGTGTACATTGTACATTCTTACTTTGTAATCCTCATGTT

CTCTCCGGTGTATTCGGCAGTTCCCCGTTTCGAACATCATCTTGTCGTATTGTTTTAATTGGGCAAT
CAATTATATCTTTCAGCTTCTTCCTCTCTCTCTCTCTCTCTCTCCCTCTCTCTCTCTGAGGCTTCTG
CTAGCAGAAGCCTACACGTCTTATTTGCGTTTGCCCCATGACGTGGGTCTTCGGAAATATTGTCGTA
AGTGACACAGTGGTGATAGAGGAGTCCAGGCGCGCTTACGAGGTTTGAAGTACCGTCGTCATTTGGG
AAACATCCCCGGCACGGCGGCGCTAGATCGCAGCAACTCGCCGGATACCCGCCCAGCGCATAGCCAG
CGCCGTTGTCTCCAGCGTCGCTTGGCGGCGCGGAAACTGTTGCACCTGAAATTAAAGCGACAACCGC
GAAAATGATTGTTTTCGATTTCATCTTCATGAATAATCTCCTAAAAAGAGCCTAAAATCGGTCACAG
TTCTCTAATGCGCGGGGACATTGACAAAAATCGCCCAATATTACTTAACGGAGGGCTCTTTCACAAT
CACGGAAACTGGGATCCGGATTTGTGCGTCAGCGGCACCCGGCGGGGTTAGAACA

>APBMS56
GGTTCGACGCGTCGCCCGCCCCGACGTTCGTGTTCAGAGAGGCGTTGATGGACGTCATGTCCACTTG
CGGCTTCATGGTCAGAGCGGCTTTCCCTCGCTCGATCTTCAGCACGCGCACCGTCACCTCCTGACCC
ACCGTCAGCAGACTGTCCACGGACACGTTCTCGCCGCCTTCCGTGATCTCAGACCCGCGCAGGAAAC
CCTCCTCGCCCTCCTCCGCGAACGCGATGAACGCCCCATTCCGGATGATGTTCTTCACCGTACCCTT
CAGGCTTCTGCTAGCAGAAGCCTACTGTCTAATGGAGGGGGGGGGGACAAGGGAGGCCTAACAGGAA
TGGGAGCGGGATTTGAAAGAGGAGGCACTAGACAGGGCTCATCAAAGAGGAGAGGAGGGATCAAGAG
GGAGGAAAGGTATCAAAGAGTGGATGATCAAGAGAGAGAGTGAGAGTGAGAGTGAGAGAGAGAGAGA
GGTAAAACAACGCCCAATATGGGCTGGAGTACTTACAGCATGTACCAACGTTTCAAGTCTTAGTTAT
ACACGATATCAGATCTCAATCATACTGCACGGACTACTGCCATCTCCACAAGTCAAAAAAGAGCAAG
AATGGTCCATTCAACCAGTTGCATAACCATACAGCATCAGGAGTCGGCCCATTCAAAAGTATATTCA
TATTGACGCTATAGAGGAATTTATTTTGGTCCACGTTCTATAAGTCTGGTCTGTTGGCCATGAACC

>APBMS68
CTAATTACTCTCGTTCTTCTCCCTAACGTGGACTAATTGGAGAGAAAGAGAGAGAAAGTGAGGTGAC
ATTGACGCCAAACTTTCCCGAACATCAGTTGAAAAGATGAAGCATAGTAGTATTCGATCTTGTTCGA
TTGTGAAAACATTGATTGATGTGTTAAGGAGAGAACAAGACAAGAGAAGGAAAAATAGCCATTCGGC
TTAACACTCTCTTCTGTTCCTGTTCGCAAACGGAAGCGAGAGAGAGAGAGAGAGAGAGAGAAAGCGT
TAATGAACACTTTTGTCATTCTCCTTCTCTATTATGGTCAAGAAAAGACAAGTGTTGATATTTTTCT
ATTCTTAGATATTCTCTCGAATGTATCTTCATTTGTGTTTGTTCTTTTATTTCTTCCAAATTTGTCT
CTCCACAACATCCATCACAACATTTCCCACTGTTAATCTCACTGAAAA

>APBMS71
ATTTCAAATTGGGCATTGCAAGGGCCTATAGCATGGAACATATTGATGTGAACAAGATGTTGCTTTG
ACAAGTGAAGTCTCTCTCTCTCTTTCTCTCTCTCTCTCTCTATATCTTTATTTCCAAAACTAAATGT
ATGGATACATAATTTGTGATAGGAAAACCCTCAATTAAGCATGCATGTCAAGAGAGGTCAAAGAACT
GAACCTATTCATTTCATAGACCACACTCTATAAAATATAAAGGGGCCAATACTACAAGAAGTATTTA
G

>APBMS105
CTGGAGGATTTCCTGACCGTTTCCGCAAGGCCAGACGTTCATCCACACTCAGATGCTTCTCACCGCC
TTCGCCATCACTTAAGGAAGCTACTTTATTGTTTTTCTTCGGATGATCTTCTTCATGATGACGATGA
TGGCGGTCCTCGGTCTTCTCTTCATGATTCTCTTCGGTGTATTTCTGGTAAATCAGCAGATCGGCGG
TGCTGGTTCCTCGATCTTTCATTGCCGCCTTGGACTTGGAAATTCCTCCTCCTCCTCCTCCTCCTCC
TCTTTCTCTTTCTCTGTCACTGTATTGCTGCTGTTGTTGTTCAGTATCCGAAACCGAGGCGAACA

Appendix F

ACANTHORRHYNCHIUM PAPILLATUM MICROSATELLITE MARKERS

The following are sequences submitted to GenBank (Bilofsky & Burks, 1998) from both APMS and APBMS libraries that were used as microsatellite markers. They are in GenBank flatfile format showing various information including GenBank accession numbers and publication data.

```
LOCUS       DQ092744                207 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APMS14 microsatellite sequence.
ACCESSION   DQ092744
VERSION     DQ092744  GI:70797658
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 207)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..207
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APMS14"
     repeat_region   51..143
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ag"
BASE COUNT       80 a     23 c     73 g     31 t
ORIGIN
        1 gcagcaatat taacctcgct cgacgagtaa caaatatata cataagttaa agagagagag
       61 agagagagag agagagagag agagagagag agagagagag agagagagag agagagagag
      121 agagagagag agagagagag agaaggatct cgagctgagg gctggacgga ttcgaagtga
      181 atttttctgt tgtcccaggc ctttcag
//
LOCUS       DQ092745                258 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APMS4 microsatellite sequence.
```

```
ACCESSION   DQ092745
VERSION     DQ092745  GI:70797659
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 258)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..258
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APMS4"
     repeat_region   25..68
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ag"
BASE COUNT       77 a     54 c     58 g     69 t
ORIGIN
        1 ctgcagcctc ctaccagcat tgaaagagag agagagagag agagagagag agagagagag
       61 agagagagtg tatgaatgaa atcagggtta acttaactta acctgtcata gattatgttg
      121 acttcaaatt cttcgcacta tttcatccaa tcaaaacctt tttctgccca gcccttacat
      181 gcactgaatt tataatcctc cccaggtcag gtcaggtcag gtctgatctg gtctggtctg
      241 gtcttgaact cctgaaat
//
LOCUS       DQ092746                374 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APMS28 microsatellite sequence.
ACCESSION   DQ092746
VERSION     DQ092746  GI:70797660
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 374)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
```

```
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..374
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APMS28"
     repeat_region   77..123
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ga"
BASE COUNT      105 a     85 c    108 g     76 t
ORIGIN
        1 aattcatcag tgaacgcaac agtaaataag ctgcgacact ttgccgacta cagttgtaac
       61 aagtgtgtgc acagatgaga gagagagaga gagagagaga gagagagaga gagagagaga
      121 gagcaagtta cgtgatcaga cctggacgga agaatgtcag cagctacggg actaacagtg
      181 ccactaggca cttgccccgc agaacctgtc catggctgtt gacctggtgc accactgttg
      241 cgtaacttca acgagctttg aggcactaga actctctggg gagaaccctc aatacgccgg
      301 ctgctccgag gaacattgtt ggttacccca ttctcttggt gacttcctgt gggaggagcc
      361 tcggaatgca tgct
//
LOCUS       DQ092747                 330 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APBMS3 microsatellite sequence.
ACCESSION   DQ092747
VERSION     DQ092747  GI:70797661
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 330)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..330
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APBMS3"
     repeat_region   226..266
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ag"
BASE COUNT       93 a     53 c     84 g    100 t
ORIGIN
```

```
        1 ctacttgagt tagggagtgc gtactaagtt ttaatcattt ctaggtcagt tgcagatcca
       61 ctgagttgat atatatactg tactatccag tgctgccctt ggcatttggg gccggatgac
      121 tgaattaatt tatgatgtaa aatcggcaat gacagttgct tacaacttct acggtgctct
      181 ctttcgtttg ttcagacctt ttgtacgtac taaagcaaat taattagaga gagagagaga
      241 gagagagaga gagagagaga gagagatgaa ttggtataat tgcaagccga ttggcggtga
      301 actagcggac tgcatgtctt cctttacgta
//
LOCUS       DQ092748                 327 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APBMS72 microsatellite sequence.
ACCESSION   DQ092748
VERSION     DQ092748  GI:70797662
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 327)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..327
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APBMS72"
     repeat_region   76..135
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="tc"
BASE COUNT       45 a     99 c     52 g    131 t
ORIGIN
        1 gtttcttttt tccttaattt tttttattct tcaaaatgat taatctgtag gcaatctctt
       61 gcgtttgctt gatcttctct ctctctctct ctctctctct ctctctctct ctctctctct
      121 ctctctctct ctctccgttg ctctccaagt gtggacgctc ttcttgcgtg ctaaatcata
      181 taataagtct ccctccctgc cccctccccg ccccatcaat tctgtttctg agttgtgttg
      241 tatctcagtg ccgttgttgt tgtgacactg ttgctcatcg tcgtgggcaa atcgcttgca
      301 gccaacaacg ccgtatgctg gtcgtag
//
LOCUS       DQ092749                 119 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APBMS14 microsatellite sequence.
ACCESSION   DQ092749
VERSION     DQ092749  GI:70797663
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
```

```
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 119)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 119)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..119
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APBMS14"
     repeat_region   34..61
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ga"
BASE COUNT       33 a     15 c     56 g     15 t
ORIGIN
        1 ctggaaggct tgcgaagccg gtggtaagtc ccggagagag ggggagagag agagagagag
       61 acggtgaggg agaacgagcg cgagcgagtg tatttggagg agagaggacc ggtggaatt
//
LOCUS       DQ092750                 191 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APBMS23 microsatellite sequence.
ACCESSION   DQ092750
VERSION     DQ092750  GI:70797664
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 191)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 191)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..191
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APBMS23"
     repeat_region   116..147
```

```
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit_seq="ga"
BASE COUNT       54 a     48 c     52 g     37 t
ORIGIN
        1 tgttctccgt gttacatctg ctccgcccaa aacaggggcg aaccgcgaat gcaagcatca
       61 gtcctcgccg ctcgaaatcg tcagtacttt ccagcgaacg ggacaatctg atggtgagag
      121 agagtgagag agagagtgag agagtgagca atcatcatca acgacctcga actcagagga
      181 tctcgtgaat c
//
LOCUS       DQ092751                 399 bp    DNA     linear   PLN 22-MAY-2006
DEFINITION  Acanthorrhynchium papillatum clone APBMS61 microsatellite sequence.
ACCESSION   DQ092751
VERSION     DQ092751  GI:70797665
KEYWORDS    .
SOURCE      Acanthorrhynchium papillatum
  ORGANISM  Acanthorrhynchium papillatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Bryidae; Hypnales; Sematophyllaceae; Acanthorrhynchium.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Development of microsatellite markers for the tropical moss,
            Acanthorrhynchium papillatum
  JOURNAL   Mol. Ecol. Notes 6 (2), 396-398 (2006)
REFERENCE   2  (bases 1 to 399)
  AUTHORS   Leonardia,A.A.P., Kumar,P.P. and Tan,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2005) Department of Biological Sciences, National
            University of Singapore, 14 Science Drive 4, Singapore 117543,
            Singapore
FEATURES             Location/Qualifiers
     source          1..399
                     /organism="Acanthorrhynchium papillatum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:142992"
                     /clone="APBMS61"
     repeat_region   260..302
                     /note="microsatellite"
                     /rpt_type=tandem
BASE COUNT      101 a    146 c     53 g     99 t
ORIGIN
        1 agttctgcac aacccctgca gttcggggta ataacaaaac ctctgatcct aagagcagta
       61 gcaacattct cacgagccag aatcgcaccc tcgagaaaat gtagagaacc gcatcattct
      121 aaaactgaag ttcctccact cttttccagc acattcatgg atcgttatcc cccccaatcc
      181 gacaccttcc accgttctca ttttctcgtc gctcttccct cggtactctc accttcatct
      241 gcttctctcc gcgctttttc acacacatct ctctctctct ctccttcccc cctcctctct
      301 ctccctaact atccgcacaa aaagccagga ttcgcaacaa gattctaaga agcacctcca
      361 cgaagcccaa caccaaagga cttcaacgcc ctgccaaag
//
```

Appendix G

CHARACTERIZING CANDIDATE MICROSATELLITE MARKERS

Table G.1: Microsatellite marker alleles of 98 samples of *Acanthorrhynchium papillatum* used to characterize the candidate markers. Header rows indicate names of candidate markers. Alleles are PCR amplicon lengths in bp.

Sample	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APBL06	174	78	122	95	127	108	237	79
APBL07	172	81	115	93	124	124	237	79
APBL08	205	83	107	124	126	120	239	111
APBL09	176	75	82	114	127	108	235	79
APBL10	176	74	81	114	128	108	235	79
APBL11	175	87	132	128	126	120	238	90
APBL12	199	83	97	124	94	126	233	111
APBL15	176	81	108	102	122	108	237	83
APBL17	175	85	108	95	123	108	237	79
APBL18	170	86	79	95	125	122	235	106
APBL23	174	83	124	95	124	108	237	79
APBL41	177	81	116	114	123	120	237	85
APBL43	175	82	87	104	126	122	237	79
APBL47	166	78	101	106	115	108	237	79
APBL51	202	82	103	124	125	120	239	111
APBL54	174	78	115	102	130	108	237	79
APBL59	175	86	129	128	125	120	238	69
APBL65	217	82	126	132	126	126	233	90
APBL66	175	126	136	117	114	122	237	80
APBL68	175	88	86	124	124	108	237	79
APBL69	169	78	128	95	126	108	239	79
APBL72	197	83	79	126	126	124	239	90
APBT01	173	78	98	124	99	122	239	79
APBT02	173	78	97	124	99	122	239	79
APBT03	173	78	98	124	99	122	239	79
APBT04	173	78	98	124	99	122	239	79
APBT05	173	77	98	124	99	122	239	79
APBT12	173	79	98	124	99	122	239	79
APBT13	176	87	142	97	123	120	237	79
APBT15	169	84	90	163	126	124	237	83
APBT16	169	85	90	163	127	124	237	83
APBT25	174	79	102	93	114	122	237	79
APBT27	172	86	81	104	121	108	235	79
APBT28	175	82	81	92	121	108	235	79
APBT31	174	79	81	93	158	108	239	79
APBT34	172	87	119	124	130	124	237	79
APBT35	175	90	79	99	120	130	235	101
APBT36	172	90	85	93	121	122	237	79
APBT37	172	90	85	93	120	122	237	79
APBT38	181	86	83	102	119	120	237	79
APBT39	172	86	156	104	121	108	235	79
APBT40	175	83	87	116	121	122	239	79
APBT44	184	84	134	95	123	108	237	79
APBT45	184	85	134	95	124	108	237	79
APBT51	169	79	134	108	122	108	237	79
APBT53	176	67	81	162	121	124	237	79
APBT54	176	68	81	164	122	124	237	79
APMR01	177	81	126	114	124	122	239	79
APMR02	176	81	126	114	124	122	239	79
APMR03	172	80	107	95	119	122	237	79
APMR04	172	79	107	95	119	122	237	79
APMR05	172	79	107	95	119	122	237	79
APMR06	172	79	107	95	119	122	237	79
APMR07	172	80	107	95	119	122	237	79

continued on next page

continued from previous page

Sample	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APMR09	173	73	118	120	127	120	237	79
APMR10	172	79	87	95	130	108	237	79
APMR11	172	78	87	95	130	108	237	79
APMR12	172	78	87	95	130	108	237	79
APMR13	172	78	87	95	130	108	237	79
APMR14	176	86	146	97	124	120	239	81
APMR15	172	78	89	104	124	120	237	76
APMR16	172	78	89	104	124	120	237	76
APMR17	172	79	89	104	124	120	237	76
APMR19	173	84	81	97	125	122	237	72
APMR20	176	81	127	114	124	122	239	79
APMR21	176	80	128	114	124	122	239	79
APMR22	176	80	128	114	124	122	239	79
APMR23	176	80	126	114	124	122	239	79
APMR24	176	80	128	95	124	122	237	79
APMR25	176	78	88	95	124	124	237	79
APMR26	176	78	88	112	121	124	237	79
APMR27	179	78	88	112	121	124	237	79
APMR28	176	80	128	114	124	122	239	79
APMR29	176	80	128	114	124	122	239	79
APMR30	176	86	148	97	124	120	239	81
APMR31	176	80	128	114	124	122	239	79
APMR32	176	87	148	97	124	120	239	81
APMR33	178	79	142	110	126	124	237	79
APMR35	176	87	144	97	124	120	239	81
APMR36	176	87	145	97	124	120	239	81
APMR37	176	87	148	97	124	120	239	81
APMR38	176	87	148	97	124	120	239	81
APMR39	176	87	148	97	124	120	239	81
APMR40	176	87	148	97	124	120	239	81
APMR41	176	87	146	97	124	120	239	81
APMR42	176	86	145	97	124	120	239	81
APMR43	176	86	146	97	124	120	239	81
APMR44	176	87	145	97	124	120	239	81
APMR45	176	87	146	97	124	120	239	81
APMR46	176	87	147	97	124	120	239	81
APMR47	180	87	130	104	128	108	237	76
APMR49	176	87	146	97	124	120	239	81
APMR51	176	86	145	97	124	120	239	81
APMR52	176	87	147	97	124	120	238	81
APMR56	176	81	128	114	124	122	239	79
APMR57	177	84	87	97	124	122	237	77
APMR58	176	81	126	114	124	122	239	79
APMR63	175	80	136	128	119	120	237	81

Appendix H

COMPUTATION OF LINKAGE DISEQUILIBRIUM

This is the results file of the computation of linkage disequilibrium for the 8 microsatellite loci using the software package GENEPOP (Raymond & Rousset, 1995b).

```
GenePop web version 3.4 Genotypic disequilibrium

File: 224413 (Microsatellites of Acanthorrhynchium papillatum)

Number of population detected: 1
Numer of loci detected:        8

Markov chain parameters
            Dememorization:       1000
            Batches:              100
            Iterations per batch: 1000
```

Pop	Locus#1	Locus#2	P-Value	S.E.
APBL06	APMS4	APMS14	0.00000	0.00000
APBL06	APMS4	APMS28	0.00000	0.00000
APBL06	APMS14	APMS28	0.00000	0.00000
APBL06	APMS4	APBMS3	0.00000	0.00000
APBL06	APMS14	APBMS3	0.00000	0.00000
APBL06	APMS28	APBMS3	0.00000	0.00000
APBL06	APMS4	APBMS14	0.00000	0.00000
APBL06	APMS14	APBMS14	0.00000	0.00000
APBL06	APMS28	APBMS14	0.00000	0.00000
APBL06	APBMS3	APBMS14	0.00000	0.00000
APBL06	APMS4	APBMS23	0.00000	0.00000
APBL06	APMS14	APBMS23	0.00000	0.00000
APBL06	APMS28	APBMS23	0.00000	0.00000
APBL06	APBMS3	APBMS23	0.00000	0.00000
APBL06	APBMS14	APBMS23	0.00000	0.00000
APBL06	APMS4	APBMS61	0.00000	0.00000
APBL06	APMS14	APBMS61	0.00000	0.00000
APBL06	APMS28	APBMS61	0.00000	0.00000
APBL06	APBMS3	APBMS61	0.00000	0.00000
APBL06	APBMS14	APBMS61	0.00000	0.00000
APBL06	APBMS23	APBMS61	0.00000	0.00000
APBL06	APMS4	APBMS72	0.00000	0.00000
APBL06	APMS14	APBMS72	0.00000	0.00000
APBL06	APMS28	APBMS72	0.00000	0.00000
APBL06	APBMS3	APBMS72	0.00000	0.00000
APBL06	APBMS14	APBMS72	0.00000	0.00000
APBL06	APBMS23	APBMS72	0.00000	0.00000

```
APBL06     APBMS61    APBMS72     0.00000      0.00000


P-value for each locus pair across all populations
(Fisher's method)
------------------------------------------------------------
Locus pair               Chi2        df       P-value
-----------------        ---------  ----      ----------
APMS4     &  APMS14      Infinity    2          Highly sign.
APMS4     &  APMS28      Infinity    2          Highly sign.
APMS14    &  APMS28      Infinity    2          Highly sign.
APMS4     &  APBMS3      Infinity    2          Highly sign.
APMS14    &  APBMS3      Infinity    2          Highly sign.
APMS28    &  APBMS3      Infinity    2          Highly sign.
APMS4     &  APBMS14     Infinity    2          Highly sign.
APMS14    &  APBMS14     Infinity    2          Highly sign.
APMS28    &  APBMS14     Infinity    2          Highly sign.
APBMS3    &  APBMS14     Infinity    2          Highly sign.
APMS4     &  APBMS23     Infinity    2          Highly sign.
APMS14    &  APBMS23     Infinity    2          Highly sign.
APMS28    &  APBMS23     Infinity    2          Highly sign.
APBMS3    &  APBMS23     Infinity    2          Highly sign.
APBMS14   &  APBMS23     Infinity    2          Highly sign.
APMS4     &  APBMS61     Infinity    2          Highly sign.
APMS14    &  APBMS61     Infinity    2          Highly sign.
APMS28    &  APBMS61     Infinity    2          Highly sign.
APBMS3    &  APBMS61     Infinity    2          Highly sign.
APBMS14   &  APBMS61     Infinity    2          Highly sign.
APBMS23   &  APBMS61     Infinity    2          Highly sign.
APMS4     &  APBMS72     Infinity    2          Highly sign.
APMS14    &  APBMS72     Infinity    2          Highly sign.
APMS28    &  APBMS72     Infinity    2          Highly sign.
APBMS3    &  APBMS72     Infinity    2          Highly sign.
APBMS14   &  APBMS72     Infinity    2          Highly sign.
APBMS23   &  APBMS72     Infinity    2          Highly sign.
APBMS61   &  APBMS72     Infinity    2          Highly sign.

Normal Ending
```

Appendix I

REPORT ON MICROSATELLITE MARKERS FOR *ACANTHORRHYNCHIUM PAPILLATUM*

Molecular Ecology Notes (2006) **6**, 396–398 doi: 10.1111/j.1471-8286.2006.01246.x

PRIMER NOTE

Development of microsatellite markers for the tropical moss, *Acanthorrhynchium papillatum*

ALFREDO AMIEL P. LEONARDÍA, PRAKASH P. KUMAR and BENITO C. TAN
Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore S117543

Abstract

We isolated and characterized microsatellite loci for the palaeotropic moss, *Acanthorrhynchium papillatum*. Eight loci tested on 98 gametophytic samples generated four to 26 alleles per locus with genetic diversities ranging from 0.578 to 0.936. These microsatellite loci are now being used as genetic markers for studies on the effects of deforestation on moss populations in South-East Asia.

Keywords: *Acanthorrhynchium papillatum*, microsatellites, Sematophyllaceae

Received 08 October 2005; revision accepted 14 November 2005

The forests of South-East Asia continue to be threatened hotspots of biodiversity. Although there are projects in the area that study the effects of deforestation on plant diversity, these efforts mainly focus on flowering plants. Little is known of the effects of deforestation on the genetic diversity of mosses in South-East Asia.

Acanthorrhynchium papillatum (Sematophyllaceae, Musci) is a pleurocarpous moss distributed from East Africa, Indo-Malesia, Australia to Oceania. It is commonly found on forest floors and rotting logs in mature forests and highly disturbed woodlands. We developed microsatellite markers for this species as part of a larger project to assess the effects of habitat changes on moss genetic diversity.

Genomic DNA was extracted from pooled individuals of a single clump of *A. papillatum* using a cetyltrimethyl ammonium bromide (CTAB) protocol of Doyle & Doyle (1990). A genomic library enriched for microsatellites was generated using a modified protocol of Hamilton *et al*. (1999). Approximately 20 μg of genomic DNA was digested to completion with *Alu*I, *Hae*II and *Nhe*I (New England Biolabs) and ligated to SNX linkers. 3′-biotinylated oligos (Research Instruments) of $(GA)_{15}$ $(GT)_{15}$ $(CTC)_{10}$ and $(CAGA)_8$ motifs were allowed to hybridize with the fragments. Streptavidin beads (MagPrep, Novagen) were used to bind the hybrids while high stringency washes were performed to purify the enrichments. The resulting fragments were amplified, trimmed and ligated to pBluescript II SK + vector (Stratagene), cloned and then sequenced with T7 and T3 primers in an ABI PRISM 3100 Genetic Analyser (Applied Biosystems).

Correspondence: Benito C. Tan, Fax: +65 6779 2486; E-mail: dbsbct@nus.edu.sg

Primers were designed with the programmes GENERUNNER (1994, version 3.00, Hastings Software) and PRIMER 3 (Rozen & Skaletsky 2000). Amplifications were carried out in 10 μL volumes of 1 × F-511 buffer (FinnZymes), 250 μM dNTP (Promega), 1 μM each primer (1st BASE), 0.4 U DyNAzyme II Recombinant DNA polymerase (FinnZymes) and approximately 20 ng genomic DNA. The thermal profile used was initial denaturation for 4 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 30 s at 58 °C, 30 s at 72 °C, and final extension for 10 min at 72 °C. Following successful amplification, products were resolved on silver-stained polyacrylamide (Caetano-Anollés & Gresshoff 1994) or 1 × TBE 3% Metaphor (Seakem) gels to check for consistency of amplification and polymorphism.

Of 170 inserts sequenced, 98 possessed microsatellites, of which 45 had flanking regions that were too short for designing primers. Twenty-nine primer pairs were designed, of which eight consistently amplify loci and produce length-polymorphic products (Table 1). Although amplifications for all eight primers were performed at an annealing temperature of 58 °C, all of them amplified with no loss of efficiency or specificity at annealing temperatures from 54 to 58 °C. Several combinations of multiplex polymerase chain reaction (PCR) were attempted through this temperature range with no production of PCR artefacts. However, concentrations of primers needed to be adjusted (Table 2).

Table 1 Characteristics of eight microsatellite loci of *Acanthorrhynchium papillatum* tested on 98 gametophytic samples collected from Singapore and Peninsular Malaysia

Locus	Primer sequences (5′–3′)	Repeat motif in clone	T_a (°C)	Size range (bp)	A	H_E	GenBank Accession no.
APMS14	6-FAM-ATATTAACCTCGCTCGACG CAGCTCGAGATCCTTCTCTC	$(GA)_{46}$	58	67–126	11	0.852	DQ092744
APMS4	HEX-CCTCCTACCAGCATTGAAAG ACCTGGGGAGGATTATAAATTCAG	$(GA)_{22}$	58	166–217	15	0.736	DQ092745
APMS28	HEX-CACTTTGCCGACTACAGTTG GTCCAGGTCTGATCACGTAAC	$(GA)_{23}$	58	79–156	26	0.936	DQ092746
APBMS3	GCTCTCTTTCGTTTGTTCAG 6-FAM-CAATCGGCTTGCAATTATAC	$(GA)_{20}$	58	92–164	19	0.874	DQ092747
APBMS72	HEX-CAATCTCTTGCGTTTGCTTG AGAAGAGCGTCCACACTTGG	$(GA)_{30}$	58	69–111	11	0.602	DQ092748
APBMS14	NED-CGAAGCCGGTGGTAAGTCC CGGTCCTCTCTCCTCCAAAT	$(GA)_3(G)_4(GA)_9$	58	94–158	11	0.763	DQ092749
APBMS23	NED-TCGAAATCGTCAGTACTTTCCAG ATTCACGAGATCCTCTGAGTTCG	$(GA)_4GT(GA)_5$ $GT(GA)_3$	58	108–130	6	0.736	DQ092750
APBMS61	AACTGAAGTTCCTCCACTCTTTTC 6-FAM-AGGTGCTTCTTAGAATCTTGTTGC	$(CA)_4(TC)_8$ $CTT(C)_6TC(CT)_4$	58	233–239	4	0.578	DQ092751

T_a, annealing temperature; A, number of alleles; H_E, genetic diversity

Table 2 Multiplex reactions used. Products of similarly labelled primers in the same multiplex set do not overlap throughout their size ranges

Multiplex set	Locus	Fluorophore dye	Primer concentration (μM)
A	APMS14	6-FAM	1
	APMS28	HEX	1
	APBMS23	NED	0.5
	APBMS61	6-FAM	1
B	APMS4	HEX	1
	APBMS72	HEX	1
	APBMS3	6-FAM	1
	APBMS14	NED	0.5

Working primers were optimized for multiplex PCR by trial and error. Half of each primer pair was labelled with HEX, 6-FAM (1st BASE) or NED (Applied Biosystems) fluorophores (Table 1). The PCR extension step was performed for 2 min at 65 °C to more evenly amplify bands (Henegariu *et al*. 1997), and final extension was prolonged to 30 min at 72 °C to facilitate the uniform addition of the T-overhang. Products were purified (CleanSeq, Agencourt), resolved in an ABI 377 PRISM DNA Sequencer (Applied Biosystems) and analysed using GENESCAN version 3.1.2 (Applied Biosystems). Ninety-eight gametophytic (haploid) samples collected from Singapore and Peninsular Malaysia were genotyped.

Genetic diversities (Nei 1973) were computed with the aid of a spreadsheet. Linkage equilibrium was determined using the programme GENEPOP (Raymond & Rousset 1995). Since many of the samples shared the same multilocus genotype, the high incidence of linkage disequilibrium found by the programme may be attributed to the possible clonal nature of the organism. These microsatellite markers are now being used in studies of the population genetics and the effects of deforestation on the genetic diversity of *A. papillatum*.

Acknowledgements

This research was funded in part by the National University of Singapore Research Grant R-154-000-185-112. We would like to thank Yap Von Bing of the Department of Statistics and Applied Probability at the NUS for discussions concerning the needed statistics. We are also grateful to Helena Korpelainen and her laboratory at the Department of Applied Biology at the University of Helsinki for their advice and support during the lab consultation of the first author.

References

Caetano-Anollés G, Gresshoff PM (1994) Staining nucleic acids with silver: an alternative to radioisotopic and fluorescent labeling. *Promega Notes Magazine*, **45**, 13–20.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques*, **27**, 500–507.

Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23**, 504–511.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3323.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): a population genetics software for exact test and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rozen S, Skaletsky HJ (2000) PRIMER 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

Appendix J

INTERCLUMP DATA

Table J.1: Microsatellite marker alleles of samples for among-clump diversity studies. Header rows indicate names of markers. Alleles are PCR amplicon lengths in bp. Null alleles are indicated by '-'.

Sample Name	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APBK01	174	78	77	103	124	105	234	77
APBK02	174	78	77	103	124	105	234	77
APBK03	174	64	108	-	126	118	234	89
APBK04	172	78	103	95	123	130	234	77
APBK07	174	82	77	113	123	179	236	133
APBK08	169	91	77	103	124	105	234	128
APBK09	174	76	85	89	127	105	236	77
APBK10	201	77	124	93	123	128	232	93
APBK11	174	82	103	90	123	130	236	87
APBK12	205	75	77	101	117	105	234	110
APBK13	172	80	88	89	129	122	236	77
APBK14	172	80	88	89	129	122	236	77
APBK15	174	76	122	91	95	105	236	77
APBK17	174	82	108	89	95	118	234	77
APBK18	174	74	108	103	113	133	234	107
APBK19	175	78	81	91	96	105	236	77
APBK20	177	77	77	91	124	120	236	77
APBK21	172	83	128	120	126	105	234	77
APBK22	-	75	126	126	123	105	236	77
APBK23	171	80	134	91	121	118	236	77
APBK24	178	78	79	126	115	105	236	77
APBK25	174	78	77	118	124	128	234	109
APBK26	161	74	83	103	104	105	236	77
APBK27	174	78	124	89	125	105	236	77
APBK28	161	74	124	89	122	105	234	77
APBK29	161	74	83	103	104	105	236	77
APBK30	170	74	97	160	118	105	236	77
APBK31	171	78	91	-	126	105	236	77
APBK32	163	78	103	105	123	118	234	77
APBK33	169	89	116	93	123	105	236	77
APBK34	165	78	85	103	126	124	236	77
APBK35	176	78	77	93	124	152	234	128
APBK36	172	78	106	95	123	130	234	-
APBK37	174	78	124	118	124	128	234	-
APBK38	178	78	120	105	123	105	236	77
APBK39	169	118	77	93	124	148	236	105
APBK40	174	87	77	95	123	105	232	109
APBK41	174	82	77	113	124	148	232	109
APBK42	172	65	77	124	124	128	234	99
APBK43	174	82	77	113	124	148	232	109
APBK44	172	88	101	113	123	130	236	114
APBK45	174	87	77	95	123	191	232	109
APBK46	174	80	83	93	126	105	234	77
APBK47	174	74	114	93	122	105	236	77
APBK48	174	80	83	93	126	105	234	77
APBK49	174	109	89	91	122	114	236	77
APBK50	174	122	126	138	111	105	236	79
APBK51	174	87	106	109	136	105	236	77
APBK52	161	80	81	152	118	105	236	77
APBK53	175	70	81	93	124	118	236	79
APBK54	175	78	81	93	124	105	236	77
APBL01	177	-	116	89	126	105	238	77
APBL02	178	87	77	120	123	120	232	87
APBL03	203	78	99	93	93	124	238	109
APBL04	178	87	77	95	123	120	232	77
APBL05	161	74	88	124	126	105	236	77
APBL06	174	74	118	91	125	105	236	77

continued on next page

continued from previous page

Sample Name	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APBL07	171	76	114	89	122	122	236	77
APBL08	205	78	106	122	124	118	238	109
APBL09	175	70	79	111	125	105	234	77
APBL10	175	70	79	111	125	105	234	77
APBL11	174	82	130	126	124	118	237	87
APBL12	199	78	95	122	93	124	232	109
APBL13	163	78	103	103	122	120	236	77
APBL14	175	74	106	128	126	105	236	77
APBL15	175	76	108	99	120	105	236	81
APBL16	164	74	114	91	122	120	236	77
APBL17	174	80	99	91	122	105	236	77
APBL18	169	82	77	103	124	120	234	93
APBL19	169	78	126	91	117	122	234	77
APBL20	164	-	99	91	126	105	236	75
APBL21	174	109	99	120	126	105	236	77
APBL22	174	77	95	91	122	124	236	77
APBL23	172	78	122	91	122	105	236	77
APBL24	177	78	112	120	129	120	235	77
APBL25	174	-	120	120	117	120	240	77
APBL26	174	-	120	120	117	120	240	77
APBL27	174	78	95	91	95	124	238	91
APBL28	177	74	116	132	111	105	236	77
APBL29	169	78	79	120	97	120	240	77
APBL30	175	74	120	93	111	105	236	77
APBL31	182	78	77	88	110	122	238	89
APBL32	182	78	77	103	110	122	238	89
APBL33	165	64	120	91	121	105	236	77
APBL34	177	82	99	89	119	105	234	77
APBL35	177	64	83	130	124	97	234	77
APBL36	177	99	83	105	124	97	234	77
APBL37	175	-	93	89	122	105	236	77
APBL38	163	74	89	89	129	122	236	77
APBL39	174	78	122	91	126	122	236	77
APBL40	177	75	93	91	122	105	236	77
APBL41	175	87	114	111	122	118	236	83
APBL42	174	87	126	107	126	105	236	77
APBL43	174	78	85	101	124	120	236	77
APBL44	170	82	93	116	120	118	238	87
APBL45	171	78	112	126	124	105	236	77
APBL46	175	74	133	136	119	120	236	77
APBL47	165	74	99	103	113	105	236	77
APBL48	175	88	83	93	127	97	234	77
APBL49	167	68	134	-	122	105	236	77
APBL50	171	76	91	91	126	105	236	77
APBL51	201	78	101	122	124	118	238	109
APBL52	174	75	83	91	120	105	236	77
APBL53	174	78	122	126	126	118	238	79
APBL54	174	74	114	99	128	105	236	77
APBL55	177	74	130	103	122	105	236	77
APBL56	174	82	124	126	124	118	237	77
APBL57	181	78	81	122	93	124	238	109
APBL58	174	78	101	91	126	118	238	77
APBL59	174	82	126	126	124	118	237	87
APBL60	174	64	93	126	125	118	238	77
APBL61	182	78	97	-	93	124	238	109
APBL62	207	78	101	122	117	124	232	87
APBL63	203	78	97	122	93	124	238	109
APBL64	171	76	114	103	122	118	240	77
APBL65	215	78	122	128	124	124	232	87
APBL66	174	122	77	114	111	120	236	77
APBL67	174	78	93	89	122	124	236	77
APBL68	174	85	83	122	122	105	236	77
APBL69	168	74	124	91	124	105	238	77
APBL70	175	-	128	91	120	120	236	77
APBL71	174	76	126	105	118	105	236	77
APBL72	197	78	77	124	124	122	238	87
APKT01	174	90	118	124	97	105	236	67
APKT02	176	76	120	124	124	105	236	77
APKT03	171	74	128	106	113	120	236	77
APKT04	168	76	85	91	124	118	236	77
APKT05	174	76	128	91	120	105	236	118
APKT06	174	93	120	126	97	105	236	67
APKT07	174	85	118	120	97	105	236	67
APKT08	174	93	112	89	97	105	236	67
APKT09	174	93	118	126	97	105	236	67
APKT10	174	93	118	126	97	105	236	67
APKT11	174	93	118	124	97	105	236	67
APKT12	174	95	114	132	97	105	236	67

continued on next page

continued from previous page

Sample Name	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APKT13	174	95	114	130	97	105	236	67
APKT14	174	95	112	132	97	105	236	67
APKT15	174	95	112	132	97	105	236	67
APKT16	174	95	77	103	97	105	236	67
APKT17	174	95	112	132	97	105	236	67
APKT18	175	82	118	120	118	105	236	77
APKT19	174	75	81	124	118	122	238	77
APKT20	174	76	81	145	118	114	236	77
APKT21	171	78	-	91	129	118	236	77
APKT22	174	76	114	89	122	105	236	77
APKT23	177	-	124	93	121	105	242	77
APKT24	174	76	81	103	118	114	236	77
APKT25	177	80	77	105	122	105	234	77
APKT26	164	78	-	105	122	105	236	77
APKT27	165	78	79	106	126	120	236	77
APKT28	177	80	77	105	122	105	234	77
APKT29	177	80	124	105	122	105	234	77
APKT30	175	80	120	103	115	118	236	77
APKT31	175	74	114	91	95	105	234	79
APKT32	174	78	-	107	126	120	236	77
APKT33	178	74	77	93	117	105	234	79
APKT34	176	76	77	120	131	105	234	77
APBT01	172	74	95	122	97	120	238	77
APBT02	172	74	95	122	97	120	238	77
APBT03	172	74	95	122	97	120	238	77
APBT04	172	74	95	122	97	120	238	77
APBT05	172	74	97	122	97	120	238	77
APBT06	165	76	83	124	113	122	236	77
APBT07	165	75	83	124	113	122	236	77
APBT08	165	75	83	124	113	122	236	77
APBT09	165	78	83	124	113	122	236	77
APBT10	165	75	83	124	113	122	236	77
APBT11	165	75	83	124	113	122	236	77
APBT12	172	74	95	122	97	120	238	77
APBT13	175	82	81	93	121	118	236	77
APBT14	174	78	85	120	120	120	238	77
APBT15	168	80	88	162	125	122	236	81
APBT16	168	80	88	162	125	122	236	81
APBT17	174	74	122	107	121	138	238	75
APBT18	177	76	120	148	128	124	236	77
APBT19	167	82	112	89	-	105	236	77
APBT20	167	80	112	89	-	105	236	77
APBT21	167	80	112	89	-	105	236	77
APBT22	167	80	112	89	-	105	236	77
APBT23	177	74	132	127	123	122	236	77
APBT24	163	74	132	114	123	118	234	75
APBT25	174	74	99	89	111	120	236	77
APBT26	170	76	114	148	127	122	238	77
APBT27	171	78	79	101	119	105	234	77
APBT28	174	78	79	89	120	105	234	77
APBT29	175	78	79	89	110	118	240	77
APBT30	175	78	79	89	110	118	240	77
APBT31	174	74	79	89	156	105	238	77
APBT32	171	82	116	122	128	122	236	77
APBT33	169	82	75	122	128	105	236	77
APBT34	171	82	116	122	128	122	236	77
APBT35	174	87	77	95	119	128	234	99
APBT36	171	87	83	89	120	120	236	77
APBT37	171	87	83	89	120	120	236	77
APBT38	181	82	81	99	118	118	236	77
APBT39	171	82	154	101	119	105	234	77
APBT40	174	78	85	113	120	120	238	77
APBT41	174	80	132	93	121	118	236	77
APBT42	174	59	-	93	121	118	236	77
APBT43	169	78	95	89	120	122	234	76
APBT44	184	80	132	91	122	105	236	77
APBT45	184	80	132	91	122	105	236	77
APBT46	184	80	132	91	122	105	236	77
APBT47	175	78	77	95	126	105	238	77
APBT48	172	65	108	107	128	122	236	77
APBT49	174	77	110	145	95	124	236	77
APBT50	171	74	106	93	128	122	238	77
APBT51	168	76	77	105	121	105	236	77
APBT52	175	64	79	76	120	122	236	77
APBT53	175	64	79	-	120	122	236	77
APBT54	175	64	79	91	120	122	236	77
APMR01	175	76	88	111	122	120	238	77
APMR02	175	76	88	111	122	120	238	77

continued on next page

continued from previous page

Sample Name	APMS4	APMS14	APMS28	APBMS3	APBMS14	APBMS23	APBMS61	APBMS72
APMR03	171	76	106	91	117	120	236	77
APMR04	171	76	106	91	117	120	236	77
APMR05	171	76	106	91	117	120	236	77
APMR06	171	76	106	91	117	120	236	77
APMR07	171	76	106	91	117	120	236	77
APMR08	165	78	83	91	123	120	236	74
APMR09	172	70	116	118	125	118	236	77
APMR10	171	74	85	120	115	105	236	77
APMR11	171	74	85	91	128	105	236	77
APMR12	171	74	85	91	128	105	236	77
APMR13	171	74	85	91	128	105	236	77
APMR14	175	82	83	93	122	118	238	79
APMR15	171	74	88	101	122	118	236	74
APMR16	171	74	88	101	129	118	236	77
APMR17	171	74	88	101	122	118	236	74
APMR18	174	65	106	101	121	105	234	77
APMR19	172	80	79	93	123	120	236	75
APMR20	175	76	122	111	122	120	238	77
APMR21	175	76	126	111	122	120	238	77
APMR22	175	76	126	111	122	120	238	77
APMR23	175	76	124	111	122	120	238	77
APMR24	175	76	126	91	122	120	236	77
APMR25	175	74	85	91	122	122	236	77
APMR26	175	74	85	109	119	122	236	77
APMR27	178	74	85	109	119	122	236	77
APMR28	175	76	126	111	122	120	238	77
APMR29	175	76	126	111	122	120	238	77
APMR30	175	82	77	93	122	118	238	79
APMR31	175	76	126	111	122	120	238	77
APMR32	175	82	77	93	122	118	238	79
APMR33	177	74	140	107	125	122	236	77
APMR34	175	82	-	93	122	118	238	79
APMR35	175	82	-	93	122	118	238	79
APMR36	175	82	-	93	122	118	238	79
APMR37	175	82	-	93	122	118	238	79
APMR38	175	82	-	93	122	118	238	79
APMR39	175	82	-	93	122	118	238	79
APMR40	175	82	77	93	122	118	238	79
APMR41	175	82	77	93	122	118	238	79
APMR42	175	82	77	93	122	118	238	79
APMR43	175	82	77	93	122	118	238	79
APMR44	175	82	77	93	122	118	238	79
APMR45	175	82	77	93	122	118	238	79
APMR46	175	82	77	93	122	118	238	79
APMR47	178	82	126	101	126	105	236	74
APMR48	178	78	79	101	126	118	236	74
APMR49	175	82	77	93	122	118	238	79
APMR50	164	78	132	122	124	118	238	77
APMR51	175	82	77	93	122	118	238	79
APMR52	175	82	-	93	122	118	238	79
APMR53	174	76	136	126	117	118	236	79
APMR54	174	105	132	95	118	122	236	89
APMR55	174	105	132	95	118	122	236	89
APMR56	175	76	126	111	122	120	238	77
APMR57	176	80	85	93	123	120	236	75
APMR58	175	76	124	111	122	120	238	77
APMR59	175	76	124	111	122	120	238	77
APMR60	174	76	134	126	117	118	236	79
APMR61	174	76	134	126	117	118	236	79
APMR62	174	105	132	95	118	122	236	89
APMR63	174	76	134	126	117	118	236	79
APMR64	174	76	134	126	117	118	236	79
APMR65	174	76	134	126	117	118	236	79
APMR66	174	76	134	126	117	118	236	79
APMR67	174	76	134	126	117	118	236	79
APMR68	174	76	134	126	117	118	236	79

Appendix K

ALIGNED *ITS2* SEQUENCES

The following are the aligned *ITS*2 sequences of 193 samples of *Acanthorrhynchium papillatum*. They are presented in FASTA format for better readability. Single nucleotide gaps are represented by '-'.

```
>APBK1
GCGCCCCCCCACCGATCCCC------GCTCTAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK2
GCGCCCCCCCACCGATCCCC------GCTCTAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK3
GCGCCCCCCCACCGATCCCC------GCTCTAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK4
GCGCCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK7
GCGCCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK8
GCGCCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
```

```
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCAAA-----CTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK9
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APBK10
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK11
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK12
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAA-----CTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK17
GCGCCCCCCACCGATCCCC------GCTCTAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK18
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT
```

```
>APBK20
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCG----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTGCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC---------TAAGCGGCGCGCGTCGTC
ACGCCT

>APBK25
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK26
GCGCCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBK31
GCGCCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCTA-CTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBK32
GCGCCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCCCGCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBK33
GCGCCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGC----------------AAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCTAG---------GCGGCGCGCGTCGTC
ACGCCT

>APBK35
GCGCCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
```

```
ACGCCT

>APBK36
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK37
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCCAGCCAGT----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK38
GCGCCCCCCACCGATCCCT------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBK39
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK40
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK41
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK42
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
```

```
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK43
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK44
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG-GGGG-AGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK45
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBK47
GCGCCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGTGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTTGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGG-TCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCTAA---------GCGGCGCGCGTCGTC
ACGCCT

>APBK50
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCCCGCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBK52
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBK53
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL1
GCGCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAC---TTTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL2
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGC--CAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL3
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL4
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL6
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL8
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT
```

>APBL9
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL10
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL11
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCCCGCCAGC----GCGCGAAG-----CGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCGTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL12
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL14
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL15
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCGCCCCCCAAC---TTTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL17
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC

```
ACGCCT

>APBL18
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----CAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL20
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCA---ACTTTGGAGGGCGCGCGTCGTC
ACGCCT

>APBL21
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL22
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL26
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGTGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCCTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL28
GCGCCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCCAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL30
GCGCCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
```

```
CTCAGGCCGGGACGGAGAGAGGGGTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL31
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG----------------AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL32
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG----------------AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL33
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL34
GCGCCCCCCACCGATCCCC------GCTCCAGTC-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APBL35
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL37
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAAC---TTTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL39
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGTGCCAGC----GCGCGAAGTGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAC-TCAAGGAAGGGCGCGCGTCGTC
ACGCCT

>APBL41
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAATCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL42
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTTCCCTTGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL43
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCTTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL44
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAAC-----TGAAGGGCGCGCGTCGTC
ACGCCT

>APBL45
GCGCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAC---TTTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL46
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCC-----AACTGAAGGGCGCGCGTCGTC
ACGCCT
```

>APBL47
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL50
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL51
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCGC------------GGCGCGCGTCGTC
ACGCCT

>APBL52
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAACGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL53
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCGC------------GGCGCGCGTCGTC
ACGCCT

>APBL56
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCCCGCCAGC----GCGCGAAG-----CGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCGTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL57
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCGC------------GGCGCGCGTCGTC

```
ACGCCT

>APBL58
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCCCGGCGCGCGCCAGTGAGC----GCGCG---CGAAGGGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCGC------------GGCGCGCGTCGTC
ACGCCT

>APBL59
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCCCGCCAGC----GCGCGAAG-----CGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCGTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL61
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG----------------AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL62
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCCAGCCAGT----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL64
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAACGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL65
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBL66
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGTGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
```

```
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL68
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL69
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTTCC--GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBL71
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBL72
GCGCCCCCCACCGATCCCC------GCTCCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAG----------------AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APKT1
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT2
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGATGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGTCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT3
GCGCCCCCCACCGATCCCC------GCTCCAGTG-----------AAGTTGGGGTC-------GAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT4
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT6
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT7
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT8
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT9
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT11
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT
```

>APKT12
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT13
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT14
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT15
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT16
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT17
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGTCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APKT20
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACT-----GAAGGGCGCGCGTCGTC

```
ACGCCT

>APKT22
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT23
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT24
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAACT-----GAAGGGCGCGCGTCGTC
ACGCCT

>APKT25
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT28
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT29
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APKT32
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
```

```
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACGGAAGGGCGCGCGTCGTC
ACGCCT

>APBT1
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT4
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT6
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT7
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT8
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT9
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT10
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT11
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT15
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGTGCCAGC----GCGCGAAGTGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAGC------TAAGGGCGCGCGTCGTC
ACGCCT

>APBT16
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGTGCCAGC----GCGCGAAGTGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCCAGC------TAAGGGCGCGCGTCGTC
ACGCCT

>APBT17
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBT18
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT19
GCGCCCCCCACCGATCCGCGCTCCCGCTCC------GTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT
```

>APBT20
GCGCCCCCCACCGATCCGCGCTCCCGCTCC------GTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT21
GCGCCCCCCACCGATCCGCGCTCCCGCTCC------GTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT22
GCGCCCCCCACCGATCCGCGCTCCCGCTCC------GTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT23
GCGCCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT25
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBT26
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT29
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGC--CAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGTTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC

```
ACGCCT

>APBT30
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGC--CAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGTTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT32
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGTGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTTAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBT35
GCGCCCCCCACCGATCCCC------GCTGCAGTT----------------GGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
---AGGCCGGGACGGAGAGAGGGGGAGCTA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC-----AAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APBT36
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT37
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT38
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAAACT---TTGAAGGGCGCGCGTCGTC
ACGCCT

>APBT44
GCGCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGCCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
```

```
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT45
GCGCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGCCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT46
GCGCCCCCCACCGATCCCCTCTGCCGCTCCCGCTCCGCCGGGAGTGAGTTGGGGTCCTC----GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT49
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APBT50
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APBT51
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTTGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCTAA---------GCGGCGCGCGTCGTC
ACGCCT

>APBT53
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR1
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR2
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR3
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR4
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR5
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR6
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR7
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT
```

```
>APMR9
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR10
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GTGGCGCGCGTC---
ACGCCT

>APMR11
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GTGGCGCGCGTC---
ACGCCT

>APMR12
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GTGGCGCGCGTC---
ACGCCT

>APMR13
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GTGGCGCGCGTC---
ACGCCT

>APMR14
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR16
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGC-------------------GCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCCAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
```

```
ACGCCT

>APMR18
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR19
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR20
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR21
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR22
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR23
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR24
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
```

```
CTCAGGCCGGGACGGAGAGAGGGCTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR25
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APMR26
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCCCAACTAAACTGAAGGGCGCGCGTCGTC
ACGCCT

>APMR27
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG---AACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR28
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR29
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR30
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR31
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
```

```
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR32
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR34
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR35
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR36
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR37
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR38
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT
```

>APMR39
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR40
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR41
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR42
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR43
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR44
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR45
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC

```
ACGCCT

>APMR46
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATCG--G-ACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR47
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR48
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGGGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGCTCGCTAGCTAGAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR58
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR59
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGCCAGCGCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR60
GCGCCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APMR61
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
```

```
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APMR62
GCGCCCCCCACCGATCCCC------GCTCCAGTT-----------AAGTTGGGGTC-------GAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGAAGTCGCGTGCCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGACGGAGAGAGGGGTCGCCA----GAGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCCCC-GCTTCCTTGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTCGTC
ACGCCT

>APMR63
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APMR64
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APMR65
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT

>APMR66
GCGCCCCCCACCGATCCGCGCTCCCGCTCCCGCTCCGTCGGGAGTGAGTTGGGCTCCTCGATCGAGTGCAAGTGG
CCGTCCCCGCGGCGCGCGCGCGCCAGC----GCGCGAAGCGAAGCGAGGGTCGGCTGAAATGGATGGAACGGGGC
CGCCGTGGTGCGCCGCAGCCCGCGATCGGGTGACGTCGCGGGTCGCTCCCCTCGAATGGTGGAACCCTCCTCCTC
CTCAGGCCGGGATGGAGAGAGGGGTCGCTAG----AGGGCGAGCGTCCCGCGCTCTTTCTAAGTGCGTGGCTGCG
TCCCCCAGTCGGTTCC---GCTTCCAAGTGCAGGTCCCGCGCGCCCCC------------GCGGCGCGCGTC---
ACGCCT
```

Appendix L

WITHIN-CLUMP DATA - MAIN

Table L.1: Microsatellite marker alleles of main samples used in analysis of within-clump diversity of *Acanthorrhynchium papillatum*. Header rows indicate names of markers. Alleles are PCR fragment lengths in bp. Null alleles are indicated by '-'.

Sample	APBMS23	APBMS61	APMS14	APMS28	APBMS14	APBMS3	APBMS72	APMS4
APBK101A	126	236	80	120	123	105	77	163
APBK101B	105	236	74	126	122	105	77	178
APBK101C	114	236	80	130	123	105	77	175
APBK101D	105	236	80	-	108	122	77	163
APBK102A	105	236	80	-	108	-	77	-
APBK102B	105	236	80	-	108	-	77	-
APBK102C	105	236	80	-	108	-	77	-
APBK102D	105	236	80	-	108	-	77	-
APBK104A	105	236	-	-	123	126	77	-
APBK104B	105	236	-	-	123	126	77	-
APBK104C	105	236	-	-	123	126	77	-
APBK104D	105	236	-	-	123	126	77	-
APBK107A	105	240	78	99	122	76	77	168
APBK107B	105	240	78	99	122	76	77	168
APBK107C	105	240	78	99	122	76	77	168
APBK107D	105	240	78	99	122	76	77	168
APBK108A	105	236	-	-	136	109	77	174
APBK108B	105	236	-	-	136	109	77	174
APBK108C	105	236	-	-	136	109	77	174
APBK108D	105	236	-	-	136	109	77	174
APBK111A	105	236	80	89	126	89	77	168
APBK111B	105	236	80	91	126	124	77	168
APBK111C	105	236	80	91	126	124	77	168
APBK111D	105	236	80	89	126	89	77	168
APBK113A	130	236	78	79	115	91	77	174
APBK113B	130	236	78	79	115	91	77	174
APBK113C	105	236	80	106	117	91	77	161
APBK113D	130	236	78	79	115	91	77	174
APBK115A	105	236	80	91	121	89	77	168
APBK115B	105	236	80	91	121	89	77	168
APBK115C	105	236	80	91	121	89	77	168
APBK115D	105	236	80	91	121	89	77	168
APBL102A	118	236	74	116	125	89	77	168
APBL102B	118	236	74	116	125	89	77	168
APBL102C	118	236	74	116	125	89	77	168
APBL102D	118	236	74	116	125	89	77	168
APBL103A	105	236	74	101	121	91	77	175
APBL103B	105	236	78	120	125	120	77	164
APBL103C	105	236	78	120	125	120	77	164
APBL103D	105	236	78	120	125	120	77	164
APBL105A	105	236	74	116	121	89	77	156
APBL105B	105	236	74	116	121	89	77	156
APBL105C	105	236	74	116	121	89	77	156
APBL105D	105	236	74	116	121	89	77	156
APBL106A	105	236	76	-	124	101	77	177
APBL106B	105	236	76	-	124	101	77	177

continued on next page

continued from previous page

Sample	APBMS23	APBMS61	APMS14	APMS28	APBMS14	APBMS3	APBMS72	APMS4
APBL106C	105	236	76	-	124	101	77	177
APBL106D	105	236	76	-	124	101	77	177
APBL107A	105	236	74	91	122	91	77	170
APBL107B	105	236	74	91	122	91	77	170
APBL107C	105	236	74	91	122	91	77	170
APBL107D	105	236	74	91	122	91	77	170
APBL108A	105	236	70	122	117	105	77	171
APBL108B	105	236	70	122	117	105	77	171
APBL108C	105	236	70	122	117	105	77	171
APBL108D	105	236	70	122	117	105	77	171
APBL109A	105	236	-	89	126	-	77	170
APBL109B	105	236	-	89	126	-	77	170
APBL109C	105	236	-	89	126	-	77	170
APBL109D	120	236	76	116	127	93	77	177
APBL110A	105	236	76	97	123	-	77	165
APBL110B	105	236	76	97	123	-	77	165
APBL110C	105	236	80	114	111	101	77	177
APBL110D	105	236	76	97	123	-	77	165
APKT05A	105	236	76	128	120	91	118	174
APKT05B	105	236	76	128	120	91	118	174
APKT05C	105	236	76	128	120	91	118	174
APKT05D	105	236	76	128	120	91	118	174
APKT10A	105	236	93	118	97	126	67	174
APKT10B	105	236	93	118	97	130	67	174
APKT10C	105	236	93	118	97	130	67	174
APKT10D	105	236	93	118	97	130	67	174
APKT12A	105	236	95	114	97	132	67	174
APKT12B	105	236	95	114	97	93	67	174
APKT12C	105	236	95	114	97	132	67	174
APKT12D	105	236	95	114	97	132	67	174
APKT15A	105	236	95	114	97	132	67	174
APKT15B	105	236	95	114	97	132	67	174
APKT15C	105	236	95	114	97	132	67	174
APKT15D	105	236	95	114	97	132	67	174
APKT17A	105	236	95	112	97	132	67	174
APKT17B	105	236	95	112	97	132	67	174
APKT17C	105	236	95	112	97	132	67	174
APKT17D	105	236	95	112	97	132	67	174
APKT18A	105	236	82	118	118	120	77	175
APKT18B	105	-	82	118	118	120	77	175
APKT18C	105	236	82	-	118	120	77	175
APKT18D	105	236	82	118	118	120	77	175
APKT19A	122	238	75	81	118	124	77	174
APKT19B	122	238	91	81	118	124	77	174
APKT19C	-	-	-	-	118	124	77	174
APKT19D	122	238	82	81	118	124	77	174
APKT28A	105	234	80	77	122	105	77	177
APKT28B	105	234	80	122	122	105	77	177
APKT28C	105	234	80	122	122	105	77	177
APKT28D	105	234	80	118	122	105	77	177
APBT102A	120	238	78	85	120	113	77	174
APBT102B	122	236	75	83	113	124	77	165
APBT102C	122	236	75	83	113	124	77	165
APBT102D	122	236	75	83	113	124	77	165
APBT103A	118	236	75	81	118	99	77	174
APBT103B	118	236	75	81	118	99	77	174
APBT103C	118	236	75	81	118	99	77	174
APBT103D	118	236	75	81	118	99	77	174
APBT104A	105	236	80	132	122	91	77	184
APBT104B	105	236	80	132	122	91	77	184
APBT104C	105	236	80	132	122	91	77	184
APBT104D	105	236	80	132	122	91	77	184

continued on next page

continued from previous page

Sample	APBMS23	APBMS61	APMS14	APMS28	APBMS14	APBMS3	APBMS72	APMS4
APBT105A	105	236	80	132	122	91	77	184
APBT105B	105	236	80	132	122	91	77	184
APBT105C	105	236	80	132	122	91	77	184
APBT105D	105	236	80	132	122	91	77	184
APBT107A	122	236	80	88	125	162	81	168
APBT107B	122	236	80	88	125	162	81	168
APBT107C	122	236	80	88	125	162	81	168
APBT107D	122	236	80	88	125	162	81	168
APBT108A	122	236	76	-	125	-	77	175
APBT108B	122	236	76	-	125	-	77	175
APBT108C	122	236	76	152	125	-	77	175
APBT108D	122	236	76	-	125	-	77	175
APBT109A	105	238	82	112	122	95	77	168
APBT109B	105	238	82	112	122	95	77	168
APBT109C	105	238	82	112	122	95	77	168
APBT109D	105	238	82	112	122	95	77	168
APBT110A	122	236	78	88	127	111	77	175
APBT110B	122	236	78	88	127	111	77	175
APBT110C	122	236	78	88	127	111	77	175
APBT110D	122	236	78	88	127	111	77	175
APMR103A	118	236	76	130	117	126	79	174
APMR103B	118	236	76	130	117	126	79	174
APMR103C	118	236	76	130	117	126	79	174
APMR103D	118	236	76	130	117	126	79	174
APMR105A	118	236	76	132	117	126	79	174
APMR105B	118	236	76	132	117	126	79	174
APMR105C	118	236	76	132	117	126	79	174
APMR105D	118	236	76	132	117	126	79	174
APMR106A	118	236	76	132	117	126	79	174
APMR106B	118	236	76	132	117	126	79	174
APMR106C	118	236	76	132	117	126	79	174
APMR106D	118	236	76	132	117	126	79	174
APMR108A	118	236	76	132	117	126	79	174
APMR108B	118	236	76	132	117	126	79	174
APMR108C	118	236	76	132	-	-	-	-
APMR108D	118	236	76	132	117	126	79	174
APMR109A	118	236	76	132	117	126	79	174
APMR109B	118	236	76	132	117	126	79	174
APMR109C	118	236	76	132	117	126	79	174
APMR109D	118	236	76	132	117	126	79	174
APMR111A	105	236	74	89	122	93	79	175
APMR111B	105	236	74	89	122	93	79	175
APMR111C	105	236	74	89	122	93	79	175
APMR111D	105	236	74	89	122	93	79	175
APMR113A	118	238	82	-	122	93	79	175
APMR113B	118	238	82	-	122	93	79	175
APMR113C	118	238	82	-	122	93	79	175
APMR113D	118	238	82	-	122	93	79	175
APMR115A	120	238	76	124	122	111	77	175
APMR115B	122	236	74	85	119	109	77	178
APMR115C	122	236	74	85	119	109	77	178
APMR115D	122	236	74	85	119	109	77	178

Appendix M

WITHIN-CLUMP DATA - SUPPLEMENTARY

Table M.1: Microsatellite marker alleles of supplementary samples used in analysis of within-clump diversity of *Acanthorrhynchium papillatum*. Header rows indicate names of markers. Alleles are PCR fragment lengths in bp. Null alleles are indicated by '-'.

Sample	APBMS23	APBMS61	APMS14	APMS28	APBMS14	APBMS3	APBMS72	APMS4
APBK112A	105	234	80	83	126	93	77	174
APBK112B	105	234	80	83	126	93	77	174
APBK112C	105	234	80	81	123	93	77	174
APBK112D	105	234	80	83	126	93	77	174
APBK112E	105	234	80	83	126	93	77	174
APBK112F	105	234	80	83	126	93	77	174
APBK112G	105	234	80	83	126	93	77	174
APBK112H	105	234	80	83	126	93	77	174
APBK112I	105	234	80	81	123	93	77	174
APBK112J	105	234	80	83	126	93	77	174
APBK112K	105	234	80	81	123	93	77	174
APBK112L	105	234	80	83	126	93	77	174
APBK112M	105	234	80	83	126	93	77	174
APBK112N	105	234	80	83	126	93	77	174
APBK112O	105	234	80	83	126	93	77	174
APBK112P	105	234	80	83	126	93	77	174
APBK112Q	105	234	80	81	123	93	77	174
APBK112R	105	234	80	83	126	93	77	174
APBK112S	105	234	80	83	126	93	77	174
APBK112T	105	234	80	83	126	93	77	174
APBL111A	105	234	76	79	116	91	77	174
APBL111B	105	234	76	79	116	91	77	174
APBL111C	105	234	76	79	116	91	77	174
APBL111D	105	234	76	79	116	91	77	174
APBL111E	105	234	76	79	116	91	77	174
APBL111F	105	234	76	79	116	91	77	174
APBL111G	105	234	76	79	116	91	77	174
APBL111H	105	234	76	79	116	91	77	174
APBL111I	105	234	76	79	116	91	77	174
APBL111J	105	234	76	79	116	91	77	174
APBL111K	105	234	76	79	116	91	77	174
APBL111L	105	234	76	79	116	91	77	174
APBL111M	105	234	76	79	116	91	77	174
APBL111N	105	234	76	79	116	91	77	174
APBL111O	105	234	76	79	116	91	77	174
APBL111P	105	234	76	79	116	91	77	174
APBL111Q	105	234	76	79	116	91	77	174
APBL111R	105	234	76	79	116	91	77	174
APBL111S	105	234	76	79	116	91	77	174
APBL111T	105	234	76	79	116	91	77	174
APKT21A	118	236	78	-	129	91	77	171
APKT21B	118	236	78	-	129	91	77	171
APKT21C	118	236	78	-	129	91	77	171
APKT21D	118	236	78	-	129	91	77	171
APKT21E	118	236	78	-	129	91	77	171
APKT21F	118	236	78	-	129	91	77	171
APKT21G	118	236	78	-	129	91	77	171
APKT21H	118	236	78	108	129	91	77	171
APKT21I	118	236	78	108	129	91	77	171
APKT21J	118	236	78	108	129	91	77	171
APKT21K	118	236	78	-	129	91	77	171
APKT21L	118	236	78	108	129	91	77	171
APKT21M	118	236	78	108	129	91	77	171
APKT21N	118	236	78	108	129	91	77	171
APKT21O	118	236	78	108	129	91	77	171

continued on next page

continued from previous page

Sample	APBMS23	APBMS61	APMS14	APMS28	APBMS14	APBMS3	APBMS72	APMS4
APKT21P	118	236	78	108	129	91	77	171
APKT21Q	118	236	78	108	129	91	77	171
APKT21R	118	236	78	108	129	91	77	171
APKT21S	118	236	78	108	129	91	77	171
APKT21T	118	236	78	-	129	91	77	171
APBT111A	120	238	74	95	97	122	77	172
APBT111B	120	238	74	95	97	122	77	172
APBT111C	120	238	74	95	97	122	77	172
APBT111D	120	238	74	95	97	122	77	172
APBT111E	120	238	74	95	97	122	77	172
APBT111F	120	238	74	95	97	122	77	172
APBT111G	120	238	74	95	97	122	77	172
APBT111H	120	238	74	95	97	122	77	172
APBT111I	120	238	74	95	97	122	77	172
APBT111J	120	238	74	95	97	122	77	172
APBT111K	120	238	74	95	97	122	77	172
APBT111L	120	238	74	95	97	122	77	172
APBT111M	120	238	74	95	97	122	77	172
APBT111N	120	238	74	95	97	122	77	172
APBT111O	120	238	74	95	97	122	77	172
APBT111P	120	238	74	95	97	122	77	172
APBT111Q	120	238	74	95	97	122	77	172
APBT111R	120	238	74	95	97	122	77	172
APBT111S	120	238	74	95	97	122	77	172
APBT111T	120	238	74	95	97	122	77	172
APMR101A	118	236	76	134	117	126	79	174
APMR101B	118	236	76	134	117	126	79	174
APMR101C	118	236	76	134	117	126	79	174
APMR101D	118	236	76	134	117	126	79	174
APMR101E	118	236	76	134	117	126	79	174
APMR101F	118	236	76	134	117	126	79	174
APMR101G	118	236	76	134	117	126	79	174
APMR101H	118	236	76	134	117	126	79	174
APMR101I	118	236	76	134	117	126	79	174
APMR101J	118	236	76	134	117	126	79	174
APMR101K	118	236	76	134	117	126	79	174
APMR101L	118	236	76	134	117	126	79	174
APMR101M	118	236	76	134	117	126	79	174
APMR101N	118	236	76	134	117	126	79	174
APMR101O	118	236	76	134	117	126	79	174
APMR101P	118	236	76	134	117	126	79	174
APMR101Q	118	236	76	134	117	126	79	174
APMR101R	118	236	76	134	117	126	79	174
APMR101S	118	236	76	134	117	126	79	174
APMR101T	118	236	76	134	117	126	79	174

Appendix N

PAIRWISE GENETIC DISTANCES WITHIN A CLUMP, MAIN DATA - METHOD 1

Table N.1: Pairwise genetic distances of samples within a clump as computed using the method of Excoffier *et al.* (2006). Clumps from the same sampling locality are arranged together in the same set of columns. Column entries "A", "B", "C" and "D" refer to the corresponding samples within a clump.

APBK					**APBL**					**APKT**					**APBT**					**APMR**				
	A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
101A	*				102A	*				05A	*				102A	*				103A	*			
101B	5	*			102B	0	*			05B	0	*			102B	7	*			103B	0	*		
101C	3	5	*		102C	0	0	*		05C	0	0	*		102C	7	0	*		103C	0	0	*	
101D	3	4	4	*	102D	0	0	0	*	05D	0	0	0	*	102D	7	0	0	*	103D	0	0	0	*
102A	*				103A	*				10A	*				103A	*				105A	*			
102B	0	*			103B	5	*			10B	1	*			103B	0	*			105B	0	*		
102C	0	0	*		103C	5	0	*		10C	1	0	*		103C	0	0	*		105C	0	0	*	
102D	0	0	0	*	103D	5	0	0	*	10D	1	0	0	*	103D	0	0	0	*	105D	0	0	0	*
104A	*				105A	*				12A	*				104A	*				106A	*			
104B	0	*			105B	0	*			12B	1	*			104B	0	*			106B	0	*		
104C	0	0	*		105C	0	0	*		12C	0	1	*		104C	0	0	*		106C	0	0	*	
104D	0	0	0	*	105D	0	0	0	*	12D	0	1	0	*	104D	0	0	0	*	106D	0	0	0	*
107A	*				106A	*				15A	*				105A	*				108A	*			
107B	0	*			106B	0	*			15B	0	*			105B	0	*			108B	0	*		
107C	0	0	*		106C	0	0	*		15C	0	0	*		105C	0	0	*		108C	0	0	*	
107D	0	0	0	*	106D	0	0	0	*	15D	0	0	0	*	105D	0	0	0	*	108D	0	0	0	*
108A	*				107A	*				17A	*				107A	*				109A	*			
108B	0	*			107B	0	*			17B	0	*			107B	0	*			109B	0	*		
108C	0	0	*		107C	0	0	*		17C	0	0	*		107C	0	0	*		109C	0	0	*	
108D	0	0	0	*	107D	0	0	0	*	17D	0	0	0	*	107D	0	0	0	*	109D	0	0	0	*
111A	*				108A	*				18A	*				108A	*				111A	*			
111B	2	*			108B	0	*			18B	0	*			108B	0	*			111B	0	*		
111C	2	0	*		108C	0	0	*		18C	0	0	*		108C	0	0	*		111C	0	0	*	
111D	0	2	2	*	108D	0	0	0	*	18D	0	0	0	*	108D	0	0	0	*	111D	0	0	0	*
113A	*				109A	*				19A	*				109A	*				113A	*			
113B	0	*			109B	0	*			19B	1	*			109B	0	*			113B	0	*		
113C	5	5	*		109C	0	0	*		19C	0	0	*		109C	0	0	*		113C	0	0	*	
113D	0	0	5	*	109D	4	4	4	*	19D	1	1	0	*	109D	0	0	0	*	113D	0	0	0	*
115A	*				110A	*				28A	*				110A	*				115A	*			
115B	0	*			110B	0	*			28B	1	*			110B	0	*			115B	5	*		
115C	0	0	*		110C	4	4	*		28C	1	0	*		110C	0	0	*		115C	5	0	*	
115D	0	0	0	*	110D	0	0	4	*	28D	1	1	1	*	110D	0	0	0	*	115D	5	0	0	*

Appendix O

PAIRWISE GENETIC DISTANCES WITHIN A CLUMP, SUPPLEMENTARY DATA - METHOD 1

Table O.1: Pairwise genetic distances of 20 samples from a single clump, APBK1112, from Sungei Bantang Recreational Forest as computed using the method of Excoffier *et al.* (2006). Row and column entries "A", "B", "C", etc. refer to the corresponding samples within a clump.

	APBK112																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	*																			
B	0	*																		
C	2	2	*																	
D	0	0	2	*																
E	0	0	2	0	*															
F	0	0	2	0	0	*														
G	0	0	2	0	0	0	*													
H	0	0	2	0	0	0	0	*												
I	2	2	0	2	2	2	2	2	*											
J	0	0	2	0	0	0	0	2	0	*										
K	0	0	2	0	0	0	0	2	0	0	*									
L	0	0	2	0	0	0	0	2	0	0	0	*								
M	0	0	2	0	0	0	0	2	0	0	0	0	*							
N	0	0	2	0	0	0	0	2	0	0	0	0	0	*						
O	0	0	2	0	0	0	0	2	0	0	0	0	0	0	*					
P	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	*				
Q	2	2	0	2	2	2	2	0	2	2	2	2	2	2	2	2	*			
R	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	2	0	*		
S	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	*	
T	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	*

Appendix P

PAIRWISE GENETIC DISTANCES WITHIN A CLUMP, MAIN DATA - METHOD 2

Table P.1: Pairwise genetic distances of samples within a clump as computed using the method of Slatkin (1995). Clumps from the same sampling locality are arranged together in the same set of columns. Column entries "A", "B", "C" and "D" refer to the corresponding samples within a clump.

APBK					APBL					APKT					APBT					APMR				
	A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
101A	*				102A	*				05A	*				102A	*				103A	*			
101B	739	*			102B	0	*			05B	0	*			102B	272	*			103B	0	*		
101C	388	143	*		102C	0	0	*		05C	0	0	*		102C	272	0	*		103C	0	0	*	
101D	955	746	739	*	102D	0	0	0	*	05D	0	0	0	*	102D	272	0	0	*	103D	0	0	0	*
102A	*				103A	*				10A	*				103A	*				105A	*			
102B	0	*			103B	1355	*			10B	16	*			103B	0	*			105B	0	*		
102C	0	0	*		103C	1355	0	*		10C	16	0	*		103C	0	0	*		105C	0	0	*	
102D	0	0	0	*	103D	1355	0	0	*	10D	16	0	0	*	103D	0	0	0	*	105D	0	0	0	*
104A	*				105A	*				12A	*				104A	*				106A	*			
104B	0	*			105B	0	*			12B	1521	*			104B	0	*			106B	0	*		
104C	0	0	*		105C	0	0	*		12C	0	1521	*		104C	0	0	*		106C	0	0	*	
104D	0	0	0	*	105D	0	0	0	*	12D	0	1521	0	*	104D	0	0	0	*	106D	0	0	0	*
107A	*				106A	*				15A	*				105A	*				108A	*			
107B	0	*			106B	0	*			15B	0	*			105B	0	*			108B	0	*		
107C	0	0	*		106C	0	0	*		15C	0	0	*		105C	0	0	*		108C	0	0	*	
107D	0	0	0	*	106D	0	0	0	*	15D	0	0	0	*	105D	0	0	0	*	108D	0	0	0	*
108A	*				107A	*				17A	*				107A	*				109A	*			
108B	0	*			107B	0	*			17B	0	*			107B	0	*			109B	0	*		
108C	0	0	*		107C	0	0	*		17C	0	0	*		107C	0	0	*		109C	0	0	*	
108D	0	0	0	*	107D	0	0	0	*	17D	0	0	0	*	107D	0	0	0	*	109D	0	0	0	*
111A	*				108A	*				18A	*				108A	*				111A	*			
111B	1229	*			108B	0	*			18B	0	*			108B	0	*			111B	0	*		
111C	1229	0	*		108C	0	0	*		18C	0	0	*		108C	0	0	*		111C	0	0	*	
111D	0	1229	1229	*	108D	0	0	0	*	18D	0	0	0	*	108D	0	0	0	*	111D	0	0	0	*
113A	*				109A	*				19A	*				109A	*				113A	*			
113B	0	*			109B	0	*			19B	256	*			109B	0	*			113B	0	*		
113C	1531	1531	*		109C	0	0	*		19C	0	0	*		109C	0	0	*		113C	0	0	*	
113D	0	0	1531	*	109D	1004	1004	1004	*	19D	49	81	0	*	109D	0	0	0	*	113D	0	0	0	*
115A	*				110A	*				28A	*				110A	*				115A	*			
115B	0	*			110B	0	*			28B	2025	*			110B	0	*			115B	1555	*		
115C	0	0	*		110C	593	593	*		28C	2025	0	*		110C	0	0	*		115C	1555	0	*	
115D	0	0	0	*	110D	0	0	593	*	28D	1681	16	16	*	110D	0	0	0	*	115D	1555	0	0	*

Appendix Q

PAIRWISE GENETIC DISTANCES WITHIN A CLUMP, SUPPLEMENTARY DATA - METHOD 2

Table Q.1: Pairwise genetic distances of 20 samples from a single clump from Sungei Bantang Recreational Forest as computed using the method of Slatkin (1995). Column entries "A", "B", "C", etc. refer to the corresponding samples within a clump.

	APBK112																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	*																			
B	0	*																		
C	16	16	*																	
D	0	0	16	*																
E	0	0	16	0	*															
F	0	0	16	0	0	*														
G	0	0	16	0	0	0	*													
H	0	0	16	0	0	0	0	*												
I	16	16	0	16	16	16	16	16	*											
J	0	0	16	0	0	0	0	16	0	*										
K	0	0	16	0	0	0	0	16	0	0	*									
L	0	0	16	0	0	0	0	16	0	0	0	*								
M	0	0	16	0	0	0	0	16	0	0	0	0	*							
N	0	0	16	0	0	0	0	16	0	0	0	0	0	*						
O	0	0	16	0	0	0	0	16	0	0	0	0	0	0	*					
P	0	0	16	0	0	0	0	16	0	0	0	0	0	0	0	*				
Q	16	16	0	16	16	16	16	0	16	16	16	16	16	16	16	16	*			
R	0	0	16	0	0	0	0	16	0	0	0	0	0	0	0	16	0	*		
S	0	0	16	0	0	0	0	16	0	0	0	0	0	0	0	16	0	0	*	
T	0	0	16	0	0	0	0	16	0	0	0	0	0	0	0	16	0	0	0	*