Chapter 6

Phylogeography of the sponge *Suberites diversicolor* in Indonesia: insights into the evolution of marine lake populations

Leontine E. Becking, Dirk Erpenbeck, Katja T.C.A. Peijnenburg, Nicole J. de Voogd

Submitted manuscript

Abstract

Phylogeographic studies of taxa inhabiting marine lakes provide excellent opportunities to study biogeographical relationships and population structures of marine species in isolated habitats. Marine lakes are landlocked water bodies that maintain a marine character through narrow submarine connections to the sea and could be regarded as the marine equivalents of terrestrial islands. The sponge *Suberites diversicolor* (Demospongiae: Suberitidae) is typical of marine lake habitats in the Indo-Australian Archipelago. We employed four molecular markers (two mitochondrial and two nuclear) to study its genetic structure within and between marine lakes in Indonesia and three coastal locations in Indonesia, Singapore and Australia. Within the populations of *S. diversicolor* there were two strongly divergent lineages (A & B) (COI: *p*=0.4% and ITS: *p*=7.3%), that may constitute cryptic species. Lineage A only occurred in Kakaban lake (East Kalimantan), while lineage B was present in all sampled populations. Within lineage B, we found low levels of genetic diversity in lakes and strong spatial genetic population structuring. The Australian population is genetically differentiated from the Indonesian populations. Within Indonesia with we did not record an East-West barrier which has frequently been reported for other marine invertebrates. Kakaban lake is the largest and most isolated marine lake in Indonesia and contains the highest genetic diversity with genetic variants not observed elsewhere. Kakaban may be an area where multiple putative refugia populations have come into secondary contact, resulting in the high genetic diversity and the high number of endemics.

Keywords

anchialine • cryptic species • Indo-Australian Archipelago • marine lakes • COI • ITS

Introduction

It has long been hypothesized that marine species have large geographic ranges with large population sizes, and are faced with weaker barriers to dispersal than terrestrial organisms, thus resulting in relatively slow rates of speciation (Palumbi 1994). The assumed presence of circum-tropical species has for long supported this view. However, recent phylogeographic and population genetic studies on marine taxa portray a situation of ecologically heterogeneous environments on small spatial scales with several morphologically cryptic species instead of cosmopolitan species (e.g. Knowlton 2000, Barber et al. 2002, Peijnenburg et al. 2004, Nuryanto & Kochzius 2009, Malay & Paulay 2010, Reveillaud et al. 2010, Xavier et al. 2010a, Gaither et al. 2011). These results suggest that there may be many more barriers to dispersal at small spatial scales than we are able to observe (Palumbi 1994, Carpenter et al. 2011, Goetze & Peijnenburg *submitted*). The existence of multiple independently derived populations in landlocked marine lakes provides an opportunity for fundamental research into the role of isolation in population divergence and speciation in marine taxa (Dawson & Hamner 2005). Marine lakes are anchialine systems, which are landlocked water bodies that maintain a marine character through narrow submarine connections to the sea (Holthuis 1973, Hamner & Hamner 1998). The lakes are formed in natural inland depressions and display a tidal regime which is typically delayed and damped compared to the adjacent sea (Holthuis 1973, Hamner & Hamner 1998, Becking et al. 2011, *CHAPTER 1*). The level of obstruction of water echange, i.e. the degree of isolation, differs per lake as does the salinity and environmental regimes within the lakes (Hamner & Hamner 1998, Becking et al. 2011, *CHAPTER 1*). The number of marine lakes worldwide is estimated at approximately 200 with clusters of ten or more lakes occurring in areas with a karstic limestone landscape such as Croatia, Bermuda, Vietnam, Palau, and Indonesia (Dawson 2009).

The marine lakes share many characteristics with island systems (Dawson 2006): they are well-defined geographically (Hamner & Hamner 1998, Colin 2009, Becking et al. 2011, *CHAPTER 1*), harbor unique biota with high endemism and/or an abundance of species rare that are elsewhere (Tomascik & Mah 1994, Tomascik et al. 1997, Azzini et al. 2007, Dawson 2009, Becking et al. 2011, *CHAPTER 1, 3, 4 & 5*), and isolated populations (Dawson & Hamner 2005, Gotoh et al. 2009, Goto et al. 2011). The marine lakes in the Indo-Pacific were formed less than 15000 years ago (Dawson 2006, Sathiamurthy & Voris 2006), yet their biodiversity is unique. Recent comprehensive studies of sponge assemblages of marine lakes, coastal mangroves and coral reefs in Berau (East Kalimantan, Indonesia) indicated that these lakes harbor a significantly different assemblage consisting of a subset of the fauna of the adjacent (de Voogd et al. 2009, Becking et al. *accepted & unpublished data, CHAPTERS 2 & 3*). Extensive surveys have shown that the lake assemblages represent three groups of sponge species: 1. widespread species known from various coastal locations in Indo-Pacific reefs, 2. lake species that only occur in lake systems, 3. endemic species restricted to only one single lake (de Voogd et al. 2009, Becking et al. *accepted & unpublished data, CHAPTERS 2 & 3*). Marine lakes can significantly contribute to the regional diversity because they harbor predominantly lake and endemic species (Becking et al. *accepted & unpublished data, CHAPTER 3*). Consistent with the island biogeography theory (MacArthur & Wilson 1967, Whittaker & Fernandez-Palacios 2007, Rosindell et al. 2011) larger lakes have more species than smaller ones and the most isolated lakes contain few reef species and the highest number of putative endemics, while the more connected lakes are dominated by reef species (Becking et al. 2011, Becking et al. *accepted& unpublished data, CHAPTERS 1, 2 & 3*). In the present study our overall aim was to obtain insight into the role of isolation in genetic diversity of marine lake populations by studying population structures of a selected sponge species among Indonesian lakes and between the lakes and the sea.

Phylogeographic studies of anchialine systems across the world typically show high levels of genetic differentiation between marine lake populations, suggesting little to no gene flow at small spatial scales ranging from 10 to 100 km (see Table 1). Furthermore, molecular markers revealed the presence of highly divergent, but morphologically cryptic species in a number of taxa such as cnidarians, crustaceans, fish and mollusks (Dawson & Hamner 2005, Santos 2006, Craft et al. 2008, Page et al. 2008, Gotoh et al. 2009, Botello & Alvarez 2010, Bauzà-Ribot et al. 2011, Goto et al. 2011). There are, however, exceptions to this general pattern which have been interpreted as resulting from life history strategies involving greater dispersal capabilities (Russ et al. 2009). Too limited sampling may also have prevented any observation of patterns (Kano & Kase 2004). Here we have conducted the first phylogeographic study of Indonesian marine lake populations and this is also the first phylogeographic study on sponges in Indonesia. Sponges are one of the most dominant taxa in marine lakes in terms of biomass and species diversity (Tomascik & Mah 1994, Becking et al. 2011, *CHAPTER* 1). The sponge species *Suberites diversicolor* (Demospongiae: Suberitidae) is an ideal taxon to pursue this study as it allows comparison of multiple lakes at various scales and with varying degrees of connection to the sea. There are few other species that are prevalent in marine lakes (Becking et al. 2011, *CHAPTER* 1). *Suberites diversicolor* occurs in most moderately to highly isolated marine lakes in Indonesia (Becking et al. 2011), as well as in limited numbers of small populations in sheltered bays in Singapore, Indonesia and Australia (Becking & Lim 2009, *CHAPTER 4*). *Suberites diversicolor* shows great plasticity in adapting to harsh environments (low salinity and exposure to air) yet is absent in coral reefs (Azzini et al. 2007, de Voogd et al. 2009, Lim et al. 2009, Becking et al. 2011, *CHAPTER 1, 2 & 4*).

Our specific aims were: 1) to estimate levels of diversity and divergence of seven marine lake populations and three coastal populations using two mitochondrial and two nuclear markers, 2) to study the phylogeography of *S. diversicolor* populations in marine lakes across Indonesia, 3) to investigate the relationship between the genetic diversity and the level of isolation of the lakes.

Table 1. Overview of intra-specific genetic structure of populations in anchialine systems from literature.

Anchialine system	Location	Taxon	Marker(s)	Structure	Scale of differentiation	Reference
lake	Palau	*Mastigias papua*	mtDNA COI & nDNA ITS	yes, each lake private haplotypes	1-50km	Dawson & Hamner 2005, Dawson 2005
lake	Palau	*Brachidontes* sp.	mtDNA COI	yes, divergent species; each lake private haplotypes	1-50km	Goto et al. 2011
lake	Palau	*Sphaeramia orbicularis*	mtDNA control region	yes, lakes reduced diversity, private haplotypes	1-50 km	Gotoh et al. 2009
pool	Hawaii island	*Holocaridina rubra*	mtDNA COI	yes, each pool private haplotypes	30-50 km	Santos 2006
pool	Hawaii Archipelago	*Halocaridina rubra*	mtDNA COI	yes, each pool private haplotypes	10-50 km	Craft et al. 2008
pool	Maui &Hawaii	*Halocaridina rubra*	mtDNA COI	yes, each pool private haplotypes	1-100 km	Santos & Weese 2011
pool	Hawaii Archipelago	*Metabenaeus lohena*	mtDNA COI	panmixia	25-300 km	Russ et al. 2009
cave	Philippines	*Neritilia cavernicola*	mtDNA COI	panmixia	200 km	Kano & Kase 2003
cave	Australia	*Stygiocaris lancifera*	mtDNA COI, 16S	yes, divergent species	10-100 km	Page et al. 2008
cave	Spain	*Metacrangonyx longipes*	mtDNA COI 16S, histone	yes, divergent species	20-100 km	Bauza-Ribot et al. 2011
cave	Mexico	*Creaseria morleyi*	mtDNA COI, 16S	yes, divergent populations	10-100 km	Botello & Alvarez 2010

Sampling

Twenty four marine lakes and adjacent coastal habitats in Indonesia were thoroughly surveyed by snorkeling for the presence of the sponge *Suberites diversicolor.* Populations of *Suberites diversicolor* were located in seven marine lakes (29% of all surveyed lakes) in the region of Berau, East Kalimantan province (Kakaban lake, Haji Buang lake, Tanah Bamban lake) and the regions of Northern Raja Ampat (Cassiopeia lake, Urani lake, Sauwandarek lake) and Southern Raja Ampat (Misool Jellyfish lake) in West Papua province, and in mangroves along the coast of the island of Maratua in the region of Berau, East Kalimantan province. Additional coastal populations were sampled from Johor Straight in Singapore (collected by S.C. Lim) and the man-made Lake Alexander in Darwin, Australia (collected by B. Alvarez), resulting in a total of seven marine lake populations and three coastal populations sampled for this study (Figure 1A). The lakes Kakaban, Tanah Bamban, Haji Buang and Misool house immense perennial populations of the jellyfish *Mastigias papua* such as those that have been extensively documented in five marine lakes in Palau. For a full description of the sampled marine lakes, see Becking et al. (2011, *CHAPTER 1*). These landlocked pools of water are subjected to a tidal regime which is typically delayed in phase (ranging from 20 minutes to 4 hours) and dampened in amplitude (tide ranging from 20 cm to 1.5 m) compared to the adjacent sea (Tomascik & Mah 1994, Becking et al. 2011, *CHAPTER 1*). We used the amount of tidal delay and dampening as a proxy for the degree of connection between the lake and the adjacent sea (Hamner & Hamner 1998, Becking et al. 2011, *CHAPTER 1*). The relative degree of isolation of each marine lake is provided in Table 2.

Collections were made randomly along the entire coastline of each of the lakes and specimens were collected at least 25m distance from each other to avoid collecting clone siblings. Our aim was to collect 20 individuals per location, but in most locations the resident population size was too small to attain this target (see Table 2). Hence, sample sizes are small for some locations. The color and substrate of each specimen was recorded, and a photograph was taken either *in situ* or within 2 hours after collection. After collection, portions of the choanosome were cut into approximately 125 mm^3 cubes, avoiding the surface to minimize potential contamination with protists or other sponge associates, and preserved in 96% ethanol, which was refreshed after 24 hours. The remainder of the samples were preserved in 70% ethanol and deposited in the Porifera collection of the Naturalis Biodiversity Center, The Netherlands (RMNH POR.) as voucher specimens. The investigated specimens are listed in the Appendix 1 (page 167).

Table 2. Sample localities of ten *Suberites diversicolor* populations from marine lakes and coastal locations in Berau (East Kalimantan) and Raja Ampat (West Papua) in Indonesia, Darwin in Australia, and Singapore. Per locality relative connection to the adjacent sea is provided, and for the marine lake size, number of species and proportion of putative endemics. In addition, the density of the target sponge species *Suberites diversicolor,* color morphs, and number of samples per genetic marker (COI, COII, 28S, ITS) is provided per location.

Code	Location	Region	Latitude	Longitude	Connection	Size lake (1000 m^3)	# sponge species	% putative endemic sponge species	density *S.diversicolor* (ind. 50m^2)	color morphs *S.diversicolor*	nCOI	nITS	nCOII	n28S
Lake														
KKB	Kakaban lake	Berau	N02° 08' 23.5"	E118° 30' 31.9"	most isolated	4000	67	49%	1-15	green, red	22	21	21	5
HBL	Haji Buang lake	Berau	N02° 12' 30.4"	E118° 35' 40.8"	isolated	140	53	34%	15-50	green, red, blue, purple, yellow	20	20	20	2
TBB	Tanah Bamban lake	Berau	N02° 13' 50.0"	E118° 34' 50.7"	least isolated	120	17	0	0-2	green, red	4	4	4	2
RAJ	Sauwandarek lake	Raja Ampat	S0° 35' 19.6"	E130° 35' 48.8"	very isolated	84	20	20%	0-10	purple, blue, green	21	21	21	2
CAS	Cassiopeia lake	Raja Ampat	N0° 08' 36.6"	E130° 04' 39.8"	least isolated	13	15	0	0-10	green	10	10	10	2
URA	Urani lake	Raja Ampat	N0° 06' 05.1"	E130° 15' 05.5"	isolated	68	25	12%	0-2	green	8	8	8	1
MIS	Misool Jellfish Lake	Raja Ampat	S01° 55'	E130° 20'	isolated	12	8	25%	0-2	green	7	7	7	1
Coastal														
BER	Maratua mangrove	Berau	N02° 12' 52.3"	E118° 35' 34.1"	open				0-1	green, yellow	3	3	3	1
DAR	Lake Alexander, Darwin	Australia	S12° 25'	E130° 50'	open				0-1	green	6	6	6	2
SIN	Johor Strait	Singapore	N 01° 26'02.34"	E104°02'54.31"	open				0-1	purple, blue, green	4	4	4	2

DNA extraction, amplification and sequencing

Total DNA was extracted from 105 specimens using DNeasy tissue kit (Qiagen), following the instructions of the manufacturer. Partitions of four markers were amplified: two mitochondrial genes, cytochrome oxidase subunit 1 (COI) and subunit 2 (COII), and two nuclear markers, the nuclear ribosomal operons consisting of partial 18S rDNA, full-length internal transcribed spacer 1 and 2, 5.8S, and partial 28S rDNA fragments (ITS) and the D3-D5 region of the nuclear ribosomal 28S gene (28S). The nuclear markers are independent from the mitochondrial markers and therefore provide extra support in case of congruent results.

The standard DNA-barcoding fragment of COI was amplified by using a specific forward primer designed for *Suberites* SUB-COI-F: GGAATGATCGGGACAGCTTTTAGCATG and the degenerated reverse primer from Folmer et al. (1994) designed by Meyer et al. (2005): dgHCO2198:TAA ACT TCA GGG TGA CCA AAR AAY CA. COII was amplified with the primers from Rua et al. (2011): CO2F: TTTTTCACGATCAGATTATGTTTA and CO2R: ATACTCGCACTGAGTTTGAATAGG. ITS amplified with primers from Wörheide (1998) RA2: GTCCCTGCCCTTTGTACACA and ITS2.2: CCT GGT TAG TTT CTT TTC CTC CGC). 28S was amplified in a subset of samples with primers from McCormack and Kelly (2002) RD3A: GACCCGTCTTGAAACACGA and RD5B2: ACACACTCCTTAGCGGA . Amplifications were carried out in 25μl reaction volumes containing 5μl Phire® Reaction Buffer,3 μl dNTPs (1mM), 0.625 μl of each primer (10μM), 0.25 μl Phire® Hotstart-*Taq* polymerase DNA (Thermo Scientific, Finnzymes), and 1 μl of DNA (10-20 *n*g/μl). The temperature regime for amplification: 94°C for 30s; followed by 35 cycles of 94°C for 5 s; 50°C for 5 s; 72°C for 12 s; followed by 72°C for 1 min. PCR products were purified and sequenced by Macrogen Inc (Korea and The Netherlands).

Data analysis

Genetic diversity (COI, COII, ITS, 28S)

The poriferan origin of the obtained sequences was verified through BLAST searches (http://blast.ncbi.nlm.nih.gov/blast.cgi). Sequences were handled in SEQUENCHER 4.10.1 (Gene Codes Corporation) and aligned with CLUSTALW and MUSCLE implemented in DAMBE (Xia & Xie 2001) and SEAview v 4.3.0 (Gouy et al. 2010). Alignment was conducted under default settings and optimized by eye. Alignments were collapsed to contain only unique sequence types in DAMBE. Haplo/genotypes and nucleotide diversity as well as Tajima's D neutrality test was calculated per population with Arlequin v. 3.11 (Excoffier et al. 2005).

Phylogeography (COI, ITS, color morphs, substrate preference)

We used ITS outgroup sequences obtained from Genbank from the family Halichondriidae (Figure 1), as the available sequences for ITS of other Suberitidae were more distant than those from Halichondriidae. Several studies have shown that the families Suberitidae and Halichondriidae are sistergroups (Chombard et al. 1998, Chombard & Boury-Esnault 1999, Morrow et al. 2012). To be consistent we also used species of the family Halichondriidae for the outgroup of the COI phylogram. The best-fit DNA substitution model was selected by the Akaike Information Criterion deployed in jMODELTEST v. 0.1.1 (Posada 2008) and this model (COI: HKY and ITS:GTR+G+I for both markers) was used for subsequent Bayesian and maximum likelihood phylogeny inferences. Phylogenetic reconstructions were performed under Bayesian inference criteria implemented in MrBayes v. 3.1.2. (Huelsenbeck & Ronquist 2001). Each analysis consisted of two independent runs of four Metropolis-coupled Markov-chains, sampled at every 1000th generation at the default temperature (0.2). Analyses were terminated after the chains converged significantly as indicated by an average

standard deviation of split frequencies <0.001. Convergence was also checked in Tracer v. 1.5.0 (Rambaut & Drummond 2007). For comparison, maximum likelihood bootstrap analyses were conducted using MEGA v. 5.01 (Tamura et al. 2011) using a heuristic search with 1000 bootstrap replicates. The Bayesian and maximum likelihood phylograms were combined and visualized using Treegraph 2 (Stöver & Müller 2010). Within group *p*-distance (uncorrected), as well as net nucleotide divergence between groups were calculated in MEGA. A Kruskall-Wallace test was performed to test whether color or substrate preference significantly differed between lineages. To test for spatial structuring of samples we performed an analysis of molecular variance (AMOVA) and calculated pairwise *Φst* values between separate populations using Arlequin 3.5.1.2 (Excoffier & Lischer 2010). Significance of pairwise *Φst* values (based on *p*-distances) was determined by 10000 permutations and exact tests of population differentiation in Arlequin.

Table 3. Genetic diversity indices based on ITS sequences per population of *Suberites diversicolor* of lineage A and B (location codes indicated in Table 2); gene diversity (*h*), nucleotide diversity (π), Tajima's D neutrality test. The majority of populations had only one haplotype resulting in 0 values for all indices calculated. All Tajima D values are not significant.

Code	Lineage	n ITS	*h* ITS	π ITS	Tajima's D
KKB	A	13	0.8242 +/- 0.0567	0.005656 +/- 0.003392	1.3927
KKB	B	8	0.5357 +/- 0.1232	0.001578 +/- 0.001318	1.4488
HBL	B	20	0	0	0
TBB	B	4	0	0	0
RAJ	B	21	0	0	0
CAS	B	10	0	0	0
URA	B	8	0	0	0
MIS	B	7	0.2857 +/- 0.1964	0.000842 +/- 0.000879	-1.23716
BER	B	3	0	0	0
SIN	B	4	0	0	0
DAR	B	6	0.3333 +/- 0.2152	0.000980 +/- 0.000997	-1.13197

Table 4. Pairwise *Φst* values between all populations of lineage B based ITS sequences (location codes indicated in Table 2). Values in bold and with asterisk indicate significant values (p<0.05).

	KKB	HBL	TBB	RAJ	CAS	URA	MIS	BER	SIN
HBL	**0.60591***								
TBB	0.30435	0							
RAJ	**0.60591***	**1***	**1***						
CAS	**0.4702***	**1***	**1***	0					
URA	0.42857	**1***	**1***	0	0				
MIS	0.13514	0.16749	-0.09804	**0.91393***	**0.86315***	**0.84466***			
BER	0.25	**1***	**1***	0	0	0	**0.76136***		
SIN	0.30435	**1***	**1***	0	0	0	**0.78544***	0	
DAR	**0.53451***	**0.9512***	**0.86348***	**0.83584***	**0.74359***	**0.71049***	**0.77327***	0.55882	0.60396

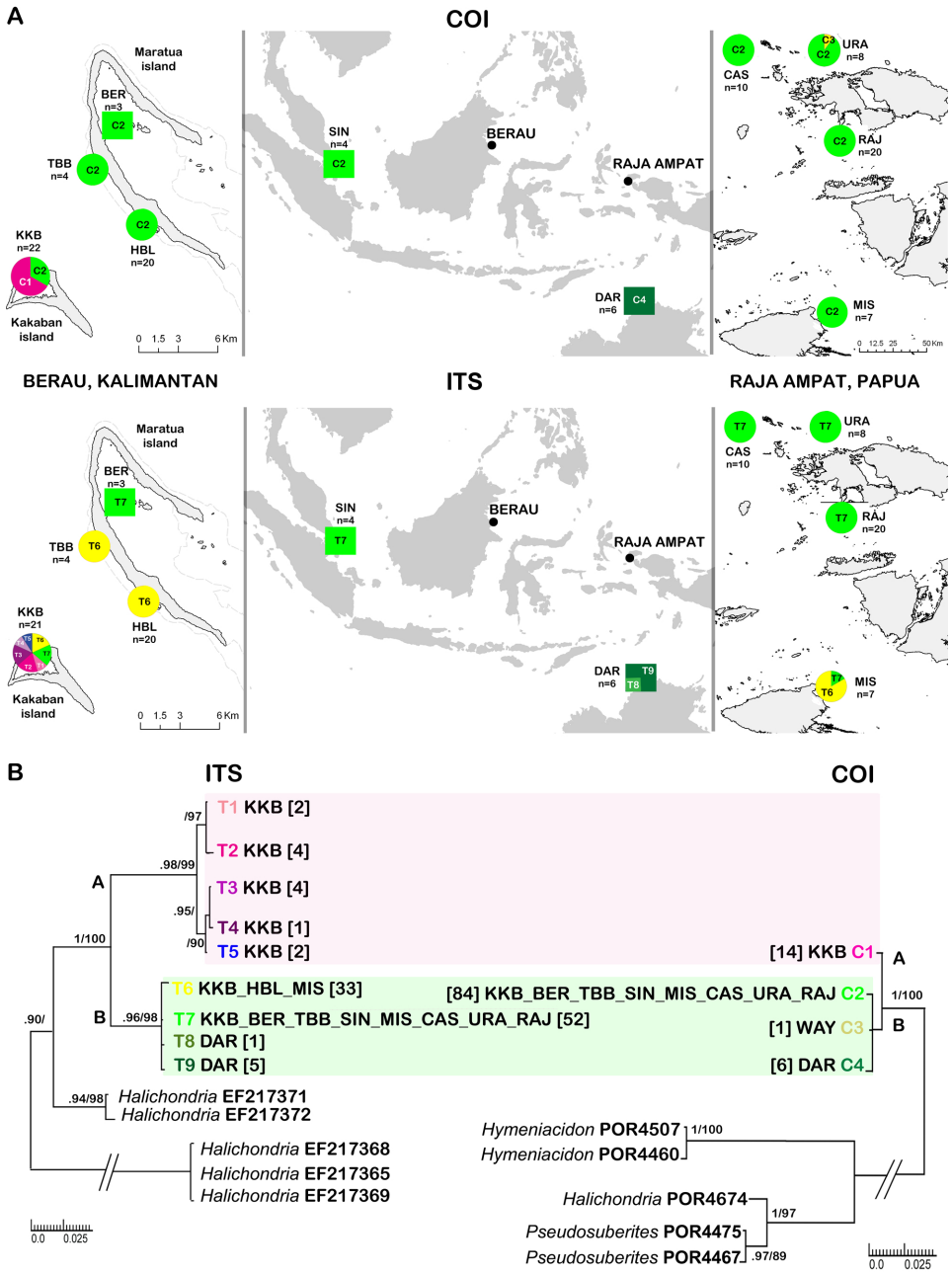

Figure 1. A Sample locations of the sponge *Suberites diversicolor*: top map represents distribution and frequencies of haplotypes for partial Cytochrome Oxidase I (COI) and below of genotypes of internal transcribed spacer region of nuclear ribosomal operons (ITS) in Indonesia, Singapore and Australia with insets of Berau, East Kalimantan (left) and Raja Ampat, West Papua (right) in Indonesia; location codes are explained in Table 2; circles represent marine lakes and squares are outer coastal populations; haplo/genotypes are indicated by number code (COI: C1-4 and ITS: T1-9) and color codes as provided in B. Note that scale differs per map. **B** Bayesian/ maximum likelihood phylogram of 105 COI sequences (right) and 104 ITS sequences (left); each haplo/genotype indicated by specific color, followed by location code and total number of samples in squared brackets. Only posterior probabilities of >90 and maximum likelihood values of >70 are indicated. Color blocks represent the same individuals for both molecular markers (i.e. lineage A (pink) and B (green) represented by the same individuals with both COI and ITS markers). Scale bars indicate substitutions/site.

Results

Sequence variation (COI, COII, ITS, 28S)
We obtained final alignments (excluding primers) for the sponge *S. diversicolor* of 519 bp for COI with four haplotypes (C1-4, 105 individuals), 331 bp for COII with one haplotype (105 individuals), 689 bp of ITS with nine genetic variants (T1-9, 104 individuals) (Table 2, Appendix 1). For a subset of 20 specimens we obtained 574 bp for 28S resulting in one genetic variant.

Divergent lineages in S. diversicolor
All COI and ITS sequences were obtained from the same specimens and fall apart into two major lineages, termed A and B. These lineages represent reciprocally monophyletic groups for both markers and were strongly supported by both Bayesian and maximum likelihood inference methods (Fig. 1B). Lineage A was represented by haplotype C1 for COI and genotypes T1-5 for ITS. Lineage B is represented by COI haplotypes C2-4 and ITS genotypes T6-9 (Figure 1B). Within lineage A there was no sequence variation in COI (n=14), while the average *p*-distance within lineage B was 0.25% (n= 91). The net nucleotide divergence between lineage A & B for COI was 0.38%. Haplotype C1 (lineage A) differed by two basepairs from C2 (the dominant haplotype from lineage B) of which one resulted in a non-synonymous substitution between two unpolar amino acids, from isoleucine to valine, when translated into proteins. For ITS the average *p*-distance within lineage A was 0.44% (n=13), while the average *p*-distance within lineage B was 0.29% (n=91). The net nucleotide divergence between lineages A & B for ITS was 7.26%. Several indels of 1-3 bp length were observed and were consistent within lineages and differed between lineages. There were insertions in lineage A with respect to lineage B from bp102-103 (either CT or TT), 380-381 (CA), 470-473 (GGA or GAA). There were gaps in lineage A with respect to lineage B from bp139, 178-180, 549-55. No double peaks were observed, and it is assumed that no intragenomic polymorphisms occur within this species. The level of intragenomic polymorphisms differ per species (Wörheide et al. 2004). We consider the risk of analyzing paralogous rDNA sequence types to be minimal as we see genealogical concordance across two unlinked loci. We did not, however, detect a significant difference between lineage A & B in color (p=0.249) or substrate preference (p=0.100) using the independent samples Kruskal-Wallis Test.

Diversity and spatial population structuring (COI & ITS)
Lineage A was only present in Kakaban lake while lineage B was present in all populations. The geographical distribution of COI haplotypes is shown in Fig. 1A. Of the four detected haplotypes in COI, haplotype C1 was restricted to Kakaban lake (East Kalimantan). Haplotype C3 only occurred in one individual in Urani lake (West Papua). The Darwin population was represented by haplotype C4 which was shared with no other population. Haplotype C2 was the most abundant haplotype, occurring in all populations except Darwin and was the dominant haplotype in the populations of Berau mangroves, Singapore, Sauwandarek lake, Cassopeia lake, Urani Lake, Misool Jellyfish lake. Of the nine detected genotypes of ITS, five were restricted to Kakaban lake (genotypes T1-5), all representatives of lineage A. Genotype T7 (lineage B) was the most abundant and was shared by all sampled populations except Haji Buang Lake and Tanah Bamban lake (Kalimantan) and Darwin (Australia). Darwin was represented by the private haplotypes T8-9. Haji Buang lake and Tanah Bamban lake were represented fully by a single genotype (T6) that was shared by Kakaban lake (Kalimantan) and Misool Jellyfish lake (Papua).

Within lineage A in Kakaban lake there was only a single haplotype of COI while the ITS gene diversity was 0.8242 +/- 0.0567, and ITS nucleotide diversity was 0.005656 +/- 0.003392. Within lineage B all populations contained a single COI haplotype except Urani lake, which had two haplotypes with a haplotype diversity of 0.3333 +/- 0.2152 and nucleotide diversity of 0.000624 +/- 0.000822. For ITS, the majority of the populations contained only a single genotype, except for Kakaban lake, Darwin and Misool Jellyfish lake. The population in Kakaban lake had the highest gene and nucleotide diversity in lineage B, followed by Darwin and Misool Jellyfish lake (Table 3). Tajima's D tests of neutrality were carried out per population. The majority of the populations had a zero value due to the presence of only one genetic variant. Values of Tajima's D for ITS were negative, but not significant ($p>0.1$) in Misool lake and Darwin (Table 3).

Spatial analysis of genetic structure of lineage B haplo/genotypes of COI and ITS showed that the Darwin population was strongly (*Φst* between 0.53-1) and significantly different from all marine lakes populations (Table 4 & Appendix 2). Besides Darwin there was no significant variation in COI between the different populations (Appendix 2). ITS showed more structure among the populations than COI (Figure 1, Table 3). The Berau lakes (East Kalimantan) Kakaban and Haji Buang lakes were significantly differentiated. The Raja Ampat lakes (West Papua) were not genetically differentiated from each other except Misool which was differentiated from all Raja Ampat lakes, yet not from the populations of the lakes Kakaban, Haji Buang and Tanah Bamban (East Kalimantan). The AMOVA analyses revealed that significant portions of the total variance within lineage B can be attributed to differences among the following three groups 1. Berau coast, Singapore coast, Northern Raja Ampat lakes (Sauwandarek, Cassopeia, Urani), 2. Berau lakes (Kakaban, Tanah Bamban, Haji Buang), Southern Raja Ampat (Misool), 3. Darwin. The among group variation was 84.6% ($p<0.001$) and the within population variation was 10% ($p<0.001$).

Discussion

Two divergent lineages

Two major lineages were uncovered in the populations of the sponge *S. diversicolor*. The congruent patterns of COI and ITS genetic markers and the degree of divergence between the two lineages (COI: 0.4% and ITS: 7.3%) is indicative of reproductive isolation, and thus we suggest that the two lineages (A and B) constitute sibling species. We searched for morphological and ecological characters to distinguish the two lineages, but genetic divergence can preclude morphological distinction. The skeletal structure and spicule lengths do not differ between lineages and fall within the natural variation of this species (Becking & Lim 2009, *CHAPTER 4*). The color and substrate preference are variable, but not consistent within a particular lineage. A related *Suberites* from Satonda lake (Sumbawa, Indonesia) displays different colors at different depths as a result of a symbiosis with the unicellular green algae *Chlorella* and a symbiotic bacteria (Arp et al. 1996). Phylogeographic studies in the Indo-Australian-Archipelago have uncovered numerous lineages in diverse marine taxa that may represent undescribed cryptic species (e.g. Barber et al. 2002, Crandall et al. 2008, Malay & Paulay 2010). Within sponges, molecular studies have revealed a high prevalence of morphologically cryptic sponge species (see review in Xavier et al. 2010a, Reveillaud et al. 2011).

The divergence between lineage A and B, points to a long standing isolation in spite the fact that they are sympatric in Kakaban lake (East Kalimantan). Within sponges there are several reports of sympatric cryptic species: *Tedania* spp. in mangroves (Wulff 2006), *Scopalina lophyropoda* (Blanquer et al. 2008), *Cliona* spp. (Xavier et al. 2010a), *Hexadella* spp. (Reveillaud et al. 2010). Differential reproductive traits and output can promote the co-existence of sibling species (Blanquer et al. 2008, Pérez-Porro et al. 2012). This observation of divergent lineages sympatric within one lake is, however, not common in the phylogeographic studies conducted thus far on populations in the marine lakes in Palau (Dawson 2005, Dawson & Hamner 2005, Gotoh 2009, Goto 2011). The Palauan studies on three distinct taxa (jellyfish, fish and bivalves) mostly show a pattern of one lineage occupying one lake (Dawson 2005, Dawson & Hamner 2005, Gotoh 2009, Goto 2011). One reason why Kakaban lake may contain two lineages is the sheer size of the lake – at almost 4 km^2 it is tenfold larger than any of the other lakes in Indonesia and the majority of Palau (Table 2, Colin 2009, Becking et al. 2011). Alternatively lineage B could be an introduction to the lake. Sponge fragments are known to be transported by waterfowl (Pronzato & Manconi 1994) and workers from the neighboring island Maratua who stay on Kakaban for short periods to attend small crops also may act as possible vectors of *Suberites diversicolor* from the Maratua lakes or the mangroves near their village.

Phylogeography

Lineage A is only present in Kakaban lake, while lineage B is present in all populations. Within lineage B the spatial genetic population structure shows three groups: 1. the three Berau lakes and southern Raja Ampat lake, 2. Berau coast, Singapore coast and the three northern Raja Ampat lakes, 3. Darwin. At present there is no comprehensive phylogeographic study of sponges spanning across the Indonesian archipelago, yet pronounced genetic differences in populations of other marine invertebrates and vertebrates are present between the Java Sea, the Indonesian Through Flow, and the seas of East Sulawesi (e.g. Barber et al. 2000, Barber et al. 2006, Timm & Kochzius 2008, Nuryanto & Kochzius 2009, Carpenter et al. 2011). The marine phylogeographic patterns of these studies strongly support the existence of a barrier in the area between the Sunda and Sahul shelves, where populations from Kalimantan are genetically isolated from those in Papua. Our data does not show a clear East to West geographic break, but the Darwin population, though small in

sample size, is clearly genetically differentiated from the other populations. Dispersal potential and habitat specialization may determine how lineages are distributed and how fauna of different geographic regions are connected (e.g. Carpenter et al. 2011). The majority of sponge population genetic and phylogeographic studies based on less conserved fragments of mitochondrial DNA and nuclear markers reveal structured populations with in some cases evidence of (occasional) long distance dispersal events (Wörheide et al. 2002, 2005, 2008, Lopez-Legentil & Pawlik 2009, DeBiasse et al. 2010, Xavier et al. 2010b). This pattern is congruent with philopatric, shortlived larvae that recruit at short distances from the parental locations (Mariani et al. 2005, Mariani et al. 2006) together with the ability of sponges to disperse as viable fragments in the currents or rafting on various floating material (Wulff 1991, 1995, Maldonado & Uriz 1999). The reproductive cycle and larvae of *S. diversicolor* are unknown, but this species does produce asexual buds (Becking pers. obs.) which may survive a considerable amount of time in the plankton or by rafting before colonizing distant locations as proposed by Wörheide et al. (2008) for *Leucetta chagosensis*.

Within lineage B none of the Indonesian marine lakes contained private genotypes (or unique genetic diversity), and many lakes were identical in composition. The only studies on marine lake phylogeography have been in the islands of Palau on the jellyfish *Mastigias papua* (Dawson & Hamner 2005, Dawson 2005), the fish *Sphaeramia orbicularis* (Gotoh et al. 2009), and the mussel *Brachidontes* sp. (Goto et al. 2011). These studies show extreme genetic isolation, low genetic diversity, and in the cases of *Mastigias papua* and *Brachidontes* sp. rapid morphological evolution in the marine lakes (Dawson & Hamner 2005, Dawson 2005, Gotoh et al. 2009, Goto et al. 2011). The lack of a strong population structure between many of the Indonesian lakes of the present study may be caused by recurrent (recent and historic) geneflow among lakes. We suggest, however, that the lack of structure is a result of small sample sizes of some populations and that the genetic markers we used do not evolve fast enough for mutations to have accumulated to show the differentiation. It is still possible that all these lakes are completely isolated, i.e. do not exchange any migrants. Intraspecific variation in demosponge COI is generally low (e.g. Erpenbeck et al. 2006, Wörheide 2006) and interspecific variation of COI in sponges can be as low as 0-0.4% (*p*-distances) (e.g. Pöppe et al. 2011). However, COI can provide low but sufficient genetic variation in populations of some sponge taxa over relatively short geographic distances (Duran & Rützler 2006, DeBiasse et al. 2010). Of the four molecular markers used in the present study, ITS evolves the fastest and therefore gave the highest resolution of spatial genetic structure. There was no sequence variation in 28S and COII between any of the populations or between the two lineages. The D3-D5 of the 28S fragment has been used to distinguish genera and species of a wide range of demosponge taxa including halichondrids (e.g. McCormack & Kelly 2002, Erpenbeck et al. 2005), but also can be too conserved to discriminate closely related species in others (Reveillaud et al. 2010). COII was proposed as a polymorphic mitochondrial marker for sponge phylogeography by Rua et al. (2011). Rua et al. (2011) indicated that the variation of this marker could be low in halichondrid species *Hymeniacidon heliophila* but attributed their results to the collection of clone-mates. In the present study COII showed no variation between any of the samples spanning a wide geographic range. We conclude that COII is not a suitable marker for intraspecific variation or distinction between closely related species of the genus *Suberites* in particular, and probably more generally for the families Suberitidae and Halichondriidae (Morrow et al. 2012).

Isolation & genetic diversity

Kakaban lake is the largest and most isolated lake in Indonesia, and houses the highest proportion of endemic sponge species (Table 2). Within this lake the population of *Suberites diversicolor* displayed the highest

genetic diversity with unique genetic variants that were not shared with the two lakes at just 6 km distance (Figure 1). These results are concordant with results from Palau where the degree of genetic distance between marine lake and adjacent sea populations was strongly correlated to the degree of connection from the lake to the sea and not the actual geographic distance between the populations (Dawson & Hamner 2005, Goto et al. 2011). Isolation acts to decrease the rate of immigration and thus to decrease the genetic diversity and the number of species expected at equilibrium in an island system (MacArthur & Wilson 1967, Whittaker & Fernandez-Palacios 2007, Chen & He 2009, Rosindell et al. 2011) Yet isolation can also enhance species formation, with the diminished gene flow allowing populations to diverge and ultimately form new species if they remain isolated (Emerson & Gillespie 2008, Chen & He 2009, Rosindell et al. 2011). With the molecular markers we used there was, however, not a direct relationship between moderate levels of isolation and the genetic diversity of the lakes. For example, the populations in northern Raja Ampat lakes (West Papua) are not genetically differentiated, despite the limited physical connections to each other and the adjacent sea.

Biogeographic scenario

Kakaban lake was probably filled with sea water less than 15000 years ago (Dawson 2006, Sathiamurthy & Voris 2006). Considering the deep divergence between lineages A & B, this divergence likely occured well before the formation of Kakaban. Wörheide et al. (2004) estimated an evolutionary rate of 1% per million years for ITS in a suberitid sponge *Prosuberites 'laughlini'* based on the formation of Isthmus of Panama. Implementing the 1 % mutational rate would mean that the two lineages diverged approximately 7 million years ago. Though this is a rough estimation with great error bars and rates of evolution may be higher for recently diverged lineages (Ho et al. 2011), the age is consistent with recent phylogeographic studies that suggest that many endemics from the Indo-Australian-Archipelago have origins in the early Pliocene-Miocene (3-20 million years ago) (e.g. Renema et al. 2008, Bellwood & Meyer 2009, Cowman & Bellwood 2011).
Kakaban lake houses a genetic and species diversity of sponges, that appears to be absent from the surrounding sea (Becking et al. *accepted & unpublished data, CHAPTERS 2, 3, 4, 5 & 6*). Each lake is ephemeral, but the marine lakes ecosystem probably has occurred in various locations during the past glacial-cycles (Sathiamurthy & Voris 2006). The Sunda shelf, which includes Borneo (Kalimantan), was exposed during the Last Glacial Maximum (LGM) when sealevels are estimated to have been approximately 110-140m lower than modern sea levels (Geyh et al. 1979, Voris 2000, Hoeksema 2007). Multiple larger and smaller depressions in the shelf have been recorded which presumably represented palaeo-lakes during the LGM (Sathiamurthy & Voris 2006) that could have become brackish marine lakes with the increase in sea level. What is more, during the LGM the Sunda Land region was dominated by mangroves (Morley 2000), and the water around the Sunda area would have been brackish due to the multiple river outlets (Hoeksema 2007), which are both environments amenable for *S. diversicolor.* Ancient lineages/endemics may have 'hopped' from lake to lake or from mangrove to lake, as the lakes formed and subsequently disappeared with the rise and fall in sealevel during the Plio-Pleistocene glacial cycles. Genetic signatures for glacial refugia are expected to be characterized by high genetic diversity and mixture of ancestral and private haplotypes (Hewitt 2000, Maggs et al. 2008). Kakaban matches this pattern. While Kakaban itself could not have been a refugium during LGM (it was dry), there may have been palaeo-lakes in the vicinity that served as such. Kakaban may be an area where multiple putative refugia populations have come into secondary contact, resulting in the high genetic diversity and the high number of endemics. Molecular studies on co-distributed taxa at larger scales including lakes from adjacent regions in Palau and Vietnam will enhance our understanding of the processes behind the unique marine lake diversity.

Acknowledgements

M. Ammer and B. Hoeksema were invaluable sources of information for the fieldwork. We would also like to thank the following people for their help with logistics of the fieldwork and samples: Suharsono, Y. Tuti, C. Huffard, M. Erdmann, S. Manugbai, Bahruddin, Estradivari, N. Santodomingo, E. Dondorp, W. Renema, B. Alvarez, S.-C. Lim, and the staff of TNC/WWF Berau Office, of Nabucco Island Dive Resort, of Derawan Dive Resort, of Conservation International Wayag station, and of The Nature Conservancy Misool station. We are grateful to the following people who facilitated the laboratory work: G. Wörheide, S. Menken, H. Breeuwer, B. Voetdijk, P. Kuperus. The present study is part of a PhD-project of LEB funded by Netherlands Organisation for Scientific Research (NWO), division Earth and Life Sciences (ALW IPJ-07002; # 817.01.008). Fieldwork in Indonesia was made possible through additional financial support of Van der Hucht de Beukelaar Stichting, World Wildlife Foundation Netherlands-INNO Fund, the Schure-Beijerinck-Popping Fund of the Royal Dutch Academy of Science (KNAW), Conservation International Ecosystem Based Management program (funded by The David and Lucile Packard Foundation), the Treub-Maatschappij Fund, The Netherlands WWF-INNO Subsidy, the Leiden University Fund (LUF)/Slingelands, Singapore Airlines, the A.M. Buitendijk Fund and the J.J. ter Pelkwijk Fund (Naturalis Biodiversity Center). We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian State Ministry of Research and Technology (RISTEK) for providing research permits in Indonesia. E. Gittenberger, W. Renema and C. Vogler provided valuable comments on the original manuscript.

Appendix 1.

Lineage	COI	ITS	Location	Color	Substrate	Fieldcode	RMNH POR
A	C1	*	KKB	Green	root	KKB905	6400
A	C1	H1	KKB	Green	root	SKKB03_120	6485
A	C1	H1	KKB	Green	root	SKKB1067	6401
A	C1	H2	KKB	Green	rock	SKKB1617	6402
A	C1	H2	KKB	Red	rock	SKKB790	6403
A	C1	H2	KKB	Green	rock	SKKB791	6404
A	C1	H2	KKB	Green	mud/sand	SKKB880	6405
A	C1	H3	KKB	Green	root	SKKB05	1397
A	C1	H3	KKB	Green	rock	SKKB1605	6406
A	C1	H3	KKB	Green	rock	SKKB826	6407
A	C1	H3	KKB	Green	rock	SKKB831	6408
A	C1	H4	KKB	Green	rock	SKKB1616	6409
A	C1	H5	KKB	Green	rock	SKKB797	6410
A	C1	H5	KKB	Green	rock	SKKB807	6411
B	C2	H6	HBL	Green	rock	MA1529	6430
B	C2	H6	HBL	Green	rock	SMA02	1438
B	C2	H6	HBL	Green	root	SMA1003	6431
B	C2	H6	HBL	Purple	rock	SMA1014	6432
B	C2	H6	HBL	Green	rock	SMA1018	6433
B	C2	H6	HBL	Green	rock	SMA1022	6434
B	C2	H6	HBL	green	root	SMA1028	6435
B	C2	H6	HBL	Green	rock	SMA1031	6436
B	C2	H6	HBL	Green	root	SMA1037	6437
B	C2	H6	HBL	Yellow	rock	SMA1043	6438
B	C2	H6	HBL	Green	rock	SMA1056	6439
B	C2	H6	HBL	Green	rock	SMA1510	6440
B	C2	H6	HBL	Green	rock	SMA1511	6441
B	C2	H6	HBL	Green	rock	SMA1514	6442
B	C2	H6	HBL	Green	rock	SMA1528	6443
B	C2	H6	HBL	Blue	rock	SMA709	6444
B	C2	H6	HBL	Green	rock	SMA948	6445
B	C2	H6	HBL	Green	rock	SMA951	6446
B	C2	H6	HBL	Green	rock	SMA959	6447
B	C2	H6	HBL	Green	mud/sand	SMA973	6448
B	C2	H6	KKB	Red	root	SKKB1559	6449
B	C2	H6	KKB	green	rock	SKKB1615	6450
B	C2	H6	KKB	Green	root	skkb747	6451
B	C2	H6	KKB	Green	root	skkb781	6452
B	C2	H6	MIS	Green	rock	SMIS1760	6457
B	C2	H6	MIS	Green	rock	SMIS1761	6458
B	C2	H6	MIS	Green	rock	SMIS1762	6459
B	C2	H6	MIS	Green	rock	SMIS1764	6460
B	C2	H6	MIS	Green	rock	SMIS1765	6461
B	C2	H6	MIS	Green	rock	SMIS1766	6462
B	C2	H6	TBB	Green	mud/sand	SMB01	2438
B	C2	H6	TBB	Green	mud/sand	SMB02	4670
B	C2	H6	TBB	Red	mud/sand	SMB03	2433
B	C2	H6	TBB	Red	mud/sand	SMB04	2434
B	C2	H7	BER	Green	mud/sand	SBER1128	6412
B	C2	H7	BER	Green	mud/sand	SBAY1637	6413
B	C2	H7	BER	Green	mud/sand	SBAY1645	6414

Lineage	COI	ITS	Location	Color	Substrate	Fieldcode	RMNH POR
B	C2	H7	CAS	Green	mud/sand	SWAY1720	6415
B	C2	H7	CAS	Green	mud/sand	SWAY1721	6416
B	C2	H7	CAS	Green	mud/sand	SWAY1722	6417
B	C2	H7	CAS	Green	mud/sand	SWAY1723	6418
B	C2	H7	CAS	Green	mud/sand	SWAY1724	6419
B	C2	H7	CAS	Green	mud/sand	SWAY1725	6420
B	C2	H7	CAS	Green	mud/sand	SWAY1726	6421
B	C2	H7	CAS	Green	mud/sand	SWAY1727	6422
B	C2	H7	CAS	Green	mud/sand	SWAY1728	6423
B	C2	H7	CAS	Green	mud/sand	SWAY1746	6424
B	C2	H7	KKB	Green	root	SKKB1066	6453
B	C2	H7	KKB	Red	mud/sand	SKKB1560	6454
B	C2	H7	KKB	Red	mud/sand	SKKB1561	6455
B	C2	H7	KKB	Green	rock	SKKB820	6456
B	C2	H7	MIS	Green	mud/sand	SMIS1763	6463
B	C2	H7	RAJ	Green	root	SRAJ07	6464
B	C2	H7	RAJ	Purple	root	SRAJ08	6465
B	C2	H7	RAJ	Blue	mud/sand	SRAJ10	4680
B	C2	H7	RAJ	Blue	root	SRAJ11	6466
B	C2	H7	RAJ	Blue	root	SRAJ13	6467
B	C2	H7	RAJ	Green	root	SRAJ1667	6468
B	C2	H7	RAJ	Green	mud/sand	SRAJ1668	6469
B	C2	H7	RAJ	Green	mud/sand	SRAJ1670	6470
B	C2	H7	RAJ	Green	mud/sand	SRAJ17	6471
B	C2	H7	RAJ	Green	root	SRAJ19	6472
B	C2	H7	RAJ	Purple	root	SRAJ1923	6473
B	C2	H7	RAJ	Green	root	SRAJ20	6474
B	C2	H7	RAJ	Green	mud/sand	SRAJ21	6475
B	C2	H7	RAJ	Purple	root	SRAJ26	6476
B	C2	H7	RAJ	Blue	root	SRAJ32	6477
B	C2	H7	RAJ	Green	root	SRAJ43	4681
B	C2	H7	RAJ	Purple	mud/sand	SRAJ37	4672
B	C2	H7	RAJ	Blue	root	SRAJ40	6478
B	C2	H7	RAJ	Green	mud/sand	SRAJ47	6479
B	C2	H7	RAJ	Green	root	SRAJ471	4682
B	C2	H7	RAJ	Blue	root	SRAJ35	4673
B	C2	H7	SIN	Green	mud/sand	SIN01	4675
B	C2	H7	SIN	Green	mud/sand	SIN02	6480
B	C2	H7	SIN	Green	mud/sand	SIN03	6481
B	C2	H7	SIN	Green	mud/sand	SIN04	6482
B	C2	H7	URA	Green	root	SW1740	6483
B	C2	H7	URA	Green	rock	SW1742	6425
B	C2	H7	URA	Green	rock	SW1743	6426
B	C2	H7	URA	Green	rock	SW1744	6427
B	C2	H7	URA	Green	root	SW1745	6428
B	C2	H7	URA	Green	root	SW1746	4677
B	C2	H7	URA	Green	root	SW1747	6429
B	C3	H7	URA	Green	rock	SW1741	6484
B	C4	H8	DAR	Green	mud/sand	SDAR09	6486
B	C4	H9	DAR	Green	mud/sand	SDAR01	6487
B	C4	H9	DAR	Green	mud/sand	SDAR02	6488
B	C4	H9	DAR	Green	mud/sand	SDAR03	6489
B	C4	H9	DAR	Green	mud/sand	SDAR07	6490
B	C4	H9	DAR	Green	mud/sand	SDAR05	6491

Appendix 2.

Code	KKB	HBL	TBB	RAJ	CAS	URA	MIS	BER	SIN
HBL	0								
TBB	0	0							
RAJ	0	0	0						
CAS	0	0	0	0					
URA	0,05138	0,22581	-0,08108	0,22581	0,09091				
MIS	0	0	0	0	0	0,02778			
BER	0	0	0	0	0	-0,15385	0		
SIN	0	0	0	0	0	-0,08108	0	0	
DAR	**1***	**1***	**1***	**1***	**1***	**0.85714***	**1***	**1***	**1***