CHAPTER 5

Introduction to comparative phylogeography of several *Naso* species (section B)

5.1 Introduction

The species-level phylogeny of *Naso* identified several sister species pairs. In this section (section B), I carry out comparative phylogeographic studies for a widely distributed species, *N. vlamingii*, which occurs throughout the Indo-Pacific (see chapter 6) and two pairs of sister species, *N. lituratus – N. elegans* and *N. tuberosus – N. tonganus*, which are distributed in distinct ocean basins of the Indo-Pacific (chapter 7). Such a comparative approach enables me to determine what underlying factors may be important in structuring the genotype of the study species and will allow me to infer likely modes of speciation.

This chapter will:

1. Present an overview of marine phylogeographic studies. As Losos and Glor (2003) point out it is not possible to infer confidently modes of speciation from phylogeographic data alone. Therefore, I describe briefly the five phylogeographic categories suggested by Avise (2000) (Figure 5.1). I also introduce genetic diversity indices (haplotype and nucleotide) and explain how these indices can be used to infer evolutionary histories of populations as proposed by Grant and Bowen (1998).
2. Briefly introduce current theories on modes of speciation with a focus on marine species.

5.2 Marine phylogeographic studies

Phylogeographic studies of marine species have been carried out using a variety of different techniques, for example allozymes (e.g.Doherty et al. 1994; Planes et al. 1996; Planes and Doherty 1997; Williams and Benzie 1998; Benzie 1999; Benzie 2000), RFLPs (e.g. Bermingham and Lessios 1993; Grant et al. 1998; Graves 1998; Williams and Benzie 1998; Benzie 2000; Benzie et al. 2002), microsatellites (e.g. Markert et al. 1999; Benzie 2000; van Herwerden et al. 2000) and mtDNA sequences from the control region (Bernardi 2000; Terry et al. 2000; Bernardi et al. 2001; Lessios et al. 2001; Stepien et al. 2001; Barber et al. 2002; Bernardi et al. 2002; Ovenden et al. 2002; Planes and Fauvelot 2002; Bernardi et al. 2003; Fauvelot et al. 2003; Keeney et al. 2003; Lessios et al. 2003; Bay et al. online) or cytochrome *b*, cytochrome oxidase I (COI) and nuclear genes (Duda and Palumbi 1999; Nelson et al. 2000; Bowen et al. 2001; Colborn et al. 2001; Muss et al. 2001; Planes et al. 2001; Chenoweth et al. 2002; Rocha et al. 2002).

Most of these phylogeographic studies have involved species (invertebrate and vertebrate) with a bipartite life cycle which involves a pelagic larval and a sedentary adult phase. The pelagic larval phase is the main mechanisms by which dispersal of marine species occurs (Palumbi 1994; Palumbi 1997; Knowlton 2000; Lessios et al. 2001; Lessios et al. 2003). This dispersive stage affects the level of genetic connectivity among various populations for marine species, which is enhanced by the lack of known extrinsic barriers. Several reviews and commentaries, for example by Knowlton (2000), Hewitt (2001; 2003) and Palumbi (1994; 1997; 2003), emphasise the importance of a dispersive larval stage and its implication to gene flow among populations of marine species.

The term population refers to geographically defined groups of individuals (sometimes called subpopulations), for example a western Australian population (WA), a Great Barrier Reef (GBR) population, a Papua New Guinea (PNG) population, etc.

In general, invertebrate phylogeographic studies (e.g. starfish, sea urchins, prawns) have displayed distinct population structures across different ocean basins (e.g. between Indian – Pacific Ocean Williams and Benzie 1998; Benzie 1999; Duda and Palumbi 1999; Benzie et al. 2002) despite the presence of dispersive larvae, but weaker or no population structures at smaller scales (e.g. within the Great Barrier Reef Williams and Benzie 1998; Benzie 1999). At the other extreme no population structure across different ocean basins has been reported for species of sea urchins (*Diadema savignyi*, Lessios et al. 2001; *Tripneustes* species, Lessios et al. 2003), where genetic differentiation was minimal suggesting gene flow across both ocean basins (from the West Indian Ocean to the East Pacific Ocean). Marine fish species with dispersive larvae may therefore display similar population structures as marine invertebrates (*ie*. distinct population structure across ocean basins, but higher levels of gene flow at smaller scales). Indeed, some phylogeographic studies of fish have found genetic differentiation, hence low levels of gene flow across ocean basins (between West Indian Ocean and East Indian Ocean and/or West Pacific Ocean) (e.g. Bernardi et al. 2001; Bay et al. online), but lack of genetic differentiation, hence high levels of gene flow at smaller scales (e.g. within West Pacific Ocean, Caribbean) (e.g. Bernardi et al. 2000; Bernardi et al. 2001; Bowen et al. 2001; Planes and Fauvelot 2002) (see Table 5.1). Some fish studies have found little or no genetic structure across different ocean basins suggesting high levels of gene flow. One example is a species of trumpetfish,

Table 5.1: Population studies that used direct sequencing of d-loop and cyt *b* for reef fishes. Species, no. of samples (*n*), no. of haplotypes (n*h*), geographic regions sampled, haplotype diversity (*h*) and percent nucleotide diversity (% π, F_{st} or Φ_{st} (a measure of gene flow) and sources are indicated.

Control region (d-loop)

Species	*n*	n*h*	Region	*h*	% π	F_{st} or Φ_{st}	Source
Dascyllus trimaculatus	56	51	within Moorea	0.98	2.5	0.048	(Bernardi et al. 2001)
(damselfishes)	62	56	French Polynesia	0.98	2.5	0.01	
Overall	98	87	Indo-Pacific	0.99	5.1	0.72	
Dascyllus trimaculatus	101	97	Indo-Pacific	0.99	6.5	no gene flow	(Bernardi et al. 2002)
Dascyllus albisella	10	9		0.96	2.4	between clades	see also
Dascyllus auripinnis	6	6		1.00	1.8	(different species)	(Bernardi et al. 2003)
Dascyllus strasburgi	5	5		1.00	3.6		
All species	122	115		0.99	6.9	0.67[a]	
Embiotoca jacksoni	152	31	Mainland California	0.81	1.1	0.02-0.94*	(Bernardi 2000)
(black surfperch)	88	25	Channel Islands	0.85	0.7	0.02-0.93*	
Overall	240	54	All regions	0.88	1.1	0.21-0.70	
Pristipomoides multidens			Exmouth (WA)				(Ovenden et al. 2002)
(goldband snapper)	111	61	to Arafura Sea (NT)	0.73-0.89	1.9-3.0	0.03-0.06	
Paralabrax maculatofasciatus			California to		0.0-0.3[b]		(Stepien et al. 2001)
(spotted sand bass)	63	7	Baja California	0.0-0.36	0.1-1.8[c]	0.32-1.00	
Girella nigricans	43	38	Sea of Cortez	0.96	2.7	0.02-0.13	(Terry et al. 2000)
(opaleye)	64	48	Pacific Coast	0.97	5.6	0.01-0.25	
Overall	107	87	All regions	0.99	6.3	--	

* Table 2, p.231 in Bernardi (2000)

[a] among all species

[b] within locations

[c] amongst locations

Table 5.1 cont./…

Control region (d-loop)

Species	*n*	n*h*	Region	*h*	% π	F_{st} or Φ_{st}	Source
Chlorurus sordidus	31	29	Amirante, Sey.	1.0-0.975	2.3-3.2		(Bay et al. online)
(parrotfish)	44	36	WA	0.98-0.99	2.6-3.2		
	42	36	GBR	0.98-0.99	2.5-3.0		
	43	40	PNG, Rota Isl.	0.99-1.0	2.6-3.6		
	25	17	Hawaii	0.94	2.8		
			WPO vs Hawaii			0.20	
	185	158	Among Oceans	0.99	4.5	0.61	
Chaetodon quadrimaculatus	46	42	Outer slope	0.996	2.2	0.0159	(Fauvelot et al. 2003)
Chromis xanthura	49	47	Outer slope	0.998	3.5	0.001	
Forcipiger flavissimus	34	33	Outer slope	0.998	1.9	0.0364	
Chaetodon citrinellus	39	25	Lagoon	0.949	1.4	0.0156	
Chrysiptera glauca	30	14	Lagoon	0.754	0.4	0.0009	
Dascyllus aruanus	45	38	Lagoon	0.986	2.0	0.0195	
Pomacentrus pavo	49	14	Lagoon	0.572	0.3	-0.0092	
Carcharhinus limbatus	34	2	Sth Carolina, Atlantic	0.371	0.035	0.087-0.129	(Keeney et al. 2003)
(blacktip shark)	45	10	PI,Fl. Gulf of Mexico	0.785	0.12	Atlantic vs. Gulf[1]	PI, Pine Island, FL
	45	8	TC,Fl. Gulf of Mexico	0.72	0.106	0.002-0.005	TC, Terra Ceia, FL
	45	9	YT,Fl. Gulf of Mexico	0.796	0.134	within Gulf[1]	YT, Yankeetown, FL

[1] Gulf: Gulf of Mexico

Table 5.1: cont./…

Cytochrome *b*

Species	*n*	n*h*	Region	*h*	% π	F_{st} or Φ_{st}	Source
Acanthurus bahianus	112	41	Tropical	0.69-0.90	0.35-0.64	0.01-0.84	(Rocha et al. 2002)
Acanthurus chirurgus	48	30	Atlantic	0.0-0.99	0.0-0.53	0.002-0.05	
Acanthurus coeruleus	82	20		0.73-0.94	0.19-0.42	0.004-0.5	
(surgeonfishes)							
Acanthochromis polyacanthus	54	32	Coral Sea	0.97	2.4	0.865	(Planes et al. 2001)
	72	41	GBR	0.86-0.97	0.6-0.7	--	
(damselfish)	126	73	All regions			0.926	
Amphiprion ocellaris	166	27	Indo-	0.98	0.5	0.12	(Nelson et al. 2000)
(clownfish)			Malayan reg.				
Ophioblennius atlanticus	121		Atlantic	0.41-1.0	0.11-1.3	0.053-0.3	(Muss et al. 2001)
Ophioblennius steindachneri	50		East. Pacific	0.84-0.97	0.8-1.3	0.0-0.4	
(blennies) Overall	171	122				--	
Albula glossodonta	33	8	Indo-Pacific	0.39	0.1	0.05	(Colborn et al. 2001)
Albula neoguinaica	15	12	Indo-Pacific	0.94	1.8	--	
Albula vulpes	47	9	Atlantic	0.54	0.1	0.007	
Albula species A-E	79	59	worldwide	0.86-0.97	0.4-0.8	0.14-0.68	
(bonefishes) Overall	174	88					
Aulostomus chinensis	68	8	Indo-Pacific	0.0-0.76	0.0-0.05	0.093	(Bowen et al. 2001)
Aulostomus strigosus	106	3	E. Atlantic	0.0-0.46	0.0-0.2	0.585	
Aulostomus maculatus	22	2	W. Atlantic	0.0-0.13	0.0-0.03	--	
(trumpetfish) Overall	196	13					
Urocampus carinirostris	90	49	GBR	--	0.32-4.64	--	(Chenoweth et al. 2002)
(pipefish)							

Aulostomus chinensis, which displayed weak population structure between the Indian- and Pacific Oceans (Bowen et al. 2001). At the other extreme two studies involving unusual fish species that lack a pelagic larval stage (*Embiotoca jacksoni*, Bernardi 2000; and *Acanthochromis polyacanthus*, Planes et al. 2001), demonstrated genetic structure (therefore low levels of gene flow) within regions (Bernardi 2000; Planes et al. 2001). Another study found genetic structure within the Caribbean region for a reef shark species, *Carcharhinus limbatus*, which gives birth to live young (Keeney et al. 2003) (see Table 5.1). The majority of fish studies (see fish references above and Table 5.1) however, found moderate to high levels of gene flow among populations (at small to moderate geographic scales), indicated by low levels of genetic differentiation, F_{st} or $\Phi_{st} \leq 0.05$ (Hartl and Clark 1997).

Fixation indices, such as Φ-statistics, a derivation of *F*-statistics, measure genetic differentiation using an hierarchical approach to examine molecular variance among populations of a species. The difference between Φ- and *F*-statistics is that the former incorporates information about DNA haplotype diversity (a measure of differences among haplotypes) (Excoffier et al. 1992; Weir 1996), whereas the latter is based on gene diversity (a measure of heterozygosity or variance in allele frequency) (Nei and Kumar 2000). The fixation indices are calculated at different levels of hierarchical subdivision (Φ_{ct}: a measure of genetic differentiation among regions, relative to total of a species, Φ_{sc}: a measure of genetic differentiation among populations within regions and Φ_{st}: a measure of genetic differentiation within populations, relative to the total of a species) (see Alvarado Bremer et al. 1995) using an Analysis of Molecular Variance (AMOVA) (Schneider et al. 2000). In the extreme case, a F_{st} or Φ_{st} value of 1 ("theoretical maximum" Hartl and Clark

1997) indicates extensive genetic differentiation (hence no gene flow), whereas values close to 0 (or negative values) ("theoretical minimum" Hartl and Clark 1997) indicate no genetic differentiation (therefore very high levels of gene flow) among populations (Hartl 2000). Hartl and Clark (1997; Hartl 2000) proposed a range of *F*-statistic values describing gene flow: an intermediate range of 0.05 to 0.15 suggests that gene flow among populations may be low to moderate; a range of 0.15 – 0.25 indicates intermediate genetic differentiation and values above 0.25 indicate high levels of genetic differentiation (i.e. very low/no gene flow) (Hartl 2000).

With increasing numbers of published phylogeographic studies, several phylogeographic patterns have emerged relating to genetic differentiation among populations of a species (intraspecific difference), categorised by Avise (2000). In general, five categories recognized by Avise (2000) can be partitioned into two groups. The first group incorporates category I and II and explains intraspecific patterns of deep lineages (strong genetic structure) for either allopatric or sympatric distributed species respectively (Figure 5.1). The second group, categories III, IV and V distinguishes between shallow lineages (with weaker genetic structure) between intraspecific lineages that are allopatric, sympatric or varied in their distributions (which includes common and widespread lineages/clades and closely related clades that form geographically distinct pockets) (Figure 5.1) (see pages 136-147 in Avise 2000). Therefore, deep gene trees are characterised by genetically divergent populations of a species (Figure 5.1) with little or no gene flow between populations from different regions (category I) or with extensive gene flow among populations from different regions (category II). These distinct lineages/clades may have arisen either by accumulating new mutations in isolation (genetic drift), and/or permitting

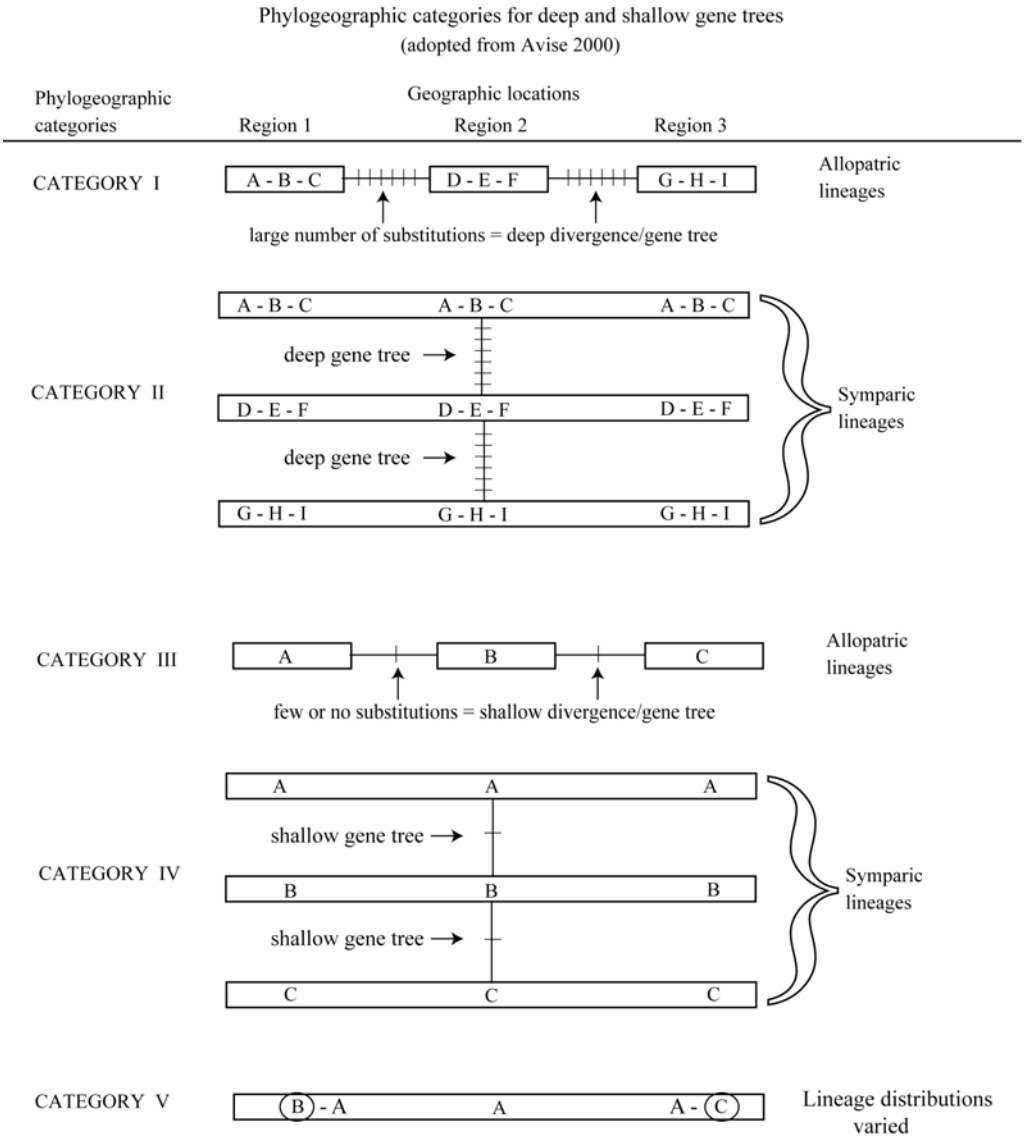

Figure 5.1: Illustration showing the five categories for deep and shallow lineages of population structure as defined by Avise (2000). Capital letters represent particular haplotypes, geographic regions are indicated as regions 1,2 & 3. The haplotypes are connected with lines forming a network and mutational changes (substitutions) between haplotypes are indicated by slashes across connecting lines. Large numbers of substitutions indicate deep divergences, few substitutions indicate shallow divergences.

lineage sorting of mtDNA over long evolutionary time (from a polymorphic ancestor) (Avise 2000). Shallow gene trees characterise populations with little divergence that have either no gene flow between populations from different regions (category III) or are highly connected (high levels of gene flow among populations from different regions, category IV), with little effect of- or rapid lineage sorting over a shorter evolutionary time (Figure 5.1) (Avise 2000). Therefore, if category I populations are sampled from different regions (e.g. 1, 2 3 in Figure 5.1) then there will be region-specific clades that are quite divergent. Category II populations on the other hand, when sampled, will have divergent clades, but these will not be confined to particular regions. Rather, all divergent lineages will be represented in all regions (broadly sympatric lineages, Avis 2000). The same goes for categories III and IV respectively, but for shallow divergences (Figure 5.1).

In addition to measuring genetic differentiation as a tool to infer levels of gene flow among populations, diversity indices (haplotype and nucleotide) can be calculated to determine the genetic architecture of populations and suggest possible historical events (e.g. isolation, secondary contact between diverged lineages, bottlenecks) that may have influenced the observed genetic structure.

Haplotype diversity (h) is calculated using the equation: $h = n\,(1 - \sum x_i^2)/(n\text{-}1)$, where n is the number of samples and x_i is the frequency of the ith haplotype (Nei 1987). The value of h ranges from 0.0 to 1.0, where a value of 0.0 indicates that all haplotypes are identical (no diversity) and a value of 1.0 (very high diversity) indicates that every individual has a unique haplotype (Grant and Bowen 1998). The nucleotide diversity (π), also termed the mean sequence divergence (Bowen and Avis 1990; Finnerty and Block 1992) is the average

sequence divergence among haplotypes (Grant and Bowen 1998) and indicates the relative evolutionary 'age' (deep or shallow) of the species. Therefore, small nucleotide diversities would suggest recently diverged (or shallow) lineages (e.g. due to rapid lineage sorting between small founder populations, or recent bottlenecks) and large nucleotide diversities indicate deep divergences (i.e. lineages accumulated over long evolutionary time, permitting lineage sorting and/or accumulation of mutations in isolation over longer periods of time). Values range from 0 to > 0.1 (over 10%), where values close to 0.0 indicate no or very little sequence divergence between haplotypes and values >> 0.01 (over 1%) suggest deeper divergences between haplotypes and/or secondary contact between allopatrically differentiated lineages (Grant and Bowen 1998).

Grant and Bowen (1998) reviewed a number of studies of marine fishes with broad distribution ranges, displaying high connectivity and shallow population histories in deeply divergent lineages (low gene structure). The authors introduced a framework of categories to describe these shallow population histories of marine fishes according to their haplotype and nucleotide diversity indices (Table 5.2). Grant & Bowen (1998 p.422) suggested the cut-off value for category 1 diversity indices (small values for h and π) to be $h < 0.5$ and $\pi < 0.5\%$. This category includes marine fish from temperate to tropical regions (e.g. cod, grouper, snapper, damselfish), which have either undergone a recent population bottleneck or founder event of a few lineages. Diversity indices for category 2 (large value for h and small value for π) range from $h > 0.5$ and $0.5\% < \pi \leq 1\%$; a large number of temperate to tropical marine and pelagic fish species fit this category (e.g. marlin, herring, sardine, goby, squirrelfish). These species may have undergone a population bottleneck followed by rapid population growth.

Table 5.2 adopted from Grant and Bowen based on data from either mt RFLPs or cyt *b* sequences (1998, p:423).

	Haplotype diversity index: *h*	
Nucleotide diversity index: π	**Small *h*** (*h*<0.5)	**Large *h*** (*h*>0.5)
Small π (π<0.5%)	**1**. Recent population bottleneck **or** founder event by single or few mtDNA lineages	**2**. Population bottleneck followed by rapid population growth and accumulation of mutations
Large (π>1%)	**3**. Few highly divergent haplotypes between geographically subdivided populations (secondary contact or strong bottleneck in previously large, stable populations)	**4**. Large stable population with long evolutionary history **or** secondary contact between differentiated lineages (allopatrically)

Category 3 diversity indices (small value *h* large value π, $h < 0.5$ and $\pi > 1\%$) were not represented in any of the fish population studies reviewed. Species belonging to this category are identified by having few highly divergent haplotypes between geographically distinct populations. According to Grant & Bowen (1998), the lack of representative fish in this category is due to the fact that this category may fit marine fish species with low haplotype diversity (small effective population size) a condition that may occur in very few inshore and some freshwater fish. Category 4 (large values for *h* and π) diversity indices $h > 0.5$ and $\pi >> 1\%$ are largely characteristic of pelagic marine fishes (Bluefish, Atlantic menhaden, anchovies) and a tropical reef fish (Caribbean blenny). These species have either had large stable populations with long evolutionary histories or have allopatrically differentiated lineages that have come into secondary contact.

5.3 Modes of speciation

Recent reviews of the mechanisms promoting speciation (allopatric and sympatric) demonstrate that examining the intra-specific relationships of closely related species pairs helps to understand speciation events (e.g. Turelli et al. 2001; Losos and Glor 2003). Allopatric speciation occurs when species have diverged in isolation (Turelli et al. 2001); sympatric speciation occurs when species have diverged whilst co-occurring in a particular geographic area (Losos and Glor 2003). Some reviews emphasize a number of factors may influence speciation, e.g. evolutionary history of the species in combination with past and present patterns of the establishment of reproductive isolation and ecological differentiation (Palumbi 1992; Barraclough and Vogler 2000; Schluter 2001; Schluter et al. 2001; Via 2001; Losos and Glor 2003).

The marine environment however, presents evolutionary scientists with a challenge (Palumbi 1997). About 70% of marine organisms reproduce by external fertilization with high fecundities and dispersive larvae (Knowlton 2000; Feral 2002). Despite this propensity for dispersal, tropical reef systems support a high diversity of fish species, many of which are closely related co-occurring species with broad distribution ranges. During the pelagic larval phase, reef fish have the ability to disperse over large distances (Palumbi 1992; Palumbi 1994) whereas adults are generally more sedentary. These life history features and the lack of known extrinsic barriers in the marine environment are expected to promote high levels of gene flow amongst populations.

Despite this dispersive capacity, mounting evidence suggests that cryptic speciation (species similar or identical morphologically but not interbreeding) in the marine environment is more common than previously thought, especially when examining broadly

distributed taxa (Knowlton 1993; Knowlton and Weigt 1998; Knowlton 2000; Norris 2000; Lessios et al. 2001; Bernardi et al. 2003). An ongoing study (3 year period) by Bernardi et al. (2001; 2002; 2003) recently recognized that the broadly distributed coral reef fish species *Dascyllus trimaculatus* (damselfish) actually consists of 5 distinct species clades, 3 of which are cryptic. The study by Bernardi et al. (2001; 2002; 2003) clearly demonstrated the usefulness of sampling a number of individuals throughout the vast distribution range of a species, as well as including the examination of sister species using phylogeographic methods.

The paradox to be explained therefore, is how such a large number of closely related, ecologically similar coral reef fish species have arisen and what processes are driving this divergence given the absence of "hard barriers" and capacity for long distance dispersal by pelagic larvae? The main mechanism of speciation for reef fish is considered to be allopatric, where populations diverge in isolation by drift and under different selective pressures, due to the absence of gene flow for extended periods of time at times of low sea level.

In the following two chapters (6 & 7) diversity indices and inferred levels of gene flow are examined in several species of *Naso* with different traits. These diversity indices are related to Avise's (2000) five and Grant and Bowen's (Grant and Bowen 1998) four categories. The study species are then compared to other species for which the same gene region (d-loop) has been characterised in phylogeographic analyses. The presence/absence of cryptic speciation is explored in a widely distributed species, *N. vlamingii* (chapter 6). Finally, with the aid of comparative studies, the mode of speciation for member of closely related species pairs will be investigated in chapter 7.