Molecular Ecology (2006) **15**, 4317–4329 doi: 10.1111/j.1365-294X.2006.03073.x

Population structure between environmentally transmitted vibrios and bobtail squids using nested clade analysis

B. W. JONES, J. E. LOPEZ, J. HUTTENBURG and M. K. NISHIGUCHI
Department of Biology, MSC 3AF, New Mexico State University, PO Box 30001, Las Cruces, NM 88003-8001, USA

Abstract

Squids from the genus *Euprymna* (Cephalopoda: Sepiolidae) and their symbiotic bacteria *Vibrio fischeri* form a mutualism in which vibrios inhabit a complex light organ within the squid host. A host-mediated daily expulsion event seeds surrounding seawater with symbiotically capable *V. fischeri* that environmentally colonize newly hatched axenic *Euprymna* juveniles. Competition experiments using native and non-native *Vibrio* have shown that this expulsion/re-colonization phenomenon has led to cospeciation in this system in the Pacific Ocean; however, the genetic architecture of these symbiotic populations has not been determined. Using genetic diversity and nested clade analyses we have examined the variation and history of three allopatric *Euprymna* squid species (*E. scolopes* of Hawaii, *E. hyllebergi* of Thailand, and *E. tasmanica* from Australia) and their respective *Vibrio* symbionts. *Euprymna* populations appear to be very genetically distinct from each other, exhibiting little or no migration over large geographical distances. In contrast, *Vibrio* symbiont populations contain more diverse haplotypes, suggesting both host presence and unidentified factors facilitating long-distance migration structure in Pacific *Vibrio* populations. Findings from this study highlight the importance of how interactions between symbiotic organisms can unexpectedly shape population structure in phylogeographical studies.

Keywords: Cephalopoda, Indo-West Pacific, nested clade analysis, phylogeography, Sepiolidae, *Vibrio fischeri*

Received 7 March 2006; revision received 9 April 2006; accepted 15 April 2006

Introduction

Non-random mating, the precursor to genetic subdivision, can be selected by several biological (mating systems, dispersal methods, etc.) and/or abiotic factors (geological processes, distribution of suitable habitat; (Achmann *et al*. 2004; Cherry 2004; Marko 2004; Miller & Ayre 2004; Ovenden *et al*. 2004; Palo *et al*. 2004; Whiteley *et al*. 2004). Interestingly, while much work has been completed to identify co-evolving host/symbiont assemblages through studies of parallel cladogenesis (Munson *et al*. 1991, 1992; Brooks 1997; Nishiguchi *et al*. 1998; Nishiguchi 2001, 2002), few studies have examined how host population structure may dictate symbiont genetic distribution in these intimate associations. This being so, there is a need to better understand the influences of host population structure and abiotic factors affecting distribution between these populations on symbiont phylogeography and subsequently, speciation.

Correspondence: M. K. Nishiguchi, Fax: (505) 646–5665; E-mail: nish@nmsu.edu

The association between bobtail squids (Cephalopoda: Sepiolidae) and the luminescent bacterium *Vibrio fischeri* has proved to be an innovative and successful model for examining the population biology of cospeciating organisms according to their phylogeographical distribution (Nishiguchi *et al*. 1998; Kimbell *et al*. 2002). In this environmentally transmitted mutualism, axenic juvenile *Euprymna* hatch and soon obtain their *Vibrio* symbionts from surrounding waters (Ruby & McFall-Ngai 1992; McFall-Ngai & Ruby 1998). Upon acquisition of these bacteria, both host and symbiont undergo a number of structural and physiological changes (Montgomery & McFall-Ngai 1993; Foster & McFall-Ngai 1998; Visick *et al*. 2000), the result of which is a highly specific association where *Vibrio*-produced light is emitted from the host light organ in a camouflage behaviour termed counterillumination (Nishiguchi *et al*. 1998; Nishiguchi 2002; Jones & Nishiguchi 2004). In the mature association, the light organ is inhabited by a monoculture of *Vibrio fischeri* in Pacific populations (Ruby & McFall-Ngai 1992).

Despite being an environmentally transmitted association, a high degree of specificity has arisen in Pacific *Euprymna*/

Vibrio assemblages (Nishiguchi *et al.* 1998; Nishiguchi 2002). This is likely due to the biology of this association, where vibrios are cycled through hosts within local populations. This cycling is mediated by a daily host behaviour where approximately 90–95% of the symbiont population is vented into the environment (Ruby & Lee 1998), seeding vibrios into local populations (Lee & Ruby 1994; Jones *et al.*, in review). Expelled vibrios subsequently become an available inoculum for newly hatched juveniles in the same areas, permitting the opportunity for the evolution of specificity in these populations. Although these 'viable' free-living vibrios are capable of colonizing naïve juvenile hosts, they are also subject to abiotic factors such as currents, which may provide an opportunity for free-living vibrios to invade areas where other populations occur, as well as colonize hosts from distantly related host species.

This present study seeks to examine this unique intraspecific association by employing a population genetics approach to examine evolutionary relationships among host squids and their associated *Vibrio* symbionts. Specifically, we quantified population structure and patterns of gene flow among populations in three *Euprymna* species found in the Indo-West Pacific region (Table 1; Fig. 1) by examining the cytochrome *c* oxidase subunit I (COI) locus. These three species were chosen based on distributions of their populations, which represent three distinct geographical areas that may lead to unique patterns of gene flow and population structure. Importantly, no other species of *Euprymna* exist sympatrically with the three species examined here, eliminating the possibility of horizontal transfer of *Vibrio* between host species. *Euprymna scolopes*, which is found along the coastal waters of the Hawaiian archipelago, represents a species where populations are in relatively close proximity to each other, allowing for the possibility of panmictic associations between populations (Fig. 1). Conversely, *Euprymna hyllebergi* (Thailand) and *Euprymna tasmanica* (Australia) represent two species whose populations are more geographically isolated (Fig. 1). *Euprymna hyllebergi* is the more extreme case of the two; populations in this species are separated by a geographical barrier (Malaysian peninsula) making it virtually impossible for migration of individuals between the two populations.

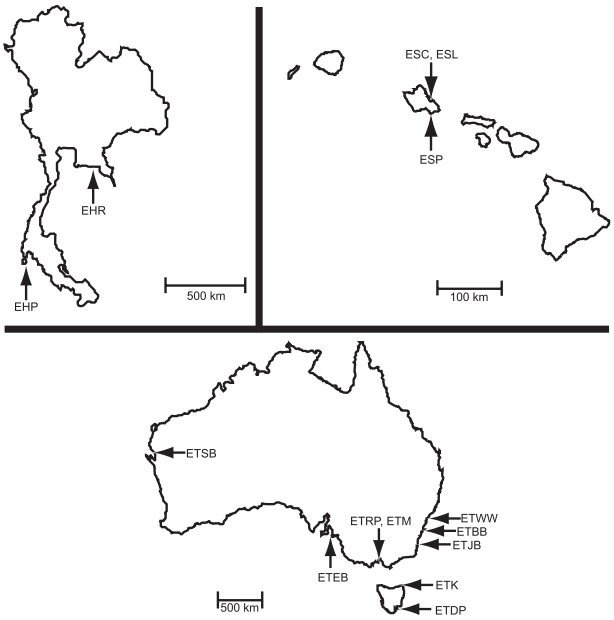

Fig. 1 Graphical representation of Hawaii, Australia, and Thailand sites sampled in this study. Refer to Table 1 for specific site information.

Table 1 Populations and collection site information for all sites in this study. The measurements of within population genetic diversity theta (per-base-pair) are listed for both *Euprymna* host and *Vibrio* symbiont at each location. Refer to Fig. 1 for mapped locations

Host	Name	Population location	Host *N*	Latitude	Longitude	Host theta	*Vibrio* theta
Euprymna	ESP	Hawai'i Kai, Honolulu, Hawai'i, USA	15	N 21°16′	W 157°45′	0.0023	0.0033
scolopes	ESL	Lilli'puna Pier, Kane'ohe Bay, Hawai'i, USA	10	N 21°25′	W 157°47′	0.0016	0.0027
	ESC	Coconut Island, Kane'ohe Bay, Hawai'i, USA	11	N 21°26′	W 157°47′	0.0041	0.0021
Euprymna	ETWW	Woywoy, NSW, Australia	1	S 33°29′6″	E 151°19′47″	0.0000	0.0052
tasmanica	ETBB	Kurnell, Botany Bay, Sydney, NSW, Australia	29	S 34°00′	E 151°12′	0.0236	0.0093
	ETJB	Sanctuary Point, Jervis Bay, NSW, Australia	2	S 35°06′	E 150°39′	0.0015	0.0022
	ETM	St. Leonard's Pier, Melbourne, VIC, Australia	10	S 38°10′14″	E 144°43′8″	0.0005	0.0087
	ETRP	Rye Pier, Melbourne, VIC, Australia	7	S 38°22′10″	E 144°49′20″	0.0006	0.0058
	ETNO	Narooma, NSW, Australia	1	S 33°29′05″	E 150°06′54″	0.0000	NA*
	ETEB	Adelaide, SA, Australia	1	S 34°58′	E 138°16′	0.0000	0.0000
	ETK	Kelso Point, TAS, Australia	13	S 41°3′24″	E 146°47′52″	0.0005	0.0132
	ETDP	Drew Point, TAS, Australia	9	S 43°01′38″	E 147°16′45″	0.0011	0.0079
	ETSB	Shark Bay, WA, Australia	3	S 25°56′	E 113°30′	0.0860	0.0007
Euprymna	EHR	Rayong, Thailand	13	N 12°38′	E 101°14′	0.0010	0.0027
hyllebergi	EHP	Phuket, Thailand	14	N 7°38′	E 98°24′	0.0033	0.0052

*, no *Vibrio* was obtained from ETNO hosts.

To examine *Vibrio* phylogeographical structure, an 891-bp fragment of the glyceraldehyde phosphate dehydrogenase (*gap*A) was examined. This gene was chosen because previous studies have shown this locus to be variable enough to discriminate between *Vibrio* strains (Nishiguchi & Nair 2003). We hypothesized that local cycling of *Vibrio* through *Euprymna* hosts has led to the specificity seen in Pacific populations, and this cycling has created concordant phylogenies that can be detected through the examination of haplotype networks. If, however, *Vibrio* genetic distribution is random, or if genetically distinct *Vibrio* strains not directly involved in local cycling are able to environmentally colonize juvenile hosts, the structure of host and symbiont haplotype networks will be markedly different. This study therefore hopes to expand our understanding of the intraspecies relationships within the sepiolid squid–*Vibrio* mutualism and provide a unique view into the phylogeography of environmentally transmitted symbiosis.

Materials and methods

Specimen collection

Table 1 lists all sites sampled in this study. Thailand populations were collected from coastal waters near Rayong and Phuket (Fig. 1). Rayong is located on the northern shoreline of the Gulf of Thailand, while Phuket is an island located in the Andaman Sea that is connected to Thailand's southwestern coast by means of a land bridge. Both Rayong and Phuket populations are separated from each other by approximately 2100 km following the coastline around the Malaysian peninsula (see Fig. 1). A total of 10 populations were sampled in Australia, encompassing multiple locations along southeastern coast ranging from Woy Woy to Adelaide, Tasmania, and Shark Bay in Western Australia (Fig. 1). In Hawaii, three populations were sampled around the island of O'ahu. Two of the populations (Coconut and Lili-puna) are located in Kanéohe Bay. The Lili-puna population is located on the southern coastline of Kanéohe Bay and the Coconut population is located around Coconut Island, which is about 480 m (offshore) in Kanéohe Bay. The third Hawaiian population is located on the southeastern tip of O'ahu and is approximately 49 km southeast of Kanéohe Bay in Nui Valley (Paiko; Fig. 1).

DNA isolation and sequencing

Individual specimens collected from *Euprymna hyllebergi*, *Euprymna tasmanica*, and *Euprymna scolopes* populations were preserved in 95% ethanol after dissection of the light organs. To isolate *Vibrio* from each host, the light organ was homogenized in sterile seawater, serially diluted, and plated on 15% agar-supplemented seawater tryptone (SWT; 0.5% tryptone, 0.3% yeast extract, 0.3% glycerol, 70% seawater). After 12–18 h of growth, 10–20 *Vibrio fischeri* colonies were isolated from each plate and transferred to 5 mL SWT for 12 h. These clones were then frozen as stocks at −80 °C in SWT in 20% glycerol.

Vibrio fischeri DNA was extracted from each strain using a rapid DNA isolation protocol. One and a half millilitres of 8- to 12-h culture grown in SWT was centrifuged at 14 000 × *g* for 5 min. The resulting pellet was re-suspended in sterile TE, pH 8.0 (10 mM Tris-HCl, 1 mM EDTA) and placed at 100 °C for 3 min, followed by a 2-min centrifugation at 6000 *g*. One microlitre of the DNA-containing supernatant was then used as a template in the polymerase chain reaction (PCR; see below for protocol) to amplify an 891-bp fragment of the glyceraldehyde phosphate dehydrogenase (*gap*A) locus using *V. fischeri* specific primers (Table S1, Supplementary material). These primers were different from previous *gap*A primers in that both primer sets are internal to those used [16 bp (*gap*AF1) and 20 bp (*gap*AR1)] in prior studies (Lee & Ruby 1994; Nishiguchi & Nair 2003). These primer sets were created due to the difficulty in amplification using the original primers for the wide variety of strains tested.

To isolate *Euprymna* DNA, approximately 25 mg of tissue was removed from either the mantle or gill inside the body cavity of each squid. DNA was then isolated from the tissue using the QIAGEN DNeasy Isolation Kit (QIAGEN). Following DNA extraction, 1–10 ng of each DNA sample was used as template for PCR to amplify the 658-bp cytochrome *c* oxidase subunit I (COI; Lopez 2004; Nishiguchi *et al*. 2004).

All loci were amplified using a 50-μL PCR containing 1–10 ng of DNA template, 0.2 μM of each forward and reverse primer, 2.5 mM of $MgCl_2$, 200 μM of each dNTP, 1× buffer (10 mM Tris-HCl, pH 9.0 50 mM KCl, and 0.1% Triton X-100), and 0.2 U of Promega *Taq* polymerase. All PCR amplifications were performed with an MJ Research Dyad thermocycler. PCR products of all loci were purified using the GENECLEAN II DNA purification kit (BIO 101). PCR products were then presequenced using Applied Biosystems BigDye Terminator version 3.1. Excess fluorescent dNTPs were removed from the presequencing reactions using DTR V3 96-well short spin plates (Edge Biosystems). All samples were sequenced using an Applied Biosystems 3100 Automated Capillary Sequencer. Forward and reverse sequences of each individual were combined into contigs using SEQUENCHER 4.1. Sequences were then examined for base miscalls and edited using the GENETIC DATA ENVIRONMENT (GDE) program (Tulio de Oliveria, University of Natal, South Africa).

Analysis of molecular variance

Analysis of molecular variance (AMOVA) was calculated using the population genetics program ARLEQUIN version

2.0 (Schneider *et al*. 2000) to examine population structure with variation partitioned among each country (Australia, US, and Thailand), among populations within each country, and within populations. Within-population genetic diversity was estimated with theta, the per-base-pair nucleotide polymorphism.

Nested clade analysis and secondary contact analysis

Several programs were used to test and analyse genetic variation among individuals in the three species being studied. For both host and symbiont populations, haplotype networks were created using a statistical parsimony procedure (Templeton *et al*. 1992; Crandall *et al*. 1994) performed by the program TCS version 1.12 (Clement *et al*. 2000). Haplotype network ambiguities (closed loops) were resolved using the empirical predictions derived from the frequency, topological, and geographical coalescent theory criterion outlined by Pfenninger & Posada (2002). The haplotype network was then nested and ambiguities were resolved using the nesting procedures given by Templeton (Templeton *et al*. 1987; Templeton & Sing 1993; Templeton *et al*. 1995). The nested clade information was input into GEODIS version 2.0 (Posada *et al*. 2000) which performed a contingency analysis of the nested clades in the haplotype network, and provided an analysis of the categorical variation for the association between clades and their geographical locations.

To determine if secondary contact had occurred for *Vibrio* lineages that displayed fragmentation, the procedures of Templeton (2001) were followed. Briefly, the average pairwise distance from the geographical centres of all haplotypes and clades was calculated for each population and nesting level. Theoretically, panmictic populations would display near-identical or decreasing average distances for increasing clade levels, while haplotypes that were previously fragmented and occurring in the parent population would have high or increasing distances with increasing clade level (Templeton 2001). An unpublished program by David Posada (GEOLOC version 3.1) was used to implement the above algorithm of Templeton (2001). This program calculates the statistical significance of the calculated clade distance values through 10 000 random permutations of each clade against geographical location (Pfenninger & Posada 2002).

Tests for recombination

Given the commonality of bacterial recombination in nature (Feil *et al*. 2000; Feil *et al*. 2001; Spratt *et al*. 2001; Mu *et al*. 2005), the programs PAIRWISE and PERMUTE in the LDHAT software package were used to investigate linkage disequilibrium in *Vibrio* populations (McVean *et al*. 2002). These programs estimated the population-scaled recombination rate ($2N_e r$) using a modified version of the composite likelihood method of Hudson (2001). In addition, these programs were used to implement a nonparametric permutation test (using 1000 permutations) for recombination by computing the linkage disequilibrium between pairs of sites through calculating the correlation of both summary statistics r^2 and $|D'|$ with physical distance. These observed statistics are compared to null distributions containing randomized single nucleotide polymorphisms to detect nonzero recombination (McVean *et al*. 2002). Finally, the minimum number of recombination events (R_m) was calculated in PAIRWISE using the algorithm of Hudson & Kaplan (1985).

Phylogenetic analysis

Phylogenetic analysis of distinct *Vibrio* haplotypes was conducted using maximum likelihood as an optimality criterion, as implemented in PAUP* (Swofford 2002). The best-fit model of sequence evolution was selected under the Akaike Information Criterion (AIC) as implemented in MODELTEST version 3.06 (Posada & Crandall 1998; Posada & Crandall 2001), and that model was used to infer the phylogenetic tree via a heuristic search of 100 replicates of random addition sequence followed by tree-bisection–reconnection (TBR) branch swapping. Nodal support was assessed with nonparametric bootstrapping obtained from 1000 heuristic replicates under TBR branch swapping.

Results

Genetic variation and nested clade analysis

For each host, a 658-bp region of COI was amplified, while an 891-bp region of the *gap*A gene was amplified for 200 *Vibrio* symbionts isolated directly from 139 field-caught sepiolid squid hosts. A total of 46 different symbiont haplotypes (GenBank Accession nos DQ646741–DQ646786), and 38 different host haplotypes were observed (GenBank Accession nos DQ646703–DQ646740). Non-synonymous mutations were identified in 18 of the 68 *Vibrio* polymorphic sites (Table S2, Supplementary material), and seven *Euprymna* polymorphic sites (Table S3, Supplementary material). There were no nonsynonymous mutations in Hawaiian *Euprymna scolopes* sequences. For *Vibrio* populations, the Thailand, Hawaiian, and Australian populations contained 7, 6, and 32 distinct haplotypes, respectively. One distinct haplotype was found within both Hawaiian and Australian populations (Figs 2 and 3). *Euprymna* populations consisted of 10 Hawaiian, 18 Australian, and 10 Thailand haplotypes (Fig. 4).

Vibrio sequence data yielded theta ranging from 0 to 0.0132 between all sites (Table 1). *Euprymna* theta ranged from 0.00 to 0.0860 between all sites. For hosts, the small number of samples from the sites containing no genetic

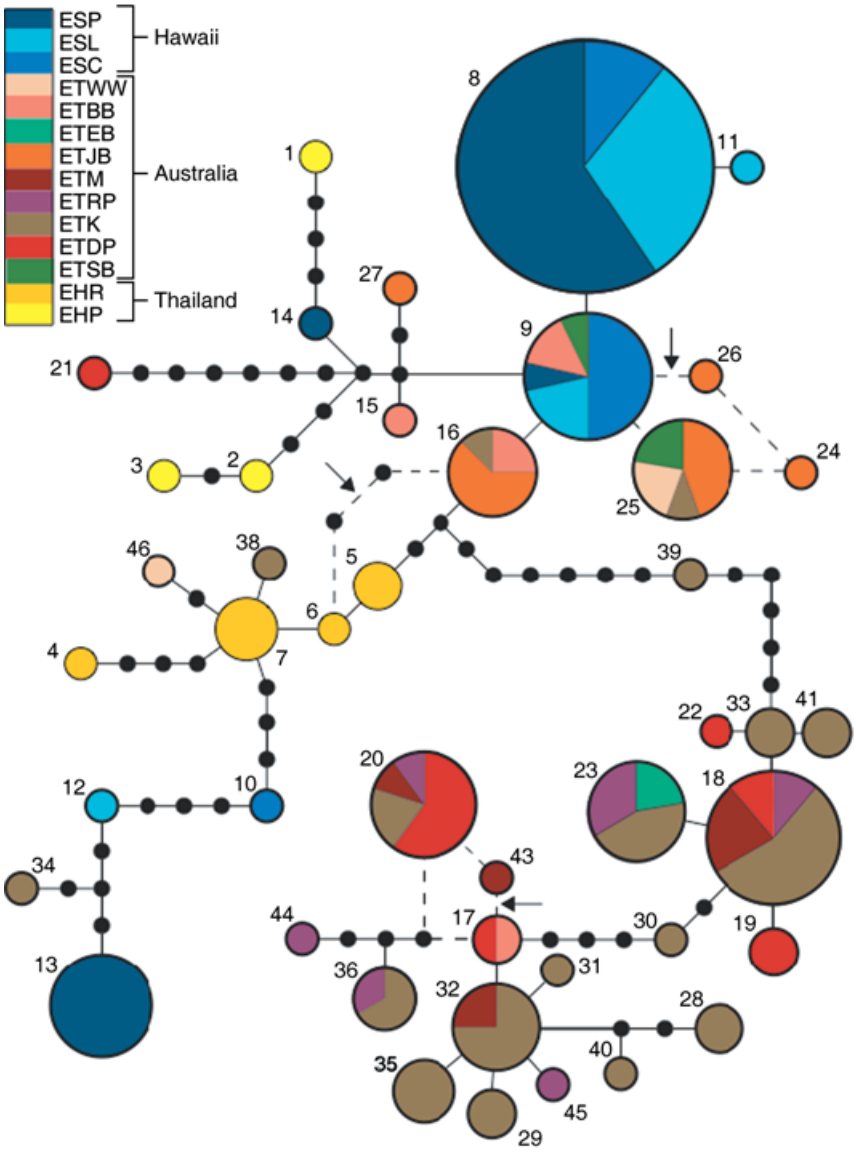

Fig. 2 Haplotype network for Indo-west-Pacific symbiotic vibrios. Black circles represent unsampled haplotypes, while colours in each circle represent the proportion of locations having that specific haplotype. Each line represents one mutational step. Loop structures are represented with a dashed line, and the point at which they were resolved is marked with an arrow. The size of each circle is proportional to the number of samples with that haplotype. Haplotype numbers are listed beside each haplotype.

variation (ETEB, ETWW, and ETNO) resulted in values of 0.00 for theta. Interestingly, within-population genetic divergence had a much greater range in the Australian populations than Hawaiian and Thailand populations. The extraordinary high values of theta for ETSB hosts are from two very distinct haplotypes found at this location (see below).

Vibrio Pacific-wide haplotype distribution contains a southern network of haplotypes exclusive to Australia (the lower right-hand portion of Fig. 2). However, a number of haplotypes more closely related to haplotypes from Hawaii and Thailand were also found in Australia. These included almost all of the Northern sites of Botany Bay, Jervis Bay, and Woy Woy (Fig. 2). For *Vibrio* haplotype networks, nesting clade 1-12 with 2-7 provided more parsimonious inference for population-level processes. This is because clade 1-12 borders a major split between clades at the total cladogram level. Preliminary analysis demonstrated that if clade 1-12 was nested within 2-6 instead of 2-7, a secondary colonization event of Thailand would have been inferred, which is clearly not the case since all EHR strains are closely related (Figs 2 and 3).

The original *Vibrio* haplotype network contained three closed-loop structures, which were resolved according to Templeton & Sing (1993). The points of resolution have been marked with an arrow on Fig. 2. The final nested network is displayed in Fig. 3. The fourth level nesting displays differentiation of southern Australian haplotypes (clade 4-3), which contain no Hawaiian or Thailand haplotypes. The null hypothesis of no geographical structure within a single clade was rejected for seven clades and at the total cladogram level (Table 2). Clade 1-10 was an internal clade exhibiting geographical structure. Since the

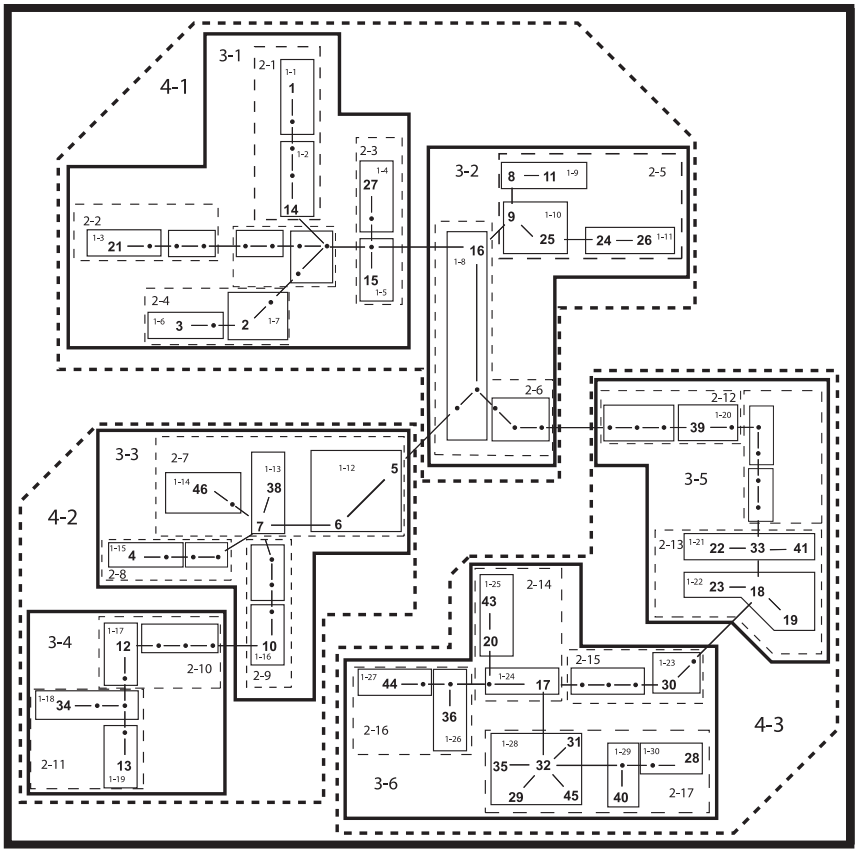

Fig. 3 Nesting diagram for the haplotype network in Fig. 2. The network has been simplified, using only haplotype numbers. Ambiguities were resolved using the rules of Templeton & Sing (1993). Clades are designated with two digit identifiers within each boxed clade. As in Fig. 2, one line represents one mutational step, but the length of the line has been varied for illustrative purposes.

interior-tip distinction of nested groups could not be determined for this clade, no inference could be made about its population structure and history. Clades 1-22 and 3-2 exhibited contiguous range expansion, while 3-6, 4-1, and the total cladogram illustrate restricted gene flow with isolation by distance. Clade 4-2 demonstrates past fragmentation followed by range expansion.

Secondary analysis shows significantly large population distances for populations at Woy Woy and Botany Bay from the 2-step to 4-step clades. Significantly, large distances were observed for Jervis Bay at the 2-step clade level. The Paiko population in Hawaii also had a significantly large population distance at the 3-step clade level (Fig. 5).

AMOVA for *Vibrio* populations failed to reveal significant genetic structure among continents, but did so among populations within continents and within populations. Sixty-four percent of the genetic variation was within populations, while 33% of the variation was explained between populations (Table 3).

Euprymna genetic data resulted in five different haplotype networks and three haplotypes that did not fit into any network (Fig. 4). Two Thailand networks contained haplotypes exclusive to Rayong and Phuket. Since there was no geographical variation in these networks, they were not included in NCA. The three networks that did contain phylogenetically informative information were exclusive to Hawaii (clades identified with Hx-x), eastern Australia (EAx-x), and southern Australia (Sax-x). The southern Australia network contained one closed-loop ambiguity, which used the criterion described above. In addition, three Botany Bay haplotypes (23, 24, and 25) grouped with the southern Australia network.

In examining the *Euprymna* haplotype network, the null hypothesis of no genetic structure was rejected for the Hawaiian clades H1-1 and H2-1, which both illustrated isolation by distance. In addition, the EA2-1 exhibited restricted gene flow with isolation by distance as well. SA1-1 showed contiguous range expansion, but there was insufficient genetic resolution to differentiate between range expansion/colonization and restricted dispersal/gene flow in clade SA2-1 (Table 4). AMOVA for *Euprymna* populations showed 67.22% genetic variation between continents and 25.44% genetic variation between populations. Within populations, 7.34% of total variation was observed (Table 3).

Tests for recombination

For *Vibrio* populations, the estimated population mutation rate (θ) was 0.012, and the population rate of recombination ($2N_e r$) was found to be 5.0. The estimated number of recombination events (assuming absence of homoplasies

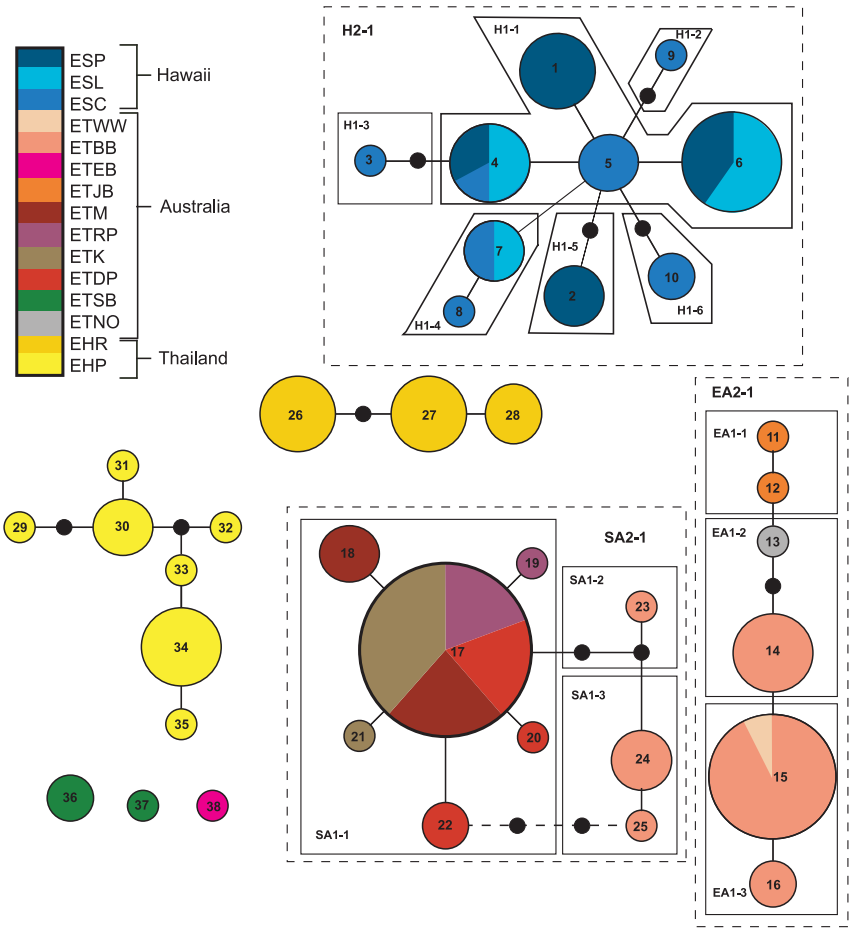

Fig. 4 Haplotype network for *Euprymna* hosts. Hawaii (H), East Australia (EA), and South Australia (SA) networks have been nested as in Fig. 3.

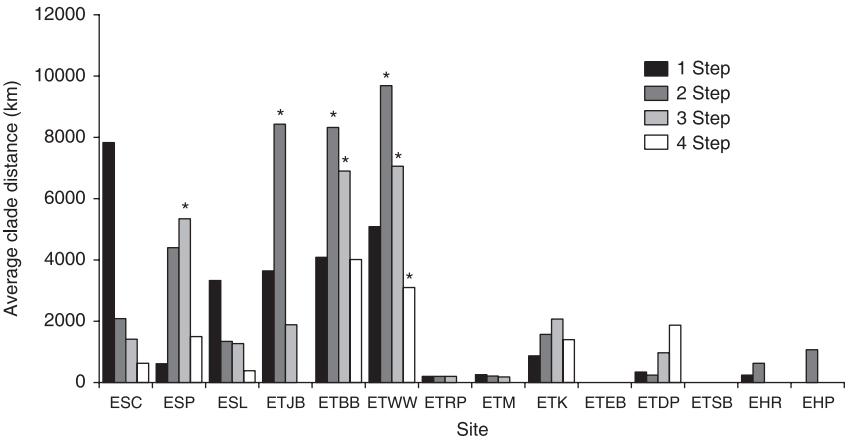

Fig. 5 Secondary contact analysis for the nested clade analysis data set. Bars represent the average clade distance for each clade level. Each asterisk represents significantly large clade distances ($P < 0.01$).

in the data), R_m, was 13. Both the r^2 and $|D'|$ measures of linkage disequilibrium provided evidence of recombination ($r^2 = 0.12$, $P = 0.004$; $|D'| = -0.135$, $P = 0.001$).

Phylogenetic analysis

The best-fit model under AIC was the Hasegawa–Kishino–Yano (HKY) model with a correction for invariant sites and a discrete gamma distribution (HKY + I + G) for *Vibrio* sequence data. The HKY model assumes two substitution types, one for transversions and one for transitions, and unequal base frequencies. The substitution model assumes unequal base frequencies, which are estimated as follows: $\pi_A = 0.3024$, $\pi_C = 0.\ 1880$, $\pi_G = 0.\ 1906$, $\pi_T = 0.3190$. The proportion of invariant sites was set to 0.8220. The α parameter of the discrete gamma distribution (four rate categories) was estimated to 0.5992. 80% of the 100 replicates yield a single unrooted tree of –ln L = 1628.87297 (Fig. 6).

Table 2 *Vibrio* significant clades and inference for population genetic data. Listed are clades (or haplotypes) exhibiting significant distributions, whether the significant values are within the clade (D_c) or nested clade (D_n), and the *P* value associated with the observed data

Clade	Nested clade	Dist.	Value (S or L)	*P*	Inference key steps	Inference
1-10	9	D_c	15083 L	0.0059	1, 2	Inconclusive
		D_n	9982 L	0.0151		
	25	D_c	1176 S	0.0008		
		D_n	6140 S	0.0252		
1-22	18 (I)	D_c	206 S	0.0167	1, 2, 11, 12, No	Contiguous range expansion
		D_n	228 S	0.0008		
	19 (T)	D_c	0.0 S	0.0000		
		D_n	536 L	0.0679		
	I-T	D_c	−296 S	0.0021		
2-5	1-9 (T)	D_c	8.2 S	0.0000	1, 2, 3, 5, 6, 7, 8, No	Sampling inadequate to discriminate between isolation by distance and long-distance dispersal
		D_n	16098 L	0.0000		
	1-11 (T)	D_c	0.0 S	0.0337		
	I-T	D_c	7787 L	0.0000		
		D_n	−1403 S	0.0063		
3-2	I-T	D_c	−14999 S	0.0161	1, 2, 11, 12, No	Contiguous range expansion
3-6	2-14 (I)	D_n	376 L	0.0458	1, 2, 3, 4, No	Restricted gene flow with isolation by distance
	2-17 (T)	D_c	179 S	0.0136		
		D_n	184 S	0.0001		
	I-T	D_c	183 L	0.0065		
		D_n	173 L	0.0016		
4-1	3-1 (T)	D_n	9471 S	0.0002	1, 2, 3, 4, No	Restricted gene flow with isolation by distance
	3-2 (I)	D_c	15201 L	0.0001		
		D_n	14698 L	0.0001		
	I-T	D_n	4956 L	0.0001		
4-2	3-3 (I)	D_c	3018 S	0.0079	1, 2, 11, 12, 13, Yes	Past fragmentation followed by range Expansion
	3-4 (T)	D_c	3474 S	0.0104		
		D_n	14149 L	0.0000		
	I-T	D_n	−8118 S	0.0000		
Total	4-1 (I)	D_c	13043 L	0.0000	1, 2, 3, 4, No	Restricted gene flow with isolation by distance
		D_n	9044 L	0.0073		
	4-3 (T)	D_c	352 S	0.0000		
		D_n	7281 S	0.0058		
	I-T	D_n	11046 L	0.0000		
		D_c	1809 L	0.0018		

Source of *Vibrio* variation	d.f.	Variance components	% Variation
Among continents	2	0.01^{NS}	1.80
Among pops. within continents	11	0.17*	33.33
Within populations	186	0.32*	64.86
Overall F_{ST}		0.35*	
Source of *Euprymna* Variation	d.f.	Variance components	% Variation
Among continents	2	30.44*	67.22
Among pops. within continents	12	11.52*	25.44
Within populations	124	3.32*	7.34
Overall F_{ST}		0.93*	

Table 3 AMOVA results for *Vibrio fischeri gap*A and *Euprymna* COI genetic subdivision between populations from Hawaii, Thailand, and Australia

*, $P < 0.001$; NS, not significant.

Table 4 *Euprymna* significant clades and inference for population genetic data

Clade	Nested clade	Dist.	Value (S or L)	*P*	Inference key steps	Inference
H1-1	1 (T)	D_c	0 S	0.0240	1, 2, 3, 4, No	Restricted gene flow with IBD
H2-1	1-2 (T)	D_c	0 S	0.0340	1, 2, 3, 4, No	Restricted gene flow with IBD
		D_n	11.95 L	0.0270		
	I-T	D_n	7.993 L	0.0020		
EA2-1	I-T	D_n	77.84 L	0.0170	1, 2, 11, 17, 4, No	Restricted gene flow with IBD
SA1-1	21 (T)	D_c	338 L	0.0474	1, 2, 11, 12, No	Contiguous range expansion
	I-T	D_n	222 L	0.0245		
SA2-1	1-1 (T)	D_c	231 S	0.0000	1, 2, 3, 5, 6, Too	Insufficient genetic resolution to
		D_n	275 S	0.0000	few clades	discriminate between range
	1-3 (T)	D_c	0 S	0.0011		expansion/colonization and
		D_n	665 L	0.0000		restricted dispersal/gene flow
	I-T	D_c	–205 L	0.0000		
		D_n	345 L	0.0000		

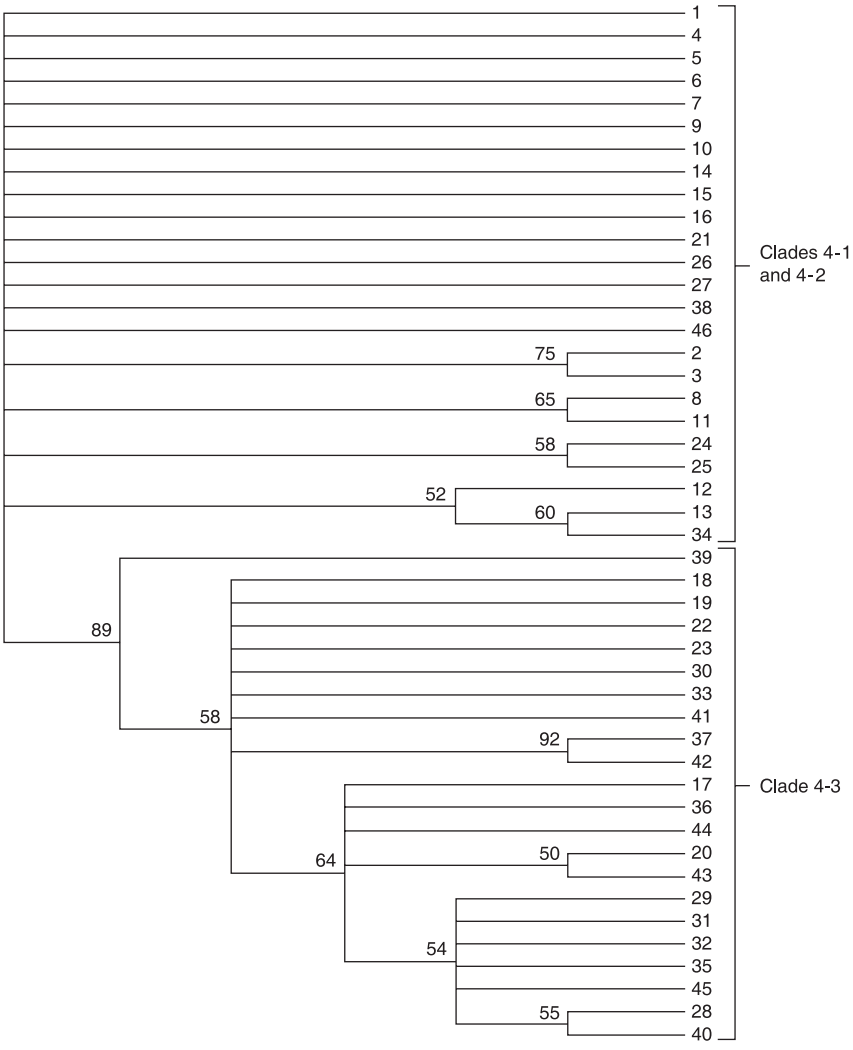

Fig. 6 Phylogenetic analysis of *Vibrio fischeri* haplotypes. Refer to Fig. 2 for haplotype designations. There was not sufficient resolution to distinguish between closely related haplotypes, but the tree supports the separation between clades 4-3 and 4-1/4-–2.

Discussion

Sampling

Sampling strategies for *Euprymna scolopes* in O'ahu were more exhaustive than *Euprymna hyllebergi* and *Euprymna tasmanica* populations. This was due to the heavy sampling of *E. scolopes* populations over the past 20 years. Sampling efforts have searched for other populations between Kanéohe Bay and Nui Valley (Paiko), but no specimens have been found due to inappropriate habitat with rough surf in these areas.

Results of the analysis for Australian *E. tasmanica* and Thailand *E. hyllebergi* yielded two distinct haplotype networks between populations in each country that cannot be connected due to the large number of mutational steps between populations. Further sampling between Rayong and Phuket should allow connection between the two Thailand networks. Likewise, future sampling between Shark Bay, Adelaide, and Melbourne may also connect the Australian networks, providing a more clear view of gene flow within and between these distant populations.

Euprymna population structure

In Hawaii, host networks differ dramatically from the *Vibrio* network. The host network shows restricted gene flow with isolation by distance in both clades H1-1 and H2-1 (Fig. 4). Previous work examining *E. scolopes* geographical structure supports these results as Paiko (ESP) and Lilli-puna, O'ahu (ESL, ESC) were also shown to be genetically distinct (Kimbell *et al*. 2002) based on allozyme data. Since little is known regarding active dispersal of benthic *Euprymna* species, the probability of introgression between relatively close populations may be less frequent compared to other invertebrate species, and may be affected more by abiotic factors such as currents and tidal changes.

In contrast to the relatively simple Hawaiian network, NCA and AMOVA revealed some interesting trends in *E. tasmanica* host populations. First, Melbourne and Tasmania populations share nearly all haplotypes (clade SA1-1). This suggests that a large amount of gene flow is occurring between these two populations, and NCA corroborates this by inferring contiguous range expansion in clade SA1-1. Being that *E. tasmanica* is also benthic and is capable of only swimming short distances, the probability that individuals migrate between the populations (290 km across the Bass Straight) is more likely a phenomenon of currents than active dispersal. Oceanographic data suggests that currents in the Bass Straight are turbulent, characterized by many underwater cascades and eddy or vortex patterns due to the convergence of several oceanic currents at this geographical region (Fandry *et al*. 1997). In addition, the Bass Straight is part of the Australian continental shelf with shallow waters of 50–70 m in depth. This combination of turbulent current and shallow depth may account for mixing between Melbourne and Tasmania populations via juvenile migrants carried in these cascade and eddy/vortex currents across the Bass Straight.

The second interesting aspect of the Australian haplotype networks is that individuals from Botany Bay (ETBB) were represented in both the SA and EA haplotype networks (haplotypes 14, 15, 16, 23, 24, 25). As demonstrated from phylogenetic inference, there was insufficient genetic resolution to determine what processes have led to this variation (Table 4). While the individuals found in Botany Bay with South Australian-like haplotypes could represent a small number of migrants, it is clear that some sort of barrier exists between the South Australian and East Australian networks.

Symbiotic Vibrio *population structure*

A small amount of recombination within the *Vibrio* data set was detected using three different methods. Recombination could potentially disrupt inference of NCA due to construction of an incorrect network. In TCS, homoplasies created by recombination are displayed as closed loops, and existing procedures were used to resolve these ambiguities (Templeton & Sing 1993). Even so, inference based on any network should be treated with caution due to the fact that no estimated network is completely correct. Strengthening the data set will include future analysis examining other informative loci in addition to *gap*A.

Potential caveats aside, the *Vibrio* network illustrates Hawaii *Vibrio fischeri* grouping together with East Australian congeners (Fig. 2). In addition, clade 4-2 in the *Vibrio* network contains a number of Hawaiian strains that are more closely related to the Rayong strains than the other Hawaiian strains found in clade 2-5 (Fig. 2). The inference provided here is past fragmentation, followed by range expansion. Result of this historical process therefore may be a secondary colonization event in Hawaii at the Paiko population. This can be visualized on the cladogram (Fig. 2) by the large sample from Paiko (ESP, dark blue) that is more closely related to Rayong isolates (EHR, orange-yellow). Indeed, secondary contact analysis supports this position as 3-step clade population distances are significantly high when compared to randomly distributed populations (Fig. 5).

Given that there is no connection between the Hawaii and Thailand networks based on COI data (Fig. 4) and that *Euprymna* hosts are different species in Thailand, Hawaii, and Australia, it is important to consider how *Vibrio* symbionts may have migrated between host populations. *Vibrio* may conceivably travel by oceanographic currents through the South China Sea, ride the Kuroshio current North past Japan, where it would follow the North Equatorial current North of Hawaii, and then down to the

Hawaiian archipelago. A more plausible hypothesis suggests that *Vibrio* strains were moved by ballast water from large tankers. One problem with this hypothesis, however, is that secondary contact occurs deep at higher nesting levels in the cladogram, suggesting that the event happened long ago. Additionally, there may be populations of *Euprymna* that exist along many of the South Pacific Islands, allowing *Vibrio* bacteria to leap frog from one population to the next. Examination of both additional *Euprymna* populations within these areas, as well as *V. fischeri* haplotypes from both hosts and surrounding seawater may support either of these hypotheses, and would also explain why Australian strains appear scattered throughout the cladogram.

Unlike the COI host networks, the complex *gap*A network allows inference at higher clade levels. At the total cladogram level of the *Vibrio* network, analysis suggests restricted gene flow with isolation by distance between clades 4-1 and 4-3 (Fig. 3, Table 3). This inference is mainly due to the presence of only Australian haplotypes in clade 4-3 and the mixture of Australian and Hawaiian haplotypes in clade 4-1. Interestingly, northeast Australian strains (with the exception of one ETBB strain) group with the Hawaiian or Thailand strains (Fig. 2). Significantly large population clade distances for Jervis Bay, Woy Woy, and Botany Bay are the results of isolates from these populations grouping with Hawaiian strains. Results of the phylogenetic analysis on individual haplotypes supported the observed high-level division between clades 4-3 and 4-2/4-2 (Fig. 6), despite little or no resolution between other haplotypes. Since *Vibrio* moving with a host is not a possibility at these high clade levels, the observed structure may be due to current activity and/or ballast water.

Population structure and cospeciation

Studies in other systems suggest there is a genetic barrier between Sydney and Melbourne for other types of invertebrates. Work with the asteroid sea-star *Coscinasterias muricata* identified a split between northern and southern haplotypes in Mallacoota, Victoria, which is approximately 425 km east of Melbourne and 425 km south of Sydney on the Southeast Coast of Australia (Waters & Roy 2003). This site corresponds to the eastern-most edge of the Maugean marine biogeographical province proposed by Bennett & Pope (1953). The Maugaen biogrographical zone extends from Mallacoota to Adelaide, and includes all of Tasmania, which corresponds to the observed data for *V. fischeri*, as these populations are almost exclusively contained in clade 4-3 (Fig. 3). Ecological factors that have structured the establishment of this biogeographical zone are still under contention, but the zone appears to be the result of temperature gradients and ocean currents (O'Hara & Poore 2000). Given the data at hand, the structure of *V. fischeri* populations in clade 4-3 and the EA and SA host networks raises a number of interesting questions regarding the biogeography of this mutualism. For example, if temperature is keeping these strains mostly restricted to the Maugaen biogeographical zone, is adaptation to these colder waters reflected in *V. fischeri* host-associated growth rates of these strains? If so, how might this cold-water adaptation affect squid-host colonization in warmer waters where *V. fischeri* must compete with other strains?

At first glance, our results seem to contradict previous studies demonstrating competitive dominance of native over non-native *V. fischeri* strains in *E. scolopes*. In Nishiguchi *et al*. (1998), however, *V. fischeri* strains used in colonization studies were from Melbourne, Australia and Lillipuna, O'ahu, Hawaii. Although the present study did not examine the original strains used in the competition studies, the haplotype network presented here show no shared haplotypes between the ETM and ESL populations (Fig. 2). Nonetheless, results of colonization studies using strains collected in this study (particularly between those from Australia and Hawaii that share identical haplotypes) will provide unique insight to the nature of competitive dominance and local host adaptation by *V. fischeri*. Future studies that include not only allopatric but sympatric populations of sepiolid squids and their *Vibrio* symbionts will shed light on how quickly adaptation occurs between co-occurring partners, and which factors predominate in selection and the overall evolution of this dynamic mutualism.

Acknowledgements

Funding was provided by NSF SBE-0123690, NSF DBI-0079820, NSF DEB-0316516, and NIH S06-GM08136-26 to M. K. N. and NIH-RISE GM 61222–01 in support of J. E. L. We would like to thank Sandra Duran, David Posada, and Tom Parchman for all their assistance with the population genetics programs. Finally, we would like to thank the Ruby and McFall-Ngai laboratories, Natalie Moltschaniwskyj, Geoffrey Woolcott, Jarawat Nabhitabhata, and Anuwat Nateewathana for all their help in collecting the squid specimens.

Supplementary material

The supplementary material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3073/MEC3073sm.htm

Table S1 Bacterial *gap*A, eukaryotic COI primer sequences

Table S2 DNA sequence variation in *Vibrio* haplotypes. Refer to Figure 2 for haplotype number designation. Bases in bold represent nonsynonymous mutations

Table S3 DNA sequence variation in *Euprymna* haplotypes. Due to the large amount of variation in COI, haplotypes are divided into Hawaiian, East Australian, South Australian, Phuket, and Rayong populations to examine variation. Listed is

each polymorphic site with associated base pair location for each haplotype. Haplotypes from Adelaide and Shark Bay, Australia are not included due to their high level of variation from all other sequences. Nonsynonymous sites are in bold. Refer to Figure 4 for haplotype number designation

References

Achmann R, Curik I, Dovc P *et al.* (2004) Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Animal Genetics*, **35**, 285–292.

Bennett I, Pope EC (1953) Intertidal zonation of the exposed rocky shores of Tasmania and its relationship with the rest of Australia. *Australian Journal of Marine and Freshwater Research*, **11**, 182–221.

Brooks D (1997) Testing the context and extent of host-parasite co-evolution. *Systematic Zoology*, **28**, 229–249.

Cherry JL (2004) Selection, subdivision and extinction and recolonization. *Genetics*, **166**, 1105–1114.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Crandall KA, Templeton AR, Sing CF (1994) *Intraspecific Cladogram Estimation: Problems and Solutions*. Clarendon Press, Oxford.

Fandry C, Cresswell G, Baines PG, Edwards RJ (1997) *The Facts: Ocean Currents Around Tasmania. Three Oceans Influence the Coastal Environment of the Island State, its Climate and Wealth and Distribution of its Natural Life*. CSIRO, Australia.

Feil EJ, Smith JM, Enright MC, Spratt BG (2000) Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics*, **154**, 1439–1450.

Feil EJ, Holmes EC, Bessen DE *et al.* (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proceedings of the National Academy of Sciences, USA*, **98**, 182–187.

Foster JS, McFall-Ngai MJ (1998) Induction of apoptosis by cooperative bacteria in the morphogenesis of host epithelial tissues. *Development Genes and Evolution*, **208**, 295–303.

Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics*, **159**, 1805–1817.

Hudson RR, Kaplan N (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Jones BW, Nishiguchi MK (2004) Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* (Mollusca: Cephalopoda). *Marine Biology*, **144**, 1151–1155.

Jones BW, Maruyama A, Ouverney CC, Nishiguchi MK (in review) Development and application of phylogenetic probes for the *in situ* identification of symbiotic *Vibrio fischeri* and *Vibrio logei* (Proteobacteria: Vibrionaceae), in review.

Kimbell JR, McFall-Ngai MJ, Roderick GK (2002) Two genetically distinct populations of bobtail squid, *Eyprymna scolpes*, exist on the island of O'ahu. *Pacific Science*, **56**, 347–355.

Lee K-H, Ruby EG (1994) Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Applied and Environmental Microbiology*, **60**, 1565–1571.

Lopez JE (2004) *Symbiosis among sepiolid squid: origins of the symbiotic light organ and examination of host population structure at the intraspecies level*, MS Thesis, New Mexico State University, New Mexico.

Marko PB (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology*, **13**, 597–611.

McFall-Ngai MJ, Ruby EG (1998) Sepiolids and vibrios: when first they meet. *Bioscience*, **48**, 257–265.

McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231–1241.

Miller KJ, Ayre DJ (2004) The role of sexual and asexual reproduction in structuring high latitude populations of the reef coral *Pocillopora damicornis*. *Heredity*, **92**, 557–568.

Montgomery MK, McFall-Ngai M (1993) Embryonic development of the light organ of the sepiolid squid *Euprymna scolopes* Berry. *Biology Bulletin*, **184**, 296–308.

Mu J, Awadalla P, Duan J *et al.* (2005) Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biology*, **3**, e335.

Munson MA, Baumann P, Clark MA *et al.* (1991) Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *Journal of Bacteriology*, **173**, 6321–6324.

Munson MA, Baumann P, Moran NA (1992) Phylogenetic relationships of the endosymbionts of mealybugs (Homoptera: Pseudococcidae) based on 16S rDNA sequences. *Molecular Phylogenetics and Evolution*, **1**, 26–30.

Nishiguchi MK (2001) Co-evolution of symbionts and hosts: the sepiolid–*Vibrio* model. In: *Symbiosis: Mechanisms and Model Systems* (ed. Seckbach J), pp. 757–774. Cole-Kluwer Academic Publishers, Dordrecht, The Netherlands.

Nishiguchi MK (2002) Host-symbiont recognition in the environmentally transmitted sepiolid squid–*Vibrio* mutualism. *Microbial Ecology*, **44**, 10–18.

Nishiguchi MK, Nair VS (2003) Evolution of symbiosis in the Vibrionaceae: a combined approach using molecules and physiology. *International Journal of Systemic Bacteriology*, **53**, 2019–2026.

Nishiguchi MK, Ruby EG, McFall-Ngai MJ (1998) Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid–*Vibrio* symbioses. *Applied and Environmental Microbiology*, **64**, 3209–3213.

Nishiguchi MK, Lopez JE, Boletzky Sv (2004) Enlightenment of old ideas from new investigations: more questions regarding the evolution of the bacteriogenic light organs in squids. *Evolution & Development*, **6**, 41–49.

O'Hara TD, Poore GCB (2000) Patterns of distribution for southern Australian marine echinoderms and decapods. *Journal of Biogeography*, **27**, 1321–1335.

Ovenden JR, Salini J, O'Connor S, Street R (2004) Pronounced genetic population structure in a potentially vagile fish species (*Pristipomoides multidens*, Teleostei; Perciformes; Lutjanidae) from the East Indies triangle. *Molecular Ecology*, **13**, 1991–1999.

Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.

Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migrations, and secondary contact. *Evolution*, **56**, 1776–1788.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations.

Proceedings of the National Academy of Sciences USA, **98**, 13751–13756.

Ruby EG, Lee K-H (1998) The *Vibrio fischeri-Euprymna scolopes* light organ association: current ecological paradigms. *Applied and Environmental Microbiology*, **64**, 805–812.

Ruby EG, McFall-Ngai MJ (1992) A squid that glows in the night: development of an animal-bacterial mutualism. *Journal of Bacteriology*, **174**, 4865–4870.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN, version 2.000: A Sofware for Population Genetics Data Analysis*. University of Geneva, Switzerland.

Spratt BG, Hanage WP, Feil EJ (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Current Opinion in Microbiology*, **4**, 602–606.

Swofford DL (2002) *PAUP*v4.08b Phylogenetic analysis using parsimony (*and other methods)*. Sinauer Associates: Sunderland MA.

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

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.

Templeton AR, Beorwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Abystoma tigrunum*. *Genetics*, **140**, 767–782.

Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG (2000) *Vibrio fischeri lux* genes play an important role in colonization and development of the host light organ. *Journal of Bacteriology*, **182**, 4578–4586.

Waters JM, Roy MS (2003) Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography*, **30**, 1787–1796.

Whiteley AR, Spruell P, Allendorf FW (2004) Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology*, **13**, 3675–3688.

B.W. Jones is currently a postdoctoral research associate at University of Calgary Medical School, and is interested in the evolution of pathogenesis. J.E. Lopez is a medical student at the University of New Mexico, and has interests in eukaryotic-bacterial interactions. J. Huttenberg is a PhD student at the University of Wisconsin, Madison, and has interests in microbial symbiosis. M.K. Nishiguchi is an associate professor of Biology at New Mexico State University, whose research encompasses the evolution of symbiotic associations.
