

UC San Diego

Research Theses and Dissertations

Title

Genetic Analysis of Recruitment and Dispersal Patterns in the Porcelain Shore Crab, *Petrolisthes cinctipes*

Permalink

<https://escholarship.org/uc/item/1z97c98j>

Author

Toonen, Robert J.

Publication Date

2001

Peer reviewed

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

Genetic Analysis of Recruitment and Dispersal Patterns in the Porcelain Shore Crab,
Petrolisthes cinctipes

BY

ROBERT JOHN TOONEN
B.S. (University of Alberta) 1991
M.S. (University of North Carolina, Wilmington) 1993

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Population Biology

in the

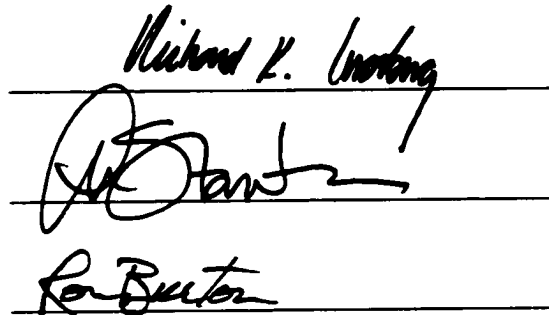
OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Three handwritten signatures are written on three horizontal lines. The top signature is 'Michael L. Unwin', the middle signature is 'R. J. Toonen', and the bottom signature is 'R. Burton'.

Committee in Charge

2001

UMI Number: 3039178



UMI Microform 3039178

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Parents are wonderful people, too often misunderstood,
and I have not sung your praises nearly as often as I should.

You have always been there for us, and you taught us right from wrong.
You have always believed that I would accomplish this, even when it took so long.

So many memories together, most happy, some sad,
but all things considered, it's a wonderful life I've had.

Thanks for the rules I resented and wished were not there,
They were for my own good, and are a sign of how much you both care.

You held our hands, you eased our pain, and soothed all of our fears.
Thinking back now, I can only say thank you for all of those years.

As children, Mark and I didn't realize how good our life was,
It was simply our life, there was no 'why' or 'because'.

Dad was working every day in the rat-race of life,
Trying to make a good living, to support his kids and his wife.

So, somehow it always seemed to be Dad's job to work to pay the bills,
While Mom's job was to be there to kiss skinned knees, and nurse all our ills.

Perhaps that is the reason that we somehow got the notion,
That Dad was so tough, that he seemed to lack emotion.

But I know now that deep inside Dad's heart, where no one can see,
He's a sentimental old guy, loving, caring, and as soft as can be.

That is why I want to dedicate this work to my Dad
To tell my parents how much I love them, and to make them both glad

That despite my short-comings, and the all times they wanted to kill me,
They restrained themselves, and waited to see what kind of man I'd turn out to be.

It may not always have shown, but my love and appreciation has always been strong,
it has never diminished, if anything increased, as time moves along.

I may not have understood at the time, but as I grow older your wisdom appears,
and again I can only say thank you for all of those years.

– Rob Toonen

Acknowledgments

As I sit trying to write this section I wonder if anyone will ever really read it, and I suspect that even those who do will more than likely just skim through to see if their name is recorded on these pages before moving on. I hope that is not the case, because the work that follows would not have been possible without the help of the friends, mentors, labmates and family that have supported, encouraged, helped and guided me through this experience.

In the time honored tradition of first thanking my major professor, I want to thank Rick Grosberg. I could not have asked for a better advisor than Rick. For all the “tough guy” facade that he can sometimes show, Rick cares very deeply about the well-being and success of his students, and has been a superb mentor. He has been wonderfully supportive through some tough times for me, and after the fiasco of my oral qualifying examination, I would have quickly left Davis had it not been for Rick and the other members of the Grosberg lab. Rick has left me to my own devices when I wanted that space, but has always been there to advise or direct me when I needed him. Rick has been an outstanding role-model and mentor, and I will be lucky if even a small fraction of his talents have rubbed off on me during my time in his lab.

I would also like to thank Mau Stanton and Ron Burton for agreeing to be part of my dissertation committee and for their encouragement and support. I especially appreciate their understanding and acceptance of the time crunch and their quick

responses to my chapter revisions in order to get everything finished and submitted on time.

Thanks to Kelly Batchelor and my wife, Carol Fong, for enduring the elements and helping me to collect crabs in a number of surreal and sometimes hellish trips along the coast. Not many people were willing to come out and sample animals on both the morning and nighttime tides for five consecutive days with little sleep, but Carol and Kelly even managed to somehow convince me that they enjoyed it. I thank Kelly Batchelor, Carol Fong, Melissa Locke, and Maureen Cahill for their help in labeling vials, measuring crab chelae, extracting DNA and doing PCR amplifications on the thousands of crabs included in this dissertation study (and the many thousands more that ended up in the freezer unanalyzed). Without their help, I would still be doing extractions and PCR rather than writing this!

Coming from a long background of latex-glove-phobia, I owe a great debt of thanks to Brenda Cameron and Peter Marko for teaching me almost everything I know about the molecular techniques that I eventually used to complete my dissertation. I thank Brenda for being such a great lab-mom – for the numerous times that she ended up finishing things that I forgot in lab, for contributing to my ever-expanding waistline with her wonderful baking, and for not getting upset with me when I re-invented the Molotov Cocktail late one night in the gel lab (or any of the other times I started something in the lab on fire, for that matter). Despite my cowboy attitude to molecular work, Brenda just sighed and said that her children would never get a chance to turn her grey, because I had beat them to it. ☺

The parade of collaborators, “Gros-docs” and fellow graduate students that have passed through Rick’s lab during my time in Davis have all contributed greatly to my development and enjoyment of my time here, and I thank David Ayre, Lisa Borghesi, Dave Carlon, Sean Craig, Emmet Duffy, Olaf Ellers, Melissa Frey, Sarah Gilman, Mike Graham, Dennis Hedgecock, Jamie King, Susan Keen, Don Levitan, Jon Lewis, Kyle Summers, Neil Tsutsui and especially Mike Hart, Mike Hellberg, Ed Pearson, Peter Marko, and Alex Wilson for their suggestions, guidance, assistance, and lively discussion that have so greatly influenced me (probably even more than I realize). I also thank Paul Bentzen for letting me come up to Seattle and hang out at the Marine Molecular Biotechnology Laboratory at University of Washington while I tried to learn as much as I could from Pam Jensen and Patrick O’Reilly. In addition to the people above, I also thank Paul Sunnucks for listening to me complain about my data not making any sense, and helping to keep me enthusiastic about my project despite the difficulties. I thank Chuck Langley, Dennis Hedgecock, Steve Morgan and Megan Donahue for their contributions in helping me to finally figure out why my data didn’t make sense!

Thanks to the staff of the Section of Evolution and Ecology (especially Linda Arnold, Michelle Davenport, Janet Dillon, Carla Muñoz, Marina Rumiansev, and Mike Zumalt) for always being there whenever I needed something, and for helping to get me out of all of the many messes that I got myself into over the years.

This project was supported by funding from a variety of sources, primarily sponging off NSF grants for the *Hydractinia* project, and a California Sea Grant

(R/F-177) Rick and I were awarded to study Dungeness Crab. I also received funding from the Jastro Shields Committee, Center for Population Biology, and Graduate Division at the University of California, Davis. The Province of Alberta awarded me a Sir James Lougheed Award of Distinction, and the Natural Sciences and Engineering Research Council of Canada funded me through a Post-Graduate Scholarship B despite the fact that I was in California. Visa and MasterCard picked up the slop where necessary.

Last, I save the traditional place of honor for thanking my family. Family has always been extremely important to me, and more than ever as I write this. Thanks, Mom and Dad, for always believing in me and supporting me, even if you haven't always understood quite what I was doing or why. I love you both very much. Thanks to Mark and Sanne for your encouragement, love and support. I was able to concentrate on my research exclusively due to the generous spousal RA and computer allowance that Carol has provided me over the years. Much more than that, however, I owe Carol a great deal of thanks for her love, support, and understanding during the times when I was stressed and bitchy; her stoic acceptance of the times that I was too busy to go shopping or sight-seeing with her; and her ability to make me smile and forget all my troubles for a while anytime.

Abstract

Understanding the role of planktonic larval dispersal remains one of the major hurdles facing marine population biologists. Using microsatellite markers, I characterized genetic structure within and among 12 geographic populations that span nearly the entire species range (from Morro Bay, CA to Neah Bay, WA.) of the porcelain shore crab, *Petrolisthes cinctipes*. Retention zones associated with headlands (the upwelling-relaxation flow hypothesis) cannot account for the pattern of genetic structure observed among geographic populations of *P. cinctipes*. Inferred levels of effective gene flow among populations decrease with geographic distance of separation between sites, but this relationship appears nonlinear. This curvilinear relationship suggests that pelagic larvae of *P. cinctipes* settle with greatest frequency at an intermediate distance from the site of release.

Estimates of population subdivision (using both θ and ρ) among northern California sites from collections made in each of 1997, 1998, and 1999 are significant and consistent. However, these single-year estimates appear misleading relative to the overall data set in which the observed genetic variance is partitioned by sampling year, site, and age class (recruits, subadults and adults). Geographic population structure changes from year-to-year in such a way that the overall estimate of population subdivision among sites (pooling the three sampling years) is not significantly different from zero. Instead, collection year and age class appear to account for the significant differentiation among samples. This result argues that evolutionary inferences drawn from single-time samples of *P. cinctipes* are unreliable. If this pattern proves to be a

general phenomenon among marine invertebrates with planktonic larvae, studies of the genetic structure of marine organisms that sample populations only once through time should be viewed with caution.

These data are consistent with genetic structure in *P. cinctipes* being a combined result of 1) variation in the source of larval recruits across years, 2) individual variation in reproductive success, and 3) temporally and spatially variable natural selection.

Further understanding will require refined theoretical expectations for marine populations that are far from equilibrium, linked by complex patterns of dispersal, and in which a variety of contemporary and historical processes are likely acting synergistically.

Chapter 1.

Isolation and inheritance patterns of microsatellite markers in the intertidal crab *Petrolisthes cinctipes*.

Robert J. Toonen

**Center for Population Biology,
Section of Evolution and Ecology
University of California, Davis,
Davis, CA, USA. 95616.**

Abstract

I isolated and developed primers for 17 microsatellite loci in the anomuran crab *Petrolisthes cinctipes* as a prelude to a comprehensive analysis of genetic structure in this species. I excluded from subsequent analyses any loci that 1) produced unscorable multiband amplification products; 2) showed evidence of null alleles; 3) failed Slatkin's (1994, 1996) exact test for a fit to the Ewens sampling distribution for selectively neutral alleles; 4) did not amplify a product of the expected size; 5) failed to amplify any product from DNA extracted from the original crab from which the library was developed; or 6) exhibited non-Mendelian patterns of inheritance. Non-Mendelian inheritance was assessed by comparing the genotypes of brooded embryos to their mothers. Amplification of two loci (Pc160b, and Pc36L) from brooded embryos revealed heterozygous amplification products that did not match either apparent allele amplified from the mother. On the basis of these criteria, 15 of the 17 loci isolated in this study are excluded. The data generated from testing for non-Mendelian inheritance also allows an inference about the mating structure of this crab. Using the two remaining loci, I found that eight of ten broods analyzed show evidence of multiple paternity. This result suggests that multiple paternity of broods is widespread in *Petrolisthes cinctipes*.

Introduction

Approximately 80% of marine organisms (about 90,000 species), both vertebrate and invertebrate, have a biphasic life-cycle and produce planktonic larvae that spend some variable amount of time (ranging from minutes to months) developing in the water column before settling and metamorphosing into adults (Thorson 1961, 1964). The potential for dispersal of motile larvae during this planktonic stage appears great; consequently gene flow among populations of benthic marine species with long-lived pelagic larvae ought to be high. The relationship between gene flow and population genetic structure is well known: "gene flow, if unopposed by other factors, homogenizes the populations of a species" (Futuyma 1998). Numerous empirical studies find that the degree of genetic divergence among populations of species with long-lived planktonic larvae is generally lower than among species lacking such larvae (e.g., Berger 1973; Waples 1987; Hellberg 1996, reviewed by Burton 1983; Hedgecock 1986; Palumbi 1995; Bohonak 1999). However, there are many exceptions to this general pattern (e.g., Benzie and Williams 1997; Todd et al. 1998; Kyle and Boulding 2000). Despite the potential for extensive dispersal via planktonic larvae, significant population structure over very small spatial scales has been found repeatedly among marine populations (reviewed by Avise 1994; Hedgecock 1994; Grosberg and Cunningham 2001). Perhaps the most consistent result from empirical studies of marine populations is that small-scale genetic heterogeneity occurs even among populations lacking obvious large-scale structure (e.g., Johnson and Black 1982;

Johnson and Black 1984a, b; Lacson et al. 1989; Karl and Avise 1992; Kordos and Burton 1993; Edmands et al. 1996; Todd et al. 1998; Kyle and Boulding 2000; Moberg and Burton 2000). In fact, many of these studies have revealed that estimates of genetic differentiation among adjacent populations (tens of km) is of the same, or greater, magnitude than values across the species range (thousands of km) (e.g., Johnson and Black 1982; Johnson and Black 1984a, b; Kordos and Burton 1993; Edmands et al. 1996; Moberg and Burton 2000).

How such fine-scale genetic structure can accumulate in species with apparently high potential for dispersal remains a major void in our understanding of the evolution of marine organisms (e.g., Caley et al. 1996; Benzie 1999; Moberg and Burton 2000, reviewed by Hedgecock 1994; Bossart and Prowell 1998; Grosberg and Cunningham 2001). Some data on the genetic structure of marine species suggests that populations are not at equilibrium (e.g., Hellberg 1995; Hellberg 1996; Benzie and Williams 1997; Marko 1998; Benzie 1999). Other possibilities that have been proposed to explain this common pattern of fine-scale differentiation include variation in the source populations from which larvae originate (e.g., Kordos and Burton 1993), sweepstakes reproductive success (e.g., Hedgecock 1994), and pre- or post-settlement natural selection (e.g., Johnson and Black 1984a; Schmidt and Rand 2001). In order to address such questions, fine-scale sampling of age-classes through time are necessary, but such studies are rare, and even the few that have been done to date were limited by the sensitivity of the genetic markers chosen (e.g., Kordos and Burton 1993; Edmands et al. 1996; Moberg and Burton 2000).

Regardless of the cause of fine-scale population structure, differences among populations at equilibrium can only be detected with a selectively-neutral genetic marker that mutates at a rate greater than the rate of migration among populations (reviewed by Burton 1994, 1996; Hedgecock 1994; Grosberg and Cunningham 2001). For species with pelagic larvae, the power to detect population structure and to identify the critical stages that influence structure and dynamics must ultimately be limited by the mutation rate of a marker (Grosberg and Cunningham 2001). As such, the development of ever more variable genetic markers has become a holy grail of population geneticists and resource managers alike. Microsatellites (also known as simple sequence repeats, or SSR loci) have rapidly become the genetic marker of choice for population biology studies because mutation rates (estimated to range from 10^{-3} - 10^{-5} per locus per generation) and levels of polymorphism (up to 60 alleles or more per locus) may be an order of magnitude or greater than other classes of genetic markers (e.g., allozymes, mtDNA sequence and RFLPs)(Fleischer 1996). Thus, microsatellites are the most sensitive molecular genetic tools currently available to characterize spatial and temporal patterns of genetic variation within and among populations (Goldstein and Schlötterer 1999; Luikart and England 1999; Chambers and MacAvoy 2000; Sunnucks 2000a). In addition, because microsatellite loci are amplified via the polymerase chain reaction (PCR), numerous loci can be assayed from very small amounts of tissue (such as individual larvae). These attributes of microsatellites potentially allow analysis of all life history stages, and make microsatellites the natural choice to address the causes of

fine-scale genetic structure in the face of extensive dispersal potential of marine invertebrates.

As a prelude to a comprehensive analysis of genetic structure in the anomuran crab *Petrolisthes cinctipes*, I isolated and developed primers for microsatellite markers. In this chapter, I describe this process, along with a subsequent analysis of the inheritance patterns of these loci.

Materials and methods

Study organism

The family Porcellanidae is one of the most diverse groups of nearshore decapod crustaceans, containing 23 genera. *Petrolisthes* is currently the largest genus, with over 100 described species (Haig 1960). Recent molecular work, however, suggests that the genus *Petrolisthes* contains two distinct clades, and that many of the currently recognized nominal species require further examination (Stillman and Reeb 2001). Two species of *Petrolisthes* occur in the Eastern North Pacific: *P. cinctipes*, and *P. eriomerus*. These two species, however, are not among those taxa in question (Stillman and Reeb 2001).

Petrolisthes cinctipes is a small suspension-feeding crab that gains most of its nutrition from planktonic diatoms (Knudsen 1964). It is restricted to a relative narrow vertical zone of the rocky intertidal (~ 0 to 1.8 m above MLLW)(Jensen and Armstrong

1991). Individuals of both sexes and all size classes, from newly settled postlarvae (~1.5 mm) to large adults (up to 20 mm carapace width), live together in large aggregations (up to 3933 individuals / m²) (Jensen 1990). Aggregations of these small crabs are found in virtually all areas of appropriate habitat ranging from roughly the Dixon Entrance, British Columbia (54°71' N) to Point Conception, CA (34°26' N).

P. cinctipes made an ideal choice to study the effect of larval dispersal on population structure for a variety of reasons. First, with such high densities, *P. cinctipes* is one of the most abundant decapod crustaceans in the Eastern North Pacific (Jensen 1990). *P. cinctipes* occurs sympatrically with the congener *P. eriomerus* along the North American coastline; however, *P. eriomerus* inhabits the low intertidal and subtidal zones, whereas *P. cinctipes* is restricted to areas of the middle and high intertidal (Jensen and Armstrong 1991). Unlike *P. eriomerus*, which could not survive transplants as little as 30 cm above its normal vertical limit (Jensen and Armstrong 1991), *P. cinctipes* can survive the mid- to high intertidal zone due to a variety of physiological adaptations to emergent stresses (Stillman and Somero 1996; Stillman 1998, 2000; Stillman and Somero 2000). *P. cinctipes* thrives in transplants below its normal tidal range, and is competitively dominant to *P. eriomerus* in laboratory experiments (Jensen and Armstrong 1991). These results suggest that its lower range limit is not set by competitive interactions with its congener. Instead, *P. cinctipes* appears to be restricted to a narrow band of the intertidal zone throughout its geographic range by substrate preference. Abrupt substrate changes in the mid-intertidal set the lower distribution because individual crabs actively avoid rocks resting on sand or silt

(Jensen and Armstrong 1991). Thus, *P. cinctipes* is both abundant and easy to locate across a broad geographic range.

Second although lab crosses have failed (Jensen 1990), females of *Petrolisthes cinctipes* brood their eggs through the initial stages of development, making it possible to collect a dam and her offspring from the field. Sexes are separate in *P. cinctipes*, but sexual dimorphism is slight. Females molt prior to copulation, and fertilization of the eggs is likely external after they are released, but before being attached to pleopodal setae for brooding (Nyblade 1987). Females typically produce broods from February through mid-April. Newly extruded broods may contain from 10 to 1580 embryos, each roughly 0.80 - 0.84 mm in diameter (Knudsen 1964). Some females may produce a second brood between May and September.

Third, identification of new recruits is straightforward in *P. cinctipes*. Feeding larvae spend several months in the plankton prior to molting into megalopae, but settle quickly in response to a waterborne cue associated with live conspecific adults (Jensen 1989, 1991a, b). In contrast, megalopae held in a bare container can delay settlement by as much as 45 days before dying in culture (Jensen 1991a). Megalopae of *P. cinctipes* do not molt at the time of settlement but rather lose the ability to swim due to degeneration of the pleopods; metamorphosis to the pigmented first instar juvenile takes place a week or more later (Jensen 1991a).

DNA extraction

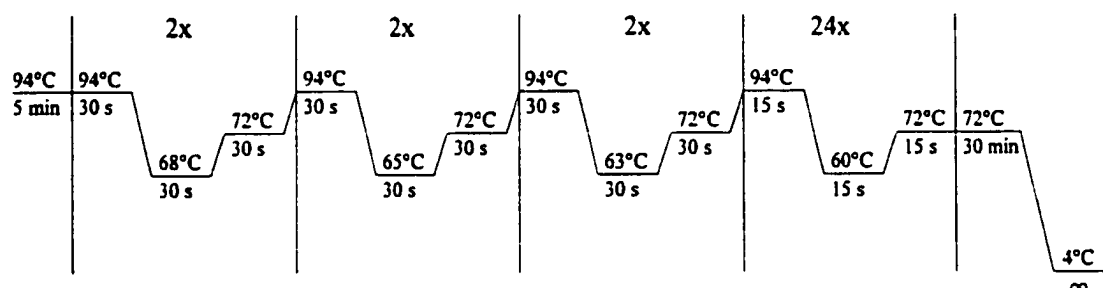
Petrolisthes readily self-amputate a cheliped (claw-bearing forearm) at a specialized break point in the limb as a defensive strategy when threatened (autotomy), making non-lethal sampling quite simple. Adult crabs (5 - 15 mm CW) were collected from Bodega Bay, CA and returned to the lab. Live crabs were harassed to stimulate self-amputation of the cheliped. After dissection of 3-5 mg of muscle tissue from the carpus (wrist), each cheliped was placed into a uniquely numbered vial filled with 95% ethanol, and stored at -40°C as a backup. DNA was extracted from the dissected muscle tissue using a modification of the 2x CTAB protocol outlined in Grosberg et al. (1996). Addition of 0.2% SDS and 2.0 mg / ml proteinase-K to the CTAB lysis buffer, together with constant agitation during an overnight incubation at 55°C, resulted in much more consistent amplification of samples than the standard 2X CTAB protocol (Toonen 1997, Appendix I). I discovered that minimizing the amount of carapace included in the extractions also consistently increased the quality and quantity of extracted DNA.

PCR amplification and electrophoresis

Primers for amplifying microsatellite loci used in this study were isolated and designed following the Microsatellites for Ecologists protocol presented in Appendix I (Toonen 1997). A total of 21 primer pairs flanking 17 simple sequence repeat (SSR or

microsatellite) loci were scored for variability using a sample of six individual crabs from each of four sites in northern California (Table 1).

PCR reaction mixes contained Perkin Elmer 10X Buffer II at 1X concentration, 2.5 mM MgCl₂, 0.1 mM dNTPs, 1X BSA, 0.5 mM of each the forward and reverse primer, 1 unit of Taq polymerase, and 1-50 ng of template DNA adjusted to a final volume of 10 µl (Toonen 1997, Appendix I). PCR amplifications were performed using a touchdown PCR protocol on a Perkin Elmer 9700 thermocycler as follows:



Loci Pc156s, Pc170s, Pc160b and Pc36L were amplified in a single PCR reaction using the same recipe and touchdown PCR protocol outlined above. Amplified PCR products were electrophoresed on an Applied Biosystems ABI 377 XL automated sequencer and scored with STRand (Hughes 1998), as outlined in Toonen and Hughes (in press, Appendix II).

STRand stores allelic data generated by the ABI 377XL sequencer in Microsoft's relational database program, Access, which allows direct export into the spreadsheet program MS Excel. Allelic data were then translated into Arlequin (Schneider et al. 2000) input format using the MS toolkit (Park 2001) macro for MS

Excel 2000. Summary statistics were all calculated with the Arlequin population genetics data analysis program.

Data analyses

The use of molecular genetic markers to portray population structure rests on at least four basic premises: 1) selective neutrality, 2) equilibrium between genetic drift and migration, 3) absence of null alleles, and 4) independent assortment of loci (reviewed by Grosberg and Cunningham 2001). Estimates of gene flow calculated from Wright's hierarchical F-statistics, or Weir and Cockerham's (1984) comparable estimators, should be robust to unmeasured selective forces if allelic frequencies do not deviate significantly from the Ewens sampling distribution (N. Barton, pers. comm.). Populations that have not reached equilibrium also generally show significant departures from the Ewens sampling distribution (Watterson 1986; Hutchison and Templeton 1999; Schneider et al. 2000). All loci were therefore tested for significant deviations from neutrality and equilibrium using Slatkin's (1994, 1996) exact test program, Enumerate (Slatkin 1997).

Null alleles are most easily detected by examining patterns of inheritance (e.g., Callen et al. 1993; Paetkau and Strobeck 1995; Pemberton et al. 1995). Given the obstacles to performing laboratory crosses of *P. cinctipes*, the only way to exclude loci with non-amplifying alleles is through empirical discovery. I excluded any loci for

which I was consistently unable to amplify a product from any individual within the initial survey of crabs (Table 1).

Testing conformity to Hardy-Weinberg (H-W) expectations is not straightforward for SSR loci, because with sample sizes less than thousands of individuals, the G-statistic does not follow a standard distribution when there are many alleles per locus (Edwards et al. 1992; Guo and Thompson 1992). Significant deviations from H-W expectations must instead be determined by permutation tests (Deka et al. 1991; Guo and Thompson 1992). I performed global tests for deviation from H-W equilibrium using Arlequin (Schneider et al. 2000). Tests for deviation from H-W expectations employ a Markov-chain method as described in Guo and Thompson (1992). I discovered that the 10,000 step procedure default in Arlequin (Schneider et al. 2000) was unreliable: significance of H-W proportions varied from 0.00 to 0.80 in 50 trials using 10,000 steps. Even the 100,000 step procedure recommended by Schneider et al. (2000) proved insufficient; I found that 500,000 steps were required to produce consistent results to two decimal places in 50 trials with these data. Therefore, I increased the length of the chain to 500,000 steps, with a 10,000 step dememorization for all tests of H-W proportions. Independent assortment of SSR loci was also tested in Arlequin, using a likelihood-ratio test (Slatkin and Excoffier 1996) against an empirical distribution that was likewise obtained by 500,000 permutations.

I calculated Weir and Cockerham's (1984) unbiased estimator of Wright's F_{ST} (θ) on a locus-by-locus basis using the Tools for Population Genetic Analyses (TFPGA) software (Miller 1997). Means and standard error of the means (SE) for each

estimate of the hierarchical F -statistics were obtained by jackknifing over loci; 95% confidence intervals (CI) were obtained from 10,000 bootstrapping replicates. TFPGA uses the method and terminology outlined in (Weir and Cockerham 1984; Weir 1996). Goodman's ρ (Goodman 1997), an unbiased estimator of Slatkin's R_{ST} (Slatkin 1995), was calculated using R_{ST} Calc 2.2 (Goodman 1997), by averaging variance components. Estimates of the effective number of migrants per generation (N_m) are also calculated by averaging variance components in R_{ST} Calc. The P -value for $\rho > 0$ is determined from 10,000 permutations of the data. Means, SE, and 95% CIs were obtained by 10,000 bootstrapping replicates.

Inheritance of microsatellite markers

To test for Mendelian segregation, loci were scored in 12 offspring from each of ten brooding females provided by S. Morgan and M. Donahue from the Bodega Bay Marine Lab, CA. Chelae were removed from each female and muscle tissue was dissected for DNA extraction as outlined above. Brooded embryos were extracted using the Nucleon PhytoPure plant DNA extraction kit (Amersham Biosciences), with the exception that embryos were not ground under liquid nitrogen as recommended for plant tissue. Instead embryos were removed individually and placed whole into extraction buffer, where each was ground with a clean pestle.

Multiple paternity of broods

Using only loci that did not show significant deviations from expected Mendelian ratios of codominant autosomal loci (Pc156s and Pc170s) allows me to infer the minimum number of males with which a female must have mated. Although the genotypes of the sires for each brood are unknown in these field-collected animals, the minimum number of mates can be estimated from the number of non-maternal alleles detected among the offspring. In several cases, the number of non-maternal alleles is increased by homozygous offspring, for which the sire is inferred to have shared at least one allele with the mother. Thus, if every female mated with only heterozygous males, the minimum number of mates required to generate the allelic distributions detected in each family is estimated as half the number of non-maternal alleles. For cases in which there were an odd number of non-maternal alleles detected among the brooded offspring, estimates of the minimum number of males required were rounded up (because it is not possible to mate with half a male, the minimum estimate must be rounded up). For example, if three non-maternal alleles are detected, the minimum estimated number of mated must be rounded up from 1.5 to two.

Given that loci Pc156s and Pc170s have 34 and 47 alleles, respectively, and high levels of heterozygosity (0.70 and 0.83, respectively), the chance for two individuals to share a particular two-locus genotype by chance is on the order of 10^{-4} to 10^{-5} (Avice 1994).

Results

Microsatellite variation

Sequencing of 200 positive clones from a genomic library isolated from *Petrolisthes cinctipes* yielded 24 clones with dinucleotide motifs (primarily (CA)_n (GA)_n or (GT)_n motifs), seven clones with trinucleotide repeats ((GTT)_n (GAT)_n (TGT)_n or (GGT)_n), one clone with a tetranucleotide repeat ((TTTA)_n), and 27 clones that contained mixed- or complex-repeat motifs. The remaining clones were either false positives (29), could not be sequenced reliably due to drop off or uncalled bases (16), or contained microsatellite-like sequences that did not contain any long strings of tandem repeats. Only 17 of the positive clones possessed sufficient flanking sequence around the repeat motif from which PCR primers could be developed (Table 1). I designed primer pairs for these clones, and each locus was scored in a sample of at least 24 crabs (Table 1). Primers that 1) produced unscorable multiband amplification products; 2) showed evidence of null alleles; 3) failed Slatkin's (1994, 1996) exact test for a fit to the Ewens sampling distribution for selectively neutral alleles; 4) did not amplify a product of the expected size; or 5) failed to amplify any product from DNA extracted from the original crab from which the library was developed were excluded from further analysis (Table 1). After this initial screening, only four of the 17 loci (Pc 156s, Pc170s, Pc160b and Pc36L) remained for subsequent analyses.

Hardy-Weinberg expectation and allelic frequency distributions

The remaining four microsatellite loci were then amplified in a multiplex reaction to genotype 4612 *P. cinctipes* for the studies presented in Chapters 2 and 3. All four markers turned out to be highly polymorphic, with 26 (Pc156s) to 55 (Pc160b) alleles per locus overall and expected heterozygosities ranging from 0.65 to 0.95 (Table 2).

Across all 4612 crabs sampled during these studies, there was a significant deficit of heterozygotes at two loci: Pc36L and Pc160b (Table 2). Given that many individuals failed to amplify a scorable product during PCR, these data are consistent with null alleles at these loci (Callen et al. 1993; Paetkau and Strobeck 1995; Pemberton et al. 1995). The allelic frequency distribution for each locus from all 4612 crabs genotyped during these studies is presented in Figure 1. Loci Pc160b and Pc36L did not differ significantly from the normal distribution expected under a single-stepwise mutational model, whereas Pc156s and Pc170s did (e.g., Shriver et al. 1993; Estoup et al. 1995; Kimmel and Chakraborty 1996; O'Connell et al. 1997). For both loci Pc170s and Pc156s, three alleles represent > 80% of the total allelic frequency detected among crabs throughout the species range (Fig. 1).

Estimates of F_{ST} and R_{ST}

Each of the two loci that exhibit a unimodal distribution of allelic variation provide a similar picture of the pattern of geographic structure among populations of *P. cinctipes*. Estimates of each component of Wright's hierarchical F -statistics (Wright 1951) are similar for loci Pc160b and Pc36L (Table 3). Using Weir and Cockerham's (1984) estimators, θ from these two loci was essentially zero, whereas f , the inbreeding coefficient, was significantly higher (Table 3). However the pattern of population structure revealed by these two loci is contradicted by the inference drawn from the two loci that show multimodal distributions of allelic frequencies (Pc156s and Pc170s, Fig. 1). As with estimates from the two unimodal loci, Weir and Cockerham's (1984) estimators were similar for the two multimodal loci. However, estimates of Wright's F -statistics derived from loci Pc156s and Pc170s provide essentially the opposite picture of population structure in *P. cinctipes* from loci Pc160b and Pc36L. Estimates of f , using only loci Pc156s and Pc170s, were essentially zero, whereas estimates of θ were significantly higher (Table 3).

As with the estimates of F_{ST} , estimates of R_{ST} and gene flow (Nm) were highly divergent between the two unimodal loci (Pc160b and Pc36L) and those that exhibit multimodal distributions (Pc156s and Pc170s). The overall estimate of ρ derived from the unimodal loci (Pc160b and Pc36L) was significantly greater than zero (Table 4), but also significantly lower than estimates derived from the two multimodal loci (Pc156s

and Pc170s). Consequently, the 95% CI of estimates of gene flow based on these two pairs of loci did not overlap (Table 4).

Patterns of inheritance

There was no evidence of deviation from expected Mendelian ratios of codominant autosomal loci for either locus Pc170s or locus Pc156s. All genotyped offspring contained at least one of the maternal alleles (Table 5), and across the 10 families examined, the ratio of maternal alleles detected among the offspring did not differ from 1:1 for either Pc170s ($df = 1$, $\chi^2 = 0.405$, $p > 0.05$) or Pc156s ($df = 1$, $\chi^2 = 0.348$, $p > 0.05$).

At the other two loci [Pc36L (Fig 2) and Pc160b (Fig. 3)], however, analysis of brooded embryos revealed bands which did not match either amplification product scored from the mother. For example, the apparently heterozygous offspring #5 (418, 420 bp) and #9 (416, 418 bp) each amplify two bands that are not shared with their apparently heterozygous mother (412, 414 bp) at locus Pc36L (Fig. 2). For presentation, only the five most extreme families are presented in the figures. The pattern of inheritance for these two loci remains unexplained, but is inconsistent with simple Mendelian patterns of allelic inheritance required for population genetic analyses.

Multiple paternity of broods

Despite the small sample size of brooded offspring genotyped from each female, eight of the ten families examined in this study show evidence for multiple paternity of broods (Table 5). Given that the minimum number of mates underestimates the contribution of any homozygous males or those that share an allele with the female, this estimate of the minimum number of males (2-3) is conservative. Furthermore, each female carried hundreds of brooded embryos, of which only 12 were genotyped. Even with this tiny sample, the genetic analysis revealed as many as five non-maternal alleles among these offspring (Table 5).

Discussion

Many studies of population genetic structure among marine species with planktonic larvae have ultimately raised more questions than they have answered (e.g., Johnson and Black 1982; Johnson and Black 1984a, b, reviewed by Hedgecock 1994; Grosberg and Cunningham 2001). Among the most perplexing of these questions is how fine-scale population structure persists despite the extensive dispersal potential of planktonic larvae. Many authors concur that addressing this question depends on the development of highly polymorphic molecular markers, coupled with finer-scale spatial and temporal sampling than currently used (e.g., Benzie 1999; Benzie 2000; Moberg and Burton 2000). Microsatellites are widely regarded as the most powerful

tool currently available for population genetic studies (Goldstein and Schlötterer 1999). Thus, prior to conducting an extensive survey of population structure in the anomuran crab *Petrolisthes cinctipes*, I developed microsatellite DNA markers (also known as simple sequence repeats, or SSR loci) for *Petrolisthes*, because none exist currently. This study is the first to isolate microsatellite markers in an anomuran crab.

My attempts to develop microsatellite markers for *P. cinctipes* revealed a number of widely acknowledged, but rarely addressed, problems with SSR loci. First, there is considerable variability among taxa in the ease with which microsatellite markers can be isolated (reviewed by Chambers and MacAvoy 2000; Nève and Meglécz 2000; Sunnucks 2000a, b). Marine invertebrates are proving to be among the more challenging for the development of molecular genetic markers. I have spoken with many researchers who have extensive experience isolating microsatellites from insects or vertebrates, but have been unsuccessful at isolating SSR loci from a marine invertebrate species. In my experience, the number of positive clones isolated from vertebrates and even terrestrial invertebrates (ants and termites) has been more than an order of magnitude higher than from a marine invertebrate done simultaneously (Toonen, 1997, Appendix I). Even among vertebrates that are considered very difficult (e.g., the Lemon Shark, *Negaprion brevirostris* – Feldheim et al. 2001), success with primer development is typically higher than that obtained from an ‘easy’ marine invertebrate such as *Petrolisthes*. This observation is supported by recent reviews (Amos 1999; Chambers and MacAvoy 2000), which conclude that microsatellite loci are significantly longer and more common in vertebrates than invertebrates (primarily

insects), and that among the vertebrates, ectothermic species (such as turtles) tend to have longer repeat arrays. However, such comparisons are heavily biased towards insects as the only invertebrate taxon, and studies with marine crustaceans such as *Petrolisthes cinctipes* (present study) and Dungeness crab, *Cancer magister* (R. Toonen and P. Jensen unpubl. data) suggest that this pattern may not hold true for marine invertebrates. For example, a complex dinucleotide isolated from *Petrolisthes* in this study was 137 repeats, and an imperfect tetranucleotide locus isolated from *C. magister* had as many as 90 repeats (P. Jensen, unpubl. data). Overall, the mean number of repeats for all alleles detected in the initial survey of only 24 crabs for each of the 17 loci presented in Table 1 was 37 ± 1.89 SE, which falls well above the confidence intervals for vertebrates presented in Amos (1999, Fig. 6.1). The data for microsatellite loci among marine invertebrate taxa remain sparse (e.g., Brockmann et al. 1994; Tam and Kornfield 1996; Brooker et al. 2000; Supungul et al. 2000; Xu et al. 2001), but allelic size ranges are frequently on the order of 50-100 bp and the number of alleles per locus typically averages more than 20 among published studies. For example, in a sample of 312 individuals of the giant tiger prawn, *Penaeus monodon*, Brooker et al. (2000) found 34 - 84 alleles per locus. These initial studies suggest that the generalization that microsatellites are longer and more diverse in vertebrates than invertebrates may be premature.

Statistics for analyzing population genetic data and making inferences about population structure make a number of assumptions, including 1) populations are in genetic equilibrium between genetic drift and migration; 2) loci are selectively neutral

and not closely linked to any gene under selection; 3) all alleles in the population are amplifiable (i.e., there are no null alleles); 4) loci assort independently of one another; and 5) alleles are inherited in a Mendelian fashion (reviewed by Goldstein and Schlötterer 1999; Luikart and England 1999; Chambers and MacAvoy 2000; Sunnucks 2000a, b). The utility of microsatellite markers for characterizing genetic structure, inferring levels of gene flow and describing mating patterns is only as good as the extent to which they conform to these assumptions. Despite a growing number of cautionary notes regarding microsatellite studies that fail to conform to these assumptions (e.g., Pemberton et al. 1995; Garza and Freimer 1996; Gullberg et al. 1997; Orti et al. 1997; Valsecchi et al. 1997; Clisson et al. 2000; Harris and Crandall 2000; Noor et al. 2001; Pogson et al. 2001), studies that test these assumptions for markers prior to analyses remain relatively uncommon. In part, this likely derives from the fact that testing these assumptions is costly and time-consuming. Also, exclusion of loci that do not meet these criteria forces researchers to develop and test more microsatellite loci, further increasing the time and cost of microsatellite-based studies. However, I believe that widespread testing of these assumptions is also hampered by greatly varying opinions on how serious a problem is posed by our incomplete understanding of mutational processes occurring at microsatellite loci and the impact of violations of the assumptions on population studies (reviewed by Goldstein and Schlötterer 1999; Luikart and England 1999; Chambers and MacAvoy 2000; Sunnucks 2000a, b). For the study of geographic population structure in *Petrolisthes cinctipes*, I decided to exclude any loci that deviated significantly from any of these assumptions.

After excluding loci that did not meet the assumptions of the anticipated population genetics analyses to be used the study, four loci remained. These four loci are highly polymorphic, with between 34 and 63 alleles each detected among the 4612 crabs genotyped during the course of this study. One pair of loci (Pc170s and Pc156s) amplified from almost every crab, but the allelic frequency distributions were multimodal, and do not match the expectation for a purely single-stepwise mutational model (Shriver et al. 1993). The other pair of loci (Pc160b and Pc36L) exhibit a unimodal distribution of allelic frequencies, more consistent with the expectations under a stepwise mutational model. However, > 10% of individuals assayed during the course of this study failed to amplify a scorable product. On a locus-by-locus basis, each of these two loci that exhibit a similar allelic frequency distribution yield consistent results in terms of estimates of population substructure, genetic distance and UPGMA clustering patterns (see Chapter 2). However, the results of the two loci that exhibit unimodal allelic frequency distributions (Pc160b and Pc36L) are at odds with the two that are multimodal (Pc156s and Pc170s), and the two pairs of loci give conflicting pictures of the genetic structure of populations of *P. cinctipes*. Inferences based on either F_{ST} or R_{ST} drawn from only loci Pc160b and Pc36L suggest a considerable degree of inbreeding, but little geographic population structure among populations. Loci Pc156s and Pc170s, on the other hand, suggest that there is no significant degree of inbreeding in *P. cinctipes*, but that there exists a significant amount of geographic population structure.

Microsatellites are generally accepted to be co-dominant markers that are inherited independently of one another in a Mendelian fashion (reviewed by Amos 1999; Goldstein and Schlötterer 1999; Luikart and England 1999; Chambers and MacAvoy 2000; Sunnucks 2000a). Although confirmation of Mendelian inheritance of microsatellite markers by pedigree analyses is reported in relatively few studies (e.g., Estoup et al. 1993; Garcia De Leon et al. 1995; Banks et al. 1996; Houlden et al. 1996; Eggleston-Stott et al. 1997; Aliah et al. 1999; Ardren et al. 1999; Khasa et al. 2000; McGoldrick et al. 2000), most uphold this generalization. In the case of *Petrolisthes cinctipes*, however, non-Mendelian patterns of inheritance among the offspring of brooding mothers provide a plausible explanation for the disagreement among the two pairs of loci. For both loci Pc160b and Pc36L, several families include heterozygous offspring that do not share any amplification product with their heterozygous mother. Obviously this result is inconsistent with simple Mendelian inheritance of “alleles” at these loci. The most likely explanation for this pattern is amplification of a multi-copy segment of DNA exhibiting intragenomic variation (e.g., Harris and Crandall 2000). I therefore excluded these two loci (Pc160b and Pc36L) from any subsequent analyses.

The inheritance data also provided novel information about the mating system of *Petrolisthes cinctipes*. Previous studies of paternity in several species of brachyuran crabs suggest that multiple paternity is rare despite the presence of sperm storage organs in females of these species (e.g., Diesel 1989, 1990; Koga et al. 1993; Jensen et al. 1996; Urbani et al. 1998). In contrast, using only the loci Pc156s and Pc170s, this study shows that females of *P. cinctipes* mate more than once and that at least some sperm

from different males fertilizes the brood of most females. However, it remains unclear whether multiple paternity involves sperm storage or repeated mating during egg extrusion. Regardless of the mechanism, multiple paternity of broods seems widespread in *Petrolisthes*. Even when only 12 offspring per brood were sampled, eight of 10 females carried broods from which the offspring genotypes exclude a single father.

This study highlights the importance of confirming Mendelian patterns of inheritance for loci prior to use in population studies. Studies that perform a systematic screening of microsatellite loci before use in population studies remain uncommon, and even after my initial screening, two of the four loci chosen for the study of *Petrolisthes cinctipes* proved to be inappropriate for such analyses. If the four loci that remained after the initial screening had not provided conflicting patterns of genetic structure, I may not have tested the pattern of inheritance for these loci. Fortunately, the majority of published studies that assess patterns of inheritance generally find that microsatellites act as co-dominant markers that are inherited independently of one another in a Mendelian fashion (e.g., Estoup et al. 1993; Garcia De Leon et al. 1995; Banks et al. 1996; Aliah et al. 1999; Ardren et al. 1999; Khasa et al. 2000; McGoldrick et al. 2000). However, without such confirmation, inferences drawn from patterns of allelic frequency and distribution, especially when loci suggest differing patterns of genetic structure, should be regarded with caution.

Table 1. Microsatellite loci characterized from a genomic library of the porcelain crab *Petrolisthes cinctipes*. Six crabs each from four sites (Bodega, Salt Point, Pt. Arena & Crescent City) were scored with each primer pair to determine whether the locus was suitable for inclusion in the study. Repeat motifs were determined from the sequence of the cloned allele, and '...' in the motif denotes a non-repeat segment of not more than 15 bp. Locus names all begin with the species abbreviation (Pc), followed by the bacterial clone number from which the microsatellite sequences were obtained. An 's' or 'L' following the clone number signifies a locus for which there was sufficient flanking sequence to design both a short (s) and a long (L) amplification product primer pair, to provide flexibility in designing overlapping multiplex PCR reactions. Loci labeled with a 'b' denotes those for which the primer was redesigned in an effort to eliminate null alleles or produce a more easily scorable amplification product. It is important to note that the number of alleles listed here reflects only those detected in the initial sample of 24 crabs. Loci in the excluded column denoted with 'untested' are newly developed to increase the number of markers for subsequent studies.

Locus name	Primer Sequence (5' → 3')	Cloned repeat motif	Number of alleles	Excluded? (If so, reason)
Pc33	F: gaa agt gat tgc aaa tgc agt g R: aca ttc acg ccc aat act cc	(AT) ₅ ... (GT) ₁₄	8	Yes (failed neutrality test)
Pc34	F: cac aca cac gcg agc ttt tat agg R: cat gaa ggg aat aca ggc cga g	(TA) ₄ (CA) ₄ (CG) ₂ ... (AC) ₂₀	10	Yes (product too large)
Pc35	F: ttc aca ctg gta aca cat tta tag gag R: atc atc tea cca gtc ttc ttc g	(GTT) ₁₂	14	Yes (apparent null alleles)
Pc35b	F: cac act ggt aac aca ttt ata gga g R: acc agt ctt gtt cga cca tt		?	Untested
Pc36s	F: atg ttg acc act gca tgg cg R: cgc gcc tct cac ccc taa c	(TTTAA) ₂ TTTA (TTAA) ₂ (TTAT) ₃ TTGT TTAT C (TATT) ₃ TTTT	2	Yes (Used Pc36L instead)
Pc36L	F: atg ttg acc act gca tgg cg R: ggg caa gtt tgg aag agg tgg		8	No

Pc37	F: gtc cat cga cga agc aga cg R: ctg acg gag acg gga ctg g	(GGTGAT) ₃ (GAT) ₂₁	5	Yes (required 50°C annealing)
Pc102	F: tag gcg acg ata cag gga gg R: cag ctt gtg aga ggc tgg c	(GGT) ₃ GT (GA) ₃₄ (AC) ₄	?	Yes (Unscorable, multibanded)
Pc102b	F: ggc gac gat aca ggg agg tc R: gct tgt gag agg ctg gct tg		?	Untested
Pc103s	F: gga agt ggt ggt ggt cat tgg R: gcc taa ttt ctg ttc ccc tca ctt c	(AG) ₈₇	7	Yes (apparent null alleles)
Pc103L	F: gga agt ggt ggt ggt cat tgg R: ggt cag gcc tgg tat cct tgg		12	Yes (apparent null alleles)
Pc108	F: gtt tgg tgt tta aaa tgg tg R: tta aat cta tat ata atc cac tgt gtc	274 bp Complex: (CT) _n (TG) _n (AG) _n	0	Yes (No amplification of control)
Pc110	F: tga gtc ctt caa ata ttg cca cg R: gcc acg aaa caa aca agc tcc	(CA) ₂₀	9	Yes (Unscorable, multibanded)
Pc110b	F: cct cct gac aat acc acg aaa caa g R: ctc ctg act atg cca cga aac aaa c		?	Untested
Pc112	F: aag gct ctg tcc cca ccc tc R: ggg aaa tgg gtt gtg cat tag g	(AG) ₁₄ AA (AG) ₆ TG AG GG (GA) ₂	7	Yes (Unscorable, multibanded)
Pc112b	F: tag aag gct ctg tcc cca ccc R: ggg aaa tgg gtt gtg cat tag g		?	Untested
Pc156s	F: ttg gct ttg aag acc ctg tgg R: cgg ggg atc att gct ttg tc	(TG) ₂₆ TA(TG) ₁₁	8	No

Pc156L	F: cag agt gag gtt tgc cgt acc c R: cgg ggg atc att gct ttg tc		13	Yes (Used Pc156s instead)
Pc160	F: tgc tgc tcc gtg tct taa ccc R: agg tga gtg cag gag tgg gg	(TG) ₁₆	9	Yes (product too large)
Pc160b	F: gct gtt ttc tag gca tgg ttt gg R: agg tga gtg cag gag tgg gg		12	No
Pc168s	F: aag gtg taa gtt aaa tgg gg R: gat caa aag gtt acc ctg ac	(GTT) ₄ GTC CTG (GTT) ₅ GT	0	Yes (No amplification of control)
Pc168L	F: aaa tat cga att ttc ggt tg R: gat caa aag gtt acc ctg ac		0	Yes (No amplification of control)
Pc170s	F: tgg cgg ttg ctg ttg ttg tc R: ggc acc agt cat tcc cag ttg	TCTAGA (TGT) ₆ TTGTAA TCTAGA (TGT) ₆ ... (TG) ₄	7	No
Pc170L	F: tgc ggt ttg ata tgt cga cgc R: ggc acc agt cat tcc cag ttg		7	Yes (Used Pc170s instead)
Pc188	F: ggg ata acg aaa gga aaa tag gac ag R: cgg ggg atc tca act cac tg	(GT) ₃ (GA) ₃₈	?	Untested
Pc191s	F: ttc agg ttc ctt ccc cct cc R: gat gcc gga gtg cat agc ag	(GA) ₂₄	6	Yes (failed neutrality test)
Pc191L	F: ttc agg ttc ctt ccc cct cc R: acg ata ccc gca tac ccc ac		6	Yes (failed neutrality test)
Pc192b	F: ctg cca acc cca caa ggt tc R: ctc tct cac cga cgg caa gg	(TC) ₂ (GA) ₂₅ GT GA GT	?	Yes (Unscorable, multibanded)

Table 2. Summary of the four polymorphic microsatellite loci isolated from *Petrolishes cinclipes* originally used in this study. Shown are the forward and reverse primer sequences (5' → 3') and the fluorescent dye with which each locus was labelled, motif of the original cloned microsatellite sequence, number of alleles observed in the sample of 4612 crabs used in this study, the number of crabs for which each locus was scored (N), the observed and expected heterozygosities ($\pm$ SE) for each locus across all populations studied, and the probability value ($\pm$ SE) for conformity to Hardy-Weinberg equilibrium (see text).

Locus	Primer Sequence	Cloned repeat motif	No. of alleles	N	Heterozygosity		H-W <i>P</i> -value
					Observed	Expected	
Pc156s	HEX - ttg gct ttg aag acc ctg tgg cgg ggg atc att gct ttg tc	(TG) ₂₆ TA(TG) ₁₁	34	4612	0.63	0.70	0.06 ($\pm$ 0.005)
Pc160b	NED - gct gtt ttc tag gca tgg tt gg agg tga gtg cag gag tgg gg	(TG) ₁₆	63	4328	0.78	0.95	0.01 ($\pm$ 0.005)
Pc170s	6FAM - tgg ccg ttg ctg ttg ttg tc ggc acc agt cat tcc cag ttg	TCTAGA (TGT) ₆ TTGTAA TCTAGA (TGT) ₆ ... (TG) ₄	47	4598	0.77	0.83	0.07 ($\pm$ 0.004)
Pc36L	6FAM - atg ttg acc act gca tgg cg ggg caa gtt tgg aag agg tgg	(TTTAA) ₂ TTAA (TTAA) ₂ (TTAT) ₃ TTGT TTAT C (TATT) ₃ TTTT	45	3743	0.60	0.85	0.04 ($\pm$ 0.009)

Table 3. Weir and Cockerham (1984) estimates of Wright's F -statistics calculated separately for each locus scored from *Petrolisthes cinctipes* collected from 12 geographic locations throughout its range. Means, standard error (SE) and 95% confidence intervals were determined by jackknifing and bootstrapping (see text). Estimates of F -statistics calculated for pairs of loci pooled by allelic frequency distribution are also presented.

Locus	$F = F_{IT}$ (within total)	$\theta = F_{ST}$ (among sites)	$f = F_{IS}$ (within sites)
Pc156s	0.07	0.10	- 0.03
Pc160b	0.17	0.002	0.17
Pc170s	0.05	0.12	- 0.07
Pc36L	0.29	0.01	0.29
Mean	0.15	0.05	0.11
SE	0.01	0.01	0.02
95% C.I.	0.06 – 0.24	0.004 – 0.11	- 0.05 – 0.23
Pairs of loci grouped by allelic frequency distribution			
Pc156s & Pc170s	0.06	0.10	- 0.04
95% C.I.	0.04 – 0.07	0.08 – 0.11	- 0.02 – - 0.08
Pc160b & Pc36L	0.24	0.005	0.23
95% C.I.	0.18 – 0.30	0.001 – 0.01	0.16 – 0.29

Table 4. Goodman's (1997) ρ and estimate of gene flow (Nm), calculated separately for each locus scored from all *Petrolisthes cinctipes* that amplified a scorable product at all four loci. Means, SE, 95% CIs and P-values (not shown, all estimates of ρ and Nm were significant at $\alpha = 0.01$) were obtained by permutation and bootstrapping (see text).

Locus	ρ	Nm
Pc156s	0.07	1.95
Pc160b	0.009	11.45
Pc170s	0.04	5.20
Pc36L	0.006	12.98
Mean	0.03	7.07
SE	0.001	0.02
95% C.I.	0.02 – 0.05	5.69 – 8.59
Pairs of loci grouped by allelic frequency distribution		
Pc156s & Pc170s	0.06	4.49
95% C.I.	0.04 – 0.07	3.34 – 5.41
Pc160b & Pc36L	0.008	32.28
95% C.I.	0.006 – 0.010	13.27 – 38.73

Table 5. Summary of allelic inheritance data from 12 brooded offspring from each of ten gravid females of *Petrolisthes cinctipes*. In each case, mother refers to the dam from which brooded embryos were collected. Alleles are identified by the length of the amplification fragment (in bp). The minimum number of males with which each female mated is estimated by one-half the maximum number of non-maternal alleles observed at either locus among the offspring (see text).

Family	Locus 156s		Locus 170s		min # of males
	mother	offspring	mother	offspring	
1	76 120	76 78 82 86 100 120	126 128	126 128 134	2
2	76 120	76 120	116 126	90 116 126 128 142	2
3	76 120	76 78 120	126 128	90 100 116 126 128 142	2
4	76 120	76 78 90 94 120	126 142	96 126 128 134 142 144	2
5	76 120	76 84 86 98 106 120	126 128	90 112 126 128 140 142 200	3
6	76 120	76 78 82 108 120	116 128	116 124 128 140 142	2
7	76 120	76 78 88 120	126 128	116 126 128 142	1
8	76 120	76 94 98 106 120	128 128	98 126 128 140 142	2
9	76 76	76 78 92	126 128	92 126 128	1
10	76 94	76 78 86 94 120	126 128	126 128 132 134	2

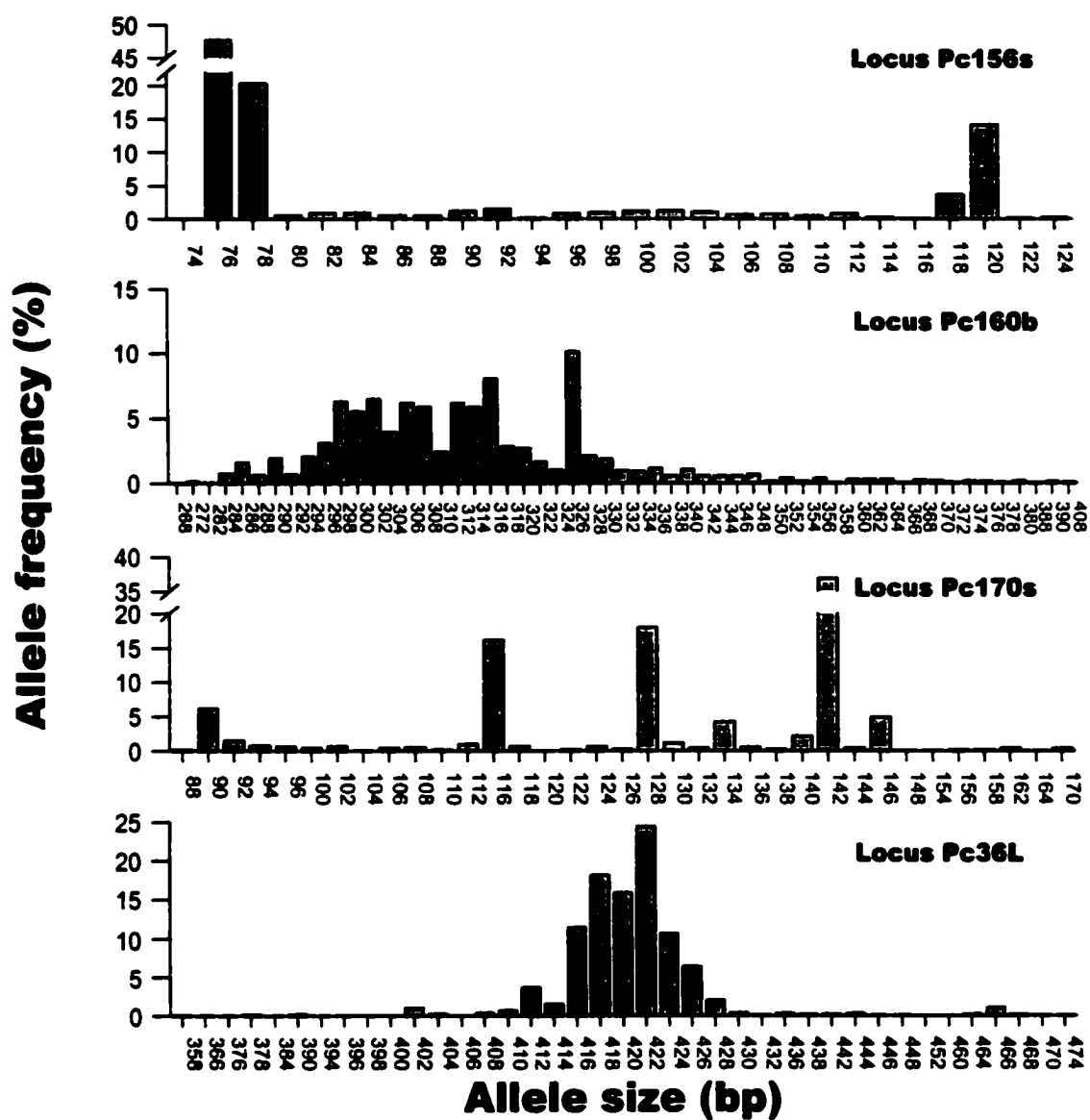


Figure 1. Allele frequency distributions of the four polymorphic microsatellite loci amplified from 4612 *Petrolisthes cinctipes*.

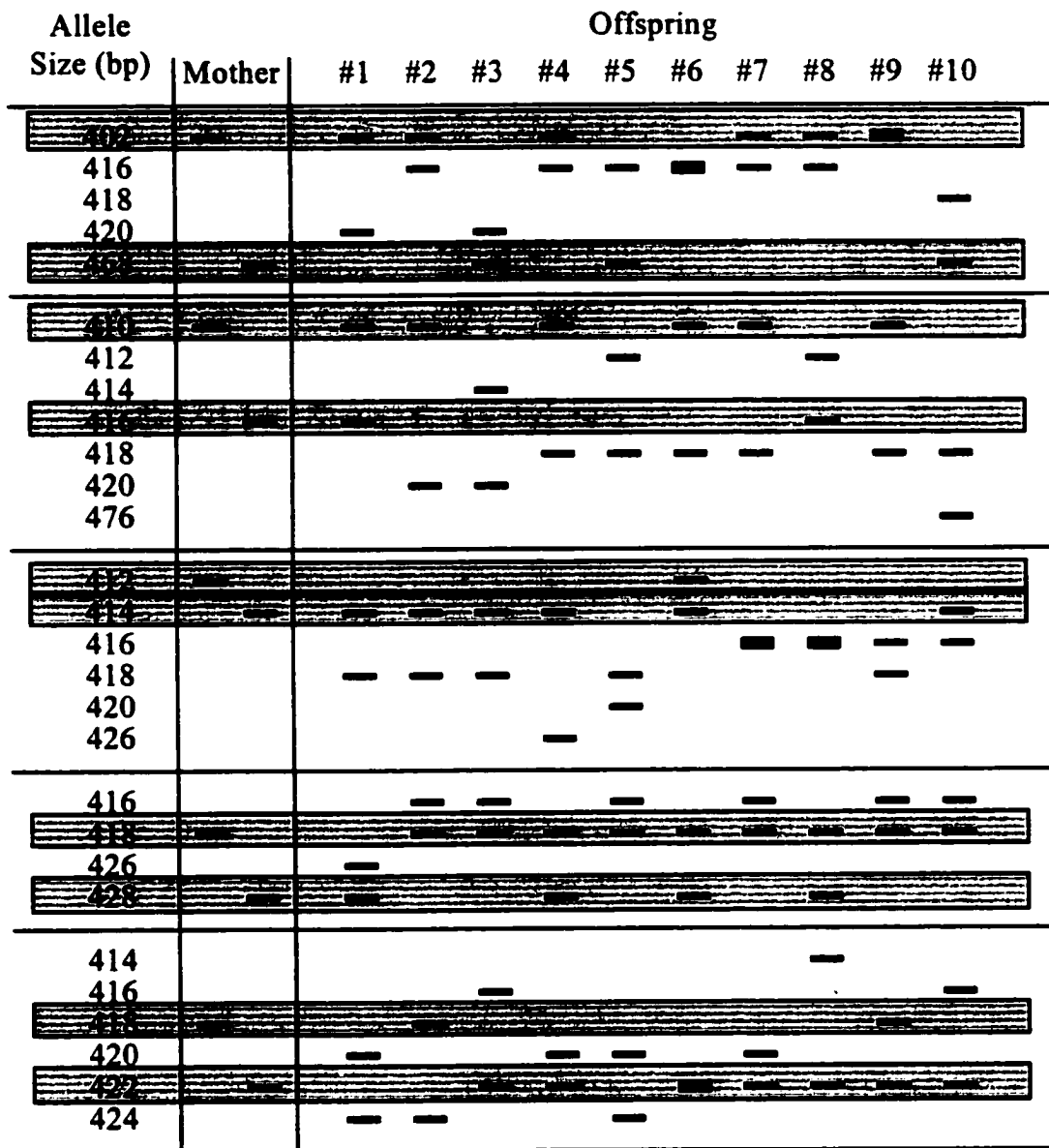


Figure 2. Schematic representation of maternal genotype for locus Pc36L compared to the genotypes from 10 of her brooded embryos. Size of amplification products is given in base pairs (bp) and apparent homozygotes are represented by thickened lines. Maternal “alleles” are highlighted with grey bars across all offspring amplification products.

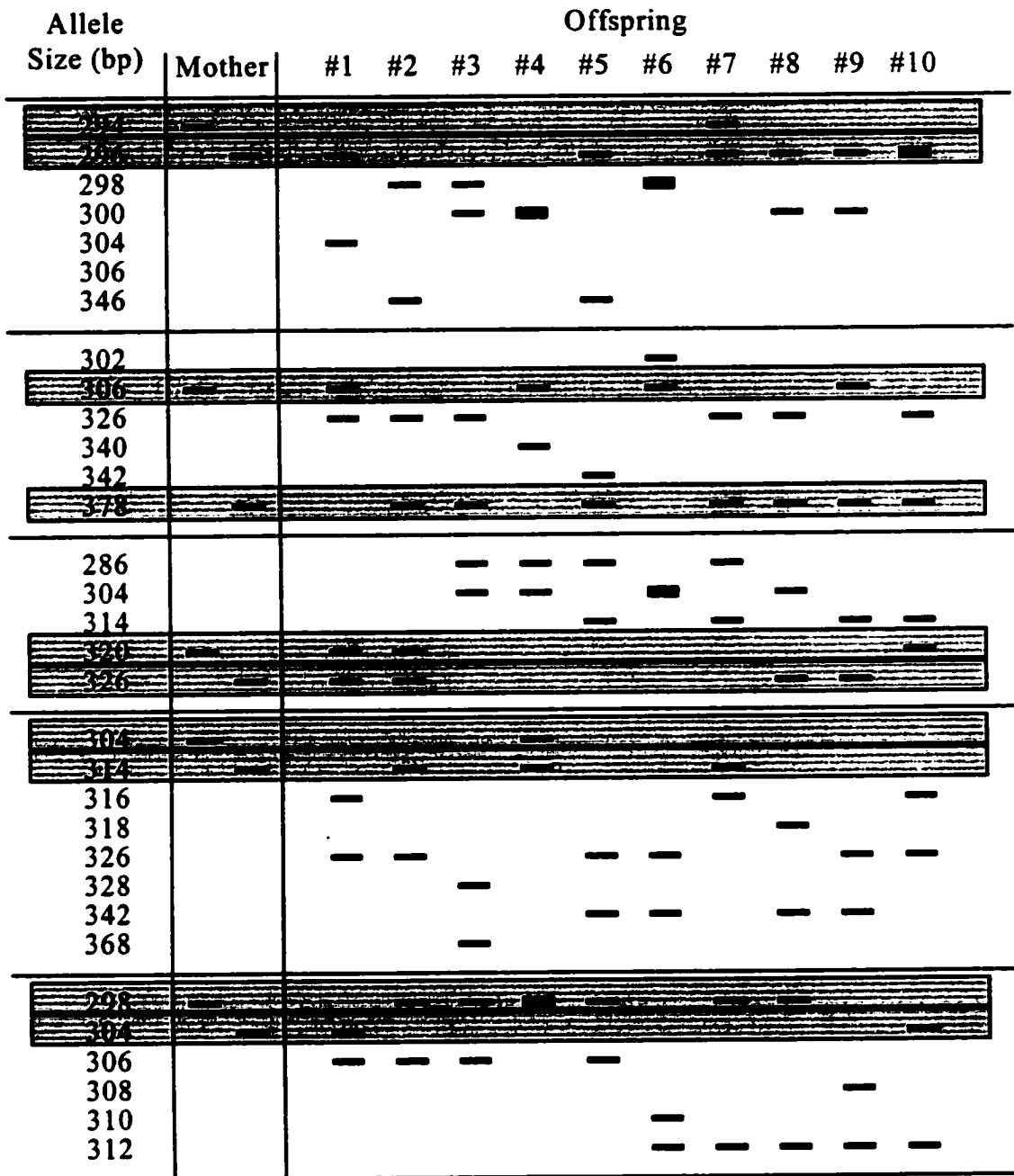


Figure 3. Schematic representation of maternal genotype for locus Pc160b compared to the genotypes from 10 of her brooded embryos. Size of amplification products is given in base pairs (bp) and apparent homozygotes are represented by thickened lines. Maternal "alleles" are highlighted with grey bars across all offspring amplification products.

Literature Cited

- Aliah, R. S., M. Takagi, S. Dong, C. T. Teoh, and N. Taniguchi. 1999. Isolation and inheritance of microsatellite markers in the common carp *Cyprinus carpio*. *Fisheries Science* 65:235-239.
- Amos, W. A. 1999. A comparative approach to the study on microsatellite evolution. Pp. 66-79 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford.
- Ardren, W. R., S. Borer, F. Thrower, J. E. Joyce, and A. R. Kapuscinski. 1999. Inheritance of 12 microsatellite loci in *Oncorhynchus mykiss*. *Journal of Heredity* 90:529-536.
- Awise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. 511 pp. Chapman and Hall, New York, NY.
- Banks, M. A., B. A. Baldwin, and D. Hedgecock. 1996. Research on chinook salmon (*Oncorhynchus tshawytscha*) stock structure using microsatellite DNA. *Bulletin of National Research Institute of Aquaculture, Supplement* 2:5-9.
- Benzie, J. A. H. 1999. Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoologist* 39:131-145.
- Benzie, J. A. H. 2000. The detection of spatial variation in widespread marine species: Methods and bias in the analysis of population structure in the crown of thorns starfish (Echinodermata: Asteroidea). *Hydrobiologia* 420:1-14.

- Benzie, J. A. H., and S. T. Williams. 1997. Genetic structure of giant clam (*Tridacna maxima*) populations in the West Pacific is not consistent with dispersal by present-day ocean currents. *Evolution* 51:768-783.
- Berger, E. M. 1973. Gene-enzyme variation in three sympatric species of *Littorina*. *Biological Bulletin* 145:83-90.
- Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 74:21-45.
- Bossart, J. L., and D. P. Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution* 13:202-206.
- Brockmann, H. J., T. Colson, and W. Potts. 1994. Sperm competition in horseshoe crabs (*Limulus polyphemus*). *Behavioral Ecology and Sociobiology* 35:153-160.
- Brooker, A. L., J. A. H. Benzie, D. Blair, and J.-J. Versini. 2000. Population structure of the giant tiger prawn *Penaeus monodon* in Australian waters, determined using microsatellite markers. *Marine Biology* 136:149-157.
- Burton, R. S. 1983. Protein polymorphisms and genetic differentiation of marine invertebrate populations. *Marine Biology Letters* 4:193-206.
- Burton, R. S. 1994. Inferring the genetic structure of marine populations: a case study comparing allozyme and DNA sequence data. *CalCOFI Report* 35:52-60.
- Burton, R. S. 1996. Molecular tools in marine ecology. *Journal of Experimental Marine Biology and Ecology* 200:85-101.

- Caley, M. J., M. H. Carr, M. A. Hixon, T. P. Hughes, G. P. Jones, and B. A. Menge. 1996. Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics* 1996:477-500.
- Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, J. C. Mulley, and G. R. Sutherland. 1993. Incidence and Origin of Null Alleles in the Acn Microsatellite Markers. *American Journal of Human Genetics* 52:922-927.
- Chambers, G. K., and E. S. MacAvoy. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology, Part B* 126:455 - 476.
- Clisson, I., M. Lathuilliere, and B. Crouau-Roy. 2000. Conservation and evolution of microsatellite loci in primate taxa. *American Journal of Primatology* 50:205-214.
- Deka, R., R. Chakraborty, and R. E. Ferrell. 1991. A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics* 11:83 - 92.
- Diesel, R. 1989. Structure and function of the reproductive system of the symbiotic spider crab *Inachus phalangium* (Decapoda: Majidae): Observations on sperm transfer, sperm storage, and spawning. *Journal of Crustacean Biology* 9:266-277.
- Diesel, R. 1990. Sperm competition and reproductive success in decapod *Inachus phalangium* (Majidae): A male ghost spider crab that seals off rivals' sperm. *Journal of Zoology* 220:213-224.

- Edmands, S., P. E. Moberg, and R. S. Burton. 1996. Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. *Marine Biology* 126:443-450.
- Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241 - 253.
- Eggleston-Stott, M. L., A. Delvalle, S. Dileanis, E. Wictum, and A. T. Bowling. 1997. A single base transversion in the flanking region of an equine microsatellite locus affects amplification of one allele. *Animal Genetics* 28:438-440.
- Estoup, A., L. Garnery, M. Solignac, and J. M. Cornuet. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* 140:679-695.
- Estoup, A., P. Presa, F. Krieg, D. Vaiman, and R. Guyomard. 1993. (CT)-n and (GT)-n microsatellites: A new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* 71:488-496.
- Feldheim, K. A., S. H. Gruber, and M. V. Ashley. 2001. Population genetic structure of the lemon shark (*Negaprion brevirostris*) in the western Atlantic: DNA microsatellite variation. *Molecular Ecology* 10:295 - 303.
- Fleischer, R. C. 1996. Application of molecular methods to the assessment of genetic mating systems in vertebrates. Pp. 133-161 in J. D. Ferraris and S. R. Palumbi, eds. *Molecular Zoology: advances, strategies, and protocols*. Wiley and Sons, Inc., New York, NY.

- Futuyma, D. J. 1998. *Evolutionary Biology*. 763 pp. Sinauer Associates, Inc., Sunderland, MA.
- Garcia De Leon, F. J., J. F. Dallas, B. Chatain, M. Canonne, J. J. Versini, and F. Bonhomme. 1995. Development and use of microsatellite markers in sea bass, *Dicentrarchus labrax* (Linnaeus, 1758) (Perciformes: Serranidae). *Molecular Marine Biology and Biotechnology* 4:62-68.
- Garza, J. C., and N. B. Freimer. 1996. Homoplasmy for size at microsatellite loci in humans and chimpanzees. *Genome Research* 6:211-217.
- Goldstein, D. B., and C. Schlötterer. 1999. *Microsatellites: Evolution and Applications*. 352 pp. Oxford University Press, Oxford.
- Goodman, S. J. 1997. R-ST Calc: A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6:881-885.
- Grosberg, R. K., and C. W. Cunningham. 2001. Genetic Structure in the Sea: From Populations to Communities. Pp. 61-84 in M. D. Bertness, S. D. Gaines and M. E. Hay, eds. *Marine Community Ecology*. Sinauer Associates, Inc., Sunderland, MA.
- Grosberg, R. K., D. R. Levitan, and B. B. Cameron. 1996. Characterization of genetic structure and genealogies using RAPD-PCR markers: A random primer for the novice and nervous. Pp. 67-100 in Ferraris, J. D. and S. R. Palumbi, eds. *Molecular zoology: Advances, strategies, and protocols*. Wiley-Liss, Inc., New York, NY.

- Gullberg, A., H. Tegelstrom, and M. Olsson. 1997. Microsatellites in the sand lizard (*Lacerta agilis*): Description, variation, inheritance, and applicability. *Biochemical Genetics* 35:281-295.
- Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48:361-372.
- Haig, J. 1960. The Porcellanidae (Crustacea: Anomura) of the Eastern Pacific. *Alan Hancock Pacific Expeditions* 24:1-440.
- Harris, D. J., and K. A. Crandall. 2000. Intragenomic Variation Within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies. *Molecular Biology and Evolution* 17:284-291.
- Hedgecock, D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science* 39:550-564.
- Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? Pp. 122-134 in A.R. Beaumont, ed. *Genetics and Evolution of Aquatic Organisms*. Chapman and Hall, London, UK
- Hellberg, M. E. 1995. Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: Equilibrium and nonequilibrium at different spatial scales. *Marine Biology* 123:573-581.
- Hellberg, M. E. 1996. Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. *Evolution* 50:1167-1175.

- Houlden, B. A., P. England, and W. B. Sherwin. 1996. Paternity exclusion in koalas using hypervariable microsatellites. *Journal of Heredity* 87:149-152.
- Hughes, S. 1998. STRand Nucleic Acid Analysis Software. Regents of the University of California, Davis. Available at <http://www.vgl.ucdavis.edu/STRand>
- Hutchison, D. W., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898-1914.
- Jensen, G. C. 1989. Gregarious settlement by megalopae of the porcelain crabs *Petrolisthes cinctipes* (Randall) and *P. eriomerus* Stimpson. *Journal of Experimental Marine Biology and Ecology* 131:223-231.
- Jensen, G. C. 1990. Intertidal zonation of porcelain crabs: Resource partitioning and the role of selective settlement. Ph.D. Dissertation, Zoology, University of Washington, Seattle.
- Jensen, G. C. 1991a. Competency, settling behavior, and postsettlement aggregation by porcelain crabs megalopae (Anomura: Porcellanidae). *Journal of Experimental Marine Biology and Ecology* 153:49 - 61.
- Jensen, G. C. 1991b. Gregarious settlement by porcelain crabs, *Petrolisthes* (Anomura: Porcellanidae). *Memoirs of the Queensland Museum* 31:390.
- Jensen, G. C., and D. A. Armstrong. 1991. Intertidal zonation among congeners: Factors regulating distribution of porcelain crabs *Petrolisthes* spp. (Anomura: Porcellanidae). *Marine Ecology Progress Series* 73:47 - 60.

- Jensen, P. C., J. M. Orensanz, and D. A. Armstrong. 1996. Structure of the female reproductive tract in the Dungeness crab (*Cancer magister*) and implications for the mating system. *Biological Bulletin* 190:336-349.
- Johnson, M. P., and R. Black. 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* 70:157-164.
- Johnson, M. S., and R. Black. 1984a. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intretidal limpet. *Evolution* 38:1371-1383.
- Johnson, M. S., and R. Black. 1984b. The Wahlund effect and the geographical scale of variation in the intertidal limpet *Siphonaria* sp. *Marine Biology* 79:295-302.
- Karl, S. A., and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. *Science* 256:100 - 102.
- Khasa, P. D., C. H. Newton, M. H. Rahman, B. Jaquish, and B. P. Dancik. 2000. Isolation, characterizaton, and inheritance of microsatellite loci in alpine larch and western larch. *Genome* 43:439-448.
- Kimmel, M., and R. Chakraborty. 1996. Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theoretical Population Biology* 50:345-367.
- Knudsen, J. W. 1964. Observations of the reproductive cycles and ecology of the common Brachyura and crablike Anomura of Puget Sound, Washington. *Pacific Science* 18:3-33.

- Koga, T., Y. Henmi, and M. Murai. 1993. Sperm competition and the assurance of underground copulation in the sand-bubbler crab *Scopimera globosa* (Brachyura: Ocypodidae). *Journal of Crustacean Biology* 13:134-137.
- Kordos, L. M., and R. S. Burton. 1993. Genetic differentiation of Texas Gulf Coast populations of the blue crab *Callinectes sapidus*. *Marine Biology* 117:227-233.
- Kyle, C. J., and E. G. Boulding. 2000. Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Marine Biology* 137:835-845.
- Lacson, J. M., V. M. Riccardi, S. W. Calhoun, and D. C. Morizot. 1989. Genetic differentiation of bicolor damselfish (*Eupomacentrus partitus*) populations in the Florida Keys (USA). *Marine Biology* 103:445-452.
- Luikart, G., and P. R. England. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* 14:253 - 256.
- Marko, P. 1998. Historical allopatry and the biogeography of speciation in the Prosobranch snail genus *Nucella*. *Evolution* 52:757-774.
- McGoldrick, D. J., D. Hedgecock, L. J. English, P. Baoprasertkul, and R. D. Ward. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): Selection and null alleles. *Journal of Shellfish Research* 19:779-788.
- Miller, M. P. 1997. Tools for Population Genetic Analyses (TFPGA): A Windows program for the analysis of allozyme and molecular population genetic data. Available at <http://www.public.asu.edu/~mmille8/tfpga.htm>

- Moberg, P. E., and R. S. Burton. 2000. Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. *Marine Biology* 136:773-784.
- Nève, G., and E. Megléc. 2000. Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution* 15:376 - 377.
- Noor, M. A. F., R. M. Kliman, and C. A. Machado. 2001. Evolutionary history of microsatellites in the obscura group of *Drosophila*. *Molecular Biology and Evolution* 18:551-556.
- Nyblade, C. F. 1987. Phylum or Subphylum Crustacea, Class Malacostraca, Order Decapoda, Anomura. Pp. 441 - 450 in M. F. Strathmann, ed. *Reproduction and development of marine invertebrates of the Northern Pacific Coast: data and methods or the study of eggs, embryos, and larvae*. University of Washington Press, Seattle.
- O'Connell, M., R. G. Danzmann, J. M. Cornuet, J. M. Wright, and M. M. Ferguson. 1997. Differentiation of rainbow trout (*Oncorhynchus mykiss*) populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. *Canadian Journal of Fisheries and Aquatic Sciences* 54:1391-1399.
- Orti, G., D. E. Pearse, and J. C. Avise. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the United States of America* 94:10745-10749.
- Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology* 4:519-520.

- Palumbi, S. R. 1995. Using genetics as an indirect estimator of larval dispersal. Pp. 369-387 in L. R. McEdward, ed. *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, FL.
- Park, S. 2001. *Microsatellite Toolkit for MS Excel*. Smurfit Institute of Genetics, Trinity College, Dublin. Available at <http://acer.gen.tcd.ie/~sdepark/ms-toolkit/>
- Pemberton, J. M., J. Slate, D. R. Bancroft, and J. A. Barrett. 1995. Nonamplifying alleles at microsatellite loci - A caution for parentage and population studies. *Molecular Ecology* 4:249-252.
- Pogson, G. H., C. T. Taggart, K. A. Mesa, and R. G. Boutilier. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* 55:131-146.
- Schmidt, P. S., and D. M. Rand. 2001. Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of *Mpi* genotypes. *Evolution* 55:1336-1344.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva. Available at <http://anthro.unige.ch/arlequin>
- Shriver, M. D., L. Jin, R. Chakraborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the SMM: A computer simulation approach. *Genetics* 134:983-993.
- Slatkin, M. 1994. An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* 64:71-74.

- Slatkin, M. 1995. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics* 139:457-462.
- Slatkin, M. 1996. A correction to the exact test based on the Ewens sampling distribution. *Genetical Research* 68:259 - 260.
- Slatkin, M. 1997. Enumerate. Department of Integrative Biology, University of California, Berkeley. Available at http://ib.berkeley.edu/labs/slatkin/monty/Ewens_exact.program
- Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* 76:377 - 383.
- Stillman, J. H. 1998. A comparative analysis of morphological, physiological, and biochemical adaptation to abiotic stress in intertidal porcelain crabs, genus *Petrolisthes*. Ph.D. Dissertation, Zoology. Oregon State University, Corvallis, Corvallis.
- Stillman, J. H. 2000. The evolutionary history and adaptive significance of secondary respiratory structures in intertidal crabs: Relationships with body size and vertical distribution. *Physiological and Biochemical Zoology* 73:83-96.
- Stillman, J. H., and C. A. Reeb. 2001. Molecular phylogenetics of Eastern Pacific Porcelain crabs, Genera *Petrolisthes* and *Pachycheles*, based on the mtDNA 16S rDNA sequence: Phylogeographic and Systematic Implications. *Molecular Phylogenetics and Evolution* 19:236-245.
- Stillman, J. H., and G. N. Somero. 1996. Adaptation to temperature stress and aerial exposure in congeneric species of intertidal porcelain crabs (genus *Petrolisthes*):

- correlation of physiology, biochemistry and morphology with vertical distribution. *Journal of Experimental Biology* 199:1845-1855.
- Stillman, J. H., and G. N. Somero. 2000. A comparative analysis of the upper thermal tolerance limits of Eastern Pacific porcelain crabs, genus *Petrolisthes*: Influences of latitude, vertical zonation, acclimation, and phylogeny. *Physiological and Biochemical Zoology* 73:200-208.
- Sunnucks, P. 2000a. Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15:199-203.
- Sunnucks, P. 2000b. Reply from P. Sunnucks. *Trends in Ecology and Evolution* 15:377.
- Supungul, P., P. Sootanan, S. Klinbunga, W. Kamonrat, P. Jarayabhand, and A. Tassanakajon. 2000. Microsatellite polymorphism and the population structure of the Black Tiger Shrimp (*Penaeus monodon*) in Thailand. *Marine Biotechnology* 2:339-347.
- Tam, Y. K., and I. Kornfield. 1996. Characterization of microsatellite markers in *Homarus* (Crustacea, Decapoda). *Molecular Marine Biology and Biotechnology* 5:230-238.
- Thorson, G. 1961. Length of pelagic larval life in marine bottom invertebrates as related to larval transport by ocean currents. Pp. 455-474 in M. Sears, ed. *Oceanography*. Horn Shafer, Baltimore, MD.
- Thorson, G. 1964. Light as an ecological factor in the dispersal and settlement of larvae of marine bottom invertebrates. *Ophelia* 1:167-208.

- Todd, C. D., W. J. Lambert, and J. P. Thorpe. 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktrophic and pelagic lecithotropic larval stages: Are pelagic larvae "for" dispersal? *Journal of Experimental Marine Biology and Ecology* 228:1-28.
- Toonen, R. J. 1997. Microsatellites for ecologists: Non-radioactive isolation and amplification protocols for microsatellite markers. Available at <http://biogeek.ucdavis.edu/Msats/index.html> (Appendix 1).
- Toonen, R. J., and S. Hughes. in press. Increased throughput for fragment analysis on an ABI 377 automated sequencer using a membrane comb and STRand software. *Biotechniques* (Appendix 2).
- Urbani, N., B. Sainte-Marie, J.-M. Sevigny, D. Zadworny, and U. Kuhnlein. 1998. Sperm competition and paternity assurance during the first breeding period of female snow crab (*Chionoecetes opilio*) (Brachyura: Majidae). *Canadian Journal of Fisheries and Aquatic Sciences* 55:1104-1113.
- Valsecchi, E., P. Palsboll, P. Hale, D. Glockner-Ferrari, M. Ferrari, P. Clapham, F. Larsen, D. Mattila, R. Sears, J. Sigurjonsson, M. Brown, P. Corkeron, and B. Amos. 1997. Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution* 14:355-362.
- Waples, R. S. 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution* 41:385-400.

- Watterson, G. 1986. The homozygosity test after a change in population size. *Genetics* 112:899 - 907.
- Weir, B. S. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. 445 pp. Sinauer Associates, Inc., Sunderland, MA.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323-354.
- Xu, Z., J. H. Primavera, L. D. de la Pena, P. Pettit, J. Belak, and A. Alcivar-Warren. 2001. Genetic diversity of wild and cultured Black Tiger Shrimp (*Penaeus monodon*) in the Philippines using microsatellites. *Aquaculture* 199:13-40.

Chapter 2.

Isolation Genetic differentiation among populations of the intertidal crab *Petrolisthes cinctipes*.

Robert J. Toonen

Center for Population Biology,
Section of Evolution and Ecology
University of California, Davis,
Davis, CA, USA. 95616.

Abstract

I used microsatellite markers to characterize genetic structure within and among geographic populations of the porcelain shore crab *Petrolisthes cinctipes*. Adult crabs were sampled from 12 geographic regions covering nearly the entire species range, from Morro Bay, CA to Neah Bay, WA. Significant population differentiation was observed for nearly all pair-wise comparisons between crabs sampled from all geographic sites across all spatial scales. Retention zones associated with headlands cannot account for the pattern of genetic structure observed among geographic populations of *P. cinctipes*. Pair-wise genetic distances between sites within embayments were often much larger than genetic distances between sites separated by similarly short geographic distances (< 50 km) on opposite sides of a headland. An analysis of molecular variance indicates that northern California sites are genetically differentiated from sites in Washington State and central California. The long planktonic larval periods of many marine organisms, such as *P. cinctipes*, are generally thought to confer high dispersal potential. The significant genetic differentiation among regions revealed by this study suggests that larvae of *P. cinctipes* do not disperse to the extent inferred by laboratory studies of their pelagic lifespan. Inferred levels of effective gene flow among populations decrease with geographic distance of separation between sites, but this relationship appears nonlinear. 35% of the variation in gene flow among geographic populations is explained by a curvilinear relationship between the log of M and distance of separation.

This curvilinear relationship suggests that pelagic larvae of *P. cinctipes* settle with greatest frequency at an intermediate distance from the natal site.

Introduction

The relationship between dispersal and population genetic structure is well-known: gene flow, if unopposed by other factors, quickly homogenizes the populations of a species. Obvious barriers to dispersal are rare in the sea, and approximately 80% of marine organisms, both vertebrate and invertebrate, maintain a pelagic larval stage in their life cycle (Thorson 1964). Thus, owing to their apparently great potential for dispersal via planktonic larvae that may remain in the water column for days to months, populations of benthic marine invertebrates with pelagic larvae ought to be genetically homogeneous over broad geographic regions. Indeed, numerous empirical studies on species with long-lived pelagic larvae confirm that across similar geographic scales, these species generally show significantly lower levels of genetic structure than species lacking such larvae (reviewed by Burton 1983; Hedgecock 1986; Palumbi 1995; Bohonak 1999; Grosberg and Cunningham 2001). However, despite the expectation that marine invertebrate species with planktonically dispersing larvae should be characterized by high levels of gene flow, empirical studies frequently reveal significant degrees of genetic structure among geographic sites (e.g., Lacson et al. 1989; Karl and Avise 1992; Li and Hedgecock 1998; Todd et al. 1998; Kyle and Boulding 2000; reviewed by Avise 1994). In fact, some such studies (e.g., Johnson and Black 1982; Johnson and Black 1984a, b; Kordos and Burton 1993; Edmands et al. 1996) show that estimates of F_{ST} among adjacent populations (tens of km) is of the same or greater magnitude than values across the species range (thousands of km).

Studies of some taxa suggest that patterns of genetic structure better reflect historic oceanic circulation patterns than of contemporary paths of gene flow (e.g., Benzie and Williams 1997; reviewed by Bossart and Prowell 1998). In other cases, fine-scale genetic structure appears to result from post-settlement selection (e.g., Koehn et al. 1980; Hilbish and Koehn 1985) or temporal variation among new recruits (e.g., Johnson and Black 1984a, b; Kordos and Burton 1993; Li and Hedgecock 1998).

Despite the lack of obvious physical barriers to dispersal among contemporary populations, current patterns may restrict the movement of larvae through coastal oceans. For example, the concordant breaks in genetic structure of many unrelated taxa along the Gulf of Mexico and Atlantic coasts may result from strong oceanic currents around Cape Canaveral, FL (Avisé 1994). Likewise, current patterns around Pt. Conception, CA appear to limit the passage of pelagic larvae traveling northward, such that many species with pelagic larvae have their northern range limit at this well-known biogeographic boundary (Wares et al. 2001). Other headlands along the California coastline also have associated physical oceanographic features that retain larvae (Wing et al. 1995a; Wing et al. 1995b; Wing et al. 1998; Wares et al. 2001) and may limit gene flow among populations separated by these promontories.

Oceanic currents along the northern California coastline are predominated by upwelling, driven by strong equatorial winds from April through August, and punctuated by short relaxation events when the winds cease or reverse direction (e.g., Send et al. 1987; Hayward and Mantyla 1990; Schwing and Husby 1991; Largier et al. 1993; reviewed by Huyer 1983). During periods of upwelling, headlands deflect

southward currents offshore to form gyres. Pelagic larvae appear to be transported offshore from their natal sites and accumulate in gyres associated with coastal promontories, such as Pt. Reyes and Pt. Arena (Wing et al. 1995a; Wing et al. 1995b; Wing et al. 1998). When upwelling winds subside, water from these offshore gyres flows onshore. This onshore flow distributes the accumulated larvae northward throughout the embayment depending on the duration of the relaxation event. Pulses of larval settlement along the California coast coincide with the arrival of warm, buoyancy-driven currents from these gyres (e.g., Wing et al. 1998; Lundquist et al. 2000). However, alongshore relaxation flows do not reach all populations within an embayment during every relaxation event. For example, winds driving upwelling often resume before the relaxation flow reaches more northerly sites (Wing et al. 1995a; Wing et al. 1995b). Thus, the upwelling-relaxation flow hypothesis (Botsford et al. 1994; Wing et al. 1995a; Wing et al. 1995b) predicts differential delivery of larvae to sites with increasing distance northward from a headland. This differential delivery of larvae to sites within an embayment should result in recruitment being most common and predictable on the northward side of promontories where relaxation currents regularly reach, but becoming sporadic with increasing distance northward from the headland (e.g., Botsford et al. 1994; Wing et al. 1995a; Morgan et al. 2000).

Such advances in our understanding of nearshore oceanographic processes have led to the formation of mechanistic hypotheses predicting how larval and adult population dynamics may be connected (e.g., Gaines et al. 1985; Roughgarden et al. 1987; Roughgarden et al. 1988; Bertness et al. 1992; Gaines and Bertness 1992; Wing

et al. 1998). If these physical oceanographic processes associated with promontories along the northern California coastline are important determinants of larval dispersal, then larvae should be retained in the gyre associated with a headland south of their natal population, and redistributed northward during relaxation events. If the relaxation event were to last long enough that northward flow reached the next promontory, the flow would be deflected offshore, preventing further larval settlement. This should result in sites between headlands being genetically homogeneous, but could lead to differentiation of populations on opposite sides of retention zones. If, on the other hand, some proportion of larvae bypass retention zones to disperse and recruit beyond their natal embayment, then genetic homogeneity among sites should be equally likely between headlands and across them. However the question of whether oceanographic circulation patterns leave a genetic signature that could have evolutionary ramifications remains unanswered. In this study, I examined patterns of genetic differentiation in a species with long-lived planktonic larvae along the Pacific coast of North America to assess support for these predictions. In this chapter, I use microsatellites to characterize genetic structure among geographic populations of the porcelain shore crab, *Petrolisthes cinctipes*, across approximately 1500 km (Neah Bay, WA to Morro Bay, CA), encompassing nearly the entire species range (Stillman and Reeb 2001). I then examine clustering patterns among populations and use an analysis of molecular variance approach to weigh support for the upwelling-relaxation flow hypothesis and a simple isolation-by-distance hypothesis for the pattern of genetic variation characterized in *P. cinctipes*.

Materials and methods

Study organism

Two species of porcelain crabs (family Porcellanidae) occur in the Eastern North Pacific: *Petrolisthes cinctipes*, and *Petrolisthes eriomerus*. The family Porcellanidae is one of the most diverse groups of nearshore decapod crustaceans, containing 23 genera of which *Petrolisthes* is the largest, with over 100 currently described species (Haig 1960). Recent molecular work, however, suggests that the genus *Petrolisthes* contains 2 distinct clades, and that many of the currently recognized nominal species require further examination (Stillman and Reeb 2001). These two Eastern North Pacific species are not among those in question (Stillman and Reeb 2001).

Petrolisthes cinctipes is restricted to a relatively narrow band of the intertidal zone (from roughly 0 to 1.8 m above MLLW), but is found in virtually all appropriate habitat throughout its range from roughly the Dixon Entrance (British Columbia, Canada, 54° 71' N) to just north of Point Conception, CA (34° 26' N) (Jensen and Armstrong 1991). Despite the apparent motility of porcellanid crabs, adults are quite sedentary in the field. Mark-release-and-recapture data show that >50% of crabs (348 of 600) were recovered within a few meters of the point of release after two months (Jensen and Armstrong 1991). Individuals of both sexes and all size classes from newly settled postlarvae (~1.5 mm) to large adults (up to 20 mm carapace width) live together

in large aggregations. *P. cinctipes* is one of the most abundant decapod crustaceans in the Eastern North Pacific, occurring in densities as high as 3933 individuals / m² (Jensen 1990).

Although the chelae (claws) of *P. cinctipes* may account for nearly 50% of the body mass, these crabs are suspension feeders that gain most of their nutrition from diatoms (Knudsen 1964). The chelae are used primarily for territorial disputes and mating displays. Chelae are important determinants of mate choice in other crabs (Sainte-Marie et al. 1999), and the same is true of *P. cinctipes* (Jensen, pers. comm.).

The sexes are separate in *P. cinctipes*, but sexual dimorphism is slight. Fertilization of the eggs occurs externally after they are released, but prior to being attached to the female's pleopodal setae for brooding (Nyblade 1987). Females typically produce broods from February through mid-April. Newly extruded broods may contain from 10 to 1580 embryos, each roughly 0.80 - 0.84 mm in diameter (Knudsen 1964). Some females may produce a second brood between May and September. Feeding larvae spend several months in the plankton before becoming competent to settle. Assuming an average current speed of roughly 10 cm / s, larvae have the potential to drift passively > 1000 km during their larval development. After larvae molt into megalopae, they become capable of settling quickly in response to a waterborne cue associated with live conspecific adults (Jensen 1989, 1991a, b). In contrast, megalopae held in a bare container lacking the settlement cue delay settlement by as much as 45 days before dying in culture (Jensen 1991a). Megalopae of *P. cinctipes* do not molt at the time of settlement but rather lose the ability to swim due to

degeneration of the pleopods; metamorphosis to the pigmented first instar juvenile takes place a week or more later (Jensen 1991a). This delay of metamorphosis after settlement makes identification of newly settled postlarvae straightforward.

Collection of samples and DNA extraction

Adult *Petrolisthes cinctipes* were collected from rocky cobble around large boulders at each site in Figure 1. Boulders were selected based on the presence of dense aggregations of crabs. During collection, live crabs were stored in a mesh container filled with fresh kelp, in order to minimize stress. *Petrolisthes* readily self-amputate a cheliped at a specialized break point in the limb as a defensive strategy when threatened, making non-lethal sampling quite simple. Adult crabs (5 - 15 mm CW) at each site were harassed to stimulate self-amputation of the cheliped. Detached chelipeds were immediately put into 95% ethanol. Following the collection, crabs were all released to their original location. Samples were kept on ice in the field for roughly 4-5 days, but upon return to the laboratory, the 95% ethanol was decanted and replaced. Vials containing ethanol-preserved chelipeds were subsequently stored at -40°C.

In the laboratory, each cheliped was assigned a unique identification number; up to 90 individuals from each site were selected randomly from the numbered cheliped samples for subsequent DNA extraction and genetic analysis. If fewer than 90 adults were collected at a site, all samples were included in the study. Unlike the sampling reported in other chapters, there is no temporal component to sampling for this study,

and this collection was made in the manner typical of other snap-shot population genetic surveys. Samples from Crescent City, CA to Dillon Beach, CA were collected on a single tidal series (5 days) in August of 1997. Samples from Neah Bay, WA were collected in August 1999. Samples from Monterey and Morro Bay, CA were collected in January and March 2000, respectively. DNA extraction and subsequent PCR amplification were done blindly, and crab identity was reassigned to numbered samples only after all samples had been scored.

Approximately 5 mg of muscle tissue was dissected from each cheliped, and DNA was extracted from the dissected tissue using a modification of the 2x CTAB protocol outlined in Grosberg et al. (1996). Addition of 0.2% SDS and 2.0 mg / ml proteinase-K to the CTAB lysis buffer, together with constant agitation during an overnight incubation at 55°C, resulted in much more consistent amplification of samples than the standard 2X CTAB protocol (Toonen 1997), Appendix I]. I discovered that minimizing the amount of carapace included in the extractions also led to consistently higher quality and quantity of DNA extracted.

Amplification and electrophoresis

Primers for amplifying microsatellite loci used in this study were isolated and designed following the protocol from Appendix I, as outlined in Chapter 1 (Toonen 1997). Four highly polymorphic microsatellite loci (Pc36L, Pc156s, Pc160b & Pc170s) were initially used to genotype all individuals in this study (Table 1). However, I

subsequently determined that two of the loci (Pc160b & Pc36L) were not inherited in a Mendelian fashion (Chapter 1). Therefore, data from loci Pc160b and Pc36L were excluded, and only locus Pc156s and locus Pc170s are included in these analyses.

In a final volume of 10 μ l, PCR reaction mixes contained Perkin Elmer 10X Buffer II at 1X concentration, 2.5 mM MgCl₂, 0.1 mM dNTPs, 1X BSA, 0.5 mM of each the forward and reverse primer, 1 unit of Taq polymerase, and 1-50 ng of template DNA (Toonen 1997). PCR amplifications were performed using a touchdown PCR protocol on a Perkin Elmer 9700 thermocycler as outlined in Chapter 1.

Electrophoresis of amplified PCR products was done on an Applied Biosystems ABI 377 XL automated sequencer and scored with the software program STRand (Hughes 1998), as outlined in Toonen and Hughes (In press, Appendix II).

Data analysis

STRand (Hughes 1998) stores allelic data generated by the ABI 377XL sequencer in Microsoft's relational database program, Access, which allows direct export into the spreadsheet program Excel. Allelic data were then translated into the input format for data analysis programs using the MS toolkit (Park 2001) macro for Excel 2000.

The use of molecular genetic markers to portray population structure rests on at least three basic premises: 1) selective neutrality, 2) absence of null alleles, and 3) independent assortment of loci (reviewed by Grosberg and Cunningham 2001). The

operation of natural selection is difficult to assess with anonymous markers, and genetic markers are typically assumed to be selectively neutral in the absence of data to the contrary (e.g., Karl and Avise 1992). Although statistical analyses of population genetic data are relatively insensitive to weak selection (Slatkin and Barton 1989), strong selection can lead to serious biases in the interpretation of genetic data (Gillespie 1986). If, however, allelic frequencies do not deviate significantly from the Ewens sampling distribution (Slatkin 1994, 1996), then estimates of gene flow calculated from Wright's hierarchical F-statistics, or Weir and Cockerham's (Weir and Cockerham 1984) comparable estimators, should be robust to unmeasured selective forces (N. Barton, pers. comm.). All loci were tested for significant deviations from neutrality using Slatkin's exact test program, Enumerate (Slatkin 1997). This test also provides some confidence that populations are not so far from genetic equilibrium as to invalidate population genetic analyses (Hutchison and Templeton 1999). Populations that have not reached equilibrium also generally show significant departures from the Ewens sampling distribution (Watterson 1986; Schneider et al. 2000). Therefore, confirming a statistical fit of allelic frequencies to the Ewens sampling distribution appears a reasonable first step prior to application of genetic analyses that assume equilibrium and selective neutrality.

Null alleles are most easily detected by examining patterns of inheritance (e.g., Callen et al. 1993; Paetkau and Strobeck 1995; Pemberton et al. 1995). Unfortunately, lab crosses with intertidal crabs are extremely difficult, and long-term sperm storage in other crabs (Jensen et al. 1996) suggests that a long period of female isolation would be

required prior to initiating such crosses. Thus, analysis of inheritance patterns along known lineages is simply not possible within the scope of this project. Therefore, I excluded any locus for which I was consistently unable to amplify a product from any individual within a random sample of crabs from four sites across Northern California (Chapter 1).

Testing conformity to Hardy-Weinberg (H-W) expectations is not straightforward for SSR loci, because the G-statistic does not follow a standard distribution when there are many alleles per locus unless the sample size is extremely large (Edwards et al. 1992; Guo and Thompson 1992). Significant deviations from H-W expectations must instead be determined by permutation tests (Deka et al. 1991; Guo and Thompson 1992). I performed global tests for deviation from H-W equilibrium using the Arlequin Population Genetics Data Analysis Program (Schneider et al. 2000). Tests for deviation from H-W expectations employ a Markov-chain method as described in Guo and Thompson (1992). I used a Markov chain length of 500,000 steps, with a 10,000 step dememorization, for all tests of H-W proportions presented herein. Independent assortment of SSR loci was also tested in Arlequin, using a likelihood-ratio test (Slatkin and Excoffier 1996) against an empirical distribution which was likewise obtained by 500,000 permutations.

There remains considerable uncertainty about both the generality of mutational processes underlying microsatellite evolution among diverse taxa and the appropriate statistical theory for data analyses (reviewed by Luikart and England 1999; Chambers and MacAvoy 2000; Nève and Meglécz 2000; Sunnucks 2000a, b). Wright's F -

statistics are based on an infinite alleles model (IAM). However, microsatellites are widely thought to evolve in a stepwise manner, and R_{ST} (Slatkin 1995) is an alternative hierarchical statistic that incorporates the stepwise mutational model (SMM).

Empirical data are mounting that no single statistical estimator has superior performance in all situations. For example, Shriver et al. (Shriver et al. 1993) found the mutational model that best fit a sample of about 50 tandem repeat loci depended on the size of the repeat: all 12 of the 3-5 bp repeats fit the SMM, 20 of 31 mono- and dinucleotide repeats fit the SMM while the remainder fit the IAM, and 8 of 11 minisatellite loci (15-70 bp) fit the IAM. However, for both empirical microsatellite data from the Atlantic cod, *Gadus morhua*, (Ruzzante 1998), and SSR simulation data (Takezaki and Nei 1996), SMM-based measures of genetic distance and population structure did not appear to perform consistently better than did IAM-based statistics regardless of the repeat motif in question. Because the most appropriate analysis for microsatellite data remains unresolved, I have included both IAM and SMM analyses.

I calculated Weir and Cockerham's (1984) unbiased estimator of Wright's F_{ST} (θ) on a locus-by-locus basis using the Tools for Population Genetic Analyses (TFPGA) software (Miller 1997). TFGPA uses the method and terminology outlined in Weir and Cockerham (1984; Weir 1996). Because only two loci are included in this data set, bootstrapping and jackknifing are not meaningful, and therefore no mean, standard error or confidence intervals for θ are presented. Goodman's ρ (Goodman 1997), an unbiased estimator of Slatkin's R_{ST} (Slatkin 1995), and $(\delta\mu)^2$, an SMM-based

genetic distance measure preferable for microsatellite data (Goldstein et al. 1995a, b; Takezaki and Nei 1996), were both calculated using R_{ST} Calc 2.2 (Goodman 1997).

Assuming that genes move with equal likelihood among all populations, the effective number of migrants exchanged between populations each generation (N_m) can be estimated using the relationship $N_m = (1 - F_{ST}) / 4 F_{ST}$ (Wright 1978). To infer the pattern of gene flow between geographic populations from the spatial pattern of genetic differentiation, I used M , the estimate of N_m calculated separately for each pair of populations. Although the absolute value of N_m may be unreliable because of the myriad of untested assumptions required for this calculation (Whitlock and McCauley 1999), the relative average amount of exchange between populations should be accurately estimated using M (Hutchison and Templeton 1999). Simulations (Slatkin and Maddison 1990; Slatkin 1993; Hellberg 1994) suggest that for populations distributed linearly and obeying the assumptions of a one-dimensional stepping stone, the expected slope of a regression of $\log_{10}(M)$ versus $\log_{10}(\text{geographic distance of separation})$ is -1.0 (i.e., gene flow between localities correlates inversely with the distance separating those populations); for a two-dimensional stepping stone model the expected slope is -0.5 (i.e., gene flow varies as the inverse square of distance). Standard regression analyses are inappropriate for these nonindependent, pair-wise comparisons, and therefore Mantel's test (Mantel 1967) was used to determine the significance of these relationships (Hellberg 1994). Ordinary least squares (OLS) regression slope (b) is more appropriate for significance testing, while the reduced major axis (RMA) regression slope (v) better estimates the relationship between

inferred gene flow when geographic distance is used as a proxy for the number of dispersal steps between populations (Hellberg 1994).

The two genetic distance measures that most often returned the correct tree topology in simulations based on both IAM and SMM (Takezaki and Nei 1996) were Cavalli-Sforza and Edwards' (1967) chord distance (D_{CE}) (Edwards and Cavalli-Sforza 1965; Cavalli-Sforza and Edwards 1967) and D_A (Nei et al. 1983). However, Nei's standard genetic distance, D_S (Nei 1972), gave the best approximation of branch lengths for IAM simulations, while $(\delta\mu)^2$ (Goldstein et al. 1995b) best approximated branch lengths in SMM simulations (Takezaki and Nei 1996). I used PHYLIP (Felsenstein 1993) to calculate the C-S chord distance (Cavalli-Sforza and Edwards 1967). D_A (Nei et al. 1983) was calculated in Arlequin. PHYLIP was also used to construct UPGMA (Unweighted Pair Group Method with Arithmetic mean) dendrograms (Michener and Sokal 1957) using both D_{CE} and D_A . UPGMA dendrograms give an indication of the degree of genetic dissimilarity among populations; the longer the branches separating a pair of populations, the more genetically dissimilar are individuals sampled from those geographic locations. If gene flow is strictly a function of the number of dispersal steps between populations, then sites that are geographically proximate should cluster more closely than with distant sites.

Finally, I used Arlequin to perform an analysis of molecular variance (AMOVA), which partitioned observed variance into components due to inter-individual differences and inter-population differences (Schneider et al., 1997). Using the AMOVA framework within Arlequin, I partitioned observed genetic variation

components and calculated hierarchical F -statistic analogs (Φ -statistics). The significance of these values were tested using Arlequin by performing 500,000 matrix permutations (Excoffier et al. 1992). AMOVA allows me to determine the proportion of the observed genetic variation explained by differences between geographic regions, sampling sites within geographic regions, and individuals within sampling sites.

Results

Microsatellite variation

The four microsatellite loci used in this study are highly polymorphic, with 26 (Pc156s) to 55 (Pc160b) alleles per locus and expected heterozygosities ranging from 0.65 to 0.95 (Table 1). Tests for heterozygote deficiencies across all individuals sampled revealed significant departure from Hardy-Weinberg expectations at only Pc36L, although Pc160b is marginal (Table 1). A significant deficiency of heterozygotes at only a single locus suggests the presence of null alleles at that locus (Callen et al. 1993; Paetkau and Strobeck 1995; Pemberton et al. 1995). Given that many individuals failed to amplify a scorable product during PCR (Table 1), these data are consistent with null alleles at this locus. However, the non-Mendelian inheritance at these loci (Chapter 1) poses a much more serious violation of the statistical assumptions upon which these analyses are based. I have therefore excluded these loci from the following analyses. Again, it is important to emphasize that loci were only excluded

from these analyses when they failed to meet the basic assumptions upon which the genetic analyses used herein are based, not because of the patterns that they revealed.

The exclusion of loci that failed to fit the basic assumptions of neutral genetic analyses left only locus Pc156s and locus Pc170s for characterization of genetic structure in *P. cinctipes*. The allelic frequency distributions of both loci are multimodal, and do not fit the expectations of a single-stepwise mutational model (Fig. 2). As expected for highly polymorphic SSR loci, the average heterozygosity within sampling sites was relatively high ($\bar{x} = 0.73 \pm 0.03$ SE). The number of individuals genotyped from each site ranged from 54 to 95, which should be sufficiently large to minimize bias associated with calculating measures of genetic distance and population structure from microsatellite data (Ruzzante 1998).

The mean number of alleles per locus varied significantly among sites ($F = 10.1$, $df = 11$, $p < 0.01$), ranging from a low of 3.5 ± 0.4 SE in Morro Bay to a high of 23.0 ± 1.4 SE at Cape Mendocino (Fig. 3). There were significant changes in the number of alleles per locus across sites ($F = 2.59$, $df = 11$, $p = 0.01$) that were similar for both loci ($F = 0.99$, $df = 1$, $p > 0.05$), and there was no significant interaction between locus and site ($F = 1.20$, $df = 11$, $p > 0.05$). Parallel changes in the number of alleles per locus across loci merits further attention, but there is no obvious geographic pattern to the distribution (Fig. 3).

Tests of conformity to H-W expectations among all individuals sampled at these two loci provided no evidence for a deficiency or excess of heterozygotes overall (Table 1). However, even after Bonferoni correction, seven of 24 locus-by-locus tests at each

site deviated significantly from H-W expectations (Table 2). In six of the cases where departure from H-W expectation was significant, there was an excess of heterozygotes, whereas only one site (Albion for Pc170s) showed a significant heterozygote deficit (Table 2). Significant departures from H-W expectations were split almost evenly between the two loci (3 for Pc156s, 4 for Pc170s).

Genetic structure

I used Arlequin to perform exact tests of population differentiation (Raymond and Rousset 1995; Goudet et al. 1996). All pair-wise comparisons of populations showed significant differentiation using either locus singly, or both loci together ($\alpha = 0.01$, analyses not shown). For both loci, one allele was relatively common among all sites, and three alleles together accounted for approximately 80% of the allelic variation across all sites (Figs. 2 & 4). Private alleles at low frequencies were found at each locus, with usually only a single private allele at any given site. The two exceptions were Cape Mendocino (site #3) and Shelter Cove (site #4), at which multiple private alleles occurred (Fig. 2). All private alleles detected in this study occurred in the northern California sites between Cape Mendocino and Salt Point (Fig. 2).

Differences in the frequency of the three most common alleles at each locus show no obvious pattern across northern California (sites #2 - 10 on Fig. 1). However, allele frequency differences between northern California sites and those from central California (Monterey and Morro Bay) and Washington (Neah Bay) sites are striking at

both loci (Fig. 4). It is also important to note that these three sites (Monterey, Morro and Neah Bays) were all sampled in a different year than were the central California sites, and temporal variation in allelic frequencies may contribute to these differences among sites as well (see Chapter 3).

Because there are only two loci included in these analyses, jackknifing and bootstrapping across loci to calculate confidence intervals and standard errors for these estimates is trivial. Therefore, only the locus-by-locus results and means are presented for these analyses of population subdivision. However, in all analyses the general trend revealed by both loci was concordant, and differences in estimates of population subdivision and genetic distance generated by each locus were slight. Weir and Cockerham's (1984) estimator of population subdivision, θ ($= F_{ST}$), averaged 0.11 overall, while f ($= F_{IS}$) brackets zero (Table 3). Estimates are similar across loci, and jackknifing across populations suggests that no single population contributes disproportionately to the overall estimates (Table 3). Overall, these results suggest a moderate amount of population differentiation among sites, but no apparent deviation from panmixia within sites. Goodman's (1997) estimator of R_{ST} , ρ , similarly suggests significant genetic differentiation ($\bar{\rho} = 0.13$, $p < 0.01$) among populations.

Estimates of the effective number of migrants per generation (M , the estimate of Nm calculated separately for all population pairs, see Slatkin and Maddison 1990; Slatkin 1993), were 4.72 ± 0.91 SE and 1.77 ± 0.005 SE for θ and ρ , respectively. The calculation of M relies on a number of underlying assumptions that decrease confidence in the absolute value (reviewed by McCauley 1993; Whitlock and McCauley 1999).

However, the relative influence of gene flow and drift on population structure can still be inferred reliably using this approach, and estimates of M should be viewed in that context (Hutchison and Templeton 1999).

All pair-wise exact tests of population differentiation between sites are significant using θ ($\alpha = 0.05$), but using ρ , seven of the 66 pair-wise comparisons were not: Neah Bay / Morro Bay; Crescent City / Ft. Bragg and Salt Point; Ft. Bragg / Salt Point and Monterey; and Albion / Pt. Arena and Morro Bay. Like the results of similar studies (e.g., Kordos and Burton 1993; Hellberg 1994, 1996; Moberg and Burton 2000), both θ and ρ correlate significantly with untransformed pair-wise geographic distance among sites (Mantel test, $r^2 = 0.41$ and 0.30 , respectively, $p < 0.05$). The relationship between the base-10 log-transformed estimates of M_θ and the geographic distance of separation (km) between collection sites is significant (Mantel test, $r^2 = 0.18$, $p < 0.01$), but the relationship with M_ρ is not (Fig. 5). Following the technique of Hellberg, the slope of the RMA regression for $\log(M_\theta)$ versus $\log(\text{geographic distance of separation})$ is -0.43 (95% CI = $-0.74 \leq v \leq -0.23$), whereas the slope for M_ρ is -0.17 (95% CI = $-0.24 \leq v \leq -0.04$). However, like the Mantel test, the OLS slopes suggest that while the relationship between the logarithms of distance and M_θ is significant ($-0.79 \leq b \leq -0.32$), the relationship with M_ρ is not ($-0.45 \leq b \leq 0.11$).

Interestingly, a quadratic regression provides a greatly improved parabolic fit to the data for both M_θ and M_ρ . Higher order (cubic and quartic) polynomial regression produced almost identical curves that did not further improve the fit (analyses not shown). In the case of M_θ , quadratic regression explains an additional 17% of the

variation in the data ($r^2 = 0.35$, $p < 0.01$), as compared to M_p in which the relationship between inferred gene flow and geographic distance becomes significant when an additional 7% of the variation is explained by quadratic regression ($r^2 = 0.09$, $p < 0.05$). In every analysis, the SMM-based estimates (ρ) explain less of the observed variation than the IAM-based estimates (θ). Although ρ is generally considered the more appropriate statistic for microsatellites, the multi-modal allelic frequency distributions of loci Pc156s and Pc170s are inconsistent with the expectations under a single step-wise mutational model.

All measures of population differentiation and genetic distance between sites generally increase with geographic distance to both the north and to the south of a minimum in northern California (Tables 4 & 5 & 6). Pair-wise comparisons of population subdivision and genetic distance between all populations sampled are generally lowest somewhere around Pt. Arena (#7) or Salt Point (#8). This minimum results in a pattern in which estimates of population differentiation and genetic distance are consistently lower among sites separated by hundreds of km than among sites separated by tens of km. For example, pair-wise estimates of both population subdivision and genetic distance between Bodega Bay (#9) and the remainder of the sites is lowest around Pt. Arena (#7, roughly 100 km) ($\theta = 0.02$, $\rho = 0.04$, $D_{CE} = 0.12$, $(\delta\mu)^2 = 4.42$), whereas comparisons with the adjacent site, Dillon Beach (#10, roughly 10 km), are usually much higher ($\theta = 0.16$, $\rho = 0.29$, $D_{CE} = 0.18$, $(\delta\mu)^2 = 56.80$). This discontinuity in the relationship between genetic and geographic distance across the range of *P. cinctipes* is repeated across all site comparisons such that the column and

row values are generally lowest around Pt. Arena or Salt Point (Tables 4 & 5).

Furthermore, genetic distances between geographic sites on opposite sides of a headland were lower than approximately equidistant sites adjacent to one another within an embayment in more than 80% of the pair-wise comparisons (e.g., Pt. Arena/Albion versus Ft. Bragg/Albion).

Based on D_A (Nei et al. 1983), the overall pattern of clustering in the UPGMA dendrogram matches the geographic position of collection sites quite well, with the northernmost site being genetically distinct from all others (Fig. 6). The central California sites (Monterey and Morro Bay) also form a well-supported cluster which is distinct from the remaining northern California sites (Fig. 6). Within the northern California sites, however, Dillon Beach appears to be an outlier in terms of genetic similarity to the remaining sites. Also, the discontinuity in the pair-wise genetic distances results in Albion clustering with the sites to the south of Pt. Arena, whereas Salt Point clusters with sites much farther to the north (Fig. 6). There is a significant correlation between D_A and every other genetic distance measure I have calculated for these data, including D_{CE} (Edwards and Cavalli-Sforza 1965; Cavalli-Sforza and Edwards 1967), allele sharing (Bowcock et al. 1994), Nei's standard (Nei 1972) and unbiased (Nei 1978) genetic distances, Roger's genetic distance (Rogers 1972) and Wright's (Wright 1978) subsequent modification of it, and coancestry (Reynolds et al. 1983) distance. All seven of these genetic distance measures show a significant correlation ($r^2 = 0.87 - 0.99$, $p < 0.001$) and all produce similar overall pattern of clustering in UPGMA dendrograms.

I used an analysis of molecular variance (AMOVA) to characterize the distribution of genetic variation within and among various sampling units, and to weigh the merits of predictions generated from three dispersal hypotheses (Table 6). The first hypothesis predicts that oceanic gyres resulting from headlands along the Northern Pacific coastline (Botsford et al. 1994; Wing et al. 1995a; Wing et al. 1995b) should limit gene flow from gyre-to-gyre, leaving detectable genetic differentiation among embayments (such as between Pt. Reyes, Pt. Arena and Cape Mendocino). More than twice as much genetic variation occurs among sites within embayments as among embayments (7.66% versus 3.55%). This is also the only pooling of geographic sites into groups which results in a nonsignificant value of Φ value (Table 6). In contrast, the simple geographic region hypothesis, which compares all of the northern California sites (#2-10) to those in central California (#11 & 12) and Washington (#1), explains 9.65% of observed genetic variation, and the genetic break among these geographic regions is highly significant ($\Phi_{CT} = 0.10$, $p < 0.01$, Table 6). The final hypothesis, that migration does not occur in a simple stepwise pattern either by geographic site or by embayment, has nearly infinite alternative predictions. I used a post-hoc grouping of sites based on UPGMA clustering patterns (Fig. 6) to determine if more of the observed genetic variation was explained among clusters than among sites within clusters. Although this final hypothesis explains the highest proportion of among group variation of the three (11.30%), more than 87% of observed variation still occurs within sites rather than among clusters (Table 6).

The Dillon Beach site seems to be an outlier in terms of genetic distance, and the fact that this site has not been recolonized since the local extinction during the winter of 1997 suggests that it is not subject to regular recruitment. When Dillon Beach is excluded from the analyses in Table 6, the proportion of the variation among groups is identical for hypotheses one and actually drops for hypothesis three (analyses not shown), but not the simple geographic distance hypothesis. In fact, with Dillon Beach excluded from the analysis, with the exception of the post-hoc parsing of sites based on the UPGMA clusters, the simple geographic region hypothesis explains more of the among group variation in the data than any other possible parsing of sites in all combinations of 2, 3, 4, 5 and 6 sites per group (analyses not shown).

Discussion

There is a growing consensus that oceanographic circulation patterns influence the genetic structure of marine invertebrates with planktonic larvae (e.g., Reeb and Avise 1990; Planes 1993; Scheltema 1995; Benzie and Williams 1997; reviewed by Burton 1994; Hedgecock 1994a; Burton 1996; Grosberg and Cunningham 2001). Perhaps the best described of near-shore oceanographic processes affecting larval dispersal is the coastal upwelling zone of northern California (e.g., Send et al. 1987; Hayward and Mantyla 1990; Schwing and Husby 1991; Largier et al. 1993; Wing et al. 1995a; Wing et al. 1995b; reviewed by Huyer 1983; Wing et al. 1998). The upwelling-relaxation flow hypothesis (Botsford et al. 1994; Wing et al. 1995a; Wing et al. 1995b)

predicts that larvae produced within an embayment should be retained by gyres associated with headlands. The question remains whether these retention zones reduce gene flow among embayments sufficiently that they could have evolutionary ramifications.

To address this question, I chose microsatellite DNA markers (also known as simple sequence repeats, or SSR loci), because they are widely regarded as the most powerful tool currently available for population genetic studies (Goldstein and Schlötterer 1999). Despite the difficulty in developing microsatellite markers for marine invertebrates (see Chapter 1), there is great incentive to do so for studies such as this. If a genetic marker is selectively neutral, genetic differences among populations at equilibrium can only be detected if the marker mutates at a rate greater than the rate of migration between populations (reviewed by Burton 1994; Hedgecock 1994a; Burton 1996; Grosberg and Cunningham 2001). From this perspective, microsatellites have rapidly become the preferred genetic marker among the molecular genetic tools currently available to characterize spatial and temporal patterns of genetic variation within and among natural populations. Together with the fine-scale sampling of sites, the highly polymorphic microstellite loci used in this study provide the best chance currently possible of detecting a genetic signature of the upwelling-relaxation flow hypothesis.

No matter how the data presented herein are analyzed, retention of larvae by gyres associated with headlands cannot account for the structure observed among populations of *Petrolisthes cinctipes*. Pair-wise estimates of genetic distance between

sites separated by similarly short geographic distances (< 50 km) on opposite sides of a headland (e.g., Pt. Arena/Albion) were often much less than genetic distances between adjacent sites within an embayment (e.g., Albion/Ft. Bragg). The analysis of molecular variance (AMOVA) (Excoffier et al. 1992) provides no support for a genetic effect of the upwelling-relaxation flow hypothesis. Instead, the AMOVA supports grouping the northern California sites (#2-9) as a single region that shows significant genetic differentiation from both Neah Bay, WA (#1) and central California sites (#11-12). UPGMA clustering suggests that the northern California sites form three distinct groups, but AMOVA using the post-hoc clusters determined from the UPGMA dendrogram only slightly increased the proportion of variation explained among clusters. Furthermore, the proportion of variation within sites in the AMOVA using the post-hoc grouping of sites is actually higher than the simple geographic hypothesis. These results are inconsistent with a strong genetic consequence of retention zones. I hesitate to define this region as a single population, however, because virtually all pair-wise comparisons of genetic structure by any metric provide evidence for significant differentiation among northern California sites.

Support for the upwelling-relaxation flow hypothesis was also lacking in the allozyme studies of Edmands et al. (1996) and Moberg and Burton (2000) who characterized genetic structure of urchins in the genus *Strongylocentrotus* sampled in the same geographic region. Like both of these studies on urchins, the data presented here show that genetic differentiation between northern California sites separated by short distances often exceeded differentiation over much larger geographic distances.

In fact, the smallest pairwise estimate of $(\delta\mu)^2$ in this study occurs between the two sites at opposite ends of the species range (Neah Bay/Morro Bay). In their study of the urchin *S. franciscanus*, Moberg and Burton (2000) found that the single largest pair-wise genetic distance (Bodega Bay/Pt. Arena) occurred within the same embayment, whereas the smallest pair-wise genetic distance (Gualala/Albion) occurred across a headland. This is not the case with the *Petrolisthes* data, where both the largest (Neah Bay/Dillon Beach) and smallest (Neah Bay/Morro Bay) pair-wise $(\delta\mu)^2$ distances include the northernmost site sampled. However, Moberg and Burton (2000) did not include sites north of California, and if only the northern California (#2-10) sites are considered, without exception, pair-wise genetic distances between sites separated by headlands are always exceeded by at least one of the comparisons between sites within the same embayment. Thus, the recurring pattern from these genetic surveys is that genetic differentiation between embayments is of the same or lesser magnitude than among sites within embayments.

Although it falls roughly in the middle of the sampling area, Dillon Beach stands apart from the remaining northern California sites. Regardless of the metric of population subdivision or genetic distance, most of the largest values in this study involve comparisons with the Dillon Beach population. Dillon Beach is at the mouth of Tomales Bay, in the lee of the tip of Pt. Reyes. This position may explain the degree of isolation it experiences, and why, following a local extinction of the population during the winter of 1997, the site has not been recolonized. Exclusion of Dillon Beach from the simple geographic hypothesis in an AMOVA analysis increases the proportion of

the variation explained among regions. However, excluding Dillon Beach from other analyses did not qualitatively affect the overall pattern of genetic subdivision, percent variation explained or UPGMA clustering (analyses not shown). Interestingly, Moberg and Burton (2000) also discovered a genetic incongruity in this area; including samples from Bodega Bay (< 10 km away) resulted in significant genetic differentiation among populations of adult urchins sampled in 1996, but excluding this site resulted in F_{ST} values that did not differ significantly from zero. This recurrent genetic outliers in the lee of Pt. Reyes suggests some dispersal anomaly associated with this area, and warrants further study.

Like previous studies of genetic structure in marine species (e.g., Johnson and Black 1982; Johnson and Black 1984a, b; Watts et al. 1990; Kordos and Burton 1993; Edmands et al. 1996; Moberg and Burton 2000), estimates of population subdivision among *P. cinctipes* at sites separated by only tens of kilometers often exceed those between sites separated by many hundreds of kilometers. For example, the Bodega Bay and Dillon Beach sites are separated by only 10 km, and yet estimates of θ (0.13) and ρ (0.29) both exceed the estimate for population subdivision across the entire species range ($\theta = 0.09$, $\rho = 0.13$). Such a pattern could be generated by high variance and sweepstakes success of reproduction among adults (e.g., Hedgecock 1994a), variation in source populations from which recruits originate (e.g., Kordos and Burton 1993), or pre- or post-settlement natural selection (e.g., Koehn et al. 1980; Johnson and Black 1982; Johnson and Black 1984a, b; Hilbish and Koehn 1985).

Because the microsatellite loci used in this study are anonymous, it is conceivable that they are within, or linked to, a gene under natural selection that varies in space or time, producing a pattern consistent with simple population differentiation. The parallel changes among loci in the number of alleles at each site revealed by this study may indicate natural selection, recent range expansion, or local population bottlenecks (e.g., England et al. 1996; Houlden et al. 1996; Marko 1998; Luikart et al. 1999; Ramey et al. 2000; Wares et al. 2001). Unfortunately, given our ignorance of the potential selective forces, and the genes (if any) to which these anonymous markers might be linked, it is impossible to eliminate spatially or temporally variable selection as a force influencing genetic structure in this species. Nevertheless, the data fit the Ewen's sampling distribution. Moreover, there is no consistent pattern of sharp genetic discontinuity (e.g., Reeb and Avise 1990; Karl and Avise 1992) or geographic clines (e.g., Koehn et al. 1980) to suggest that selection shapes genetic structure of *P. cinctipes* populations. Therefore, it seems likely that factors other than natural selection must contribute to genetic structure in *P. cinctipes*.

Likewise, high variance and sweepstakes success of reproduction among adults seems an unlikely explanation for generating genetic structure among populations of *P. cinctipes*. Females brood less than 1000 offspring on average, and the sweepstakes model is generally thought to apply only to highly fecund animals such as urchins or oysters (Hedgecock 1994a), which average hundreds of thousands of offspring per spawn.

Variation in the source population from which recruits derive is impossible to address without knowing the genetic structure of the recruits, and there are no recruits included in this study. However, for this hypothesis to be plausible, the degree of genetic differentiation among populations must be relatively high (Kordos and Burton 1993). Comparisons of F_{ST} values from study-to-study are complicated by differences in the choice of genetic markers used, sampling scale and arrangement, hence it is difficult to evaluate the impact of these differences on such comparisons (Bossart and Prowell 1998; Whitlock and McCauley 1999). In the studies most similar to this one in terms of sampling sites, scale and arrangement, overall estimates of F_{ST} for species with pelagic larvae ranged from roughly 0.001 to 0.05, whereas species with crawl-away larvae ranged from 0.1 to 0.3 (Hellberg 1995; Edmands et al. 1996; Hellberg 1996; Moberg and Burton 2000). To complicate the comparison further, each of these studies were based on allozymes, whereas this study uses microsatellite markers. Comparative studies using estimates of F_{ST} based on allozymes and on microsatellites have in some cases found estimates derived from allozymes to be significantly higher (e.g., Scribner et al. 1994; Lehmann et al. 1996), in other cases significantly lower (e.g., Pihkanen et al. 1996; Sanchez et al. 1996), and yet others to be equivalent (e.g., Seppa and Gertsch 1996). Given these differences, comparisons of F_{ST} values among these studies should be considered tenuous. Furthermore, it is important to remember that the central California and Washington State samples were not collected in the same year as the northern California samples. This temporal separation of samples may also affect estimates of genetic distance between sites, and the relative contribution of spatial and

temporal separation among samples is considered in greater detail in Chapter 3. With these caveats, when compared to other north-western Pacific species, the magnitude of population subdivision among even northern California *P. cinctipes* is much higher than expected for a species with long-lived planktonic larvae. Overall, the average amount of gene flow decreased significantly with distance, and 18% of the variation in gene flow among populations is explained by a simple linear regression with geographic distance of separation. Despite the potential for larvae to disperse across great distances, there remains significant genetic structure among geographic sites.

Given that *P. cinctipes* inhabits a very narrow (~2m) band of the rocky intertidal along the North American coastline, the dispersal of this species is likely to be as close to a theoretical one-dimensional linear array of populations as any in nature. The predicted slope from simulations that relate M to geographic distance for a one-dimensional stepping stone model is -1.0, but neither F_{ST} or R_{ST} estimates include -1 within the 95% or 99% confidence interval calculated for the slope of these regressions. Instead, the relationship between the log-transformed values of M and geographic distance for *P. cinctipes* more closely matches the predictions of a two-dimensional stepping-stone model (Slatkin and Maddison 1990; Slatkin 1993; Hellberg 1994). Inferences of gene flow from F -statistics (θ or ρ) rest on a number of assumptions (Hellberg 1994), most of which are very difficult to test. A departure from selective neutrality, random mating, genetic equilibrium, equivalent population sizes, or exchange rates could all conceivably affect this estimate. Fortunately, such calculations are relatively insensitive to minor violations of these assumptions (Slatkin and Barton

1989; Hutchison and Templeton 1999), and loci that fit the Ewen's sampling distribution should not significantly bias these calculations (N. Barton, pers. comm.).

With the caveat that it is the relative magnitude of estimates of M that are of interest, rather than the absolute values (Hutchison and Templeton 1999; Whitlock and McCauley 1999), calculation of this statistic allows me to consider hypotheses about the role of geographic separation in the observed pattern of population subdivision. The simple linear regression between $\log_{10} M$ and $\log_{10}$ geographic distance is significant for only those values estimated from the IAM- (θ) but not the SMM-based (ρ) versions of Wright's hierarchical F -statistics. Many empirical studies have now found that microsatellites appear to fit the IAM-based expectations as well as or better than the SMM (e.g., Estoup et al. 1995; Garcia De Leon et al. 1997; O'Connell et al. 1997; Ruzzante 1998). Regardless, a quadratic regression gives a best fit to the data, and doubles the proportion of the variation explained for estimates derived from both estimators (θ and ρ). The improved parabolic fit makes the regression significant for both M_θ and M_ρ . This curvilinear relationship between the log of M and distance of separation suggests that pelagic larvae of *P. cinctipes* settle with greatest frequency at an intermediate distance from the natal site. Although such a relationship has not been previously demonstrated for any species with long-lived planktonic larvae, it is intuitive. All things being equal, long-lived pelagic larvae are just as unlikely to settle very close to the site of release as they are to settle very far away from it.

The results presented herein suggest that variation in source populations from which recruits originate and pre- or post-settlement natural selection are the most likely

source of genetic variation among populations of *P. cinctipes*. To address these hypotheses, however, patterns of allelic identity among recruits and adults at each site need to be quantified over time. Previous studies of population structure among urchins of the genus *Strongylocentrotus* (Edmands et al. 1996; Moberg and Burton 2000) reveal significant temporal genetic variation within sites and among size classes at each site. Moberg and Burton (2000) emphasize that data such as these do not constitute a strong test of the upwelling-relaxation flow hypothesis, however, because the temporal and spatial scale of sampling for recruits may not be fine enough. They point out that the predictions of the upwelling-relaxation flow hypothesis may only apply at the temporal scale of individual relaxation events, and not over the scale of seasons or years. However, if these retention zones associated with headlands are leaky (i.e., a small proportion of larvae are exchanged regularly across headlands), there is no reason to expect any genetic signature of retention zones (reviewed by Hedgecock 1994a, b; Burton 1996; Grosberg and Cunningham 2001). Furthermore, the strength of retention zones likely varies from year-to-year. A single year of low intensity upwelling could lead to gene flow across headlands; if this occurs regularly within the lifespan of these crabs (i.e., at least once every 8 years or so), this periodic mixing should erase any signature that may have developed in the interim. Thus, a more detailed study that includes temporal samples of age classes at each site is needed to examine adequately the role of retention zones in contributing to the genetic structure observed in species with extensive planktonic larval periods, such as *P. cinctipes*. Likewise, inferences about individual reproductive success, larval habitat choice and mixing are impossible

without data regarding the genotypic identity of recruits through time. These questions are addressed in detail in Chapter 3.

Together, these data show that retention zones associated with headlands cannot account for the pattern of genetic structure observed among geographic populations of *Petrolisthes cinctipes*. Predictions based on oceanographic processes that prove important determinants of demographic structure (Wing et al. 1995a; Wing et al. 1995b; Wing et al. 1998; Morgan et al. 2000) are not consistent with the observed pattern of genetic structure among these geographic populations. This suggests that retention zones are sufficiently leaky or impermanent that, even using microsatellite markers, they do not leave a detectable genetic signature. Thus, simple models of gene flow available presently do not adequately account for the complexity of marine larval dispersal, and suggests that much remains to be learned about the processes structuring marine invertebrate populations.

Table 1. Summary of the four polymorphic microsatellite loci isolated from *Petrolisthes cinctipes* originally used in this study. Forward and reverse primer sequences are given in Table 1. Shown are the fluorescent dye with which the forward primer for each locus was labelled, the size range of the amplification products, the motif of the original cloned microsatellite sequence, and the number of alleles observed in the sample of 1010 crabs genotyped in this study. The number of crabs for which the locus was scorable (N), the observed and expected heterozygosities ($\pm$ SE) for each locus across all populations, and the probability value ($\pm$ SE) for conformity to Hardy-Weinberg equilibrium are also presented. Tests of H-W (pooled across all sites) were done using a Markov-chain length of 500,000 steps and a 10,000 step dememorization for each locus. After discovering that locus Pc160b and locus Pc36L are not inherited in a Mendelian fashion (see Chapter 1), these loci were excluded from all subsequent analyses.

Locus	Dye label	Locus size range (bp)	Cloned repeat motif	No. of alleles	N	Heterozygosity		H-W <i>P</i> -value
						Observed	Expected	
Pc156s	HEX	72 - 124	(TG) ₂₆ TA(TG) ₁₁	26	1010	0.67	0.65	0.25 ($\pm$ 0.005)
Pc170s	6-FAM	86 - 182	TCTAGA (TGT) ₆ TTGTAA TCTAGA (TGT) ₆ ... (TG) ₄	35	1010	0.77	0.71	0.19 ($\pm$ 0.001)
Pc160b	NED	272 - 408	(TG) ₁₆	55	959	0.78	0.95	0.09 ($\pm$ 0.002)
Pc36L	6-FAM	358 - 474	(TTTTA) ₂ TTTA (TTAA) ₂ (TTAT) ₃ TTGT TTAT C (TATT) ₃ TTTT	39	854	0.59	0.85	0.004 ($\pm$ 0.001)

Table 2. Summary of microsatellite data collected for each geographic site sampled, arranged north to south corresponding to Figure 1. The number of individual crabs scored at each site is presented (N) along with the observed and expected heterozygosities for each locus, and the probability value for conformity to Hardy-Weinberg equilibrium (see text).

Geographic location	Locus	N	Heterozygosity		H-W p -value
			Observed	Expected	
Neah Bay	Pc156s	90	0.87	0.50	< 0.001
	Pc170s		0.57	0.60	0.12
Crescent City	Pc156s	75	0.64	0.71	0.34
	Pc170s		0.83	0.72	0.17
Cape Mendocino	Pc156s	89	0.75	0.72	0.58
	Pc170s		0.91	0.85	0.52
Shelter Cove	Pc156s	90	0.78	0.82	0.29
	Pc170s		0.76	0.76	0.34
Ft. Bragg	Pc156s	95	0.52	0.79	0.10
	Pc170s		0.80	0.67	0.72
Albion	Pc156s	90	0.33	0.35	0.58
	Pc170s		0.64	0.77	< 0.001
Pt. Arena	Pc156s	90	0.48	0.47	0.61
	Pc170s		0.64	0.73	0.11
Salt Point	Pc156s	90	0.79	0.81	0.18
	Pc170s		0.79	0.78	0.06
Bodega	Pc156s	79	0.53	0.60	0.56
	Pc170s		0.75	0.85	0.33
Dillon Beach	Pc156s	78	0.83	0.83	0.09
	Pc170s		0.87	0.57	< 0.001
Monterey Bay	Pc156s	90	0.70	0.65	< 0.001
	Pc170s		0.77	0.63	< 0.001
Morro Bay	Pc156s	54	0.87	0.54	< 0.001
	Pc170s		0.94	0.63	< 0.001

Table 3. Weir and Cockerham (1984) estimates of Wright's F -statistics calculated separately for each locus scored from *Petrolisthes cinctipes* collected from 12 geographic locations throughout its range. Means and standard errors (SE) for each value were obtained by jackknifing over geographic populations.

Locus	$F = F_{IT}$ (within total)	$\theta = F_{ST}$ (among sites)	$f = F_{IS}$ (within sites)
Pc156s	0.07	0.10	0.01
Pc170s	0.05	0.12	-0.05
Mean	0.06	0.11	-0.03
SE	0.003	0.003	0.005

Table 4. Matrix of estimators of population subdivision (θ (Weir and Cockerham 1984) below the diagonal and ρ (Goodman 1997) above it) between pairs of geographic sites from which *Petrolisthes cinctipes* were collected. Sites are arranged from north to south, corresponding to Figure 1.

Site	Neah Bay	Crescent City	Cape Mendocino	Shelter Cove	Ft. Bragg	Albion	Pt. Arena	Salt Point	Bodega Bay	Dillon Beach	Monterey Bay	Morro Bay
Neah Bay	—	0.27	0.23	0.32	0.23	0.03	0.06	0.28	0.15	0.53	0.09	-0.01
Crescent City	0.26	—	0.05	0.02	0.01	0.16	0.14	-0.01	0.10	0.07	0.03	0.17
Cape Mendocino	0.20	0.01	—	0.03	0.09	0.14	0.10	0.05	0.02	0.19	0.07	0.16
Shelter Cove	0.24	0.02	0.02	—	0.06	0.21	0.18	0.03	0.09	0.08	0.08	0.22
Ft. Bragg	0.25	0.04	0.04	0.01	—	0.12	0.12	0.01	0.11	0.10	0.01	0.14
Albion	0.19	0.08	0.06	0.13	0.15	—	0.01	0.15	0.07	0.39	0.04	0.01
Pt. Arena	0.13	0.10	0.07	0.14	0.16	0.01	—	0.13	0.04	0.37	0.04	0.02
Salt Point	0.18	0.01	0.01	0.01	0.03	0.07	0.07	—	0.09	0.10	0.02	0.16
Bodega	0.14	0.05	0.02	0.07	0.09	0.02	0.02	0.03	—	0.29	0.07	0.09
Dillon Beach	0.31	0.07	0.08	0.04	0.03	0.22	0.23	0.06	0.16	—	0.16	0.37
Monterey Bay	0.27	0.08	0.07	0.07	0.06	0.16	0.17	0.07	0.09	0.13	—	0.05
Morro Bay	0.26	0.13	0.10	0.12	0.09	0.19	0.19	0.11	0.11	0.18	0.02	—

Table 5. Matrix of estimators of genetic distances between pairs of geographic sites from which *Petrolisthes cinctipes* were collected. Values represent $(\delta\mu)^2$ (Goldstein et al. 1995) below the diagonal, and D_{c-E} (Cavalli-Sforza and Edwards 1967) above it (see text). Sites are arranged from north to south, corresponding to Figure 1.

Site	Neah Bay	Crescent City	Cape Mendocino	Shelter Cove	Ft. Bragg	Albion	Pt. Arena	Salt Point	Bodega Bay	Dillon Beach	Monterey Bay	Morro Bay
Neah Bay	—	0.47	0.40	0.47	0.48	0.33	0.32	0.39	0.33	0.40	0.46	0.45
Crescent City	32.71	—	0.13	0.13	0.14	0.20	0.23	0.12	0.17	0.14	0.20	0.27
Cape Mendocino	27.30	9.26	—	0.16	0.18	0.17	0.21	0.10	0.13	0.15	0.22	0.25
Shelter Cove	46.97	3.47	6.28	—	0.07	0.27	0.30	0.18	0.22	0.10	0.17	0.23
Ft. Bragg	24.75	1.85	14.73	10.33	—	0.28	0.31	0.20	0.23	0.11	0.17	0.23
Albion	1.12	21.87	18.19	33.60	16.07	—	0.07	0.14	0.10	0.24	0.28	0.31
Pt. Arena	2.47	19.49	13.37	28.69	15.50	0.46	—	0.17	0.12	0.26	0.30	0.33
Salt Point	27.34	0.25	8.75	4.94	1.06	17.56	15.61	—	0.10	0.16	0.25	0.30
Bodega	12.66	17.09	3.90	18.46	18.77	7.73	4.42	14.61	—	0.18	0.24	0.28
Dillon Beach	80.84	11.97	37.11	14.01	16.27	63.87	61.10	15.07	56.80	—	0.18	0.25
Monterey Bay	10.56	7.21	14.25	18.07	3.00	5.36	5.59	4.80	11.67	33.20	—	0.11
Morro Bay	0.06	32.69	28.69	47.75	24.22	1.26	2.90	27.29	13.97	80.07	10.18	—

Table 6. Hierarchical analysis of molecular variance (AMOVA) results for *Petrolisthes cinctipes* under three hypotheses of population structure. Sites (in numerical order from north to south, Fig .1) were grouped to assess the proportion of observed genetic variation that could be explained by each of these hypotheses. Probability of obtaining a more extreme variance components and Φ statistic by chance alone was determined by permutation test in Arlequin (500,000), and significant values ($\alpha = 0.01$) are denoted by an asterisk.

Hypothesis 1: Upwelling-relaxation flow hypothesis (Oceanic gyres associated with headlands restrict gene flow)					
Population Grouping	Source of Variation	d.f.	Variance component	% of variation	Φ statistic
(1); (2,3); (4,5,6); (7,8,9,10); (11,12)	Among embayments	4	0.027	3.55	$\Phi_{CT} = 0.04$
	Among sites, within embayments	7	0.059	7.66	$\Phi_{SC} = 0.08^*$
	Within sites	2008	0.68	88.80	$\Phi_{ST} = 0.11^*$
Hypothesis 2: Simple geographic hypothesis (geographically proximate sites in central California are linked genetically, irrespective of embayments)					
(1); (2,3,4,5,6, 7,8,9,10); (11,12)	Among geographic regions	2	0.078	9.65	$\Phi_{CT} = 0.10^*$
	Among sites, within regions	9	0.049	6.05	$\Phi_{SC} = 0.07^*$
	Within sites	2008	0.68	84.31	$\Phi_{ST} = 0.16^*$
(1); (2,3,4,5,6, 7,8,9); (11,12)	Among geographic regions	2	0.081	10.10	$\Phi_{CT} = 0.10^*$
	Among sites, within regions	8	0.039	4.92	$\Phi_{SC} = 0.05^*$
	Within populations	1853	0.679	84.98	$\Phi_{ST} = 0.15^*$
Hypothesis 3: Migration does not occur in a stepwise fashion (post-hoc grouping of sites based on UPGMA clustering pattern)					
(1); (2,3,4,5,8); (6,7,9); (10); (11,12)	Among clusters	4	0.088	11.30	$\Phi_{CT} = 0.13^*$
	Among sites, within clusters	7	0.013	1.62	$\Phi_{SC} = 0.02^*$
	Within sites	2008	0.680	87.08	$\Phi_{ST} = 0.11^*$

1) Neah Bay, WA

48:21N, 124:36W

2) Crescent City

41:45N, 124:11W

3) Cape Mendocino

40:27N, 124:21W

4) Shelter Cove

40:01N, 124:04W

5) Ft. Bragg

39:26N, 123:48W

6) Albion

39:13N, 123:41W

7) Pt. Arena

38:54N, 123:41W

8) Salt Point

38:30N, 123:14W

9) Bodega Bay

38:19N, 123:01W

10) Dillon Beach

38:15N, 122:57W

11) Monterey Bay

36:37N, 121:49W

12) Morro Bay

35:22N, 120:51W

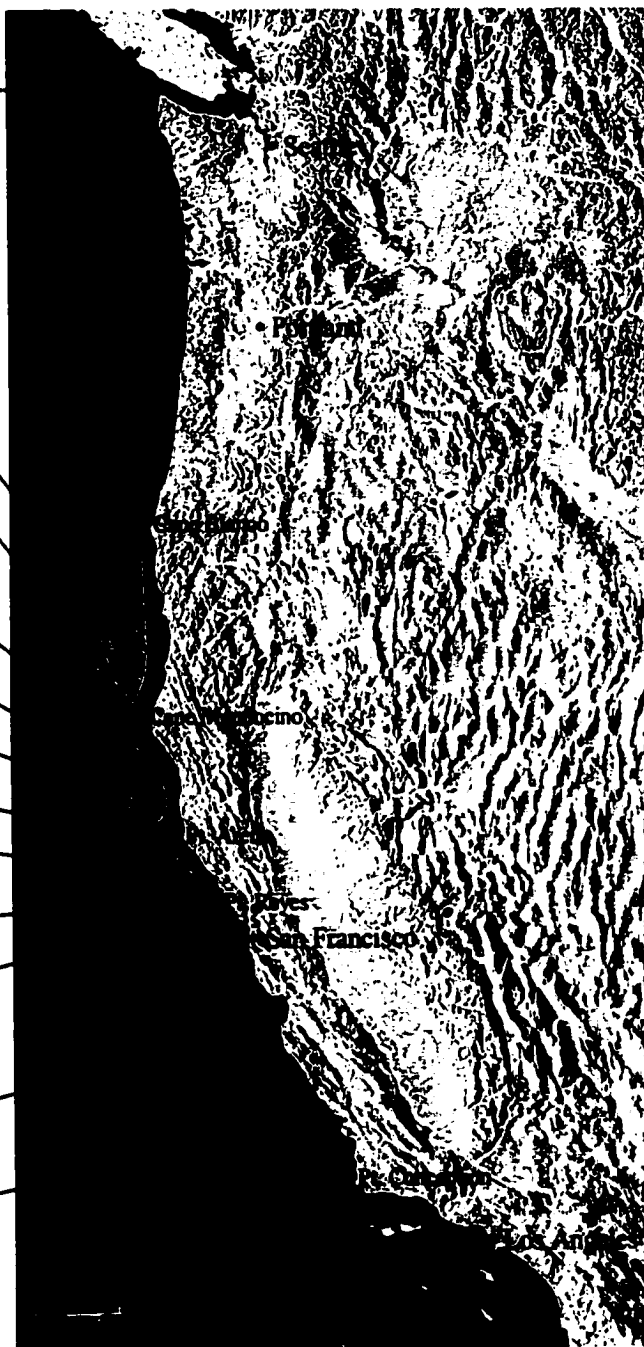


Figure 1. Collection sites for the crab, *Petrolisthes cinctipes*. Each site is numbered from north to south, and the GPS coordinates for each is presented below the location. Retention zones associated with headlands at Cape Mendocino, Pt. Arena and Pt. Reyes are indicated by curled arrows. Site markers that pass through each arrow are those within the same embayment.

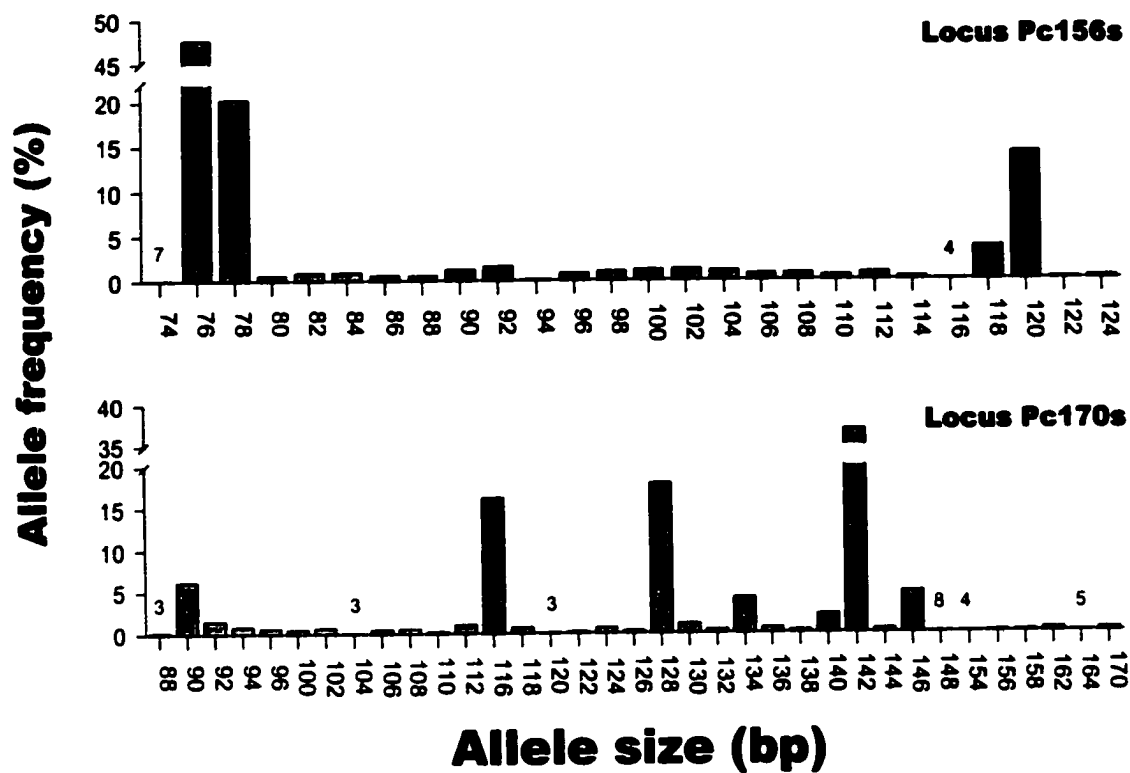


Figure 2. Allele frequency distributions of microsatellite loci amplified from the 1010 individual *Petrolisthes cinctipes* genotyped in this study. For both loci, the three most common alleles represented more than 80% of those scored. Private alleles (those detected in individuals from only a single site) are identified with the number of the site, corresponding to Figure 1.

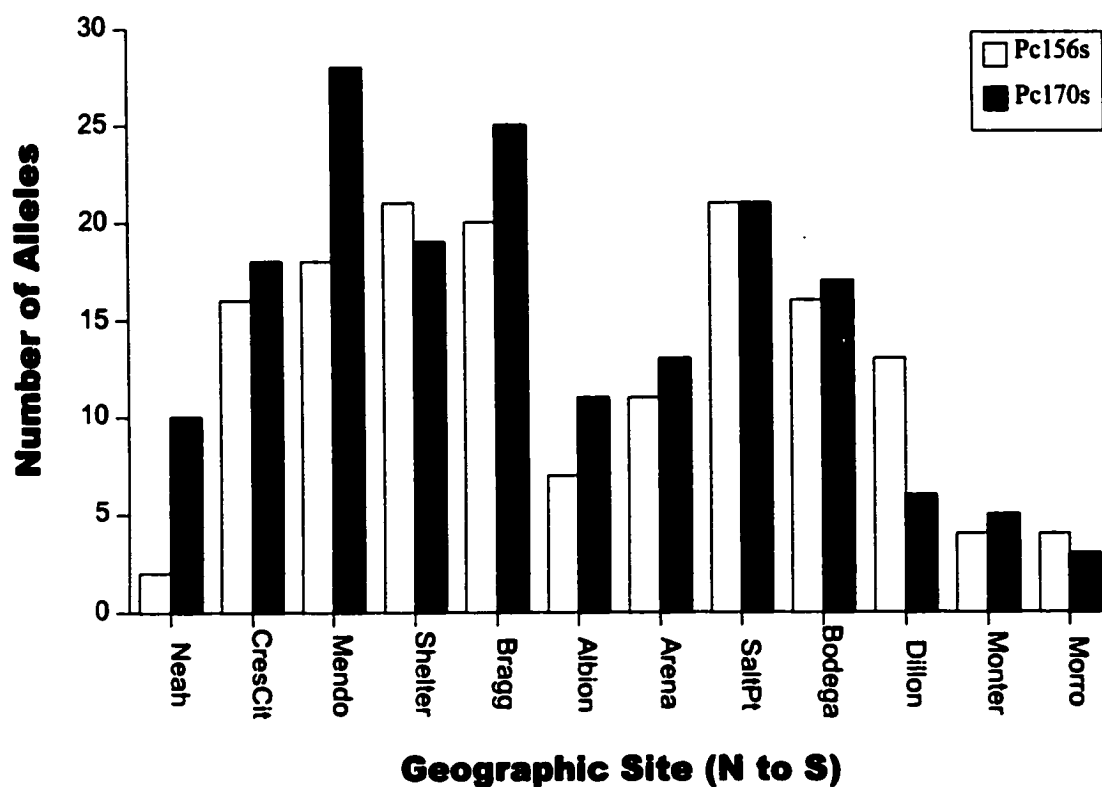


Figure 3. Number of alleles detected at each locus among samples of *Petrolisthes cinctipes*. Sites are presented from north to south, corresponding to Figure 1.

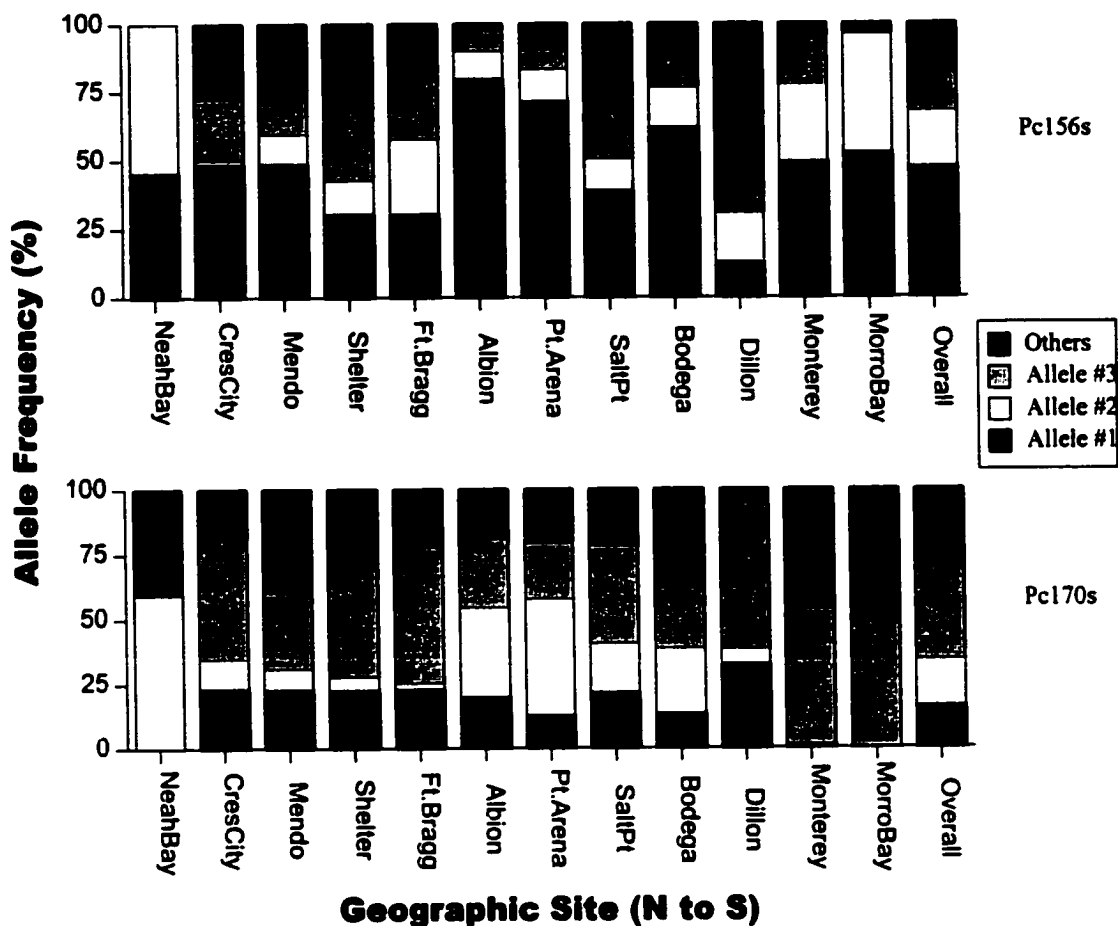


Figure 4. Allele frequencies of the three most common alleles at each locus for which *Petrolisthes cinctipes* were scored. The most common allele is listed as Allele #1, the second most common as Allele #2, and the third as Allele #3. All other alleles are lumped into the “Others” category for presentation. Alleles #1, 2 & 3 are 76, 78, and 120 bp, respectively for locus Pc156s (upper plot), and 142, 128, and 116 for locus Pc170s (lower plot).

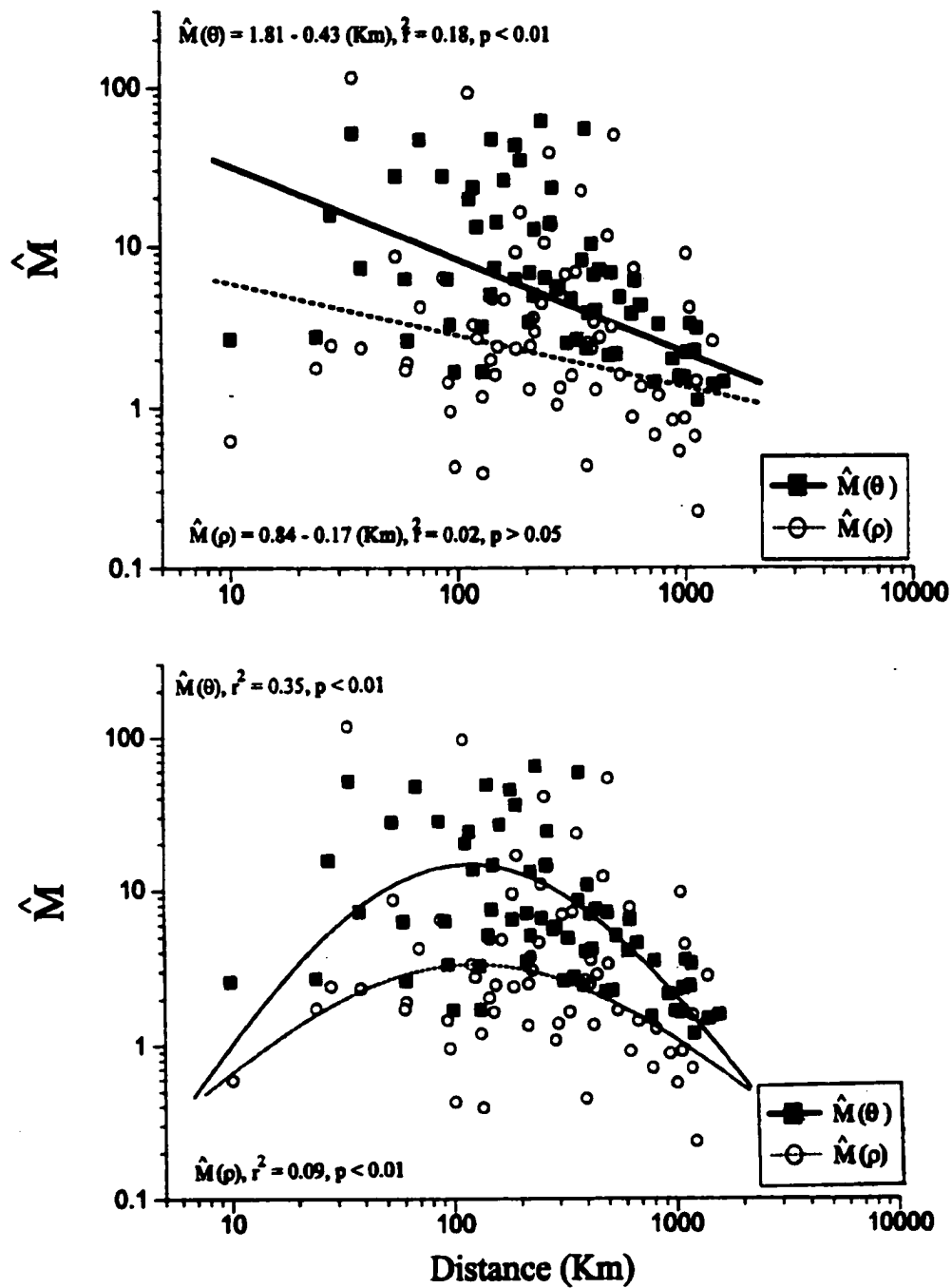


Figure 5. Log-log plot of pair-wise estimates of gene flow ($\hat{M}$) and geographic distance of separation between sites from which *Petrolisthes cinctipes* were collected. Linear regression equations (upper plot) explain significantly less of the variation than do the quadratic regressions (lower plot) for estimates of $\hat{M}$ derived from both θ and ρ .

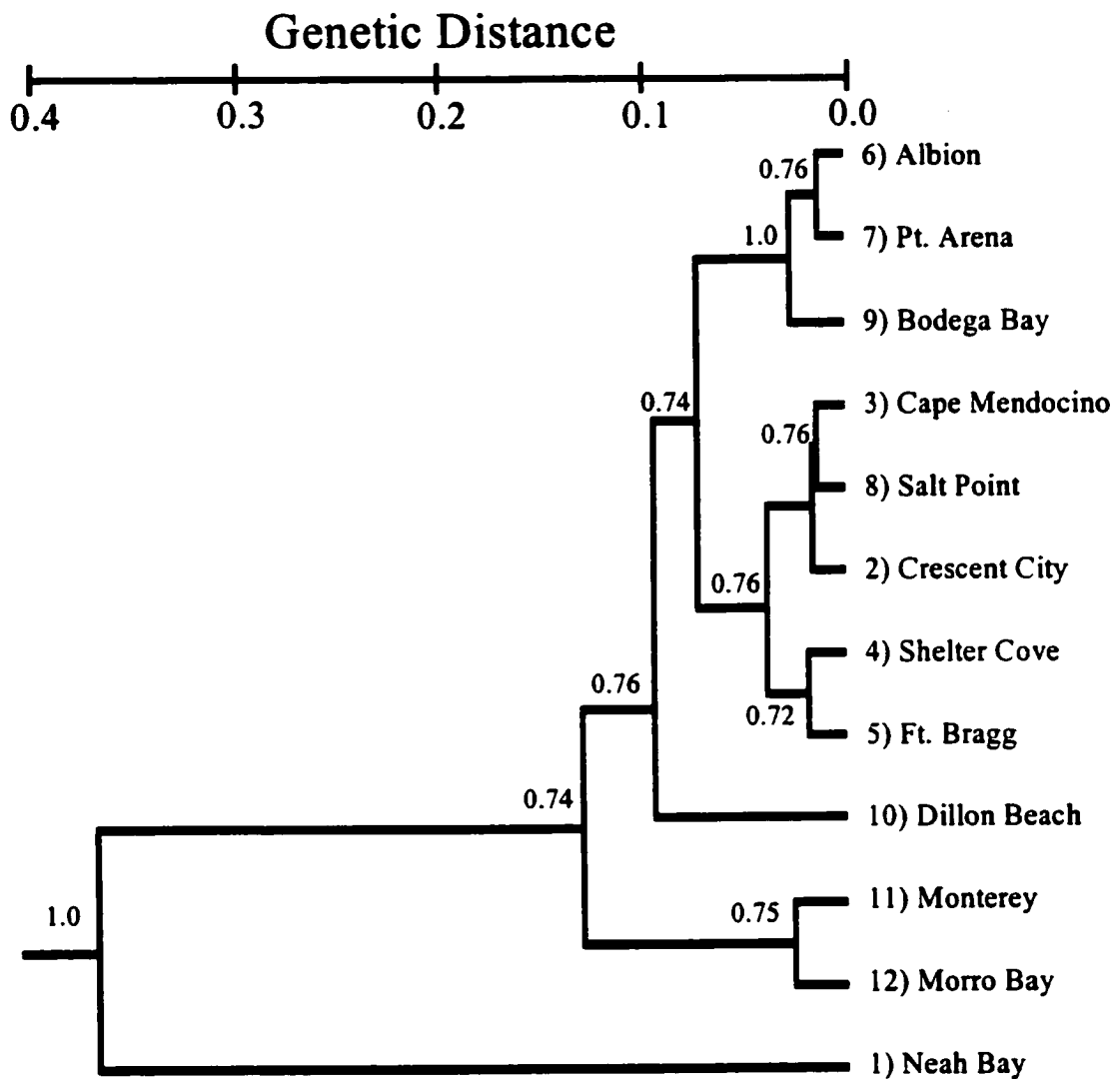


Figure 6. UPGMA dendrogram for adult *Petrolisthes cinctipes* sampled at each geographic site, based on the D_A genetic distance of Nei et al (1983). Sites are numbered from north to south, corresponding to Figure 1, and values from 10,000 bootstrap replicates are presented next to each node. The same UPGMA dendrogram topology is produced using most common measures of genetic distance (see text).

Literature Cited

- Avise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. 511 pp.
Chapman & Hall, New York, NY.
- Benzie, J. A. H., and S. T. Williams. 1997. Genetic structure of giant clam (*Tridacna maxima*) populations in the West Pacific is not consistent with dispersal by present-day ocean currents. *Evolution* 51:768-783.
- Bertness, M. D., S. D. Gaines, E. G. Stephens, and P. O. Yund. 1992. Components of recruitment in populations of the acorn barnacle *Semibalanus balanoides* Linnaeus. *Journal of Experimental Marine Biology and Ecology* 156:199-215.
- Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 74:21-45.
- Bossart, J. L., and D. P. Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution* 13:202-206.
- Botsford, L. W., C. L. Moloney, A. Hastings, J. L. Largier, T. M. Powell, K. Higgins, and J. F. Quinn. 1994. The influence of spatially and temporally varying oceanographic conditions on meroplanktonic metapopulations. *Deep-Sea Research Part II Topical Studies in Oceanography* 41:107-145.
- Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455-457.

- Burton, R. S. 1983. Protein polymorphisms and genetic differentiation of marine invertebrate populations. *Marine Biology Letters* 4:193-206.
- Burton, R. S. 1994. Inferring the genetic structure of marine populations: a case study comparing allozyme and DNA sequence data. *CalCOFI Report* 35:52-60.
- Burton, R. S. 1996. Molecular tools in marine ecology. *Journal of Experimental Marine Biology and Ecology* 200:85-101.
- Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, J. C. Mulley, and G. R. Sutherland. 1993. Incidence and Origin of Null Alleles in the Acn Microsatellite Markers. *American Journal of Human Genetics* 52:922-927.
- Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 32:550-570.
- Chambers, G. K., and E. S. MacAvoy. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology, Part B* 126:455 - 476.
- Deka, R., R. Chakraborty, and R. E. Ferrell. 1991. A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics* 11:83 - 92.
- Edmands, S., P. E. Moberg, and R. S. Burton. 1996. Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. *Marine Biology* 126:443-450.
- Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241 - 253.

- Edwards, A. W. F., and L. L. Cavalli-Sforza. 1965. A method for cluster analysis. *Biometrics* 21:362--375.
- England, P. R., D. A. Briscoe, and R. Frankham. 1996. Microsatellite polymorphisms in a wild population of *Drosophila melanogaster*. *Genetical Research* 67:285-290.
- Estoup, A., L. Garnery, M. Solignac, and J. M. Cornuet. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* 140:679-695.
- Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Feldheim, K. A., S. H. Gruber, and M. V. Ashley. 2001. Population genetic structure of the lemon shark (*Negaprion brevirostris*) in the western Atlantic: DNA microsatellite variation. *Molecular Ecology* 10:295 - 303.
- Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Department of Genetics, University of Washington, Seattle, WA. Available at <http://evolution.genetics.washington.edu/phylip.html>
- Gaines, S., S. Brown, and J. Roughgarden. 1985. Spatial variation in larval concentrations as a cause of spatial variation in settlement for the barnacle, *Balanus glandula*. *Oecologia* 67:267-272.
- Gaines, S. D., and M. D. Bertness. 1992. Dispersal of Juveniles and Variable Recruitment in Sessile Marine Species. *Nature* 360:579-580.

- Garcia De Leon, F. J., L. Chikhi, and F. Bonhomme. 1997. Microsatellite polymorphism and population subdivision in natural populations of European sea bass *Dicentrarchus labrax* (Linnaeus, 1758). *Molecular Ecology* 6:51-62.
- Gillespie, J. H. 1986. Rates of molecular evolution. *Annual Reviews of Ecology and Systematics* 17:637 - 665.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463-471.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* 92:6723-6727.
- Goldstein, D. B., and C. Schlötterer. 1999. *Microsatellites: Evolution and Applications*. 352 pp. Oxford University Press, Oxford.
- Goodman, S. J. 1997. R-ST Calc: A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6:881-885.
- Goudet, J., M. Raymond, T. de Meeiis, and F. Rousset. 1996. Testing differentiation in diploid populations. *Genetics* 144:1933-1940.
- Grosberg, R. K., and C. W. Cunningham. 2001. Genetic Structure in the Sea: From Populations to Communities. Pp. 61-84 in M. D. Bertness, S. D. Gaines and M.

- E. Hay, eds. *Marine Community Ecology*. Sinauer Associates, Inc., Sunderland, MA.
- Grosberg, R. K., D. R. Levitan, and B. B. Cameron. 1996. Characterization of genetic structure and genealogies using RAPD-PCR markers: A random primer for the novice and nervous. Pp. 67-100 *in* Ferraris, J. D. and S. R. Palumbi, eds. *Molecular zoology: Advances, strategies, and protocols*. Wiley-Liss, Inc., New York, NY.
- Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48:361 - 372.
- Haig, J. 1960. The Porcellanidae (Crustacea: Anomura) of the Eastern Pacific. *Alan Hancock Pacific Expeditions* 24:1 - 440.
- Hayward, T. L., and A. Mantyla. 1990. Physical, chemical and biological structure of a coastal eddy near Cape Mendocino. *Journal of Marine Research* 48:825-850.
- Hedgecock, D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science* 39:550-564.
- Hedgecock, D. 1994a. Does variance in reproductive success limit effective population sizes of marine organisms? Pp. 122-134 *in* A.R. Beaumont, ed. *Genetics and Evolution of Aquatic Organisms*. Chapman & Hall, London, UK.
- Hedgecock, D. 1994b. Population genetics of marine organisms. *U.S. Globec News*, No. 6:1-3,11

- Hellberg, M. E. 1994. Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, *Balanophyllia elegans*. *Evolution* 48:1829 - 1854.
- Hellberg, M. E. 1995. Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: Equilibrium and nonequilibrium at different spatial scales. *Marine Biology* 123:573-581.
- Hellberg, M. E. 1996. Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. *Evolution* 50:1167-1175.
- Hilbish, T. J., and R. K. Koehn. 1985. Exclusion of the role of secondary contact in an allele frequency cline in the mussel *Mytilus edulis*. *Evolution* 39:432-443.
- Houlden, B. A., P. R. England, A. C. Taylor, W. D. Greville, and W. B. Sherwin. 1996. Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck. *Molecular Ecology* 5:269-281.
- Hughes, S. 1998. STRand Nucleic Acid Analysis Software. Regents of the University of California, Davis. Available at <http://www.vgl.ucdavis.edu/STRand>
- Hutchison, D. W., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898-1914.
- Huyer, A. 1983. Coastal upwelling in the California Current system. *Progress in Oceanography* 12:259-284.

- Jensen, G. C. 1989. Gregarious settlement by megalopae of the porcelain crabs *Petrolisthes cinctipes* (Randall) and *P. eriomerus* Stimpson. *Journal of Experimental Marine Biology and Ecology* 131:223-231.
- Jensen, G. C. 1990. Intertidal zonation of porcelain crabs: Resource partitioning and the role of selective settlement. Ph.D. Dissertation, Zoology. University of Washington, Seattle.
- Jensen, G. C. 1991a. Competency, settling behavior, and postsettlement aggregation by porcelain crabs megalopae (Anomura: Porcellanidae). *Journal of Experimental Marine Biology and Ecology* 153:49 - 61.
- Jensen, G. C. 1991b. Gregarious settlement by porcelain crabs, *Petrolisthes* (Anomura: Porcellanidae). *Memoirs of the Queensland Museum* 31:390.
- Jensen, G. C., and D. A. Armstrong. 1991. Intertidal zonation among congeners: Factors regulating distribution of porcelain crabs *Petrolisthes* spp. (Anomura: Porcellanidae). *Marine Ecology Progress Series* 73:47-60.
- Jensen, P. C., J. M. Orensanz, and D. A. Armstrong. 1996. Structure of the female reproductive tract in the Dungeness crab (*Cancer magister*) and implications for the mating system. *Biological Bulletin* 190:336-349.
- Johnson, M. P., and R. Black. 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* 70:157-164.
- Johnson, M. S., and R. Black. 1984a. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intretidal limpet. *Evolution* 38:1371-1383.

- Johnson, M. S., and R. Black. 1984b. The Wahlund effect and the geographical scale of variation in the intertidal limpet *Siphonaria* sp. *Marine Biology* 79:295-302.
- Karl, S. A., and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. *Science* 256:100 - 102.
- Knudsen, J. W. 1964. Observations of the reproductive cycles and ecology of the common Brachyura and crablike Anomura of Puget Sound, Washington. *Pacific Science* 18:3-33.
- Koehn, R. K., R. Milkman, and J. B. Mitton. 1976. Population genetics of marine pelecypods part 4: Selection, migration and genetic differentiation in the blue mussel, *Mytilus edulis*. *Evolution* 30:2-32.
- Kordos, L. M., and R. S. Burton. 1993. Genetic differentiation of Texas Gulf Coast populations of the blue crab *Callinectes sapidus*. *Marine Biology* 117:227-233.
- Kyle, C. J., and E. G. Boulding. 2000. Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Marine Biology* 137:835-845.
- Lacson, J. M., V. M. Riccardi, S. W. Calhoun, and D. C. Morizot. 1989. Genetic differentiation of bicolor damselfish (*Eupomacentrus partitus*) populations in the Florida Keys (USA). *Marine Biology* 103:445-452.
- Largier, J. L., B. A. Magnell, and C. D. Winant. 1993. Subtidal circulation over the northern California shelf. *Journal of Geophysical Research - Oceans* 98:18147-18179.

- Lehmann, T., W. A. Hawley, L. Kamau, D. Fontenille, F. Simard, and F. H. Collins. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: Comparison of microsatellite and allozyme loci. *Heredity* 77:192-200.
- Li, G., and D. Hedgecock. 1998. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variances in reproductive success. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1025-1033.
- Luikart, G., J.-M. Cornuet, and F. W. Allendorf. 1999. Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology* 13:523-530.
- Luikart, G., and P. R. England. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* 14:253 - 256.
- Lundquist, C. J., L. W. Botsford, L. E. Morgan, J. M. Diehl, T. Lee, D. R. Lockwood, and E. L. Pearson. 2000. Effects of El Niño and La Niña on local invertebrate settlement in northern California. *California Cooperative Oceanic Fisheries Investigations Reports* 41:167-176.
- Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. *Cancer Research* 27:209-220.
- Marko, P. 1998. Historical allopatry and the biogeography of speciation in the Prosobranch snail genus *Nucella*. *Evolution* 52:757-774.

- McCauley, D. E. 1993. Genetic consequences of extinction and recolonization in fragmented habitats. Pp. 217-233 in P. M. Kareiva, J. G. Kingsolver and R. B. Huey, eds. Biotic interactions and global change. Sinauer Associates, Inc., Sunderland, MA.
- Michener, C. D., and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130-162.
- Miller, M. P. 1997. Tools for Population Genetic Analyses (TFPGA): A Windows program for the analysis of allozyme and molecular population genetic data. Available at <http://www.public.asu.edu/~mmille8/tfpga.htm>
- Moberg, P. E., and R. S. Burton. 2000. Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. *Marine Biology* 136:773-784.
- Morgan, L. E., S. R. Wing, L. W. Botsford, C. J. Lundquist, and J. M. Diehl. 2000. Spatial variability in red sea urchin (*Strongylocentrotus franciscanus*) recruitment in northern California. *Fisheries Oceanography* 9:83-98.
- Nei, M. 1972. Genetic distance between populations. *American Naturalist* 106:283-292.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.
- Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 9:153-170.
- Nève, G., and E. Megléc. 2000. Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution* 15:376 - 377.

- Nyblade, C. F. 1987. Phylum or Subphylum Crustacea, Class Malacostraca, Order Decapoda, Anomura. Pp. 441 - 450 in M. F. Strathmann, ed. *Reproduction and development of marine invertebrates of the Northern Pacific Coast: data and methods or the study of eggs, embryos, and larvae*. University of Washington Press, Seattle, WA.
- O'Connell, M., R. G. Danzmann, J. M. Cornuet, J. M. Wright, and M. M. Ferguson. 1997. Differentiation of rainbow trout (*Oncorhynchus mykiss*) populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. *Canadian Journal of Fisheries and Aquatic Sciences* 54:1391-1399.
- Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology* 4:519-520.
- Palumbi, S. R. 1995. Using genetics as an indirect estimator of larval dispersal. Pp. 369-387 in L. R. McEdward, ed. *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, FL.
- Park, S. 2001. Microsatellite Toolkit for MS Excel. Smurfit Institute of Genetics, Trinity College, Dublin. Available at <http://acer.gen.tcd.ie/~sdepark/ms-toolkit/>
- Pemberton, J. M., J. Slate, D. R. Bancroft, and J. A. Barrett. 1995. Nonamplifying alleles at microsatellite loci - A caution for parentage and population studies. *Molecular Ecology* 4:249-252.
- Pihkanen, S., R. Vainola, and S. Varvio. 1996. Characterizing dog breed differentiation with microsatellite markers. *Animal Genetics* 27:343-346.

- Planes, S. 1993. Genetic differentiation in relation to restricted larval dispersal of the convict surgeonfish *Acanthurus triostegus* in French Polynesia. *Marine Ecology Progress Series* 98:237-246.
- Ramey, R. R., II, G. Luikart, and F. J. Singer. 2000. Genetic bottlenecks resulting from restoration efforts: The case of bighorn sheep in Badlands National Park. *Restoration Ecology* 8:85-90.
- Raymond, M., and F. Rousset. 1995. An exact test for population differentiation. *Evolution* 49:1280-1283.
- 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.
- Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.
- Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. Pp. 145-154. *Studies in Genetics VII*. University of Texas, Publication No. 7213, Austin, TX.
- Roughgarden, J., S. Gaines, and H. Possingham. 1988. Recruitment Dynamics in Complex Life Cycles. *Science* 241:1460-1466.
- Roughgarden, J., S. D. Gaines, and S. W. Pacala. 1987. Supply-side ecology: The role of physical transport processes. Pp. 491-518 in J. H. R. Gee and P. S. GILLER, eds. *Organization of communities: Past and present*. Blackwell Scientific Publications, London, GB.

- Ruzzante, D. E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1-14.
- Sainte-Marie, B., N. Urbani, J.-M. Sevigny, F. Hazel, and U. Kuhnlein. 1999. Multiple choice criteria and the dynamics of assortative mating during the first breeding season of female snow crab *Chionoecetes opilio* (Brachyura, Majidae). *Marine Ecology Progress Series* 181:141-153.
- Sanchez, J. A., C. Clabby, D. Ramos, G. Blanco, F. Flavin, E. Vazquez, and R. Powell. 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). *Heredity* 77:423-432.
- Scheltema, R. S. 1995. The relevance of passive dispersal for the biogeography of Caribbean mollusks. *American Malacological Bulletin* 11:99-115.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva. Available at <http://anthro.unige.ch/arlequin>
- Schwing, F. B., and D. M. Husby. 1991. Mesoscale oceanic response to wind events off central California in spring 1989: CTD surveys and AVHRR imagery. *CalCOFI Report* 32:1-16.
- Scribner, K. T., J. W. Arntzen, and T. Burke. 1994. Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. *Molecular Biology and Evolution* 11:737-748.

- Send, U., R. C. Beardsley, and C. D. Winant. 1987. Relaxation from upwelling in the Coastal Ocean Dynamics Experiment. *Journal of Geophysical Research - Oceans* 92:1683-1698.
- Seppa, P., and P. Gertsch. 1996. Genetic relatedness in the ant *Camponotus herculeanus*. A comparison of estimates from allozyme and DNA microsatellite markers. *Insectes Sociaux* 43:235-246.
- Shriver, M. D., L. Jin, R. Chakraborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the SMM: A computer simulation approach. *Genetics* 134:983-993.
- Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47:264-279.
- Slatkin, M. 1994. An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* 64:71-74.
- Slatkin, M. 1995. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics* 139:457-462.
- Slatkin, M. 1996. A correction to the exact test based on the Ewens sampling distribution. *Genetical Research* 68:259 - 260.
- Slatkin, M. 1997. Enumerate. Department of Integrative Biology, University of California, Berkeley. Available at http://ib.berkeley.edu/labs/slatkin/monty/Ewens_exact.program
- Slatkin, M., and N. H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.

- Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* 76:377 - 383.
- Slatkin, M., and W. P. Maddison. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* 126:249-260.
- Stillman, J. H., and C. A. Reeb. 2001. Molecular phylogenetics of Eastern Pacific Porcelain crabs, Genera *Petrolisthes* and *Pachycheles*, based on the mtDNA 16S rDNA sequence: Phylogeographic and Systematic Implications. *Molecular Phylogenetics and Evolution* 19:236-245.
- Sunnucks, P. 2000a. Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15:199 - 203.
- Sunnucks, P. 2000b. Reply from P. Sunnucks. *Trends in Ecology and Evolution* 15:377.
- Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389-399.
- Thorson, G. 1964. Light as an ecological factor in the dispersal and settlement of larvae of marine bottom invertebrates. *Ophelia* 1:167-208.
- Todd, C. D., W. J. Lambert, and J. P. Thorpe. 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktrophic and pelagic lecithotropic larval stages: Are pelagic larvae "for" dispersal? *Journal of Experimental Marine Biology and Ecology* 228:1-28.
- Toonen, R. J. 1997. Microsatellites for ecologists: Non-radioactive isolation and amplification protocols for microsatellite markers. Available at <http://biogeek.ucdavis.edu/Msats/index.html> (Appendix 1).

- Toonen, R. J., and S. Hughes. In press. Increased throughput for fragment analysis on an ABI 377 automated sequencer using a membrane comb and STRand software. *Biotechniques* (Appendix 2).
- Wares, J. P., S. D. Gaines, and C. W. Cunningham. 2001. A comparative study of asymmetric migration events across a marine biogeography boundary. *Evolution* 55:295-306.
- Watterson, G. 1986. The homozygosity test after a change in population size. *Genetics* 112:899 - 907.
- Watts, R. J., M. S. Johnson, and R. Black. 1990. Effects of recruitment on genetic patchiness in the urchin *Echinometra mathaei* in Western Australia. *Marine Biology* 105:145-151.
- Weir, B. S. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. 445 pp. Sinauer Associates, Inc., Sunderland, MA.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{ST} \approx 1/(4Nm + 1)$. *Heredity* 82:117-125.
- Wing, S. R., L. W. Botsford, J. L. Largier, and L. E. Morgan. 1995a. Spatial structure of relaxation events and crab settlement in the northern California upwelling system. *Marine Ecology Progress Series* 128:199-211.

- Wing, S. R., J. L. Largier, L. W. Botsford, and J. F. Quinn. 1995b. Settlement and transport of benthic invertebrates in an intermittent upwelling region. *Limnology and Oceanography* 40:316-329.
- Wing, S. R., L. W. Botsford, S. V. Ralston, and J. L. Largier. 1998. Meroplanktonic distribution and circulation in a coastal retention zone of the northern California upwelling system. *Limnology and Oceanography* 43:1710-1721.
- Wright, S. 1978. *Evolution and the Genetics of Populations: Variability Within and Among Natural Populations*. University of Chicago Press, Chicago, IL.

Chapter 3.

Patterns of spatial and temporal genetic heterogeneity among adults, subadults, and recruits of the intertidal crab *Petrolisthes cinctipes*.

Robert J. Toonen

**Center for Population Biology,
Section of Evolution and Ecology
University of California, Davis,
Davis, CA, USA. 95616.**

Abstract

Population genetic surveys frequently assume that patterns of genetic structure are stable over ecological time scales. This assumption makes possible inferences about the evolutionary importance of gene flow from populations sampled only a single time, but remains infrequently tested. Using two highly polymorphic microsatellite loci, I characterize the temporal and spatial patterns of genetic structure in the porcelain shore crab, *Petrolisthes cinctipes*. Groups of recruits, subadults and adults collected from northern California sites in each of 1997, 1998, and 1999 showed no significant departures from H-W expectations, although the overall pooled data suggests a deficit of heterozygotes. Estimates of both θ and ρ are highly consistent across loci and years from these samples. Estimates of θ are range from 0.07 to 0.08, whereas estimates of ρ are more variable from year-to-year, and range from 0.03 to 0.09 across the three sampling years. In contrast, estimates of both θ and ρ are significantly lower (0.02 and 0.01, respectively) when estimated from the overall data set from pooling the three years of samples. Comparing the genetic structure among years and sites, partitioned by size class, reveals that the among year component is greatest for recruits ($\theta = 0.12$), but erodes with increasing age classes (subadults = 0.09, adults = 0.08). Conversely, the among sites component of the analysis is least among the recruits ($\theta = 0.01$), but increases with the age class of the samples (subadults = 0.03, adults = 0.05). In an hierarchical analysis that includes all samples partitioned by year, site, and size class, estimates of population subdivision (θ) among sites is not significantly different from

zero. In contrast, subdivision among both the sampling year and size classes were significantly greater than zero. An AMOVA reveals that across the three years sampled, the among sites component is the only non-significant effect, although there is evidence that sites differ significantly within each sampling year. These data suggest that despite the remarkable similarity of estimates of both θ and ρ from the three years sampled here and the geographical survey presented in Chapter 2, the inferences drawn from single-year sampling are misleading relative to the complete analysis of the multi-year, age-class structured sampling analysis. Clustering analyses suggest recruits are most closely related to southern populations in 1997 and 1998, and to northern populations in 1999. Recruits show the greatest year-to-year variation in allelic frequencies, and heterozygosities of recruit samples are significantly lower than that of adult populations with which they cluster most closely. Across all samples, both allelic frequencies and the mean number of alleles per locus change significantly across sampling years; year-to-year differences at a site generally exceed those across all sites sampled within each year. These data are consistent with there being 1) variation in the source of larval recruits across years, 2) variation among individuals in their contribution to the larval pool, and 3) temporally and spatially variable natural selection that results in local adaptation among sites, overlaid by strong annual variation among samples. Taken together, these data suggest that evolutionary inferences drawn from geographic samples of populations of *Petrolisthes cinctipes* sampled a single time are potentially misleading relative to inferences drawn from the detailed study of size-structured samples analyzed through time.

Introduction

Over microevolutionary time, rates and patterns of gene flow determine the potential for, and scale at which, local adaptation can evolve; over macroevolutionary time, rates and patterns of gene flow should be related to evolutionary persistence and rates of speciation (Hartl and Clark 1997; Futuyma 1998; Lynch and Walsh 1998).

Direct measures of gene flow are very difficult or impossible in many natural populations (reviewed by Crochet 1996; Neigel 1997; Ouborg et al. 1999; Cain et al. 2000; Grosberg and Cunningham 2001). Indirect measures of gene flow inferred from the spatial distribution of alleles present an attractive alternative to costly, difficult and time-consuming direct measures of gene flow (e.g., Hellberg 1996; Lewis et al. 1997; Stanton et al. 1997; Petit and Mayer 1999; Ayre and Hughes 2000; Vos et al. 2001, reviewed by Palumbi 1994; Crochet 1996; Palumbi 1996; Neigel 1997; Bohonak 1999; Ellstrand et al. 1999; Ouborg et al. 1999; Grosberg and Cunningham 2001). Even in cases where direct estimates of dispersal are possible (e.g., Carlon and Olson 1993; Lewis et al. 1997; Petit and Mayer 1999; Michels et al. 2001), direct estimates of gene flow usually underestimate the effects of long-distance dispersal of individuals relative to the results obtained from indirect methods (Lewis et al. 1997; Petit and Mayer 1999). Although rare dispersal events are unlikely to be detected by direct methods, they may prove extremely important determinants of the biology and evolutionary history of natural populations (reviewed by Crochet 1996; Neigel 1997; Ouborg et al. 1999; Cain et al. 2000; Grosberg and Cunningham 2001). Thus, indirect methods of estimating

gene flow have the distinct advantage of providing estimates of the ultimate consequences of dispersal rather than the distribution or frequency of dispersal events themselves (Ouborg et al. 1999).

Such indirect approaches generally assume that patterns of genetic structure among geographic populations are stable over ecological time scales (reviewed by Lessios et al. 1994; Crochet 1996; Neigel 1997; Ouborg et al. 1999; Grosberg and Cunningham 2001). In the absence of natural selection, mutation, and migration, allelic frequencies in an infinite and randomly mating population remain constant from generation to generation (Hartl and Clark 1997; Futuyma 1998; Lynch and Walsh 1998). In nature, however, spatial patterns of genetic structure must change through time as species respond to a suite of stochastic (e.g., sampling error and genetic drift) and deterministic (e.g., selection and gene flow) forces (reviewed by Burton and Feldman 1982; Hedgecock 1994; Lessios et al. 1994; Palumbi 1994; Koehn and Hilbish 1995; Crochet 1996; Palumbi 1996; Neigel 1997; Bohonak 1999; Ellstrand et al. 1999; Ouborg et al. 1999; Grosberg and Cunningham 2001). Despite the operation of stochastic and deterministic evolutionary forces, does genetic structure remain sufficiently constant that one can reliably infer patterns and levels of gene flow from a single time frame?

In most cases, we simply don't know the answer, because most studies of spatial patterns of genetic structure are based on populations that are sampled only once. There are relatively few studies that examine whether the spatial distribution of alleles or allelic frequencies vary significantly through time (e.g., Gold et al. 1993; Kordos and

Burton 1993; Piertney and Carvalho 1995; Viard et al. 1997; Begg et al. 1998; Jollivet et al. 1998; Tarr et al. 1998; Tremblay et al. 1998; White et al. 1998; Banks et al. 2000; Carlsson and Nilsson 2000). For example, a preliminary search of the BIOSIS database revealed 5126 studies, published between 1985 and 2000, categorized using the keywords “population genetic structure.” In contrast, only 360 of those studies included temporal as an additional keyword. Most studies that assess genetic stability among populations through time have been used allozymes, and are more-or-less evenly divided between those documenting that the genetic constitution of populations tends to remain relatively stable (e.g., Berger 1973; Burton and Feldman 1981; Lessios 1985; McClenaghan et al. 1985; Waples 1990b; White and Svendsen 1990), and those that find significant changes in allele frequencies among samples drawn from the same population over relatively short periods of time (e.g., Gaines and Krebs 1971; Powers and Place 1978; Johnson and Black 1984a; Piertney and Carvalho 1995; de Sousa et al. 1997; Brykov et al. 1999; Kusumo and Druehl 2000; Gustine and Sanderson 2001).

Allele frequencies are expected to show the greatest rate of change in species with a short generation time, small and fluctuating population sizes, and with low levels of gene flow among populations (Lessios et al. 1994; Raijmann and Menken 2000). If any species should show temporal stability of allelic frequencies, the most likely candidates would be marine invertebrate taxa with high fecundity, long-lived planktonic larvae, few obvious barriers to dispersal, broad species ranges, and very large and relatively stable population sizes. However, the few studies of marine taxa that demonstrate temporally stable genetic structure come from species such as the tidepool

copepod *Tigriopus* (e.g., Burton and Feldman 1981; Burton and Lee 1994; Ganz and Burton 1995; Burton 1997, 1998; Burton et al. 1999), or the intertidal isopod *Excirolana* (Lessios et al. 1994), that do not possess these characteristics likely to ensure temporal stability of gene frequencies. Other studies (e.g., Johnson and Black 1984a; de Sousa et al. 1997; Li and Hedgecock 1998; Johnson and Wernham 1999) suggest that gene frequencies can change significantly through time, even in species that possess planktonic dispersal, broad species ranges and large population sizes.

If allele frequencies of marine populations are subject to substantial recurrent or stochastic change, then conclusions regarding local adaptation, gene flow, the degree of genetic structuring or any other reconstruction of the genetic history or evolutionary trajectory of a species drawn from studies in which populations are sampled only once may be unreliable (reviewed by Lessios et al. 1994; Raijmann and Menken 2000). Despite growing evidence that temporal genetic analysis can contribute substantially to our understanding of recruitment processes (e.g., Koehn et al. 1976; Todd et al. 1988; Lacson and Morizot 1991; Gold et al. 1993; Hedgecock 1994; Lewis and Thorpe 1994; Piertney and Carvalho 1995; Johnson and Wernham 1999), studies that compare the genetic composition of populations through time, or of different age classes within a population remain the exception rather than the rule. Some such studies examine patterns of genetic structure among cohorts, some examine only the temporal pattern of change among adults, but few do both (e.g., Kordos and Burton 1993; Edmands et al. 1996; Li and Hedgecock 1998; Johnson and Wernham 1999; Moberg and Burton 2000). This distinction is important because both types of data are needed to distinguish

between age-class differences that are consistent and do not change through time as a result of age-specific processes and those that differ through space or time as a result of population or year-specific processes (e.g., Burton 1994; Burton and Lee 1994; Ganz and Burton 1995; Burton 1997; Johnson and Wernham 1999, reviewed in Burton 1998).

This study presents a size-structured analysis of temporal genetic variation that builds on the initial examination of geographic genetic structure among populations of the porcellanid crab *Petrolisthes cinctipes* presented in Chapter 2. Here I use two highly polymorphic microsatellite loci to characterize the year-to-year variation in genetic structure among the recruits, subadults and adults of *P. cinctipes*. Given that *P. cinctipes* is a sedentary marine invertebrate that produces long-lived planktonic larvae, has an enormous census population size (on the order of 10^7 or more), a relatively high fecundity, an essentially continuous geographic range from the Dixon Entrance (British Columbia, Canada, $54^{\circ} 71' \text{ N}$) to just north of Point Conception, CA ($34^{\circ} 26' \text{ N}$), and a maximum lifespan on the order of eight to ten years (Jensen 1990), this species seems an obvious choice to expect temporal stability of gene frequencies over ecological time scales. This study highlights the importance of examining patterns of geographic genetic structure among populations through time and among age classes in order to evaluate the relative contribution of stochastic (e.g., individual reproductive success) and deterministic (e.g., selection and gene flow) forces in shaping the evolutionary trajectory of species.

Materials and Methods

Study organism and sample collection

Petrolisthes cinctipes ranges from roughly the Dixon Entrance (British Columbia, Canada, 54:71N) to Point Conception, CA (34:26N). Although restricted to a relatively narrow vertical range of the rocky intertidal (from roughly 0 to 1.8 m above MLLW), *P. cinctipes* is one of the most abundant decapod crustaceans in the Eastern North Pacific, occurring in densities as high as 3933 individuals / m² (Jensen 1990). Males and females of all size classes, from newly settled postlarvae (~1.5 mm) to large adults (up to 20 mm carapace width), live together in large aggregations among rocky cobble. Dense aggregations result largely from individual substrate preference (lower bound) and physiological tolerance (upper bound) of the adults, reinforced by gregarious settlement of the larvae (Jensen 1989, 1991a, b; Jensen and Armstrong 1991). Females typically extrude externally fertilized broods from February through mid-April. Newly extruded broods may contain from 10 to 1580 embryos, each roughly 800 - 840 µm in diameter (Knudsen 1964). Some females produce a second brood between May and September. Feeding larvae spend several months in the plankton before becoming competent to settle. During that time they have the potential to drift passively > 100 km, and potentially > 1000 km, depending on current patterns and velocity. After larvae molt into megalopae, they settle quickly in response to a waterborne cue associated with live conspecific adults (Jensen 1989, 1991a, b).

Megalopae of *P. cinctipes* do not molt at the time of settlement but rather lose the ability to swim due to degeneration of their swimming appendages (pleopods); metamorphosis to the pigmented first instar juvenile occurs a week or more later (Jensen 1991a). This delay of metamorphosis after settlement makes identification of newly settled postlarvae straightforward. New recruits take advantage of the defended space beneath adults for about a year after settlement, until reaching roughly 5 mm carapace width (CW), when adults actively begin to expel them (Jensen 1991a). Colonization of new habitats occurs most frequently by crabs of ~ 5mm CW, presumably because they are forced to seek new shelter when expelled by adults (Jensen 1991a). Despite the potential motility of porcellanid crabs, adults are quite sedentary in the field. Mark-release-recapture data show that >50% of crabs were recovered within a few meters of the point of release after two months (Jensen and Armstrong 1991).

Petrolisthes cinctipes were sampled from rocky cobble around the base of large boulders at nine sites distributed across three embayments in northern California (Fig. 1). Crabs were collected and chelipeds (claw-bearing forearms) were preserved, measured and stored as described in Chapter 2. Each site from Crescent City, CA to Dillon Beach, CA was visited on a single tidal series in each sampling year. Samples were collected in August of 1997, September of 1998 and July of 1999.

At each site, I tried to collect between 50 and 100 recruits, subadults and adults, classified according to their CW. Carapace width was estimated as the average of a regression using each claw width and claw length, as described in Chapter 2. If the CW

estimated from the claw length regression model differed from that estimated from claw width by more than 1.0 mm, the sample was excluded from further analysis. Recruits were defined as non-pigmented postlarvae that must have settled within the two weeks prior to collection. Subadults were defined as individuals between 5 and 7 mm carapace width, likely corresponding to individuals at least one year of age. Adults were those individuals > 9 mm CW, corresponding to individuals at least 3 years old. The relationship between carapace width (size class) and the age is, at best, an approximation, because there is considerable variation in conditions among sites and individual growth rates within a site.

I collected all crabs encountered at a site, regardless of size. I stored the live crabs in a mesh container filled with fresh kelp (to minimize stress), until at least 50 individuals of each size class were found or the tide forced me to stop collecting. Subadults and adults were individually harassed to stimulate self-amputation of a cheliped (claw-bearing forearm), which was placed immediately into 95% ethanol. Live crabs were then released at the site of collection. New recruits were preserved whole in 95% ethanol. Samples were kept on ice in the field for roughly 4-5 days, after which the 95% ethanol was decanted and replaced. Separate vials containing preserved chelipeds from subadult and adult crabs, as well as whole recruits from each site were subsequently stored at -40°C.

DNA extraction

Prior to DNA extraction, each sample was removed from the ethanol and the length and width of the chela (claw) was measured to the nearest 0.01 mm with digital calipers. After dissection of 3-5 mg of muscle tissue from the claw, each sample was placed into vial filled with 95% ethanol, and stored at -40°C as a backup. In the laboratory, each cheliped was assigned a unique identification number; up to 50 individuals from each site were selected randomly from the numbered cheliped samples for subsequent DNA extraction and genetic analysis. If fewer than 50 individuals of a given size class were collected at a site, all samples were included in the study. DNA extraction, PCR amplification and electrophoresis were all done blindly using these unique crab ID numbers. Carapace width and site of collection for individual crabs were reassigned to numbered samples only after all samples had been scored.

DNA was extracted from the dissected muscle tissue using a modification of Gentra System's PUREGENE® D-T200 protocol. I discovered that minimizing the amount of carapace included in the extractions led to consistently higher quality and quantity of DNA extracted. I modified the mouse tail DNA isolation protocol from Gentra by doubling all reagent volumes for cell lysis, RNase treatment, protein precipitation, and DNA precipitation steps. Tissue was digested overnight at 55°C on a low speed vortex with 0.2 mg / ml of proteinase-K. After adding 200 µl of protein precipitation solution, samples were incubated at -20°C for 5-10 minutes to maximize precipitation efficiency. I precipitated the DNA with 600 µl of ice-cold 100%

isopropanol, and incubated samples at -20°C for 5-10 minutes prior to centrifugation. Samples were centrifuged at 13,000 RCF for 10 min at 4°C, and the pellet was washed twice with 600 µl of ice-cold 70% ethanol. Precipitate was then dried under vacuum. Desiccated DNA pellets were rehydrated in 50 µl of low-TE (10mM Tris, 0.1mM EDTA) prior to being stored at -40°C for subsequent amplification.

PCR amplification and electrophoresis

The procedure for isolating SSR loci and developing primers for amplifying these loci in *Petrolisthes cinctipes* is described in detail in Chapter 1. Primer sequences for amplifying the two highly polymorphic microsatellite loci used in this study, Pc156s and Pc170s, are presented in Table 1. In a final volume of 10 µl, PCR reaction mixes contained Perkin Elmer 10X Buffer II at 1X concentration, 2.5 mM MgCl₂, 0.1 mM dNTPs, 1X BSA, 0.5 mM of each the forward and reverse primer, 1 unit of Taq polymerase, and 1-50 ng of template DNA (Toonen 1997). PCR amplifications were performed using the touchdown protocol detailed in Chapter 1. Amplified PCR products were electrophoresed on an Applied Biosystems ABI 377 XL automated sequencer and scored with STRand (Hughes 1998), as outlined in Toonen and Hughes (in press, Appendix II).

Data analysis

I performed global tests for deviation from Hardy-Weinberg expectations using the Arlequin Population Genetics Data Analysis Program (Schneider et al. 2000). Tests for H-W equilibrium employ a Markov-chain method (Guo and Thompson 1992). Markov chain lengths for these tests were 500,000 steps with a 10,000 step dememorization as explained in Chapter 1.

I calculated Weir and Cockerham's (1984) unbiased estimator of Wright's F_{ST} (θ) on a locus-by-locus basis using the Tools for Population Genetic Analyses (TFPGA) software (Miller 1997). TFPGA uses the method and terminology outlined in Weir and Cockerham (1984; Weir 1996). Because only two loci are included in this data set, however, bootstrapping and jackknifing are not meaningful, and therefore no mean, standard error or confidence intervals for θ are presented. Goodman's ρ (Goodman 1997), and $(\delta\mu)^2$ (Goldstein et al. 1995a, b), were both calculated using R_{ST} Calc 2.2 (Goodman 1997).

I constructed UPGMA (Unweighted Pair Group Method with Arithmetic Mean) dendrograms (Michener and Sokal 1957) using the genetic distance D_A (Nei et al. 1983), based on its superior performance in simulations (Takezaki and Nei 1996). D_A (Nei et al. 1983) was calculated in Arlequin, and UPGMA dendrograms were drawn using PHYLIP (Felsenstein 1993) or TFPGA (Miller 1997). UPGMA dendrograms give an indication of the degree of genetic dissimilarity among populations; the longer the branches separating a pair of populations, the more genetically dissimilar are

individuals sampled from those geographic locations. If gene flow depends primarily on the number of dispersal steps between populations, then sites that are geographically proximate should cluster more closely than with distant sites.

Finally, I used Arlequin to perform an analysis of molecular variance (AMOVA), which partitions observed genetic variance into components analogous to traditional analysis of variance analyses (Excoffier et al. 1992). Although AMOVA is based on an analysis of variance of gene frequencies, this framework partitions the total variance into covariance components due to inter-individual differences, intra-population differences and inter-population differences (Excoffier 2000; Rousset 2000). There are six different types of hierarchical AMOVA calculated by Arlequin (Schneider et al. 2000), but it is not straightforward to adapt the AMOVA framework to incorporate a nested design analysis of variance, or to test interaction terms (Excoffier, pers. comm.). Therefore, I was constrained to test the relative contributions of year, site and size class in a pairwise fashion, examining the relative proportion of variation among groups within populations. Using the AMOVA framework within Arlequin, I partitioned observed genetic variation components and calculated hierarchical F -statistic analogs (Φ -statistics). The significance of these values was tested by 100,000 matrix permutations (Excoffier et al. 1992).

Standard statistical analyses (ANOVA, post-hoc comparisons among means, and linear regression) were all performed using JMP 4.0.2 (SAS Institute 2000).

Results

The analysis of size-structured genetic samples has the potential to significantly enhance our understanding of recruitment processes even in species that lack strong population differentiation as adults (Johnson and Wernham 1999; Moberg and Burton 2000). Determining how genetic variation is partitioned among age classes, and how patterns of geographic structure change among them and through time within them can provide crucial insights into processes such as larval-dispersal-mediated gene flow, and local adaptation. Thus, in the sections that follow, the data are analyzed by site, year and size class to compare the relative contributions of each factor to the overall pattern of geographic population structure. In order to isolate the effect of a given factor in contributing to the overall pattern of geographic population structure, analyses are also included that pool the data across site, year or size class. Results from hierarchical analyses that include each year, site and size class can then be compared to the pooled results. Cases where any data are pooled are indicated in the subheadings for each section of the results that follow.

Overall patterns of microsatellite variation and conformity to Hardy-Weinberg expectations – all samples pooled

In order to test for global deviations from Hardy-Weinberg equilibrium, all samples were pooled and the complete data set was tested. Across all crabs genotyped

in this study, both loci only marginally conform to Hardy-Weinberg (H-W) expectations (Table 1).

The number of individuals genotyped varied by size class and site. Sample sizes for residents and recruits are presented by locus in figures 2 and 3. Overall there were 978 crabs genotyped from the 8 sites sampled in 1997, 1372 crabs genotyped from the 8 sites sampled in 1998, and 1252 crabs genotyped from the 6 sites sampled in 1999. In most cases, the number of individuals of each size class scored at each site was sufficient to minimize bias in estimates of genetic structure and distance (Ruzzante 1998). As with the survey of geographic genetic structure presented in Chapter 2, the microsatellite loci used in this study are highly polymorphic, with 34 alleles for locus Pc156s, and 47 alleles at locus Pc170s. Expected heterozygosities ranged from 0.68 to 0.83 across all 3602 crabs genotyped in this study (Table 1).

The overall allelic frequency distributions for each locus were similar to those reported in Chapter 2 (Fig. 2), but include a greater size range (52 - 124 bp for Pc156s and 86 - 196 bp for Pc170s in the current study as compared to 74 - 124 bp and 88 - 170 bp, respectively, in the previous chapter). As with the survey presented in Chapter 2, the three most common alleles at each locus accounted for more than 80% of the allelic variation across all sites. Private alleles occurred at very low frequencies at each locus. Interestingly, all private alleles in this study were outside the size range of alleles reported for each locus in Chapter 2 (Fig. 2). Seven of the 10 private alleles detected in this study were found among recruits, but as with the geographic survey presented in

Chapter 2, all private alleles occurred in the central portion of the northern California range, between Shelter Cove (#3) and Salt Point (#7).

Geographic patterns of variation in the number of alleles per locus at each site were similar for both loci (ANOVA, $F = 1.12$, $df = 1$, $p > 0.05$). Averaged across loci, the mean number of alleles per locus varied from 3.0 (1998, residents at Albion) to 22.0 (1999, residents at Ft. Bragg), but differences among sites (ANOVA, $F = 1.86$, $df = 7$, $p > 0.05$) and size classes (ANOVA, $F = 0.08$, $df = 2$, $p > 0.05$) were not significant (Fig. 3). The mean number of alleles per locus (13.90 ± 1.06 SE in 1997, 9.17 ± 0.95 SE in 1998, and 13.59 ± 1.12 SE in 1999) did vary significantly across years (ANOVA, $F = 11.30$, $df = 2$, $p < 0.05$). The mean number of alleles per locus also varied greatly among residents and recruits at a site (Figs. 4 & 5). Parallel changes in the number of alleles per locus merit further attention, but there is no obvious geographic or temporal pattern to the distribution (Fig. 3).

Conformity to Hardy-Weinberg expectations according to size class, sites and years

Although the overall data suggest marginal conformity to H-W expectations, there were significant deviations from H-W expectations for some size classes at a number of sites sampled. Locus Pc156s showed five significant deviations from H-W expectations among the residents (adults + subadults) and recruits tested across all sites (Fig. 4). Four of the five significant deviations from H-W expectations occurred among the residents at each site, whereas only one occurred among the new recruits (1997

Albion). For locus Pc170s, on the other hand, there were 20 significant deviations from H-W expectations, even after sequential Bonferroni correction (Rice 1989). Deviations from H-W expectations at Pc170s were split nearly evenly between residents and recruits; 11 of the significant results occurred among the residents, whereas nine were among recruits (Fig. 5).

Conformity to Hardy-Weinberg expectations among size classes and years – sites pooled

When pooled across sites, heterozygosities are much closer to H-W expectations for each size class and year of sampling (Table 2). After applying sequential Bonferroni correction (Rice 1989), there were no significant departures from H-W expectations among the recruits, subadults and adults collected within any year of sampling (Table 2).

Allele frequencies among sites and years – size classes pooled

When size class samples from each site are pooled, frequencies of the three most common alleles at each locus show dramatic annual fluctuations (Fig. 6). As with the allelic frequency data presented in figures 3 and 4, however, there are no obvious patterns of variation either among sites or among years (Fig. 6). Changes in allelic frequencies are greatest among recruits, and least among the adults, although all size

classes exhibit significant annual variation in allelic frequencies across sites (analyses not shown).

Genetic structure among sites and years – adult crabs only

In order to compare results with the analyses presented in Chapter 2, I used only the adult crabs to examine patterns of geographic population structure in each of the three years of study. Weir and Cockerham's (1984) estimator of population subdivision, θ ($= F_{ST}$), averaged 0.07 in 1997, 0.08 in 1998, and 0.08 in 1999 (Table 3). Each of these estimates is approximately equal to that revealed in the geographic survey of the same area presented in Chapter 2. Because there are only two loci included in these analyses, jackknifing and bootstrapping across loci to calculate confidence intervals and standard errors for these estimates is meaningless. Therefore, only the locus-by-locus results and means are presented for all analyses of population subdivision. However, in most cases the general trend revealed by both loci was concordant, and differences in estimates of population subdivision and genetic distance generated by each locus were generally slight.

In each year of sampling, there is a significant amount of population differentiation among sites, but the overall estimate of θ when the samples from all three sampling years are pooled is much lower than any of the single-year samples analyzed individually. Goodman's (1997) estimator of R_{ST} , ρ , shows a pattern similar to that of θ , although the variation from year-to-year is greater (Table 3). Despite

general agreement of both IAM- (θ) and SMM-based (ρ) estimators that there is significant population subdivision among sites in each year sampled, when the three years of samples are pooled, both θ ($F = 14.24$, $df = 3$, $p = 0.01$) and ρ ($F = 78.0$, $df = 3$, $p < 0.01$) differ significantly from estimates based on any single year of data (Table 3). Estimates of population subdivision among sites are remarkably consistent across each sampling year, but inferences regarding population structure differ when the three years of samples are pooled. Annual estimates of F and f are less consistent, and inferences about inbreeding based on these statistics vary greatly depending on the year in question. For example, in 1997 there appears to be a moderate amount of inbreeding within sites, whereas in 1998, the data suggest a moderate degree of outcrossing (Table 3).

Estimates of gene flow among sites also vary significantly between sampling years. Based on the relationship $N_m = (1 - F_{ST}) / 4 F_{ST}$ (Wright 1978), the effective number of migrants per generation (using θ) was estimated to be 2.49 ± 0.02 SE in 1997, 7.06 ± 0.08 SE in 1998, and 3.83 ± 0.03 SE in 1999. The 95% confidence intervals for 1997 (1.75 - 3.43) and 1998 (4.39 - 11.34) do not overlap, but the 1999 interval (2.78 - 5.37) overlaps both. When the overall data set is analyzed with samples from all three years pooled, the estimate of N_m is 18.79 ± 0.10 SE, and the 95% confidence interval (11.80 - 24.93) does not overlap with any of the single-year estimates derived from samples collected in 1997, 1998 or 1999.

Genetic structure among size classes and years – sites pooled

Estimates of subdivision among recruits, subadults, and adults are all of roughly the same magnitude in each sampling year (Table 4). These estimates of θ and ρ for each size class show relatively similar patterns in each year, and are also consistent with the findings from the previous chapter. By either measure, subdivision among sites averages approximately 0.08 within each size class in each year of sampling. Interestingly, estimates of population subdivision are highly consistent among all size classes in 1997, but not 1998 or 1999 (Table 4).

Genetic structure among size classes – years and sites pooled

Estimates of θ for each size class when samples from all three years are pooled (Table 5) are again significantly lower ($F = 4.0$, $df = 3$, $p = 0.05$) than estimates based on single-year samples (Table 4). Using Weir and Cockerham's θ in the hierarchical nesting scheme of Weir (Weir 1996), subdivision among subpopulations within populations (sites) and among populations (years) can be estimated simultaneously. The hierarchical analysis suggests that for each size class, there is far greater similarity among geographic sites than among years (Table 5). Estimates of θ_s (among sites) range from 0.0005 to 0.06 ($\bar{x} = 0.03 \pm 0.002$ SE), whereas estimates of θ_y (among years) range from 0.07 to 0.12 ($\bar{x} = 0.10 \pm 0.002$ SE). Subdivision among sites is lowest for recruits ($\theta_s = 0.01$), but increases among subadults ($\theta_s = 0.03$) and adults (θ_s

= 0.05). Estimates of θ among years, on the other hand, are highest for recruits (θ_r = 0.12), but decrease among subadults and adults (Table 5).

Genetic structure among sites, size classes, and years – all samples

In each year of sampling, there is more genetic structure among size classes than among sites (Table 6). Comparison of the relative effects of year, site and size class in a single hierarchical analysis of all data collected in this study indicates that the observed population subdivision results from variation among size classes and years rather than among sites (Table 6).

Genetic distance and clustering analyses – among sites and size classes by year

UPGMA dendrograms based on genetic distance, D_A (Nei et al. 1983), reveal little consistency among sampling years in the overall patterns of genetic similarity within and among sites (Figs. 7-9). For example, residents (adults and subadults) at Salt Point show little genetic differentiation ($D_A < 0.10$) from residents at most other northern California sites in 1997 and 1999 (Figs. 7 & 9). In 1998, however, residents at Salt Point are separated from residents at all other sites sampled by the greatest genetic distance ($D_A > 0.15$) measured in that year (Fig. 8). Crabs collected from Albion are even more variable in their relation to other northern California sites: in 1998, residents at Albion cluster in the middle of the sampling sites, whereas in 1999 these same

residents show the greatest genetic distance to other sampling sites detected in this study (Figs. 8 & 9). Neither the magnitude of differentiation among samples (ranging from $D_A \approx 0.17$ in 1998 to > 0.40 in 1999), nor the pattern of clustering of sites or size classes are consistent from year-to-year.

The trend for recruit samples to cluster most closely with residents of other sites shows an interesting pattern across sampling years. I classified recruits into one of four categories: 1) those that cluster with residents of the same site from which they were collected (self-seeding), 2) those that cluster with residents of a site to the south of the site from which they were collected (south), 3) those that cluster with residents of a site to the north of the site from which they were collected (north), or 4) those recruit samples that did not cluster closely with residents from any site (unknown). Using these categories, six of the eight samples of recruits from 1997 were either self-seeding (3) or derived from the south (3), and none clustered with residents from a site to the north (Fig. 7). In 1998, three samples of recruits were again derived from the south, but only one was self-seeding, and again none clustered with residents from a site to the north (Fig. 8). In contrast, although there were only five samples of recruits available in 1999, three of them derived from residents of sites to the north, and one was self-seeding (Fig. 9). 1999 was the only year of this study in which recruits showed a general tendency to cluster with residents of sites located north of the site from which the recruits were collected.

Genetic distance and clustering analyses among sites, size classes, and years

An overall UPGMA clustering analysis, based on the complete data set for all the sites, size classes and sampling years in this study does not reveal any obvious patterns with respect to sites or size classes (Fig. 10). In general, the clustering pattern of sites is similar to that of the single year UPGMA dendrograms (Figs. 7 - 9). However, there are some differences in the pattern of clustering that result from the inclusion of some samples from other years. For example, Shelter Cove and Salt Point form the same cluster in the 1999 and overall UPGMA dendrograms, but the Ft. Bragg samples cluster more closely with 1997 and 1998 samples, altering their position in the overall clustering diagram (Fig. 10), relative to that of the 1999 UPGMA dendrogram (Fig. 9).

Regardless of the differences from the single-year dendrograms, the most striking pattern to emerge from the overall UPGMA clustering analysis is that all samples collected in each sampling year tend to be more similar to one another, regardless of site or size class, than samples collected in other sampling years (Fig. 10). Given that recruits show the greatest differentiation among sampling years (Table 5), it is possible that this variation is sufficient to swamp the weaker signal among residents. This is not the case, however, because dendrograms based upon only residents show a similar trend for samples to cluster by year (analysis not shown). This pattern of samples tending to cluster together by year rather than by geography is consistent

whether the UPGMA is constructed with 1) all size classes, 2) residents only, 3) all alleles, or 4) only the three most common alleles at each locus.

Genetic distance and clustering analyses among sites and years – size classes pooled

The pattern of clustering among samples collected in each year is even more striking if the size classes collected at each site are pooled. The resulting UPGMA dendrogram produces three relatively well-supported clusters that contain all the geographic sites sampled within each year of the study (Fig. 11). Although the clustering of sites within years varied for each measure of genetic distance, samples clustered by year, regardless of whether the UPGMA dendrogram was constructed using Cavalli-Sforza and Edwards chord distance (Edwards and Cavalli-Sforza 1965; Cavalli-Sforza and Edwards 1967), Nei's genetic distance (Nei 1972, 1978), Rogers' genetic distance (Rogers 1972) or Wright's (Wright 1978) modification of it, coancestry (Reynolds et al. 1983), or $(\delta\mu)^2$ (Goldstein et al. 1995a).

I also randomly subsampled 90 crabs (all three size classes, without replacement) from each site ten times to compare the UPGMA dendrogram produced by the mixed-size sample to that of the pooled data. Although bootstrap support is considerably lower than that for Figure 11, each of the ten dendrograms similarly clustered by year of sampling rather than by geography (analyses not shown).

Analysis of Molecular Variance (AMOVA) among size classes, sites, and years

The analysis of molecular variance (AMOVA) shows that little of the observed genetic variation can be accounted for by differences among geographic sites.

Regardless of the grouping chosen, variation among sites explains less than 1% of the overall genetic variation (Table 7). Differences among size classes within sites exceed differences among geographic sites, although the proportion of variation explained by both is only about 1% (Table 7).

Sampling year, on the other hand, explains approximately 5% of the observed genetic variation, whether sites or size classes are analyzed by year (Table 7). Size classes within years appear to explain the greatest proportion of the genetic variation in these data, with roughly 5% of the observed genetic variation accounted for by differences among years and among size classes within years. Although size classes within years explains the greatest proportion of the observed genetic variation in these data, more than 90% of the genetic variation remains unexplained in each case (Table 7).

Discussion

Interpreting patterns of genetic structure in natural populations is problematic because a particular pattern may be generated by both historical and contemporary processes, acting singly or in combination, at a variety of different spatial and temporal

scales (reviewed by Grosberg and Cunningham 2001). Furthermore, most methods for detecting patterns of genetic structure and inferring gene flow assume that any genetic signal from historical events has been erased by the combined effects of migration and drift (reviewed by Luikart and England 1999; Grosberg and Cunningham 2001).

However, few studies of population structure among marine organisms test this assumption explicitly (e.g., Lessios et al. 1994; Hellberg 1995). Many authors have argued that few marine species satisfy fully the expectations of equilibrium between genetic drift and migration across the entire species range (reviewed by Palumbi 1994; Hilbish 1996; Benzie 1999; Grosberg and Cunningham 2001).

In sand-free areas of the rocky intertidal, densities of *Petrolisthes cinctipes* can reach > 3500 individuals per square meter, making it the most common decapod crustacean throughout its range from the Dixon Entrance (BC, Canada) to Pt. Conception (CA) (Jensen 1990). Coupled with a relatively high fecundity, a generation time on the order of 3-8 years and long-lived pelagic larvae, this seems a species in which temporal stability of allelic frequencies is highly likely. Despite the striking similarity of F_{ST} and R_{ST} estimates from the initial geographic survey (Chapter 2) and each of the three annual samples presented in this study, the overall data set pooled across years provides little support for a temporally persistent pattern of geographic structure among populations for the species. The overall hierarchical analysis of all samples collected in this study (Table 6) suggests that the most important source of variation among samples of *P. cinctipes* are year ($\theta_r = 0.05$) and size class ($\theta_s = 0.08$) effects, rather than geographic sampling site effects ($\theta_g \approx 0$).

Insofar as the two loci used in this study provide an accurate representation of the genetic history of *Petrolisthes cinctipes*, these data suggest that inferences regarding genetic structure of this species drawn from a single time sample of populations are unreliable. Inferences drawn from each of four independent surveys in which populations are sampled only a single time are highly consistent, but also misleading relative to the inference of genetic structure drawn from the sampling of age-structured samples through time.

Lessios et al. (1994) suggest that allozyme stability through time of populations of the marine isopod *Excirolana* provides confidence in evolutionary conclusions reached by studies that have taken only a single genetic sample through time. The temporal variability of the microsatellite data presented herein do not support a similar conclusion for *Petrolisthes cinctipes*. If temporal genetic variation within populations of benthic invertebrates proves to be common, then our ability to draw evolutionary inferences based on populations that have been sampled only once is seriously compromised. However, until further data are available on the generality of this pattern, these data provide a cautionary note regarding the reliability of using single-time snapshot surveys to partition genetic structure in marine species, and emphasize the importance of sampling populations more than once before drawing evolutionary inferences from allelic frequency data.

These data beg the question of how such patterns of temporal and spatial variation arise among populations of *Petrolisthes cinctipes*. Although many recent reviews have found a significant relationship between larval dispersal potential and

degree of genetic subdivision (e.g., Burton 1983; Hedgecock 1986; Ayre 1990; Palumbi 1994; Bohonak 1999), each has also identified numerous cases in which dispersal potential only weakly predicted genetic structure (reviewed by Cunningham and Collins 1994, 1998; Benzie 1999; Grosberg and Cunningham 2001). Perhaps the most consistent result from surveys of genetic structure among marine invertebrates is that differentiation of adjacent sites can often equal or exceed differentiation between sites across the entire species range (e.g., Johnson and Black 1982; Johnson and Black 1984a, b; Watts et al. 1990; Kordos and Burton 1993; Edmands et al. 1996; Moberg and Burton 2000). These exceptions highlight that our understanding of processes generating genetic structure among marine populations, especially the role of contemporary dispersal, is far from complete. For example, genetic breaks revealed by surveys often do not correspond to any known barriers to dispersal (e.g., Reeb and Avise 1990; Karl and Avise 1992, reviewed by Avise 1994; Benzie 1999; Grosberg and Cunningham 2001). Conversely, genetic homogeneity occurs among populations for which there are no obvious contemporary pathways for dispersal (reviewed by Avise 1994; Cunningham and Collins 1994; Shulman and Bermingham 1995; Palumbi 1997; Cunningham and Collins 1998; Benzie 1999; Benzie 2000; Grosberg and Cunningham 2001). In general, the present-day genetic structure of most marine species seems to result from the combined effects of contemporary gene flow and historical factors, such as range expansions, that can be greatly complicated by a lack of equilibrium among populations throughout the species' ranges (e.g., Hellberg 1995; Marko 1998, reviewed by Cunningham and Collins 1994, 1998; Grosberg and Cunningham 2001).

Several hypotheses have been advanced to explain how such small-scale genetic structure can accumulate in the face of extensive pelagic larval dispersal. One possibility is that substantial microgeographic population structure exists. A Wahlund effect, the departure from Hardy-Weinberg equilibrium due to mixing of individuals from groups with different allelic frequencies, could account for heterozygote deficiencies on small scales in some species (e.g., Johnson and Black 1984b; Kijima et al. 1987; Waples 1990a). In the case of the limpet *Siphonaria*, there is little geographic variation in allozyme frequencies at the scale of the species range (Johnson and Black 1982; Johnson and Black 1984b). Despite this large-scale uniformity, however, there exists significant fine-scale genetic patchiness, which is repeated, rather than accumulated, on the larger scale (Johnson and Black 1984b). Neither geographic or temporal variation in allelic frequencies can explain the pattern. Instead, the fine-scale structure appears to result from binomial sampling variance among small local breeding groups that results in a Wahlund effect (Johnson and Black 1984a, b). This hypothesis predicts that there should be a significant deficiency of heterozygotes across broad scales that disappears as the scale of sampling is reduced to the point that inadvertent mixing of samples from smaller panmictic groups is eliminated.

A second possibility is that the source population from which planktonic larvae recruit varies through time. Differences in seasonal current patterns, spawning season or developmental times among geographic populations may result in temporal variation in the allelic frequencies of recruits arriving at a given site (e.g., Kordos and Burton 1993). Such genetic variation may accumulate within populations and lead to small-

scale genetic structure if inter-population gene flow is not sufficient to overcome population differentiation resulting from genetic drift and/or natural selection (e.g., Koehn and Williams 1978; Kordos and Burton 1993). The larval source hypothesis predicts that the genetic affinity of recruits to the geographic populations of spawners should vary with time and location, but further requires that adult populations exhibit sufficient differentiation to account for the variation detected among recruits.

In addition to variation among larval sources made possible by long-lived pelagic larvae, variation in individual reproductive success may be great due to the considerable fecundity and high early mortality of many marine species. In highly fecund species, a small minority of individuals may effectively contribute all recruits for the entire population in each generation by a sweepstakes-chance matching of reproductive activity with oceanographic conditions conducive to spawning, fertilization, larval survival and successful recruitment (Hedgecock 1994). The sweepstakes-reproductive success hypothesis can account for both the widespread, large discrepancies commonly found between effective and census population estimates and for local genetic differentiation despite high gene flow (Hedgecock 1994). This hypothesis predicts that allelic frequencies should exhibit significant temporal variation owing to genetic drift, and that genetic diversity should be significantly reduced among new recruits relative to the spawning population (e.g., Moberg and Burton 2000).

Finally, small-scale population structure may result from pre- or post-settlement natural selection among populations (e.g., Koehn et al. 1976; Powers and Place 1978; Johnson and Black 1984a; Hilbish and Koehn 1985; Gardner and Palmer 1998).

Despite geographic proximity of sites, abiotic (such as wave exposure, salinity, nutrient supply, etc.) and biotic differences (such as competitors and predators) have the potential to exert differential selective pressure over small spatial scales. Although it is common to assume selective neutrality of the genetic markers used in surveys of population genetic structure such as this, markers may be linked or directly subject to natural selection that can alter the allelic frequencies among samples obtained from a given site (e.g., Koehn et al. 1976; Powers and Place 1978; Hilbish and Koehn 1985; Karl and Avise 1992; McGoldrick et al. 2000), or even among size-classes within a site (e.g., Gardner and Palmer 1998). In general, researchers have inferred the action of natural selection when population surveys using putatively neutral genetic markers revealed clinal or habitat-specific patterns of allelic frequency change (e.g., Koehn et al. 1976; Powers and Place 1978; Hilbish and Koehn 1985; Schmidt and Rand 2001), or marked differences among loci or classes of genetic markers (e.g., Reeb and Avise 1990; Karl and Avise 1992; Schmidt and Rand 2001). The ecological genetics required to test this hypothesis are not trivial, and with more complex spatial patterns, isolating the specific causes and effects of natural selection can be even more difficult (e.g., Gardner and Palmer 1998).

The results presented in this study of *Petrolisthes cinctipes* are bolstered by their remarkable concordance with the findings of Moberg and Burton (2000). In an allozyme-based survey of recruits, subadults and adults of the red sea urchin *Strongylocentrotus franciscanus* across the same geographic region as this study of *P. cinctipes*, Moberg and Burton (2000) also found that levels of genetic differentiation

among adult populations were too low to account for the amount of temporal differentiation among recruits through changes in source population alone. Even more striking is the similarity of clustering patterns in the UPGMA dendrograms from 1997 (the only year in which samples for both studies overlap). Both studies include samples from similar, though not identical, locations; the Anchor Bay site sampled by Moberg and Burton is roughly 18 km from my site at Salt Point, and Caspar is about 12 km from my site at Albion. If I consider such proximate geographic sites as equivalent for the sake of comparison, there is only a single difference in the general pattern of clustering among residents and recruits between the two studies. In *S. franciscanus*, Bodega Bay recruits from 1997 are the most divergent sample and do not cluster with any resident population (Moberg and Burton 2000), whereas recruits of *P. cinctipes* to Bodega Bay in 1997 cluster with the residents sampled from Bodega. Nevertheless, the similarity in clustering in the UPGMA analyses between these two studies using different genetic markers on different species is notable. Overall, both studies suggest that the pattern of clustering among adults and recruits varies across sampling years, and sites that cluster together in one year may be among the most divergent the next.

This study cannot exclude a Wahlund effect, variation in the source population from which larvae recruit, sweepstakes-reproductive success, or natural selection as contributors to the pattern of genetic structure in *Petrolisthes cinctipes*. Instead, the data are consistent with each of the possible mechanisms acting to a greater-or-lesser degree to produce the overall pattern of genetic structure revealed by this survey. The first possibility, that there exists a Wahlund effect, is supported by both loci showing a

slight excess of homozygotes overall; however, no significant deficiencies of heterozygotes are detected if samples are analyzed by size class within each year. This argues that the homozygote excess is unlikely to be due to a null allele. Furthermore there were no individuals that failed to amplify a scorable product from the 3602 crabs genotyped for locus Pc156s. Locus Pc170s, on the other hand, failed to amplify any product in 14 of the 3602 crabs (~ 0.4%), and a rare null allele is certainly possible, but the effect of an allele at such low frequency cannot alone account for the excess of homozygotes at this locus. The consistency of this effect between the loci suggests that the causes of the heterozygote deficits are likely to be populational, rather than locus-specific. These data are consistent with age classes exhibiting significant differentiation in allele frequencies, and a Wahlund effect results from mixing size classes in a pooled analysis. However, genetic differentiation accumulates among sites and erodes among years with increasing age class. Thus, the variation among size classes cannot alone account for the pattern of genetic structure in *P. cinctipes*.

Similar deviations from H-W equilibrium are also expected if the spawning populations from which larvae recruit are genetically differentiated from recipient populations (Kordos and Burton 1993). However, geographic variation of allelic frequencies within a sampling year are inadequate to produce these deficits, and all samples conform to H-W expectations among size classes and years when geographic sites are pooled. This suggests that differences in the source of recruits through time alone cannot account for the degree of differentiation observed among populations of *P. cinctipes* sampled in this study. However, there is strong evidence from these data that

the source of recruits does vary through time. The UPGMA dendrograms from 1997 and 1998 do not cluster recruits with a single population of adults to the north of the site at which recruits were collected. In striking contrast, none of the recruit samples collected in 1999 cluster with adult populations located to the south, and three of the five cluster with populations to north.

Superficially, the change in source of recruits could be taken as support for a genetic consequence of the upwelling-relaxation flow mechanism proposed by Wing, Botsford and coworkers (Wing et al. 1995a; Wing et al. 1995b; Wing et al. 1998; Morgan et al. 2000). Characterized by weak upwelling and warmer sea water temperatures, 1997 and 1998 were El Niño years, whereas 1999 was a La Niña year, dominated by unusually low temperatures and strong upwelling. Thus, the change in recruit source corresponds to the frequency and intensity of upwelling during these years. However, the pattern does not strictly conform to predictions of the upwelling-relaxation hypothesis. In 1997, although recruits generally arrive from the south, they also appear to originate consistently from residents of a different embayment. In contrast, each case where recruits cluster significantly with a resident sample in 1998 includes residents from within the same embayment. Obviously the effect of retention zones, if they have any detectable genetic consequence, is highly variable by year. Nevertheless, Moberg and Burton (2000) emphasize that data such as these do not constitute a strong test of the upwelling-relaxation flow hypothesis, because the temporal and spatial scale of sampling for recruits may not be fine enough. They point out that the effects of retention zones may only apply at the temporal scale of individual

relaxation events, and not over the scale of seasons or years. As such, they constitute an important demographic structuring force, but appear to have minimal genetic effects. If the genetic effects of upwelling zones associated with headlands are not detectable with hypervariable microsatellite markers on the scale of years, it seems unlikely that such retention zones will have significant evolutionary ramifications for benthic marine organisms with long-lived pelagic larvae.

The sweepstakes-reproduction hypothesis applies to organisms with high fecundity, such as oysters, where a single female can produce over 90 million eggs per month (e.g., Galtsoff 1930). From this perspective, *P. cinctipes*, in which a female averages < 1000 offspring each (Knudsen 1964; M. Donahue, unpubl. data), does not make a likely candidate for high variance in individual reproductive success contributing to temporal variation in allelic frequencies. However, the fact that there are significant temporal fluctuations in allelic frequencies of resident populations suggests a greatly reduced effective population size relative to the numerical abundance of crabs throughout the range. In all comparisons, recruits differed most between years, whereas adults differed least. Furthermore, if recruits and residents that cluster together in the UPGMA dendrograms accurately reflect the source of recruits to a site, the heterozygosity of recruits is significantly reduced relative to the heterozygosity of residents which produced them for both loci (analyses not shown). If recruits are pooled across years and sites, however, the heterozygosity of recruits is significantly higher than that of resident populations. Each of these results are consistent with variation in adult reproductive success leading to reduced effective population size.

Although the sweepstakes-reproduction hypothesis along may be unable to account for the pattern of genetic structure in *P. cinctipes*, these data are consistent with a reduced number of breeders producing the recruits sampled in each year.

Finally, there are concordant changes in both space and time with regard to the number of alleles detected per locus, a pattern consistent with the action of natural selection (e.g., Sun et al. 2001). In this case, however, the differences between years swamp the differences among sites; changes in the number of alleles per locus across years at each site usually exceed that detected across all sites sampled within a year. Interannual variability likewise dominates changes in allelic frequencies at each site. In general, there are greater differences in the frequency of alleles between sampling years within a given site than among all geographic sites sampled within a year. There is no obvious correlation with weather patterns across years to suggest a simple environmental mechanism of natural selection, however. Given that 1997 and 1998 were El Niño years, whereas 1999 was a La Niña year, the simplest prediction based on abiotic natural selection would be that 1997 and 1998 should be more similar to one another than either is to 1999. However, in terms of numbers of alleles per locus, allelic frequencies and genetic similarity, 1997 and 1999 samples consistently resemble one another more closely than samples drawn from 1998. This result certainly does not eliminate natural selection as an importance source of variation in these data, but it suggests that the mechanism by which selection affects allelic frequencies in *P. cinctipes* is not a simple consequence of predominant coastal weather patterns.

Thus, patterns of genetic structure in *Petrolisthes cinctipes* are consistent with there being significant effects of variation in the source population from which larvae recruit, sweepstakes-reproductive success, and natural selection/local adaptation. In part, this is a failing of the current hypotheses to provide mutually exclusive predictions, and further refinement of these hypotheses to generate unique and falsifiable predictions is needed to tease apart the relative influence of each factor in contributing to the overall pattern of genetic structure in marine species. In addition, we need more detailed empirical studies to confirm the pattern that these hypotheses are trying to explain. Are there general patterns consistent among groups of marine organisms, or does each species present a unique combination of life history mode, natural history and evolutionary background that effectively prevent useful generalizations being made? In order to refine theoretical expectations for marine populations that may often be far from equilibrium, which are linked by complex patterns of dispersal, and in which a variety of contemporary and historical processes are likely to acting synergistically, much more detailed empirical studies are needed to characterize the important generalities and differences among species.

Table 1. Summary of genetic information based on two microsatellite loci isolated from *Petrolisthes cinctipes*. Forward and reverse primer sequences (5' → 3'), fluorescent labeling, number of alleles per locus, the number of crabs for which the locus was scored (N), the observed and expected heterozygosities for each locus pooled across residents of all populations studied, and the probability value for conformity to Hardy-Weinberg equilibrium (see text).

Locus	Primer Sequence	N	No. of alleles	Heterozygosity		H-W P-value (± SE)
				Observed	Expected	
Pc156s	HEX - TTG GCT TTG AAG ACC CTG TGG CGG GGG ATC ATT GCT TTG TC	3602	34	0.60	0.68	0.07 (± 0.008)
Pc170s	6FAM - TGG CCG TTG CTG TTG TTG TC GGC ACC AGT CAT TCC CAG TTG	3588	47	0.76	0.83	0.02 (± 0.004)

Table 2. Summary of microsatellite data collected for each size class of *Petrolisthes cinctipes* pooled across sites in Northern California in each of 1997-99. The number of individual crabs genotyped from each size class at each site is presented (N) along with the observed and expected heterozygosities for each locus, and the probability value for conformity to Hardy-Weinberg equilibrium (see text). *P*-values less than 0.5 are highlighted in bold, although after Bonferroni correction none of the deviations from H-W expectation are significant.

Year	Size class	N	Heterozygosity		H-W <i>p</i> -value ± SE
			Observed	Expected	
Locus Pcl56s					
1997	recruits	186	0.65	0.76	0.25 ± 0.01
	subadults	402	0.70	0.74	0.18 ± 0.01
	adults	390	0.54	0.76	0.03 ± 0.01
1998	recruits	518	0.51	0.58	0.11 ± 0.01
	subadults	428	0.51	0.57	0.35 ± 0.02
	adults	426	0.47	0.50	0.24 ± 0.01
1999	recruits	320	0.66	0.71	0.27 ± 0.01
	subadults	492	0.67	0.77	0.24 ± 0.01
	adults	440	0.65	0.78	0.11 ± 0.01
Locus Pcl70s					
1997	recruits	184	0.65	0.91	0.10 ± 0.01
	subadults	396	0.84	0.87	0.13 ± 0.01
	adults	390	0.81	0.88	0.06 ± 0.01
1998	recruits	516	0.74	0.78	0.01 ± 0.01
	subadults	428	0.72	0.80	0.06 ± 0.01
	adults	426	0.74	0.80	0.24 ± 0.01
1999	recruits	316	0.72	0.82	0.20 ± 0.01
	subadults	492	0.79	0.78	0.16 ± 0.01
	adults	440	0.80	0.80	0.24 ± 0.02

Table 3. Annual variation in genetic structure among adult *Petrolisthes cinctipes* populations.

Locus	N	<i>F</i> (within total)	θ (among sites)	<i>f</i> (within sites)	ρ (among sites)
1997					
Pc156s	390	0.23	0.07	0.18	0.09
Pc170s	390	0.07	0.06	0.01	0.08
Overall	390	0.15	0.07	0.09	0.09
1998					
Pc156s	426	- 0.07	0.07	- 0.16	0.03
Pc170s	426	0.05	0.08	- 0.04	0.03
Overall	426	0.01	0.08	- 0.08	0.03
1999					
Pc156s	440	0.19	0.07	0.12	0.06
Pc170s	440	0.03	0.09	- 0.06	0.06
Overall	440	0.11	0.08	0.03	0.06
Overall (pooled years)					
Pc156s	1256	0.20	0.02	0.18	0.01
Pc170s	1256	0.06	0.03	0.03	0.02
Overall	1256	0.12	0.02	0.10	0.01

Table 4. Variation in genetic structure among years partitioned by size class (pooled across loci and sites) for *Petrolisthes cinctipes* populations.

Size class	N	F (within total)	θ (among sites)	f (within sites)	ρ (among sites)
1997					
Recruits	186	0.20	0.08	0.14	0.14
Subadults	402	0.08	0.08	0.00	0.06
Adults	390	0.15	0.07	0.09	0.10
1998					
Recruits	518	0.09	0.10	- 0.01	0.07
Subadults	428	0.06	0.06	0.02	0.02
Adults	426	0.01	0.08	- 0.08	0.07
1999					
Recruits	320	0.09	0.12	- 0.03	0.12
Subadults	492	0.04	0.05	- 0.01	0.07
Adults	440	0.11	0.08	0.03	0.06

Table 5. Variation in genetic structure among years partitioned by size class and loci (pooled across sites) for *Petrolisthes cinctipes* populations.

Locus	N	F (within total)	θ_r (among years)	θ_s (among sites)	f (within years & sites)
Recruits					
Pc156s	1024	0.15	0.11	0.03	0.05
Pc170s	1016	0.12	0.12	0.0005	- 0.001
Overall	1016	0.13	0.12	0.01	0.02
Subadults					
Pc156s	1322	0.22	0.11	0.04	0.02
Pc170s	1316	0.05	0.08	0.02	- 0.01
Overall	1316	0.13	0.09	0.03	0.002
Adults					
Pc156s	1256	0.10	0.09	0.06	0.12
Pc170s	1256	0.06	0.07	0.05	- 0.03
Overall	1256	0.08	0.08	0.05	0.04

Table 6. Variation in genetic structure by year and locus for *Petrolisthes cinctipes* populations.

Year	Locus	N	<i>F</i> (within total)	θ_r (among sites)	θ_s (among size classes)	<i>f</i> (within sites & size classes)
1997	Pc156s	978	0.18	0.04	0.07	0.12
	Pc170s	970	0.10	0.06	0.08	0.03
	Overall	970	0.14	0.05	0.08	0.07
1998	Pc156s	1372	0.04	0.02	0.07	- 0.04
	Pc170s	1370	0.07	0.05	0.08	- 0.02
	Overall	1370	0.06	0.04	0.08	- 0.02
1999	Pc156s	1252	0.13	0.04	0.06	0.08
	Pc170s	1248	0.04	0.04	0.10	- 0.07
	Overall	1248	0.08	0.04	0.08	0.01
Overall results using all samples						
Locus	<i>F</i>	θ_r (years)	θ_s (sites)	θ_t (size classes)	<i>f</i>	
Pc156s	0.13	0.05	- 0.01	0.08	0.06	
Pc170s	0.08	0.04	0.01	0.08	- 0.02	
Overall	0.10	0.05	- 0.0001	0.08	0.02	

Table 7. Hierarchical analysis of molecular variance (AMOVA) results for *Petrolisthes cinctipes* examining the relative contribution of size class, sampling year, and site. The relative proportion of genetic variation explained by years, sites, and size class are assessed. Probability of obtaining a more extreme variance component and Φ statistic by chance alone was determined by permutation test (500,000) in Arlequin, and significant values ($\alpha = 0.01$) are denoted by an asterisk.

Source of Variation	d.f.	Variance component	% of variation	Φ statistic
Sites within Years				
Among years	2	19547.953	5.29	$\Phi_{CT} = 0.05^*$
Among sites, within years	18	2660.934	0.52	$\Phi_{SC} = 0.005^*$
Within sites	7170	501213.447	94.19	$\Phi_{ST} = 0.06^*$
Size Classes within Sites				
Among sites	7	6837.629	0.83	$\Phi_{CT} = 0.008$
Among size classes, within sites	16	6756.074	1.31	$\Phi_{SC} = 0.01^*$
Within size classes	7164	509842.098	97.86	$\Phi_{ST} = 0.02^*$
Size Classes within Years				
Among years	2	19551.381	4.67	$\Phi_{CT} = 0.05^*$
Among size classes, within years	6	25504.378	5.13	$\Phi_{SC} = 0.05^*$
Within size classes	7194	478379.715	90.20	$\Phi_{ST} = 0.10^*$

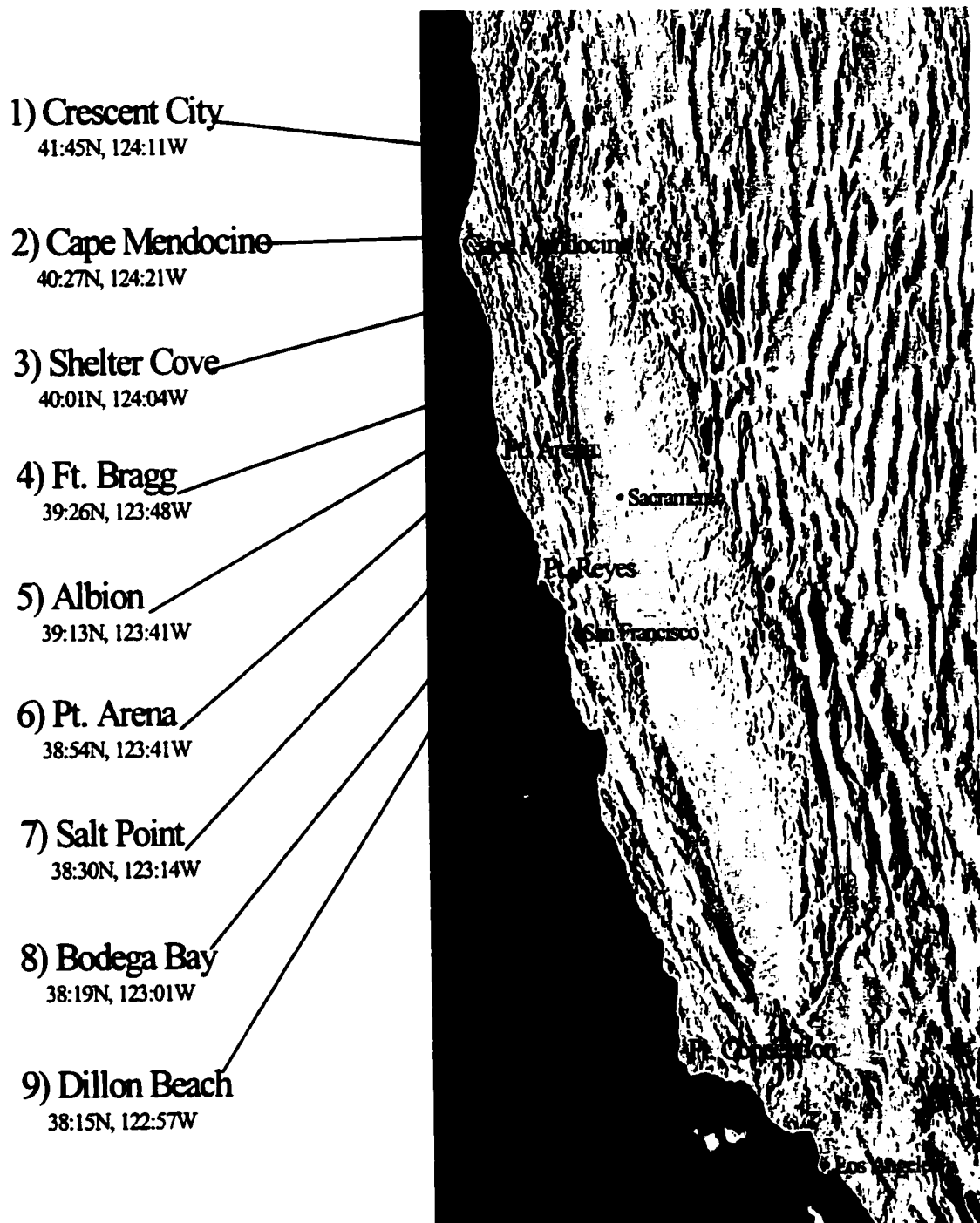


Figure 1. Northern California collection sites for *Petrolisthes cinctipes*. Each site is numbered from north to south, and the GPS coordinates for each are presented below the location.

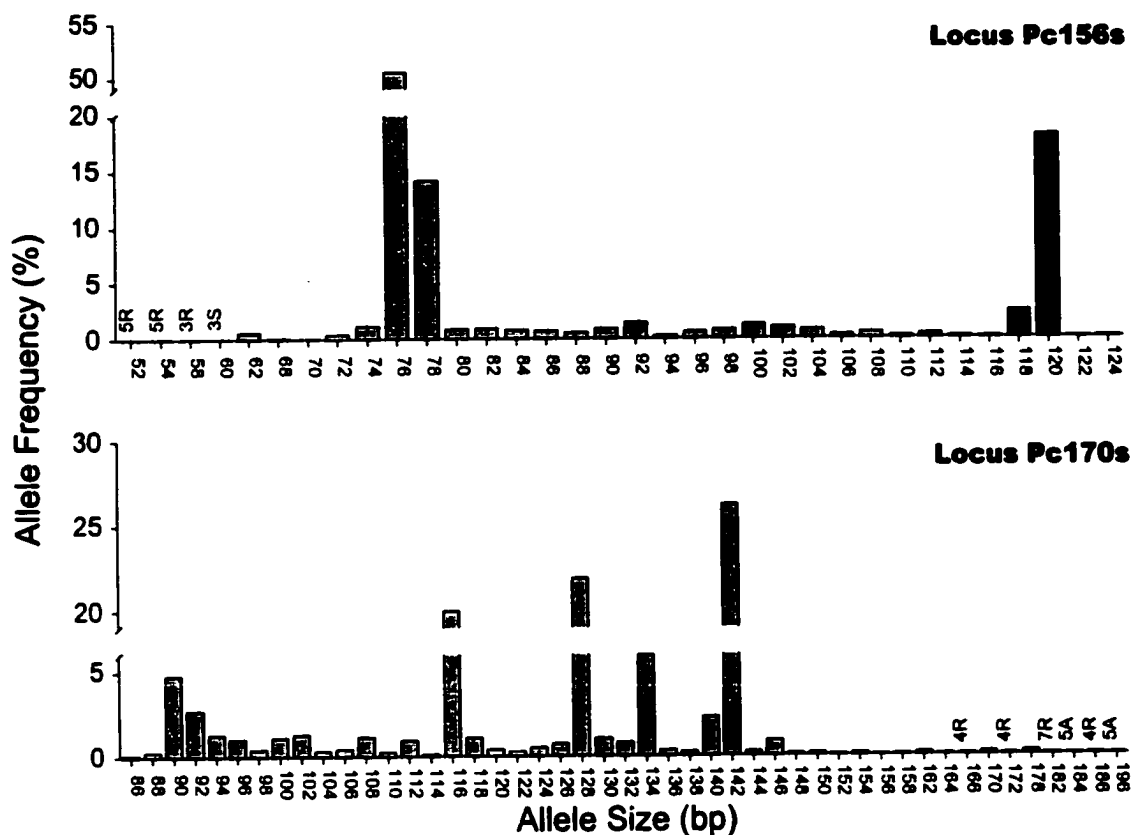


Figure 2. Allele frequency distributions of the two polymorphic microsatellite loci amplified from the 3602 individual *Petrolisthes cinctipes* genotyped in this study. For both loci, the three most common alleles represented more than 80% of those scored. Private alleles (those detected in individuals from only a single site) are identified with the number of the site (corresponding to Figure 1), and the size class (A = adults, S = subadults, R = recruits) in which the private allele was detected.

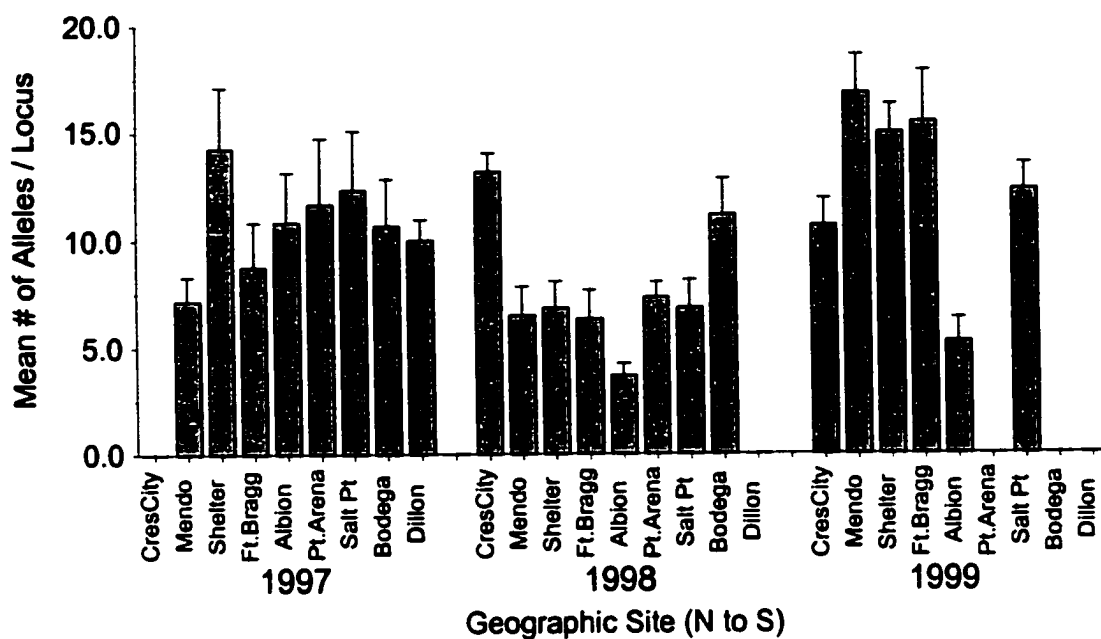


Figure 3. Number of alleles detected at each locus among samples of *Petrolisthes cinctipes* collected for each year of this study. Mean number of alleles per locus does not differ significantly among sites ($F = 1.86$, $df = 7$, $p > 0.05$) or size classes ($F = 0.08$, $df = 2$, $p > 0.05$), but differences between years, however, were significant ($F = 11.30$, $df = 2$, $p < 0.01$). Missing bars for a site represent samples that were not available.

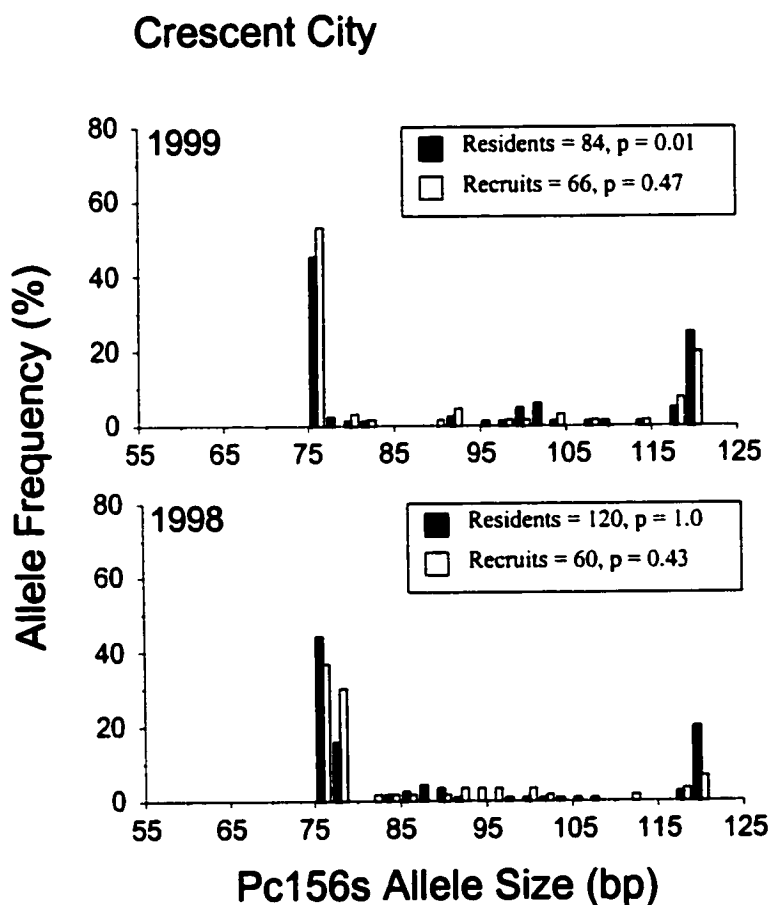


Figure 4A. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Crescent City for 1998 and 1999. Residents consist of both subadults and adults. Sample size is presented along with the probability value for conformity to Hardy-Weinberg expectations, derived from Markov-chain permutation as outlined in the text (500,000 steps and a 10,000 step dememorization). The remaining sites (except Dillon Beach, for which only 1997 samples were collected) are presented on the following pages in order from north to south corresponding to Figure 1.

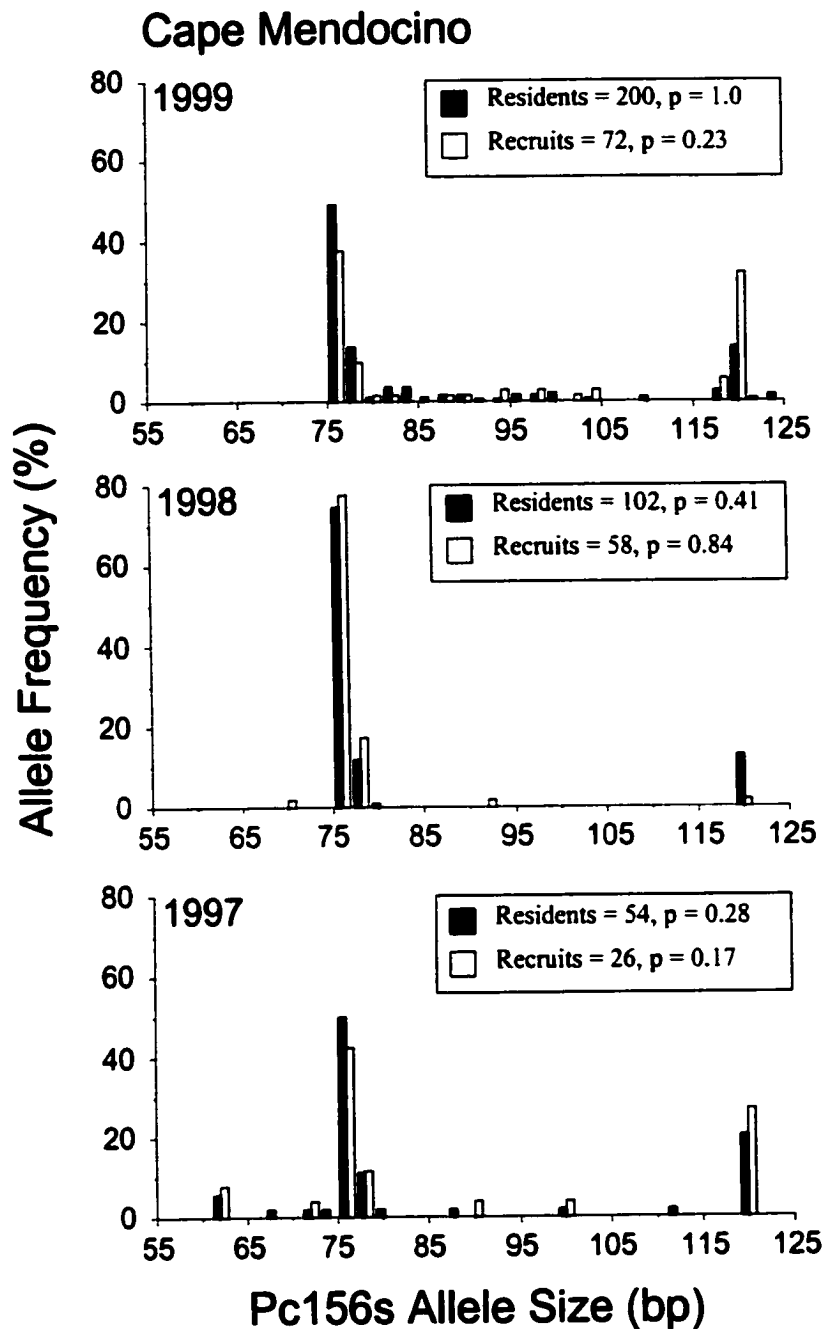


Figure 4B. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Cape Mendocino for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

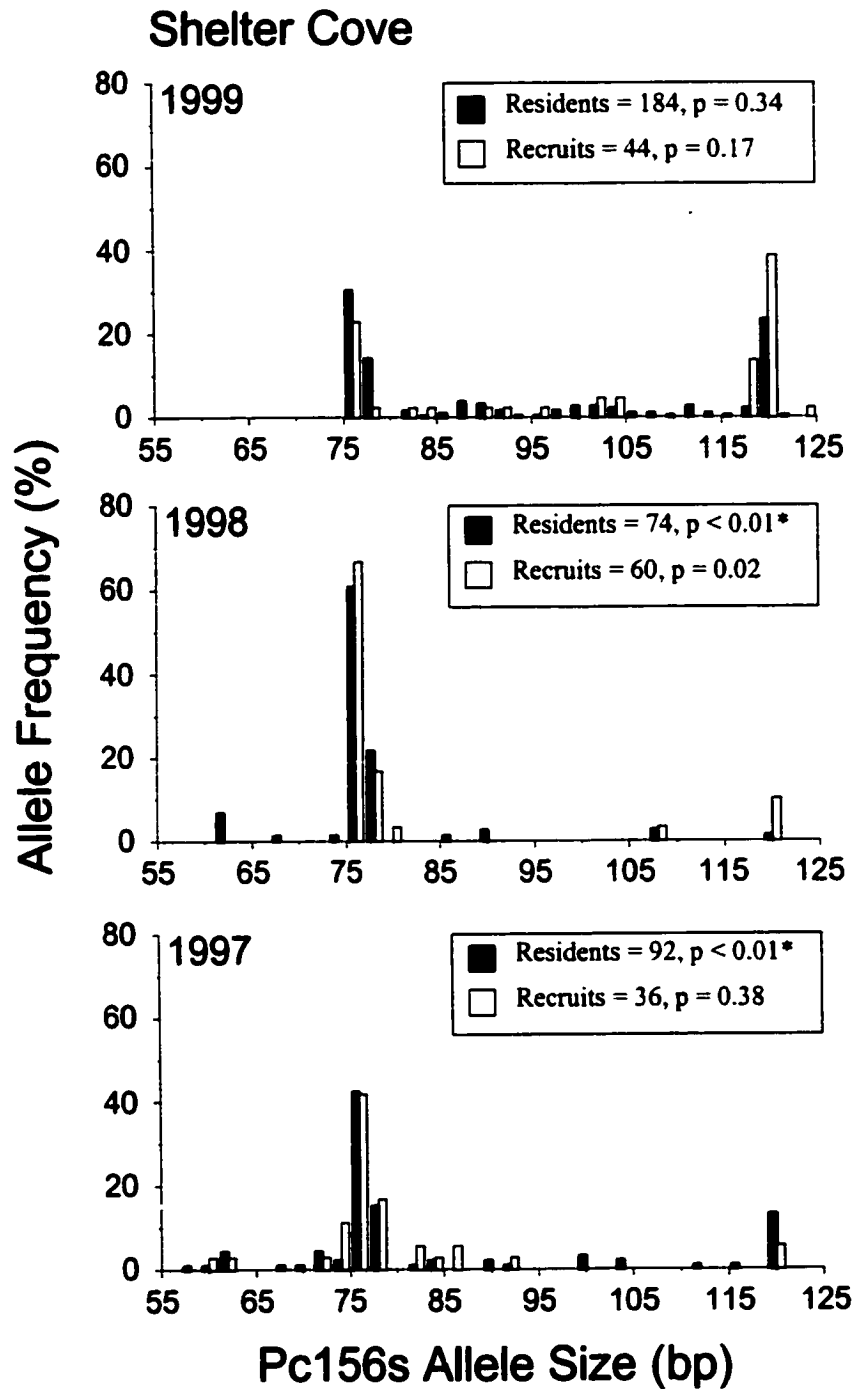


Figure 4C. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Shelter Cove for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

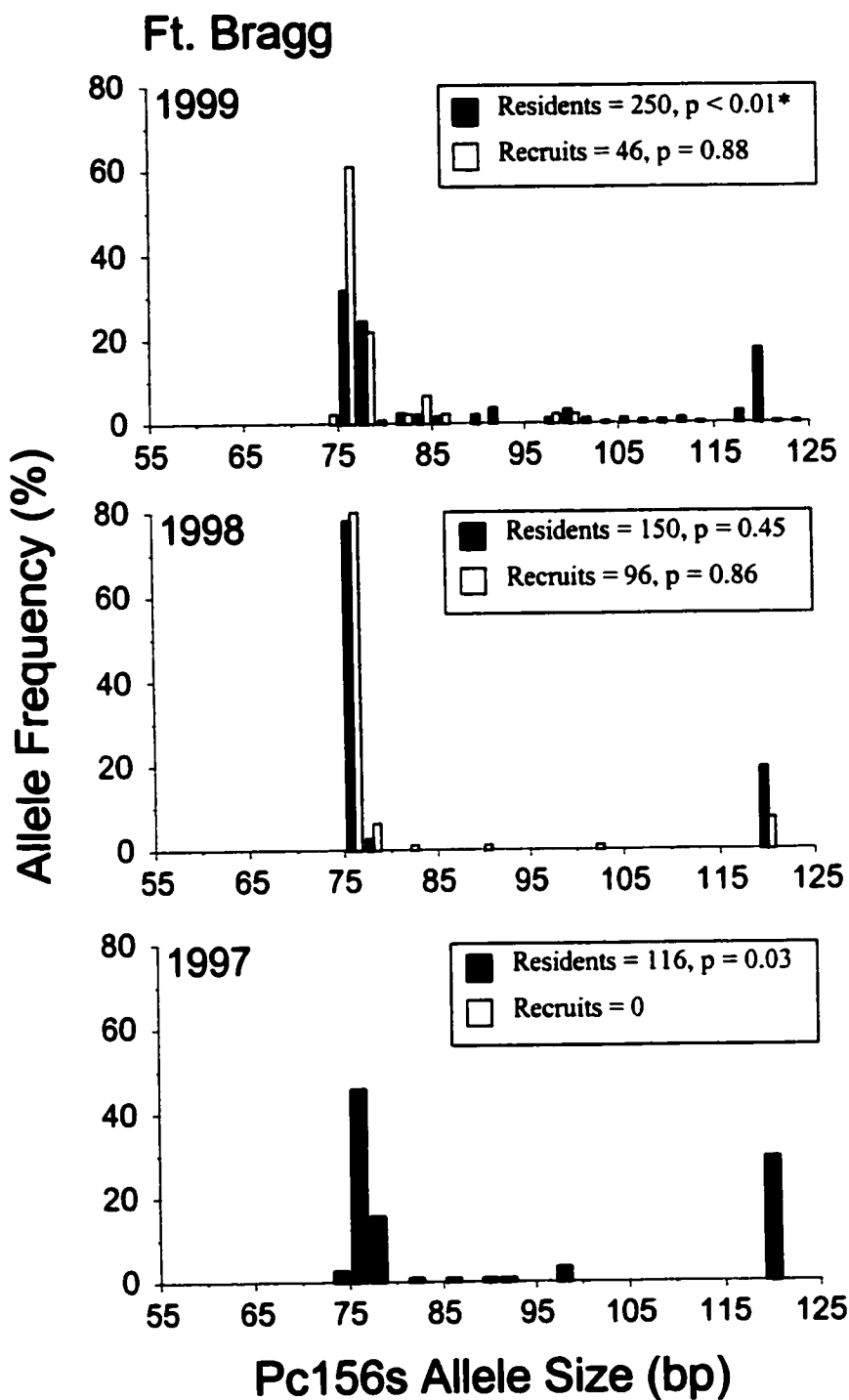


Figure 4D. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Ft. Bragg for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

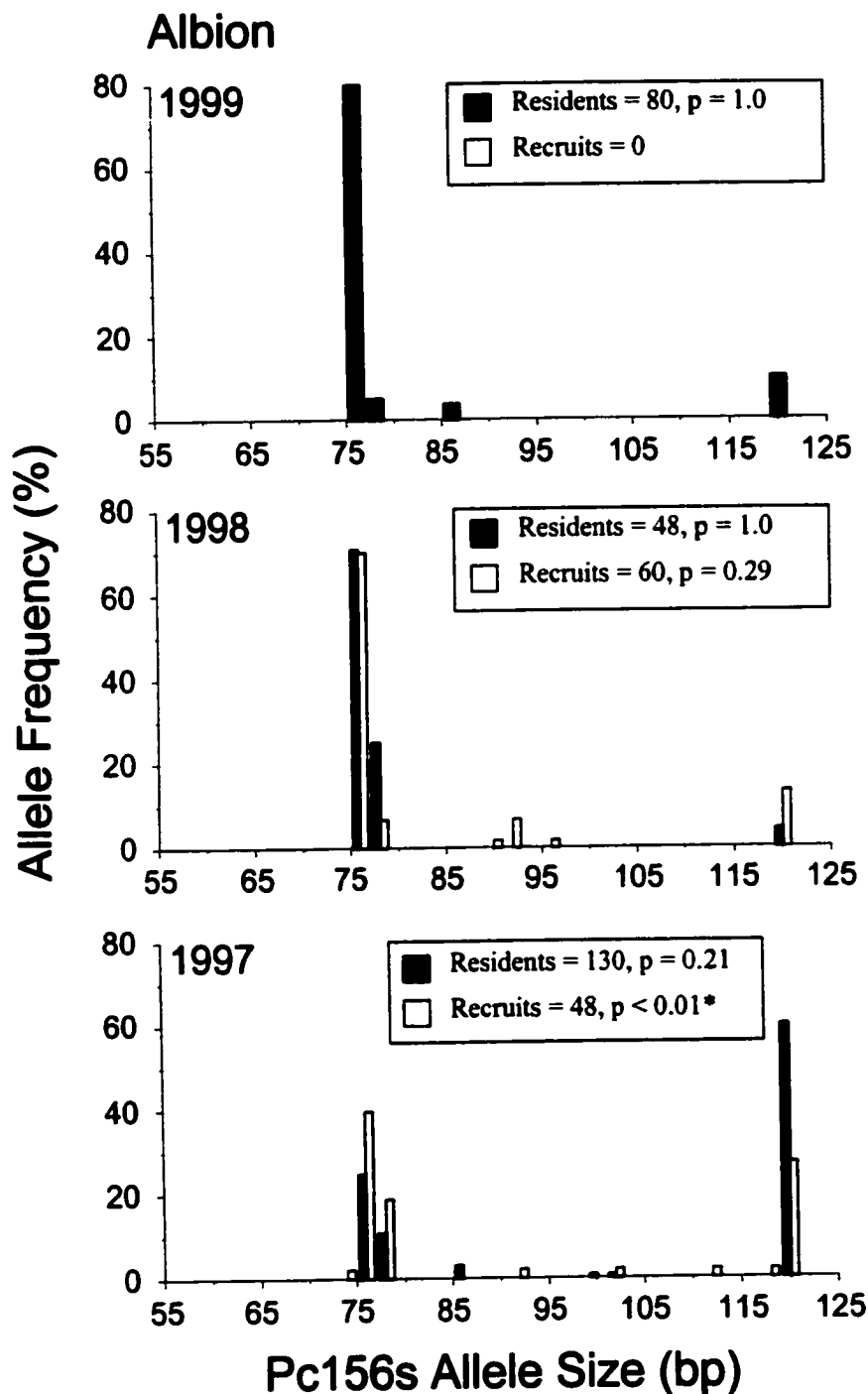


Figure 4E. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Albion for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

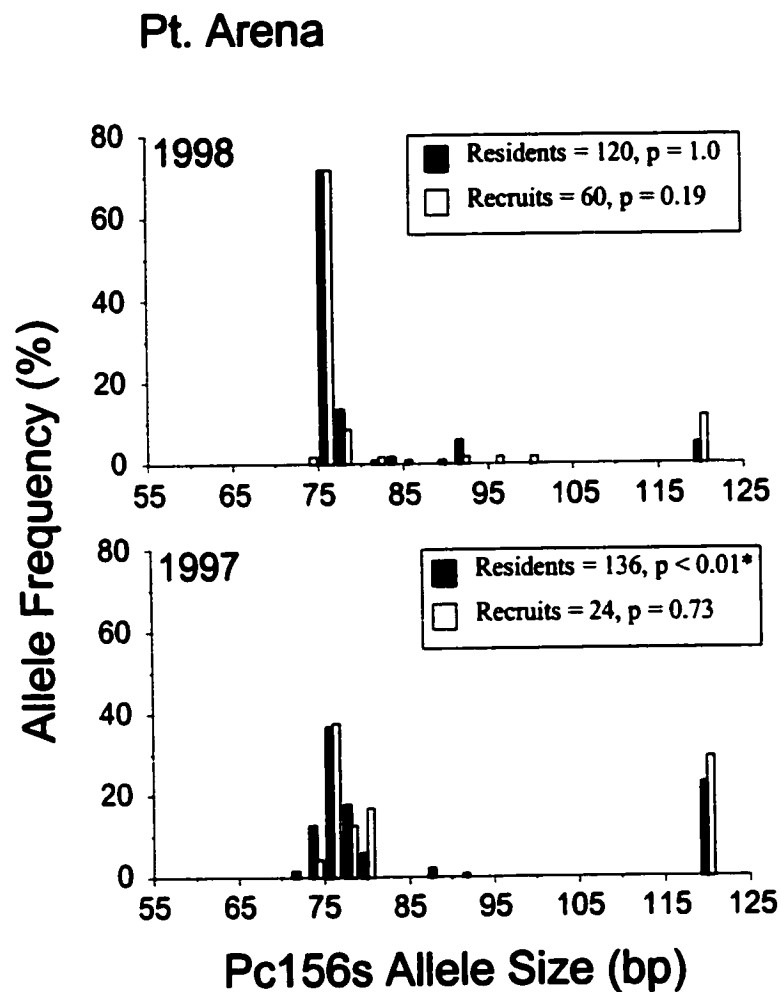


Figure 4F. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Pt. Arena for 1997 and 1998. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

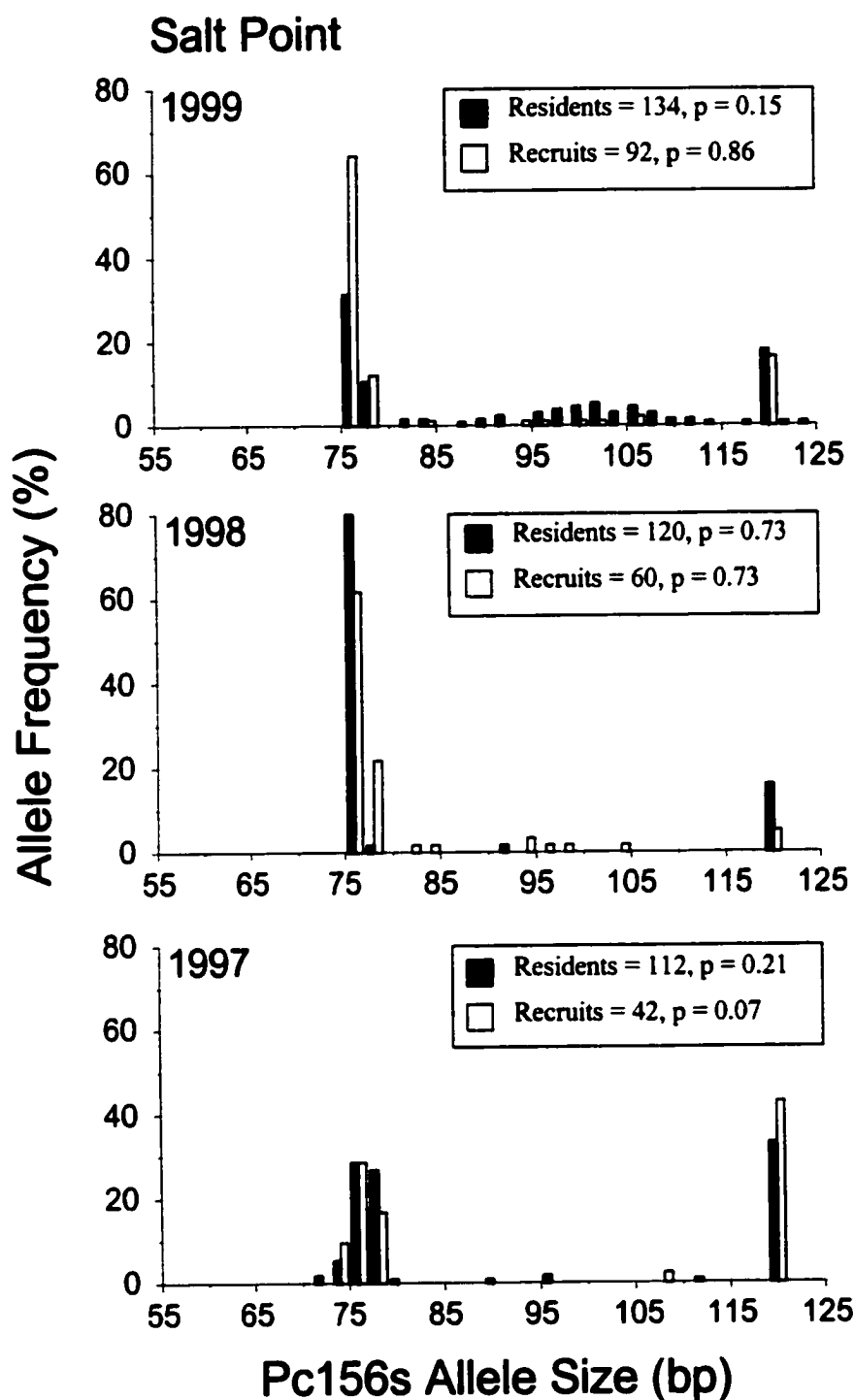


Figure 4G. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Salt Point for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

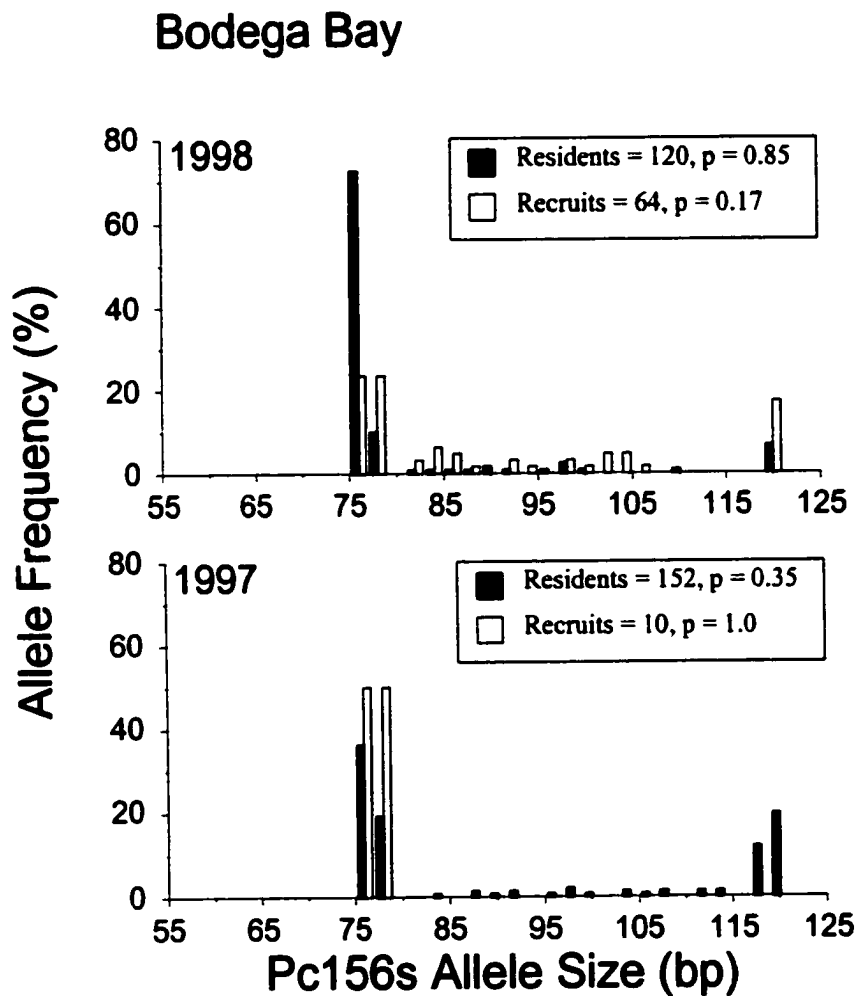


Figure 4H. Allele frequency distributions for locus Pc156s amplified from resident and recruit *Petrolisthes cinctipes* at Bodega Bay for 1997 and 1998. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

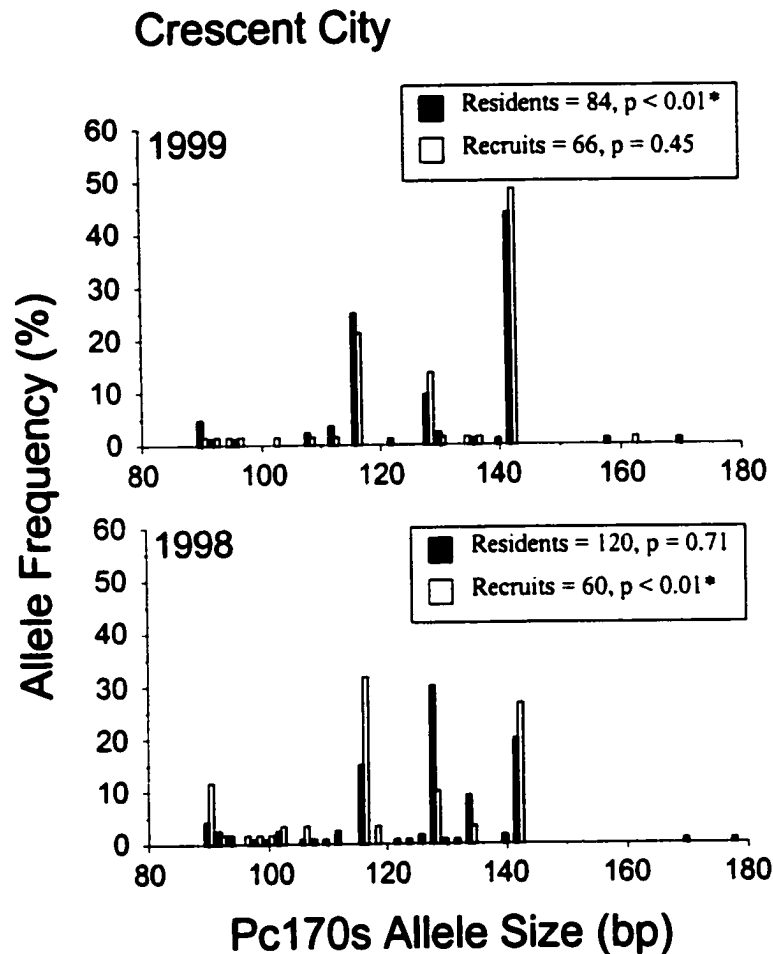


Figure 5A. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Crescent City for 1998 and 1999. Residents consist of both subadults and adults. Sample size is presented along with the probability value for conformity to Hardy-Weinberg expectations, derived from Markov-chain permutation as outlined in the text (500,000 steps and a 10,000 step dememorization). The remaining sites (except Dillon Beach, for which only 1997 samples were collected) are presented on the following pages in order from north to south corresponding to Figure 1.

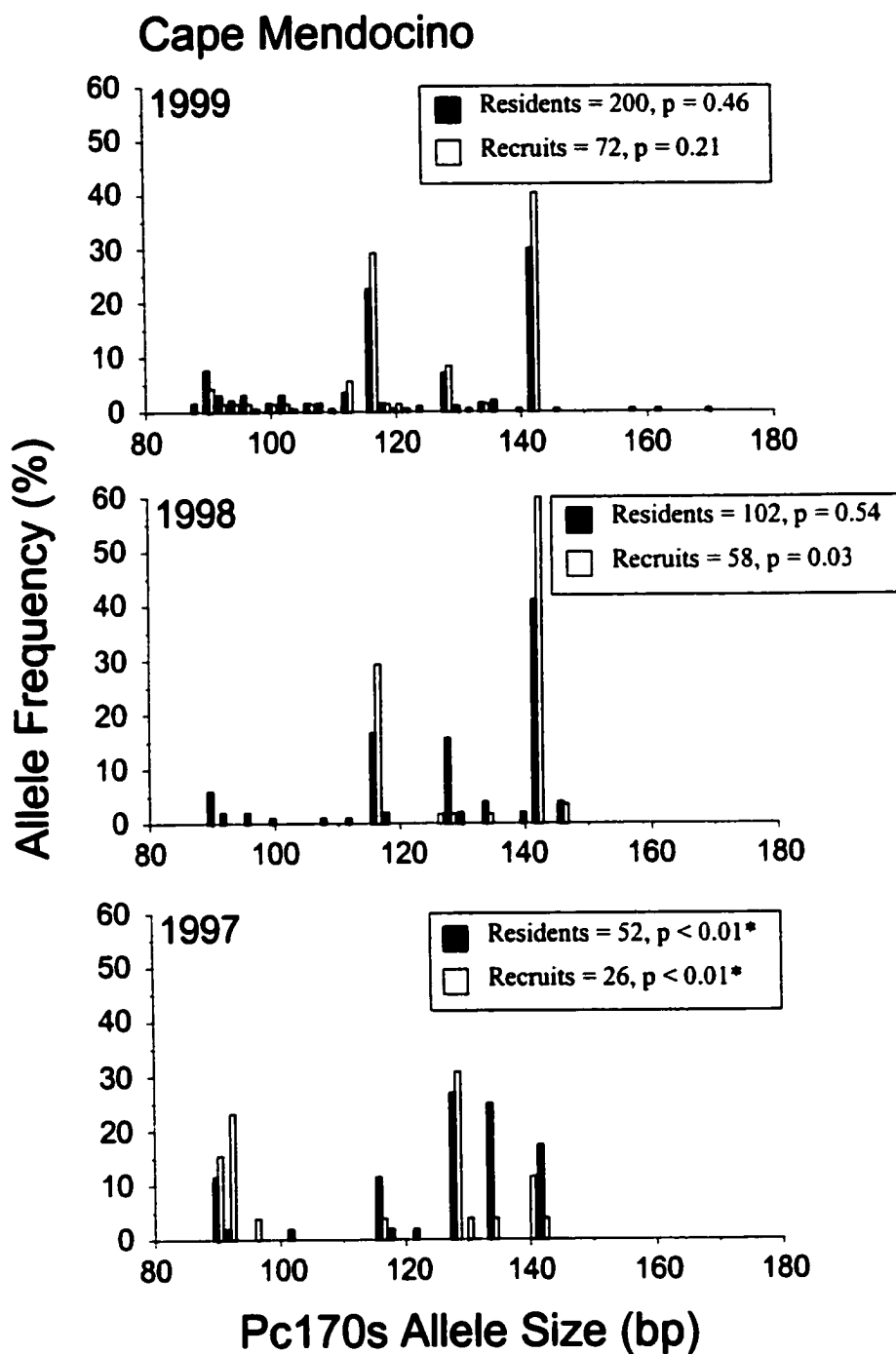


Figure 5B. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Cape Mendocino for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

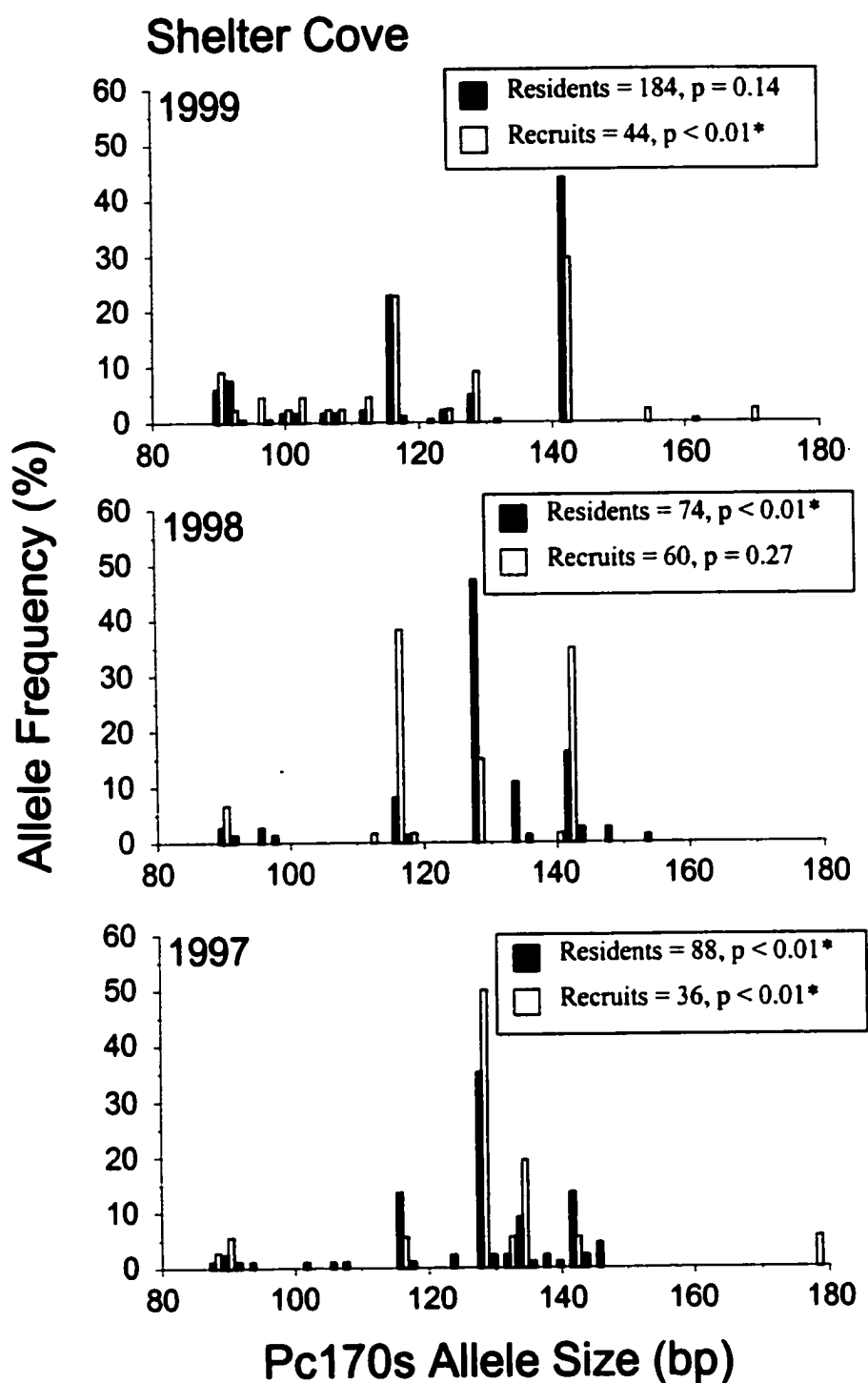


Figure 5C. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Shelter Cove for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

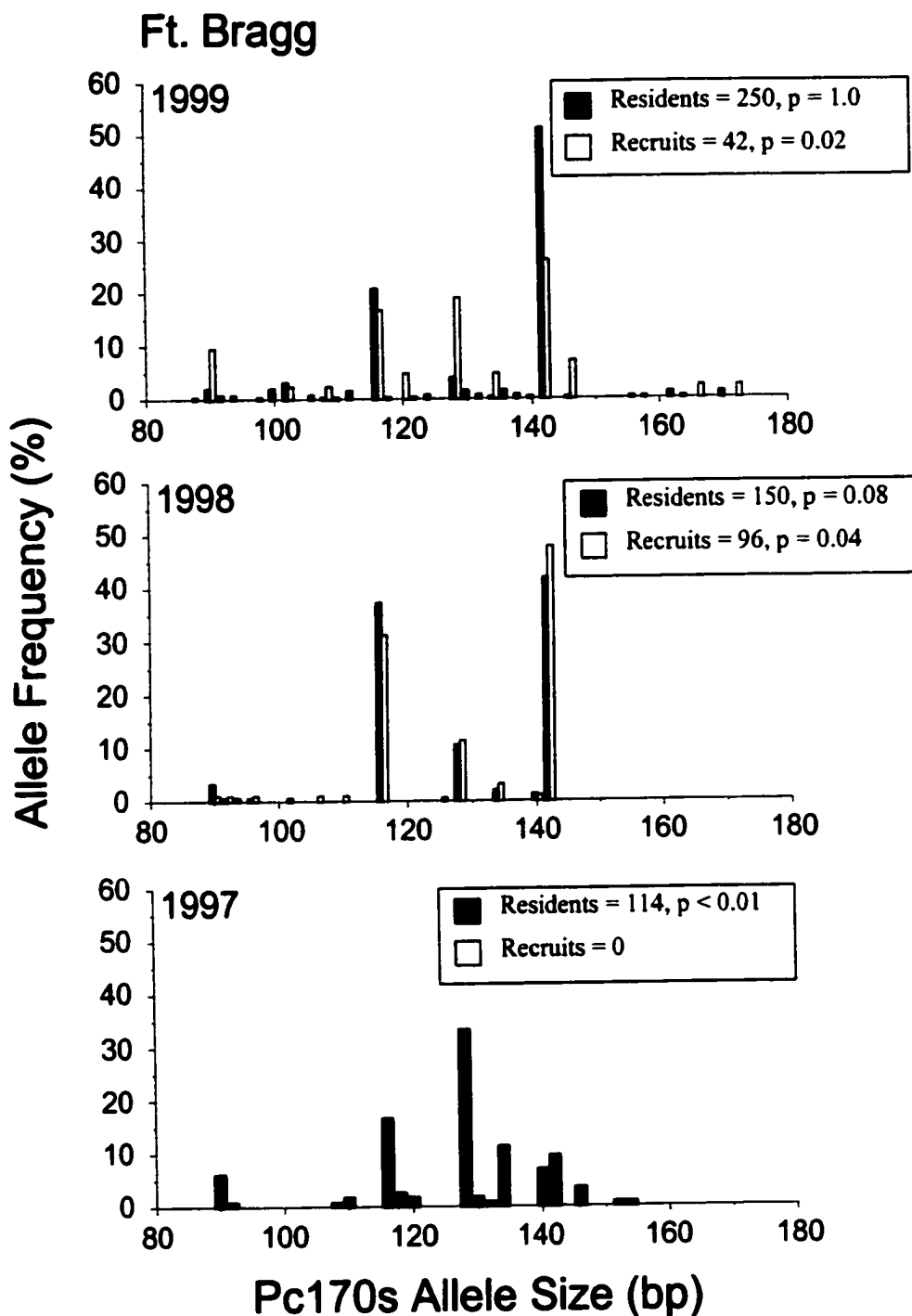


Figure 5D. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Ft. Bragg for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

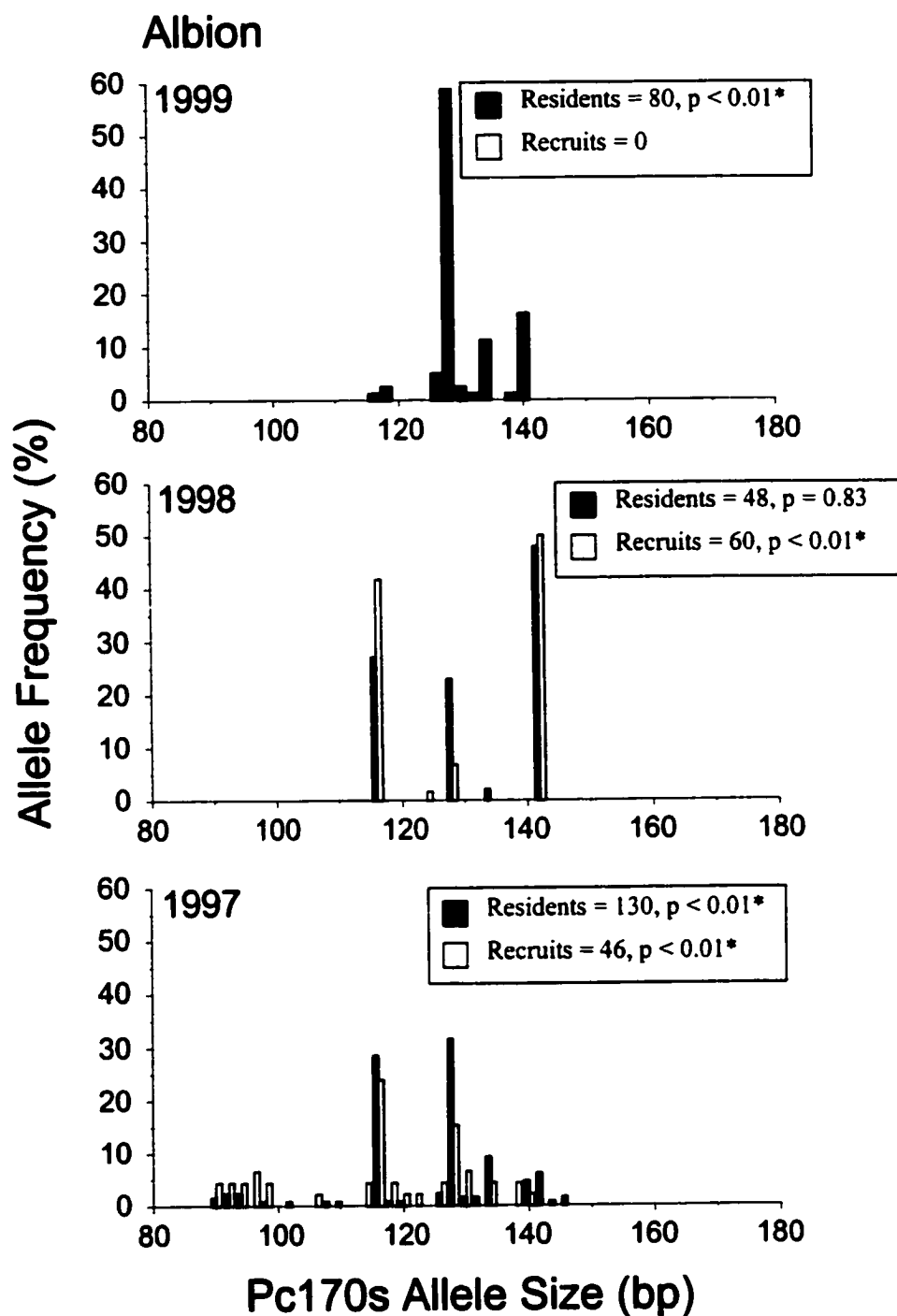


Figure 5E. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Albion for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

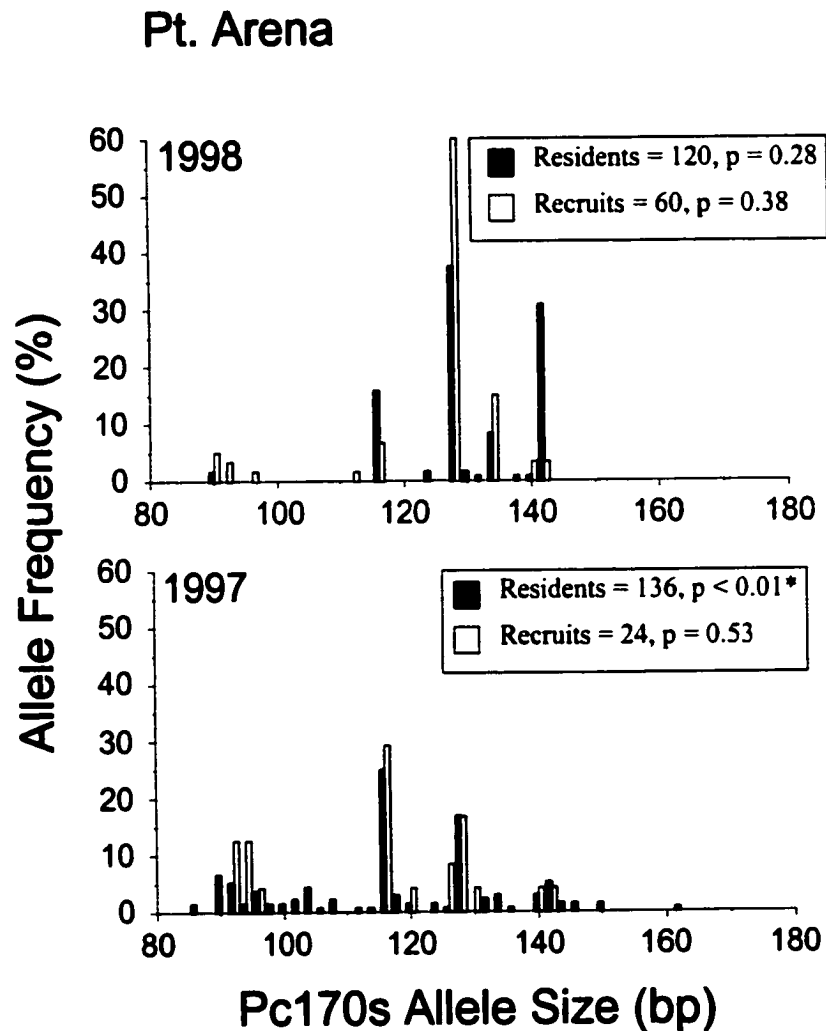


Figure 5F. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Pt. Arena for 1997 and 1998. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

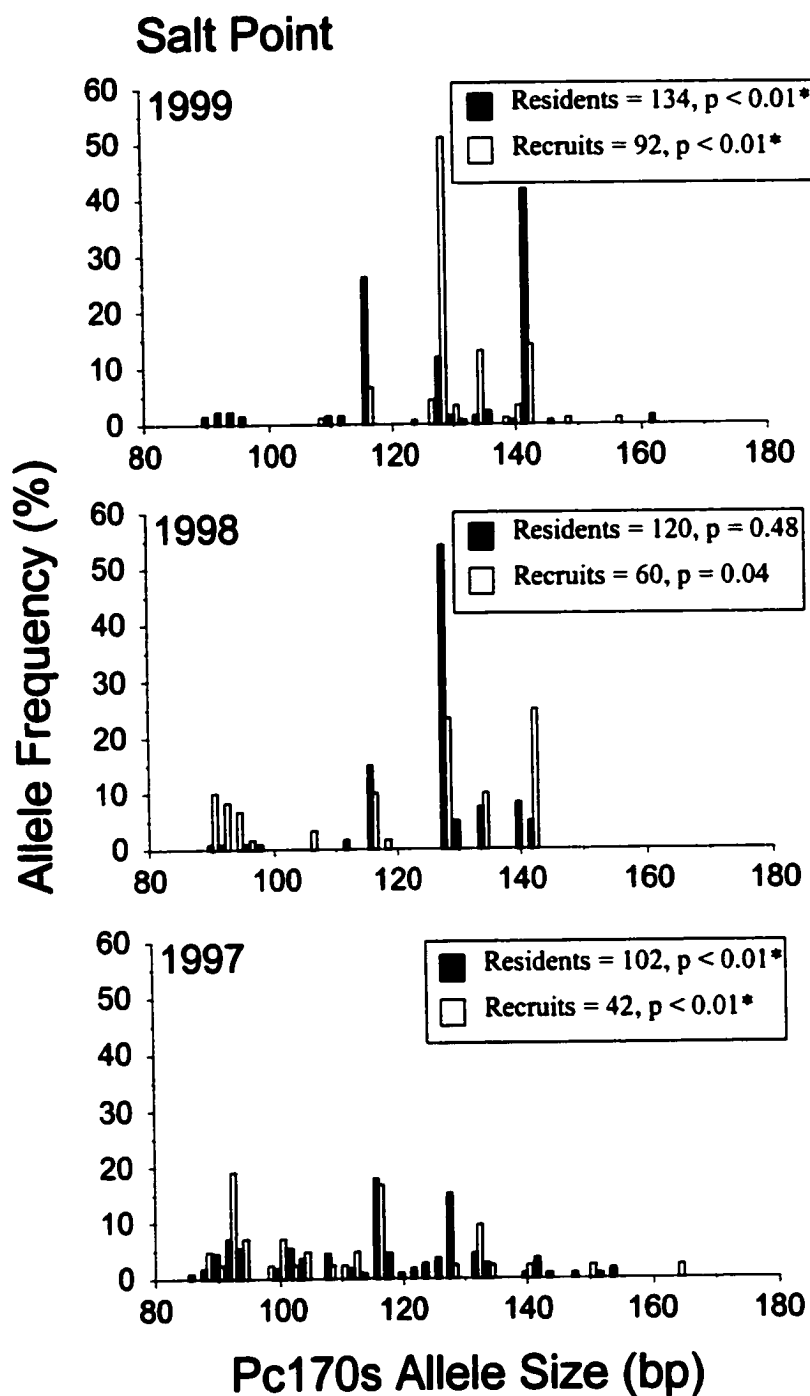


Figure 5G. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Salt Point for each year of the study. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

Bodega Bay

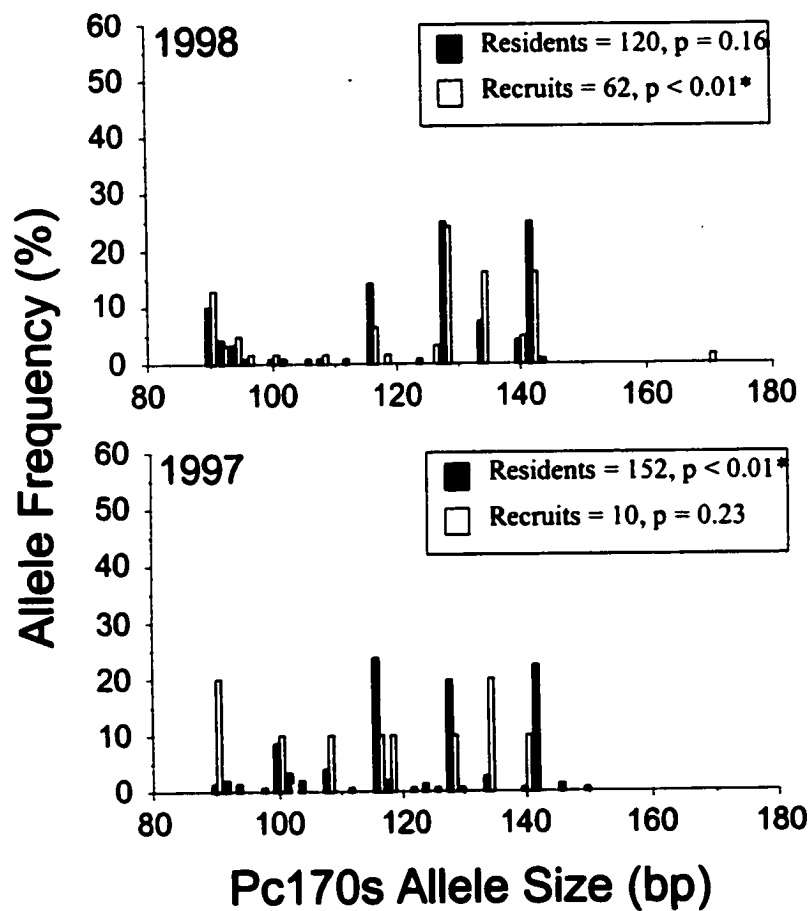


Figure 5H. Allele frequency distributions for locus Pc170s amplified from resident and recruit *Petrolisthes cinctipes* at Bodega Bay for 1997 and 1998. Sample size for residents (both subadults and adults) and recruits are presented along with the probability value for conformity to Hardy-Weinberg expectations.

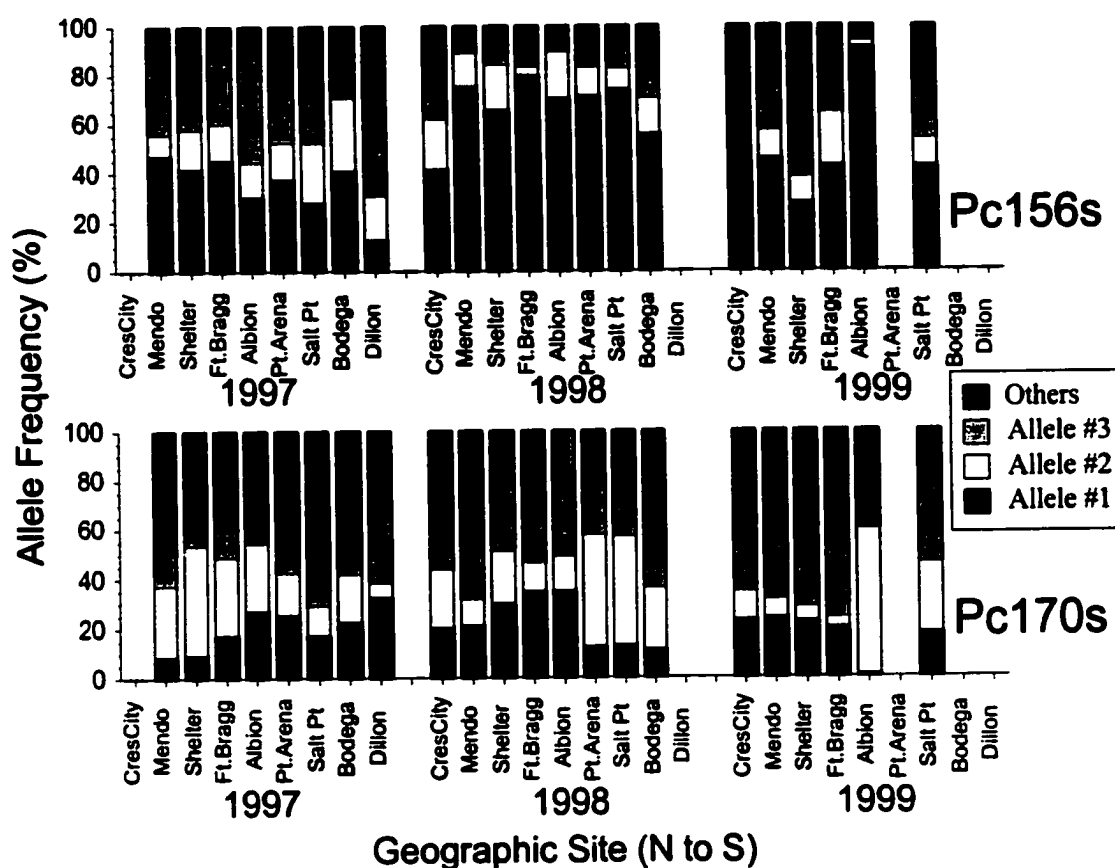


Figure 6. Allele frequencies of the three most common alleles at each locus for which *Petrolisthes cinctipes* were scored. The most common allele is listed as Allele #1, the second most common as Allele #2, and the third as Allele #3. All other alleles are lumped into the Others category for presentation. Corresponding to the allelic frequency distributions presented in Figure 2, Alleles #1, 2 & 3 are 76, 120 & 78 bp, respectively, for locus Pc156s, and 142, 128 & 116 bp, respectively, for locus Pc170s.

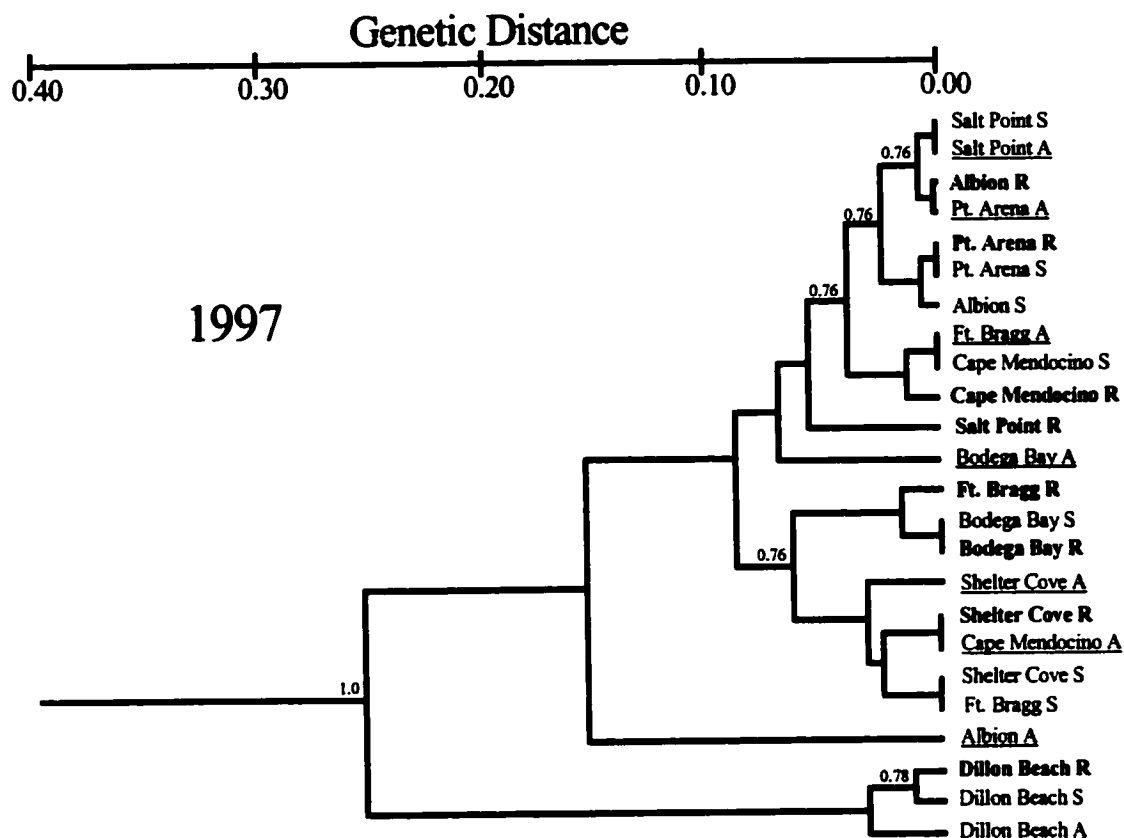


Figure 7. UPGMA dendrogram for *Petrolisthes cinctipes* from each geographic site sampled in 1997, based on the D_A genetic distance of Nei et al (1983). Recruits sampled from each site are denoted in bold (**Site R**), adults are underlined (Site A), and subadults are presented in normal text. Values from 10,000 bootstrap replicates that equaled or exceeded 0.75 are presented next to the nodes.

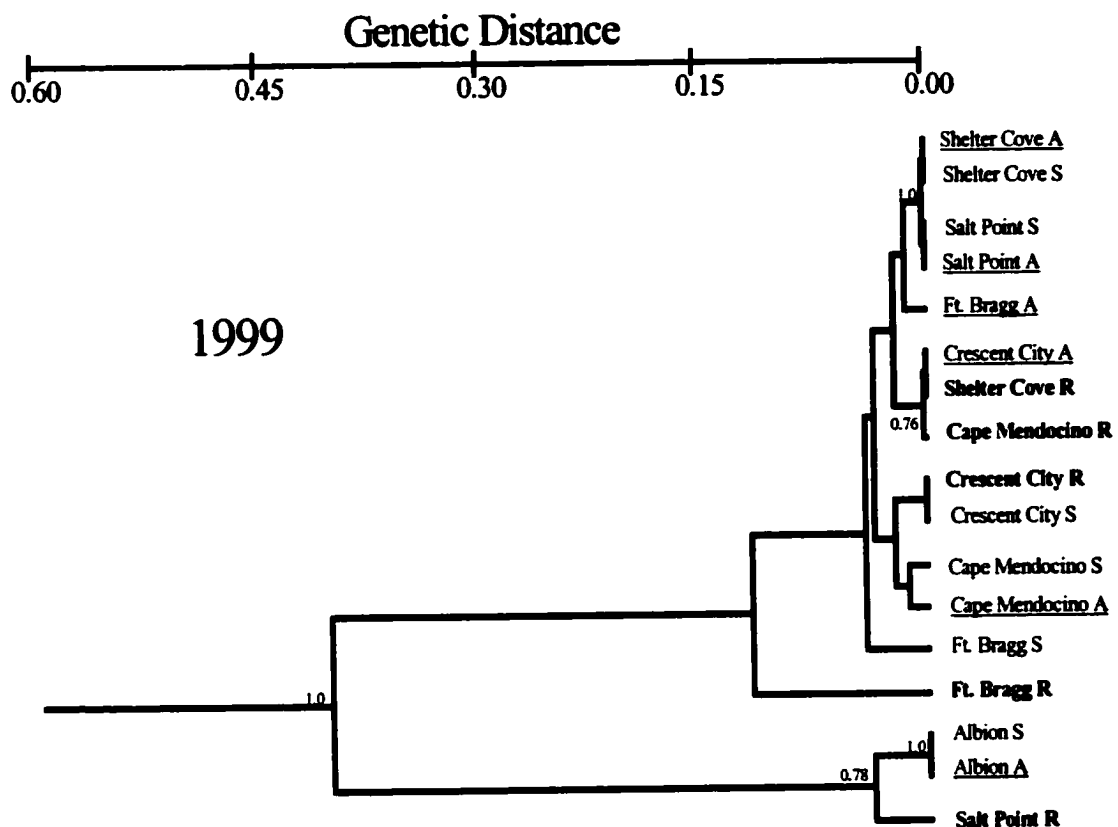


Figure 9. UPGMA dendrogram for *Petrolisthes cinctipes* from each geographic site sampled in 1999, based on the D_A genetic distance of Nei et al (1983). Recruits sampled from each site are denoted in bold followed by 'R' (Site R), adults are underlined followed by 'A' (Site A), and subadults are presented in normal text followed by 'S' (Site S). Values from 10,000 bootstrap replicates that exceeded 0.75 are presented next to the nodes.

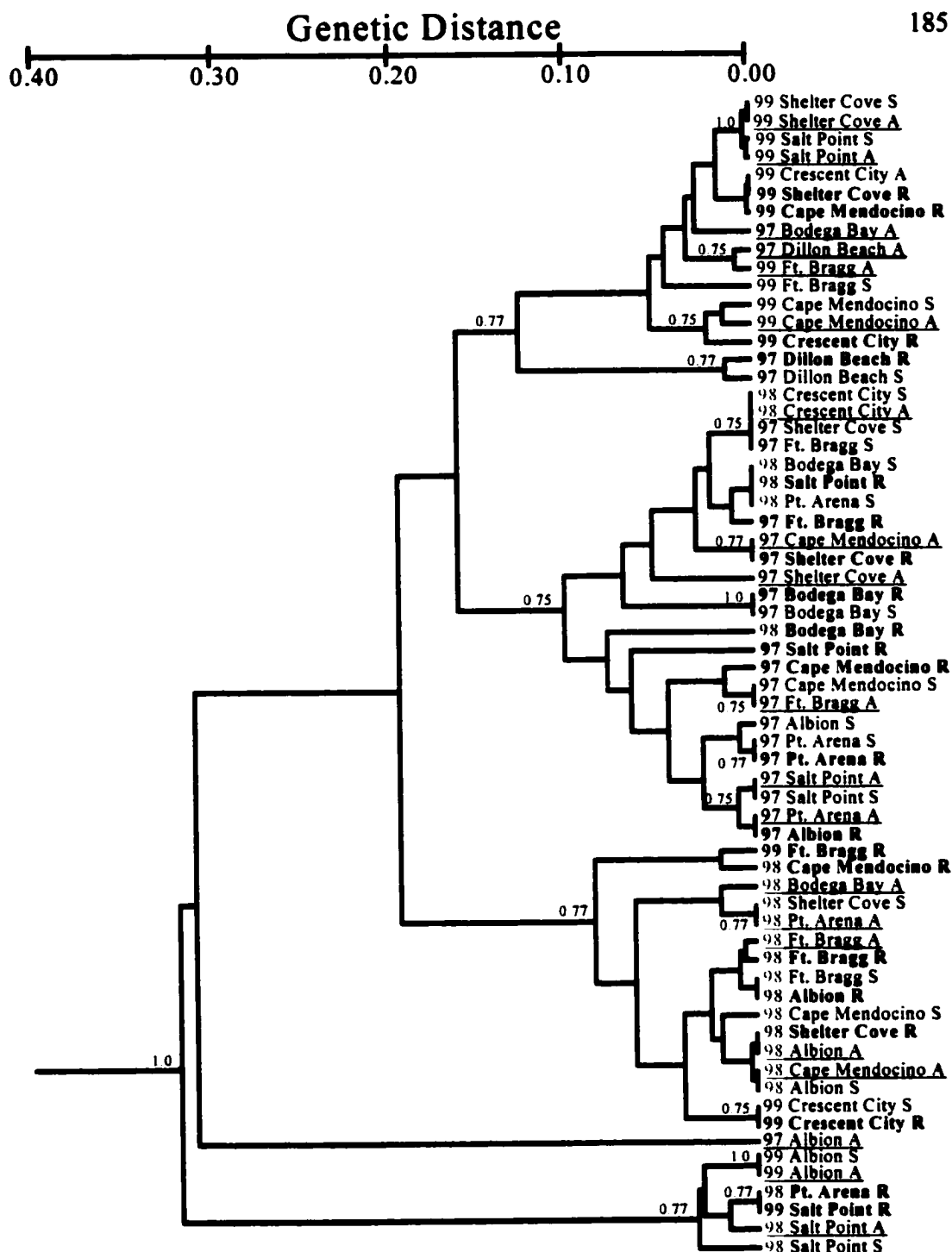


Figure 10. UPGMA dendrogram for *Petrolisthes cinctipes* from each geographic site sampled across all three years of the study, based on the D_A genetic distance of Nei et al (1983). Again, recruits sampled from each site are denoted in bold followed by 'R' (Site R), adults are underlined followed by 'A' (Site A), and subadults are presented in normal text followed by 'S' (Site S). Additionally, the year in which samples were collected is presented, with 1999 in green, 1998 in pink, and 1997 in black. Values from 10,000 bootstrap replicates that equaled or exceeded 0.75 are presented next to the nodes.

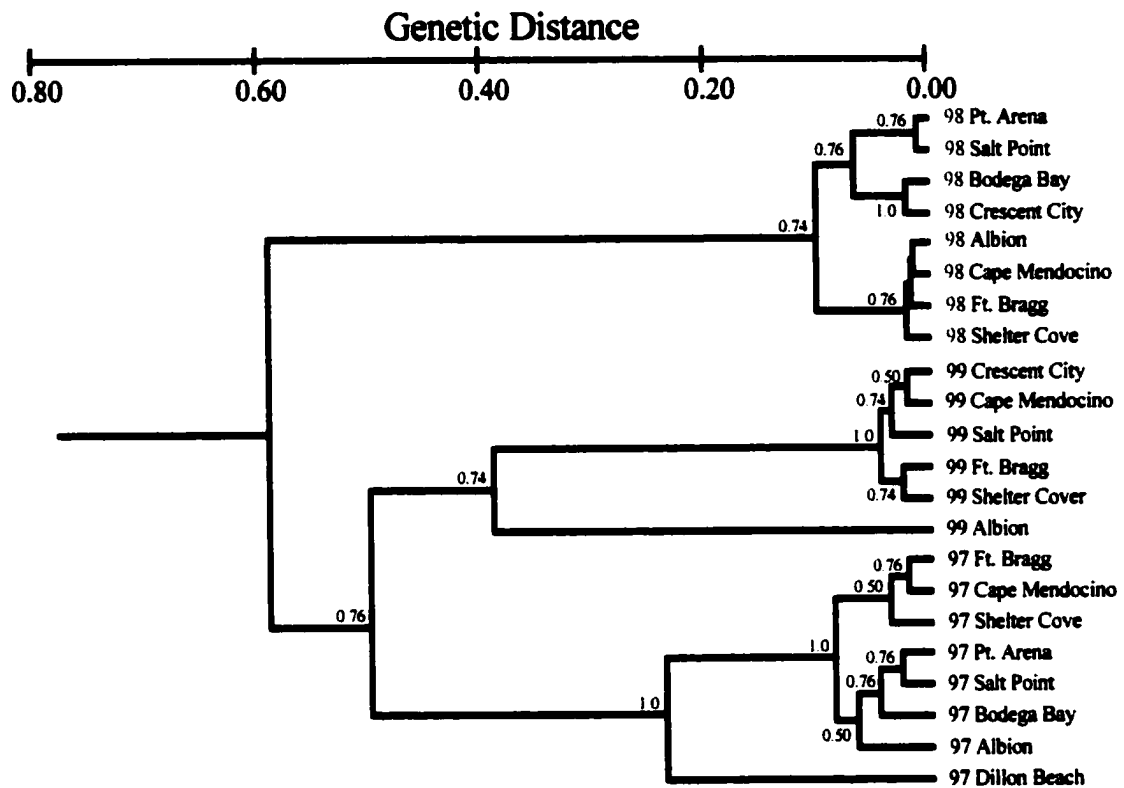


Figure 11. UPGMA dendrogram for *Petrolisthes cinctipes* from each geographic site sampled in 1997, based on the D_A genetic distance of Nei et al (1983). Size classes at each site are pooled, so only site and year labels are included. Again, 1999 samples are presented in green, 1998 in pink, and 1997 in black. Values from 10,000 bootstrap replicates that equaled or exceeded 0.75 are presented next to the nodes.

Literature Cited

- Awise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. 511 pp. Chapman & Hall, New York, NY.
- Ayre, D. J. 1990. Population subdivision in Australian temperate marine invertebrates: Larval connections versus historical factors. *Australian Journal of Ecology* 15:403-412.
- Ayre, D. J., and T. P. Hughes. 2000. Genotypic diversity and gene flow in brooding and spawning corals along the Great Barrier Reef, Australia. *Evolution* 54:1590-1605.
- Banks, M. A., V. K. Rashbrook, M. J. Calavetta, C. A. Dean, and D. Hedgecock. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. *Canadian Journal of Fisheries and Aquatic Sciences* 57:915-927.
- Begg, G. A., C. P. Keenan, and M. J. Sellin. 1998. Genetic variation and stock structure of school mackerel and spotted mackerel in northern Australian waters. *Journal of Fish Biology* 53:543-559.
- Benzie, J. A. H. 1999. Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoologist* 39:131-145.
- Benzie, J. A. H. 2000. The detection of spatial variation in widespread marine species: Methods and bias in the analysis of population structure in the crown of thorns starfish (Echinodermata: Asteroidea). *Hydrobiologia* 420:1-14.

- Berger, E. M. 1973. Gene-enzyme variation in three sympatric species of *Littorina*.
Biological Bulletin 145:83-90.
- Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. Quarterly Review
of Biology 74:21-45.
- Brykov, V. A., N. Polyakova, L. A. Skurikhina, A. D. Kukhlevsky, O. N. Kirillova, D.
Churikov, A. I. Pudovkin, and A. J. Gharrett. 1999. Analysis of mtDNA
indicates weak temporal genetic heterogeneity in pink salmon spawning runs in
two rivers on Sakhalin Island. Journal of Fish Biology 55:617-635.
- Burton, R., and M. W. Feldman. 1981. Population genetics of *Tigriopus californicus*.
II. Differentiation among neighboring populations. Evolution 35:1192-1205.
- Burton, R. S. 1983. Protein polymorphisms and genetic differentiation of marine
invertebrate populations. Marine Biology Letters 4:193-206.
- Burton, R. S. 1994. Inferring the genetic structure of marine populations: a case study
comparing allozyme and DNA sequence data. CalCOFI Report 35:52-60.
- Burton, R. S. 1997. Genetic evidence for long term persistence of marine invertebrate
populations in an ephemeral environment. Evolution 51:993-998.
- Burton, R. S. 1998. Intraspecific phylogeography across the point conception
biogeographic boundary. Evolution 52:734-745.
- Burton, R. S., and M. W. Feldman. 1982. Population genetics of coastal and estuarine
invertebrates: Does larval behavior influence population structure? Pp. 709 in V.
S. Kennedy, ed. Estuarine comparisons: Proceedings of the 6th Biennial
International Estuarine Research Conference. Academic Press, New York, NY.

- 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*. *Proceedings of the National Academy of Sciences of the United States of America* 91:5197-5201.
- Burton, R. S., P. D. Rawson, and S. Edmands. 1999. Genetic architecture of physiological phenotypes: Empirical evidence for coadapted gene complexes. *American Zoologist* 39:451-462.
- Cain, M. L., B. G. Milligan, and A. E. Strand. 2000. Long-distance seed dispersal in plant populations. *American Journal of Botany* 87:1217-1227.
- Carlson, D. B., and R. R. Olson. 1993. Larval dispersal distance as an explanation for adult spatial pattern in two Caribbean reef corals. *Journal of Experimental Marine Biology and Ecology*. 173:247-263.
- Carlsson, J., and J. Nilsson. 2000. Population genetic structure of brown trout (*Salmo trutta* L.) within a northern boreal forest stream. *Hereditas* 132:172-181.
- Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 32:550-570.
- Crochet, P.-A. 1996. Can measures of gene flow help to evaluate bird dispersal? *Acta Oecologica* 17:459-474.
- Cunningham, C. W., and T. M. Collins. 1994. Developing model systems for molecular biogeography: Vicariance and interchange in marine invertebrates. Pp. 405-433 in B. Schierwater, B. Streit, G. P. Wagner and R. DeSalle, eds. *Molecular*

ecology and evolution: Approaches and applications. Birkhaeuser Verlag, Basel, Switzerland.

- Cunningham, C. W., and T. M. Collins. 1998. Beyond area relationships: Extinction and recolonization in molecular marine biogeography. Pp. 297-321 in R. DeSalle and B. Schierwater, eds. Molecular approaches to ecology and evolution. Birkhaeuser Verlag, Basel, Switzerland.
- de Sousa, G. B., A. Jimenez, A. Blanco, and C. N. Gardenal. 1997. Temporal variation of allozyme frequencies in *Aedes albifasciatus* (Diptera: Culicidae) from Argentina. *Biochemical Genetics* 35:339-349.
- Edmands, S., P. E. Moberg, and R. S. Burton. 1996. Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. *Marine Biology* 126:443-450.
- Edwards, A. W. F., and L. L. Cavalli-Sforza. 1965. A method for cluster analysis. *Biometrics* 21:362--375.
- Ellstrand, N. C., H. C. Prentice, and J. F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30:539-563.
- Excoffier, L. 2000. Analysis of population subdivision. Pp. 271-308 in D. Balding, M. Bishop and C. Cannings, eds. *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd., New York, NY.

- Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Falconer, D. S., and T. F. C. MacKay. 1996. *Introduction to Quantitative Genetics*. Longman, New York, NY.
- Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Department of Genetics, University of Washington, Seattle, WA. Available at <http://evolution.genetics.washington.edu/phylip.html>
- Futuyma, D. J. 1998. *Evolutionary Biology*. Sinauer Associates, Inc., Sunderland, MA.
- Gaines, M. S., and C. J. Krebs. 1971. Genetic changes in fluctuating vole populations. *Evolution* 25:702-723.
- Galtsoff, P. S. 1930. The fecundity of the oyster. *Science* 72:97-98.
- Ganz, H. H., and R. S. Burton. 1995. Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod *Tigriopus californicus*. *Marine Biology* 123:821-827.
- Gardner, J. P. A., and N. L. Palmer. 1998. Size-dependent, spatial and temporal genetic variation at a leucine aminopeptidase (LAP) locus among blue mussel (*Mytilus galloprovincialis*) populations along a salinity gradient. *Marine Biology* 132:275-281.
- Gold, J. R., L. R. Richardson, T. L. King, and G. C. Matlock. 1993. Temporal stability of nuclear gene (allozyme) and mitochondrial DNA genotypes among red drums

- from the Gulf of Mexico. *Transactions of the American Fisheries Society* 122:659-668.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463-471.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* 92:6723-6727.
- Goodman, S. J. 1997. R-ST Calc: A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6:881-885.
- Grosberg, R. K., and C. W. Cunningham. 2001. Genetic Structure in the Sea: From Populations to Communities. Pp. 61-84 in M. D. Bertness, S. D. Gaines and M. E. Hay, eds. *Marine Community Ecology*. Sinauer Associates, Inc., Sunderland, MA.
- Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48:361-372.
- Gustine, D. L., and M. A. Sanderson. 2001. Quantifying spatial and temporal genotypic changes in white clover populations by RAPD technology. *Crop Science* 41:143-148.

- Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. Sinauer Associates, Inc., Sunderland, MA. 481 pp.
- Hedgecock, D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science* 39:550-564.
- Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? Pp. 122-134 in A. R. Beaumont, ed. *Genetics and Evolution of Aquatic Organisms*. Chapman & Hall, London, UK.
- Hellberg, M. E. 1995. Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: Equilibrium and nonequilibrium at different spatial scales. *Marine Biology* 123:573-581.
- Hellberg, M. E. 1996. Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. *Evolution* 50:1167-1175.
- Hilbish, T. J. 1996. Population genetics of marine species: The interaction of natural selection and historically differentiated populations. *Journal of Experimental Marine Biology and Ecology* 200:67-83.
- Hilbish, T. J., and R. K. Koehn. 1985. Exclusion of the role of secondary contact in an allele frequency cline in the mussel *Mytilus edulis*. *Evolution* 39:432-443.
- Hughes, S. 1998. STRand Nucleic Acid Analysis Software. Regents of the University of California, Davis. Available at <http://www.vgl.ucdavis.edu/STRand>

- Jensen, G. C. 1989. Gregarious settlement by megalopae of the porcelain crabs *Petrolisthes cinctipes* (Randall) and *P. eriomerus* Stimpson. *Journal of Experimental Marine Biology and Ecology* 131:223-231.
- Jensen, G. C. 1990. Intertidal zonation of porcelain crabs: Resource partitioning and the role of selective settlement. Ph.D. Dissertation, Zoology. University of Washington, Seattle.
- Jensen, G. C. 1991a. Competency, settling behavior, and postsettlement aggregation by porcelain crabs megalopae (Anomura: Porcellanidae). *Journal of Experimental Marine Biology and Ecology* 153:49 - 61.
- Jensen, G. C. 1991b. Gregarious settlement by porcelain crabs, *Petrolisthes* (Anomura: Porcellanidae). *Memoirs of the Queensland Museum* 31:390.
- Jensen, G. C., and D. A. Armstrong. 1991. Intertidal zonation among congeners: Factors regulating distribution of porcelain crabs *Petrolisthes* spp. (Anomura: Porcellanidae). *Marine Ecology Progress Series* 73:47 - 60.
- Johnson, M. P., and R. Black. 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* 70:157-164.
- Johnson, M. S., and R. Black. 1984a. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intretidal limpet. *Evolution* 38:1371-1383.
- Johnson, M. S., and R. Black. 1984b. The Wahlund effect and the geographical scale of variation in the intertidal limpet *Siphonaria* sp. *Marine Biology* 79:295-302.

- Johnson, M. S., and J. Wernham. 1999. Temporal variation of recruits as a basis of ephemeral genetic heterogeneity in the western rock lobster *Panulirus cygnus*. *Marine Biology* 135:133-139.
- Jollivet, D., L. R. J. Dixon, D. Desbruyeres, and D. R. Dixon. 1998. Ribosomal (rDNA) variation in a deep sea hydrothermal vent polychaete, *Alvinella pompejana*, from 13 degrees N on the East Pacific rise. *Journal of the Marine Biological Association of the United Kingdom* 78:113-130.
- Karl, S. A., and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. *Science* 256:100 - 102.
- Kijima, A., N. Taniguchi, N. Mori, and J. Hagiwara. 1987. Genetic variability and breeding structure in *Ruditapes philippinarum*. *Reports of the USA Marine Biological Institute Kochi University*:173-182.
- Knudsen, J. W. 1964. Observations of the reproductive cycles and ecology of the common *Brachyura* and crablike *Anomura* of Puget Sound, Washington. *Pacific Science* 18:3 - 33.
- Koehn, R. K., and T. J. Hilbish. 1995. The adaptive importance of genetic variation. Pp. 182-189 in M. Slatkin, ed. *Exploring Evolutionary Biology: Readings from American Scientist*. Sinauer Associates, Inc., Sunderland, Mass.
- Koehn, R. K., R. Milkman, and J. B. Mitton. 1976. Population genetics of marine pelecypods part 4: Selection, migration and genetic differentiation in the blue mussel, *Mytilus edulis*. *Evolution* 30:2-32.

- Koehn, R. K., and G. C. Williams. 1978. Genetic differentiation without isolation in the American eel, *Anguilla rostrata*. II. Temporal stability of geographic patterns. *Evolution* 32:624-637.
- Kordos, L. M., and R. S. Burton. 1993. Genetic differentiation of Texas Gulf Coast populations of the blue crab *Callinectes sapidus*. *Marine Biology* 117:227-233.
- Kusumo, H. T., and L. D. Druehl. 2000. Variability over space and time in the genetic structure of the winged kelp *Alaria marginata*. *Marine Biology* 136:397-409.
- Lacson, J. M., and D. C. Morizot. 1991. Temporal genetic variation in subpopulations of bicolor damselfish (*Stegastes partitus*) inhabiting coral reefs in the Florida Keys (USA). *Marine Biology* 110:353-358.
- Lessios, H. A. 1985. Genetic consequences of mass mortality in the Caribbean sea urchin *Diadema antillarum*. *Proceedings of the Fifth Coral Reef Congress* 4:119-126.
- Lessios, H. A., J. R. Weinberg, and V. R. Starczak. 1994. Temporal variation in populations of the marine isopod *Excirolana*: How stable are gene frequencies and morphology? *Evolution* 48:549-563.
- Lewis, O. T., C. D. Thomas, J. K. Hill, M. I. Brookes, T. P. R. Crane, Y. A. Graneau, J. L. B. Mallet, and O. C. Rose. 1997. Three ways of assessing metapopulation structure in the butterfly *Plebejus argus*. *Ecological Entomology* 22:283-293.
- Lewis, R. I., and J. P. Thorpe. 1994. Temporal stability of gene frequencies within genetically heterogeneous populations of the queen scallop *Aequipecten (Chlamys) opercularis*. *Marine Biology* 121:117-126.

- Li, G., and D. Hedgecock. 1998. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variances in reproductive success. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1025-1033.
- Luikart, G., and P. R. England. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* 14:253 - 256.
- Lynch, M., and B. Walsh. 1998. *Genetics and Analysis of Quantitative traits*. Sinauer Associates, Inc., Sunderland, MA. 980 pp.
- Marko, P. 1998. Historical allopatry and the biogeography of speciation in the Prosobranch snail genus *Nucella*. *Evolution* 52:757-774.
- McClenaghan, L. R., Jr, M. H. Smith, and M. W. Smith. 1985. Biochemical genetics of mosquitofish (*Gambusia affinis*): IV. Changes of allele frequencies through time and space. *Evolution* 39:451-460.
- McGoldrick, D. J., D. Hedgecock, L. J. English, P. Baoprasertkul, and R. D. Ward. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): Selection and null alleles. *Journal of Shellfish Research* 19:779-788.
- Michels, E., K. Cottenie, L. Neys, and L. De Meester. 2001. Zooplankton on the move: First results on the quantification of dispersal of zooplankton in a set of interconnected ponds. *Hydrobiologia* 442:117-126.
- Michener, C. D., and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130-162.

- Miller, M. P. 1997. Tools for Population Genetic Analyses (TFPGA): A Windows program for the analysis of allozyme and molecular population genetic data. Available at <http://www.public.asu.edu/~mmille8/tfpga.htm>
- Moberg, P. E., and R. S. Burton. 2000. Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. *Marine Biology* 136:773-784.
- Morgan, L. E., S. R. Wing, L. W. Botsford, C. J. Lundquist, and J. M. Diehl. 2000. Spatial variability in red sea urchin (*Strongylocentrotus franciscanus*) recruitment in northern California. *Fisheries Oceanography* 9:83-98.
- Nei, M. 1972. Genetic distance between populations. *American Naturalist* 106:283-292.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.
- Nei, M., F. Tajima, and Y. Tatenno. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 91:153-170.
- Neigel, J. E. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics* 28:105-128.
- Ouborg, N. J., Y. Piquot, and J. M. Van Groenendael. 1999. Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* 87:551-568.
- Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* 25:547-572.

- Palumbi, S. R. 1996. Macrospatial genetic structure and speciation in marine taxa with high dispersal abilities. Pp. 101-117 in J. D. Ferraris and S. R. Palumbi, eds. Molecular zoology: Advances, strategies, and protocols. Wiley-Liss, Inc., New York, NY.
- Palumbi, S. R. 1997. Molecular biogeography of the Pacific. Coral Reefs 16 (Suppl):S47-S52.
- Petit, E., and F. Mayer. 1999. Male dispersal in the noctule bat (*Nyctalus noctula*): Where are the limits? Proceedings of the Royal Society Biological Sciences Series B 266:1717-1722.
- Piertney, S. B., and G. R. Carvalho. 1995. Microgeographic genetic differentiation in the intertidal isopod *Jaera albifrons* Leach. II. Temporal variation in allele frequencies. Journal of Experimental Marine Biology and Ecology 188:277-288.
- Powers, D. A., and A. R. Place. 1978. Biochemical genetics of *Fundulus heteroclitus* (L.). I. Temporal and spatial variation in gene frequencies of *Ldh-B*, *Mdh-A*, *Gpi-B*, and *Pgm-A*. Biochemical Genetics 16:539-607.
- Raijmann, L. E. L., and S. B. J. Menken. 2000. Temporal variation in the genetic structure of host-associated populations of the small ermine moth *Yponomeuta padellus* (Lepidoptera, Yponomeutidae). Biological Journal of the Linnean Society 70:555-570.
- 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.

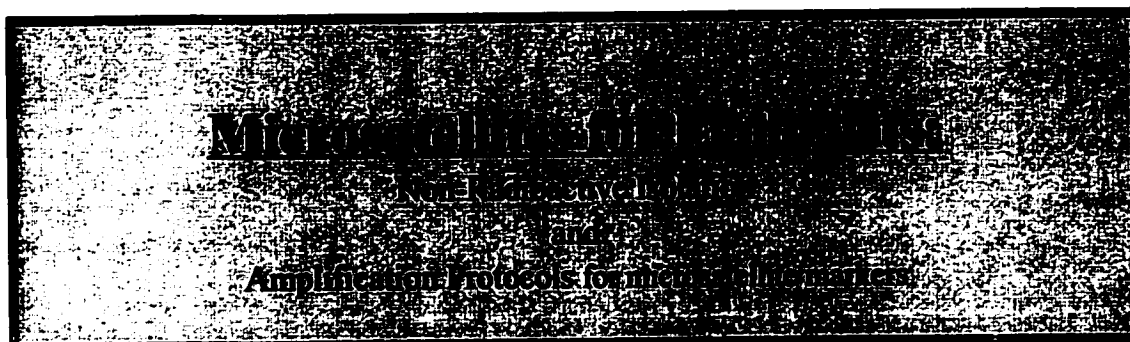
- Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 42:223-225.
- Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. Pp. 145-154 in *Studies in Genetics VII*. University of Texas, Publication No. 7213, Austin, TX.
- Rousset, F. 2000. Inferences from spatial population genetics. Pp. 239-270 in D. Balding, M. Bishop and C. Cannings, eds. *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd., New York, NY.
- Ruzzante, D. E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1-14.
- SAS Institute, Inc. 2000. JMPin. SAS Institute, Cary, NC, USA.
- Schmidt, P. S., and D. M. Rand. 2001. Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of *Mpi* genotypes. *Evolution* 55:1336-1344.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva. Available at <http://anthro.unige.ch/arlequin>
- Shulman, M. J., and E. Bermingham. 1995. Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* 49:897-910.

- Stanton, M. L., C. Galen, and J. Shore. 1997. Population structure along a steep environmental gradient: Consequences of flowering time and habitat variation in the snow buttercup, *Ranunculus adoneus*. *Evolution* 51:79-94.
- Sun, C. Q., X. K. Wang, Z. C. Li, A. Yoshimura, and N. Iwata. 2001. Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theoretical and Applied Genetics* 102:157-162.
- Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389-399.
- Tarr, C. L., S. Conant, and R. C. Fleischer. 1998. Founder events and variation at microsatellite loci in an insular passerine bird, the Laysan finch (*Telespiza cantans*). *Molecular Ecology* 7:719-731.
- Todd, C. D., J. N. Havenhand, and J. P. Thorpe. 1988. Genetic differentiation, pelagic larval transport and gene flow between local populations of the intertidal marine mollusc *Adalaria proxima* (Alder and Hancock). *Functional Ecology* 2:441-452.
- Toonen, R. J. 1997. Microsatellites for ecologists: Non-radioactive isolation and amplification protocols for microsatellite markers. Available at <http://biogeek.ucdavis.edu/Msats/index.html> (see Appendix I).
- Toonen, R. J., and S. Hughes. in press. Increased throughput for fragment analysis on an ABI 377 automated sequencer using a membrane comb and STRand software. *Biotechniques* (see Appendix II).

- Tremblay, R., B. Myrand, and J.-M. Seigny. 1998. Genetic characterization of wild and suspension-cultured blue mussels (*Mytilus edulis* Linneaus, 1758) in the Magdalen Islands (southern Gulf of St. Lawrence, Canada). *Journal of Shellfish Research* 17:1191-1202.
- Viard, F., F. Justy, and P. Jarne. 1997. Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. *Genetics* 146:973-982.
- Vos, C. C., A. G. Antonisse-De Jong, P. W. Goedhart, and M. J. M. Smulders. 2001. Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity* 86:598-608.
- Waples, R. S. 1990a. Conservation genetics of Pacific salmon: II. Effective population size and the rate of loss of genetic variability. *Journal of Heredity* 81:267-276.
- Waples, R. S. 1990b. Conservation genetics of Pacific salmon: III. Estimating effective population size. *Journal of Heredity* 81:277-289.
- Watts, R. J., M. S. Johnson, and R. Black. 1990. Effects of recruitment on genetic patchiness in the urchin *Echinometra mathaei* in Western Australia. *Marine Biology* 105:145-151.
- Weir, B. S. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Inc., Sunderland, MA. 445 pp.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38:1358-1370.

- White, M., and G. E. Svendsen. 1990. Genic homogeneity among temporal populations of eastern chipmunk (*Tamias striatus*). *Canadian Journal of Zoology* 68:2169-2173.
- White, S. E., P. K. Kennedy, and M. L. Kennedy. 1998. Temporal genetic variation in the raccoon. *Journal of Mammalogy* 79:747-754.
- Wing, S. R., L. W. Botsford, J. L. Largier, and L. E. Morgan. 1995a. Spatial structure of relaxation events and crab settlement in the northern California upwelling system. *Marine Ecology Progress Series* 128:199-211.
- Wing, S. R., L. W. Botsford, S. V. Ralston, and J. L. Largier. 1998. Meroplanktonic distribution and circulation in a coastal retention zone of the northern California upwelling system. *Limnology and Oceanography* 43:1710-1721.
- Wing, S. R., J. L. Largier, L. W. Botsford, and J. F. Quinn. 1995b. Settlement and transport of benthic invertebrates in an intermittent upwelling region. *Limnology and Oceanography* 40:316-329.
- Wright, S. 1978. *Evolution and the Genetics of Populations: Variability Within and Among Natural Populations*. University of Chicago Press, Chicago, IL.

Appendix 1.



Version 1.0, July 1997.

Rob Toonen

*University of California, Davis
Section of Evolution and Ecology,
Center for Population Biology
Davis, CA 95616*

*Ph: (916) 752-2937 Fax: 752-1449
rob@biogeek.ucdavis.edu*

adapted from:

<u>The Microsatellite Manual</u> Version 6, July 27, 1996 Travis Glenn Laboratory of Molecular Systematics - MRC 534, MSC Smithsonian Institution, Washington DC 20560	<u>Microsatellite Markers: Isolation with Non-radioactive Probes and Amplification.</u> Version of 12/1996 Arnaud Estoup and Julie Turgon Laboratoire de Génétique des Poissons INRA 78352 Jouy-en-Josas FRANCE	<u>The Estimation of Genetic Distance and Population Substructure from Microsatellite allele frequency data.</u> by Brent W. Murray (May 1996) Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1. (http://helix.biology.mcmaster.ca/brent/brent.html)	<u>Reduction of Background Problems in Nonradioactive Northern and Southern Blot Analyses Enables Higher Sensitivity than ³²P-based Hybridizations.</u> Gabi Engler-Blum, Monika Meier, Jürgen Frank & Gerhard A. Müller. <i>Analytical Biochemistry</i> 210:235-244 (1993).
Kyle Summers & Peter Marko's Protocols	Boehringer-Mannheim Genius Kits	Stratagene XL1 & XL2 Blue Kits	Brenda Johnson's Protocols

Copyright & Fair Use

This document is protected by the copyright laws of the United States of America. No portion of this document may be reproduced without acknowledgment of the author; the title page and running header constitute appropriate acknowledgment. I hereby grant individuals and organizations permission to reproduce up to 10 (ten) copies of this document (or portions of this document) per version. My permission to produce more than 10 copies of this document must be obtained, prior to such reproduction. No portion of this document may be reproduced or distributed for profit.

This protocol was developed in order to assist non-specialists in the development of "low-tech" microsatellite facilities. If you use any of the information herein, please acknowledge Rob Toonen – something to the effect of: Toonen, R.J. Microsatellites for Ecologists: Non-Radioactive Isolation and Amplification Protocols for microsatellite markers, Unpublished manuscript, available from the author or via anonymous FTP from the Microsats for Ecologists web page (<http://biogeek.ucdavis.edu/Msats/>).

Thanks, and if you have any questions or problems with the techniques/protocols listed herein, or just want to make some suggestions, please feel free to drop me a note (rob@biogeek.ucdavis.edu). Actually, let me state that a little more forcefully: if you notice any errors, omissions, problems, etc. PLEASE contact me so that I can correct them for the next version!

Rob Toonen

CONTENTS

FOREWORD AND BACKGROUND **GRAPHICAL BREAKDOWN OF PROTOCOL STRUCTURE** **PRELIMINARY TASKS**

- A. Label Oligonucleotide Probes with Digoxigenin pg 216
 B. Extract DNA for Cloning & Verify Presence of Microsatellites pg 220

MAKE A PARTIAL GENOMIC LIBRARY

- C. Digest Extracted DNA with a Restriction Enzyme pg 227
 D. Recover the $\approx$ 350-850 Bp DNA Fragments pg 230
 E. Ligate Size-selected DNA Fragments into Cloning Vector pg 233
 F. Transform Competent Cells pg 237
 G. Replate Colonies on New "Transfer" Petri Dishes pg 243

LIBRARY SCREENING

- H. Lift Colonies onto Membranes and Fix DNA to Filters pg 246
 I. Screen for Inserts with Microsatellite Sequences (Positive Clones) pg 250
 J. Colorimetric Detection and Picking and Storing Colonies pg 254
 K. Estimate Insert Sizes of Putative Positives (Optional) PCR of Cloned DNA
 with Universal Primers pg 261

SEQUENCING

- L. Miniprep: Plasmid Isolation from Picked Colonies pg 265
 M. Sequence Miniprep Products pg 268
 N. Read Sequence and Select Primers pg 269

PRIMER TESTING and PCR TUNE-UP

- O. Amplifying Microsatellites and Optimizing PCR pg 272

APPENDICES

- APPENDIX A** - Oligonucleotide Probes Used for Screening pg 283
APPENDIX B - Bacterial Media Recipes pg 285
APPENDIX C - Sequence Migration on Acrylamide Gel pg 289
APPENDIX D - Rapid Dna Extraction Protocols pg 293
APPENDIX E - Proposed Calendar pg 295
APPENDIX F - List of Necessary Supplies pg 299
APPENDIX G - Solutions and Buffers pg 302
APPENDIX H - Estimated Cost pg 306

FOREWORD AND BACKGROUND

I will start with a brief excerpt from an excellent review of the microsatellite literature done by Brent Murray at McMasters University in 1996. The included section below is only his General Introduction section, and I encourage you to check out his site (<http://helix.biology.mcmaster.ca/brent/brent.html>) for the remainder of the review (although it is already becoming somewhat dated).

Genetic analysis of natural populations has allowed biologists to ask a wide variety of questions which previously could only be answered by extensive observation of the group in question. A number of genetic markers have proven to be useful, including mitochondrial DNA, Major Histocompatibility Complex loci, allozyme loci, and Variable Number of Tandem Repeat (VNTR) markers. VNTR markers are characterized by a core sequence which consists of a number of identical repeated sequences. They can be divided into two categories based on the repeat length: a) minisatellites, 15-70 base pair (bp) repeats, which can be quite large, and microsatellites, 2-6 bp repeats, which seldom exceed 200 bp in total length. These sequences are regularly distributed throughout the genome and are found, on average, about every 10-20 Kb in Eukaryotes. At a given locus, there may be an important length polymorphism defined by the variable number of repeated motifs. They are usually flanked by conserved sequences which allow specific amplification of each microsatellite locus within a species. This conservative DNA allows the development of specific primers for PCR reactions that allow one to specifically amplify the VNTR sequence between the conserved area on either side of the repeat. The applicability of specific primers developed from the flanking sequence of a microsatellite varies among taxa, but in general a given primer will also amplify VNTR loci in congeneric species. As the taxonomic relationship among taxa becomes more distant, the utility of primers quickly decreases: with very few exceptions primers do not amplify species from different families, and even when amplification is possible, the VNTR loci are seldom variable.

Recently microsatellites have been increasingly used as the marker of choice for population genetic applications. There are some advantages to utilizing microsatellites over the other markers, which make them desirable. Microsatellite loci are found in large numbers and are relatively evenly spaced throughout the genome. For example, Edwards et al. (1991) examined the frequency of 5 microsatellite loci (tri and tetra repeats) on the X chromosome. They found, for either the tri or tetra microsatellite loci, that any given repeat was found every 300 to 500 Kb. From this they estimated that for all the 44 possible unique trimeric and tetrameric repeats there are 400,000 loci or about

1 every 10 to 20 Kb (Edwards et al. 1991). Of the class of loci examined by Edwards et al. (1991) 50% are polymorphic. Further, most of these loci are selectively neutral which makes them compatible with the assumptions of most population genetic theory. Technically microsatellites are more desirable than the larger VNTR loci because they can be analyzed via the Polymerase Chain Reaction (PCR) (although improvements to the thermal stable polymerases are enabling the amplification of larger fragments of DNA) and the alleles can be unambiguously sized on polyacrylamide gels. The lack of exact size resolution of larger VNTR loci has lead to procedures such as binning which reduce the statistical power of the analysis (Budowle et al. 1991). PCR analysis of small fragments also allows the analysis of degraded samples in which the mean fragment size of the genomic DNA has been severely reduced through environmental insult (e.g. Paabo et al. 1989). Finally, microsatellites have been found to be variable even in populations which have low levels of allozyme and mitochondrial variation (Estoup et al. 1995a, 1995b, 1996; Paetkau and Strobeck, 1994).

Microsatellites are useful for a number of analyses. They were originally utilized for genetic mapping (e.g. Weissenbach et al. 1992) and have been extensively used for linkage analyses in the association with disease susceptibility genes (e.g. Robinson et al. 1996). In addition they have proven useful in the analysis of paternity and kinship (Queller et al. 1993) and in the probability of sample identity at both the individual (Edwards et al. 1992) and population levels (Paetkau et al. 1995). In the study of entire populations microsatellites are also very useful (Bruford and Wayne 1993). Microsatellite variation has been used to study the amount of hybridization between closely related species (Gottelli et al. 1994; Roy et al. 1994). Comparison of levels of variation between species and populations have also proven useful in the assessment of overall genetic variation (Gottelli et al. 1994; Paetkau and Strobeck 1994; Taylor et al. 1994). They can be used to estimate effective population size (Allen et al., 1995) and to gain insight into the degree of population substructure including both the amount of migration between subpopulations (Allen et al. 1995; Gottelli et al. 1994;) and genetic relationships among the various subpopulations (e.g. Bowcock et al. 1994; Forbes et al. 1995; Estoup et al. 1996; Lade et al. 1996).

Presently there are a wide variety of procedures which can be utilized to analyze microsatellite data. These include various estimates of genetic distance between populations, such as average shared allele distance, D_{SA} , (Chakraborty and Jin 1993; Bowcock et al. 1994; for examples of their use with empirical data see Bowcock et al. 1994; Estoup et al. 1995a, 1995b, 1996), Chord distance, D_{CH} , (Cavalli-Sforza and Edwards 1967; for examples of their use with empirical data see Estoup et al. 1995a, 1995b, 1996), Nei's standard genetic distance, D_S , (Nei 1984; for examples of their use with empirical data see Gottelli et al. 1994; Taylor et al. 1994; Scribner et al. 1994; Paetkau et al. 1995; Lade et al. 1996), and average sum of squares of the differences in allele size, S_B , (Goldstein et al. 1995a; Slatkin 1995; for

examples of their use with empirical data see Forbes et al. 1995). In addition various estimators of F_{ST} can also be utilized, including G_{ST} (Nei 1984; for examples of their use with empirical data see Roy et al. 1994), θ_{ST} (Weir and Cockerham 1984 for examples of their use with empirical data see Roy et al. 1994), and R_{ST} (Slatkin 1995 for examples of their use with empirical data see Allen et al. 1995; Forbes et al. 1995).

Although these techniques have been applied to a variety of applications for the analysis of parentage and population genetic substructure, few people are currently discussing the problems with homoplasy common among microsatellite markers. At the Evolution meetings this year in Boulder, CO, every person who looked at the sequences of many microsatellite size homologues discovered that there was hidden variation at that locus (e.g., out of 50-some amplified individuals with an imperfect $(CA)_{21}$ repeat there were about 8 different alleles -- the inserted base pairs were different or in different locations -- for at least one trout species). Orti et al. (1997) further demonstrated that a single size-class represented as many as 21 distinct sequence variants, and suggested that absolute fragment size (as generally scored with microsatellite assays) is an unreliable indicator of historical affinities among alleles. There is currently no consensus for how to deal with such homoplasy, nor is there any consensus for how common it is. At least some authors (e.g., Bernie May) claim that homoplasy is such a substantial problem that the utility of microsatellites is limited at best. Whether or not that opinion is accurate, one must be aware of the problems of homoplasy and be prepared to deal with it in some way if your planned use of these markers is sensitive to the bias introduced by non-homology of size-based alleles. Obviously potential homoplasy of size-class fragments severely limits the utility of microsatellites for phylogenetic or phylogeographic applications.

Another significant problem (which Brent discusses in his review on the web page), is the application of an appropriate model for use in analyzing the data generated by

microsatellite assays. As mentioned above, there are a wide variety of techniques that can be applied to analyze microsatellite data (including D_{SA} , D_{CH} , D_S , D_{SW} , S_B , $(\delta\mu)^2$, F_{ST} , G_{ST} , θ_{ST} , and R_{ST}), and the application of these metrics is variously justified by different arguments about the proper mutational model to be applied to microsatellite evolution. A review of the mutational processes suspected to play a role in microsatellite evolution, and a theoretical critique of the various methods of phylogenetic analysis can be found at (<http://lotka.stanford.edu/microdist.html>). Estimates of genetic distance can be biased by small sample sizes, however, and because biases and sampling variances can be affected by the number of, and distances between alleles, microsatellites may be particularly prone to such bias effects. To the best of my knowledge, however, there has been only one study of the fit of these various metrics to empirical data (Ruzzante, CJFAS in press), and that study suggests that virtually all of these metrics are particularly biased at small sample sizes. Of the seven genetic distance and two structure metrics examined, only $(\delta\mu)^2$, D_{SW} , R_{ST} , and F_{ST} performed satisfactorily in terms of bias and variance, and both $(\delta\mu)^2$ and D_{SW} were significantly biased at small (< 50) sample sizes. The general finding of Ruzzante's work was that "sample sizes of $50 \leq N \leq 100$ individuals per site were generally necessary for precise estimation of genetic distances and this value depended on number of loci, number of alleles and range in allele size." Perhaps the most surprising result of this work was that the original formulation of Wright's F_{ST} (which is IAM-based) generally outperformed all other measures with the exception of a few specific cases where Slatkin's R_{ST} (SMM-based modification of F_{ST}) had a lower variance. These methodological difficulties are likely to soon be addressed by the numerous theoreticians working on the problem, after which the accepted methods for analysis of microsatellite data will hopefully become standardized.

Prior to using microsatellite markers, these sequence repeats and their flanking sequence must be isolated from the genome. When no microsatellite sequences are available in the literature, specifically cloning such sequences becomes an initial procedure that can not be avoided. As this procedure requires a good practical knowledge of several molecular techniques, it often prevents less experimentally oriented biologists from undertaking this task. Therefore, one of the objectives in writing out this protocol is to provide interested researchers with a reliable and relatively straightforward method for the isolation of microsatellites and the determination of basic conditions necessary to routinely use them as molecular markers. Another constraint related to the isolation of microsatellites has traditionally been the required use of radioactive molecules to screen the partial genomic library. Recently, a screening technique using non-radioactively labelled probes (Boehringer Mannheim's digoxigenin labelling technique) has been developed to circumvent this problem.

This document outlines the protocols I have modified to allow (1) the quick and reliable cloning of microsatellite markers from decapod crustaceans without the use of radioactivity and, (2) the characterization of microsatellite markers such that conditions for their routine use in other taxa can be developed. For each task, one or several tests are proposed to verify that step has been successfully completed before continuing. In order to keep the protocols accessible to those not completely familiar with molecular techniques, these protocols are somewhat overly detailed.

Although these protocols were adapted from a number of existing protocols specifically for the isolation and amplification of microsatellites in decapod crustaceans, they have also been successfully applied to at least one species of hydrozoan, bryozoan, gastropod

and insect. Similar protocols (see Estoup & Turgeon, 1996 and Glenn, 1996) have been applied to a wide variety of organisms, including:

➔ Vertebrates:

- ① Fishes: Common (brown) and rainbow trout, Salmon, European catfish, Flounder, Cod, Cisco and Tilapia
- ② Birds: Chickens, Black-throated blue warblers, Swainson's warblers, golden-collared manakins, chickadees, marsh wrens, Whooping cranes, and Spotted Owls.
- ③ Lizards
- ④ Mammals: Rabbits, Bats (2 species), Elk, Mice, and Humans.
- ⑤ Amphibians: "Poison-arrow" frogs.

➔ Invertebrates:

- ① Bees
- ② Polychaetes
- ③ Gastropods
- ④ Ants
- ⑤ Aphids
- ⑥ Dipteran flies
- ⑦ Decapod crustaceans
- ⑧ Hymenopterans
- ⑨ Bryozoans
- ⑩ Cnidarians

This is a relatively short list of the species of which I can remember off the top of my head, but I am sure that there are MANY species for which primers exist, but of which I am unaware -- I have heard rumors that Travis Glenn is going to start a web page that will include a search engine for primers and contacts at some point, but I haven't seen it yet.... As I have already said, but wish to re-emphasize, this manual is a compilation and modification of existing protocols that have been taken from several sources. I listed the primary sources from which I have borrowed in the development of this procedure, but the protocol has been mostly adapted from Travis Glenn's (1996) Microsatellite Manual and from Engler-Blum et al. (1993); the general design format of this document has been unshamefully copied from Estoup

& Turgeon (1996).

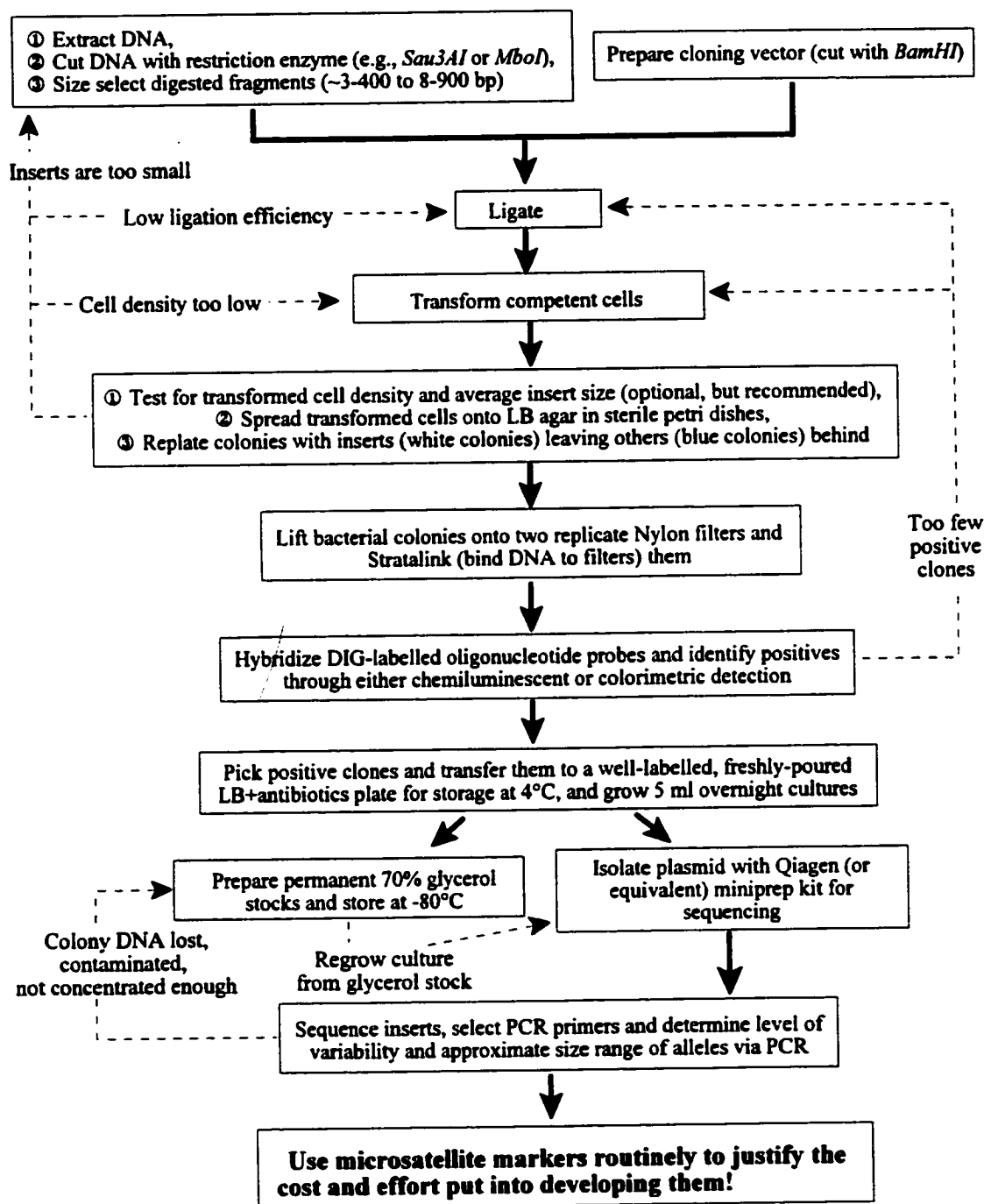
A copy of Travis Glenn's Microsatellite Manual (Version 6, 1996) is available from the Smithsonian FTP server at: <ftp://onyx.si.edu/protocols/MsatManV6.rtf> (MsatManV6.rtf -- RTF format for non-MAC users).

A compressed version of the protocol written by Estoup & Turgeon is available at: <http://www.inapg.inra.fr/dsa/microsat/microsat.htm> (microev1.zip -- Word 6 format, Windows 3.11).

This manual is available in WordPerfect 7/8 format from the Grosberg Lab server at: <ftp://biogeek.ucdavis.edu/Msats/MsatProtocolV1.wpd> or you can download it in Adobe PDF format from the web page: <http://biogeek.ucdavis.edu/Msats/index.html/>

A quick flow-sheet guide to the proposed steps in the procedure follows – for some reason this next page does not display properly in some installations of the Envoy viewer. I will continue to try to fix this, but it does not affect the utility of the remainder of the manual – it's really just for putting everything into the big picture of what you're trying to accomplish here. If you read through the manual, you could draw this yourself on the back of an envelope...

Breakdown of Protocol for the Isolation, Characterization and Amplification of Microsatellite Markers.



A. LABEL OLIGONUCLEOTIDE PROBES WITH DIGOXIGENIN

RECOMMENDATIONS

You'll need a series of custom oligonucleotides to make probes for microsatellites. Travis Glenn recommends a number of "common and useful" di- and tri-nucleotide repeats. There are smaller (mono-) and larger (4,5,6, etc..) VNTRs, but the di- and trimeric repeats are both more abundant and useful according to Glenn. By selecting the repeats that are the most "abundant and useful" from his list, I have made three sets of mixed probes which require different temperatures: a low, medium and high temperature mixed probe. The low temperature probe is composed of five DIG-labelled oligonucleotides: $(AAC)_8$, $(ATC)_8$, $(AAG)_8$, $(AAT)_8$ and $(ACT)_8$. The base pairs within parentheses is the actual repeat sequence, and the subscript is the number of times that sequence is repeated (i.e. $(AC)_8 = ACACACACACACACAC$). The medium temperature probe is composed of two DIG-labelled oligos: $(AC)_{12}$ and $(AG)_{12}$. The high temperature probe is composed of five DIG-labelled oligos: $(ACC)_8$, $(AGC)_8$, $(ACG)_8$, $(CCG)_8$ and $(AGG)_8$. The Low Temperature Probe is hybridized at 55°C, the Medium Temperature Probe is hybridized at 65°C, and the High Temperature Probe is hybridized at 75°C (These details are all explained below). I have also added a mixed probe used by Estoup & Turgeon [$(TC)_{10}$, $(TG)_{10}$, $(CAC)_5CA$, $CT(CCT)_5$, $CT(ATCT)_6$, $(TGTA)_6TG$] that they claim is hybridized at 45°C, but the mean T_m is actually ~60°C (57-62°C), so I use it at the low (55°C) temperature as well. Basically all you need to do at this stage is make sure that you order (or "borrow") at least the following oligonucleotides (or any other specific ones that you need for your purposes) in preparation for probing:

(AAC)₈, (ATC)₈, (AAG)₈, (AAT)₈, (ACT)₈, (ACC)₈, (AGC)₈,
 (ACG)₈, (CCG)₈, (AGG)₈, (CAC)₅CA, CT(CCT)₅, CT(ATCT)₆,
 (TGTA)₆TG, (TC)₁₀, (TG)₁₀, (AC)₁₂ and (AG)₁₂.

Labelling the oligonucleotide probes with digoxigenin (DIG) may be performed at any time, but takes most of a day; therefore you should consider doing it on a day when you are waiting for some other step (such as an overnight digestion with restriction enzymes or ligation reaction) early in the procedure. Labelled probes can be kept for months at -20°C (I have stored and successfully used some after nearly a year). Simultaneous hybridization of up to 6 membranes in a container (the maximal capacity of most classical hybridization tubes is usually 6 membranes, but this protocol is modified so that limit no longer applies) requires 10 pMol/ml of mixed probes (which can consist of several mixed oligonucleotides of similar melting temperature). I do not use a traditional hybridization oven, but rather an orbital shaker table in a VMR incubator which allows for more filters per container. I still try to keep the filter number low to keep hybridization efficiency high, so I typically use 15 ml of hybridization buffer/probe to hybridize up to 6 filters simultaneously; that requires a total of 150 pMol of probe, which could, for example, be 75 pMol each, if you were to mix (AC)₁₂ and (AG)₁₂ oligonucleotides in the same buffer. I will present the protocol in detail for the “medium” temperature probe that I use [(AC)₁₂+(AG)₁₂], and assume that the reader will be able to modify the protocol for labelling other probe combinations.

One other step that can also be performed early on, in tandem with other steps that necessitate prolonged incubations is the preparation of the vector. Although I have not included that step in the protocol below, it can be found at the beginning of the ligation protocol section, and I encourage the completion of vector and probe preparation during the

early stages of this procedure (*because if you're anything like me, you'll be REALLY anxious to go on once you reach the later steps, and will kick yourself for not having completed these steps during the "down time" at the outset when you reach them!*).

PROTOCOL

● **DIG Oligonucleotide 3'-End Labelling Kit (Boehringer Mannheim Genius 5 Kit* -- Cat. No. 1362 372)**

- **For each probe, mix in a total volume of 30 μ l:**
! Work on ice throughout this protocol !

Vial 1* (5x Reaction Buffer)	6 μ l
Vial 2* (CoCl₂ solution, 25 mM)	6 μ l
Vial 3* (DIG-ddUTP solution, 1 mM)	1.5 μ l
Vial 4* (Terminal Transferase, 50 units/μl)	1.5 μ l
150 pMol Oligonucleotide probes (see Appendix A)	xx μ l (total of 150 pMol)
H₂O	xx μ l (to final volume of 30 μ l)

● **Incubation**

- Combine the above ingredients into a 0.5 ml sterile snap-cap centrifuge tube.
- ↓
- Vortex the mixture, then pulse spin in bench-top centrifuge (*to recover droplets on the tube sides*)
- ↓
- Incubate at 37°C -precisely- for 15 minutes (*the efficiency of the DIG labelling depends on the incubation time and temperature*)
- ↓
- Place on ice to slow the DIG tail incorporation reaction.
- ↓
- Add **3 μ l of a STOP- Solution Mix** (stop solution is made by adding 200 μ l EDTA 0.2mM, pH 8.0 + 1 μ l glycogen (**kit Vial 8***) together and vortexing) to definitely stop the DIG labelling reaction. (**Caution: DO NOT** use phenol or PCI to stop the reaction, because this treatment leads to partitioning of the DIG-labelled oligonucleotides into the organic phase!)

● **Precipitation**

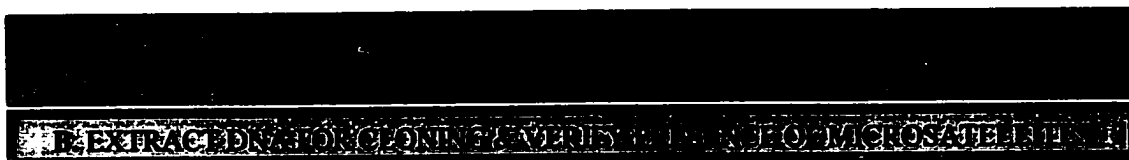
- Add 3.75 μ l of 4M LiCl + 125 μ l of ice-cold 100% Ethanol

- ↓
- **Mix well** by inverting several times or vortexing GENTLY.
 - ↓
 - **Precipitate labelled oligonucleotides:** Leave for at least 20 minutes at -80°C or 2 hours at -20°C (if you are in a REAL rush, you can try precipitating for 3 minutes in liquid nitrogen).
 - ↓
 - **Spin** at 12 000 g (~12,000 RPM) for 25 minutes at 4°C (*in principle, no pellet should be visible*)
 - ↓
 - **Slowly pour off the ethanol**, and dab tube on clean Kimwipe to remove as much EtOH as possible.
 - ↓
 - **Wash pellet** by adding 100 μl of ice-cold 70% ethanol (*Agitate GENTLY by inverting for ~1-2 min*).
 - ↓
 - ! DO NOT vortex or otherwise attempt to dislodge the pellet or your sample will likely be lost !**
 - ↓
 - **Spin** at 12,000 RPM again for 20 minutes at 4°C
 - ↓
 - **Discard the supernatant**, then set the tube upside down on a Kimwipe© to drain as much EtOH as possible.
 - ↓
 - **Vacuum dry** (dessicator or speedvac, if available) to eliminate all traces of ethanol

● Resuspension and storage

- **Resuspend** the labelled probe in an appropriate (small, 100-300 μl) volume of UV-treated double-distilled water for 30 to 60 minutes, mixing once in a while (*if you chose to vortex the sample to mix it, do not do so for at least 2 minutes after first adding water*). **Store at -20°C** if probe is not used immediately. Probe labelling efficiency should **always** be checked against the provided labelled control-oligonucleotide (**kit Vial 6***) before use in hybridizations, because labelling efficiency can vary. (*Using a poorly labelled probe generally leads to missing many positive clones wasting both your time and reagents -- if your labelling efficiency is low, it is best to discard that probe and try again*).





RECOMMENDATIONS

Obviously before you can verify the existence of microsatellites in the genome of your critter of interest, you need to extract it. Cloning requires a large quantity (a minimum of 2 μg – preferably 4 or 5) of good quality DNA (clean and not overly degraded). It is therefore necessary to perform several phenol washes and to use a relatively large amount of well-preserved tissues (fresh, -80°C frozen, or recently preserved in 100% EtOH). In order to facilitate the extraction, do not put too much tissue in a tube: it is better to use several tubes per species with a limited amount of tissue per tube (preferably, but not necessarily, all from a single heterogametic individual). After completing the extraction, simply resuspend the DNA in a single tube of the series, and then transfer that volume from tube to tube in order to concentrate the DNA from all extraction tubes in the series into a single sample. You should be aiming for a single clean extraction which is resuspended in at least 10-15 μl at a concentration of approximately 500 ng/ μl (or more) to develop a partial genomic library. Once you have that DNA extraction, you can test to see whether or not it is worthwhile to continue your efforts by verifying the existence of microsatellites in the species of interest. I will first describe the technique I have adapted for DNA extraction in problematic decapod crustaceans (explained below), and then explain how to test for the presence of microsatellites in the genomic extraction.

Among decapods, I have found a wide range of ease of extraction for samples. For some species, such as *Cancer antennarius* and *C. magister*, a simple short-term (2 hr) CTAB or

one of the various rapid extraction techniques outlined in the appendices work just fine. For other species, in particular *Petrolisthes cinctipes*, however a modified 2X CTAB, 0.02% SDS extraction is required to run over-night in order to get large amounts of clean DNA. I am not sure why, but *P. cinctipes* is particularly problematic, and even when reasonable amounts of DNA are obtained from short-term extractions, PCR reactions with the extracted samples fail. Given that range of success with samples, I have adopted the more stringent extraction protocol for all my work, just to maintain consistency of sample treatment within the study. I would recommend that you try a simpler protocol with your species first before resorting to my technique, but others have found that this extraction protocol works very well for other problematic species (e.g., certain bryozoans and molluscs). I will include only the longer protocol that I have resorted to for extraction of problematic species, with the assumption that standard DNA extraction protocols are readily available and can be found elsewhere if desired.

PROTOCOL

● Tissue preparation:

If possible, tissue should be extracted from live or flash frozen (liquid nitrogen or dry ice) tissue samples, but if not possible, samples preserved for short periods of time in 70-100% ethanol (EtOH) are acceptable. Alternatively, Nelson et al. (1997) have proposed immediate digestion of tissue samples (digestion buffer with DTT, SDS and Prot.K) in the field to avoid preservation problems and shown that yields from this technique can actually exceed those of cryopreserved samples (J.S. Nelson, C.W. Khiong, C.L. Ming & V.P.E. Phang. *Biotechniques* 23:224-226).

➤ On ice, remove a tissue sample about the size of a rice grain.
Add sample to a **2.0 ml sterile snap-cap centrifuge tube** containing:
750 µl of modified 2X CTAB lysis buffer* + 15 µl of Proteinase K (10 mg/ml)
briefly macerate the sample with a sterile steel pin



➤ **Incubate in a water bath with periodic (~ every 30 minutes) mixing (or light**

vortexing)
for 2 hours at 55-65°C



- Add another 10 µl of Proteinase K (10 mg/ml) and incubate on a rocker-table (e.g., Nutator®) overnight at 55-65°C

*** Modified 2X CTAB Lysis Buffer:**

- ① Combine (but do not stir or shake):
- 2 g CTAB
 - 2 g SDS (this may need to be added as liquid 20% SDS if it will not go into solution)
 - 10 ml 1M Tris, pH 8.0
 - 35 ml 4M NaCl
 - 4 ml 0.5M EDTA, pH 8.0
 - Glass-distilled H₂O to 100 ml Final Volume (F.V.)
- ② Autoclave and allow to cool (the detergents will go into solution when heated)
- ③ Add 0.2 ml Filter Sterilized (F.S.) β-mercaptoethanol.

② Phenol wash

- In a fume hood, add 1 V (750 µl in this case) of PCT⁺ (25:24:1) and gently invert the tubes several times.

“Burp” each tube (pop the caps open, then replace them) BEFORE centrifuging⁺ (*venting the tubes*

releases the pressure and prevents tubes from leaking and spraying your sample - and hazardous chemicals - all over when heated during centrifugation).



- Spin at 10,000 RPM⁺ for 10 minutes

⁺In our lab we use the Eppendorf bench-top micro-centrifuge. I will assume that other micro-centrifuges have approximately similar RPM x g ratings. If the equipment to which you have access is not the same as we use, please consult the instructions provided with the various tubes and equipment listed herein for the appropriate g-ratings for centrifugation rather than blindly use my RPM values.



- CAREFULLY collect the top (aqueous) layer without disturbing the interface

! Leave the interface - it is better to leave some sample behind than to transfer some of the flocculant impurities at the interface to the next step!



- Repeat Step 2 if, and as long as, any impurities are present (*any flocculent matter at the interface*)

±(25:24:1) PCI — Phenol:Chloroform:Isoamyl Alcohol

① Combine:

- 25 ml **Equilibrated Phenol** (*Phenol can be purchased equilibrated, or the protocol for equilibrating phenol can be located in Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.*)
- 24 ml Chloroform
- 1 ml Isoamyl Alcohol

② Always use fresh HPLC/molecular biology-grade reagents to minimize risk of sample contamination

● Phenol removal — Chloroform wash

➤ In a fume hood, add 1 V (750 µl in this case) of CI (24:1) and gently invert the tubes several times.

(“Burping” tubes is not necessary with CI -- CI is made as PCI above, but without the phenol)



➤ Spin at 12,000 RPM for 12 minutes



➤ CAREFULLY collect the supernatant (aqueous top layer) without disturbing the interface (at this point there should be no flocculent material visible at the interface)

● Precipitation

➤ Add 2 to 2.5 V (simply fill the tube) of ice-cold 95% EtOH

Mix gently by inverting several times, until solution appears homogeneous (if persistent low yield is found, addition of 10% v/v 3M Sodium Acetate (NaOAc) will increase precipitation efficiency)



➤ Store at -80°C for 20 minutes, or -20°C overnight to precipitate DNA



➤ Spin at 12,000 RPM for 15-20 minutes



➤ CAREFULLY Pour off the EtOH, and wash twice with 750-1000 µl of ice-cold 70% EtOH
! Do NOT vortex: the pellet could be detached and lost!



➤ For each wash, spin at 5,000 RPM for 5 minutes at 4°C.



➤ After the second wash, remove as much EtOH as possible, then vacuum or air dry the pellet. (Use a Kimwipe to draw off the EtOH remaining on the tube sides and, if possible, speedvac to dry)



- Resuspend the pellet in 35-50 μ l of UV-treated glass H_2O (or 1x TE for long-term storage)
(Let the DNA solubilize at room temperature for 30-60 minutes.)
(To speed the process, warm to 37°C and mix by vortexing every 5 minutes)

● Tests on extracted DNA (Optional)

- Load 5 to 10 μ l of extracted DNA solution (200 - 400 ng of DNA) + 5 μ l loading buffer[♦] on a 1% Synergel, 0.6% agarose minigel (see Section O). Electrophorese 40 minutes at 140 Volts.



➤ On the same gel, load:

- A 100 bp ladder marker to check if there is any 350-850 bp degraded DNA (fragments to be cloned).
- 400 ng of a maker such as MIII (Boehringer) in order to roughly estimate the amount of loaded DNA. This amount is estimated by appraising the intensity of the fluorescence, knowing that the top band of the MIII pattern approximately contains half of the total loaded amount.



- Also prepare a control to test the capacity of *Sau3AI* to cut the extracted DNA.
- Use 200-400 ng of extracted DNA + 1.5 μ l 10 X digestion buffer + 1.5 μ l *Sau3AI* (10 u/ μ l) + d H_2O up to 15 μ l; incubate at 37°C for 2 hours. Also incubate the two following controls under the same conditions: 1) 400 ng of DNA + 1.5 μ l 10X digestion buffer + d H_2O up to 15 μ l; 2) 400 ng of native DNA + 8 μ l d H_2O . Electrophorese a 10 μ l aliquote + 5 μ l loading blue of each solution on 1.5% agarose minigel (10x15cm) for ca. 40 minutes (see Section O). Run a MIII (Boehringer) and a 100 bp ladder marker alongside.

♦ Standard gel loading buffer for DNA samples:

0.25% Bromophenol Blue, 0.25% Xylene cyanol, 25% Ficoll in glass d H_2O

♦ Light gel loading buffer for DNA samples:

Same as above diluted 1:1 with UV-treated glycerol.

Once you have a reliable and clean genomic extraction, you will probably want to verify the presence of microsatellites in that DNA before spending several weeks and thousands of dollars trying to isolate SSR markers. This is especially true for those species for which no microsatellites are known, and for which there is little or no data on the presence and abundance of repeats in the genome (most marine invertebrates, for example). Unfortunately, this is only really possible if you have a LOT of very clean DNA for which an exact

concentration has been determined (make sure to eliminate all RNA by a short RNase treatment).

There are a variety of potential protocols to check for microsatellite presence and approximate “density.” Regardless of the method, all these protocols require a LOT of DNA (several **micrograms**) and are therefore often impossible to for species having low extractable quantities of DNA. Also, this is really only worth doing if you have DNA from at least one species for which the frequency of microsatellites in the genome is known. If several such species are available, it is best to choose the ones that give the widest possible range of frequencies against which to compare your samples, because a dot-blot will really only give you a relative impression of the control to which your sample is closest in microsatellite abundance. Fortunately, many different repeat motifs may be used to probe genomic samples simultaneously or independently.

PROTOCOL

● DNA Dot-Blot Preparation

For each species (the species of interest and controls), pipette three 5 µl spots of total genomic DNA (at a concentration of 100ng/µl, 500ng/µl and 1000ng/µl of DNA, corresponding to 0.5, 2.5 and 5 µg total DNA) on a positively charged Nylon membrane (see Section H for a brief discussion of appropriate membranes). Label the spots well using an HB pencil (an ink pen or even permanent sharpie felt marker will run in some of the solutions, so use a pencil), and if you are trying to cram many species onto a single filter, you may want to reduce diffusion by adding only 1µl of DNA suspension at a time, and allowing it to dry before subsequent additions (Estoup & Turgeon suggest drying the additions with a hairdryer after each deposit to accomplish the same task). Add a blank (the dH₂O used to resuspend your extracted DNA). Remember to always wear latex gloves when working on your samples, and to manipulate the membranes with forceps, touching only the edges of the membrane away from where the samples are located.



- Fix the DNA onto the membrane by either baking it in an 80°C incubator for 1 hour or Strata-linking it to the membrane (see Section H).



- You will basically follow the protocol outlined in Section H (lifting colonies onto

membranes), but because you are transferring genomic DNA rather than bacteria, you need not worry about washing the membranes as required in Section H (although it doesn't hurt to include it). Simply saturate several filter paper sheets with Solution 3 (see Section H or Appendix G) and incubate the membrane on these sheets for ~10 minutes.



➤ Dab excess fluid off the membrane as described in Section H, and transfer it onto several paper sheets saturated with Solution 4 (see Section H or Appendix G). Let them sit for 10 minutes.



➤ Dab the membrane again, then transfer it onto several filter paper sheets saturated with Solution 1 (see Section H). Let them sit for 5 minutes.



➤ Fix the DNA onto the membrane by putting it in an 80°C incubator for 1 hour or Stratalinking the DNA to the filter (again, see Section H for details). The membranes are then ready for hybridization (outlined in Section I, screening for positive clones, below). Once the DNA is fixed to the membranes in this way, they can be stored in the dark at room temperature (or better yet, at -20°C) for a long time.

Estoup & Turgeon also include a personal communication from Phillippe Jarne for a similar technique that uses far less DNA than the one that they and I use. I have copied it from the Estoup & Turgeon guide below:

Dot-blot DNA preparation protocol -- (Philippe Jarne, pers. comm.).

1. Prepare series of 5 µl DNA dilutions for the studied species as well as for the positive controls (a probe dilution, or DNA containing microsatellite of a similar motif). For example, prepare series of 5 µl dilutions containing 1 µg, 500 ng, 250 ng, etc..... Also prepare a 5 µl blank (water).
2. Mix each sample with 5 µl of [NaOH 0.5 M / NaCl 1.5 M].
3. Incubate 10 minutes at 68°C.
4. Deposit 3-4 µl of these samples on a piece of nylon membrane (Hybond-N+) marked with a 1 cm² grid. These membranes fix single-stranded DNA; thus making a further hybridization with another single-stranded DNA- the probe- possible. Always manipulate the membranes with forceps or with gloves, touching only the edges.
5. Let dry for 10-15 minutes.
6. Quickly wash the membranes twice (1 minute for each wash) in 2X SSPE.
7. Let dry for 30 minutes at room temperature. The membranes are ready for hybridization.

After binding the genomic DNA to a filter as outlined by one of these techniques, you are ready to attempt to hybridize a series of oligonucleotide probes to the extracts to see how

they look relative to species with known abundances of microsatellites (if there is no other quantifiable sources available, this could include human, mice or salmon, all of which have been studied extensively). Select a series of likely SSR repeat motifs and use the hybridization and detection protocols outlined below (see Sections I & J). The relative intensity of staining in your species of interest relative to that of species with known microsatellite abundances will give you some indication of how fruitful your isolation attempts are likely to be.

C. DIGEST EXTRACTED DNA WITH SITES FOR RESTRICTION ENZYMES

RECOMMENDATIONS

Digestion must be total, which implies that:

- The amount of DNA per unit volume must not be too high (too much DNA in a tube may reduce the digestion efficiency; it is therefore better to perform the digestion in several different tubes than overload a single tube). I generally try 2-5 µg of DNA for the first run, and increase or decrease that amount as necessary.
- A lot of enzyme is necessary (although a maximum amount corresponding to 1/10 of the total reaction volume must not be exceeded, for then the digestion is inhibited by the glycerol). I aim for 40 units to digest up to 5 µg of DNA. Despite the fact that this is about 10 times the required amount, I have found that it is better to use an excess of the enzyme to ensure complete digestion rather than be stingy and have the size-selection fail
- The total digestion reaction volume must not be too high (typically 30 to 50 µl) in order to ensure good heat conductivity.
- Different species and even samples appear to be digested at different efficiencies by these random cutters - you may need to try others to get a good digestion.

Therefore, if you are trying a new species or have not attempted this procedure before, it is best to try two or more restriction enzyme on the sample (as DNA concentration allows) simultaneously. I always use both *Sau3AI* and *MboI* on any new sample to be digested, then select the best digestion to use for recovery of size-selected DNA. *Sau3AI* is a particularly good choice for organisms that do not methylate their DNA, and for critters that do, *MboI* is a particularly good choice. However, many other cutters may be appropriate (any blunt cutter that produces a lot of fragments in the appropriate size range should be fine), and it may be worth your time to find out how enzymes that are already available in your lab, or are specific to your organism, work before going to purchase others. In general 6-cutters (*EcoRI*, *EcoRV*, *XhoI*, *XbaI*, etc.) tend to give fragments in the 5-6000 bp range, which is too large for automated sequencing, though, so try to stick to 4 base cutters. If you are interested in studying the frequency or distribution of microsatellites in the genome of your organism, then you must make several different COMPLETE libraries (i.e., cut with a number of different restriction enzymes and not size selected) and screen them all, because several people have found biases in repeat position in relation to cutters sites (e.g., Robert Butcher reports finding (GT)_n repeats disproportionately biased to be just 3' of *Sau3AI* sites). If this is the case, you would not get the 5' flanking sequence of that repeat for your primers. If you're simply interested in finding some primers to assess levels of polymorphism, don't worry about those sort of issues....

PROTOCOL

● Enzymatic Restriction Digestion

➡ Prepare a mix of 50 µl (DNA concentration permitting – if low DNA concentration is likely to be a problem, cut this recipe back to 35-40 µl, retaining the *proportion* of the 10X digestion buffer) in a 0.5 ml snap-cap centrifuge tube with the following ingredients:

- 5 µl *Sau3AI* 10X restriction enzyme digestion buffer
- 4 µl *Sau3AI* (@ 10 u/µl) This is obviously overkill on the enzyme,

but it ensures complete digestion.

(1 unit of restriction enzyme is defined as the amount required to completely digest 1 µg of DNA in 1 hour at 37°C)

- 2-5 µg (depending on DNA availability from earlier extractions) of genomic DNA
- XX µl UV-treated glass dH₂O (enough to make the total volume = 50 µl)



➤ Vortex briefly, then pulse spin (to recover droplets on the tube sides)



➤ Incubate at 37°C for 5-12 hours



➤ Store at -20°C if not used immediately.

– IF DIGESTION IS CHRONICALLY INCOMPLETE –

- A single mix can be prepared, and then ~5-10 µl is dispatched in each of several tubes
 - Vortex + quick spin
 - Incubate at 37°C overnight
- The (usually four) digestion tubes can then be pooled
- Store at -20°C if not used immediately.

● Negative Control (OPTIONAL – Only if you have enough DNA or are having problems):

➤ Combine the following ingredients in a 0.5 ml tube:

- 20 µl DNA
- 2.5 µl *Sau3AI* 10X digestion buffer (NOTE: do NOT add any *Sau3AI* enzyme to negative control)
- 2.5 µl H₂O



➤ Incubate as specified for previously mentioned digestion.

● Test (again OPTIONAL, but recommended for new species)

- Once the digestion is over, load an aliquot of ~7 µl from the digestion on a 1% Synergel, 1% agarose gel (which provides sufficient resolution of 350 to 850 bp fragments). Electrophorese for ~40-60 minutes at 100-140 V – see Section O for details. Add 2 µl of loading buffer with the sample



On the same gel, run alongside:

- 300 ng of 123 bp ladder marker in order to verify that the DNA smear around 350-850 bp is sufficiently dense (these are the fragments to be recovered from the gel and cloned) as well as

to verify that there are no discrete DNA bands in this zone. The presence of such band(s) (*repeated DNA*) could create problems during the cloning procedure, because they may end up representing a substantial proportion of the partial genomic library. If one or several discrete bands are noted in the 350-850 bp zone, it may be necessary to redo the digestion – which is simpler than the alternative (cleaning up the sample with the aid of DEAE membranes).

- 300 ng of a marker such as MIII (Boehringer) (*top band roughly at a position corresponding to the non-digested DNA*)

- 400 ng of the negative control DNA described above. (*The proportion of DNA digested by Sau3AI relative to that degraded by DNase or by physical breakdown during the incubation is best estimated with this control*).

Once the migration is over, stain the gel with ethidium bromide (EtBr) and take a photograph of the banding pattern under UV light for your permanent records.

D. RECOVER THE ~350-850bp DNA FRAGMENTS

RECOMMENDATIONS

The optimal insert size is around 600 bp. Such fragments are short enough to be sequenced, and long enough to:

- 1) Screen a sufficient proportion of the genome to isolate microsatellites with a moderately low number of clones.
- 2) Reduce the probability that the repeated sequence be too close to the cloning site (*which would impede the selection of primers for the microsatellite sequence amplification*).

The 3rd band on the Gibco 123 bp ladder represents 369 bp, and the 7th represents 861 bp, or better yet, use the 400 (4th band) and 900 (9th band) on the Gibco 100bp ladder — I usually cut as closely as possible to the top of each of these bands. I also use the 517 and 1018 bp bands of the 1kb ladder to judge the homogeneity of migration among wells on the gel -- running all 3 ladders spread across the width of the gel gives you a better size estimate. Experience shows that the techniques used for selectively recovering fragments, as well as for cloning and screening, often result in the over-representation of small fragments. A significant proportion of small fragments do not migrate at the expected position on the gel (they “move too far”). Moreover, small

fragments are more easily inserted in the cloning vector, so getting too many small fragments may prove problematic. For example, the expected mean fragment size is of 600 bp when 350-850 bp fragments are selectively recovered. In practice, however, the effective mean fragment size will rather turn out to be around 500 bp (sometimes even less). Thus it is important to ensure that the gel is run **slowly and steadily** (*don't cheat and turn this one up or down to fit your schedule, and make sure to use a reliable and constant power supply*) and that size-selection of fragments is accurate (i.e., use multiple ladders to estimate size and be careful with your cuts). For a reasons discussed in Section O, below, you should also use TAE rather than TBE buffer to run this gel. If small fragment length proves a chronic problem, move your cut up the gel (perhaps to the 517 - 1018 bands on the 1kb ladder).

PROTOCOL

● Separate appropriately-sized fragments by electrophoresis:

● Prepare one or several maxi-well(s) by either taping several combs together, or purchasing (making) a special thick comb to enable loading of about 40+ μ l in a single well (experience suggests that you can't load more than 100 μ l into a well and expect this to work very well). If making your own, it may be desirable to make a single maxi-well with a blank and then a mini-well on either side for loading prepared markers, or simply a very large continuous well to separate enormous quantities of DNA depending on the application for which you intend to use this. Prepare a fresh 1% agarose (**DO NOT** use any synergel -- if your separation is poor, try a 1.7% agarose gel instead) gel using TAE (although TBE generally gives better resolution of small fragments, TAE is considered better for in-gel manipulations and band recovery procedures -- see Section O for further details). Load as much digested DNA (hopefully the entire 40-50 μ l) + 5-7 μ l of **light loading buffer** (*dilute loading buffer 1:1 with UV-treated glycerol*) as possible until the well begins to overflow. Electrophorese the sample at 30-40V (about 25 mA) for a minimum of 1 hour (*until bromophenol blue dye -- the darker of the two -- has migrated about 3-4 cm*) for reasonably good size-separation of the 350-850 bp fragments (**DO NOT** allow the sample to separate too much -- the 350-850 bp region should comprise less than 1 cm of gel length). Run about 5-10 μ l of each the 123 bp and 1kb ladder marker alongside the sample, leaving one empty well between the sample and the markers, and AT LEAST one empty lane between multiple samples (if more than one sample is to be run on a single gel).



● During the electrophoresis, prepare fresh EtBr stain (0.5 mg/L) and destain (dH₂O). These

should be made fresh and only be used for a single gel in order to minimize contamination of your partial genomic library with ladder or xenic DNA.

! Exposing DNA to UV light causes a physical breakdown of the DNA which then becomes non-clonable. In all cases, keep the exposure time to an ABSOLUTE MINIMUM. While recovering the DNA fragments, avoid using a strong or poorly adjusted UV lamp. If possible, use a UV lamp with a wavelength higher than 300 nm. If not, place Saranwrap, folded several layers thick, under the plate to cut down on the UV transmission to the gel — VERY thin plexiglas also works but is so efficient that it makes visualization of bands difficult.

➤ When the migration is over, use a sterile coverslip or scalpel to quickly cut the gel at the top and bottom of the desired range (i.e., along the 369 and 861 bp — the 3rd and 7th band indicated by the 123bp ladder marker). The sides of the lanes can then be cut by simply eyeballing the lane from the edge of the wells. If the gel did not run straight, you can use the UV light very sporadically to visualise the desired zones ($\lambda > 300$ nm if possible).

➤ Using either a flamed small spatula⁺ or sterile coverslip or scalpel blade, complete excision of the desired fragments, and place the gel piece in either a piece of Saran Wrap, or directly into a 1.5 ml centrifuge tube. Freeze the gel piece at -20°C for at least 60 minutes, or at -80°C for at least 20 minutes (*if you're in a real hurry, you can also freeze it for about 5 minutes in liquid nitrogen*).

⁺Sterilizing instruments for DNA manipulation:

- ① Wash spatula or loop with 7X tissue soap, then rinse well with dH₂O
- ② Immerse instrument in 100% EtOH
- ③ Flame (hold over lit bunsen burner) instrument until EtOH burns off
- ④ Allow instrument to cool prior to use — **DO NOT** allow it to touch any surface once flamed, or start over

● Re-extract size-selected DNA from the agarose gel:

➤ Once the gel is completely frozen solid, transfer it to a centrifuge tube (if you froze it in Saran Wrap), and rapid thaw it for 5 minutes at 37°C (preferably in a water bath, otherwise hold the tube tightly in your hand and use body heat to thaw it for 5 minutes).

➤ Using either a flamed small spatula⁺ or sterile coverslip or scalpel blade, completely macerate the gel. Once the gel is homogenized into small pieces, transfer the slurry into a red-dot Millepore UltraFree MC filter and spin the filter for 20 minutes at 7,500 RPM. This treatment should remove all the buffer (which contains your DNA) from the agarose gel.

➤ Resuspend the dehydrated agarose in 200 μ l of UV-treated 1x TE buffer (*this can be accomplished with either a flamed spatula or simply the tip of the sterile pipette that you used to transfer the TE into the filter*) and allow the gel to rehydrate and elute for ~30 minutes at

room temperature. Centrifuge at 7,500 RPM for 20 minutes again.



➤ Discard the filter insert (complete with dehydrated gel), and measure the filtrate*. Add 10% of the filtrate volume of 3M Sodium Acetate (NaOAc). Mix well, then double the total volume (i.e., add 1 V of the filtrate + NaOAc) with cold ($\sim 4^{\circ}\text{C}$) 100% isopropanol. Precipitate at room temperature for at least an hour (*this is one precipitation that you don't want to cool down, or you're likely to get impurities*).

***Measuring the volume of filtrate or other unknown sample:**

Using a pipettor with a volume large enough to collect the entire sample, dial the pipette volume to some value larger than the amount of sample in the tube. Carefully aspirate the entire sample volume, making sure that you capture no air bubbles or allow any sample to run back down into the tip once collected. Carefully hold the pipettor and slowly dial the volume back down, watching the liquid move down the pipette tip as you do so, until the sample comes to the end of the tip. Read the volume at which the sample reaches the tip, and that is the total volume of your sample.



➤ Spin down the precipitate for 20 minutes at 14,000 RPM. CAREFULLY decant the liquid and add 500 μl of ice-cold 95% EtOH to wash. Gently invert the vial to wash the sample. **! DO NOT vortex or otherwise attempt to dislodge the pellet or you are likely to lose your sample !**



➤ Spin again at 14,000 RPM for 12 minutes. Carefully dump the EtOH and dab the excess off with a Kimwipe before drying the 'pellet' (*no pellet should be visible to your eye, but there is likely still DNA there*) in a desiccator.

● Resuspend partial genomic library and quantify

➤ Resuspend the sample in 25 μl of UV-treated dH_2O for about 30 minutes with periodic mixing. Quantify on the fluorometer. If you have less than about 5 $\text{ng}/\mu\text{l}$ of DNA in your size-selected library, you might need to start over and consider some of the suggestions for increasing yield along the way (some protocols recommend using 100-150 $\text{ng}/\mu\text{l}$, but I have never used more than 25 $\text{ng}/\mu\text{l}$ – even if you get virtually nothing from your gel-extraction it's always worth trying at least one ligation reaction before starting over, because I've gotten some good ligations from samples at concentrations as low as 1-3 $\text{ng}/\mu\text{l}$ before).

● Test (OPTIONAL, but recommended for beginners)

Load 2 μl of the final 25 μl DNA solution (combined with 1 μl of loading buffer) on a 1% Synergel, 0.6% agarose gel. Electrophorese for 30 minutes at 140 V.

On the same gel, run alongside:

• 3-5 μl of each 123bp and 1kb ladder markers.

Stain the gel with EtBr and take a photograph for your permanent records. Confirm the size of recovered fragments is between 350 and 850 bp. Yes, the recovered DNA is supposed to be a smear (worry if it is not!).

E. LIGATION

RECOMMENDATIONS

There are a series of ligations and controls that ought to be done each time a new vector preparation or a new tube of ligase (or a really old one that hasn't been used in quite a while) is used. This is especially important for people who are new to the technique, but is good practice for everyone. The table below gives the 4 controls (2 positive, 2 negative) that should be done in addition to the ligation "test series" the first time you try to run a new species and are unsure about the proper ratio of insert to vector. Ideally, you'd like a molar ratio of cohesive ends of insert to vector of approximately 1:1. You can easily calculate how much of each you need to achieve this goal with the simple formula:

$$\frac{\text{length of vector (~ 3000 bp for Bluescript)}}{\text{length of insert DNA (~ 500 bp here)}} = \text{Ratio of vector:insert for ligation calculations}$$

Thus, in this case, the vector is 6 times longer than the insert, so we would need to add 6 times the mass of vector to have the same number of cohesive ends as the insert (insert:vector ratio = 1:1 here). This is the formula that Travis Glenn recommends be used in his *Microsatellite Manual*, and is, in theory, the ideal conditions. However, in practice these condition often do not provide the highest ligation efficiencies, therefore I recommend that if you have never done this before, or are trying a new species, you should set up a series of

insert:vector ratio ligations starting around 2:1 or 4:1, and going as low as 1:6 or 1:8 (i.e., 4:1, 2:1, 1:1, 1:2, 1:4, 1:6, 1:8). This series presumes that you have a lot of DNA in your size-selected library and that wasting some on finding the ideal ligation ratio empirically will not prevent you from actually completing the procedure. If you do not have DNA (or time) to burn at this point, you might want to try an abbreviated series of ligations (3:1, 1:1, 1:3) — I have found that the ratios in the range of 1:2 - 1:4 generally work the best for me, but that may differ by species and by individual, so try the full series if you have the DNA (and time) to spare.

Cocktail Recipes:

Reagent	(Negative) Control 1	(Negative) Control 2	(Positive) Control 3	(Positive) Control 4	3:1 Ligation	1:1 Ligation	1:3 Ligation
5X Buffer	2 μ l	2 μ l	2 μ l	2 μ l	2 μ l	2 μ l	2 μ l
T4 Ligase	1 μ l	0	1 μ l	1 μ l	1 μ l	1 μ l	1 μ l
Vector (cut & dephosphorylated)	0.5 μ l	0.5 μ l	0	0	xx μ l	xx μ l	xx μ l
Insert DNA	0	0	0	0	xx μ l	xx μ l	xx μ l
Uncut Plasmid	0	0	0.34 μ l	0	0	0	0
Cut Plasmid	0	0	0	1 μ l	0	0	0
dH ₂ O	6.5 μ l	7.5 μ l	6.66 μ l	6 μ l	xx μ l	xx μ l	xx μ l
Total Reaction Volume	10 μ l	10 μ l	10 μ l	10 μ l	10 μ l	10 μ l	10 μ l

PROTOCOL

! Work on ice throughout this protocol!

● Prepare the Vector:

➤ This step may be done in tandem with several earlier steps, and should (hopefully) have

been completed before preparing to ligate the size-selected DNA.



➤➤ Prepare 10-20 µg of vector and store at -20°C until ready to use. The suggested mix to prepare the vector is as follows:

- 10 µl *Bam*HI 10x Restriction Buffer
- 5 µl 10x *BSA* (acetylated)
- 10-20 µg *pBluescript* KS+ Vector (supplied from Stratagene at 1 µg/µl)
- 10-20 µl *Bam*HI restriction enzyme (use 100 u/10 µg vector)
- 45-65 µl dH₂O (to make final volume 100 µl)

! Although it is fine to stock-pile prepared vector and store at -20°C, do not try to cut more than 20µg of vector in a 100µl digestion. If you want to prepare more than 20µg of vector at one time, use several different reactions !



➤➤ Incubate in a 37°C waterbath for ~4 hours.



➤➤ Prepare a 1% Synergel, 0.6% agarose minigel and run 3 µl of the digestion (mixed with 1 µl of loading buffer) on the gel along with 1 µl of the uncut plasmid as the "ladder" (in addition to the regular size standards: 123bp & 1kb) — run the gel at 100 V for about 30 minutes to see whether or not you can detect any uncut plasmid remaining in the digestion. If you see any uncut plasmid remaining, add some more restriction enzyme and continue the incubation for an additional hour, then repeat the gel run to see if there is any uncut *pBluescript* KS+ remaining. When you cannot see any evidence of uncut plasmid, add 5 µl of Shrimp Alkaline Phosphatase (SAP) to the reaction tube and continue to incubate at 37°C for 1 more hour. SAP is preferable to calf intestine (CIAP) because it is totally inactivated by heat shock. **DO NOT** use bacterial (BAP), because it is difficult to get rid of, and often renders the vector useless.



➤➤ After the incubation with SAP is completed, add 100 µl of 1x TE, and extract once with PCI and twice with CI in 1.5 ml centrifuge tubes (as outlined under DNA extraction protocols, above).



➤➤ Add 20 µl of 3M NaOAc and 450 µl of ice-cold 95% EtOH. Chill at -80°C for 20 minutes.



➤➤ Spin at 14,000 RPM for 20 minutes at 4°C



➤➤ CAREFULLY pour off the EtOH and dab excess off on a Kimwipe. Wash with 500 µl of 70% EtOH, and let sit for ~2 minutes (**DO NOT** mix).



➤➤ Spin at 14,000 for 5 minutes, pour off EtOH and dry pellet.



➤➤ Resuspend cut & dephosphorylated vector in 50-100 µl (50 µl / 10 µg) of UV-treated 1x TE.



- Store vector at -20°C if not used immediately.

If this "quick and easy" vector prep. protocol (modified from Travis Glenn's protocol) does not work, find Travis Glenn's Microsatellite Manual and follow the whole vector preparation protocol (steps C-M).

● Ligate partial genomic library (size-selected DNA) into the vector:

- Calculate the volume of vector, insert and dH_2O required for each ligation ratio.

E.g., for insert DNA @ $10\text{ ng}/\mu\text{l}$ and vector @ $200\text{ ng}/\mu\text{l}$ the calculations are as follows:

$$0.5\text{ }\mu\text{l of vector @ } 200\text{ ng}/\mu\text{l} = 100\text{ ng vector}$$

$$100\text{ ng vector} / 6\text{ (3000bp Bluescript / 500bp insert)} = 16.67\text{ ng insert}$$

$$1.67\text{ }\mu\text{l of insert @ } 10\text{ ng}/\mu\text{l} = 16.7\text{ ng.}$$

Therefore for the 1:1 ligation mix, we'd add $0.5\text{ }\mu\text{l}$ of vector, $1.67\text{ }\mu\text{l}$ of insert, and $4.83\text{ }\mu\text{l}$ of dH_2O (to make the final volume $10\text{ }\mu\text{l}$).

The calculation would be the same for the 1:3 or 3:1, but instead of dividing by 6, you'd divide by (6×3) or $(6/3)$ when determining the amount of insert to add.

(Alternatively, you can 'cheat' and make the vector concentration equal to 6 times the insert DNA concentration, and then just complete the ratios volumetrically)



- Once you have calculated and combined all the ingredients on ice, mix by gently vortexing, and then pulse spin to recover droplets clinging to the sides of the tube



- Incubate at $12-16^{\circ}\text{C}$ overnight (some protocols suggest room temperature for 2-4 hours, but I have never gotten great ligations from higher temperature protocols).

! Although the exact temperature at which the ligation is carried out does not really matter, the constancy of the temperature over the duration of incubation DOES. Make sure that the temperature does not fluctuate widely during the ligation reaction, or your ligation efficiency will be very low !



- After completing the overnight incubation, add 2 V of 1x TE ($20\text{ }\mu\text{l}$ here) and incubate for 20 minutes at 67°C to stop the reaction.



➤ After completing the 67°C incubation, remove a $2\text{ }\mu\text{l}$ aliquot to combine with $1\text{ }\mu\text{l}$ of loading buffer and run the samples on a 1% Synergel, 0.6% agarose gel for about 30 minutes at 100 V. Many samples can be loaded on the same mini-gel by placing 2 combs into the gel (one in the normal position, and the other about half-way down the gel length), because samples do not need to be well resolved or separated. Once the Bromophenol blue nearly reaches the wells of the other comb, stain the gel with EtBr and take a picture for your permanent records.



- Compare your ligation samples to the positive and negative controls. The negative controls should both remain linear DNA (producing a single bright band), while the positive controls

should both be circular DNA producing a characteristic banding pattern. Your samples should (hopefully -- if your ligation worked well) resemble the positive control (#4 above) most closely. You should have a band at the same position as the negative controls (the linear DNA that was not incorporated into a ligation product), as well as a smear behind that band (the vector+insert DNA that was successfully ligated and is now circular DNA). If you cannot see any smear behind the linear DNA band in your samples, your ligation did not work.



➤ Store ligations at -20°C immediately after stopping the ligation reaction.

TRANSFORMATIONS

RECOMMENDATIONS

The following protocol is almost identical to that proposed by Stratagene for their Epicuran coli® XL1-Blue super competent and XL2-Blue ultra competent cells. It depends on your ligation efficiency and the amount of ligation that you recover which cells are likely to be the best for you. If you have access to an electroporator, none of this is going to matter to you, but we never thought it was worth the money for the limited number of transformations that we do in our lab. If you have lousy ligations and can't seem to get a lot of good ligations, then the XL2 ultra competent cells are your best bet. If you have lots of good ligation, the XL2 cells are likely overkill, and you'll probably end up wasting both money and product by using them — buy the XL1 super competent cells instead. In order to take advantage of the blue/white selection, and to maximize your transformation efficiency, you'll want to buy the XL₁-Blue MRF' super/ultra competent cells (depending on your ligations, as described above). Follow the directions suggested by Stratagene for concentrations and volumes of things. These kits change from time-to-time (whether the β -mercaptoethanol is concentrated or not, etc.), and differ with cell line, so it is difficult for me to write a general set of instructions here. In general, I'd say that the Stratagene protocols are good ones, and that you should follow the instructions that come with your cells (but you can read through here for hints/tips, too). For

the folks in my lab, here's the protocol I use.

For the **entire** transformation procedure, it is absolutely necessary to **work on ice** (except for the heat shock at 42° and growth phase at 37°) or preferably within a walk-in freezer if available. It is really nice to be able to work **in a sterile environment** (laminar-flow fumehood), but if not make sure that **EVERYTHING** used in this protocol is sterile (autoclaved or purchased ready-to-use). Autoclaved pipette tips cooled in a freezer or on ice may be used (personally, I cool everything in a walk-in freezer for at least an hour before use).

Thawing and freezing competent cells results in a substantial loss of their competence (Stratagene uses some chemically induced competency, and thawing the cells allows some to recover). Accidental thawing must therefore be avoided, and any loss of competence needs to be compensated for by transforming cells with more ligation mix (or replacing the cells). The cells must be stored at -80°C at all times and have a limited life even at that temperature. If the cells are really old (a year or more), increase the transformation volumes to compensate for loss of competency.

Alternatively, if you're working on a really tight budget, you can try making your own competent cells using the TSB chemical competency (modified from C.T. Chang & R.H. Miller, 1988. Nucl. Acids Res. 16:3580) protocol outlined here — the transformation is nowhere near as efficient, but it's simple and gets the job done if you're tight for funding:

TSB Chemically Competent Cell Protocol:

- ① grow bacteria to early log phase in 20 ml of LB broth,
- ② Chill on ice for 10 minutes,
- ③ Spin in Sorval at 4000 RPM for 10 minutes at 4°C,

- ④ Decant supernatant, turn tube upside and allow to drain for 2-3 min on ice upside down,
- ⑤ remove all liquid and resuspend in 1/10 volume (2 ml) of Transformation Storage Buffer (TSB*),
- ⑥ Incubate 10 minutes on ice,
- ⑦ add 100 pg of plasmid DNA ($\sim 2/3$ μ l) to 1-200 μ l of cells,
- ⑧ keep on ice for 5-30 minutes (no heat shock is necessary, but you can try a 42°C bath for 60 seconds if transformation efficiency remains low despite good ligations),
- ⑨ add 8-900 μ l of glucose-enriched TSB (0.036g glucose/10 ml TSB), and grow at 37°C in an orbital shaker water bath for 45 minutes,
- ⑩ plate out the cells on LB agar and grow plates at 37°C overnight.

* Transformation Storage Buffer (TSB)

LB medium (pH 6.1)	47.5 ml
10% PEG 4000	5 g
5% DMSO	2.5 ml
10 mM MgCl ₂ [M.W.=203.3]	0.1016 g
10 mM MgSO ₄ [M.W.=246.47]	0.1232 g

PROTOCOL

LB agar plates must be prepared in advance for plating the transformed colonies and all your work will be wasted if the plates are not ready to use at the time the cells are ready to be plated. Make sure that you read this entire protocol through (especially the section on plating the cells) before you begin to ensure that you are prepared for each step in advance.

● Label a 15 ml Falcon #2059 polypropylene (the tube must be polypropylene to prevent chemical degradation of the tube by the β -mercaptoethanol, which leads to decreased or eliminated transformation efficiency) tube for each sample to be transformed. Follow Stratagene's recommendations for the β -mercaptoethanol (= 2-mercaptoethanol) addition – using 25 mM β -mercaptoethanol has been shown to increase transformation efficiencies two to threefold. Place the mercaptoethanol in the bottom of each tube, and place the tubes on ice. Store the ice container along with the pipettors and sterile wide-bore tips to be used during the transformation in the walk-in cold-room. Everything used during this procedure MUST be pre-chilled (on ice or in the cold-room) to avoid heat shock. Allow equipment to cool for approximately 20 minutes prior to thawing cells. Put competent cells (1-200 μ l per tube, depending on strain) on ice in the walk-in cold-room ($\sim 2-4^\circ\text{C}$), so that they thaw out S-L-O-W-L-Y.

Delicately mix the thawed cells by gently tapping on the tubes (you want to turn over the contents, but not form any drops that could be stranded on the sides of the tube – very slow

*pipetting with wide-bore tips is another way to accomplish this. Cells must be kept in an aerobic environment, therefore it is **essential** to periodically agitate the tubes — every 2 minutes or so — in such a way that permits air circulation. **However**, over-agitation of tubes leads to dramatic decreases in transformation efficiency.).*

! At this point, cells are very fragile: they must be handled very delicately. !



➤ Following Stratagene's directions, you should have already added the β -mercaptoethanol solution to each tube (I say to follow their directions because in some kits the mercapto. comes concentrated, while in others it comes ready-to-use). The important thing for the highest transformation efficiency is to ensure that the final concentration is exactly 25mM.



➤ Dispatch 100 μ l of super- or ultra-competent cells per tube (it is best to use only one tube per species of interest, especially if this is your first time through -- later you may add more samples or more cells if it proves necessary). If you find that transformation efficiencies are too high, decrease the cell volume to 50 μ l (remember to halve the mercapto also!); if you find that transformation efficiencies are too low, increase the cell volume to 200 μ l (and double the mercapto).



➤ Mix manually by **gently** tapping the tubes or by refluxing **SLOWLY** with a wide-bore pipette.



➤ Incubate on ice for 10 minutes,
Mix by GENTLY tapping the tubes or refluxing the cells through a wide-bore pipette every 2 minutes.



➤ Double-check that your water bath is at precisely 42°C for Step #3



● Add ~1-2 μ l of ligation mix (this is an approximate volume: depending on your efficiency of ligation, your transformation efficiency and your cell line, it may be necessary to add significantly more than 2 μ l and to increase or decrease the volume of cells to which you are adding the ligation mixture).

GENTLY tap or swirl the tubes to mix.

Use the pUC18 control plasmid as a positive control for the transformation. Without it, you'll have no idea whether your cell density is low because of loss of cell competence (or some other transformation problem that needs to be fixed) or the quality/quantity of ligation product used. If the pUC works and your transformation doesn't, try increasing both the cell (up to 200 μ l) and ligation (up to 10 μ l) volumes to see whether that increases your efficiency.



➤ Incubate on ice for 30 minutes, gently tapping or swirling every 10 minutes.
(I emphasize again that over-agitation will decrease efficiency of transformation **more** than does lack of agitation).



● **Heat shock:** Put the tubes in the exactly 42 °C water bath for the exact time recommended in the cell line protocol (i.e., exactly 30 sec for XL2 ultracompetent cells and 45 sec for XL1

supercompetent cells -- make sure you follow the Stratagene instructions for the transformation protocol!).

! The length of time and the temperature of the heat shock is critical to transformation efficiency !

Efficiency of transformations decreases **SHARPLY** with heat pulses **<30 or >45 seconds!** As soon as the heat shock is over, **immediately** put the tubes right back into ice for 2 minutes. (By this I mean move from the ice into the water bath, hold for exactly 45 seconds (if your cell line is the XL1 supercompetent cells, or 30 sec if you're using the XL2 ultracompetent cells) and then plunge the tubes directly back into the ice bucket. Both should be within immediate reach -- you should not even have to take a single step to move the tubes from the ice to the water bath and back -- the purpose of this activity is to shock the cells by moving them quickly from the heat to the cold).



● Add 0.9 ml of a glucose-enriched liquid bacterial medium **with NO antibiotics** (see Appendix B) (a short growth period in liquid SOC, or NZY* helps the cells to recover after the heat shock, and increases survival, and therefore transformation efficiency). **Incubate in a shaking/rocking incubator 1 hour at 37°C** (use light shaking, ~150-250 max rpm, because the cells are still fragile, and do not allow cells to grow for much longer or you are likely to get multiple copies of the same sequence when examining positives). While the cells are growing, prepare the agarose plates as described in the step below.

This culture may be stored at 4°C for a few days if necessary, but should not be stored if it is possible to plate immediately (because storage increases the likelihood that some cells will eject the phagemid containing your inserts). If storage is necessary, 1) mix the culture well (by **GENTLY** refluxing with a wide-bore pipette) prior to plating, because bacteria pellet very easily and you will lose the majority of your cells if you do not, and 2) make especially sure that your antibiotics are fresh and at the right concentration to ensure selection of only colonies with the phagemid intact.

It is a good idea to plate several different proportional volumes of cells the first time you work with any new species. For example, if transformation efficiency is generally high, you might want to try plating 10, 25 and 50 µl of the culture. If, on the other hand, you have chronically low transformation efficiency, you might want to try carefully spinning the culture down in a 2 ml centrifuge tube for 5-10 minutes at 1000 RPM, and remove ~750 µl of the broth. The cells can then be resuspended in remaining ~250 µl of broth (by refluxing with a wide-bore pipette), and 3 plates made from that broth: 50% (125 µl), 35% (87.5 µl), and 15% (37.5 µl). In this way you can cover the range of cell densities that may prove useful for re-plating cells during the next step of the protocol simultaneously, without necessitating any storage of the transformed cells.



● **Plate the cells:**

Plates should be poured and cooled at least 24 hours in advance if possible — **always** store poured plates at 4°C and discard any plates stored for more than 2 weeks. Allow plates to “dry” inverted at 37°C for several hours prior to beginning this protocol. Pre-dried plates make spreading much easier, but if the plates become too dry, reagents and media will be absorbed by the agar before it is possible to spread them evenly across the entire surface of the plate. While the cells are incubating for an hour, spread freshly prepared Xgal and IPTG on the plates

to allow blue/white selection (see Appendix B). Xgal should be prepared immediately prior to being plated, and should be kept on dry ice throughout – IPTG is more stable and the stock (25 mg/ml) can be dissolved in dH₂O and stored at -20°C. These reagents need to be plated onto the agar in exactly the same manner as the cells, using a sterile “hockey stick” and turning the dish while spreading. Continue spreading the reagent until the plate is well covered and begins to dry and become “sticky” (i.e., the spreader no longer glides wetly over the plate surface — there should not be any “line” of reagent left when you lift the spreader from the agar plate — although not essential here, this is especially important when spreading the cells!). Once both reagents are spread onto the plates, place the plates back into the 37°C incubator for the remainder of the 1 hour incubation.



Once the cells have been incubated, spin them down as described in the shaded box above, remove the supernatant medium, GENTLY reflux cell culture with a wide-bore pipette several times to homogenize the culture. Then spread ~ 100 µl of each culture on a Petri dish containing LB Agar+Ampicillin+Tetracycline. (see Appendix B). Use sterile “hockey sticks” (*long glass Pasteur pipettes shaped using a Bunsen burner flame into an “L”*) to spread the cell mixture homogeneously across the plate. It is important that there be no sharp edges on the stick, and that no downward pressure be applied to the spreader during this step. You should simply turn the petri dish continuously with one hand while briskly moving the stick back and forth over the surface of the plate with the other hand. If you’ve never done anything like this before, ask any microbiologist to show you some basic culturing techniques — it will make your life MUCH easier.



➤ Incubate the Petri dishes agar side up for ~16 hours at 37°C.



WHITE Colonies: Bacteria with insert

BLUE Colonies: Bacteria without insert

On average, you should expect about 70-80 % white colonies from a decent ligation if your transformation worked well. A good density is around 300-400 colonies (clones) per dish. White colonies must be sufficiently distant from one another so that they can be individually removed with a toothpick, and there should be no satellite colonies (white “smudges” or tiny colonies around your clones — these usually occur when 1) you let the plates grow too long, or 2) the cells are so dense that the Ampicillin surrounding them has been hydrolyzed and allowed untransformed XL1-blue colonies to begin to grow on the plates). The colony density must also be high enough to collect enough clones to fill another plate (thereby optimizing your effort).

If you find that colony density is lousy (too high or too low) for you to transfer to another plate, or that you have a lot of satellite colonies, you’ll need to fix that (and perhaps try the transformation again) before continuing. Possible solutions include: 1) if the colony density is too low, use 2,5 or even 10 µl of ligation mix and/or spread more of the transformation volume (don’t try to spread more than 200 µl or you’ll be swirling the petri dish forever. Concentrate the cells as described above and remove more of the supernatant broth); if the colony density is too high, spread less of the transformation mixture (you should be able to judge the “correct amount” from the series you did at first) and/or use half as many cells in the transformation

with the appropriate dilution of the ligation mix (the transformation efficiency is more or less proportional to these).



In order to check whether your partial library is really going to be useful, it is **strongly advised** that you do a PCR amplification test on 20-30 randomly chosen white clones in order to: 1) estimate the average insert size (see Part J), and 2) double-check that the blue-white selection really worked, and that the colonies are not white because your Xgal was old rather than because your colonies have inserts. It is preferable to determine the average insert size using 20-30 randomly chosen white colonies rather than using only positive clones (i.e., with colonies containing microsatellites, as determined by the screening procedure) because the positive clones tend to have a higher average insert size than randomly chosen inserted fragments. This is especially important if you are trying to infer the genomic microsatellite occurrence frequency (number of positive clones / size of genome that has been screened). Amplifying a few (~10) blue colonies will confirm whether or not they lack inserts.

G. REPLATE COLONIES ON GRID PETRI DISHES

The first time Brenda Johnson told me that she did this with all her plates, I told her that she must be insane, because it is SO tedious and time consuming. Well, it didn't take me long to decide that she was not nuts and that it was worth the time it took to transfer all those colonies to new plates prior to probing. It is well worth the effort for several reasons: 1) it allows you to select only clones with known inserts (white colonies from above) for probing; 2) the clones become much larger after you move them to new plates with a sterile toothpick and therefore are easier to probe (the detection is more sensitive because there is more DNA with which to work); 3) you can move the bacteria to newly poured plates with new antibiotics to minimize the risk of contamination from other bacteria or untransformed XL1-blue colonies; and 4) there is little doubt of which colony you are trying to pick when you compare your filter lifts to the plates (and this is the single biggest advantage to moving colonies that convinced me to always do it).

There are a couple of different strategies to replating colonies. The first is to place your plate on a piece of engineering or graphing paper with a small grid (you're aiming for

about 250 colonies per plate, so pick an appropriate grid size) on it beneath your plates. Then you simply remove white colonies from your original plate and place the bacteria into the center of each grid square as you fill the plate. Alternatively, you can start on the outside of your plate (about 3 mm from the edge to ensure that all the colonies you transfer will be contacted by the filter when you lift them), and add colonies about 1-2 mm apart in a spiral until you fill the plate. In either case, you're aiming to get about 250 colonies with inserts onto the transfer plate. Select only relatively large and isolated white colonies to replat. If you are unsure that you can remove a single colony without touching a surrounding colony or if the colonies touch or overlap, discard the clones rather than risk contamination of one cell line by touching another. If you end up with more than a single bacterial clone in your transferred colonies, you won't be able to sequence the clone, and you'll lose the colony (and all the work you did to discover that clone had some microsatellite sequence in it!). While replating, try to avoid dipping the toothpick tip into the agar because the cells do not grow well under anaerobic conditions – try to simply rub the surface of the agar with the tip of the toothpick without actually breaking the surface if you can. You must pick the source colony right in its middle, and then only lightly touch the agar with the tip of the toothpick a few times on the new plate.



If the colony density is really low, light blue colonies MAY be chosen. These often have an insert positioned in such a way that the beta-galactosidase gene phase reading has not been modified, but it is always preferable to select all white clones if available because inserts that do not knock out the beta-galactosidase gene are often too small to get enough flanking

sequence to develop primers, even if they contain microsatellites.

If your critters are anything like mine, you'll need 4-6 such plates to get enough positive clones to end up with a reasonable number of microsatellite primers being identified. Once you have transferred the white colonies to the new plate, **incubate them upside down at 37°C for about 16 hours** to allow the colonies to grow enough to be lifted.

Some people plate a few positive control clones (in general, a positive clone from a previous cloning procedure that was performed with the same vector - pUC18) of each plate just to be sure that the probing is working, and to give an idea of probe intensity. This is a good idea for beginners and any time a new probe is used, but as you become more familiar with the techniques and know what positive clones should look like, you'll probably want to put as many colonies to probe on your plates as possible.

! You want to obtain a relatively large quantity of bacterial material -- large colonies -- from the transfer plates, but you must make sure that adjacent colonies do not converge or they become useless. You can accomplish this by making a tiny smear (about this big → •) with the tip of the toothpick when transferring bacteria to the new plate, rather than allowing extended growth periods. This will ensure that you get a large volume of bacterial cells during the filter transfer, and you can still control the spacing between colonies by keeping growth times relatively short!

Colonies grown on agar may be stored at 4°C for 2 to 3 weeks, so you need to move ahead at a reasonable pace to avoid losing all your work.



TRANSFER SOLUTIONS (autoclave and store at room temperature)

Solution 1 (SSC Stock): 20X SSC, pH 7.0 NaCl (175.3 g / l)
 Sodium citrate (88.2 g / l)

Dilute to 2x SSC (300 mM NaCl , 30 mM Na Citrate) for use.

Solution 2 (SDS Stock): 10% SDS (100 g / l)

Solution 3 (Lysis and Denaturing Sol'n): NaOH 0.4 M (16 g / l)
NaCl 0.6 M (35 g / l)

Solution 4 (Neutralizing Sol'n): Tris-HCl 0.5 M (60.7 g / l)
NaCl 1.5 M (87 g / l), pH 7.5

TRANSFER OF COLONIES

There are a wide selection of appropriate membranes for colony lifts, and if you have a particular brand or supplier to which you feel a particular loyalty, by all means use them.

However, I have found that Micron Separations Inc. (MSI) makes a superb product for a surprisingly low price (MSI through Fisher, CAT#N00HY08250 – not “N04HY...” like it says in the catalog!). I have tried several other brands of membranes and have not found a product for an equivalent price that I thought superior. We use only MSI filters in our lab for that reason. Either side of the nylon filters may be used, but make sure you **always** use forceps and wear gloves – **NEVER allow anything that may transfer DNA to contact your filters.**

Prior to transferring colonies, the nylon membranes must be labeled appropriately (as to the Petri dishes from which the colonies will be lifted, the date and the lift number – e.g., Pc1:4 35% 8/7/97 #1) **with pencil** (ink and even permanent marker will run in the SDS and NaOH solutions). You will need to make **2 lifts** of each plate for probing, so make sure that you have both enough filters and LB plates on which to regrow them, and that all filters are well labelled. Once labelled, submerge filters in 2x SSC (Solution 1 above) until completely wet, and then blot off excess SSC between 2 pieces of Whatman filter paper. Using 2 pairs of forceps to hold either side of the filter, position the filter **labelled side down** directly over the

center of the plate, and slowly lower the filter onto the colonies. Place one of the protective papers used to ship the filters over the filter, and gently press the filter onto the plate with a homemade plunger (it should be completely flat, waterproof, and only slightly smaller than the petri-dish opening – it's hard to describe better – any flat surface the size of the membrane that can be used to uniformly apply pressure to the entire surface to ensure homogeneous contact with the plate will work). **Before removing the filter**, use a hypodermic with a **tiny** amount of india ink on the tip to punch 3 or 4 holes close to the edge of the filters (choose an area between bacterial colonies – do not stab through a colony, or you will transfer bacteria around on your plate and contaminate other clones – and which is about 5 mm from the edge of the membrane to minimize the likelihood that the next filter will miss any of those alignment spots) in an **assymetrical** pattern into the membrane. The holes must be assymetrically placed so that you can unambiguously line the filter up with the plate again, and once you photocopy the filter, it is VERY easy to turn the filter upside-down by accident, so it is best to make the holes something like (.:) – or (..) if you have trouble aligning your plates – to make sure you can tell left from right on the round filters. The india ink marks both the filters and the agar so that the holes are obvious, but it spreads VERY rapidly on wet filters, so you want VERY little ink on the needle (i.e., wipe it off on a Kimwipe after dipping the hypodermic into the india ink, and just brush the edge of the hole with the blackened hypo. to darken the hole). Once pressed and punched, the filter should be **CAREFULLY** lifted with 2 pairs of forceps (use one pair of forceps on either side and lift towards the middle), placed **colony side up** (from this point on, the membranes should ALWAYS be kept **colony side up**, but remember to incubate them on **inverted petri-dishes** – you don't want anything to settle on your filters while they are in contact with the agar, or you could grow alien bacteria on your membrane!) on new LB agar plates (with Amp + Tet), and incubated at 37°C for ~2-3 hours.

! Do not allow the filter to lay back down once lifted from the plates, and do not drop the filter while moving it! Doing so is likely to smear the colonies on the membrane or contaminate the clones, either of which will reduce your ability to lift and sequence positive clones.

! Do not forget to mark the membranes before lifting the filter from the plate. If this is not done correctly, spotting and identifying the positive clones may be impossible, and all your hard work to this point will be in vain. Also make sure that you do not allow your filters to contact anything, especially the original plate, after you transfer and plate the colonies !

! This is one of the easiest stages to contaminate your bacteria and lose your work !

During the incubation, cut ~12 pieces of Whatman filter paper to fit the bottoms of the large (33x23x5 cm) Pyrex casserole dishes. Close to the end of the incubation, place 2-3 pieces of filter paper into each of 4 dishes, and saturate the filter paper in each dish with 50 ml of one of the following solutions: **Dish 1**) 10% SDS (Solution 2 above), **Dish 2**) lysis and denaturing solution (Solution 3 above), **Dish 3**) neutralizing solution (Solution 4 above), and **Dish 4**) 2x SSC with Proteinase-K (2 mg /ml). Saturate the filter papers with enough of each solution (usually about 50 ml) that the filter papers are totally wet, but there is no fluid sloshing around above the filter papers and they are completely flat (this may take some tilting and straightening of the filter papers to accomplish, you should also remove all the trapped air bubbles and straighten any folds in the filter papers at the same time). Also cut a pile of ~5" square pieces of Whatman filter paper to blot the filters as they are moved from one dish to the next, and pile them between the dishes for easy access while preparing the membranes. Very gently remove the filters from the growth plates, and place them colony side up on the first saturated filter paper (10% SDS, **Dish 1**) for 5 minutes. Remove filters one-by-one, dab excess fluid off the **bottom** of the filter by placing it briefly onto a clean piece of the Whatman filter paper you set between the dishes for this purpose, and transfer the filters to the second tray

Lysis buffer, **Dish 2**). Repeat this process for the third dish, but for the last wash with Proteinase-K, allow the filters to sit for at least 15 minutes before removing the membranes. For **ONLY** this final blotting, place the filter between 2 pieces of **clean** Whatman filter paper, and **very carefully** (i.e., do not allow the filter papers to move at all relative to the nylon membrane -- hold them firmly, otherwise they will shift and cause smearing of the lysed DNA on the filters) scrape a ruler a couple of times over the surface of the upper filter paper to remove any remaining cellular debris from the filters (after the first pass, you should see the dots of the cellular debris adhering to the filter paper). In my experience, omitting the Proteinase-K and filter-paper press steps leads to a dramatic increase in the number of false positives detected in later steps, because the antibody tends to adhere to the cellular debris left on the membranes. If you find that protein residues are still interfering with your detections even after this treatment, a deproteinization step (incubate filters for 50 minutes at 45°C in 50 ml of prehybridization solution -- see below -- with 0.5 mg / ml of Proteinase K) may be necessary, but I have not personally found this to be the case.

After pressing the membranes between filter paper with a ruler, fold a large sheet of Whatman paper and place the filters **colony side up** on the filter paper and allow them to dry for 30-60 minutes (until there is no moisture remaining in the filters -- they'll look very white with a slight silvery sheen when dry) before fixing the bacterial DNA to the filters by either baking them (at 80°C for ~2 hrs) or "Stratalinking" the filters with UV light (crosslinking the DNA to the nylon filters with 60 s of exposure to 40W of 254 nm UV light or 30 s exposure to 75W). Membranes onto which DNA has been transferred and linked can be stored dry at room temperature for extended periods of time (although you should always keep them wrapped in tinfoil and prevent dust or other contaminants settling onto the filters when doing so). Once wet, filters **should never be allowed to dry out again** (filters must be kept wet to allow

stripping and reprobing of the membranes -- if you plan to do only a single detection run, this comment does not apply), but can still be stored wet with 2x SSC in a tinfoil wrapped container at room temperature for extended periods.

CONTROL

At this point, it is useful to prepare a probe control by spotting 1 μ l of each labeled probe onto a well-labeled piece of Nylon membrane (circle the spot of probe and label it directly on the membrane). Fix the probes to the nylon filter in the same manner used to fix the DNA from the lysed colonies above.

I. SCREENING FOR POSITIVE CLONES

Unlike most protocols, we do not use a hybridization oven or hybridization tubes for this step of the probing. Because we do not use any radioactive solutions, we can share a regular incubator with both the cloning and the hybridization, and we simply use an orbital shaker table in the incubator to mix the filters. This reduces the total cost of developing microsatellite markers for most labs, because it is common to have a drying oven or some sort of incubator available to you somewhere in the department, and hybridization ovens are more difficult to come by, so it is likely that use of one would require purchasing one specifically for this procedure. By using a standard VWR incubator (model 1545) and a Lab Line variable speed orbital shaker, you can do this procedure in most labs equipped for doing any molecular work, without any special equipment purchases specifically for microsatellite isolation. Instead of hybridization tubes, we use Tucker Housewares 1.5 cup (360 ml) Serve & Store bowls with screw-top lids (No. 0571-1) for our hybridizations (plastic bowl hereafter). These

bowls are only slightly larger than the nylon membranes (95mm diameter), so they do not require much hybridization solution to immerse the membranes, but are large enough to still allow free movement of the membranes when on the shaker table.

① Prepare hybridization and detection solutions (modified from Engler-Blum et al., 1993):

Note: I have suggested the amount of each buffer to have on hand prior to beginning a probe/detection series, but the recipes are all for **ONE LITER (1L)** volumes. Adjust the amounts as necessary for the volume of buffer that you are making!

Rinsing solution:

2x SSC, 0.1% SDS Autoclave, then store at room temperature

Pre-Hybridization Solution (500 ml, pH 7.2):

0.25 mM Na ₂ HPO ₄	(35.49 g / l)
10% SDS	(100 g / l)
0.5% BM Blocking Reagent	(5 g / l) - Shipped with Genius detection kits
500 mM EDTA	(2 ml of 0.5M EDTA stock)
Do <u>NOT</u> Filter, store at 4°C	

Hybridization Washing Buffer (1.5 l):

20 mM Na ₂ HPO ₄	(2.84 g / l)
500 mM EDTA	(2 ml of 0.5M EDTA stock)
1% SDS	(10 g / l)
0.2 µm Filter, then store at 4°C	

Detection Washing Buffer (1.5 l, pH 8.0):

0.1 M Maleic Acid (C ₄ H ₄ O ₄)	(11.61 g / l)
3 M NaCl	(175.32 g / l)
0.6% Tween-20	(6 ml / l)

0.2 μ m Filter, then store at 4°C

Detection Blocking Buffer (500 ml, pH 8.0):

0.1 M Maleic Acid	(11.61 g / l)	
3 M NaCl	(175.32 g / l)	
0.6% Tween-20	(6 ml / l)	
0.5% BM Blocking Reagent	(5 g / l)	BM's "blocking reagent" is primarily casein hydrolysate.

DO NOT Filter, store at 4°C

NBT/BCIP colorimetric detection buffer (500 ml, pH 9.5):

100 mM Tris	(12.11 g / l)
50 mM MgCl ₂	(11.66 g / l)
100 mM NaCl	(5.85 g / l)

0.2 μ m Filter, then store at 4°C

Most protocols suggest autoclaving all buffers for long-term storage at room temperature, but in practice, the magnesium precipitates when these solutions are autoclaved. Therefore, I filter sterilize everything and store it at 4°C to keep the magnesium in solution. I have also found that filtering these buffers to 0.4 μ m further reduces background and leads to cleaner detections and fewer false positives.

RECOMMENDATIONS

This protocol remains identical, regardless of the number of membranes being hybridized. I usually hybridize a dozen filters (lifts from 6 transfer plates) at a time, but it is simple to increase or reduce that number by simply increasing or decreasing the amount of probe solution added to the plastic bowls. It is important to make sure that an incubator and a water bath (to preheat the solutions) are available, and at the correct temperatures. If you do not have

a water bath available, one can be made cheaply by purchasing a large (~250W) heater from the local petshop and placing it in a shallow tray or aquarium surrounded by Styrofoam SM or some other efficient insulator. Ideally, you want to start hybridizations early in the afternoon, and allow them to run over night before starting the detection the next morning. Stripping and reprobing is possible, simply remove the color with dimethylformamide, and then strip the probes with 0.1% SDS (see below).

! Membranes must not dry out at any point if they are to be stripped and reprobed. If the filters require storage, place them in 2x SSC and wrap the container in tin foil - otherwise the stripping and reprobing are unlikely to be successful.

PROTOCOL

- After transferring the filters to a new plastic bowl at each step, pour the used solution directly down the sink -
- * I assume that you are starting with the Genius 3 or 7 DIG Nucleic Acid Detection kits (Boehringer Mannheim No. 1 175 041 and 1 363 514, respectively).

① Filter Wetting:

Soak membranes in rinsing solution (see above) for 5 minutes



➡ ② Prehybridization:

Incubate the membranes for 60 minutes in 10-20 ml / membrane of prewarmed prehybridization buffer (prewarmed to the temperature at which the hybridization is to be carried out — this solution is very concentrated and will require substantial warming to redissolve, make sure you leave enough time to warm and completely dissolve this buffer). This step must be done **at the temperature at which the probing is to be conducted** (i.e., leave it in the incubator!). It is OK to just pour off this buffer and replace it with the probe, rather than individually transferring the filters for probe hybridization.



➡ ③ Hybridization:

Replace the prehybridization solution with 15 ml of prewarmed hybridization buffer (prehybridization solution + 150 pMol of DIG-labeled oligonucleotide probe) (*this volume - 15 ml - and the concentration of labeled probe may be adjusted according to the number of membranes in the hybridization*) and hybridize at the appropriate temperature (*hybridization temperature must be adapted to the set of probes used for the screening procedure see Appendix A — this is generally about 5-10°C below the mean melting temperature of the mixed probes*) overnight (at least 6 hours, up to 16 hours).

! Since probes are single-stranded, no preliminary denaturation is necessary. Also, the hybridization solution containing the DIG-labeled probe may be recovered and reused. I have successfully used the same probe solution as many as 4 times, after which the recovered volume became too low to probe another set of filters. The probe does eventually go off, though, so if you have used the probe previously, or it has been stored for a long time, make sure to test it again during your detection !



➡ ④ Posthybridization Washes:

Wash membranes 3 times, each for 20 minutes, in 50-100 ml of prewarmed hybridization washing buffer.

To minimize carry-over, membranes should be individually transferred to a new tupperware bowl containing fresh washing buffer, rather than simply dumping and replacing the solution.

This wash should be performed in the incubator **at the same temperature as the hybridization** (which will actually be a degree or two lower, because the incubator will cool slightly due to the frequent opening of the door during the washes). This is a very stringent wash, and if you find that you have no success with these washes, you may wish to drop the temperature about 5-10 degrees for this step.

! The remaining steps are performed at ROOM TEMPERATURE to avoid denaturing the antibodies. Allow enough time to warm the buffers to ~65°C, and allow them to cool back down to room temperature before using them. Make sure all detection buffers are at room temperature before use !

J. COLORIMETRIC DETECTION AND PICKING & STORING COLONIES

RECOMMENDATIONS

Your choice of a detection technique depends largely on how sensitive you want your screening to be. If you are interested in the most sensitive screening possible, follow the technique proposed by Engler-Blum et al. (1993, Anal. Biochem. 210:235-244). However, in most cases, the extra sensitivity afforded by the chemiluminescent detection techniques proves a detriment rather than an asset. I have tried BM's 3' end labeling kit (Genius 5) and the 3' end-tailing kit (Genius 6) together with each of the luminiscent detection (Genius 7) and colorimetric detection (Genius 3) kits, and found that the cheaper and easier kits (Genius 5 & 3) give more reliable results because they are less sensitive. I will provide only the protocol

for the techniques that I have adopted, but if you find that it proves too insensitive, you should consider going with the chemiluminescent detection kit, and, if that still proves too insensitive, possibly try the end-tailing kit (although in my experience end-tailing produced far too many false positives). Despite the claim of other protocols to the contrary, stripping and reprobing is possible and I have had good success with reusing the same filters a number of times. I occasionally find a positive clone that apparently did not strip and lights up again on reprobing, but when I have sequenced these clones, I generally find that there are multiple repeat motifs responsible for the second positive detection, rather than unsuccessful stripping. The frequency of apparent unsuccessful stripping is about 1 clone per 500 colonies, so I am willing to suffer through that potential inconvenience rather than make numerous replicate plates to avoid reprobing. The protocol for stripping and reprobing is included below.

PROTOCOL

! I will restate here that the remaining steps are performed at room temperature to avoid denaturing the antibodies. Although most of these buffers require substantial warming to completely dissolve the solutes after refrigeration, it is essential that they be allowed to cool down to room temperature before being used for the detection steps !

Following the posthybridization washes, equilibrate the membranes briefly (~ 5 min) in 50 ml detection washing buffer warmed to ~ 37°C before use.



► Warm blocking buffer to 65°C to redissolve the solution and **allow it to cool** before incubating the membranes at room temperature in 50-100 ml of blocking buffer for 60+ minutes to block nonspecific sites. Make sure that the membranes are shaking freely and are completely covered with blocking buffer during this step to ensure adequate blocking. The minimum time required for blocking is 30 minutes, but 60 is STRONGLY recommended (to avoid false positives and background), and longer blocking times are acceptable (i.e., if you gotta go for a long lunch, this is the time!). It is acceptable to simply dump the buffer before adding the anti-DIG at this step, rather than individually transferring the membranes.



► Incubate the membranes at room temperature for 30 minutes in a 1:5,000 dilution of Anti-DIG alkaline phosphatase in blocking buffer (2 µl of Anti-DIG in 10 ml of blocking buffer per

6 membranes). Once diluted, the antibody solution is stable for about 12 hrs @ 4°C, but should be discarded after use.

(Keep 200 µl of this solution for the colorimetric detection control described below)

! It is absolutely imperative that the Anti-DIG-AP Conjugate antibody be stored at 4°C and never be allowed to warm up -- use the antibody directly out of the fridge and store it on ice if it is to be removed for more than a few seconds !



➤ Wash twice to remove any unbound antibody from the filters or the containers. Individually transfer the membranes to a new bowl (this is the single most critical step for individual transfer of membranes), with 100-200 ml of detection washing buffer. Place on the shaker table for 15-20 minutes, and repeat. These washes remove unbound antibody, so minimizing carryover is particularly desirable here.



➤ Pour 10 ml of NBT/BCIP colorimetric detection buffer per membrane into a graduated cylinder. Wrap the graduated cylinder with a paper towel or tinfoil to minimize light penetration, and add 200 µl of NBT/BCIP stock (BM Cat#1 681 451, 8 ml -- or see below for detection recipe) per 10 ml of detection buffer (this solution should look bright yellow, but should remain virtually clear when added to the colorimetric detection buffer -- if the detection solution looks powdery upon adding the NBT/BCIP stock, there is a problem with your detection buffer). Store the detection solution at 4°C until used.

NBT/BCIP Stock Solution:

- ① dissolve 187.5 mg NBT and 94.0 mg BCIP in 6.7 ml of 100% DMSO (all molecular biology grade reagents)
- ② Once both reagents are completely dissolved in the DMSO, add 3.3 ml of UV-treated dH₂O to a final volume of 10 ml.
- ③ Store stock in complete darkness at 4°C

! BCIP is insoluble in water, and must be completely dissolved in 100% DMSO before ANY water is added to bring the DMSO concentration down to 67% or you'll never get it into solution !



➤ Add each membrane, bacterial clone side up to a petri dish, and cover it. You'll need a tinfoil lined styrofoam cooler, or some other container to prevent any light penetration in which to store the petri dishes while the color is developing. During the development there can be absolutely NO movement of the membranes, or you'll get smears that make it difficult to identify which clones are the positives. Move the pile of covered petri dishes next to your light-proof box, and get the dilute NBT/BCIP stock out of the frig. Pour enough to cover the membrane into the dish (~10 ml), cover it and place it immediately into the light-proof box (after which the box does NOT move until the detection is complete). Repeat for all your membranes. The membranes can be exposed to light for very short periods to check development, but any light exposure increases the background color on the membranes.



► Allow the detection to develop undisturbed for at least a 3-6 hours, but preferably overnight (color development begins within a couple of minutes and is complete within 12 hrs, if you can wait for the precipitation to end, you will get clearer spots for colony matching and selection).

Positive clones are dark violet (the color intensity is more or less related to the microsatellite size):

negative clones are usually colorless, but could appear light blue to brown, but are never purply violet.

Wash the filters with dH₂O, and place them into a clear acetate sheet protector (e.g., K&M from Avery Economy weight side loading sheet protectors, #2VY811), and photocopy them onto overhead transparencies. This allows direct comparison of the membranes to the plates and to one another, without requiring you to dry the membranes which would prevent you from stripping the filters and reprobing.

This step is also where the importance of the assymetrically placed holes in the filters becomes obvious — in order to pick the colonies from which the positive spots on the photocopy were obtained, it is **essential** that you be able to unambiguously align the photocopy with the colonies on the plate.

! Photocopying the filters has inverted the colony positions on the membrane. You need to turn the photocopy “upside down” (with the letters and numbers as seen in a mirror) in order to align the three reference holes to one another before trying to line positive marks up with bacterial colonies. Again, make sure that the marks are asymmetrical (.:) — or (.:) if you have problems aligning the filters — on the plate, because it is easy to mistake which side of the plate is which otherwise !

Unfortunately, much of the scoring accuracy comes with experience: you need to take into consideration the size, intensity and color of each spot. You should try to avoid skipping a positive clone forming a small spot because of a poor transfer, and picking a spot that is large but whose coloration is “wrong.” Thus, although you photocopy the membranes for your permanent records, and for alignment to the plates on which the clones are grown, you should first examine the spots on the membranes themselves, and determine which spots “look good to you” before examining the photocopies. It is very useful to use positive clones containing large microsatellite (30 or more repetitions) as controls for reference. These generally produce very large and obviously colored spots against which the putative positives on your membranes can be compared. The use of replicate lifts allows one further check on the validity of apparently good spots, because only spots that appear on **BOTH** membranes are used for subsequent steps in the procedure.

CONTROLS

① Probe labelling and membrane treatment control

This control is outlined in the first step of the protocol (probe labeling), and should be done anytime that

you make a new probe or use an old one. Spotting of labeled oligos onto a control membrane is a good way to make a positive control if you don't have known microsatellites to add into

your replated colonies earlier. You need only add 2 µl of control labeled DNA onto a small piece of membrane (it will fill a spot about this large  on your filter, so I usually cut a piece of membrane into eighths for this purpose), allow it to dry, Stratalink it with UV as above, and treat it as one of your filters that just came out of the post-hybridization washes during the detection protocols. Alternatively, you could designate a small spot on your filter for the control oligo prior to starting any of the probing steps -- simply add 2 µl to that spot after having lysed the cells and treated them with proteinase-K, prior to allowing the filters to dry for cross-linking to the membranes.

② Coloration reaction control

Simply add ~200 µl of diluted NBT/BCIP detection solution to ~200 µl of the 1:5000 dilution of the blocking buffer containing Anti-DIG antibodies before discarding the solution. It should turn dark purple-blue within minutes.

③ Positive controls

Adding several known positive clones to each membrane provides the best positive control, but if that proves awkward, use control ① for your positive control.

DECOLORATION OF MEMBRANES AND STRIPPING OF PROBES:

To quote Estoup & Turgeon's microsatellite protocol, "None of the protocols proposed by Boehringer were fully satisfactory." I completely agree, however, that does not mean that it cannot be done successfully. The membranes may be decolored sufficiently to reuse by incubating the membranes on an orbital shaker in pure dimethylformamide at 55°C.

! Dimethylformamide is a hazardous solvent, and should always be used in a fume hood!
CAUTION! The flash point of dimethylformamide is 67°C, so DO NOT warm to above 55°C just to be safe !

The time required and efficacy of the color removal depends on the temperature and the frequency of solvent change. If you find that the spots are not disappearing from your membranes, change the solvent more frequently, or split the membranes into more containers.

Once the color has been completely removed from your membranes, the hybridized probes may be removed by simply incubating the membranes at ~95°C for 5 minutes in 0.1% SDS. Again, it is preferable if the membranes have room to move freely during this stripping, but it is difficult to get any incubating shaker to that temperature, so I generally use a large dish and stir it slowly during the incubation.

! Again, it is essential that the filters not be allowed to dry out at any point during the procedure if the filters are to be stripped and reprobed. Always store filters in the dark in 2x SSC when they are not in use !

PICKING AND STORING COLONIES:

Once you have identified which bacterial clones appear to contain microsatellites, it is important to make backup stocks of those clones. This is done by carefully labeling another transfer plate (a "positives plate") with the ID of each clone and a circle (e.g., P.c. #125 O) into which the colony is to be introduced. Also prepare a series of labeled test-tubes with ~5 ml of 2x YT broth (containing ~ 5 μ l (1 μ l / ml) of each Amp & Tet stocks, see. Appendix A). Make two additional test tubes of broth: 1) one negative control that lacks antibiotics, and 2) a second negative control that contains the antibiotics -- neither of these should EVER show any growth for you to be confident that your cultures contain only the positive clones that you picked to inoculate the cultures. If there is growth in only the first negative control, you may be able to get clean sequence from the cultures, but you certainly do not want to make glycerol stocks from these cultures, and it would be better if you made new broth and lifted those colonies from the "positives transfer plate" again.

To transfer the colonies, use a small pipette with a relatively wide bore (e.g., Eppendorf P10), punch out the center of the colony and then touch the agar inside the circle of the new plate several times to make a dot the size of the transfers you made with the toothpicks on the original transfer plates. Then eject the plug of agar into the antibiotic medium and reflux several times.

! It is essential that you not allow the reflux broth to contact the piston of the pipette or aspirate any of the broth, or you will contaminate all subsequent cultures with that clone, and be unable to sequence any of them for primer determination!
 Slowly refluxing the broth decreases the likelihood of contamination, and using smaller volume, long pipette tip decreases the likelihood of piston contact. If you suspect that the pipette may have been contaminated, wash it out well with 10M NaOH and allow it to dry before continuing.

Repeat for every putative positive clone so that you'll have a repeat plate of colonies from which to regrow the overnights if something fails, and if everything works, you'll have 5

ml of pure stock from which to make glycerol stocks and minipreps (see below). Grow both the positive transfer plates and the test tubes at 37°C for 12-16 hours (overnight).

Assuming that both negative controls came out clean, and that all your test tubes now look milky from bacterial growth, you'll want to make glycerol stocks of those clones, and spin down the remaining cells in the culture from which to isolate the plasmids for sequencing (see miniprep protocol, below).

To make glycerol stocks, first get a dewar of liquid nitrogen for flash freezing the stocks. Then, using a wide-bore pipette, add 70 µl of each culture to a well-labeled 0.5 ml snap-cap vial (make sure that you use good sterile technique here, you'll never get a chance to go back and salvage these stocks if you contaminate them, and all your work to this point will have been in vain!). Finally add 30µl of autoclaved and UV-treated glycerol (be careful, glycerol is very viscous, and you'll need to pipette slowly here, even though you're using wide-bore pipette tips) to each tube and mix by refluxing the contents several times (until you cannot see any remaining glycerol). Once the culture is well mixed, simply place the tube into the liquid nitrogen to flash freeze it, and when done, transfer all the stocks to a -80°C freezer for long-term storage. Just to be safe, it is recommended that you make two sets of the entire clone collection and store each set in a different freezer. Then, if you need to start up a new culture, simply thaw the glycerol stock on ice, and transfer ~5 µl to a test tube containing ~5ml of 2x YT broth with antibiotics, and incubate the tube overnight at 37°C again. Although the cells retain their original viability for several years when stored in at -80°C in glycerol, repeated thaws will quickly decrease cell viability -- if you are forced to thaw the stocks more than a couple of times, make new glycerol stocks from the overnights (use a double dose of antibiotics for the overnights if you plan to make new glycerol stocks, just to make sure that **nothing** else grows in your culture). If you anticipate having to thaw your stocks several

times in the future, make several glycerol stocks, and use the spares while keeping your "original" stock for back-up.

ESTIMATE INSERT SIZES FOR Universal Primers

White colonies do not necessarily contain insert of an appropriate size. The partial genomic library of the putative positives should be checked to determine the proportion of clones with insert of an appropriate size.

- ① Working in a sterile environment (e.g., under a laminar flow fumehood), use a sterile toothpick, and dip one tip into the overnight growth (the 5ml of 2xYT broth) from each positive clone (alternatively, you can simply touch the center of each colony on the positive transfer plate with the toothpick tip) and briefly (~2 sec) swirl the toothpick in a **different** PCR tube for each clone containing 50 µl sterile, UV-treated H₂O (if touching the clones on the plate, allow the toothpicks to soak in their tube while you complete the rest of the clones before swirling them) to detach the cells from the toothpick and leave them in the water.
- ② **Incubate the tubes in boiling (100°C) water for 5 minutes to rupture the cells.** (fill the holes in the hot block with water and let it start boiling before adding your PCR tube to the hole and timing your incubation) Place the tube directly onto ice immediately after boiling.
- ③ Spin the tube in the fridge for 5 minutes at 10 000 g to pellet cellular debris, and immediately place the tube back on ice. **Store at 4°C** (for short periods of time) or **at -20°C** (for long-term storage).
- ④ Remove **2 µl of supernatant** as the PCR template, and add it to a series of reaction tubes. Make up a PCR reaction mixture to aliquot into these tubes from the following recipe (I outline the recipe for the Perkin Elmer Ampli-Taq that we use. Other vendors, e.g., Promega and Gibco make equivalent products that can also be used -- follow their recommendations for concentrations if they differ from the recipe I present here):

PCR Mix -- For each reaction (final volume: 10 µl, including the 2µl of supernatant) combine:

Reagent	Volume (µl)	Stock concentration
10X Buffer (Perkin Elmer)	1	10X
Mg ²⁺ (Perkin Elmer)	0.5	25 mM

dNTPs (A,T,G,C combined)	1	1.0 mM per nucleotide
M13 Universal Primer	0.6	10 μ M
M13 Reverse Primer	0.6	10 μ M
Ampli-Taq Polymerase (Perkin Elmer)	0.08	5 U/ μ l
H₂O (Sterile UV-treated)	4.22	N/A
Total Volume	8.0 μ l	N/A

NOTE: In the preparation of a PCR mix, it is a good idea to always allow 5-10% additional volume for pipetting error and liquid retention in tubes/tips. On the other hand, Taq is very expensive, so don't mix up too much and waste reagents, so, for example, I usually prepare 10.5-11 volumes of the mixture for a set of 10 reaction tubes.

Add **8.0 μ l of the PCR mix to each tube containing 2 μ l of supernatant** (added per PCR tube, as described above), and amplify the vector via PCR.

REACTION CONDITIONS FOR PCR:

Depending on which thermocycler and PCR tubes you use, the cycle times you will use may vary greatly. I am not going to suggest times for every machine, but rather simply say that using the Perkin Elmer 2400 Thermocycler (accepts thin-walled tubes, and has rapid ramp times), I have great success using the following protocol:

INITIAL DENATURATION: 5 minutes at 95°C

CYCLES OF DENATURATION - PRIMER ANNEALING - ELONGATION:

3 CYCLES: 95°C - 1 minute 50°C - 1 minute, 30 seconds 72°C - 1 minute, 30 seconds +
25 CYCLES: 95°C - 30 seconds 50°C - 30 seconds 72°C - 1 minute

FINAL ELONGATION: 7 minutes at 72°C

→ Ramp down to 4°C and then refrigerate until loaded on a gel.

Note: These times may still seem a bit long to those familiar with PCR, but the long times and the high concentration of Mg²⁺ help to ensure amplification of very large (>1000 bp) fragments. Under the PCR conditions with which most people will be familiar (2 min denaturation times and lower Mg²⁺ concentration) you will most certainly not be able to amplify for small fragments but it is possible that you may be able to find the longer fragments.

bp) in which you are interested here.

- Electrophoresis of PCR products on a Synergel (1% Synergel, 0.6% Agarose): Add 3-4 μ l of loading buffer per tube. Load ~7-10 μ l of the mixture per well on a 1.6% Synergel (a minigel works fine for this), and run it at ~140 V for 30-40 minutes.
- Results: Stain the gel with EtBr, and photograph the gel using the UV lightbox.

! Due to the location of the primer sites on the vector, it is necessary to take some number of bp off the amplified fragment size to obtain the actual size of the insert (~100 bp for pUC 18 or 19 vectors; ~250 bp for Bluescript). A vector with NO insert will give a PCR product of ~250 bp (for Bluescript, or ~100 bp for pUC), so you should get sizes larger than that for all your clones. Use a clone that lacks an insert (i.e., a blue clone or an untransformed cell line) as a control, and subtract that value from your observed fragment lengths to determine the mean insert size !

It will probably be necessary to optimize the PCR reaction conditions depending on your results obtained with the protocol I have proposed here (see the PCR conditions recommended later in this manual -- Section O -- for those typical of the ones I use to amplify my microsatellite loci). Given that, it is best to perform your first PCR reaction on a small number of clones to ensure that you get the expected results before devoting a lot of time and energy to setting up every clone in this way. If you do not get amplification in a large number of lanes, you may have large fragments that require longer elongation times, and should try a more extreme protocol, such as that proposed by Estoup & Turgeon (below). You can play with all these conditions to optimize the PCR in order to give you an accurate estimate of the insert size in your plasmids. If you get results with my protocol, though, it is MUCH faster to run many samples using the shorter PCR times.

There are a number of different PCR protocols for determining the average length of inserts, but the most extreme (in terms of PCR reaction times) that I have seen is the protocol suggested by Estoup & Turgeon:

PCR Mix: for each reaction (final volume: 12.5 μ l) combine:

Reagent	Volume (μ l)	Final concentration or amount	Stock concentration
10X Buffer (Perkin Elmer)	1.25	1X	10X
Mg ²⁺ (Perkin Elmer)	1.5	3 mM	25 mM
dNTPs (A,T,G,C combined)	0.6	120 μ M	2.5 mM per nucleotide
M13 Universal Primer	0.6	480 nM	10 μ M
M13 Reverse Primer	0.6	480 nM	10 μ M
Ampli-Taq Polymerase (PE)	0.2	1 U	5 U/ μ l
BSA (1 mg/ml)	0.25	20 μ g/ml	1 mg/ml
H ₂ O (Sterile UV-treated)	5.5	N/A	N/A

INITIAL DENATURATION: 5 minutes at 96°C

CYCLES OF DENATURATION - PRIMER ANNEALING - ELONGATION:

5 CYCLES: 96°C - 1 minute 43°C - 1 minute 72°C - 4 minutes
+
25 CYCLES: 94°C - 1 minute 43°C - 1 minute 72°C - 4 minutes

FINAL ELONGATION: 7-10 minutes at 72°C

Regardless of what PCR conditions you use to get fragment amplification, once the insert sizes have been estimated, you should only select the positive clones with insert sizes greater than 400 but less than 1000 bp (**remember to subtract your control length from the total length of the fragment!**). If you get desperate for more markers, you can relax these criteria for the insert size to as short as 300 or as long as 1200 bp, but beyond those ranges, sequencing the inserts gives such low probability of getting good flanking sequence with which to develop primers that there is little point in sequencing anything <300 or >1200. The plasmids from the selected clones (those good positives with insert sizes between 400-1000 bp) are then isolated by a Miniprep (we use Qiagen's QiaPrep Spin Miniprep -- see below).

L. RECOVER PLASMIDS (MINIPREP)

PRE-CULTURE:

If you were unsuccessful in preparing the pre-culture in the previous step (i.e., your negative controls grew something), you'll need to make a new overnight growth from which to isolate the plasmids for sequencing, as outlined below (otherwise, skip ahead to the Miniprep section below).

For each plasmid (clone) to be sequenced, prepare a 16 ml polypropylene tube with:

- 2-5 ml liquid 2xYT or LB medium (See Appendix B)
- 2-5 μ l Ampicillin at 50 mg/ml (1 μ l/ml)
- 2-5 μ l Tetracycline at 25 mg/ml (1 μ l/ml)

Note: If you find that your antibiotics precipitate when mixed, try adding 10 μ l/ml of 0.1X stock concentrations (i.e, 5mg/ml Amp & 2.5 mg/ml Tet).

Innoculate each tube by picking a fraction of the clone from the positive transfer plate with a sterile toothpick (simply touch the center of the colony with the toothpick tip and drop the toothpick in the 16 ml tube).

Incubate overnight (12-16 hrs) at 37°C in shaking/rocking incubator (~225 rpm).

PLASMID ISOLATION (MINIPREP):

There are two ways to do this: 1) the fast and easy way (but also expensive), and 2) the long and hard way (but also cheaper).

The fast and easy way is to buy a prepared spin miniprep kit (the QIAprep spin miniprep kit – Cat. No. 27104 – gives us the most consistent results of any kit that we've tried yet). For this prep kit, follow the instructions exactly, but skip the optional wash with the "PB Buffer" (Step 7 in the protocol we have) and elute the DNA with 50 μ l of dH₂O rather than TE or their new "EB Buffer" (both of which seem to interfere with automated sequencing to some extent). Although the % recovery is slightly higher when the DNA is eluted with 100 μ l, the

increase in efficiency does not compensate for the decrease in concentration, so stick with 50 μ l of dH₂O for the elution step. Also allow the filters to stand for 5 minutes with the water in them, rather than for 1 minute as called for in the protocol (there is no significant difference in recovery efficiency between 5 and 30 minutes, but there is between 1 and 5 minutes).

Although using this kit runs about \$1 per prep, if you're going to spend \$30 on sequencing afterwards, in my opinion, it is much better to spend the extra dollar and have a really clean product for sequencing than to risk having to repeat the sequencing due to product quality. If you follow this advice, and buy the Qiagen kit, skip down to the section labeled Test: Digestion, below.

The long and hard way is to follow the miniprep protocol outlined in Sambrook, Fritsch & Maniatis: Molecular Cloning, a laboratory manual, or that outlined by Estoup & Turgeon (included below):

- Gently swirl the preculture and transfer it into 2 ml tubes (work in a sterile environment, such as a laminar-flow fumehood).
- Spin at 12000 rpm for 5 minutes to pellet the bacterial cells.
- Pour off the supernatant with precaution.

Optional cleaning step (to remove LB media in the pellet):

- Add 1 ml cold TE, pH=7.5 (10 mM Tris, 1 mM EDTA, pH 7.5).
- Vortex.
- Spin 5 minutes at 12000 rpm
- Carefully pour off the supernatant.

- Place on ice.
- Add 150 μ l of cold TE, pH=7.5 (10 mM Tris, 1 mM EDTA, pH 7.5). Immediately vortex each tube, then vortex each tube once again to resuspend the pellets.
- Place tubes on ice for 5 minutes.
- For the alkaline lysis, prepare, immediately prior to use:
 - 2 ml NaOH (1N) (0.2 M final)
 - H₂O sterile 7.5 ml
 - 500 μ l SDS (20%) (1% final)
 Add 260 μ l of this mix into each tube. Gently shake.

- Place on ice for 5 minutes.
- Neutralize with 195 μ l 3M NaOAc, pH 5.3 (adjusted with acetic acid). Shake gently.
- Place on ice for 5 minutes.
- Spin at 12000 rpm for 10 minutes.
- Recover the supernatant. Avoid pipetting white flocculating impurities.
- Phenol/Chloroform cleaning:
 - Add 1V Phenol - Shake - Spin at 12000 rpm for 10 minutes - Recover the supernatant.
 - Add 1V Chloroform - Shake - Spin at 12000 rpm for 10 minutes - Recover the supernatant.
- Add 2 Volumes of 100% Ethanol. Put in liquid N for 2 minutes or 30 minutes at -80°C .
- Spin at 12000 rpm for 15 minutes. Gently pour off the supernatant.
- Add 800 μ l of 70% Ethanol
- Spin at 12000 rpm for 10 minutes. Gently pour off the supernatant.
- Dry in a speedvac and resuspend the DNA pellet in 35 μ l sterile H_2O (don't use TE for sequencing, but it is better for long-term storage of the DNA if you are not going to use an automated sequencer.)

NOTE: If Miniprep products are to be sequenced using a thermostable polymerase, the phenol/chloroform steps **must absolutely be avoided** because these polymerases are extremely sensitive to very low phenol contamination. The use of potassium salts is also to be avoided. As an alternative, follow the above protocol, but neutralize the alkaline lysis with 3M NaOAc, pH=5, and perform two ethanol precipitation, adding 1/20 V of NaOAc instead of using the phenol/chloroform steps.

TEST: DIGESTION

Here are two alternative ways to remove the insert from the pUC 18 plasmid (total length= 2638 bp):

There is one specific *PvuII* restriction site on each side of the inserting point, such that a digest with *PvuII* will yield a 322 bp product, plus the length of your insert:

PvuII + *PvuII* = 322 bp

Alternatively, there is an *EcoRI* site on one side of the insertion point, and a *HindIII* site on the other side of it, such that a digest with these will yield a 51 bp product, plus the length of your insert:

HindIII + *EcoRI* = 51 bp

For the pBluescript II plasmid (total length = 2961 bp) there are a number of similar alternatives:

EcoRI + *EcoRI* = 18 bp

PstI + *XbaI* = 8 bp

EcoRV + *SpeI* = 10 bp

HindIII + *KpnI* = 32 bp

AclI + *MucI* = 30 bp

For each clone, prepare a tube with:

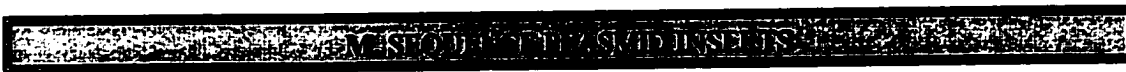
- 2 μ l of miniprep solution
- 1 μ l of a single restriction enzyme (*PvuII* or *EcoRI*) or 0.5 μ l of each pair of enzymes.
- 1 μ l 10X restriction enzyme buffer
- 0.2 μ l RNase
- dH₂O to a final volume of 10 μ l (5.8 μ l)

Incubate 1 hour at 37°C.

Load 10 μ l of digested miniprep + 5 μ l loading buffer on a 1% synergel, 0.6% agarose minigel.

Add 5-7 μ l of 100 or 123bp and 1kb ladders alongside the sample, and run the gel at ~140 V.

Stain and read the gel as described previously. In order to estimate the size of the insert from your organism, it is necessary to subtract the size of the vector fragment (see box above) between the two restriction sites from the size of the fragment seen on the agarose gel. The size estimate obtained with the Miniprep should be approximately the same as that provided by the PCR (boiling) test earlier. Inserts ranging in size from 500 to 1000 bp should be sequenced first, and shorter/longer inserts selected as necessary (see earlier discussion of limits on acceptable sized inserts).



RECOMMENDATIONS

If it is at all possible for you to send this off to an automated sequencing facility, do so (e.g., Davis Sequencing – <http://www.DavisSequencing.com> – charges \$15 and all you need to provide is 7 μ l of your miniprep template). Manual sequencing is both costly and time/labor intensive for those not already set up to do this. Also, if you need this protocol to develop microsatellites, chances are fair that you are not prepared to manually sequence your positive

clones and get sequence that is good enough that it can be used to develop unique primers. If you are unable to farm out your sequencing to an automated sequencing facility, Boehringer Mannheim makes a chemiluminescent sequencing kit (BM Cat. No. 1-449-443), that you can simply follow the direction provided with the kit to sequence your clones (also, you will need the detection set for this kit later to score your microsatellites, in any case), or you can use the standard T7 radioactive sequencing kit (BM Cat. No. 1-427-571). Manual sequencing generally limits your sequencing to reads of 2-300 bp, and will therefore make your life more difficult, but it is possible (I am told....).

N. READ SEQUENCE AND SELECT PRIMERS

If you have been able to send your sequencing to an automated facility, you will get back a simple ABI file that can be viewed in a number of software packages. Perkin Elmer provides ABI Prism software for the Macintosh, and there are at least 2 PC packages (at least there are 2 of which I am aware) that are available for downloading from the web (Chromas and ABI View). Personally, I use Chromas, but everyone has a favorite software package and a perfectly good reason why that package is better than any other. Chromas is available at: <http://trishul.sci.gu.edu.au/~conor/chromas.html>, and ABI-View is available at: <http://www.paranoia.com/~dhk/abiview.html>.

If you are forced to sequence your positive clones manually, I advise you to follow the protocols provided with your kit, or by your colleagues. I have managed to avoid manual sequencing, and can only provide the protocol suggested by Estoup & Turgeon given below:

- Preparation of 6 % acrylamide gel, gel pouring, and sequence loading, see Appendix C
- An initial 2H30 migration allows preliminary characterization of the microsatellite: motif, size, and location within the insert.

- From the bottom up, the sequence is read in the 5'⇒3' direction. Universal forward and reverse primers give the sequences of complementary strains, each starting from one extremity of the insert.
- Insert size must be taken into consideration to determine an appropriate migration time.
- Using universal, such as M13(-21), forward and reverse primers, the microsatellite must be isolated so that the sequences on each side can be read without ambiguity. Several sequencing gels are usually necessary for clean reads of the complete flanking sequences for each microsatellite.
- After the first 2H30 migration with the Universal Primers, one can:
 - Do a 7h migration with Universal Primer if the microsatellite is too high on the 2h30 gel.
 - Do a 2h30 migration with Reverse Primer if the microsatellite is too high on the 2h30 gel, then a 7h migration with Reverse Primer if the microsatellite is still too high.
 - Choose a first primer and do a second sequencing reaction with that primer: in doing so, the size of the fragment to be sequenced will be reduced, making it possible to read the region from which the second microsatellite primer is to be chosen.

Note: It is also important to determine the size of the fragment to be amplified (number of base pairs composing the microsatellite, the two primers, and the intervening sequences), because you then know the exact size of the PCR product that should be produced, and the Miniprep extraction can thereafter be used as a reference microsatellite (known fragment size of this particular allele). If a sequence is compared with another allele, the size correspondence may be biased by +/- one nucleotide because of the difference in nucleotide composition. For that reason, the most precise allele size reference is the cloned allele PCR product (Template = 1/1000 Miniprep dilution).

Assuming that you've been able to sequence your positive clones at this point, you now

know the sequence of the flanking region and the size of the repeat itself. In order to amplify that locus, you must develop two primers with the following characteristics for each microsatellite locus you hope to use:

- They should both be read without error (there must be **NO** undetermined nucleotides).
- They should neither be too close nor too far away from the SSR (aim for 10 to 60 bp).
- They should both consist of a n 18 to 27bp (20bp is "ideal") nucleotide sequence **WITHOUT** apparent repetitions and with an a priori random succession of nucleotides, and should amplify a product between 100 and 300 bp.
- They should both be rich in G and C (between 20 and 80%, but if possible, more than 50%) regularly distributed along the primer sequence (in particular, avoid series of A or T of more than 5 nucleotides).
- They should start and end with minimally one G or one C.
- They should both have T_m between 57 and 68°C -- if that yields nothing you could consider 45-65°C.
- They should both have low or no complementarity.
- The size and T_m of both primers should be roughly equal (although 3-4°C difference may still give good results), and the primers should not cross-hybridize. If possible, mispriming should also be checked, and avoided.

Naturally, it is difficult to examine all permutations of possible primers by eye and decide which is likely to be a good choice, given all these constraints. Fortunately, there are a number of computer programs that calculate all of these primer parameters: T_m , complementarity, %GC, product size, etc. **OLIGO**, is a popular program available for either the PC or Mac, and **Cprimer** is a popular freeware Mac program (available from the Indiana University Biogopher at <http://iubio.bio.indiana.edu:80/IUBio-Software%2bData/> ⇒ click on "molecular biology", and then on "mac" — it's in there somewhere), but I use **Primer3** -- a CGI script program available for free on the Web

(<http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) that includes a human mispriming library option, and both explains and allows you to manipulate the stringency of all the factors I just mentioned above.

Primers are then ordered (I currently order all my primers from either the Great American Gene Company or from Operon technologies -- <http://operon.com/custoligos.html> -- because both average about 60¢ per base, and deliver enough oligo to make about 6-8ml of primer) and resuspended in **sterile UV treated water or low TE** to a final concentration of 5 μ M for use (*I usually make several concentrated stocks -- around 500 μ M each -- for storage, and then dry them under vacuum for long-term storage at -20°C. Each of those stocks can then be resuspended at 5-50 μ M (depending on the purpose) and aliquoted for use to minimize contamination risks. You can also use aerosol-resistant tips at this step to avoid contamination of the primers stock).*



First, the cold PCR is optimized, and only then should the labeled PCR tests (whether radioactive, chemiluminescent or automated sequencer fluorescent tagged primers) be initiated.

The cold PCR will allow you to cheaply determine what conditions (if any) are suitable for the amplification of a specific microsatellite product; the labeled PCR and the migration on acrylamide gels will allow you to read allelic variability among individuals.

The first step in determining whether or not you will be able to score allelic variation (if the locus is variable) for a given microsatellite is to attempt to amplify the product from both the individual from which you originally constructed your size-selected partial genomic library, and from your transformed bacterial clone which was used for the minipreps. To do this, I use 2.5 μ l of 10ng/ μ l DNA from the original extraction as the PCR template, and add it to a series of reaction tubes. I also add 2.5 μ l of DNA template from the bacterial clone (isolated as described on Page 38) for a positive control. The PCR mix and conditions are listed below:

PCR Mix – For each reaction (final volume: 25 μ l, including the 2.5 μ l of supernatant) combine:

Reagent	Volume (μ l)	Volume (μ l)	Stock concentration
10X Buffer (Perkin Elmer)	2.5	2.5	10X
Mg ²⁺ (Perkin Elmer)	1.5	1.5	25 mM
dNTPs (A,T,G,C combined)	1.5	1.5	2.5 mM per nucleotide
Custom Forward Primer	2.5	2.5	5 μ M
Custom Reverse Primer	2.5	2.5	5 μ M
Ampli-Taq Polymerase (PE)	0.2	0.2	5 U/ μ l
H ₂ O (Sterile UV-treated)	11.8	5.55	N/A
BSA	0	6.25	1 mg/ml
Total Volume	22.5 μl		

Remember: It is a good idea to always allow 5-10% additional volume for pipetting error in making your reaction soup.

To determine the initial conditions for PCR, you can estimate the “ideal” annealing

temperature for your primers using the following formula (for primers P1 and P2 with respective T_m of T1 and T2 as determined by either the OLIGO or Primer3 program):

$$\frac{T_1 + T_2}{2} + 5^\circ C$$

I generally try a series of PCR reactions with the annealing temperature being changed by 5°C above and below that calculated annealing temperature.

PCR REACTION

The typical conditions I use for my first attempt at amplifying a new primer set are simply:

INITIAL DENATURATION: 4 minutes at 95°C

CYCLES OF DENATURATION - PRIMER ANNEALING - ELONGATION:

25 CYCLES:

94°C - 15 - 30 seconds

Calculated annealing temp. - 15 - 30 seconds

72°C - 15 - 30 seconds

FINAL ELONGATION: 7 minutes at 72°C (you will probably want to extend this elongation step to 30 minutes when amplifying your fragments to score allelic identity, because that tends to force uniform adenylation of PCR fragments which may otherwise lead to single base-pair differences in identically-sized fragments, but that is not really a concern at this stage).

→ Ramp down to 4°C and then refrigerate until loaded on a gel.

If this works fine, you may want to try to cut down the PCR times even further. Pam Jensen (in Paul Bentzen's lab at the University of Washington) uses the following protocol for amplifying her microsatellites: one cycle of 96°C for 10s; twenty-five cycles of 94°C for 10s, annealing temp for 15s, 72°C for 30s; 4°C hold). If this protocol will work well for you (as it seems to for Pam and I), it will dramatically cut your PCR times and is worth a try.

If the above conditions do not work, you can try:

INITIAL DENATURATION: 5 minutes at 96°C

CYCLES OF DENATURATION - PRIMER ANNEALING - ELONGATION:

3 CYCLES: 96°C - 2 minute 5°C below Annealing Temperature - 1 minute 72°C - 1 minute
+
25 CYCLES: 95°C - 30 seconds 5°C below Annealing Temperature - 30 seconds 72°C - 45 seconds

FINAL ELONGATION: 7 minutes at 72°C

→ Ramp down to 4°C and then refrigerate until loaded on a gel.

Depending on the results you get between these two sets of PCR conditions, you may want to try simply increasing the number of cycles on the first protocol (if the product looks clean from the latter conditions), or decreasing the number of cycles on the second protocol (if the product appears but looks 'stuttered').

LOADING PCR PRODUCTS ONTO A GEL:

Add 4 µl of loading buffer.

Load ~10µl per well on a 2% agarose minigel. Run the gel at 140 Volts for 30-40 minutes

Stain the gel with EtBr, as previously described.

The PCR product pattern obtained on the agarose gel must be clean, i.e. no DNA smear (typically called "stutter bands") and no aspecific bands, and of the correct size (determined by the product bp size predicted by your sequence when entered into Oligo, Cprimer or Primer3). The product from the individual and from the bacterial colony must be identical and of the correct

size. The concentration of agarose you'll want to use for the gel will vary by the product size you are amplifying (see below for a discussion of gel types and concentrations). If you do not get an ideal gel the first time (and you usually don't), you can play with the PCR conditions (usually the annealing temperature), the MgCl_2 concentration, and the template and primer concentrations trying to clean up the product (and I usually run all these variations regardless of what the first gel looks like just to get the best looking product I can squeeze out of a given set of primers -- see below).

- Increasing the annealing temperature increases the stringency for the binding of primers with the template DNA: nonspecific PCR products/bands may be eliminated, but the amount of desired PCR product will also be reduced, and in some cases cycle number must be increased to compensate.
- Decreasing the Mg^{2+} concentration will help eliminate nonspecific bands. Increasing the Mg^{2+} concentration will have the opposite effects. Start with an initial Mg^{2+} concentration of 1.0 mM. Mg^{2+} concentration must be kept between 0.7 mM and 1.5 mM to get amplification.
- Start with 25 cycles. This number of cycles can be reduced to eliminate nonspecific bands (*above 25 cycles, the rate of polymerization errors - substitutions and stutter bands - increases*), or increased to produce additional amplification product, depending on need.

Once you are convinced that you have the best looking product you can get, and you think that you have a scorable amplification product, you need to test whether or not the SSR is polymorphic, and if it is, whether or not there are any null alleles within the population.

To test the selected primers for polymorphism, use a population sample from the species of interest. For example, extract DNA from ca. 20 unrelated and randomly chosen individuals using DNAzol, a modified salt extraction (e.g., Gentra's PureGene kit), or the rapid Chelex extraction method (see Appendix D). There are two schools of thought on how to proceed from here. First, you could use cold primers (you are still deciding whether or not the primers are worth using, so don't waste your money ordering labelled primers until you've confirmed that they are polymorphic and there are no null alleles!) to PCR 5-10 individuals,

using 2 μ l of DNA from each individual as a DNA template. As a control, take 2 μ l of DNA extracted from the individual used for the cloning procedure (diluted to ~ 10 ng/ μ l) and 2 μ l of DNA from the miniprep (also diluted to ~ 10 ng/ μ l). Amplify these individuals in separate tubes and run them side-by-side on a gel to look for size variation among the amplification products (make sure to run the original individual on the same gel for a size standard). Alternatively, you can mix a small volume of extracted DNA from ~ 10 individuals (don't get greedy here! — I have heard that this may not work for more than about 10 individuals) and PCR a combined sample template in a single tube. I would suggest checking this technique at least a couple of times to ensure that the combined sample doesn't produce any product not seen in at least one of the constituent samples when PCR'ed individually.

In either case, if you get a clean product of the same size as the original individual from all 10 unrelated individuals, chances are that the marker is monomorphic (or at least the level of variability is so low as to be uninteresting from a population genetics standpoint), but if you get a number of different-sized bands in the first case, or a smear in the second, the locus is likely variable. By amplifying DNA from ~ 25 -30 individuals, you can estimate the frequency of at least the most common allele, and the size range of the alleles is likely close to the largest and smallest amplified fragment observed in these samples.

It is absolutely necessary that each allele be capable of being unambiguously scored before investing in labeled primers (which are expensive, and useless if you cannot score that locus). You must be able to clearly identify a band with much higher intensity, and to differentiate it from those corresponding to the stutter bands produced by polymerase slippage during the PCR reaction. The principal factors affecting the number and the intensity of stutter bands are the Mg^{+} concentration and the number of cycles of amplification. Depending on your facility and the activities of your colleagues, it may be easiest for you to use radioactivity

(^{32}P), chemiluminescent 5' labeled primers (via BM's Genius 8 kit), or fluorescently tagged primers for scoring on the ABI or LiCor automated sequencers. Your method of scoring allelic variation will vary widely depending on which of these methods you use, and because the method is most likely to be dictated by the availability of expertise in one of these techniques in your department, I will assume that you have access to a protocol for your technique of choice. If you do not currently have access to any of these options, the Genius 8 kit is likely the most reasonable choice.

There are, of course, about as many ways to score the loci as there are to develop them, but one of the simplest ways is to a variety of the high resolution agarose preparations on the market. These products include FMC NuSieve 3:1 (Amresco also makes a 3:1 agarose), Metaphor and Amresco SFR agarose. Synergel (made by Diversified Biotech) is another product that helps to resolve small fragments, and after playing around with all of these products, I think that Synergel is a worthwhile investment. I have found that a 1:1 mixture of Synergel with the high resolution agaroses listed above is not only cheaper, but also tends to give the best fragment resolution in the minimum amount of running time (compared to running a pure agarose gel of the same total concentration). The table below gives a very general guideline to follow for gel concentrations to resolve various product sizes (assuming 1X TBE buffer):

% Agarose	% Synergel	Optimal separation range
1.5	1.5	300 - 1000 bp
2.0	2.0	100 - 500 bp
2.5	2.5	50 - 250 bp
3.0	3.0	10 - 100 bp

Depending on the fragment size that you are trying to resolve, and the gel concentration

that you intend to use, your loading buffers will migrate differentially. Some manufacturers (e.g., FMC) include that information with their agarose, others do not. You can always experiment a little, but in general, the higher the percentage agarose, the smaller fragment size the dye will represent. For example, FMC provides the following information for the approximate migration of dyes in relation to DNA fragment size for their NuSieve agarose:

1X TAE Buffer		% Agarose	1X TBE Buffer	
Xylene Cyanol	Bromophenol Blue		Xylene Cyanol	Bromophenol Blue
950 bp	130 bp	2.5	700 bp	70 bp
650 bp	80 bp	3.0	500 bp	40 bp
350 bp	40 bp	4.0	250 bp	20 bp
200 bp	30 bp	5.0	140 bp	8 bp
120 bp	20 bp	6.0	90 bp	4 bp

Given the above table, this seems like a good place to discuss TAE versus TBE buffers. In general TBE buffers provide better resolution for small fragments and (unlike TAE) can be re-used a couple of times within the day before resolution begins to suffer. Given that, why would anyone use TAE? Well, TAE is much more stable at room temperature (TBE eventually starts to precipitate no matter how you treat it), and can be mixed up at 50X and stored for years without any noticeable difference. Also, acetic acid does not tend to interfere with any of subsequent techniques that boric acid does, and is therefore the buffer of choice for any gels from which DNA is to be recovered.

While playing around with the concentrations and figuring out the appropriate mixture of agarose and the concentrations to use, you'll also want to figure out the appropriate running

conditions of the gels. I usually run my 10cm long x 15 cm wide gels at 100V (~50mA) for 1 hour. You can start with those settings and optimize the running conditions from there.

Another alternative is to use a 6% native acrylamide gel. The method of scoring loci will likely depend on the expertise of people in your department and the equipment available to you, but I will provide a rough outline on the technique for running an acrylamide gel for those who lack a better source. A good place to start is to borrow or purchase a small vertical electrophoresis gel rig, such as BioRad's (1-800-4BIORAD) various Protean systems. Some are mini rigs capable of running in as little as 45 minutes, but there are also full-sized rigs capable of running 12 gels simultaneously.

For an acrylamide gel system, you'll have two similarly-sized glass plates, one with a notch (called "eared" because of the extensions on either side of the notch) and one without. You will want one of the plates to be slippery (so you can remove it later) and one to be not sticky (so the gel remains on that plate). Personally, I prefer the slippery surface to be on the eared plate, but that is a matter of preference. Regardless of which way you prefer to do it, ALWAYS use the same method to prevent confusion and end up with two slippery surfaces. To make one plate surface slippery, you simply need to wipe it down with Rain-X™ (you can buy it for about \$3 at Napa Auto Parts, and just follow the directions on the bottle). Once treated with Rain-X, simply place the spacers between the plates, and clamp them with Acco Binder Clips to hold the plates in place while taping them together. You can buy all sorts of fancy tapes (many costing \$10-15 or more per roll), but I have found that a simple roll of wide masking tape from Home Depot (for about \$1.50) works just as well, and doesn't tend to leave the sticky residue that the fancy "residue free" tapes always seem to leave (at least in my experience). Tape the plates by carefully placing one half of the tape on the backside of the 'bottom' edge of the clipped glass plates -- **make sure that there is a perfectly smooth contact, and that no air bubbles or folds form** (if you don't

get a perfect seal, take it all off and start again!). Leave yourself about 2" of tape on each end of the glass, and rub your finger along the tape to make sure that it makes contact all the way along. Carefully fold the tape over and run your finger along the bottom of the 2 glass plates to make sure that you get a perfect seal on the bottoms, while being careful not to contact the 'overhanging' tape. Fold the tape over again onto the top side of the plate, and rub your finger along it to ensure solid contact. **Again don't let the edges of the overhang touch anything** — you should be left with a perfect fit of the tape, and about a 2" 'U-shaped' overhang when you're done. Then press the center of the overhanging tape up onto the glass, making a fold as if wrapping a Christmas present. Once you've pressed it onto the glass along the edges, fold it over and wrap the tape around onto the glass plates. Do the same thing for the sides as well. Once the plates are taped, mix up the acrylamide solution. I use a 60 ml recipe:

Nanopure water	45 ml
10X TBE	6 ml
40% Bis-Acrylamide	9 ml
10% APS	300 μ l
TEMED	60 μ l

Mix the first 3 ingredients first but **do not** add the APS and TEMED until you are ready to pour the gel — these reagents will start the polymerization of the acrylamide, and your gel will form in your beaker rather than between your plates.

If you tilt the plate about 80° away from yourself (about 10° off the table) and onto one corner, then place the lip of the beaker into the space between the plates and pour the liquid acrylamide slowly along one edge of the plates. As the liquid fills the space between the plates, allow the opposite corner to come back down onto the table, and slowly tilt the plates back up towards 90° to the table. Once the area between the plates is filled, slowly insert the comb into the liquid (make sure that some liquid remains around the edges so that the comb is filled) and let the gel stand until the acrylamide polymerizes (at least 60, but usually 90 minutes). Once the gel is polymerized, you

are ready to load it.

You can add 4 μ l of (I use tartrazine to visualize the 20bp fragment front, as well as regular loading blue -- see Miniatis) loading buffer to your PCR product (loading dye with formamide) to make it easier to see and to follow on the gel.

Load 3–4 μ l per well on the gel. Electrophoresis -- the time and voltage will depend on your system and your product size. To start, try 100V (40mA) for about 3 hours. Use the loading dyes to determine when to abort the run — when the tartrazine runs off the end of the gel, the primer front will be right at the edge of the gel, and the run should be usually be stopped at this point, unless you are trying to resolve larger fragments (allow approximately 3 hours per 100 bp that you are trying to resolve at the above settings). The higher the voltage, the shorter the run, but the more the gel will heat up, and the worse the “smiling” of the products will be. If you are interested in running a lot of samples, you would be better served to get a jacketed electrophoresis rig that uses a refrigeration unit to cool the electrophoresis buffer and keep the system cool at high running voltages. The more gels you have in the same system, the more templates you can run simultaneously, and the better the comparability of the migration distance between gels. If you find that you are unable to resolve allelic differences reliably with a native acrylamide gel, you can try a denaturing acrylamide gel (add 30g of Urea to the above recipe — again, **do not** add the TEMED and APS until the solution is ready to be used; the reaction is endothermic, so some heat may need to be added to get the Urea into solution, but use the minimum setting on the hot-plate — Urea comes out of solution at high temperatures!). I will leave the mechanics of denaturing gels to the reader if it proves necessary.

Amplified fragments can vary size by 1 bp due to adenylation, and your migration on a gel can vary by as many as 2–4 bp per run depending on electrophoresis conditions. Therefore, you should always run an internal standard to check the amplicons against both your ladder and this known size product. This ‘control’ fragment should be re-amplified each time under the same

labeled PCR conditions, and run alongside the 'unknown' individuals being scored on each gel. To force adenylation on every PCR fragment, you can simply include a half-hour elongation of the products at the end of the PCR cycles used to amplify the fragments (described above).

Good luck, and happy data collection!

APPENDIX A: OLIGONUCLEOTIDE PROBES USED FOR SCREENING

Low Temperature Probes (Hybridized at ~ 50°C):

1) Estoup & Turgeon's Mixture (they use 45°C for this probe, I use 50°C):

Oligonucleotide	(TC) ₁₀	(TG) ₁₀	(CAC) ₅ CA	CT(CCT) ₅	CT(ATCT) ₆	(TGTA) ₆ TG
T _m	60°C	60°C	62°C	62°C	57°C	57°C

2) Glenn's Mixture (This probe is hybridized at 55°C):

Oligonucleotide	(AAC) ₈	(AAG) ₈	(AAT) ₈	(ATC) ₈	(ACT) ₈
T _m	64°C	64°C	48°C	58°C	58°C

Medium Temperature Probe (Hybridized at ~ 65°C):

1) Glenn's Mixture:

Oligonucleotide	(AC) ₁₂	(AG) ₁₂
T _m	65°C	65°C

High Temperature Probe (Hybridized at ~ 75°C):

1) Glenn's Mixture:

Oligonucleotide	(ACC) ₈	(AGC) ₈	(ACG) ₈	(CCG) ₈	(AGG) ₈
T _m	80°C	71°C	71°C	85°C	80°C

Note 1: T_m can be determined in a number of ways. The most accurate way is to calculate it using the formula:

$$T_m = 81.5 + (16.6 \times \log[Na^+]) + \left(\frac{41 \times (\#G + \#C)}{\text{length}} \right) - \left(\frac{500}{\text{length}} \right)$$

Alternatively, the easiest way to determine T_m is to follow this formula:

$$T_m = [4 \times (\#G + \#C)] + [2 \times (\#A + \#T)]$$

Hybridization should generally be carried out at ~10°C below T_m — much less than 10°C leads

to increased numbers of false positives due to nonspecific hybridization, and much above 10°C leads to decreased numbers of positives due to lack of specific hybridization.

Note 2: Probes with nearly equal T_m and that do not cross-hybridize may be used simultaneously. A genomic library is normally screened with such a set of probes. The mean T_m of the probes should be within about 10°C, and a mix of these probes can then be used to simultaneously screen a library for as wide a range of repeats as possible. I have used all of the above probes with varying degrees of success under different circumstances.

If just starting out, or planning to probe only a single time, a wide mixture of probes (e.g., Estoup & Turgeon's mixture) would likely be best. If you just want to "test" your filters to see whether or not you have real positives, Glenn's medium temperature probe is probably your best bet. To get the most complete screening of your library you should consider using a combination of the above mixtures that covers the entire range of oligonucleotides listed.

Because the dinucleotide repeats are so much more common than tri- or tetranucleotide repeats, you may want to design other primer mixtures to screen only for the longer repeat motifs separately, depending on your particular need. For a list of all di-, tri- and tetranucleotide SSR sequence permutations to aid in probe development, see C. Echt & P. May-Marquardt (1996) *Genome* 40:9-17 Appendix A (now available on the Web at <http://s27w007.pswfs.gov/Data/cetable.html>).

Note 3: $(ATCT)_n = (GATA)_n$ — Self- and cross-complementarity must be avoided in probe mixtures.

5' ATATATATATATATATATAT...3'

3' TATATATATATATATATATAT 5'

This example of the $(AT)_n$ probe shows how self-hybridization occurs. It is important that probes do not self- or cross-hybridize. If designing your own probes, make sure that self- and cross-hybridization are not possible (at least significant hybridization should not occur), or you will waste a lot of time and money sequencing false positives.

APPENDIX E: BACTERIAL MEDIA

Below are listed several of the recipes for the most commonly used bacterial media (if you need more, check the Difco Manual). I list only the ingredients for the liquid media because it is simple to make plating agar from any of the media listed by simply adding 7.5 g of Difco Bactoagar per 500 ml of the completed medium (enough for approximately 20 petri-dishes) just prior to autoclaving (it won't dissolve, so don't even try to mix it -- just leave the pile of agar on the bottom of the bottle and autoclave it). After autoclaving (typically at 120°C for 20 minutes), swirl the cooling agar periodically to ensure homogeneous mixing of the agar throughout the medium. If the agar is to be stored for more than 2 weeks, allow to cool completely and then store at 4°C. If the agar is to be used right away (i.e., within 2 weeks), allow the medium to cool to 50-55°C (if you're using stored agar that has been refrigerated, microwave the agar on medium power for 10 minutes or so -- until the agar becomes completely liquified -- and allow to cool as above -- **CAUTION:** Agar has a tendency to overheat and boil over during microwaving, make sure that you watch the heating agar and swirl the bottle frequently!) and then add:

- 500 µl Tetracycline at 30 mg/ml (resuspend in sterile dH₂O or ethanol -- tetracycline (TC) is not completely soluble in water, so if you chose to suspend your stock in water, make sure you mix the stock well each time before you use it -- store stock at -20°C in tinfoil). The addition of TC is technically optional: ampicillin is the antibiotic necessary in the selection of plasmid-bearing bacteria, not TC. However, TC prevents the growth of "satellite" bacterial strains which, unlike XL1-Blue, do not possess the tetracycline resistance gene. Addition of TC is a good practice because unless your sterile technique is outstanding, satellite colonies will start to grow around transformed clones as they hydrolyze the ampicillin in the plates. TC

will select for only XL1-Blue (transformed or untransformed) growth on your plates and minimize contamination.

- 500 μ l Ampicillin at 50 mg/ml (resuspend in sterile H₂O, store at -20°C in tinfoil).

This antibiotic is **ESSENTIAL** to ensure the selection of only those colonies that has been transformed. Together with the TC you should have only transformed XL1-Blue cells growing on your agar, which is the goal of this exercise.

Gently swirl the contents of the bottle to mix the antibiotics throughout the agar (but do not mix in such a way as to introduce bubbles to the cooling agar). Carefully (to avoid introducing bubbles) but quickly (to avoid hardening of the agar) pour the cooling agar+antibiotics into sterile petri dishes and remove any bubbles that you see with a sterile pipette. Allow the plates to cool at room temperature for at least 20 minutes (up to about an hour). Seal plates in a ziplock bag and store upside down at 4°C for no more than 2-3 weeks (after which the antibiotics have likely gone off, and you have no idea exactly what is growing on your plates).

When you are ready to use the plates, allow them to “dry” by placing them upside down (place the plate on the edge of the lip to keep the plates slightly open and allow condensation to escape) in a 37°C incubator for at least 1-2 hrs prior to use (but no more than 4-5). If the plates are to be used for blue/white selection of transformed bacteria, they must also contain Xgal (5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside) and IPTG (isopropyl- β -D-thio-galactopyranoside) which need to be manually added to the plates prior to plating the transformed bacteria. Close to the end of the “drying” period, make up a stock solution of 40 mg/ml of Xgal (in 100% dimethyl formamide) and a 25 mg/ml stock solution of IPTG (in dH₂O), and store that in the fridge or on ice. Xgal should be prepared immediately prior to being plated if possible, and if you plan to store excess Xgal stock, it must be kept on **dry ice**

anytime it is not in the freezer. should be kept on dry ice throughout — the stock can be stored for short periods of time at -20°C or longer periods at -80°C , but it is most effective when prepared fresh. IPTG is more stable and ready-to-use stock (25 mg/ml) can be simply be stored at -20°C . These reagents need to be plated onto the agar in exactly the same manner as the cells, using a sterile “hockey stick” and turning the dish while spreading. Using the sterile “hockey stick,” spread 25 μl of the Xgal onto each agar plate to be used (it is essential that the Xgal be spread first to avoid precipitation of the reagent and failure of the blue/white selection). Once completed, spread 25 μl of the IPTG stock onto each plate. The plates are now ready for the transformed bacteria to be plated. If you have managed to keep the Xgal on dry ice for the entire time, it can be stored for a couple of weeks at -80°C , but if the reagent warms, or is stored for extended periods of time in solution, the efficiency of blue/white selection rapidly decreases.

I am told by several people that it is possible to simply add:

- 1 ml IPTG at 30 mg/ml (resuspend in sterile H_2O ; store at -20°C)
- 1 ml Xgal at 50 mg/ml (resuspend in dimethyl formamide, stored at -80°C for relatively short periods of time) to the cooling agar at the same time as you add the antibiotics. Many people have told me that this protocol works well for them, but it always seemed very faint and unreliable to me. Blue/white selection does not work well if the Xgal is not fresh or if the reagent warms up very often. I guess that the addition of the Xgal to melted agar simply warms it too much to get really good color selection. I get better results by plating the IPTG and Xgal individually on the plates as described above, but you could try this “quick ‘n dirty” method first if you have the reagents to spare — it will certainly save you time and energy, but you use a lot more reagent and the color selection isn’t nearly as obvious.

All solutions which are not autoclaved should be filter sterilized prior to addition to other media, or use in culture, in order to minimize contamination.

LB BACTERIAL MEDIUM:

To prepare 500 ml of Medium, mix:

Difco Bacto-tryptone	5 g
Difco Bacto-yeast extract	2.5 g
NaCl	5 g

- Add sterile H₂O up to 500 ml
- pH to 7.5 (this usually requires adding about 4 drops of NaOH)
- Autoclave and store at 4°C.

2xYT BACTERIAL MEDIUM:

To prepare 500 ml of Medium, mix:

Difco Bacto-tryptone 8 g
 Difco Bacto-yeast extract 5 g
 NaCl 5 g

- Add sterile H₂O up to 500 ml
- pH to 7.5 (this usually requires adding about 4 drops of NaOH)
- Autoclave and store at 4°C.

SOC BACTERIAL MEDIUM:

To prepare 500 ml of Medium, mix::

Difco Bactotryptone 10 g
 Difco Yeast extract 2.5 g
 NaCl 0.25 g

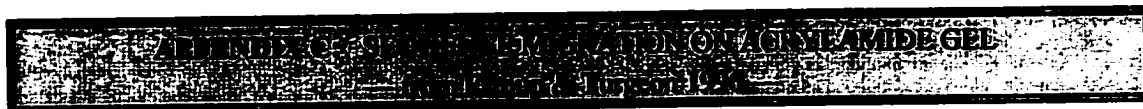
- Add sterile H₂O up to 500 ml
- pH to 7.5 (this usually requires adding about 4 drops of NaOH)
- Autoclave, then add:
 - 1M MgCl₂ 5 ml (Filter sterilized!)
 - 1M MgSO₄ 5 ml (Filter sterilized!)
 - 2M Glucose 0.5 ml (= 1.0 ml of 20% w/v) (Filter sterilized!)
- Filter sterilize entire volume prior to storage at 4°C

NZY⁺ BACTERIAL MEDIUM:

To prepare 500 ml of Medium, mix::

NZ Amine (casein hydrolysate) 5 g
 Difco Yeast extract 2.5 g
 NaCl 2.5 g

- Add sterile H₂O up to 500 ml
- pH to 7.5 (this usually requires adding about 4 drops of NaOH)
- Autoclave, then add:
 - 1M MgCl₂ 6.25 ml (Filter sterilized!)
 - 1M MgSO₄ 6.25 ml (Filter sterilized!)
 - 2M Glucose 5 ml (= 10 ml of 20% w/v) (Filter sterilized!)
- Filter sterilize entire volume again prior to storage at 4°C



PREPARING THE ACRYLAMIDE SOLUTION

!! CAUTION: Acrylamide is a potent NEUROTOXIN prior to polymerization: always handle with care, and wear a dust mask !!

For one gel:

Strongly shake under running hot water until the solution becomes clear; then keep shaking.

POURING A GEL

Note: there are several alternative ways to pour a gel: the proposed method is generally quickly mastered and has provided reliable results in our lab. However, you may want to select a method that is already familiar and/or routinely used by your colleagues (they will be able to help you out!)

Using ethanol, clean the interior side of two glass plates. Let the ethanol evaporate for 1 to 2 minutes. Under a fumehood, apply a coat of silicone - **SigmaCote** (very toxic)- on the interior side of the small glass plate. Let dry for 5 minutes under the fumehood.

[Added Note: SigmaCote is indeed very toxic but need not be used for this purpose. A product commonly sold in auto-parts shops under the product name of Rain-X also works very well, and is basically nontoxic]

Put the large glass plate on a stable supporting device, interior side up. Put spacers on each side and along the bottom of the plate (*use a bottom spacer that is longer than the width of the glass plates so that you may easily remove it after polymerization*). Put the small glass plate on the large one, interior sides facing each other, and tighten both plates together with strong paper clips. Make sure that the bottom spacer fits well and tight with the lateral spacers before your tighten the plates together.

Put the acrylamide mix on ice, keep shaking. The acrylamide polymerises more quickly at

higher temperature: cooling the acrylamide mix will prevent it from polymerising before you are done pouring it. However, do not overcool it, for crystals may form and reduce the homogeneity of the gel.

Immediately prior to pouring the gel, add 420 μ l Ammonium Persulfate 10% + 70 μ l Temed. Mix manually during a few seconds.

Hold plates tilted (but nearer to the vertical), and insert a spatula to slightly separate the glass plates at the top. Slowly pour the acrylamide mix, starting from one side. The mix must flow down as a narrow band along the side and all the way down to the bottom. Once the mix is down and still flowing along one side of the plates, let the gel fill the space between the plates as a front coming back up. The acrylamide solution must be poured continuously, with a relatively slow and regular flow. You must finish pouring the acrylamide mix before it starts polymerising. No air bubbles must remain in the gel. If any bubbles are formed while pouring, finish pouring the gel, then hold the plates vertically to allow bubbles to flow back up. If bubbles remains, tap on the gel with a hard plastic block to help them move up.

As soon as the gel is poured, remove the spatula, put the plates down at ca. 15-20°. Saturate the upper slit with acrylamide solution and insert two combs, flat side in. The front zone formed by the flat side of the combs must be perfectly linear. Once again, the formation of air bubbles must be avoided. Add five large paper clips at the top of the gel to tighten the top of the small glass plate against the large plate.

Let the gel polymerise, at ca. 15-20° of inclination, for about 30 minutes. Remove all clips, rinse the top of the gel with running tap water to remove the excess of acrylamide. Delicately remove the bottom spacer and the combs. Do not force the combs out. Combs must come out progressively and not all at once.

Rinse with water and dry the plates before putting them vertically in the sequencing cuvette,

small plate in. Secure tightly against the cuvette heating plate. Add TBE 1X Buffer in the upper reservoir of the cuvette. Make sure that there are no leaks. Fill the bottom reservoir with TBE 1X Buffer. Using a syringe filled with TBE 1X, chase away the air bubbles remaining between the plates (especially at the top, where urea and acrylamide filaments may remain, but also at the bottom). Insert the combs, teeth down, with the tips barely penetrating the gel front. Secure them with small paper clips to avoid accidentally displacing them while loading samples.

LOADING SAMPLES

Load formamide blue every other well to make sure that there are no leaks from one well to the next. If some wells leak, they may be excluded or a clip may be used to keep the plates tight (immediately after loading that zone and until after a few minutes of the migration).

Start a pre-run to heat the gel up to 48 - 50 °C (a small adhesive thermometer is usually stuck on the exterior of the large plate), i.e. for 30 to 40 minutes. Use the following conditions:

1800 - 2000 V 50 - 70 mA 100 W
--

Put the sequencing tubes at 90°C for 4 minutes (*DNA denaturation prior to loading*) and keep on ice until loading (This denaturation step must also be performed prior to loading PCR product.

Meanwhile (just prior to loading), chase away the urea from the wells, using a syringe filled with TBE 1X.

Load 3 to 4 µl per well, respecting the order G, A, T, C for each sequence. Pre-run under the following conditions until the gel temperature is back to 48 - 49°C:

1800 V 50 mA 80 - 90 W
--

When a temperature of 48 - 49°C has been reached (ca. 15 to 20 minutes), modify the running conditions to:

1500 - 1800 V

30 - 50 mA

60 - 70 W

! 48-49°C is the optimal temperature range. At a lower temperature, there is a risk of renaturation while DNA denaturation is possible at a higher temperature. Power, voltage and amperage source values must be adjusted so that this optimal temperature range is maintained!

At the end of the migration, take the plates off the cuvette and lay them down horizontally, small plate up. Pull off the combs. Let cool down for a few minutes to ease the gel detachment. Gently and progressively lift up the small plate, starting at the bottom. The silicone eases the detachment of the gel from the small plate, leaving the gel stuck to the large plate. Lay a 3M Whatman paper on the gel for 1 to 2 minutes (applying a uniformly distributed pressure with the small glass plate will help transferring the gel from the large plate onto the paper sheet), then gently lift up the paper sheet onto which the gel must stay. Cover the gel with a piece of Saran Wrap (cellophane). Cut the Whatman paper and the Saran Wrap around the gel, leaving ca. 1 cm around. Vacuum dry for 2 hours on a gel dryer at 85-90°C (gel side up). Expose the gel to a Kodak X-OMAT AR film for 1-3 days, depending on the intensity of the signal. Develop in the dark room.

Note: I have included the protocol above and the following one (Appendix D – Chelex DNA extraction) directly from the microsatellite manual written by Estoup & Turgeon, and have not modified or even tried them myself. I have include these protocols only for the sake of completeness. The original protocol is described and discussed in detail in ESTOUP A., Largiader C. R., Perrot E., Chourrout D. (1996), Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes, *Molecular Marine*

Biology and Biotechnology, 5(3): 295-298. If you decide to use either of these protocols, please cite the **original authors** rather than this manual.

APPENDIX D: RAPID DNA EXTRACTION TECHNIQUES

Chelex extraction is extremely quick and easy (DNA may be extracted from several hundreds of individuals per day) and not costly, but the trade off is that it may occasionally create problems for some loci/tissues, and DNA storage is limited to a year or so before degradation becomes a major issue. If problems occur, you should return to a more traditional extraction protocol (such as the PCI + EtOH protocol I outlined above), although this protocol is much longer and somewhat daunting for DNA extraction from a large number of individuals: nevertheless, it yields enough good quality double-stranded DNA. I prefer to use the commercially-available products sold under the name DNA-zol (MRC Reagents: <http://www.mrcgene.com>) or PureGene (a modified salt extraction kit offered for sale by Gentra -- <http://www.gentra.com/contact.htm>, which gives great yields of very high quality DNA) both of which allow high quality genomic DNA to be extracted in about 30 minutes without any of the drawbacks of the Chelex method. Personally, I found that the Gentra kit gave consistently better results, but it is a little more expensive than the DNA-zol. A more expensive, but even better performing, product has recently been released by Amersham Biosciences (<http://www.amershambiosciences.com/>) under the name Nucleon PhytoPure. For problematic marine invertebrates, this kit is the best we've found for extracting high quality DNA. If you are interested in trying these products any of the companies ought to provide you with a free sample to compare and the time saved may prove well worth the additional cost. I will not provide instructions on how to use these kits, because they are provided by the company, and I simply follow their directions.

Note: I have included the protocol below and the previous one (Appendix C -- sequence migration on acrylamide) **directly from the microsatellite manual written by Estoup & Turgeon**, and have not modified or even tried them myself. I have include these protocols only for the sake of completeness. The original

protocol is described and discussed in detail in **ESTOUP A., Largiader C. R., Perrot E., Chourrout D. (1996), Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes, *Molecular Marine Biology and Biotechnology*, 5(3): 295-298.** If you decide to use either of these protocols, please cite the original authors rather than this manual.

ESTOUP & TURGEON'S CHELEX PROTOCOL:

- Put a small piece of tissue in a 1,5 or 2 ml Eppendorf tube.
(*It is sometimes -but rarely- necessary to grind the tissue to improve digestion efficiency*)
- Add 500 µl of Chelex 10 % (Biorad, Chelex resin 100) solution. Pipette the Chelex solution while at 60°C and under agitation (*Cut 0,25 cm off the extremity of a 1000 µl pipette; pipette directly from the Chelex solution while it is being agitated*).
- Add 15 µl of Proteinase K at 10 mg/ml
- Incubate 1 hour at 55°C in a shaking/rocking incubator
- Incubate 2x15 minutes (quick manual shaking between the two periods) at 100°C

Store tubes with Chelex resin balls at 4°C or -20°C

PCR: Prior to using Chelex-extracted DNA as a PCR template, vortex and quickly spin at 10 000g (5 seconds) in order to pellet the Chelex balls. For a total PCR reaction volume of 12.5 µl, pipette 2 µl of supernatant from between the surface and the aqueous-Chelex interface (the lower Chelex phase sometimes contains lipid impurities).

APPENDIX: CALENDAR

The following calendar is based on how long it takes me (after a couple of years of practice) to accomplish each task in this microsatellite isolation procedure. Keeping up with this calendar requires some practical experience with molecular techniques in general, and with cloning procedures in particular, but don't be surprised if you need to repeat some of the steps and end up taking significantly longer than predicted in this outline. The tests suggested to accompany each task are time-consuming, but they provide a day-by-day verification of the results of each step — if you decide not to perform the tests to save time, and the procedure does not work the first time through, you'll have no idea where you went wrong, or how to remedy the situation.

Note: Allow a few additional days for the preparation of most buffers, solutions and bacterial media at the beginning of the procedure, and, if necessary, for replenishing your stock of these supplies during the procedure.

Day 1

- ① Extract DNA (PCI extraction -- Step B)

Test: Estimate the extracted DNA concentration. Run crude extracted DNA on a gel to ensure low 350-850 bp smear density.

Day 2

- ① Label oligonucleotide probes with digoxigenin (DIG labelling -- Step A)
- ② Start overnight digestions with *SAU 3AI* and *MboI* restriction enzymes (Step C)

Test: Stratalink probes to a small filter and run detection protocol to ensure probes were efficiently labeled (once in a while the labeling just doesn't work).

Day 3

- ① Run out digested genomic DNA on a gel and recover 350-850 bp fragments (Step D)

Test: Estimate digestion efficiency and 350-850 bp smear density from the overnight digestion.

Day 4

Test: Check recovered fragment size

- ① Prepare agar and media for growing and plating bacteria (requires autoclaving -- Appendix B).
- ② Ligate size-selected DNA into cloning vector (Step E)

Day 5

Test: Run out ligations on a mini-gel along with positive and negative controls to ensure high ligation efficiency prior to transformation.

- ① Transform competent cells and spread them onto LB plates -- grow overnight (Step F)

Day 6

Test: Estimate both the average size of inserted fragment (boiling+PCR) and transformed cell density. Re-transform? Re-ligate?

- ① Perform additional spreads if necessary

Days 7 & 8

- ① Replate white colonies to new LB plates at an approximate density of 250 colonies per plate (Step G)

-At this point, you should have approximately 6 Petri dishes with the expected density of white colonies. Continue with the transformations until you have enough white colonies to fill at least 6 transfer plates before beginning to replate the colonies onto transfer plates-

Day 9

- ① Lift the transferred colonies onto 2 replicate membranes (Step H)

Test: Start overnight hybridization on test membrane(s) to ensure probing works and determine the approximate proportion of positive clones per plate in your critter. Test membrane(s) should have a couple known microsatellite spots (unlabeled oligo linked onto the filter) added as a positive control for the probing procedure.

Day 10

Test: Analyse results of test-membrane screening, and depending on the number of positive clones detected on the test-membrane, determine whether or not to increase the number of replated dishes and membranes.

- ① Start screening the remaining membranes (overnight hybridization at appropriate temperature with selected probe) (Step I)

Day 11

- ① Colorimetric detection of hybridized filters (Step J) — Locate and mark positive clones for transfer to replicate “positives” plates.
- ② Strip filters and rehybridize with a second probe (if desired or required).
- ③ Repeat until you have at least 30 (and preferably 50 while at this stage to ensure you never have to go back and repeat this part of the procedure) good positive clones.

Days 11-15? (depending on the success rate and number of probes used)

- ① Repeated probing of membranes and transfer of positives to positives plates.
- At this point, one should have, for each species of interest, approximately 30-50 positive clones.-*

Test: Estimate the insert size of selected clones (via PCR) and discard those with inserts smaller than 400 or larger than 1000 bp (Step K)

- At this point, one should have, for each species of interest, approximately 25-30 positive clones possessing 400-1000 bp insert (ideal insert size: 400 to 700 bp) -

Day 16

- ① Grow small cultures in LB medium overnight for permanent glycerol stocks (in duplicate) and store at -80°C.

Day 17

- ① Grow 5 ml cultures of positive clones overnight in 2xYT or LB medium for plasmid isolation (Step L)

Day 18

- ① Isolate plasmids from positive clones with Miniprep kit (e.g., Qiagen MiniSpin) (Step L)
- ② Start sequencing inserts (Step M – hopefully you’ll have an automated sequencing facility to do this for you!)

Day 25 (after getting your sequences back)

① Determine which of your putative positive clones do actually contain microsatellites (hopefully they all do), and whether the forward (M13(-21)) sequence was sufficient to develop primers (did the sequencing run read all the way through the repeat?) or whether you need to send the sequences back for a reverse sequencing run (Step N). Unfortunately the latter is usually the case (unless you are fortunate enough to have a LiCor IR² machine using SBS chemistry). You can save a week by ordering both the forward and reverse sequences at the same time, but if your sequence a putative positive clone without a repeat, you just wasted \$30 for no reason....

② If there is sufficient flanking sequence on either side of the repeat, use a program like Oligo or Primer3 (<http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) to design ~20mer primers unique (hopefully) to the microsatellite, and order those primers (from somewhere like Gibco, Operon, or The Great American Gene Company).

Day 27 (after your primers arrive)

① Run a PCR reaction on both the single individual from which the microsatellite was isolated and a mixture of about 15-25 other individuals (Step O)

Test: Run a gel of the PCR products, and examine the product from the original individual to ensure that the amplified DNA is clean (no stutter or phantom banding) and of the exact size of the sequenced microsatellite. If not, try to optimize the PCR until you can produce a single clean band of the expected size. Once you've accomplished that, examine the mixed individual PCR to see whether you get a single band of the same size (suggesting that the microsatellite is monomorphic, or at least other allele frequencies are very low), or a smear (indicating polymorphism and giving a rough idea of the size range of alleles). Select only those primers that produce both a clean original individual product and a smeared mixed individual product to screen for levels of polymorphism and number of alleles per locus.

Day 30

① Select the 5-6 loci with the best product quality and the appropriate levels of polymorphism to start collecting your data (You're on your own from here..)

Good luck!

APPENDIX 3: ALPHABETIC LIST OF SUPPLIES

100bp DNA ladder (Gibco)

123bp DNA ladder (Gibco)

1kb DNA ladder (Gibco)

Agarose

Alkaline phosphatase, Shrimp (SAP) 10% (Boehringer Mannheim)

Ampicillin (store at 4°C resuspended in sterile dH₂O, then store at -20°C)

Bactoagar (Difco through Fisher)

Bactotryptone (Difco through Fisher)

Bacto-yeast extract (Difco through Fisher)

BSA 1mg/ml (Bovine Serum Albumin)

Chelex 10 % (Biorad, resin Chelex 100) — optional

Chloroform

DIG Luminescent Nucleic Acid Detection kit, Genius 7 (Boehringer Mannheim)

DIG Colorimetric Nucleic Acid Detection kit, Genius 3 (Boehringer Mannheim)

DIG Oligonucleotide 3' End Labeling kit, Genius 5 (Boehringer Mannheim)

DIG Oligonucleotide 3' End Tailing kit, Genius 6 (Boehringer Mannheim)

DIG Oligonucleotide 5' End Labeling kit, Genius 8 (Boehringer Mannheim)

DIG Luminescent Detection kit for sequencing (Boehringer Mannheim)

dNTP (A,T,C,G) 2.5 mM per nucleotide

DNAzol (MRC, Inc) — optional

Epicurian Coli super- or ultracompetent cells (+ β-mercapto cofactor) (Stratagene)

Ethidium Bromide

EDTA

Ethanol 70% Store at -20°C

Ethanol 100% Store at -20 °C

Glycerol

Glycogen

Glucose

IPTG 30 mg/ml (resuspended in sterile H₂O, store at -20°C)

KOAc (Potassium acetate) 3M - pH 5.2

LiCl (Lithium chloride) 4M

Maleic acid 0.1M

MgCl₂

NaOAc (Sodium acetate) 20 mM and 3M

NaCl

N-Lauroylsarkosine

NaOH

Nylon membranes (Micron Separations, Inc. through Fisher)

Nitrogen, liquid

Petri dishes 100x15 mm, sterile

Phenol

pBluescript phagemid vector (Stratagene)

Primer M13 universal reverse 10 µM

Primer M13 universal forward 10 µM

Proteinase K (Gibco)

SAU 3AI

SDS

SigmaCote — this highly toxic product can be replaced with **Rain-X** from most automotive parts shops.

SSC 10X

T4 DNA Ligase (Gibco)

TAE 50 X

Taq, AmpliTaq DNA Polymerase (Perkin Elmer)

TBE 10 X

Temed

Tetracycline (resuspended in sterile H₂O or ethanol, stored at -20°C)

Toothpicks (sterile, approximately 1500 - 2000)

Tris 50 mM pH 8

Tris-HCl 0,1 M pH 9.5

Tris-HCl 0,5 M pH 7.5

Tris-HCl 40 mM

Urea

Xgal 40 mg/ml (store dry at -20°C until use, resuspended in 100% dimethylformamide)

APPENDIX C: SOLUTIONS AND BUFFERS

Although many of these solutions and buffers are already included in the text, I thought that it would be nice to have a list of the recipes in one place, so this appendix should list all the buffer recipes that you'll need for the procedure.

A. LABELING OLIGONUCLEOTIDE PROBE WITH DIGOXIGENIN

Solution STOP: Mix 400 μ l EDTA 0.2M + 2 μ l glycogen

B. DNA EXTRACTION for CLONING

Modified 2X CTAB Lysis Buffer:

- ① Combine (but do not stir or shake):
 - 2 g CTAB
 - 2 g SDS
 - 10 ml 1M Tris, pH 8.0
 - 35 ml 4M NaCl
 - 4 ml 0.5M EDTA, pH 8.0
 - Glass-distilled H₂O to 100 ml Final Volume (F.V.)
- ② Autoclave and allow to cool (the detergents will go into solution when heated)
- ③ Add 0.2 ml Filter Sterilized (F.S.) β -mercaptoethanol.

(25:24:1) PCI — Phenol:Chloroform:Isoamyl Alcohol

- ① Combine:
 - 25 ml **Equilibrated Phenol** (*Phenol can be purchased equilibrated, or the protocol for equilibrating phenol can be located in Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold*

Spring Harbor, NY.)

- 24 ml Chloroform
- 1 ml Isoamyl Alcohol

② Always use fresh HPLC/molecular biology-grade reagents to minimize risk of sample contamination

Standard gel loading buffer for DNA samples:

0.25% Bromophenol Blue, 0.25% Xylene cyanol, 25% Ficoll in glass dH₂O

Light gel loading buffer for DNA samples:

Same as above diluted 1:1 with UV-treated glycerol.

F. TRANSFORMATIONS

Transformation Storage Buffer (TSB)

LB medium (pH 6.1)	47.5 ml
10% PEG 4000	5 g
5% DMSO	2.5 ml
10 mM MgCl ₂ [M.W.=203.3]	0.1016 g
10 mM MgSO ₄ [M.W.=246.47]	0.1232 g

H. LIFT COLONIES ONTO MEMBRANES

All 4 of these solutions should be made up in advance, autoclaved and then stored at room temperature. The solutions can either be made at ready-to-use concentrations (e.g., 2x SSC) or they can be made in concentrated form to be diluted at use (e.g., 20x SSC), depending on available storage space and the number of times you expect to go through this procedure.

Solution 1 (SSC Stock): 20X SSC, pH 7.0

NaCl (175.3 g / l), Sodium citrate (88.2 g / l)

Dilute to 2x SSC (300 mM NaCl , 30 mM Na Citrate) for use.

Solution 2 (SDS Stock): 10% SDS (100 g / l)

Solution 3 (Lysis and Denaturing Sol'n): NaOH 0.4 M (16 g / l)
 NaCl 0.6 M (35 g / l)

Solution 4 (Neutralizing Sol'n): Tris-HCl 0.5 M (60.7 g / l)
 NaCl 1.5 M (87 g / l), pH 7.5

J. SCREENING

Hybridization and Detection solutions (modified from Engler-Blum et al., 1993):

! Note: I have suggested the amount of each buffer to have on hand prior to beginning a probe/detection series, but the recipes are all for ONE LITER (1 l) volumes. Adjust the amounts as necessary for the volume of buffer that you are making !

Rinsing solution:

2x SSC, 0.1% SDS Autoclave, then store at room temperature

Pre-Hybridization Solution (500 ml, pH 7.2):

0.25 mM Na₂HPO₄ (35.49 g / l)
 10% SDS (100 g / l)
 0.5% BM Blocking Reagent (5 g / l) -Shipped with Genius 3 & 7 detection
 kits
 500 mM EDTA (2 ml of 0.5M EDTA stock)
Do NOT Filter, store at 4°C

Hybridization Washing Buffer (1.5 l):

20 mM Na₂HPO₄ (2.84 g / l)
 500 mM EDTA (2 ml of 0.5M EDTA stock)
 1% SDS (10 g / l)
0.2 µm Filter, then store at 4°C

Detection Washing Buffer (1.5 L, pH 8.0):

0.1 M Maleic Acid (C₄H₄O₄) (11.61 g / l)
 3 M NaCl (175.32 g / l)
 0.6% Tween-20 (6 ml / l)

0.2 μ m Filter, then store at 4°C

Detection Blocking Buffer (500 ml, pH 8.0):

0.1 M Maleic Acid	(11.61 g / l)
3 M NaCl	(175.32 g / l)
0.6% Tween-20	(6 ml / l)
0.5% BM Blocking Reagent	(5 g / l) BM's "blocking reagent" is primarily casein hydrolysate.

DO NOT Filter, store at 4°C

NBT/BCIP colorimetric detection buffer (500 ml, pH 9.5):

100 mM Tris	(12.11 g / l)
50 mM MgCl ₂	(11.66 g / l)
100 mM NaCl	(5.85 g / l)

0.2 μ m Filter, then store at 4°C

Most protocols suggest autoclaving all buffers for long-term storage at room temperature, but in practice, the magnesium precipitates when these solutions are autoclaved. Therefore, I filter sterilize everything and store it at 4°C to keep the magnesium in solution. I have also found that filtering these buffers further reduces background and leads to cleaner detections and fewer false positives.

APPENDIX H — ESTIMATED COST

Here is a list of most products (at least the expensive ones) involved for the cloning and sequencing of approximately 5-10 polymorphic microsatellite markers for one species. Several of these products may be stored and used again for subsequent cloning procedures, which greatly reduces the total cost of the procedure. The estimated total cost of approximately \$7000 US represents an estimate of the initial investment to begin isolation of microsatellite markers in a lab with all the equipment (e.g., gel rigs, power supplies, microcentrifuges, etc...) to do molecular biology (if you don't at least have access to any of this equipment, this and any other protocol is likely of little use to you!). This cost drops to as little as \$1500 or so (of which \$900 is the cost of sequencing) for a second species. Thus, in order to make this procedure as inexpensive as possible, and to reduce its labor cost, it is wise to find a group of people who want to clone microsatellite markers for several species at once and spread the initial outlay around among yourselves.

Products/kits	Suggested supplier	Reference	Approx Price ('97 US\$)	Storage
100bp DNA ladder	Gibco BRL	15628-019	74	-20°C
123bp DNA ladder	Gibco BRL	15613-011	72	-20°C
1kb DNA ladder	Gibco BRL	15615-016	102	-20°C
Agarose	Gibco BRL	15510-027	348	Cool & dry
Shrimp Alkaline Phosphatase	Boehringer	1 758 250	62	-20°C
Difco Bacto-Agar, Difco Bacto-tryptone & Difco Bacto-yeast extract	Fisher (Difco)	DF0145-17-0 DF0123-17-3 DF0127-17-9	76 46 64	4°C
T4 DNA Ligase	Gibco BRL	15224-025	136	-20°C
E-Coli XL1-Blue Supercompetent Cells or XL2-Blue Ultracompetent Cells	Stratagene	200236200150	155165	-80°C
<i>MboI</i> and <i>Sau3AI</i> restriction endonucleases	Gibco BRL	15248-016 15229-016	60 66	-20°C
NBT/BCIP Stock Solution	Boehringer	1 681 451	75	4°C

Acrylamide	Sigma	A 9099	108	Room temp
Genius 5 DIG oligonucleotide 3' end labeling kit	Boehringer	1 362 372	325	-20°C
Genius 6 DIG oligonucleotide 3' end tailing kit	Boehringer	1 417 231	325	-20°C
Genius 3 DIG colorimetric nucleic acid detection kit	Boehringer	1 175 041	244	-20°C
Genius 7 DIG chemiluminescent nucleic acid detection kit	Boehringer	1 363 514	416	-20°C
Genius 8 DIG oligonucleotide 5' end labeling kit	Boehringer	1 480 863	297	-20°C
Luminescent detection set for DIG sequencing	Boehringer	1 733 915	199	-20°C
QIAprep Spin Miniprep Kit	Qiagen	27104	57	Room temp
10 mM dNTP Mix	Gibco BRL	18427-013	52	-20°C
MSI Nylon Transfer Membranes	Fisher (MSI)	N00HY08250	78	Room temp
pBluescript KS ⁺ phagemid vector	Stratagene	212207	170	-20°C
Proteinase K	Gibco	25530-015	180	-20°C
Oligonucleotide probes and PCR primers		custom	~ 1100	-20°C
Xgal	Gibco	15520-018	197	-20°C
IPTG	Gibco	15529-019	30	-20°C
Other miscellaneous products (taq, etc.)			~ 1500	
Automated sequencing (@ \$15 per sample x 2 directions = \$30 each)		custom	~ 900	



Appendix 2.

**Increased throughput for fragment analysis on an
ABI 377 automated sequencer using
a membrane comb and STRand software.**

Robert J. Toonen

**Center for Population Biology,
University of California, Davis,
Davis, CA, USA. 95616.**

Voice: (916) 752-2937 Fax: (916) 752-1449

Email: rob@biogeek.ucdavis.edu

and

Shayne Hughes

**Veterinary Genetics Laboratory
University of California, Davis,
Davis, CA, USA. 95616.**

Email: shayne@ucdavis.edu

In press, Biotechniques

Key words: GeneScan, ABI 377, STRand, microsatellite fragment analysis,
membrane comb

Abstract

This manuscript outlines our protocol for using a freely downloadable fragment analysis software package (STRand) together with a 96+4 RapidLoad membrane comb (Gel Company) to increase throughput of samples for fragment analysis on ABI sequencers without costly upgrades from the manufacturer. We outline how using these products allows me to score 90 lanes of sample per gel on our ABI 377XL (64-lane sequencer), saving both time and money in the processing of samples. This protocol is a major modification to those suggested by the manufacturer. Our protocol gives more consistent results that are easier to score than standard protocols, and it reduces reagent costs. Interest in fragment analysis (primarily microsatellites and AFLPs) is steadily increasing among both population ecologists and geneticists, and methods that simultaneously increase sample throughput while reducing costs associated with these analyses by over 50% per gel should prove useful to anyone using an ABI, MJ Basestation or Licor automated sequencer for fragment analysis.

INTRODUCTION

Interest in microsatellites and AFLPs as the genetic marker of choice for population and ecological genetics studies is steadily growing, and the advantages of fragment analysis on automated sequencers has increased demands on available systems concomitantly. Given increased demands on available sequencers, issues of optimization of multiplex reactions, sample throughput and cost have become the focus of many researchers when deciding which detection technique to use for scoring fragment polymorphisms. Upgrades of the ABI 377 DNA Sequencer from Applied Biosystems Inc. (formerly PE Biosystems, Foster City, CA) can increase sample throughput to 96 lanes per gel, but are expensive and often difficult to justify due to the substantial cost of matching peripheral equipment to the 96-well format. This article describes how we use STRand software (4) and RapidLoad membrane combs (3) from the Gel Company (San Francisco, CA) to score up to 90 samples per gel using GENESCAN (PE Biosystems) collection software on an ABI 377 XL (64-lane) DNA sequencer. This method will work with either ABI 373 or 377 sequencers, and provides a cheap alternative to increase the throughput of fragment analysis with only slight modification to existing protocols and without requiring any expensive equipment upgrade.

METHODS AND MATERIALS

PCR and multiplex optimization. Using 5'-end labeled Phosphoramidite (6-FAM, HEX and NED) custom primers from Applied Biosystems, dye-labeled amplification products are generated via polymerase chain reaction (PCR). PCR protocols were modified from the GENESCAN Reference Guide provided by Applied Biosystems (1) to optimize amplification

efficiency for marine crabs. ABI slab gel sequencers read four colored dyes simultaneously, but one dye (ROX) is used for the internal size standard. Hence, three overlapping loci can be scored simultaneously in an individual via multiplex PCR amplification. Where loci with non-overlapping allelic ranges exist within the limitations of the size standard used (Applied Biosystems markets a variety of size standards), two (or more) loci can be labeled with the same dye to increase the number of loci scored in a single PCR reaction with multiple primer-pairs. Because the allelic range in the marine crabs *Petrolisthes cinctipes* and *Cancer magister* is unusually large (up to 190 bp), a maximum of seven loci can be simultaneously amplified in a single multiplex PCR reaction depending on the degree of overlap among the allelic range of amplification products.

Although the primer front can be cut from the gel in STRand or GENESCAN, the presence of unincorporated primer often obscures small amplification products. This requires either removal of unincorporated primer from the PCR products via a clean-up step or optimization of the PCR to minimize the amount of unincorporated primer remaining after amplification. We have chosen the latter strategy to eliminate the cost of PCR clean-up and increase the speed of scoring samples following PCR. By sequentially halving the primer concentration until the amount of unincorporated primer is minimized, amplification products can often be improved, and the template to primer ratios of each locus can be adjusted as necessary to equalize intensity of amplification among loci. Elimination of the majority of the primer front reduces interference with scoring small products (< 100 bp) while saving money on dye-labeled primer use, ROX size standard and PCR clean-up kits, this procedure also reduces the steps involved and reduces background, thereby making the resulting gel file easier to score.

Gel preparation. LongRanger Singel packs (BioWhittaker Molecular Applications

#50691) are cast according to the manufacturer's protocol (2) using a casting comb (Applied Biosystems #401913). Once the gel is completely polymerized, the casting comb is removed, and the well is washed with DI water to remove any remaining bits of polymerized acrylamide. The well is visually inspected for a sharp, clean and straight interface, the gel is held at a 45° angle, and a Kimwipe (Kimberly-Clarke #34120) is then used on one corner to wick all remaining DI water from the well (while avoiding contact with the gel surface). To ensure the well is completely dry prior to loading, the gel is left in a rack upside-down (with the well at the bottom) while preparing samples for loading.

Although unnecessary, we have found that pre-warming the equipment produces more consistent and easily scorable gels. We pre-warm the sequencer by using the gel and buffer from the previous run; the sequencer is simply restarted in PreRun mode during sample preparation to bring the system to temperature in advance of gel loading. Once the sequencer is pre-warmed, the new gel comes to running temperature very quickly after being inserted into the machine, precluding a lengthy pre-run.

Sample preparation. Dye-labeled amplification products are generated in advance as outlined above, and 1 µl of the PCR amplification is combined with 1.1 µl of loading buffer, which consists of 0.16 µl GENESCAN-500 ROX size standard (Applied Biosystems # 401734), 0.68 µl deionized formamide (Sigma #F 9037), 0.26 µl blue dextran/EDTA loading buffer (Applied Biosystems #402055, supplied with GENESCAN-500 ROX standard). This loading mixture is greatly diluted relative to the recommendations of the Applied Biosystems protocol provided with the GENESCAN-500 ROX size standard, and is the lowest amount of size standard per lane we could load with reliable results. At this loading volume, the size standard is detected at approximately 600-1500 relative fluorescence units (rfu). If higher signal strength

is needed, the volume of the ROX standard in the loading buffer can be easily increased. Some experimentation is required to minimize unincorporated primer in the PCR reactions and optimize the template to primer ratios for each locus to produce an amplification product detected at between 1000 and 4000 rfu (as described above). Once accomplished, this minimum loading volume (0.16 μ l per lane) of the ROX size standard works very well, and uses only about $\frac{1}{3}$ the suggested amount of this expensive reagent per gel.

After the PCR and loading buffer are mixed, the PCR plate is sealed, and the samples are denatured at 95°C for 3 min on a thermocycler before being transferred directly onto ice. Samples are kept on ice while transferring 0.9 μ l of the denatured mixture from a 96-well PCR plate into the 96+4 well RapidLoad membrane comb dipping tray (Gel Company #TAM96). The company recommends loading only 0.4 - 0.7 μ l of sample onto each tooth of the comb, but the extra volume is sometimes desirable, and the volume of sample loaded can be easily controlled during dipping (explained below). Despite the increased volume transferred to the dipping tray, between evaporation, denaturing and pipettor error, very little of the denatured samples will be left in the PCR plate after transfer, and reloading is generally not possible in case of sample loss. If multiple loads are frequently necessary, you may wish to increase your sample preparation volume.

Gel loading and running. The denatured samples in the RapidLoad tray are loaded onto the 96+4 membrane comb immediately prior to being loaded onto the gel. Samples are applied to the comb by dipping the comb into the numbered wells on the tray until the sample is absorbed. If the sample wicks up beyond $\frac{3}{4}$ of the height of any tooth, either decrease the time that you allow samples to wick into the comb (we simply withdraw the comb as soon as the samples cover $\sim \frac{3}{4}$ of the tooth), or the amount of the denatured sample transferred to the

dipping tray for future loads. In practice, $\frac{1}{4}$ of the tooth on the 96+4 RapidLoad membrane comb translates to approximately 0.7 μ l of sample, and ensures no mixing of samples due to wicking between lanes.

Without the 96-lane upgrade on the ABI 377XL, the path of the laser is not wide enough to take advantage of the entire 100 lane capacity of the RapidLoad 96+4 membrane comb (#CAM96). The laser detection zone reads only the lanes numbered 4 through 93 on the RapidLoad dipping tray. However, using STRand (described below) it is possible to take advantage of the full read path (maximum of 90 lanes) of the laser on the 377 XL sequencer. The drawback of increasing the number of lanes scored per gel is a reduced number of channels per sample (6), but in comparisons of 128 multiplex PCR reactions run with the method outlined herein to those same samples run using a traditional 64-lane GENESCAN protocol (1), we found not a single difference in allele sizes called for a given sample. Unfortunately, the 96+4 RapidLoad combs do not fit the entire width of the gel as do the standard sharktooth combs; this is problematic because 90 lanes are the absolute maximum width of the data gathering region, and any slight variation in centering the comb in the read region will cause either the first or last lane to be missed on the gel file. Therefore, it is useful to mark the comb position on the plate to ensure the same position in every run, and to run the positive controls in lanes 1 and 90 of each gel, so that a slight shift in the comb position does not result in lost data. With the two positive controls, 88 lanes of data can be reliably collected on each gel.

Although it does not appear to be necessary, we leave the RapidLoad dipping tray on ice throughout sample transfer and comb loading because in our experience it increases band resolution and uniformity slightly. Once the samples are transferred into the dipping tray, the gel from the previous run (used to pre-warm the sequencer, as described above) is removed and the new gel is loaded onto the sequencer. If any additional liquid has appeared in the well while

the plate is inverted, the gel is again tilted at a 45° angle and the liquid is removed as previously described. After confirming a flat baseline with a plate check, the front heating plate is attached and buffer is added to the correct level in the bottom buffer tray only. The upper buffer chamber is attached, but left empty. The gel run (module GS 36D-1200, 36 cm WTR XL Scan, 2.5 hr collection time) is started using a “dummy” sample sheet (we use a 64 sample sheet each time, but the sample sheet information and lane number collected for the gel are irrelevant) and then immediately paused (a No EP Current Error will pause the machine automatically if you fail to do it manually).

A pipettor is used to add approximately 400 - 500 µl of 5% ficoll loading solution (diluted with PCR-quality water from the Gel Company 20% ficoll loading solution #DAL25), to completely fill the well to the point where the solution just begins to bead between the plates. It is particularly important, however, to avoid getting any gel loading solution on the back plate while filling the well. Any liquid on the back plate will cause the comb to stick and make it loading awkward and greatly increase the likelihood of damaging the comb teeth during insertion. Bubbles in the loading solution should be removed with a gel-loading tip or micro-syringe. Once the well is filled with loading solution, we try to minimize the amount of time taken to apply the samples to the comb, insert the comb into the well, and start the electrophoresis run (it should not take more than a minute or two).

After samples have been applied to the membrane comb, it is ready for immediate loading into the loading solution filled well. This is most easily accomplished by holding the comb at a 45° angle to the back plate and allowing the teeth to first penetrate the bead of loading solution left between the plates (as described above). As the teeth penetrate the bead and then come into contact with the back plate, the angle of the comb can be increased towards 90° to allow the teeth to easily slide down between the plates until they rest just above or on the gel

surface. Over-insertion of the comb can damage the gel matrix and lead to irregular spacing between lanes or “bent” lanes that are much more difficult to score. Although the dilute loading solution can have extra blue dextran dye added to increase the resolution of the gel interface, if the gel was poured using the casting comb recommended above (Applied Biosystems #401913), the teeth of the membrane comb will just barely reach the gel surface when the upper edge is flush with the top of the back plate. After only a few loads to familiarize ourselves with the technique, we found this method of loading becomes much faster and easier than traditional methods of loading a sharktooth comb using a multichannel syringe.

Although the company protocol (3) states that samples will not diffuse into the loading solution, we try to start the run as quickly as possible after inserting the comb. That is why we start and then pause the electrophoresis run prior to filling the well and loading the samples onto the comb. Once the comb is inserted, the upper buffer chamber can be quickly (but gently) filled. As soon as the door is closed, the run can be resumed, thereby minimizing the time that samples sit in the comb. After restarting the run, we allow 2 minutes of electrophoresis (at 200W) to transfer the samples into the gel. The run is then paused, and the comb is removed and discarded. We have heard of some people reusing the membrane combs after soaking in DI water to remove any residual sample, but we have not tried this technique. The 5% ficoll loading solution is then gently rinsed from the well with 1X TBE from the upper buffer chamber using a 200 μ l pipettor. It is essential that the loading solution be completely rinsed from the well, but it is also important that the rinsing be relatively gentle to avoid irregularly spaced or wavy gels. The standard method of vigorous rinsing with a 60 cc syringe invariably produced unscorable gels for me. The 5% dilution is much more easily dislodged than the original 20% ficoll solution, so gentle rinsing several times across the width of the well is all that is needed to completely displace the loading solution. Once the well is thoroughly rinsed, the lid to the

upper buffer chamber is put into place, and the run is resumed.

RESULTS

After the gel run is finished, the gel file is transferred to a PC running Windows 98 SE (Microsoft), on which STRand (4) has been installed. The STRand software package is freely downloadable from the University of California, Davis Veterinary Genetics Laboratory web page (<http://www.vgl.ucdavis.edu/STRand> - Note: server is case sensitive), and was installed and configured following the instruction manual (5). The unanalyzed gel file produced by the Genescan collection software is copied from a Mac format zip disk onto a PC using MacDrive for Windows (Mediafour, West Des Moines IA), and the raw gel file is imported into STRand directly. Users can predefine the number of lanes to be found, or let the software determine the number of visible lanes on the gel. The tracked gel is then analyzed using a set of user-defined rules for each locus included in the multiplex reaction. These locus rules are fully customizable for specific applications. The program calls both the size standards and the allele/band sizes for each lane of data and allows the user to examine and alter any of the peak or size designations. The program displays a list of peak heights and alleles called for each locus. A separate window that provides user alerts (for problems such as having more than two alleles called for a locus) is also included along with the electropherogram for each sample in a single window, eliminating the need to switch between windows or programs to complete the analysis. Individual dye colors can be hidden as necessary, and both the X and Y axis of the electropherogram are easily adjustable with sliders for cases where zooming in on the electropherogram can help with scoring.

DISCUSSION

In our experience, results from gels run using the protocol we outline here are generally indistinguishable from those using a traditional sharktooth comb (Figure 1). We believe that the Rapid Load comb offers three primary advantages over the sharktooth comb that are unrelated to data quality. First, all 90 lanes are loaded simultaneously with the Rapid Load comb which not only decreases load times, but also eliminates the need for stagger-loading typically required with multi-channel syringes used with a sharktooth comb. Second, there is no possibility of samples leaking between wells with a Rapid Load comb. Finally, less product is required for running with the Rapid Load combs, which proves to be a significant cost savings over standard protocols using a sharktooth comb.

Although the use of the Rapid Load comb simplifies loading and reduces the cost of reagents per lane on the gel, it is the analysis software, STRand, that allows us to use the full width of the laser path on the gel, and get 88-90 lanes of scorable data from the 64-lane ABI Prism 377XL. STRand was originally designed to address the needs of the UC Davis Veterinary Genetics Laboratory (VGL) while developing its Equine DNA test using fourteen STR loci multiplexed into a single lane. The software available at the time was inefficient for that level of high-throughput genotyping, and had a tedious user interface without any relational database integration. The goal for STRand was to maximize accuracy while minimizing analysis time. Flexibility was also important; STRand is ready for new technologies through the easy addition of any gel image format, any number of lanes, any number of dyes, etc. Currently, ABI 373, 377, 310, 3100, and 3700, MJ BaseStation, and Licor formats are all supported.

STRand performs the full analysis from the original collection file to genotypes as opposed to using two programs such as Applied Biosystems' GeneScan and Genotyper. VGL

found that STRand could cut analysis time by a factor of 4 to 6 because lane tracking is done automatically, sample sheets are pulled from a database, and the analysis window shows the electropherogram along with the other data without the need to switch between windows or programs. This year VGL expects to complete 140,000 DNA types using 13 markers each for a total of almost 1.7 million genotypes all analyzed using STRand. A 96-lane gel with 13 STR loci per lane takes a skilled VGL analyst about 45 minutes to complete. Over the years, STRand may have lost a little of its lead as Applied Biosystems' product literature claims 60 minutes to analyze 96 lanes of 14 markers for GeneScan/Genotyper and 10 minutes using their new GeneMapper software (7). New features on the horizon for STRand should continue to keep it competitive in terms of performance. More attractive to individual research labs is the fact that STRand is licensed under the GNU Public License (www.fsf.org) from The Regents of the University of California which means that it is available to use for free. The source code is also available so that you can modify and improve the program as needed. The STRand homepage is located at <http://www.vgl.ucdavis.edu/STRand> - Note: server is case sensitive from which it is possible to download the software and a manual as well as subscribe to the users email list.

To manage all of the data generated, STRand stores all results in a relational database. The database can be either a Microsoft Access database file or Microsoft SQL Server (other database systems may also work, but have not been tested). Using a relational database server allows STRand users to reliably share the same database across a network so that all researchers have access to the same samples. Keeping the complete data set in a single location also provides easy access for data checking and for collecting statistics like allele frequencies. STRand exports results in a text file format similar to that used by Genotyper or results can be accessed directly from the database.

Although the protocol detailed here is specific to the ABI Prism 377 automated sequencer, it should be readily adaptable to any automated sequencer which STRand supports. In summary, the method outlined here allows users to easily increase STR sample throughput on an automated sequencer without expensive equipment upgrades, and more importantly without sacrificing data quality. The additional cost of the RapidLoad membrane combs over a reusable mylar sharktooth comb is easily offset by the savings on GENESCAN-500 ROX size standard consumed with each gel. Aside from gaining 24 additional lanes of data per gel, this protocol also reduces our cost; overall our cost per gel is reduced by over 50% with this protocol relative to the standard 64-lane GENESCAN protocol we followed originally (1). Together with the freely downloadable STRand fragment analysis software, the added benefits of the ease of sample loading and increased sample throughput of the RapidLoad membrane combs, this protocol should prove useful to anyone using an automated sequencer for fragment analysis.

Acknowledgements

We are grateful to the staff of the Division of Biological Sciences Sequencing Facility, the Veterinary Genetics Laboratory, and the Gel Company for their advice and assistance in developing and troubleshooting this protocol. We thank M. Locke, A. Wilson and R. Grosberg for their assistance and the helpful comments they provided to improve the manuscript. This research was funded in part by a grant from the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, under grant number NA06RG0142, project number R/F-177 through the California Sea Grant College System, and in part by the California State Resources Agency. The views expressed herein are those of the author and do not necessarily reflect the views of NOAA or any of its sub-agencies. The U.S. Government is authorized to reproduce and distribute for governmental purposes.

References:

1. **Applied Biosystems.** 1998. 377 DNA Sequencer: Microsatellite Analysis. AB Technical Document, Foster City, CA.
2. **BioWhittaker Molecular Applications.** 1998. Long Ranger Gel Protocols, 18 pp. BMA, Rockland, ME.
3. **Gel Company.** 1999. RapidLoad Membrane Comb Protocol for Sample Kit KAT96, 2 pp. The Gel Company, San Francisco, CA.
4. **Hughes, S.** 1998. STRand Nucleic Acid Analysis Software. Regents of the University of California, Davis. <http://www.vgl.ucdavis.edu/STRand>
5. **Locke, M., E. Baack and R.J. Toonen.** 2000. The STRand Manual. http://www.vgl.ucdavis.edu/STRand/docs/STRand_Manual.pdf
6. **Velickovic, Z.M., M.B. Velickovic and S.K.G. Grebe.** 1999. Increasing Sample Throughput on a Standard 36-Lane Model 377 DNA Sequencer to 48 or 64 Samples Per Run. *Biotechniques* 27:652-654.
7. **Applied Biosystems.** 2000. ABI Prism GeneMapper High Performance Genotyping Software for Linkage Mapping Applications. AB Technical Document, Foster City, CA.

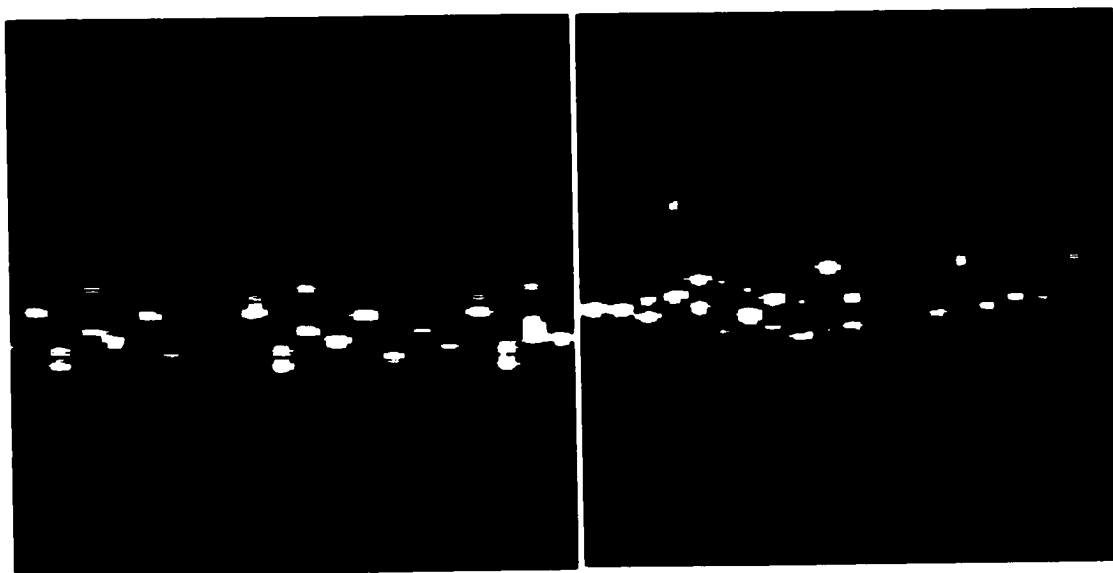


Figure 1. Magnified sections of a typical 64-lane sharktooth comb (left) and 64+4 Rapid Load membrane comb (right). Each lane is a 4-locus multiplex reaction amplified from individual anomuran crabs, *Petrolisthes cinctipes*, and uses the GENESCAN-500 ROX size standard from Applied Biosystems. We have zoomed in on the 200 — 450 bp portion of the center of each gel (lanes 22 to 42) for comparison.