INFERRING THE GENETIC STRUCTURE OF MARINE POPULATIONS: A CASE STUDY COMPARING ALLOZYME AND DNA SEQUENCE DATA

RONALD S. BURTON
Marine Biology Research Division
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California 92093-0202

ABSTRACT

The genetic structure of natural populations of marine organisms is frequently inferred from the distribution of alleles at gene loci. Until recently, most investigations relied entirely on protein electrophoretic techniques, with particular emphasis on polymorphic enzyme-coding gene loci (allozyme loci). Over the past few years, increasing use has been made of molecular techniques, including methods that allow the construction of gene genealogies. These later methods provide powerful insight into the evolutionary history of genetic variation and, under some conditions, provide valuable information concerning population structure. This paper compares results of both allozyme and DNA sequence studies for a set of populations of the intertidal copepod *Tigriopus californicus* along the California coast. The comparisons show that: (1) Allozyme frequencies distinguish almost all the study populations, whereas genealogies of the sampled DNA sequences do not. (2) Both nuclear and mitochondrial DNA sequences reveal strong population differentiation between central and southern California populations that is not apparent in the allozyme frequencies. (3) Allozymes and DNA sequences are not entirely concordant in the picture they present of population relationships. (4) The most complete analyses of population structure will require multiple genetic techniques.

RESUMEN

La estructura genética de poblaciones naturales de organismos se infiere a menudo a partir de la distribución de alelos en loci de genes. Hasta hace algunos años la mayoría de los estudios se basaban en técnicas de electroforésis de proteínas, con énfasis en loci de genes polimorfos que codifican enzimas (loci alocimos). En años recientes se ha incrementado el uso de técnicas moleculares, incluyendo el uso de métodos que permiten construir genealogías genéticas. Estos métodos ofrecen una perspectiva profunda de la historia evolutiva de la variación genética; en ciertas circunstancias, estos métodos proveen información importante de la estructura de la población. Este artículo compara los resultados obtenidos de alocimos y secuencias de ADN de un conjunto de poblaciones del copépodo intermareal *Tigriopus californicus* a lo largo de la costa de California. Las comparaciones produjeron varios resultados: (1) Las frecuencias de alocimos distinguieron a casi todas las poblaciones estudiadas, mientras que las genealogías de las secuencias de ADN no las demarcó. (2) Tanto las secuencias del ADN de las mitocondrias como las de los núcleos revelaron una marcada diferenciación poblacional entre California Sur y Central, mas este resultado no se repitió con las frecuencias de alocimos. (3) Los resultados de las secuencias de ADN y los alocimos no concuerdan en su totalidad en el esquema de las relaciones poblacionales. (4) Los análisis mas completos de la estructura de la población requerirán múltiples técnicas genéticas.

INTRODUCTION

Attempts to understand the genetic structure of marine invertebrate populations have long been hampered by our inability to directly track the dispersal of larval life stages that frequently spend substantial lengths of time in the plankton. Although current patterns and other physical factors may restrict or promote particular routes of dispersal, the oceans often appear to be barrier-free, and long-distance dispersal is clearly possible (e.g., Scheltema 1986). Despite this extensive potential for gene flow, population differentiation among marine invertebrates has now been widely documented through the use of biochemical and molecular genetic techniques (Burton 1983; Hedgecock 1986; Palumbi 1992). In particular, analyses of electrophoretically detected enzyme polymorphisms have made major contributions to our understanding of gene flow and recruitment in a diversity of marine species over the past two decades.

Several long-recognized problems arise from the use of allelic isozymes (allozymes) for the analysis of population structure. First, although protein electrophoresis allows the screening of many individuals, populations, and gene loci for analysis of population structure, its resolution is limited in that it detects only a subset of existing variants at a gene locus. Thus any putative allele may consist of multiple alleles that are indistinguishable, and differences between populations may frequently be underestimated. Second, as genetically different forms of the same enzyme, allozymes may differ in functional properties. Therefore, frequencies of alleles not only reflect patterns of gene flow and random genetic drift, but they may also result from the action of natural selection.

Finally, the evolutionary relationships among alleles are not necessarily reflected by similarity in their electrophoretic mobility; thus it is not possible to determine which alleles are derived by mutation from other alleles in the same population and which, if any, were introduced by immigrants from distant populations.

Recently, a variety of molecular techniques that assess variation in DNA sequence have supplemented or replaced allozyme techniques in the analysis of population structure (e.g., Avise 1992; Karl and Avise 1992). Of particular interest in the analysis of intraspecific population differentiation is the analysis of mitochondrial DNA (mtDNA). Unlike nuclear DNA, mtDNA is typically maternally inherited and not subject to recombination. Divergence between maternal lineages continues to increase through evolutionary time, and descendants can be assigned to lineages in nonreticulate gene genealogies (Avise 1991). In the case of nuclear genes, genealogies may be complicated by recombination between lineages, obscuring the true degree of evolutionary divergence among the sampled genes.

In addition to focusing on either nuclear or mitochondrial genomes, researchers can take two qualitatively different approaches to the use of molecular techniques for population analyses. One approach is to use restriction fragment length polymorphisms (RFLPs) or sequencing in order to find single polymorphic nucleotide sites, which are then studied in much the same way polymorphic enzymes are studied (i.e., frequencies of different alleles are determined in each study population). Although providing a wealth of markers for study, the data themselves are not qualitatively different from allozymes in that the frequencies of alleles can be estimated, but the allelic genealogies remain unknown. If restricted to nuclear markers, this approach presents nothing conceptually new to the study of population structure, and one can safely assume that the wealth of molecular markers will supplement (or replace) allozyme data in a straightforward manner.

The second approach focuses on obtaining sequences of specific fragments of DNA from relatively small numbers of individuals from each study population. In cases where a sufficient number of phylogenetically informative nucleotide sites are included in the study, sequence data can be used to infer the phylogenetic relationships among the sequenced alleles, or gene genealogy. In some cases, such information can give insight into population structure not available from allozyme frequencies. For example, the genealogical approach may reveal the ancestral lineage of an allele which may, in turn, suggest its geographic origin. The primary problem with DNA sequence data is cost—both in manpower and supplies. As a result, studies typically examine far fewer individuals, populations, and independent gene loci than do allozyme analyses. However, as discussed below, much of the power of sequence data derive from their high phylogenetic information content; depending on the types of questions to be addressed, DNA sequencing can be cost-effective.

Given the substantial differences in the nature of the data obtained from different genetic analyses, it is of great interest to carefully compare the different approaches within given study systems. In at least two marine systems, surprising results have come from comparing allozyme data with mtDNA and nuclear DNA data. In the American oyster, *Crassostrea virginica*, allozyme frequencies were found to be relatively uniform across broad geographic ranges (Buroker 1983). This result was interpreted as evidence of high levels of gene flow among populations until subsequent mtDNA data (Reeb and Avise 1990) and nuclear DNA data (RFLPs in anonymous single copy nuclear DNA; Karl and Avise 1992) revealed a region of sharp genetic differentiation where little allozyme differentiation had previously been apparent. Because of the concordance of the mtDNA and nucDNA genetic markers, it was concluded that gene flow was restricted across the phylogeographic break and that the similarity in allozyme frequencies across the break reflected the action of stabilizing natural selection favoring multilocus heterozygotes (Karl and Avise 1992).

The second study system is the focus of the current discussion. For the past fifteen years, my lab has focused attention on the population genetics of the supralittoral copepod, *Tigriopus californicus*. (Burton and Feldman 1981; Burton 1986). We have recently extended our work on the genetic structure of natural populations of this species from allozyme analyses to studies of DNA sequence variation at a mitochondrial gene (cytochrome oxidase I = COI) and a nuclear gene (histone H1) (Burton and Lee 1994). In the following paragraphs, I will review some aspects of each of our genetic data sets for California coast *T. californicus* populations. I will then attempt to draw some conclusions about the inferences of population structure that can be drawn from each type of genetic data either by itself or in conjunction with other data.

MATERIALS AND METHODS

Tigriopus californicus were collected from high intertidal rock pools at each of ten geographic sites along the central and southern California coast, including one site on the north side of Santa Cruz Island (figure 1). Population samples (typically 200–5,000 adults) were maintained in the laboratory as breeding populations in 400 ml beakers. Within two weeks of collection and before any mortality was observed among the field-collected animals, allozyme frequencies were determined

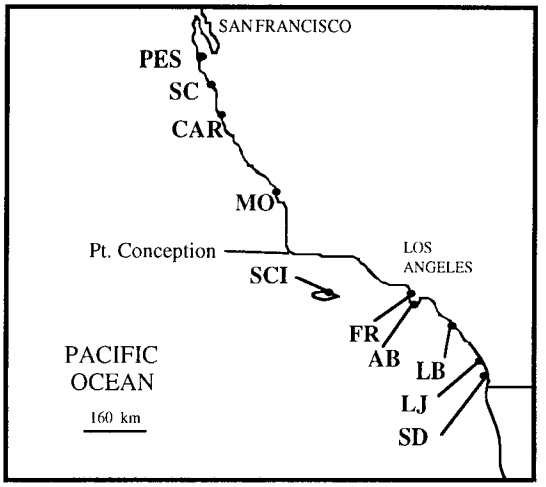

Figure 1. Map of *T. californicus* collecting sites along the central and southern California coast. PES = Pescadero Beach, SCN = Santa Cruz, CAR = Carmel, MO = Morro Bay, SCI = Santa Cruz Island, FR = Flatrock Point, AB = Abalone Cove, LB = Laguna Beach, LJ = La Jolla, SD = San Diego.

for seven gene loci by polyacrylamide gel electrophoresis, following Burton and Feldman (1981) and using stain recipes from Harris and Hopkinson (1976) and Bulnheim and Scholl (1981). A minimum of 35 individuals were scored per population for each gene locus.

Sequencing of both nuclear and mitochondrial genes was carried out on DNA extracted from 10–15 individuals from each of a number of isofemale lines established from the field-collected animals. This procedure was adopted after initial attempts to amplify COI alleles from DNA extracted from single adult *T. californicus* proved inconsistent. Because mitochondrial DNA is typically maternally inherited in animals, each isofemale line was expected to have only a single mtDNA haplotype. Although each line could contain as many as four alleles (two maternal and two paternal) for any nuclear gene, observed ambiguity in H1 sequences potentially due to multiple haplotypes segregating within isofemale lines was low, ranging from zero (13 lines) to five (2 lines) nucleotide positions per reported sequence. Since only three instances of ambiguity were observed at informative sites in the entire H1 data set (total 24 sequences with 51 informative sites), the impact of within-line polymorphism can, for the purposes of this study, be ignored.

With the exception of the Santa Cruz Island samples, the data discussed below have been previously published (see Burton and Lee 1994), so detailed protocols will not be presented here. In brief, we have used a variety of methods for DNA extraction and direct sequencing of DNA fragments amplified by the polymerase chain reaction. Our initial efforts focused on the COI gene and used total genomic DNA samples extracted by means of proteinase K digestion and phenol-chloroform extractions. Primer sequences for gene amplification and direct sequencing of the COI gene were as follows (position in the *Drosophila yakuba* mtDNA sequence [Clary and Wolstenholme 1985] are noted): primer COIK : (5'-GAGCTCCAGATATAGCATTCC-3'), 1730–1750; primer COIJ : (5'-CAATACCTGTGAGTCCTCCTA-3'), 2536–2516; primer COIL : (5'-TGAGAGAT-TATTCCAAATCC-3'), 2234–2213; primer COID: (5'-AAACCAACTGTGAACATGTG-3'), 2357–2338. Primers K and J were designed by R. Van Syoc, California Academy of Sciences. Biotinylated COIK was paired with either COID or COIJ for gene amplification by the polymerase chain reaction (PCR) using Perkin-Elmer's GeneAmp PCR Reagent Kit and the following thermocycle profile: 94°C, 1 min; 50°C, 1 min; 72°C, 2 min, for 35 cycles followed by 5 min at 72°C. PCR products were run on a 2% agarose gel and extracted from the agarose using Geneclean (Bio 101 Inc.). Solid-state sequencing (Ausubel et al. 1992) of the biotinylated strand with primers COIL and COID used Sequenase protocols (US Biochemical) following capture of the biotinylated strand on streptavidin-coated magnetic beads (Dynal, Inc.).

A previously published sequence of a histone H1 gene from *T. californicus* (Brown et al. 1992) was used to design primers for PCR amplification and sequencing of a fragment including the 5' end of the H1 coding region. Primers used were H1.5 (5'-ATATGTGTC-GAATCGAGGGC-3', position 137–156 in the published sequence) and H1.3 (5'-TCTCGACCAAGGACTTG-3', position 710–694). DNA samples used for H1 PCR were prepared by boiling 15 animals from a single isofemale line in 200 µl of 10% chelating resin (Sigma Chemical) for 8 minutes; after boiling, samples were vortexed for 10 sec and centrifuged (13,000 × g) for 2 minutes at 4°C (Walsh et al. 1991). Ten µl of the supernatant were used as template in the PCR reaction, which used the same thermocycle profile as COI PCR. We used Promega's Magic PCR Preps to purify PCR products, and carried out direct sequencing with the *f*mole Cycle Sequencing Kit (Promega) and manufacturer's protocols with ^{32}P-end-labeled primers. When possible, we analyzed the same isofemale lines for both COI and H1 sequence. Unfortunately, much of the COI sequencing had been completed and a number of the isofemale lines lost before the H1 sequencing project was initiated. We are currently trying to obtain COI sequences from the Santa Cruz Island site. As a result, both COI and H1 DNA sequences were obtained for only eight lines.

RESULTS

Allozyme data from ten population samples obtained in 1992–93 are presented in table 1. Despite the fact that each of the gene loci studied is polymorphic when the total data set is considered, within-population hetero-

TABLE 1
Allelic Frequencies for Six Polymorphic Allozyme Loci in 13 Populations of *T. californicus* along the California Coast

	Pgi					*Got2*				*ME*			
Population	F*	M	MS	S	VS	VF	F	S	VS	VF	F	S	VS
PES	0.46	0.54						1.00				1.00	
SCN		1.00						1.00				1.00	
CAR		1.00						1.00				1.00	
MO		1.00						1.00				1.00	
SCI-N		0.01	0.66	0.33				1.00				1.00	
SCI-S			0.02	0.95	0.03			1.00					
CO		1.00						1.00					
PD		0.58	0.31	0.11				1.00					
FR		1.00						1.00				1.00	
AB		1.00						0.99	0.01			0.98	0.02
LB	0.02	0.79		0.19		0.40	0.60			1.00			
LJ	0.06	0.94					1.00				1.00		
SD	0.22	0.78				1.00					1.00		

	Got1					*Gpt*			*Apk2*		*Apk1*	
Population	F	S	VS	ES	SS	F	S	VS	F	S	F	S
PES		1.00					1.00			1.00	1.00	
SCN	0.03	0.97				0.25	0.75			1.00	1.00	
CAR		0.98	0.02				1.00			1.00	1.00	
MO		0.80	0.20				1.00			1.00	1.00	
SCI-N	0.58	0.42					0.07	0.93		1.00		1.00
SCI-S		1.00					1.00			1.00	1.00	
CO		1.00					1.00			1.00	1.00	
PD		1.00									1.00	
FR	0.04	0.93	0.03				1.00			1.00	1.00	
AB		0.07		0.93			1.00			1.00	1.00	
LB	0.11	0.85			0.04		1.00		1.00		1.00	
LJ	0.19	0.79	0.02				1.00		1.00		1.00	
SD		0.88	0.12				0.94	0.06	1.00		1.00	

*Allelic designations VF, F, M, MS, S, VS, ES are in order of decreasing anodal mobility. Locus designations: Pgi = phosphoglucose isomerase, Got = glutamate-oxaloacetate transaminase, ME = Nadp-malic enzyme, Gpt = glutamate-pyruvate tranaminase, Apk = arginine phosphokinase.

TABLE 2
Nei's Genetic Distances (Standard Errors) Based on Seven Allozyme Loci and Calculated with the Jacknife Method of Mueller (1979)

	SCN	CAR	MO	SCI	FR	AB	LB	LJ	SD
PES	.043 (.006)	.032 (.005)	.039 (.006)	.586 (.057)	.033 (.005)	.178 (.025)	.647 (.064)	.686 (.065)	.664 (.065)
SCN		.009 (.001)	.014 (.002)	.593 (.058)	.009 (.002)	.146 (.023)	.626 (.061)	.649 (.063)	.669 (.064)
CAR			.004 (.001)	.621 (.059)	.000 (.000)	.133 (.022)	.582 (.059)	.608 (.061)	.627 (.062)
MO				.623 (.061)	.003 (.001)	.114 (.019)	.605 (.061)	.626 (.063)	.645 (.063)
SCI					.613 (.059)	.741 (.071)	2.44 (.20)	2.62 (.28)	2.59 (.20)
FR						.127 (.021)	.586 (.059)	.610 (.061)	.632 (.062)
AB							.819 (.078)	.820 (.078)	.878 (.081)
LB								.214 (.030)	.265 (.033)
LJ									.181 (.029)
SD									

zygosity is low ($H = 0.07 \pm 0.02$, mean $\pm$ SE), and many alleles are restricted in their distribution. Calculations of Nei's genetic distance (D) based on these relatively few loci are subject to considerable error but are useful as relative measures of population differentiation and are presented in table 2. Calculated D values range from zero (FR-CAR) to 2.62 (SCI-LJ); in the latter comparison the two populations share no alleles at four of seven loci. Although based on few loci, the observed D values are approximately an order of magnitude higher than typically encountered among conspecific populations (Thorp 1983). There is a striking lack of correspondence between the geographic distance between populations and genetic distance between the same populations (figure 2).

Twenty-one COI sequences (each was 500 base pairs long) were determined from isofemale lines of *T. californicus* derived from seven geographic populations (work on the Santa Cruz Island population is now being initi-

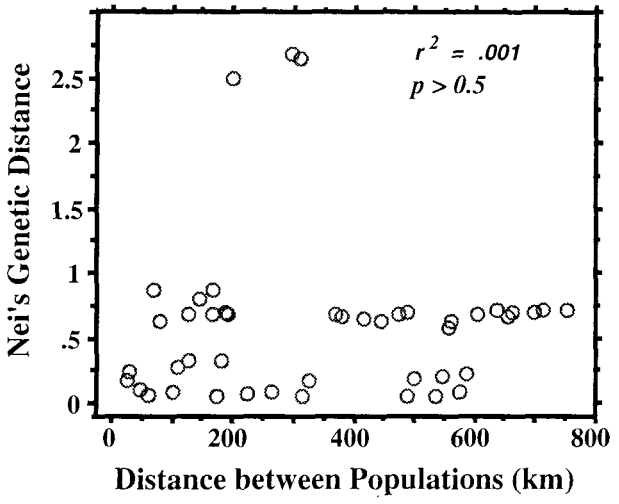

Figure 2. Relationship between geographic distance and Nei's genetic distance for all pairwise comparisons of populations in table 1. Although all points on the graph are not independent, no relationship between geographic distance and genetic distance is apparent from the allozyme data.

ated). The sequences were used to construct a maximum parsimony gene tree (Swofford 1989) that approximates the genealogical relationships among the sequences (figure 3A). Three results are apparent from the parsimony tree: (1) Extensive divergence exists between central and southern California COI sequences. Over 18% of all nucleotide sites were different in comparisons of any of the central (PES, SCN, CAR, MO) sequences to any of the southern (AB, FR, SD) sequences. (2) Although little variation was observed among isofemale lines derived from most of the natural populations, one strongly divergent haplotype was seen at SD; this cautions that additional data are needed before strong conclusions concerning population structure can be made on smaller geographic scales. (3) Bootstrap resampling of the data indicate that only the SCN population can be discriminated from other populations within a given (central or southern) region.

To assess whether or not the phylogeographic break observed in the COI data is reflected in the nuclear genome, we also sequenced approximately 500 bases of a histone H1 gene, including approximately 150 bases of the coding sequence and 350 bases of 5' flanking region (Burton and Lee 1994). In addition to the populations sampled for COI sequences, two sequences from Santa Cruz Island (SCI), two from La Jolla (LJ), and one from Laguna Beach (LB) have been obtained to date. The H1 PCR products showed a size polymorphism, with template DNA from all 11 isofemale lines from southern California sites (SD, LJ, LB, AB, FR) producing larger PCR products than that from the four central sites (MO, CA, SC, PES). PCR products from the SCI site were of the smaller size (i.e., consistent with being of the "central" type). Subsequent sequencing

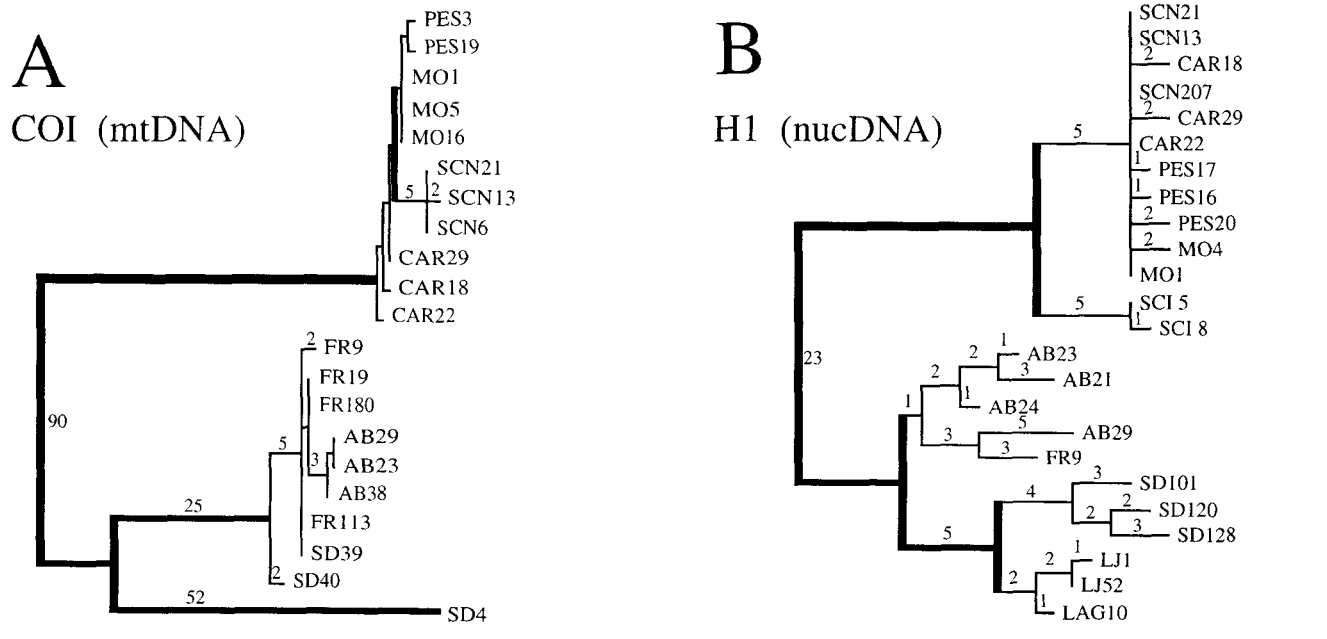

Figure 3. Genetic relationships along *T. californicus* isofemale lines. *A*, Unrooted maximum parsimony tree produced by the heuristic search procedure of PAUP 3.0r (Swofford 1989) for 21 COI sequences. Sequences are named according to source population, and numbers refer to the specific isofemale line. Branch lengths are the absolute number of nucleotide substitutions over the 500 base sequences. Tree shown is one of five equally parsimonious trees, all with similar topology. *Bold lines* indicate branches supported by bootstrap values >95%.
B, Unrooted maximum parsimony tree produced by the heuristic search procedure of PAUP 3.0r for 24 H1 sequences. Sequences are named as in *A*. Branch lengths are absolute number of nucleotide substitutions over the 425 bases sequenced (after exclusion of gaps). Tree is one of three similar, equally parsimonious trees. *Bold lines* indicate branches supported by bootstrap values >95%.

showed that two insertions (48 and 20 bases in length) in the 5' flanking region were found only in the southern samples. After these insertions/deletions are eliminated from the data set, an additional 23 of 425 bases (5.4%) distinguish between central and southern clades (figure 3B). The SCI sequences are differentiated from the central California populations, but clearly group with the central sequences in the parsimony tree; this topology is supported by 100% of bootstrap resamplings.

DISCUSSION

The Study System: *Tigriopus californicus*

The harpacticoid copepod *T. californicus* is a common inhabitant of high intertidal and supralittoral rock pools along the Pacific coast of North America, from Alaska to Baja California. The supralittoral distribution of *T. californicus* appears to result from heavy predation in the lower intertidal zone, and it is not commonly encountered in pools that harbor either fish or anemones (Dethier 1980). Restricted primarily to small ephemeral pools, *T. californicus* populations fluctuate dramatically in size; evaporative drying and extensive scouring during storms frequently lead to extinction of subpopulations inhabiting individual rock pools. The species has no resting or dormant stages capable of withstanding complete desiccation; newly refilled pools must be repopulated by immigrants. Although all life stages of *T. californicus* are free-swimming and potential dispersers, the means by which the species colonizes distant habitat is not known. Passive drifting of adults or larvae has not been documented; *T. californicus* has not (to my knowledge) been collected in plankton tows. Both larvae and adults tend to cling to the substrate in response to water turbulence, so rafting on debris washed out of inhabited pools as well as attaching to birds and crabs moving through pools are alternate means of dispersal.

Given its natural history, *T. californicus* is not representative of the ecology of species with extensive larval dispersal. However, it may provide a model system for the study of diverse benthic species that have direct development and show limited larval dispersal. In this light, many attributes of the *T. californicus* system make it particularly amenable to genetic analysis: (1) *T. californicus* is easily raised in the lab, with a generation time of about three weeks. (2) Controlled lab crosses are readily achieved, allowing formal genetic analysis of all markers employed in population studies. (3) *T. californicus* is easily collected at a variety of spatial scales, and the availability of habitat is simple to assess.

Allozymes

Although individual *T. californicus* are small (approximately 35 μg wet weight), we can consistently resolve a number of polymorphic enzyme loci on thin (0.8 mm) polyacrylamide gels. This work has demonstrated that populations inhabiting pools on a single rock outcrop are typically genetically homogeneous, whereas those separated by even short stretches of sandy beach (<500 meters) are often sharply differentiated (Burton and Feldman 1981). In fact, most of the 25 populations we have sampled along the California coast can be readily distinguished on the basis of allozyme frequencies. In many cases, alleles restricted to single populations ("private" alleles; Slatkin 1985) reach high frequency, strongly suggesting a lack of gene flow among the populations. In some cases, alleles fixed in one population have never been observed in other populations. Many of our study sites have been sampled at least five times (in some cases more than fifteen times) over periods of time extending up to 15 years; although some minor variation in allelic frequencies has been observed over this period, the geographic distribution of alleles and patterns of population differentiation have been essentially constant. For example, the Gpt^F allele has been found only in Santa Cruz County and has been consistently in the 20%–25% frequency range at the SCN site since 1979 (Burton and Feldman 1981; Burton 1986).

One striking feature of the *T. californicus* allozyme data is that, even when averaged across loci, no relationship between geographic distance and genetic distance is observed. This lack of correlation between geographic distance and genetic differentiation emphasizes a major limitation of the allozyme data: allozymes provide no phylogenetic information. When neighboring populations show one or more "private" alleles at high frequencies, it is difficult to determine whether the observed differentiation occurred *in situ* (over relatively long time periods), or if the initial immigrants colonizing the relevant geographic locales derived from already differentiated parent populations. Distinction between these hypotheses requires genealogical information. Along these lines, it is interesting to note that the SCI population appears to be equally divergent from its nearest neighboring populations to either the north or south. Was SCI colonized from northern or southern source populations? Again, resolution of this question requires genealogical information.

On the other hand, the speed and ease of allozyme analyses has allowed us to study population structure at a spatial and temporal scale that would not be practical using DNA sequencing techniques. Scoring allelic frequencies at several polymorphic gene loci with sample sizes of 50 individuals from each of ten populations can be completed in a week. The relatively few DNA sequences presented here required over a year to obtain.

Mitochondrial DNA Sequence

The most striking feature of the COI data is the great divergence between central and southern California clades. Minimum distance between sequences from the two clades exceeds 18%, a value that far exceeds any previously reported intraspecific value for a protein-encoding mitochondrial gene (Avise et al. 1987; Harrison 1989). Although we have no fossil record for establishing a molecular clock directly relevant to *T. californicus* populations, recent data on the COI gene in other crustacea yielded estimates of 2.2%–2.6% divergence per million years (Knowlton et al. 1993); at such a rate of sequence evolution, a conservative date for the split between central and southern COI clades is on the order of 7 million years. We also note that of 90 nucleotide substitutions (of the 500 bases sequenced) distinguishing the two major clades, only 3 result in diagnostic amino acid substitutions (6 amino acid residues are variable of the 165 inferred from sequence data).

A second remarkable feature of the COI data is the extreme divergence among the San Diego (SD) isofemale lines; the SD4 line differs from SD40 at 77 nucleotide sites (15.4%); notably, all these substitutions are synonymous. The existence of highly divergent mtDNA haplotypes within a population can be explained in two ways. First, it is possible that two maternal lineages have been maintained in the SD population long enough for the high level of divergence to have occurred *in situ*. The second possibility is that immigration from another population introduced divergent haplotypes into the SD population. Although neither hypothesis can be rejected, the extreme differentiation of the haplotypes argues against the former hypothesis. It seems unlikely that two lineages would have survived stochastic extinction ("lineage sorting") over millions of years in the ephemeral environment characteristically inhabited by *T. californicus* (Neigel and Avise 1986). Although more extensive within-population sampling is the subject of continuing work, the divergence at the SD site stands in marked contrast to the low levels of divergence (<1%) observed among isofemale lines from six other populations.

Nuclear DNA Sequence

An important (and perhaps unusual) feature of the H1 data presented here is the fact that no evidence for recombination between the sharply differentiated central and southern California clades has yet been observed. The simplest explanation for this lack of recombination is that the differentiation of the clades occurred allopatrically and, without subsequent gene flow, no opportunity for inter-clade recombination has existed (Burton and Lee 1994). This hypothesis seems entirely consistent with the other genetic data sets presented above. Attributes of the allozyme data (high-frequency private alleles) and the COI data (strong differentiation of central and southern clades) clearly suggest long periods of allopatric population differentiation.

The two sequences obtained from SCI suggest an answer to the question posed above concerning the origin of this population: H1 sequences from the SCI population are more closely related to those from central populations than to those from southern populations. The addition of COI sequence data may substantially strengthen support of this hypothesis. Although the SCI H1 sequences clearly group with the central clade, they are clearly differentiated from other members of the clade, as would be expected from the extensive allozyme differentiation between SCI and other central California populations.

Allozyme versus DNA Sequence Data

The relative value of the different types of genetic data discussed above depends primarily on the specific questions being asked. Many questions, some ecological and some evolutionary, can be addressed by population structure data. Neither allozyme nor sequence data alone can efficiently answer all commonly posed questions, although obviously any allozyme differentiation could ultimately be revealed at the DNA sequence level if the appropriate gene is sequenced.

One question of primary interest to both ecologists and fisheries managers is, "To what extent do geographically separated populations represent separate gene pools?" Analysis of population differentiation is inherently a one-sided experiment; i.e., although finding genetic differences between populations can often lead to strong conclusions, a lack of differentiation is typically ambiguous. This is because analysis focuses on only a tiny portion of the genome. Populations that appear to be identical on the basis of one sample of gene loci may later be found to be sharply differentiated at other loci. Because numerous loci can be surveyed quickly and with relative ease, allozyme studies remain a method of choice in assessing differentiation of geographic populations. In the *T. californicus* system, we have found allozymes to be remarkably sensitive in discriminating populations. Using from five to eight allozyme polymorphisms as markers, we can distinguish the great majority of over 20 California coast populations that we have studied over the past 15 years. In contrast with the results observed in a number of other species, most genetic diversity in *T. californicus* is revealed by between-population comparisons; heterozygosity within populations is relatively low. This might easily be explained by frequent population bottlenecks, as would be expected in an ephemeral habitat. However, closer inspection of the data adds one important feature: bottlenecks may explain why heterozygosity is low, but do not explain

the existence of high-frequency "private" alleles. Such a pattern of variation necessarily leads to the conclusion of highly restricted gene flow (Slatkin 1985).

Identification of discrete populations for ecological analysis or management purposes requires only that sufficient levels of genetic polymorphism be available for study. Morphological, allozyme, or DNA markers with simple Mendelian inheritance are potentially of equal value for addressing questions of gene pool continuity. From this perspective, the relative merits of allozyme and DNA approaches are apparent: allozymes are cheaper and faster, whereas DNA necessarily possesses a greater wealth of genetic variation. When beginning to analyze a new species or set of populations, some effort at the level of allozymes would seem a prudent methodological choice. If sufficient allozyme polymorphism is not found, a search for DNA sequence variation would be a next step. Once polymorphic nucleotide sites are identified, efficient methods for discriminating genotypes at particular nucleotide sites can be developed (see, e.g., Banks and Hedgecock 1993).

As mentioned above, two classes of hypotheses can be proposed to explain sharp differentiation of neighboring populations: (1) the populations were initially similar but have diverged via either random drift or selection *in situ* over a sufficient period of time to allow establishment of new mutant alleles; or (2) the original colonists were derived from already differentiated populations, so the present genetic differentiation simply reflects the indeterminate nature of the colonization process. Although conclusively demonstrating population differentiation, allozyme data alone cannot discriminate between these two very different explanations for the origin of the differentiation. Here, DNA sequence data clearly complement our allozyme studies. For example, the strong differentiation of the AB and FR populations (nearly fixed on alternate alleles at the *Got1* locus) appears to be the result of *in situ* differentiation, since COI and H1 DNA sequences from these populations are more closely related to each other than to those from any other population. In contrast to the lack of relationship between geographic distance and allozyme genetic distances, the DNA sequence data show clear geographic structure, with neighboring populations more similar than geographically distant populations.

When allozyme and DNA sequence data for the *T. californicus* system are compared, one striking discrepancy is immediately apparent: allozyme data suggest that Los Angeles area populations (AB and FR) are genetically similar to central California populations, whereas sequence data show exceptionally strong divergence between these same sets of populations. Although such discordance may provide insights into the natural history of conspecific populations (e.g., Reeb and Avise 1990; Karl and Avise 1992; Burton and Lee, in press), it simply emphasizes the fact that only weak inferences can be made from apparent population homogeneity. In *T. californicus*, DNA sequence data has dramatically changed our understanding of population structure by revealing a regional pattern of genetic variation that the allozyme data lacked. The geographic match between the two sets of sequence data suggests that population divergence has proceeded for enough time that concordant lineage-sorting of mitochondrial and nuclear genes has taken place (Neigel and Avise 1986; Avise and Ball 1990). The phylogeographic break in the historic distribution of the species lies somewhere between LA and the Morro Bay study sites, a region that includes Point Conception, a widely recognized zoogeographic boundary (Valentine 1966). The SCI population, lying south of Point Conception itself, appears to be derived from more northern populations based on H1 sequence data. However, the allozyme data show that SCI is highly divergent from populations to either the north or the south, indicating that the colonization of SCI occurred in the distant past and no ongoing gene flow is taking place. The genetic isolation of Santa Cruz Island was previously reported for the embiotocid fish, *Damalichthys vacca* (Haldorson 1980).

In summary, the *T. californicus* studies show that comparisons between allozyme and DNA sequence data can reveal a variety of surprises. First, despite the fact that protein electrophoresis is a relatively low-resolution technique compared to DNA sequencing, populations were generally more easily distinguished by allozyme data than by sequence data. Other molecular techniques, such as allele-specific PCR, could undoubtedly be developed to distinguish the populations based on nucleotide substitutions, but allozyme methodology has proven sufficient (and efficient) for simply demonstrating genetic isolation of populations in the *T. californicus* system. DNA sequence data, however, in revealing unprecedented levels of divergence between conspecific populations, provide a historical framework within which the allozyme data can be more fully understood.

ACKNOWLEDGMENTS

This work was supported by NSF grants BSR-8741308, DCB-8811227, and IBN-92-96185. I thank Cindy Lau and Linh Ngo for laboratory assistance.

LITERATURE CITED

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, eds. 1992. Short protocols in molecular biology. 2d ed. N.Y.: John Wiley.

Avise, J. C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Ann. Rev. Genet. 25:45–69.

———. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. Oikos 63:62–76.

Avise, J. C., and R. M. Ball, Jr. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Oxford Surveys in Evol. 7:45–67.

Avise, J. C., J. Arnold, R. M. Ball, Jr., E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489–522.

Banks, M. A., and D. Hedgecock. 1993. Discrimination between closely related Pacific oyster species (*Crassostrea*) via DNA sequences coding for large subunit rRNA. Mol. Mar. Biol. Biotech. 2:129–136.

Brown, D., A. Cook, M. Wagner, and D. Wells. 1992. Closely linked H2B genes in the marine copepod *Tigriopus californicus* indicate a recent gene duplication or gene conversion. DNA Sequence 2:387–396.

Bulnheim, H. P., and A. Scholl. 1981. Genetic variation between geographic populations of the amphipods *Gammarus zaddachi* and *G. salinus*. Mar. Biol. 64:105–115.

Buroker, N. E. 1983. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. Mar. Biol. 75:99–112.

Burton, R. S. 1983. Protein polymorphisms and genetic differentiation of marine invertebrate populations. Mar. Biol. Lett. 4:193–206.

———. 1986. Evolutionary consequences of restricted gene flow among natural populations of the copepod, *Tigriopus californicus*. Bull. Mar. Sci. 39: 526–535.

Burton, R. S., and M. W. Feldman. 1981. Population genetics of *Tigriopus californicus*. II. Differentiation among neighboring populations. Evolution 35: 1192–1205.

Burton, R. S., and B.-N. Lee. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. Proc. Natl. Acad. Sci. USA. 91:5197-5201.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22:252–271.

Dethier, M. N. 1980. Tidepools as refuges: predation and the limits of the harpaticoid copepod *Tigriopus californicus* (Baker). J. Exp. Mar. Biol. Ecol. 42:99–111.

Haldorson, L. 1980. Genetic isolation of Channel Islands fish populations: evidence from two embiotocid species. *In* The California islands: proceedings of a multidisciplinary symposium, D. M. Power, ed. Pp. 433–442.

Harris, H., and D. A. Hopkinson. 1976. Handbook of enzyme electrophoresis in human genetics, Amsterdam: North-Holland.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4:6–11.

Hedgecock, D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? Bull. Mar. Sci. 39:550–564.

Karl, S. A., and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256:100–102.

Knowlton, N., L. A. Weigt, L. A. Solorzano, D. K. Mills, and E. Bermingham. 1993. Divergence in protein, mitochondrial DNA and reproductive compatibility across the Isthmus of Panama. Science 260:1629–1632.

Mueller, L. D. 1979. A comparison of two methods for making statistical inferences on Nei's measure of genetic distance. Biometrics 35:757–763.

Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. *In* Evolutionary processes and theory. E. Nevo and S. Karlin, eds. Pp. 515–534.

Palumbi, S. R. 1992. Marine speciation on a small planet. Trends Ecol. Evol. 7:114–118.

Reeb, C. A., and J. C. Avise. 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. Genetics 124:397–406.

Scheltema, R. S. 1986. Long-distance dispersal by planktonic larvae of shoal-water benthic invertebrates among central Pacific islands. Bull. Mar. Sci. 39:241–256.

Slatkin, M. 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Swofford, D. L. 1989. Phylogenetic analysis using parsimony (PAUP), version 3.0r. Champaign, Ill.: Illinois Natural History Survey.

Thorpe, J. P. 1983. Enzyme variation, genetic distance and evolutionary divergence in relation to levels of taxonomic separation. *In* Protein polymorphism: adaptive and taxonomic significance, G. S. Oxford and D. Rollinson, eds. London: Academic Press, pp. 131–152.

Valentine, J. W. 1966. Numerical analysis of marine molluscan ranges on the extratropical northeastern Pacific shelf. Limnol. Oceanogr. 11:198–211.

Walsh, P. S., D. A. Metzger, and R. Higuchi. 1991. Chelex 100 as a medium for simple extraction of DNA for a PCR-based typing from forensic material. BioTechniques 10(4):506–513.