UNIVERSITY OF CALIFORNIA, SAN DIEGO

Evolutionary genetics of the tidepool copepod *Tigriopus californicus*

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Marine Biology

by

Jonathan M. Flowers

Committee in charge:

Professor Ronald Burton, Chair
Professor Lin Chao
Professor Brian Palenik
Professor Vic Vacquier
Professor Chris Wills

2005

The dissertation of Jonathan M. Flowers is approved, and is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2005

DEDICATION

To Mom and Dad,

my brother Jeff, and of course Bono

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It was a cool evening in a dimly lit Thai restaurant in Vancouver, British Columbia where I met Ron Burton. We sat in the corner of a large table filled with evolutionary biologists and discussed nucleotide variation in the D-loop of sea urchins and starfish. No one else seemed interested. Seven years later I look back and wonder what if Ron and I hadn't decided to share a plate of egg noodles? One thing is clear, I would not be the scientist that I am today without Ron. He has taught me much about science and being a scientist. Perhaps most importantly, Ron has taught me that enduring a career in science requires lightheartedness. As Ron often reminds us when we run into disappointing dead-ends, "It could be that the purpose of your life [or your failed experiment] is only to serve as a warning to others". I am forever indebted to Ron for giving me the opportunity to join him at SIO. Ron invested in my future by funding my research and paying my stipend for four years. If I had it to do over again, I would change only two things. I would not have used Doc Frugal *Taq* and I would never, never try to clean the Dremel tool without first consulting the instructions. Thank you Ron.

I am also grateful to my committee members, Dr. Lin Chao, Dr. Chris Wills, Dr. Vic Vacquier, and Dr. Brian Palenik. Each member contributed in countless ways to the completion of this work. I am especially indebted to Lin Chao who challenged me and made me a better scientist. I wish to thank Chris Willett and Scott Harrison who were instrumental in my academic development and also are close personal friends. I also owe much to Erin Cartaya who became a close friend during the final months of my Ph. D. Rosemary Byrne, "23", was a great help in the

lab and also provided friendship. My labmates Chris Ellison and Kristin Gruenthal were most excellent companions during our three years together. Finally, I wish to thank my family including Grandma Christoffel, Grandma Flowers, Jeff, Pam, mom, dad, and of course Melissa. I owe much to my family. I could not have completed this without their loving support.

The text of Appendix A, in full, is a reprint of the material as it appears in Flowers, J. M., S. C. Schroeter, and R. S. Burton. 2002. The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. Evolution. 56:1445-1453. The dissertation author was the primary author. Stephen C. Schroeter provided animals for the study. Ronald S. Burton funded the research in its entirety and supervised the research. Permission to include this manuscript in the dissertation of Jonathan M. Flowers was obtained from Annette Keeler, managing editor of Evolution.

VITA

2005	Ph.D., University of California, San Diego
2004	Lecturer, Principles of Genetics, University of San Diego
1999	M.S. in Zoology, Louisiana State University, Baton Rouge.
1998-1999	Teaching assistant, Principles of Genetics, Louisiana State University, Baton Rouge
1997	B.S. in Biology, Walla Walla College, College Place, Washington.

PUBLICATIONS

Flowers, J.M., S. C. Schroeter, and R. S. Burton. 2002. The recruitment sweepstakes has many winners: Genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. Evolution. 56:1445-1453.

Flowers, J. M. and D. W. Foltz. 2001. Reconciling molecular systematics and traditional taxonomy in a species-rich clade of sea stars (*Leptasterias* subgenus *Hexasterias*). Marine Biology. 139:475-483.

Burton, R. S., E. C. Metz, J. M. Flowers, and C. S. Willett. 2005. Unusual structure of ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal subrepeats. Gene. 344:105-113.

CONTRIBUTED PAPERS

Flowers, J.M. and R.S. Burton. 2004. DNA sequence polymorphism at *Pgi* in *Tigriopus californicus*: Deviations from equilibrium at an allozyme locus. California Population and Evolutionary Genetics Annual Meeting (presented talk)

Flowers, J.M. and R.S. Burton. 2004 Evolution of mitochondrial transcription genes in a species with rapidly evolving mtDNA. Society for the Study of Evolution Annual Meeting (presented talk)

Flowers, J. M. and R.S. Burton. 2002. Expression of the rRNA genes in *Tigriopus californicus*: biased transcription in inter-population hybrids. Society for the Study of Evolution Annual Meeting (presented talk)

Flowers, J. M. 2000. The recruitment sweepstakes has many winners: Genetic evidence from purple sea urchins. California Population and Evolutionary Genetics Annual Meeting (presented talk)

Hrincevich, A. W., Foltz, D.W. and J. M. Flowers*. 1998. Genetic and Morphometric Approaches to Species Identification in a Cryptic Species Complex of Sea Stars (*Leptasterias* Spp.). *(presented talk)

INVITED TALKS

California State, Los Angeles, CEA-CREST symposium May 2003

Department of Biological Sciences, La Sierra University, February 2001

FELLOWSHIPS

2002-2004 California Coastal Environmental Quality Initiative Fellowship $50,000

ABSTRACT OF THE DISSERTATION

Evolutionary genetics of the tidepool copepod *Tigriopus californicus*

by

Jonathan M. Flowers

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2005

Professor Ronald Burton, Chair

This work examines the molecular population genetics of hybrid breakdown in the harpacticoid copepod *Tigriopus californicus*. The aim of this work was to use classical population genetic analysis of natural populations together with molecular genetics to characterize the genetic basis of reduced F_2 hybrid fitness. Each of the chapters of this dissertation makes a unique contribution to our understanding of the genetic consequences of hybridization. In Chapter II, I combine time series data from natural populations with DNA sequence polymorphism data to characterize the extent of population exchange among local *T. californicus* populations. A primary result from this work is that gene flow among local populations is likely higher than previous data suggested. This has implications for our understanding of local population differentiation and the consequences of hybridization between *T. californicus* populations. In Chapters III and IV, I investigated two transcription systems, RNA polymerase I and mitochondrial RNA polymerase. Our interest in these systems primarily comes from

their unusual mode of evolution. We hypothesized that rapid evolution of these transcription systems may be detrimental to the fitness of hybrid organisms. The data presented in Chapter III provide evidence of large-scale alterations in regulation of ribosomal RNA by the RNA polymerase I system in hybrids. Although the consequences of this for hybrid fitness are not clear, our results have significant implications for the molecular mechanisms responsible for the hybrid phenomenon of nucleolar dominance. In Chapter IV, results are presented on population differentiation of mitochondrial RNA polymerase, *mtRPOL*, and a mitochondrial transcription factor, *mtTFB1*. We hypothesized that elevated rates of genome evolution in *T. californicus* mitochondrial DNA would select for accelerated changes in the region of the polymerase responsible for promoter recognition. *T. californicus* populations have diverged substantially at this locus, but we found no evidence of rapid divergence in the region responsible for promoter recognition. In a separate study, I found that the *mtRPOL* and *mtTFB1* are tightly linked. A segregation study of the linkage group marked by these genes found a large viability effect of the *mtRPOL/mtTFB1* locus on hybrid fitness. This effect is attributable to viability effects that are dependent on the cytoplasmic background. Finally, Appendix A presents work on variance in reproductive success in sea urchins that is unrelated to the main body of work. This was the product of research I conducted during the first two years of my dissertation and is included here as a supplement to the work conducted on *T. californicus*.

Chapter I

Introduction

Population demography and the fitness of hybrids

The effect of outbreeding on the fitness of hybrids depends on the complex inter-play of natural selection, population demographics, and the genetic architecture of fitness-related traits. In structured populations, hybridization frequently results in reduced hybrid fitness. This outbreeding depression, or hybrid breakdown, may be observed in the F_1 generation (Templeton 1986, Brown 1991), but more frequently reduced fitness is not observed until the F_2 (Dobzhansky 1948, Templeton 1986). This is generally attributed to the disruption of positive epistatic interactions as a result of recombination in F_1 hybrids. A significant problem in determining the causes of hybrid breakdown is distinguishing between the effects of population history (i.e., phylogeny), demographics (e.g., gene flow/drift), and environment (i.e., gene X environment interactions/local adaptation) on the evolution of multi-locus interactions. Different views concerning the importance of these factors in natural populations have contributed to different perspectives on the nature of adaptation and the role of epistasis in evolution (Coyne et al. 1997, Wade and Goodnight 1998). Establishing the relative magnitudes of these effects in natural populations is therefore critical to our basic understanding of evolution and our ability to predict the outcome of hybridization.

In small, structured populations, the evolution and maintenance of epistatic interactions is determined, in part, by the balance between gene flow and genetic drift and the temporal persistence times of local populations. When such populations occupy ephemeral environments, local extinction introduces additional dynamics of

colonization/recolonization that result in complex historical relationships of local subpopulations. Theory predicts that these dynamics will have large genetic consequences for the evolution of local populations (Whitlock and McCauley 1990). Therefore, empirical determination of the age structure of populations, their historical relationships and the relative magnitudes of gene flow and genetic drift are essential for understanding the outcome of hybridization. The harpacticoid copepod *Tigriopus californicus* is an excellent example of a species in which these dynamics are at work. *T. californicus* inhabit irregularly distributed rocky intertidal habitat along the Pacific coast of North America. In laboratory crosses, hybrid breakdown is generally observed among genetically divergent populations (Burton 1987, Burton 1990, Edmands 1999, Burton and Lee 1994, Burton 1998, Edmands 2001). Hybrid breakdown in these crosses can be explained by the evolution of positive epistatic interactions within regional populations that have long histories of isolation. However, in some cases, hybrid breakdown also occurs within regionally-distributed clades. In a series of crosses among subpopulations (< 25 km apart), Burton (1990) reported increased developmental time in F_2 hybrids. A similar observation has been made for subpopulations separated by as little as one km (Edmands 1999), but hybrid breakdown was not observed in other crosses between populations in close proximity. Therefore, in contrast to crosses between divergent clades, F_2 hybrid breakdown between populations separated by short distances is generally unpredictable. Such cases are intriguing because they may provide insight into the incipient stages of genome divergence. Distinguishing between alternative causes of the unpredictability

of the outcome of hybridization requires understanding the historical relationships of populations and the extent of on-going gene flow between them. This was the subject of the work described in Chapter II.

I expanded on previous work on local subpopulations to address problems concerning (1) the persistence times of local subpopulations, (2) the magnitude of gene flow and genetic drift in *T. californicus* populations, and (3) the role of balancing selection (see chapter 2 discussion) in generating the observed patterns of variation at the phospho-glucose isomerase (*Pgi*) allozyme locus. Studies of local population structure at *Pgi* and other allozyme loci suggest little or no gene flow even among local subpopulations (e.g., < 1 km). The magnitude of gene flow among populations has been difficult to quantify, in part, because of limited shared polymorphism among populations. A primary goal of this project was to quantify gene flow among local populations with estimates derived from the genealogy of alleles (e.g., using the method of Slatkin and Maddison 1989) at *Pgi*. Unfortunately, extensive recombination at the *Pgi* locus precluded using the coalescent to estimate migration rates because a reliable genealogy of alleles could not be reconstructed. However, a primary result of this study was that, despite limited sampling, we identified what appears to be broad sharing of nucleotide polymorphism and no fixed differences among subpopulations spanning > 60 km of coastline. This is consistent with historical gene flow that was not apparent in the allozyme data (Burton and Feldman 1981). In addition, estimates of nucleotide diversity flanking a charge-changing polymorphism suggest that a private allozyme allele in a focal population is likely

recently derived. This is significant because it indicates that the narrow distribution of private allozyme alleles may, in part, be attributed to their recent origin and is not strictly a function of gene flow.

The results in Chapter II suggest that gene flow among local populations of *T. californicus* may be higher than allozyme data had suggested. If the nucleotide polymorphism data at *Pgi* reflects on-going gene flow, then this has implications for our interpretation of the causes of hybrid breakdown in local inter-population crosses. It is possible that gene flow among local subpopulations is sufficiently high to limit the buildup of positive epistasis within populations, which would explain why reduced fitness in hybrids is generally not observed on local scales (Edmands 1999). However, this raises additional questions concerning cases where hybrid breakdown does occur. If gene flow is on-going, then the maintenance of positive epistasis within populations may in some cases be maintained by strong selection. Alternatively, the age structure of local populations that is introduced by local extinction and colonization/recolonization dynamics may produce a patch-work of populations with different histories of isolation. In this case, hybrid breakdown may occur in some inter-population crosses, but not in others as a consequence of complex historical relationships of local subpopulations. In either case, evidence of gene flow in DNA sequence data at *Pgi* has implications for (1) the demographic causes of hybrid breakdown in *T. californicus* and (2) for the design of future experiments on hybridization between local populations.

Ribosomal RNA (rRNA) transcription in inter-population hybrids

In Chapter II, I focused on classical population genetic approaches to address questions concerning the demography of *T. californicus* populations. In Chapters III and IV, I focus on the genetic basis of hybrid breakdown, which remains a significant problem in molecular population genetics. Reduced fitness in hybrids is caused by disruption of parental combinations of alleles as a consequence of recombination in the F_1 generation. The complexity of such epistatic interactions has been an impediment to determining the molecular basis of reduced hybrid fitness and, as a consequence, limited progress has been made in this area (Wu and Palopoli 1994, Oka et al. 2004). A possible strategy for beginning to address this problem is to evaluate specific genetic interactions that potentially have large downstream effects on metabolic function. One such interaction involves the transcriptional regulation of the rRNA gene family, which is necessary for ribosome biosynthesis and dictates the protein synthetic capacity of a cell. The goal of chapter 3 was to evaluate if rRNA gene regulation is altered in inter-population hybrids. As discussed below, this hypothesis was derived from observations of extraordinarily high levels of inter-population divergence in the region harboring *cis*-regulatory elements necessary for rRNA gene regulation in *T. californicus*. The results we obtained suggest dramatically altered rRNA gene expression in hybrids, but the relationship of this to hybrid breakdown is not clear.

The hypothesis that rRNA transcription may negatively impact metabolic function in hybrids was derived from three observations. First, in a number of

experimental systems, fitness-related traits such as delayed developmental time are associated with reduced rates of rRNA transcription. The most notable examples of the phenotypic effects of reduced rates of rRNA synthesis in eukaryotes are the bobbed-type mutants in *Drosophila* (Ritossa et al. 1966). The consequence of reduced rates of rRNA transcription in these mutants is a reduced capacity for protein synthesis, which has pleiotropic effects on developmental time and a number of morphological traits (Mohan and Ritossa 1970). Second, species-specificity of transcription of the rRNA gene family has been documented with *in vitro* transcription assays in which the RNA polymerase I (Pol I) apparatus from one species fails to recognize the promoter from another. This has been demonstrated in *Drosophila melanogaster* and *D. virilis* (Kohorn and Rae 1982), human and mouse (Grummt et al. 1982, Miesfeld et al. 1984, Heix and Grummt 1995), and a variety of other organisms (e.g., Fan et al. 1995). The reason for species-specificity may be related to the fact that this transcription system is dedicated exclusively to the transcription of the rRNA gene family. This one polymerase-one promoter interaction may facilitate the evolution of promoter specificity that is unlikely to occur in RNA polymerase II or III systems, which are constrained by the necessity to recognize multiple promoters. In addition, Dover and Flavell (1984) suggested that concerted evolution of the rRNA gene family could promote rapid evolution of Pol I/rRNA promoter interactions. They proposed that fixation of deleterious mutations in the rRNA gene promoter could select for compensatory changes in a component of the Pol I apparatus. Dover and Flavell (1984) hypothesized that this would result in functional incompatibility

between heterospecific combinations of the rRNA gene promoter and Pol I. A consequence of this is that promoter-specificity may evolve rapidly and therefore could lead to altered transcription of rRNA in hybrid organisms. This may have major downstream effects on rates of cell division and consequently on phenotypic traits such as increased development time.

Finally, we found that the region containing the rRNA gene promoter has diverged substantially in isolated *T. californicus* populations (Burton et al. 2005). Uncorrected inter-population divergences in this region range from 10% to over 35%. This observation raised the question of whether such large divergences impact the regulation of rRNA genes in inter-population hybrids. The work presented in chapter III was designed to test the hypothesis that rRNA gene expression was biased in inter-population hybrids.

The primary result from this work was that all F_1 hybrids show complete or nearly complete transcriptional silencing of rRNA genes from one parental population. In each cross examined, the silenced genes were always derived from the same population and were independent of the direction of the cross. In the F_2 generation, some hybrids were found to codominantly express rRNA genes, but most showed a pattern of transcriptional dominance similar to F_1 hybrids. The pattern of uniparental silencing of rRNA gene expression that we observed in F_1 hybrids is identical to the phenomenon of nucleolar dominance that has been reported in a variety of inter-specific hybrids in plants and animals (Reeder 1985, Pikaard 2000ab).

The relationship between nucleolar dominance and hybrid breakdown is not clear, but the results argue against a primary role for nucleolar dominance in reduced fitness. The reason for this is that nucleolar dominance was observed in F_1 hybrids, which in *T. californicus*, are generally vigorous (e.g., Burton 1990, Edmands 1999). Therefore, transcriptional silencing of approximately 50% of rRNA genes apparently does not limit the availability of rRNA to the extent necessary to lengthen developmental time or negatively impact other traits that show heterosis in F_1 hybrids. Similar observations have been made in hybrids between inbred lines of maize that show heterosis and also silence one rRNA gene array (Jupe and Zimmer 1993). The pattern of rRNA gene expression in the F_2 generation in *T. californicus* was consistent with our expectations of biased expression in some hybrids. However, in all cases where significant bias was observed, it was in the same direction as observed in the F_1. This suggests that the mechanism that determines dominance in F_1 hybrids, may also specify the dominance relationship in the F_2 generation.

Given the absence of a clear relationship between rRNA gene expression and hybrid breakdown, the work in Chapter III is discussed in the context of the molecular basis of nucleolar dominance. Work on model organisms including maize (Jupe and Zimmer 1993), *Xenopus* (Reeder and Roan 1984), *Drosophila* (Durica and Krider 1977), and *Arabidopsis* (Pikaard 2000b) have made progress on the genetic basis of nucleolar dominance, but the ultimate causes are still unknown. A unique feature of the *T. californicus* inter-genic spacer (IGS) provided an opportunity to evaluate a specific hypothesis that is not possible in these model organisms. As discussed in

Chapter III, a leading hypothesis for the establishment and maintenance of nucleolar dominance in hybrids involves the titration of transcription factors by enhancer-bearing subrepeat elements in the IGS (Reeder 1985). This enhancer-imbalance model proposes that a greater number of enhancers in the IGS of the dominant rRNA gene array will effectively titrate limiting sets of transcription factors thereby silencing the under-dominant array. The IGS in virtually all animals contains subrepeat elements that in some cases are known to enhance rRNA transcription and have been proposed to be responsible for nucleolar dominance. However, *T. californicus* does not contain repeat elements of any kind in the IGS (Burton et al. 2005). Therefore, a primary implication of the work in chapter 3 is that the enhancer-imbalance hypothesis does not adequately explain the uniparental silencing of rRNA gene expression in hybrid organisms.

Cyto-nuclear interactions and the evolution of mitochondrial transcription genes

Epistatic interactions between genes in the cytoplasm and the nucleus are an important class of interactions that reduce hybrid fitness (Clark and Lyckegaard 1988). An intriguing feature of such interactions is that organellar genomes encode relatively few genes that are involved in well-characterized functional interactions with nuclear encoded proteins. Therefore, genes involved in cyto-nuclear interactions provide an excellent opportunity to evaluate the molecular basis of epistatic interactions that impact hybrid fitness. Research on the molecular basis of cyto-nuclear interactions has focused on genes that are functionally integrated in the electron transport system

(ETS, Blier et al. 2001, Rand et al. 2004). This work has provided clear evidence for structural interactions in the ETS that are disrupted in hybrids and likely contribute to reductions in hybrid fitness (e.g., Edmands and Burton 1999, Rawson and Burton 2002, Sackton et al. 2003).

In chapter IV, I initiated work on the mitochondrial transcription apparatus to evaluate if cyto-nuclear incompatibilities also evolve between the rapidly evolving mitochondrial genome (mtDNA) of *T. californicus* and nuclear-encoded transcription genes. Recent developments in the study of mitochondrial transcription suggest that species-specificity of transcription has evolved during the course of mammalian evolution in a fashion that appears to be directly analogous to the Pol I transcription system. Gaspari et al. (2004) reported that the mouse mitochondrial transcription apparatus could not initiate transcription of the human promoter. In a factor-swapping experiment, they found that the specificity of promoter recognition was attributable to mitochondrial RNA polymerase and not either of the auxiliary transcription factors. While the significance of mouse-human polymerase/promoter specificity for microevolution is not clear, comparison with the related T3/T7 RNA polymerase system suggests a possible connection. Promoter specificity in bacteriophages has a simple genetic basis. Single residues in T3/T7 RNA polymerase (Raskin et al. 1992) and single bases in the promoter (Klement et al. 1990) are responsible for the species-specificity of promoter recognition (see Chapter IV discussion). Our hypothesis is that the species-specificity of transcription may evolve early in the process of species

divergence and therefore may have functional consequences for mitochondrial transcription in hybrid organisms.

The rapidly evolving mitochondrial genome of *T. californicus* and the high levels of divergence in mtDNA and nuclear genes provide an opportunity to evaluate the evolutionary significance of polymerase/promoter interactions. In chapter IV, I present the cloning of a mitochondrial transcription factor, *mtTFB1*, and the mitochondrial RNA polymerase, *mtRPOL*. An important distinction between the RNA Pol I apparatus (see above) and the mitochondrial system is that the mitochondrial apparatus is experimentally tractable in a non-model organism like *T. californicus*. The simplicity of this transcription system will facilitate future *in vitro* or *in organello* experiments on the species-specificity of transcription that are not tractable with the RNA Pol I system. A specific aim of the work presented here was to characterize the extent of inter-population divergence at these loci, especially in a region of the polymerase (i.e., the "specificity loop", Jeruzalmi and Steitz 1998) likely responsible for promoter recognition. A primary result is that these genes have diverged in isolated populations, but there has been limited evolution in the specificity loop (see Chapter IV discussion). Analysis of substitutions in other functional domains indicates that the high levels of divergence were not driven by positive selection. In addition, I show that *mtRPOL* and *mtTFB1* are tightly linked. Because of the hypothesized functional interaction of these genes with mtDNA, I hypothesized that large viability effects associated with these genes may be detected in segregation distortion in F_2 inter-cross hybrids. I genotyped F_2 hybrids at these loci and found

large reciprocal cross effects on the viabilities of *mtRPOL/mtTFB1* genotypes. These deviations from expected genotypic proportions were consistent with strong cytoplasmic effects on fitness, but higher order nuclear-nuclear and nuclear-cytoplasmic interactions likely underlie the observed effects. A significant implication of these results is that complex nuclear-cytoplasmic interactions accumulate early in the divergence of populations.

REFERENCES

Blier, P. U., F. Dufresne, and R. S. Burton. 2001. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends in Genetics. 17:400-406.

Brown, A. F. 1991. Outbreeding depression as a cost of dispersal in the harpacticoid copepod, *Tigriopus californicus*. Biol. Bull. 181:123-126.

Burton, R. S. 1986. Evolutionary consequences of restricted gene flow in the copepod *Tigriopus californicus*. Bull. Mar. Sci. 39:526-535.

Burton, R.S. 1987. Differentiation and integration of the genome in populations of the marine copepod *Tigriopus californicus*. Evolution. 41:504-513.

Burton, R.S. 1990. Hybrid breakdown in developmental time in the copepod *Tigriopus californicus*. Evolution. 44(7): 1814-1822.

Burton R.S. 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. Evolution. 52:734-745.

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

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

Burton, R. S. and S. G. Swisher. 1984. Population structure of the intertidal copepod *Tigriopus californicus* as revealed by field manipulation of allele frequencies. Oecologia. 65:108-111.

Burton, R.S., E.C. Metz, J.M. Flowers and C.S. Willett. 2005. Unusual structure of ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal subrepeats. Gene. 344:105-113.

Clark, A.G. and E. M. S. Lyckegaard. 1988. Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mitochondrial DNA in *Drosophila melanogaster*. Genetics. 118:471-481.

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. Perspective: a critique of Sewall Wright's shifting balance theory of evolution. Evolution. 51:643-671.

Crow, J. F. and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Dobzhansky, T. 1948. Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographical regions. Genetics. 33:588-602.

Dover, G. A. and R. B. Flavell. 1984. Molecular coevolution – DNA divergence and the maintenance of function. Cell. 38:622-623.

Durica, D.S. and H.M. Krider. 1977. Studies on the ribosomal RNA cistrons in interspecific *Drosophila* hybrids. I. Nucleolar dominance. Dev. Biol. 59:62-74.

Edmands, S. 1999. Heterosis and outbreeding depression in inter-population crosses spanning a wide range of divergence. Evolution. 53:1757-1768.

Edmands, S. and R. S. Burton. 1999. Cytochrome c oxidase activitiy in interpopulation hybrids of a marine copepod: a test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. Evolution. 53:1972-1978.

Edmands, S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. Molecular Ecology. 10: 1743-1750.

Fan H., K Yakura, M. Miyanishi, M. Sugita, and M. Sugiura. 1995. *In vitro* transcription of plant RNA polymerase I-dependent ribosomal RNA genes is species-specific. Plant Journal. 8:295-298.

Gaspari, M., M. Falkenberg, N. G. Larsson, C. M. Gustafsson. 2004. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. EMBO. 23:4606-4614.

Grummt, I., Roth, E. and M. R. Paul. 1982. rRNA transcription *in vitro* is species-specific. Nature. 296: 173-174.

Heix, J. and I. Grummt. 1995. Species-specificity of transcription by RNA polymerase I. Curr. Opin. Genet. Dev. 5:652-656.

Jeruzalmi, D. and T. A. Steitz. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO. 17:4101-4113.

Jupe, E. R. and E. A. Zimmer. 1993. DNase-sensitive and under-methylated rDNA is preferentially expressed in a maize hybrid. Plant Molec. Biol. 21:805-821.

Klement, J. F., M. B. Moorefield, E. Jorgenson, J. E. Brown, S. Risman, and W. T. McAllister. 1990. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. J. Mol. Biol. 215:21-29.

Kohorn, B. D. and P. M. M. Rae. 1982. Accurate transcription of truncated ribosomal DNA templates in a *Drosophila* cell-free system. Proc. Nat. Acad. Sci. USA. 79:1501-1505.

Miesfeld, R., B. Sollner-Webb, C. Croce, and N. Arnheim. 1984. The absence of a human-specific ribosomal DNA transcription factor leads to nucleolar dominance in mouse > human hybrid cells. Mol. Cell. Biol. 4:1306-1312.

Mohan, J. and F. M. Ritossa. 1970. Regulation of ribosomal RNA synthesis and its bearing on the bobbed phenotype in Drosophila melanogaster. Dev. Biol. 22:495-512.

Oka, A., A. Mita, N. Sakurai-Yamatani, H. Yamamoto, N. Takagi, T. Takano-Shimizu, K. Toshimori, K. Moriwaki, and T. Shiroishi. 2004. Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. Genetics. 166:913-924.

Wu, C. I. and M. F. Palopoli. 1994. Genetics of postmating reproductive isolation in animals. Ann. Rev. Genet. 28:283-308.

Pikaard, C. S. 2000a. The epigenetics of nucleolar dominance. Trends in Genetics. 16:495-500.

Pikaard, C. S. 2000b. Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids. Plant. Molec. Biol. 43:163-177.

Rand, D.M., R.A. Haney, and A.J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol. Evol. 19:645-653.

Raskin, C. A., G. Diaz, K. Joho, and W. T. McAllister. 1992. Substitution of a single T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. J. Mol. Biol. 228:506-515.

Rawson, P.D. and R.S. Burton. 2002. Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. Proc. Natl. Acad. Sci. USA. 99:12955-12958.

Reeder, R.H. and J. G. Roan. 1984. The mechanism of nucleolar dominance in *Xenopus* hybrids. Cell. 38: 39-44.

Reeder, R.H. 1985. Mechanisms of nucleolar dominance in animals and plants. J. Cell Biol. 101:2013-2016.

Ritossa, F. M., K. C. Atwood, and S. Spiegelman. 1966. A molecular explanation of bobbed mutants of *Drosophila* as partial deficiencies in ribosomal DNA. Genetics. 54:819-834.

Sackton, T.B., R.A. Haney and D.M. Rand. 2003. Cytonuclear coadaptation in *Drosophila*: Disruption of cytochrome c oxidase activity in backcross genotypes. Evolution. 57:2315-2325.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science. 236. 787-792.

Slatkin, M. and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 123:603-613.

Templeton, A. R. 1986. Coadaptation and outbreeding depression. Pp. 105-116 in M. Soule, ed. Conservation biology: the science of scarcity and diversity. Sinauer, Sunderland, MA.

Wade, M. J. and C. J. Goodnight. 1998. Perspective: The theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. Evolution. 52:1537-1553.

Whitlock, M. C. and D. E. McCauley. 1990. Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution. 1717-1724.

Chapter II

Population differentiation and the genealogical history of private allozyme alleles in *Tigriopus californicus*

ABSTRACT

Inferences about the demographic histories of populations can be made directly from observations over time and indirectly from genetic data. In a few rare cases, time series data on patterns of genetic change through time provide important insight into the demography of populations. When coupled with genealogical information from DNA sequence data, combined analysis can facilitate powerful inferences about population processes. Populations of the marine copepod *Tigriopus californicus* are among the most structured of any species in the marine environment. Allozyme surveys have indicated temporal stability of high frequency private alleles and stability of population structure among immediately adjacent subpopulations over periods of hundreds of generations. Here we evaluate the factors responsible for the maintenance of both the geographic structuring and temporal persistence of private alleles at the phosphoglucose-isomerase (*Pgi*) locus. We report on *Pgi* allele frequencies in focal populations spanning a period of 26 years and find that private allele frequencies in 2004 are virtually identical to frequencies in 1978. In addition, we characterize allozyme polymorphisms at the nucleotide level and evaluate patterns of polymorphism at linked sites. We find evidence that a private allele at Pescadero, California has a recent origin and is likely not at migration-drift equilibrium. Phylogeographic structuring is evident among regional populations at *Pgi*, but within regions, there is evidence of historical exchange in the nucleotide data that was not apparent in allozyme surveys. These data suggest that natural selection is unlikely to account for the temporal stability of private alleles and suggest that inferences about

population exchange based upon allozymes alone may underestimate migration rates in structured populations.

INTRODUCTION

A common indicator of restricted gene flow among populations is the presence of genetic polymorphisms with narrow geographic distributions. At migration-drift equilibrium, the average frequency of neutral private alleles is inversely related to the extent of gene flow among neighboring populations (Slatkin 1985). Therefore, the presence of high frequency alleles that are restricted in geographic distribution suggests historical restrictions in gene flow among populations. Drawing inferences concerning rates of migration from private alleles requires that such alleles are not the target of strong selection and that alleles are at migration-drift equilibrium. In most cases, however, little is known about the history of private alleles, whether they are at migration-drift equilibrium or whether they are neutral in their phenotypic effect. In rare cases, time series data on the geographic distributions and frequencies of private alleles can provide important historical information that may help address these issues (Palm et al. 2003). When coupled with genealogical information from DNA polymorphism data, such studies can provide detailed insight into the demography of isolated populations.

The marine copepod *T. californicus* inhabits tidepools along the supralittoral fringe of rocky intertidal habitats on the Pacific coast of North America. Despite having feeding larvae and the capacity of all life stages for dispersal (Vittor 1971), *T. californicus* populations are among the most sharply structured of any marine organism (Burton et al. 1979, Burton and Feldman 1981, Burton 1987, Burton 1998, Edmands 2001). DNA sequencing studies of mitochondrial DNA (mtDNA) have

indicated that regional populations in Southern California show phylogeographic structuring with uncorrected inter-population divergences frequently exceeding 15% (Burton and Lee 1994, Burton 1998, Edmands 2001). Nuclear gene genealogies similarly show deep regional divisions (Burton and Lee 1994, Willett and Burton 2004, Burton et al. 2005), and suggest that many presumably ancient, but reproductively compatible (e.g., Burton 1987, Ganz and Burton 1995, Edmands 1999, Edmands and Deimler. 2004), allopatric lineages are distributed throughout the region (Burton 1998).

Within regionally distributed clades, fine scale allozyme surveys have similarly suggested historical isolation of local subpopulations. The predominant pattern that has emerged is that high frequency private alleles are found at virtually every locus that has been surveyed (Burton et al. 1979, Burton and Feldman 1981, Ganz and Burton 1985, Burton 1986, Burton 1987, Burton 1997). In most cases, such private alleles are restricted to narrow (i.e., <25km) stretches of coastline and in exceptional cases are primarily restricted to individual rocky outcrops. Furthermore, this apparent population differentiation is stable through time. For example, a single rocky outcrop at Pescadero, California maintained a fast-migrating phosphoglucose isomerase (Pgi^{F}) allele at approximately 50% at sampling intervals between 1978 and 1996. This allele was either absent or found at extremely low frequency at the adjacent outcrop to the south (designated Site 10, Burton and Feldman 1981). Site 10 similarly maintained a slow-migrating allele (Pgi^{S}) allele that has been recorded only intermittently at very low frequency at the Pescadero site (e.g., Burton et al. 1979).

The stable geographic distribution of alleles at this location and similar observations from throughout California suggests that individual rocky outcrops are effectively closed to immigration and emigration over periods spanning hundreds of generations.

Local populations of *T. californicus* are subject to extreme fluctuations in population density and are constantly at risk of extinction due to the ephemeral nature of the tidepool habitat (Vittor 1971, Dybdahl 1994). A consequence of this is that genetic drift is expected to be an important force in isolated *T. californicus* populations. Low levels of polymorphism observed at allozyme loci in this species are generally consistent with this expectation (Burton et al. 1979, Burton and Feldman 1981). However, Burton (1997) reported long-term stability of allele frequencies over a period spanning 17 years. This suggests either that effective population sizes are sufficiently large to buffer the effects of genetic drift or stabilizing selection acts in different subpopulations to maintain polymorphism at different enzyme loci (Burton 1997).

Here we characterize allozyme polymorphisms in *T. californicus* at the DNA sequence level at *Pgi*. We address three outstanding questions in the population biology of this species: (1) are private alleles at migration-drift equilibrium, (2) are the temporal stabilities of allele frequencies in fluctuating *T. californicus* populations maintained by strictly neutral/demographic processes, (3) has natural selection contributed to the maintenance of polymorphism among divergent populations? We report DNA polymorphism data for 43 chromosomes sampled from populations throughout the southern range of the species and conduct additional allozyme surveys

of populations that have been the subject of genetic studies for a period spanning 26 years. We find long-term stability of *Pgi* alleles at focal populations and demonstrate that the stability of allele frequencies is consistent with the action of genetic drift in a population of moderate size. At Pescadero, nucleotide polymorphism data suggest that a private allele maintained at 50% since 1978 is recently derived and likely not at migration-drift equilibrium. Our results suggest that the extent of population connectivity implied by private allozyme alleles may underestimate the true level of gene flow among *T. californicus* populations.

MATERIALS AND METHODS

Sample Collection

Animals were collected from locations throughout California and Baja California, Mexico between May, 2001 and December, 2004 and transported live to Scripps Institution of Oceanography, La Jolla, California. With the exception of samples from two new sites, Mussel Rock (37° 20' 55'' N / 122° 24' 08''W) and Pomponio Beach (37° 17' 48'' N / 122° 24' 25'' W), collecting locals are the same as those reported previously (Burton et al. 1979, Burton and Feldman 1981, Burton and Swisher 1984, Ganz and Burton 1995, Burton 1997, Burton 1998). Samples from Punta Morro and Playa Altamira, Mexico, were maintained as laboratory stocks for a period of three years prior to analysis. All other samples were either prepared for molecular analysis immediately or maintained at 20° C with a 12:12 light-dark cycle for a maximum of five days. The latter samples were carefully checked to ensure that no mortality had taken place prior to preparation for molecular analysis.

Cloning of *Pgi*

Pgi (E.C. 5.3.1.9) was cloned with a standard degenerate PCR-based approach. Degenerate primers were designed from an alignment consisting of *Drosophila* (Accession number: NP_523663.1), cricket (AAG15513.1), zebrafish (AAH44450.1), and human (AAP72966.1) *Pgi* sequences. Primers were designed to match six blocks of seven or more amino acids which consisted of low degeneracy amino acids that were completely conserved among the four species in the alignment. Total RNA was

extracted with TRI Reagent (Sigma) from copepods collected from the Ocean Beach Pier collecting site, San Diego, CA. First strand cDNA synthesis was primed with an oligo(dT) primer with an extension at the 5' end to facilitate subsequent 3'RACE. Touchdown PCR was performed with cycling conditions consisting of a 30 s 95° denaturation step, an initial annealing step of 30 s at 56° C and an extension step of 2 m at 72° C. Every 2 cycles for the first 12 cycles, the annealing temperature was stepped down by 1° C. The final 23 cycles were conducted with an annealing temperature of 50° C. A second round of PCR utilizing the same pairs of PCR primers was performed with a 50° C annealing temperature for 35 cycles with 1.0 μl of the original unpurified PCR product as template. Following two rounds of PCR, two degenerate primer pairs (5'CCIYTNATGGTNACNGARGC3' / 5'TCCATRTCNCCYTGYTGRAARTA3' and 5'TAYTTYCARCARGGNGAYATGGA3' / 5'ARYTCIACNCCCCAYTGRTC3') each yielded a single PCR product of the size expected from the original *Pgi* alignment. These products were gel purified, cloned using the pCR-4-TOPO vector (TOPO TA cloning Kit, Invitrogen), and sequenced on a MegaBace sequencer (Amersham Biosciences). The section of the gene between the two original products was subsequently amplified from cDNA with *T. californicus*-specific primers. The 5' and 3' ends of the gene were obtained by RLM-RACE (Generacer Kit, Invitrogen).

Allozyme electrophoresis, PCR, and DNA sequencing

All individuals included in the sequencing analysis were genotyped by allozyme electrophoresis prior to PCR amplification. Animals were homogenized in 10 µl of chilled running buffer (0.1M Tris-Borate-EDTA, pH 8.9, with 0.12 g/ml sucrose, 10 µg/ml bromphenol blue added for loading and tracking). Five µl of the homogenate was loaded directly on an acrylamide gel for allozyme electrophoresis following the method of Burton (1990). Gels were subsequently stained with a standard recipe for *Pgi* (Harris and Hopkinson 1976) except NADP was replaced with NAD for use with a recombinant glucose-6-phosphate dehydrogenase coupling enzyme from *Leuconostoc mesenteroides* (E.C. 1.1.1.49, Sigma). The remaining five µl was prepared for PCR immediately following loading of acrylamide gels. Samples were treated with five µl of Proteinase K (0.2 mg/ml) and incubated at 65° C for one hour followed by 85° C for 15 minutes.

All *Pgi* sequences analyzed were obtained by amplifying the entire structural gene of approximately 2.5 kb from genomic DNA with primers located in the 5' and 3' untranslated regions (UTR). PCR products were then cloned to determine the gametic phase of polymorphisms. Full length *Pgi* was amplified from Carpinteria and Playa Altamira samples with primers TcPGI-5'-34F and TcPGI-STOP+21R. All other samples were amplified with primers TcPGI-5'UTR-F and TcPGI-3'UTR-R (Table 1). Initially, at least one homozygote for Pgi^{F} and Pgi^{M} (i.e., $Pgi^{1.05}$ and $Pgi^{1.00}$ in the nomenclature of Burton and Feldman 1981) allozyme alleles from San Diego and Pescadero populations were amplified and sequenced directly for characterization of

charge-changing residues. Subsequent PCR products were gel extracted, concentrated by ethanol precipitation, and cloned using either the pCR-4-TOPO or TOPO XL vectors (TOPO TA cloning kit, Invitrogen). Inserts were amplified with a set of four partially overlapping pairs of primers (Table 1) and both strands of the PCR products were sequenced on a Megabace (Amersham Biosciences) capillary sequencer.

Pgi was cloned and sequenced from a subset of the individuals that were scored for allozyme genotype. Samples from Pescadero (n = 13), San Diego (n = 14), and Santa Cruz (n = 5) were selected randomly (i.e., independent of allozyme genotype) for sequencing to allow for estimation of population genetic parameters. In contrast, of the five alleles sequenced from Site 10, two Pgi^{M}/Pgi^{S} heterozygotes were selected specifically for characterization of the Pgi^{S} ($Pgi^{\cdot 89}$, Burton and Feldman 1981). However, we were unable to unambiguously characterize the amino acid replacement responsible for this allozyme. Single Pgi^{S}/Pgi^{S} and Pgi^{M}/Pgi^{M} homozygotes from Laguna Beach were selected for sequencing based upon allozyme genotype. Single sequences were obtained from additional populations from Carpinteria, Abalone Cove, Punta Morro, and Playa Altamira populations. We assessed the level of *Taq* error in our samples by directly sequencing PCR products from two copepods and comparing these sequences to sequences obtained from a cloned PCR product amplified from the same animals. We estimated a *Taq* error rate in one clone to be 1.0×10^{-4} errors/bp/cycle and 4.5×10^{-5} errors/bp/cycle in the other.

Effective population size estimation

We obtained a single-locus estimate of the effective population size of the Pescadero population based upon temporal changes in *Pgi* allele frequencies. The 11 samples included in the analysis were collected at different intervals between April 1978 and December 2004 (Table 2). We used the maximum likelihood-based approach of Anderson et al. (2000), which uses a Monte Carlo method to determine the log-likelihood curve. The method uses multinomial sampling of a Wright-Fisher population of size N_e to calculate transition probabilities between the observed allele counts from each temporal sample (Table 2) and unobserved counts from intervening generations. The log-likelihood is then calculated as the joint probability of the data (i.e., allele counts) and the unobserved allele counts of intervening generations summed over all Monte Carlo replicates for a particular N_e.

In the laboratory at 20° C, the minimum generation time of *T. californicus* is approximately 23-26 days (Edmands and Harrison 2003) and longevity is typically 2 to 3 months (Vittor 1971). Without comparable data from natural populations concerning rates of population turnover, we conducted analyses assuming a generation time of two months. The analysis consisted of >100,000 Monte Carlo replicates for each effective population size and 95% confidence limits were calculated for the log-likelihoods of each N_e assuming an asymptotic normal approximation (Anderson et al. 2000). A modified version of the package MCLEEPSv1.1 was kindly provided by E. Anderson to facilitate analysis of the large number of generations in our dataset. The

MCLEEPS package is available at

http://www.stat.washington.edu/thompson/Genepi/Mcleeps.shtml.

Analysis of DNA sequence data

Sequences were edited with Sequencer version 4.2.2 (Genecodes, Ann Arbor, Mich.) and aligned initially with ClustalX (Thompson et al. 1997), followed by minor adjustments to the alignment made by eye. Estimation of population parameters and tests of neutrality were conducted with DnaSP version 4.0 (Rozas et al 2003). In some cases, we constructed null distributions of the parameters of interest assuming a population of constant size at neutral equilibrium using the coalescent algorithm of Hudson (1990) as implemented in DnaSP. Analysis of linkage disequilibrium between pairs of polymorphic sites was conducted using Fisher's Exact test. Neighbor-joining trees were constructed in Paupv.4.0b10 (Swofford 1998).

RESULTS

Gene structure and organization

The full length cDNA contained an open reading frame of 558 codons with a putative methionine start codon and a stop codon near the 3' end of the transcript. Comparison of the translated cDNA from the San Diego population to *Pgi* sequences from human, *Drosophila melanogaster*, and *Anopheles gambiae* revealed amino acid identities of 0.706, 0.697, and 0.697, respectively. Comparison of cDNA and genomic DNA sequences indicated the presence of ten exons and nine small introns, which were between 64 and 181 bp long in the San Diego population. Small indel polymorphisms ranging in size from one bp to approximately 20 bp were common in all nine introns in inter-population comparisons, especially when the Playa Altamira sequence was included in the alignment. One indel restricted to this population appears to result from a compound $(GT)_5(CT)_4$ short tandem repeat located in intron 2. Within populations, we observed far less length polymorphism and only one non-singleton indel, a 2 bp indel that appears to be segregating at high frequency in intron 3 of the San Diego population. Our final alignment of 2,565 bp consisted of the entire structural gene and included all 10 exons, all 9 introns, and 12 bp of the 3'UTR.

Replacement polymorphism and charge-changing residues

The Pgi^M/Pgi^F polymorphism at Pescadero is caused by an Arg/Gln replacement polymorphism (G566A, Table 3, codon position 77) in exon 3. In Pescadero, medium-migrating (Pgi^M) alleles contained an Arg at this site and fast-

migrating (Pgi^F) alleles contained a Gln at this position. In contrast, the Pgi^M and Pgi^F allele classes in San Diego are caused by Gly and Asp residues, respectively, in exon 2 (G419A, Table 4, codon position 66). In Punta Morro, Baja California, a Pgi^F allele also contained an Asp at codon position 66 and no additional charge-changing replacements relative to the San Diego Pgi^F allele consistent with a single mutational origin for the fast-migrating alleles in these populations. Unfortunately, we were unable to obtain a Pgi^M allele from this location (Ganz and Burton 1995) to confirm that the Pgi^M and Pgi^F polymorphism are in fact shared in both populations. However, the widespread Pgi^M electromorph from populations throughout California was not due to convergence on charge in different regions and a single Ser/Gly change differentiates central and southern California Pgi^M alleles. Finally, the Pgi^M/Pgi^S polymorphism at Laguna Beach appears to be due to two charge-changing replacements (Lys/Glu at codon position 463, Asp/Asn at position 318).

Allozyme allele frequencies

The Pescadero site has maintained the widespread Pgi^M and a private Pgi^F allele since 1978 (Burton et al. 1979). Allele frequencies at this site were virtually identical to frequencies in samples collected at different intervals over the past three decades (Table 2; Burton et al. 1979, Burton and Feldman 1981, Burton and Swisher 1984, Burton and Lee 1994, Burton 1997). Our sample from July 2004 contained the private Pgi^F and widespread Pgi^M alleles at frequencies of 0.51 and 0.49, respectively. These frequencies are essentially identical to the corresponding allele frequencies of

0.52 and 0.48 observed in 1978 (Table 2). In addition to stable allele frequencies, we have observed that *Pgi* genotypes at this location rarely deviate from Hardy-Weinberg expectations. The only two exceptions are the June 1978 ($P < 0.05$; G-test, $G = 10.85$, $df = 3$) and July 2004 ($P < 0.05$; G-test, $G = 5.67$, $df = 1$) samples.

The Pgi^{F} allele was not observed at the immediately adjacent outcrop (Site 10) 500 m to the south of the Pescadero site in either of our samples from July or December 2004. As in previous surveys (Burton and Feldman 1981, Burton 1997), we observed a private slow-migrating, Pgi^{S}, allele at Site 10 that was not found at the Pescadero (Table 2). With the exception of a small boulder located approximately 500 m north of Pescadero (Site 11, Burton and Feldman 1981), the closest suitable habitat to the north is at Pomponio Beach and Mussel Rock. These two locations are located 4.2 km and 9.9 km from Pescadero, respectively. These sites were found to be nearly monomorphic for the Pgi^{M} allele. An additional private allele was found segregating at low frequency at these two locations (Table 2), but the Pgi^{F} allele was not observed. *Pgi* allele frequencies at Carpinteria (fixed on Pgi^{MS}; $N = 48$), Santa Cruz (fixed on Pgi^{M}; $N = 20$), and Laguna Beach ($Pgi^{F} = .058$, $Pgi^{M} = .709$, $Pgi^{S} = .221$, $Pgi^{VS} = .012$; $N = 43$) were comparable to previous studies. A single sample from Punta Morro in 2004 was monomorphic for the Pgi^{F} allele, which contrasts with a previous report (Ganz and Burton 1995).

Nucleotide polymorphism

Inter-population distances estimated from *Pgi* after correction for within population variation are comparable to other *T. californicus* nuclear genes (Willett and Burton 2004). Similar to previous studies, there is evidence at *Pgi* of multiple regionally distributed allopatric lineages with inter-population distances of 1.5-2.5% (Table 6). For example, comparison of Pescadero and San Diego populations revealed 62 fixed differences (61 total silent, 1 replacement), or 2.5%, divergence over the length of the gene excluding indels. Within central California, F_{st} was estimated at 0.12 between Pescadero and Santa Cruz (Hudson et al. 1992). No fixed differences were observed between the three sites at Santa Cruz, Site 10, and Pescadero. Despite limited sampling, we observed 28 shared polymorphisms between Santa Cruz and Site 10/Pescadero (Table 3). A neighbor-joining tree illustrates the absence of divergence among these sites and the large divergences between regional populations and a partially reproductively isolated population from Playa Altamira, Baja California, Mexico (Fig. 1).

Estimates of nucleotide diversity (π) in the San Diego and Pescadero populations were 0.014 and 0.017 at synonymous sites, in each population respectively (Table 5). DNA polymorphism in the San Diego population exhibits a single peak in nucleotide diversity centered around the charge-changing Gly/Asp polymorphism (Fig. 2). The peak in polymorphism in the sliding window analysis is produced by a high frequency polymorphism in the 3rd position of a proline codon (position 372; Table 4), the charge-changing site (position 419), and five high

frequency polymorphisms in the adjacent intron (positions 477, 478, 481, 482, and 541 in intron 2; Table 4). We tested whether this peak had significantly elevated levels of polymorphism with a parametric bootstrap approach. We simulated 10,000 genealogies in a neutral equilibrium population with θ equal to that estimated from the entire gene in samples from San Diego. A null distribution of π and θ_W was produced from a random draw of 14 alleles from each genealogy. We then compared observed values of θ_W and π in arbitrarily chosen window sizes of 100 and 200 bp flanking the charge-changing site. The results indicated that neither π nor θ_W are significantly elevated ($P > .05$) in this region.

In Pescadero, nucleotide polymorphism differed between the widespread Pgi^M and private Pgi^F allele classes. Because of extensive recombination (or gene conversion) between Pgi^M and Pgi^F classes (see below), we examined variation in a 520 bp window that appears to be free of recombination within the private Pgi^F class (Table 3). Within this window, Tajima's D differed in sign between the two classes (Pgi^F = -1.29503 and Pgi^M = 0.16637). Estimates of π were 0.0026 in the private Pgi^F class and .0097 in the widespread Pgi^M class. We tested if polymorphism differed significantly between the two allele classes by generating 10,000 genealogies using a conservative (i.e., lower bound) value of θ estimated from both allele classes combined. This was followed by determining if the observed values of π, θ_W, and Tajima's D for the Pgi^F allele ($n = 6$) were less than those expected by a random draw of six alleles from the simulated population. Test statistics were non-significant when simulations were conducted without recombination (Table 7). However, estimates of

all three parameters estimated from the Pgi^F class were significantly less than expected from a population at neutral equilibrium when the estimated population recombination rate, C (Hudson 1987), was incorporated in the simulation (Table 7).

Recombination and Linkage disequilibrium

We found no evidence of recombination among alleles from the San Diego and Pescadero populations consistent with the reported historical isolation of these populations (e.g., Burton and Lee 1994). We subsequently analyzed LD separately for each population. The four gamete test of Hudson and Kaplan (1985) identified a minimum of five cross-over events across the length of the gene in both San Diego and Pescadero populations. Given the small sample sizes in this study, we cannot draw firm conclusions about the absence of LD, but in the event that we do detect LD, this is indicative of a minimal level of historical recombination (Lewontin 1995). In Pescadero, LD was primarily restricted to small clusters of sites less than 500 bp apart (Figure 3A). The highest levels of LD were observed in a region from intron 5 to intron 8 (positions 1518 to 2190). The general pattern of small clusters of LD among closely linked sites was also found in San Diego except there were a number of sites separated by > 1 kb which also showed significant associations (Figure 3B).

The private Pgi^F allele class in Pescadero showed no evidence of recombination within or between allele classes in a region of 520 bp flanking the charge-changing site. Outside of this window there is evidence of substantial exchange between Pgi^F and Pgi^M allele classes. A minimum of four recombination or

gene conversion events between Pgi^F and Pgi^M allele classes are required to explain the haplotype structure in the Pgi^F allele class (Table 3).

Effective population size

Figure 4 is a plot of the log-likelihood curve for N_e based upon temporal changes in *Pgi* allele frequencies in the Pescadero population (Table 3). The maximum likelihood estimate (MLE) of the effective population size at *Pgi* was 7,500 (Fig. 4).

DISCUSSION

T. californicus populations that are polymorphic for high frequency allozyme alleles frequently maintain these alleles at stable frequencies for periods spanning hundreds of generations. The population inhabiting a rocky outcrop adjacent to the parking lot at Pescadero, California has maintained the private Pgi^F allele at stable frequency for a period now spanning at least 26 years. However, despite only minor fluctuations in allele frequencies from 1978 to 2004, the temporal changes in allele frequencies are consistent with that of a neutral allele in a population of moderate size. Assuming a generation time of two months, we obtained an MLE of the effective size at *Pgi* of 7,500. This is more than an order of magnitude smaller than abundances of $> 10^5$ suggested by Burton and Feldman (1981) for this population near the beginning of the times series at Pescadero. This suggests that the size of the local population is sufficiently large to maintain stable allele frequencies for periods spanning hundreds of generations.

These considerations, together with evidence that the Pgi^F allele has a recent origin (see below), suggest that natural selection is unlikely to have contributed to the stability of *Pgi* allele frequencies at Pescadero. More generally, the temporal stability of allele frequencies in *T. californicus* populations suggests that genetic drift is a weak force over periods of hundreds of generations. Although population abundances may frequently fluctuate over several orders of magnitude, our results suggest that bottlenecked populations must remain sufficiently large that allele frequencies are not perturbed by genetic drift. The effect of recurrent bottlenecking may also be partially

offset by the capacity of *T. californicus* populations for rapid growth. For example, continuous breeding, short generation time (Vittor 1971, Harrison and Edmands 2003), high survivorship of nauplii (Harrison and Edmands 2003), and the ability of females to produce multiple broods (Burton 1985) are all characteristics of this species that are conducive to rapid recovery from reductions in size. Such features may allow populations to rebound from a crash within one or a few generations, thus buffering against the diversity-reducing effects of prolonged periods at bottlenecked population sizes (Nei et al. 1975).

A remarkable feature of local *T. californicus* populations is the presence of high frequency private alleles that are restricted in distribution to narrow (i.e., <25 km) stretches of coastline. In a survey of populations in central California, Burton and Feldman (1981) reported high-frequency private alleles at all five allozyme loci examined ($Got^{1.07}$ at Monterey (30%), $Pgm^{1.07}$ at Capitola (25%), Pgi^{F} at Pescadero (48%), $Gpt^{1.06}$ at Santa Cruz (25%), $Got^{1.03}$ at Bodega Bay (50%), and $Est^{.97}$ at Moss Beach (30%)). Numerous low frequency polymorphisms were also prominent and similarly restricted in distribution. These observations clearly indicate that *T. californicus* populations experience extremely low rates of migration. However, determining how far such alleles are from equilibrium between migration and drift is critical for understanding the extent of genetic exchange among local populations (Slatkin 1985). If private allozyme alleles are recently derived, their narrow distributions may, in part, be attributed to their recent origin and may not be strictly a function of low migration rate. This is because of the basic expectation that in

structured populations the geographic distribution of a mutation is the result of a multi-generation dispersion process. New mutations are expected to have restricted geographic distributions, while mutations that persist sufficiently long may become geographically widespread (Neigel and Avise 1993).

Our data indicate that the Pgi^F allele at Pescadero is not at mutation-drift equilibrium. This is evident in a small window flanking the Asp/Gly polymorphism that retains the signature of a recent origin of this allele. Within this window, nucleotide diversity estimated from this allele class is significantly reduced and the frequency spectrum of polymorphism is skewed from equilibrium expectations. This suggests that the region flanking the charge-changing site within the Pgi^F allele class is not at mutation-drift equilibrium and that the Pgi^F allele is likely recently derived. The significance of this is that the extremely narrow geographic distribution of this allele is likely attributable to its recent origin together with an extremely low rate of migration among immediately adjacent subpopulations. If other private allozyme alleles are similarly not at equilibrium, then this may contribute to the generally narrow distributions of allozyme alleles in this species.

Recent origins of private allozyme alleles may also explain a discrepancy in the extent of shared polymorphism and implied levels of historical exchange in allozyme and DNA sequence datasets. The discrepancy arises from the observation that DNA sequence data suggests sharing of polymorphism among distant populations that was not apparent in the allozyme data. In central California, private allozyme alleles are commonly found in subpopulations spanning the 60 km of coastline

between Pescadero and Santa Cruz and throughout the region (Burton et al. 1979, Burton and Feldman 1981). In contrast, shared polymorphisms among distant populations are rare and are consistently observed at only one allozyme locus (*Pgm*). In this study, our observation of 28 nucleotide polymorphisms shared between Santa Cruz and Pescadero populations and the absence of fixed differences at *Pgi* was therefore unexpected. This suggests a discrepancy between DNA sequence and allozyme datasets—shared DNA sequence polymorphisms among distant populations suggest historical, and possibly on-going, exchange, while restricted distributions of allozyme polymorphisms generally do not.

This qualitative argument is supported by differences in F_{st} between allozyme and DNA sequence data. F_{st} calculated from allozymes reported by Burton and Feldman (1981) and Burton (1986) yielded an estimate of 0.21, which is higher than our estimate from *Pgi* sequence data of 0.12. Although these values are not dramatically different, we expect that these values may underestimate the difference between marker classes. The estimate from allozyme data is likely to be downwardly biased because of an artifact associated with calculating F_{st} with private alleles. For example, in Santa Cruz, the frequency of the private $Gpt^{1.06}$ allele is 25% while that of the widespread $Gpt^{1.00}$ allele is 75%. In contrast, the Pescadero site is fixed on the $Gpt^{1.00}$ allele. Despite no evidence of exchange at this locus between these populations, the F_{st} at this locus assumes a low value of 0.14. Thus, levels of migration inferred from allozyme and DNA sequence datasets may significantly differ between marker classes. If additional nuclear gene loci also show this pattern, then

the extent of population connectivity suggested by the narrow distributions of allozyme alleles may lead to underestimates of the true levels of migration among *T. californicus* populations.

What factors are responsible for the predominance of private alleles in allozymes and the discrepancy in the extent of shared polymorphism between marker types? One possibility is that weak selection acts against slightly deleterious charge-changing polymorphisms (e.g., Aquadro et al. 1988, Ohta 1992). While this could contribute to narrow distributions of allozymes, it seems unlikely that mildly deleterious mutations are generally responsible for allozyme alleles that are segregating at virtually every locus in this species. Alternatively, the discrepancy may be attributed to the distribution of the ages of mutations in a population and differences in the nucleotide sampling properties of allozyme and DNA sequencing methods. Because allozymes screen only a small number of polymorphic sites in a gene, the polymorpisms are most likely to be drawn from the most abundant mutational age class--young mutations (Slatkin and Rannala 2000). In contrast, DNA sequences frequently contain many polymorphic sites and are therefore more likely to sample the entire age distribution, including young and old mutations. If new mutations in substructured populations generally have more restricted geographic distributions than old mutations (Avise and Neigel 1993), then allozyme surveys will primarily detect private alleles. By sampling many polymorphic sites, DNA sequencing studies are more likely to detect both narrowly distributed and more broadly distributed polymorphisms.

An unexpected observation was evidence of extensive recombination between Pgi^F and Pgi^M classes in the Pescadero population. Despite its recent origin, the private Pgi^F allele class retains only a small remnant block of complete LD in a window of 520 base pairs flanking the charge-changing site. LD outside this window, however, appears to have decayed substantially as a result of a minimum of four recombination (or gene conversion) events between allele classes (Table 3). This is unexpected for a very recently derived allele because a new mutation will initially be in complete LD with all polymorphisms on the chromosome on which it arises. The extent of decay of LD is not necessarily inconsistent with a recent origin of the Pgi^F allele and hints at the action of gene conversion in this region. Rates of gene conversion have only been estimated in yeast, flies, and humans, but in all three groups, estimates of the gene conversion rate exceeds that of crossing-over by two-, four-, and seven-fold, respectively (Frisse 2001 and cited references). In humans, gene conversion is the primary factor that determines the pattern of LD along small sections of chromosomes typical of sequencing studies (Przeworski and Wall 2001). Therefore, it seems likely that the extent of decay of LD within the derived Pgi^F allele is attributable to the recent action of gene conversion. However, in the combined analysis Pgi^F and Pgi^M alleles from San Diego and Pescadero a minimum of five cross-over events are necessary to explain variation within each population (Hudson and Kaplan 1985). These observations suggest that both gene conversion and crossing-over have been important factors contributing to the pattern of LD at *Pgi*.

In *T. californicus*, most allozyme polymorphisms are restricted in distribution to immediately adjacent subpopulations. In an unusual case, Ganz and Burton (1995) reported that populations from Punta Morro, Baja California, Mexico and San Diego share a high frequency Pgi^F/Pgi^M polymorphism. This is interesting because these populations are in relatively close geographic proximity (i.e.,100 km), but are reciprocally monophyletic at cytochrome oxidase I (Burton 1998, Edmands 2001) and are divergent at multiple nuclear loci (Burton et al. 2005, RSB and JMF unpublished, S. Harrison unpublished). Although we were unable to obtain a sequence for the widespread Pgi^M allele in Punta Morro, our results indicate that sharing of the Pgi^F allele is not due to convergence on charge (e.g., Katz and Harrison 1997).

These observations raise the possibility that natural selection is maintaining the Pgi^F allele in these southern populations. Polymorphisms that are maintained longer than expected under neutrality are expected to accumulate excess levels of variation at linked neutral sites (Hudson and Kaplan 1988). In the event that a polymorphism is maintained for a sufficiently long period, then the action of selection may be detected in the form of a peak of polymorphism centered around the selected site (Hudson and Kaplan 1988). In our survey of San Diego alleles, we identified a single peak of nucleotide diversity in the immediate vicinity of the charge-changing replacement polymorphism at the boundary between exon 2 and intron 2 (Fig. 2). The polymorphism in this window appears to be approximately two-fold higher than variation throughout the remainder of the gene, but is not significantly elevated based upon coalescent simulation. However, we find it intriguing that the region we

hypothesized *a priori* may be subject to balancing selection was found to have a cluster of mutation in immediately flanking sites. If selection is responsible for the increased polymorphism in San Diego, we expect that levels of heterozygosity will be increased around the charge-changing site in the Punta Morro population as well.

ACKNOWLEDGEMENTS

We are indebted to the following people who were instrumental in various aspects of this work. We thank C. Ellison for scaling a cliff in the rain to obtain a sample, E. Anderson for providing assistance with effective population size estimation, L. Dingding for help with allozyme gel electrophoresis, and R. Byrne for general laboratory assistance. H. Hoekstra, P. Andolfatto, C. Wills, S. Harrison, and C. Willett offered helpful comments on data analysis and early drafts of the manuscript. We gratefully acknowledge financial support from a National Science Foundation grant to R.S.B.

REFERENCES

Anderson, E. C., E.G. Williamson, and E. A. Thompson. 2000. Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. Genetics. 156:2109-2118

Aquadro, C. F., K. M. Lado, and W. A. Noon. 1988. The rosy region of *Drosophila melanogaster* and D. *simulans*. I. Contrasting levels of naturally-occurring DNA restriction map variation and divergence. Genetics. 119:875-888.

Ar-rushdi, A.H. 1963. The cytology of achiasmatic meiosis in the female *Tigriopus* (Copepoda). Chromosoma. 13:526

Burton, R.S. and S.G. Swisher. 1984. Population structure of the intertidal copepod *Tigriopus californicus* as revealed by field manipulation of allele frequencies. Oecologia. 65:108-111.

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

Burton, R.S. 1987. Differentiation and integration of the genome in populations of the marine copepod *Tigriopus californicus*. Evolution. 41:504-513.

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

Burton, R. S. 1985. Mating system of the intertidal copepod *Tigriopus californicus*. Mar. Biol. 86:247-252.

Burton, R. S., M.W. Feldman, and S.G. Swisher. 1981. Linkage relationships among five enzyme-coding gene loci in the copepod *Tigriopus californicus*: A genetic confirmation of achiasmatic meiosis. Biochemical Genetics. 19:1237-1245.

Burton, R.S, M.W. Feldman, and J.W. Curtsinger. 1979. Population genetics of *Tigriopus californicus* (Copepod: Harpacticoida): I. Population structure along the central California coast. Mar. Ecol. Prog. Ser. 1:29-39.

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

Burton, R.S., E.C. Metz, J.M. Flowers, and C.S. Willett. 2005. Unusual structure of

ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal subrepeats. Gene. 344:105-113.

Burton, R.S. 1990. Hybrid breakdown in developmental time in the copepod *Tigriopus californicus*. Evolution. 44: 1814-1822.

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.

Dybdahl, M. F. 1994. Extinction, recolonization, and the genetic structure of tidepool copepod populations. Evol. Ecol. 8:113-124.

Edmands, S. 1999. Heterosis and outbreeding depression in inter-population crosses spanning a wide range of divergence. Evolution. 53:1757-1768.

Edmands, S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. Molecular Ecology. 10: 1743-1750.

Edmands, S. and J. K. Deimler. 2004. Local adaptation, intrinsic coadaptation and the effects of environmental stress on interpopulation hybrids in the copeod *Tigriopus californicus*. Journal of experimental marine biology and ecology. 303:183-196.

Edmands, S. and J. S. Harrison. 2003. Molecular and quantitative trait variation within and among populations of the intertidal copepod *Tigriopus californicus*. Evolution. 57:2277-2285.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, and A. Di Rienzo. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. 69:831-843.

Ganz, H.H., and R.S. Burton. 1995. Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod *Tigriopus californicus*. Mar. Biol. 123:821-827.

Harris, H., and D.A. Hopkinson. 1976. Handbook of Enzyme Electrophoresis in Human Genetics. North-Holland, Amsterdam.

Hudson, R. R. 1987. Estimating the recombination parameter of a finite population-model without selection. Genetical Research. 50:245-250.

Hudson, R. R.. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7:1-44.

Hudson, R.R. and N.L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 111:147-164.

Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and recombination. Genetics. 120:831-840.

Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. 132:583-589.

Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. In H. N. Munro (ed.) Mammalian Protein Metabolism. Academic Press, New York. pp. 21-132.

Katz, L.A. and R. G. Harrison. 1997. Balancing selection on electrophoretic variation of phosphoglucose isomerase in two species of field cricket: *Gryllus veletis* and *G. pennsylvanicus*. Genetics. 147:609-621.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Bioinformatics Vol. 17. 12:1244-1245.

Lewontin, R. C. 1995. The detection of linkage disequilibrium in molecular sequence data. Genetics. 140:377-388.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. Bottleneck effect and genetic variability in populations. Evolution. 29:1-10.

Neigel, J.E. and J.C. Avise. 1993. Application of a random-walk model to geographic distributions of animal mitochondrial DNA variation. Genetics. 135:1209-1220.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Sys. 23: 263-286.

Palm, S., L. Laikre, P.E. Jorde and N. Ryman. 2003. Effective population size and temporal genetic change in stream resident brown trout (*Salmo trutta*, L.). Conservation genetics. 249-264.

Przeworski, M. and J.D. Wall. 2001. Why is there so little intragenic linkage disequilibrium in humans. Genet. Res. 77:143-151.

Rozas, J., J.C. Sanchez-Delbarrio, X. Messeguer, and R. Rozas. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003. 19: 2496-2497.

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

Slatkin, M. and B. Rannala. 2000. Estimating allele age. Ann. Rev. Genom. and Hum. Genet. 1:225-249.

Swofford, D. L., 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24: 4876-4882.

Vittor, B. A. 1971. Effects of the environment on fitness-related life history characteristics in *Tigriopus californicus*. Ph.D. Diss. Univ. of Oregon, Eugene.

Watterson, G. A. 1975. On the number of segregating sites in genetical models with recombination. Theor. Popul. Biol. 7:256-276.

Willett, C. S. and R. S. Burton. 2004. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. Mol. Biol. Evol. 21:443-453.

Table 1. Oligonucleotide primers used for amplification and sequencing of *Pgi*.

Primer	Sequence
TcPGI-3'UTR-R	AACTAGCGCCATAACGAATAC
TcPGI-5'UTR-F	TCCCGACGAGCAAAATAGCA
TcPGI-0.5-R	GCATCACATCCTTGCCATCCAC
TcPGI-0.5-F	CCAGAAAATCAATTTCACCGAG
TcPGI-2.25-R	CCAATCCCAAAACTCGAACAT
TcPGI-2.25-F	CCTCTGCAGTGGCGAAACAC
TcPGI-3.5-R	CAAGAAGTTACACAGAAGGAG
PA-PGI-2.25F	CCGCCGCGGTGGCGAAACAC
PA-PGI-3.5R	CTAAAATGACCGTTGTGGGTCG
TcPGI-5'-34F	GCTGCAGACGAAATACACTCTC
TcPGI-STOP+21R	AAGTCAAGTCTTGGCGAGTAA

Table 2. *Pgi* allele and genotypic frequencies in *T. californicus* populations with G-test of goodness-of-fit to Hardy-Weinberg proportions.

			Allele frequencies			Genotypic frequencies						
Site	Date	N	F	M	S	F/F	M/F	M/M	F/S	M/S	S/S	*P*
Mussel Rock[a]	12/28/04	38	-	.987	-	-	-	.974	-	-	-	n.s.
Pomponio[a]	12/28/04	36	-	.972	-	-	-	.944	-	-	-	n.s.
Pescadero[b]	4/78	50	.52	.48	-	.32	.40	.28	-	-	-	n.s.
	6/78	84	.375	.613	.012	.143	.464	.381	-	-	0.012	.013
	8/78	104	.5	.5	-	.26	.48	.26	-	-	-	n.s.
	8/79	130	.485	.515	-	.223	.523	.254	-	-	-	n.s.
	12/2/81	30	.483	.517	-	-	-	-	-	-	-	n.s.
	8/27/82	83	.515	.466	.015	-	-	-	-	-	-	n.s.
	5/10/83	97	.459	.536	.005	.268	.381	.340	-	.010	-	n.s.
	8/95	66	.38	.62	-	.121	.515	.364	-	-	-	n.s.
	7/96	70	.51	.49	.014	.286	.414	.271	.029	-	-	n.s.
	7/15/04	46	.511	.489	-	.348	.326	.326	-	-	-	.017
	12/28/04	40	.488	.513	-	.25	.475	.275	-	-	-	n.s.

Table 2. Continued.

Site	Date	N	Allele frequencies			Genotypic frequencies						*P*
			F	M	S	F/F	M/F	M/M	F/S	M/S	S/S	
Site 10[b]	8/79	70	-	.864	.136	-	-	.729	-	.271	-	n.s.
	8/95	66	.04	.84	.12	-	.076	.682	-	.242	-	n.s.
	7/15/04	48	-	.969	.031	-	-	.938	-	.063	-	n.s.
San Diego[b]	10/87	>35	.11	.89	-	-	-	-	-	-	-	n.s.
	?[c]	>50	.22	.78	-	-	-	-	-	-	-	n.s.
	9/11/04	47	.202	.798	-	.043	.319	.638	-	-	-	n.s.
	9/19/04	24	.146	.854	-	.042	.208	.750	-	-	-	n.s.
Punta Morro[b]	?[c]	>50	.31	.69	-	-	-	-	-	-	-	n.s.

[a]Samples from Mussel Rock and Pomponio contained one and two Pgi^{M}/Pgi^{MS} heterozygotes, respectively. Pgi^{MS} alleles were not observed at any other location listed in this table.
[b]Data from dates prior to 2004 are reported in Burton et al. (1979), Burton and Feldman (1981), Burton and Swisher (1984), Ganz and Burton (1995), and Burton (1997), or are unpublished.
[c]Samples collected between November 1992 and April 1994 (Ganz and Burton 1995)

Table 3. Non-singleton polymorphisms from three sites in central California. The polymorphism responsible for the *Pgi*^{F} allele is indicated in bold. The boxed region indicates the haplotype block that is conserved in all *Pgi*^{F} alleles. The electrophoretic allele class, *Pgi*^{F} or *Pgi*^{M}, of each sequence is indicated by an M or an F next to the sample ID. Sample abbreviation are PES = Pescadero, S10 = Site 10, and SCN = Santa Cruz. Singleton sites have been omitted from the alignment. Numbers at the top of the table are the positions in the global alignment of all nine populations in this study.

ID	Polymorphic sites			
	111	1111111111	1111222222	
	122333556	7788999012	2566777778	8999013345
	3634444689	2967138861	2158122995	5228691503
	0562056686	6215150282	4824518693	6271908582
PES39M	AGACTCTGCA	ATCAAGTTTT	CCATGTCTCA	TGGAGCCTCT
PES21M	....	....	....	......TAAC
PES31M	....ATC.GG	GGTG.C.C..	....	......TAAC
PES29M	....	....	....	....
PES28M	GACGA...GG	GGTG.C.C..	.T.CACACT.	.AAT.TTAAC
PES46M	GACGA...GG	GGTG.C.C..	...CACACTT	GAAT.TTAAC
PES13M	....ATC...	.......CC.	.T.CACACTT	GAAT.TTAAC
PES4F	..**A**TC..	.......CC.	.T.CACACTT	GAAT.TTAAC
PES24F	..**A**TC..	.......C..	....	......TAAC
PES41F	..**A**TC..	.......C..	....	......TAAC
PES10F	..**A**TC..	.......C..	....	......TAAC
PES2F	GA**A**TC..	GGTG.C.C..	....	....
PES42F	..**A**TC..	.......C..	.TGCACACTT	GAAT..TAAC
S1050M	.ACGA...GG	GGTG.C.C..	....	....T.....
S1051M	....	.......CC.	.T.CACA...	......TAAC
S1057*	.ACGA...GG	GGTG.C.C..	..G.......	....T.....
S1027*	.A..ATC...	.......C..	.T.CACACTT	GAAT.TTAAC
S1028M	.ACGA...GG	GGTG.C.C..	....	....T.....
SCN109M	.A..ATC.GG	GGT.GCGCCC	GT.CACACTT	GAAT.TTAAC
SCN119M	....ATC.GG	GG.....CC.	.T.CACACTT	GAAT.TTAAC
SCN120M	....ATC.GG	GG...CGCCC	GT.C.CA..T	GAAT.A...C
SCN110M	.A..ATC.GG	.GT.GC.C..	....	C
SCN118M	...G....GG	GGTG.C.C..	...CACACTT	.AAT.T...C

*Samples from Site 10, S10 27 and 57, were *Pgi*^{M}/*Pgi*^{S} heterozygotes, but the *Pgi*^{S} allele class could not be unambiguously characterized at the sequence level.

Table 4. Non-singleton polymorphisms from San Diego (SD) and Laguna Beach (LB). The polymorphism responsible for the Pgi^F allele is indicated in bold. The electrophoretic allele class, Pgi^F, Pgi^M or Pgi^S, of each sequence is indicated by an F, M, or S next to the sample ID. Singleton sites have been omitted from the alignment. Numbers at the top of the table are the positions in the global alignment of all populations.

ID	Polymorphic sites		
		11111111	111122222
	34444455	8801334557	789933345
	2671778849	2316126030	704714626
	1329781211	2719193602	520784525
SD98M	GCAGCTCAGG	GAAGGTTTCA	TGACCTATC
SD99M	....	....	....
SD100M	ATC.GAGGAC	TG.CACCA.G	.TG...GCA
SD101M	....	....	....
SD102M	....	....	........A
SD103M	....	....	......GCA
SD106M	AT......A.	TG.CACCA.G	........A
SD107F	...**A**.....	TG.CACCAT.	.TG...GCA
SD108M	....	...CACCAT.	.TG...GCA
SD109F	AT.**A**.....	....	........A
SD110M	AT........	....	........A
SD112M	....	....	........A
SD113M	ATC.GAGGAC	TGGCACCA.G	.TG...GCA
SD118M	A.C.GAGGA.	..G.......	........G
LB4M	..C.GAGGA.	TG.CACCA.G	G.GTTG..A
LB17S	..C.GAGGA.	T..CACCA.G	G.GTTG..A

Table 5. Summary of DNA polymorphism at *Pgi* in two populations of *T. californicus*

		Nucleotide diversity					
	n	π_{nonsyn}	π_{syn}	$\pi_{tot\text{-}sil}$	θ_W[a]	Tajima's D[b]	Fu and Li's D[b]
Pescadero Pgi^F	6	.0026	.0120	.0114	.0076	-0.7121	-0.6452
Pescadero Pgi^M	7	.0029	.0209	.0168	.0091	0.3019	0.2045
Pescadero	13	.0032	.0168	.0145	.0096	-.5062	-.9147
San Diego	14	.0023	.0137	.0113	.0084	-1.0134	-1.4249

[a]θ_W is the estimator of the population mutation parameter θ based upon the total number of segregating sites (Watterson 1975)
[b]All values are non-significant

Table 6. Inter-population distances at *Pgi* based on coding regions corrected by the method of Jukes and Cantor (1969). Introns and a small fragment of the 3'UTR were excluded due to alignment ambiguities among some populations. Distances are net between-population means (i.e., corrected for within population distances) as calculated in MEGA version 2.1 (Kumar et al. 2001).

	PA	PM	SD	LB	AB	CA	SCN	PES
PA	-							
PM	0.084	-						
SD	0.082	0.021	-					
LB	0.083	0.023	0.003	-				
AB	0.083	0.026	0.019	0.02	-			
CA	0.080	0.024	0.02	0.019	0.02	-		
SCN	0.078	0.02	0.015	0.015	0.016	0.014	-	
PES	0.080	0.02	0.016	0.015	0.017	0.015	0.002	-

Table 7. Test of reduced variation in the Pgi^F allele class at Pescadero in a window of 520 bp flanking the Arg/Gln polymorphism. *P* values were obtained from coalescent simulation with θ_W or π values estimated from 520 bp flanking the Q/R polymorphism in Pescadero (n = 13). The *P* values represent the probability of drawing six alleles from a population at neutral equilibrium with the π, θ, or Tajima's D estimated from the Pgi^F allele class. Probabilities are presented assuming different levels of recombination, *C*, the population recombination parameter. *C* was estimated from the entire 2.5 kb region for all *Pgi* alleles sampled from Pescadero (Hudson and Kaplan 1985).

	Probability		
Recombination rate[a]	π	θ_W	Tajima's D
No recombination	0.0786	0.1054	0.089
C/10	0.0468	0.0726	0.0578
C/8	0.0458	0.0572	0.0437
C/6	0.0407	0.0596	0.04
C/4	0.0326	0.0477	0.0316
C/2	0.0231	0.0291	0.0202
C (= 41.8)[a]	0.0123	0.0185	0.0111

[a]In *T. californicus*, $C = (1/2)4N_e c$ because of the absence of recombination females (Ar-rushdi 1963, Burton et al. 1981)

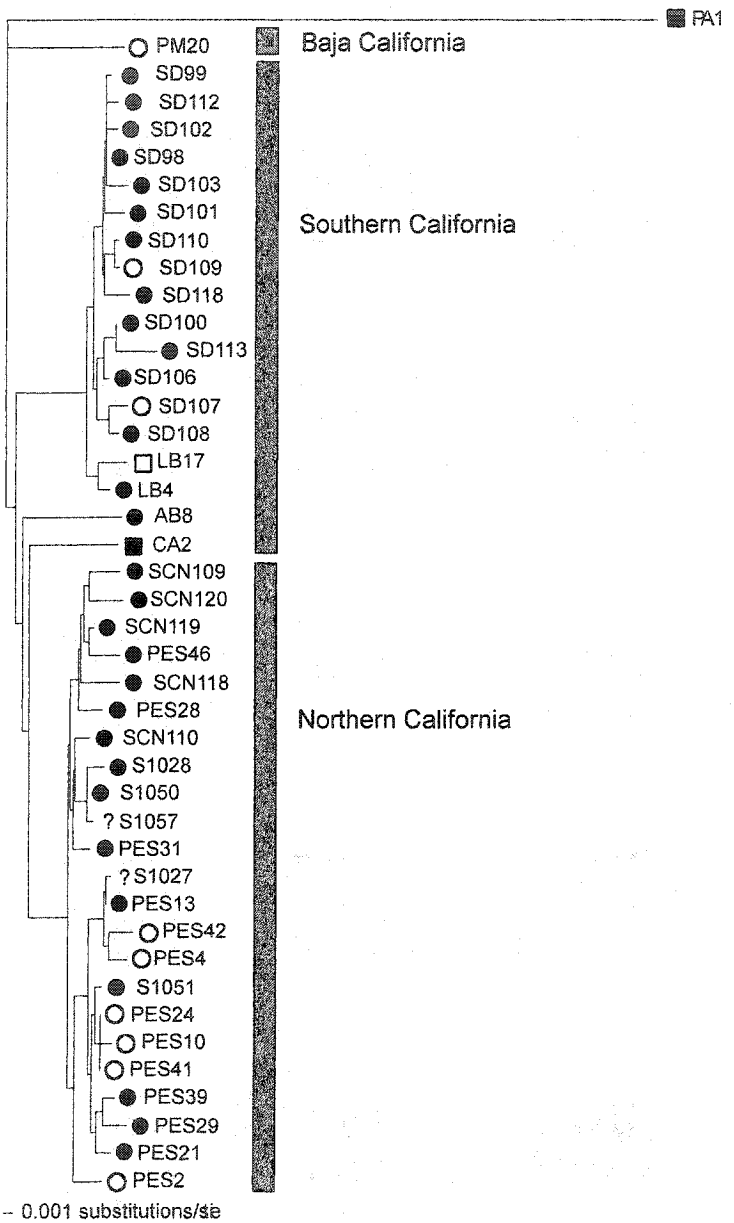


Figure 1. Neighbor-joining tree based upon 1,674 bp of coding region sequence from *Pgi*. Allozyme allele classes are indicated by the following shapes: □ = Pgi^{S}, ■ = Pgi^{MS}, ● = Pgi^{M}, ○ = Pgi^{F}

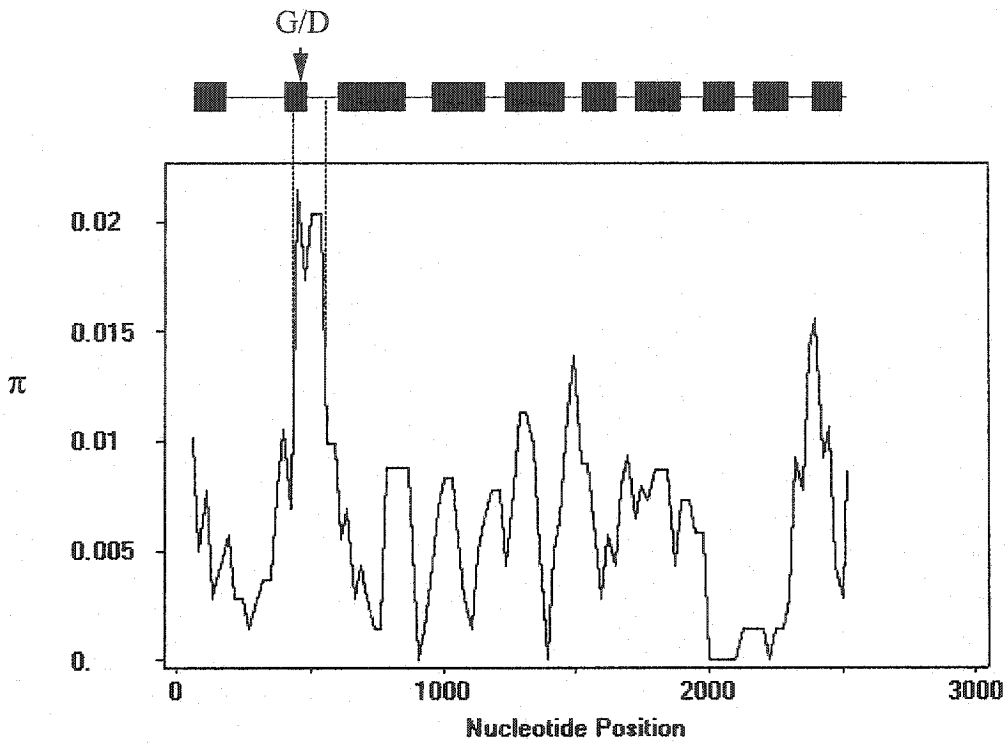


Figure 2. Sliding window of nucleotide diversity (π) from both Pgi^{M} and Pgi^{F} alleles in San Diego. The arrow indicates the Gly/Asp polymorphism responsible for the Pgi^{M} and Pgi^{F} alleles in this population. Window length is 100 bp and the step size is 25 bp.

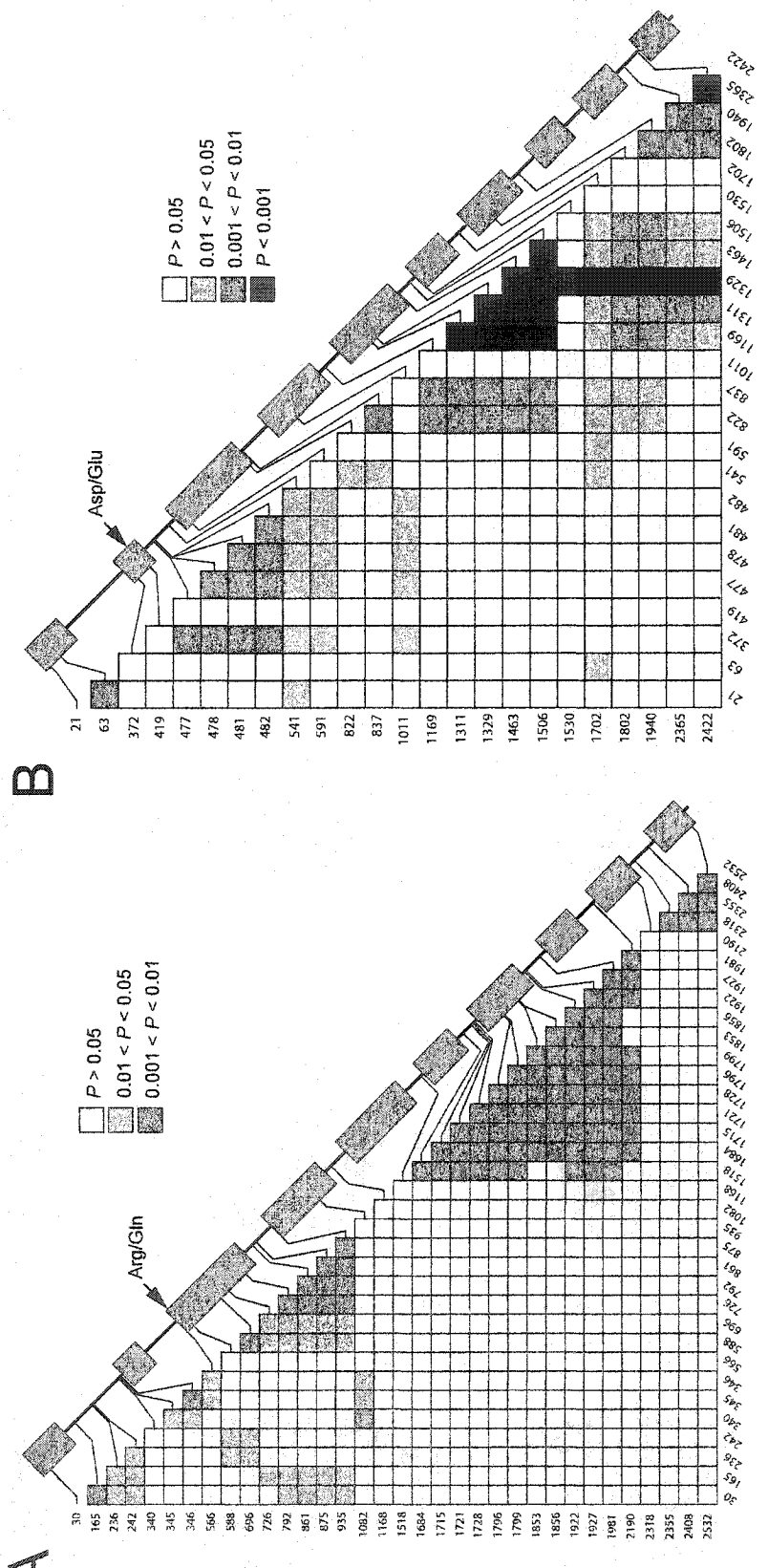


Figure. 3. Linkage disequilibrium plots from (A) Pescadero and (B) San Diego populations. Shaded boxes represent significant pairwise associations between sites that are significant by Fisher's Exact test. Thick bars on the *Pgi* cartoon represent exons and thin lines represent introns. The arrows indicate the charge-changing polymorphisms in San Diego and Pescadero populations.

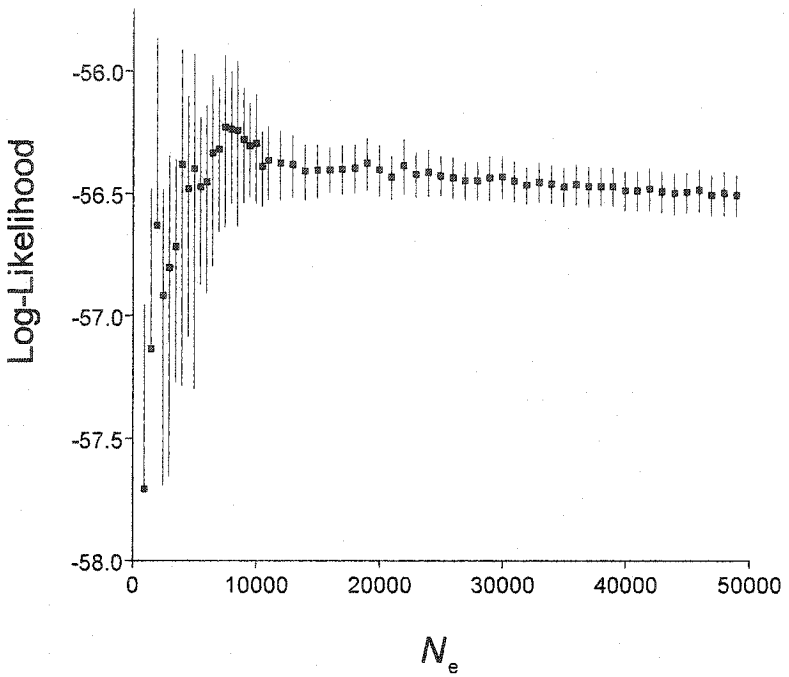


Figure 4. Log-likelihood curve of N_e based on temporal changes in allozyme allele frequencies at *Pgi* in Pescadero. Error bars are 95% confidence intervals (see Materials and Methods).

CHAPTER III

Ribosomal RNA gene silencing in inter-population hybrids: Nucleolar dominance in the absence of inter-genic spacer subrepeats

ABSTRACT

A common feature of inter-specific animal and plant hybrids is the uniparental silencing of ribosomal RNA gene transcription, or nucleolar dominance. The enhancer-imbalance model is a leading hypothesis for nucleolar dominance that was formulated from observations of hybrids between *Xenopus laevis* and *X. borealis*. In F_1 hybrids, dominance of *X. laevis* rRNA genes has been attributed to a greater number of enhancer-bearing subrepeats in the inter-genic spacer. These repeat elements are hypothesized to sequester limiting transcription factors and result in silencing of *X. borealis* genes. These repeat units are found in virtually all animal and plant species, but are absent in the copepod *Tigriopus californicus*. We examined rRNA gene expression in inter-population hybrids of *T. californicus* to determine if nucleolar dominance occurs in the absence of internal subrepeats. We report that nucleolar dominance does occur in F_1 inter-population hybrids. This is significant because it indicates that nucleolar dominance can be established and maintained without enhancer-bearing repeat elements in the inter-genic spacer and challenges the relevance of the enhancer-imbalance model for nucleolar dominance. F_2 hybrids show both dominant and codominant expression of the rRNA genes. The observed pattern of transcriptional bias in the F_2 generation is inconsistent with predictions of the enhancer-imbalance model.

INTRODUCTION

A common pattern of altered gene expression in inter-specific hybrids is the uniparental transcriptional silencing of ribosomal RNA (rRNA) genes. In non-hybrid diploid genomes, rRNA gene clusters on homologous chromosomes are both expressed, but in inter-specific hybrids, transcription of the array from one parent is frequently silenced (e.g., Blackler and Gecking 1972). This pattern of expression, called nucleolar dominance, is a hybrid phenomenon that has been documented in a large fraction of inter-specific plant and animal hybridizations (Reeder 1985, Pikaard 2000ab, Grummt and Pikaard 2003). Despite the widespread occurrence of nucleolar dominance little is known about how the pattern of dominance is established or why it occurs.

The rRNA gene family in plants and animals consists of tandemly arrayed repeats of the 18S, 5.8S, and 28S structural genes (Long and Dawid 1980). Each repeat unit consists of all three structural genes and the *cis*-regulatory elements necessary for transcription initiation and termination. Regulatory features are located in the inter-genic spacer (IGS) and typically include a promotor and arrays of internal subrepeats that act as enhancers of transcription (Grimaldi and Di Nocera 1988, Kuhn et al. 1990, Pape et al. 1989). Transcription of the ribosomal genes is accomplished by RNA polymerase I (Pol I). This polymerase and its associated transcription factors are dedicated solely to transcription of the rRNA genes and promotor recognition is specific to the promotor in the IGS (Grummt 2003). This dedication of Pol I transcription apparatus to a single family of genes may facilitate rapid coevolution of

components of the Pol I transcription apparatus to the rRNA gene promotor (Grummt et al. 1982). Tests of this hypothesis have demonstrated that the RNA Pol I transcription apparatus is not inter-changeable among distantly related species such as mouse and human (Grummt et al. 1982) and *Drosophila virilis* and *D. melanogaster* (Kohorn and Rae 1982). In most cases, it is unclear how this unusual feature of the rRNA transcription apparatus may contribute to nucleolar dominance (Reeder and Roan 1984, Miesfeld et al. 1984, Reeder 1985, Frieman et al. 1999).

Studies of inter-specific hybrids between *Xenopus laevis* and *X. borealis* have suggested a mechanism for nucleolar dominance. In hybrids, *X. laevis* rRNA genes are transcriptionally dominant over *X. borealis* (Honjo and Reeder 1973). Transcriptional silencing of *X. borealis* genes has been hypothesized to result from the ability of the *X. laevis* cluster to more effectively sequester limiting sets of transcription factors (Reeder and Roan 1984). Support for this model has come from transcription assays with artificial minigene constructs. When constructs bearing *X. laevis* regulatory elements are co-injected into oocytes with constructs containing *X. borealis* regulatory features, the *X. laevis* construct shows transcriptional dominance, thus mimicking the situation observed in inter-specific hybrids. These assays have suggested that transcriptional dominance could be conferred to *X. laevis* genes as a result of four-fold more enhancers (Reeder and Roan 1984) or stronger spacer promotors (i.e., repeated promotor-like sequences necessary for full enhancer function in *Xenopus*; Caudy and Pikaard 2002) in the *X. laevis* IGS. Either of these

mechanisms could deprive the *X. borealis* genes of transcription factors and result in transcriptional dominance of *X. laevis* arrays.

This enhancer-imbalance model is attractive because it explains all general features of the phenomenon of nucleolar dominance and simultaneously explains how transcriptional dominance is established and maintained through subsequent cell divisions (Reeder 1985). However, this model has recently come under scrutiny because of evidence that enhancer-imbalance may not contribute to nucleolar dominance in other hybrids (Frieman et al. 1999, Chen et al. 1998). Studies of allotetraploid hybrids in *Brassica* and *Arabidopsis* have established that modifications to chromatin structure are at least partially responsible for rRNA gene silencing. Nucleolar dominance in these species appears to be an epigenetic phenomenon that is inherited through chromatin modifications that are introduced by a self-reinforcing cycle of cytosine methylation and histone deacetylation (Chen and Pikaard 1997, Lawrence et al. 2004). Although it is not clear how the repressive chromatin state of the under-dominant rRNA genes is originally specified, modifications to chromatin structure account for how transcriptional silencing occurs and how the transcriptional status of genes can be transmitted through subsequent cell divisions.

The importance of chromatin configuration raises questions about the relationship between the nucleolar dominance-like pattern observed in assays with artificial minigene constructs in *Xenopus* and the phenomenon of nucleolar dominance. It may be that the the enhancer-imbalance mechanism at work in transcription assays does not contribute to nucleolar dominance in the chromosomal

context of a hybrid nucleus (Caudy and Pikaard 2002). If so, this may indicate that enhancer-bearing subrepeat elements, which are thought to be important in nucleolar dominance in *Xenopus*, may not be necessary for either the establishment or maintenance of nucleolar dominance.

An unusual feature of the rRNA gene array in the marine copepod *Tigriopus californicus* provides a means for assessing the role of subrepeat elements in nucleolar dominance. In a recent survey of genetic variation in the rRNA gene cluster, we found substantial inter-population divergence in the most conserved section of the IGS and the external transcribed spacer (ETS, Burton et al. 2005). Uncorrected inter-population divergences in this region were found to exceed 10% among all populations and exceeded 35% between northern populations and a partially reproductively isolated population in Baja California, Mexico (Ganz and Burton 1995). In addition to extensive IGS divergences, the region upstream of the transcription initiation site does not contain subrepeats (Burton et al. 2005). In other species, these subrepeats are known to contain enhancers of transcription and are otherwise a characteristic feature of animal IGS sequences (Reeder 1990). Because these repeats are hypothesized to contribute to nuclear dominance (Reeder and Roan 1984), studies of ribosomal gene expression in *T. californicus* hybrids may provide new insight into the mechanism(s) responsible for uniparental silencing of ribosomal RNA genes in hybrids.

Here we report patterns of ribosomal gene expression in inter-population hybrids of *T. californicus*. We assess whether transcription of rRNA is biased in F_1

and F_2 hybrids and demonstrate a clear pattern of nucleolar dominance in the F_1 generation. However, in F_2 hybrids nucleolar dominance appears to partially break down as some hybrids show dominant and others codominant expression of rRNA genes. Given the unusual structure of the *T. californicus* IGS, these results imply that repetitive elements in the IGS are not required for uniparental gene silencing of rRNA transcription. Furthermore, the pattern of codominant expression in relation to rRNA gene dosage in some F_2 hybrids is inconsistent with a simple transcription factor titration model for nucleolar dominance.

MATERIALS AND METHODS

T. californicus is an obligately sexual harpacticoid copepod species that inhabits intertidal rocky shelves along the Pacific Coast of North America. Limited gene flow occurs among populations of *T. californicus* and populations are frequently divergent at mitochondrial and nuclear gene loci (e.g., Burton and Feldman 1981, Burton and Lee 1994, Edmands 2001). In laboratory crosses, F_1 inter-population hybrids are vigorous, while F_2 hybrids suffer from outbreeding depression (Burton 1990, Edmands 1999, Edmands et al. 2005). The mechanism of sex determination is unknown in *T. californicus*, although it appears to be influenced by both genetic and environmental factors (Voordouw and Anholt 2002). There is no evidence of a sex chromosome in this species (Ar-rushdi 1964).

Samples were collected from Abalone Cove, Los Angeles County (AB), San Diego (SD) and Santa Cruz (SC), California, USA and maintained in laboratory cultures at 20° C with a 12:12 light-dark cycle. Collecting locales are the same as those in previous studies (e.g., Burton and Feldman 1981, Burton and Lee 1994, Ganz and Burton 1995). Inter-population hybrids were obtained from crosses between AB X SD and SD X SC populations and their reciprocals. Pair crosses were established following the method of Burton and Feldman (1981) by isolation of virgin females and mating them in mass culture with males from an alternate population. F_1 hybrid males were either isolated for DNA and RNA extraction or allowed to mate with F_1 females for production of F_2 inter-cross hybrids. Second generation hybrids were subsequently

obtained by transferring fertilized F_1 females to new cultures for production of F_2 progeny. All experiments were conducted on F_1 and F_2 adult males.

Detection of population-specific rRNA genes and rRNA transcripts in inter-population hybrids is limited in *T. californicus* by low levels of among-population divergence in the 18S and 28S structural genes (Burton et al. 2005). Among the focal populations, a single nucleotide polymorphism (SNP) at position 189 of the 18S is fixed for cytosine in SD and thymine in AB and SC populations (see Genbank Accessions: AY588143.1, AY588134.1, and AY588140.1; Burton et al. 2005). The AB and SC populations are indistinguishable in the 18S and 28S rRNA genes. This restricted our analysis to SD X AB and SD X SC inter-population hybrids.

We developed an assay that distinguishes rRNA genes from these populations based upon the T189C SNP. The method described below is a post-PCR assay that allowed us to determine if both SNP variants were present in DNA or cDNA pools derived from a single copepod and quantification of the relative proportions of the SNP variants in each pool. The method is based on the loss of fluorescence associated with the melting of fluorescence resonance energy transfer hybridization (FRET) probes from a template during melting curve analysis of PCR amplified products. This method is frequently used in SNP genotyping (Bernard et al. 1998) and has also been used to quantify the relative frequeny of SNP variants in a PCR product pool (Lyon et al. 2001).

SNP detection with this FRET method capitalizes on melting temperature differences between the homo- and heteroduplexes formed between an acceptor probe

and two SNP variants in a pool of PCR-amplified DNA template. The acceptor probe hybridizes to the SNP and flanking sites. A donor probe, with substantially higher melting temperature than the acceptor, hybridizes to the template immediately 5' of the acceptor. When both probes are hybridized to the template, excitation of a 3' fluorescein on the donor probe initiates a resonance energy transfer to a 5' fluorophore on the adjacent acceptor probe. This results in a fluorescent emission. Because the homo- and hetero-duplexes formed between the two SNP variants and the acceptor probe have different thermal stabilities, each polymorphism may be identified by the loss of fluorescence associated with the melting of the acceptor probe from each SNP variant at different temperatures. The hetero-duplex formed between the acceptor and the mismatched SNP variant is expected to have a lower melting temperature compared with the corresponding homo-duplex. Once the acceptor probe melts from the template, FRET between it and the donor probe cannot occur which results in a loss of fluorescence as temperature is increased. Fluorescence emission is monitored in a melting curve analysis in which the sample is heated from a low temperature in which both probes are hybridized to a temperature in which both probes are melted from the template.

RNA and DNA were simultaneously isolated from single adult copepods as follows. Individual animals (approximately 30 µg wet weight) were submersed live in 40 µl of TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA) with approximately 150 mg of zirconia/silica beads (Biospec, Bartlesville, OK, USA) in 500 µl RNase/DNase free microtubes. Cells were disrupted by bead-beating each sample for 20 s in a

FastPrep® 120 bead beater (Qbiogene, Carlsbad, CA, USA) with a rotor speed of 4.5 m/s. Total RNA and DNA was subsequently extracted following the TRI Reagent extraction protocol. RNA was DNase treated and first-strand cDNA synthesis was performed with Superscript II (Invitrogen, Carlsbad, CA, USA) with an 18S rRNA gene specific primer, 5'-GTGCGATCAGCAAAAAGT-3'. A 200 bp fragment containing the T189C polymorphism was PCR amplified with the primer used for cDNA synthesis and a forward primer, 5'-AGGTGAAGCTGCGAAC-3'. The cycling profile consisted of 95° 30 s, 50° 30 s, and 72° 30 s for 40 cycles in a conventional thermal cycler. Identical conditions were used to amplify genomic DNA. PCR products were visualized with ethidium bromide-stained agarose gels and quantified with Hoechst Dye on a fluorescence plate reader. PCR controls for contamination of RNA with genomic DNA were conducted for all samples. In this control, DNase-treated RNA was added directly to a PCR reaction. Samples were considered free of contaminating DNA if PCR amplification failed. In some cases, contaminating genomic DNA was detected and a second DNase treatment was conducted. These samples were checked a second time for DNA contamination prior to cDNA synthesis.

The presence or absence of each SNP variant in genomic DNA or cDNA pools was determined with a post-PCR melting curve analysis using FRET hybridization probes. The pair of probes was designed to hybridize to the T189C SNP and the region immediately adjacent to the polymorphic site. The donor probe (5'-GCGGTAATTCTGGAGCTAATACATGCTACAA-3') was end-labeled with a 3' fluorescein. The acceptor probe (5'-CCCTTCACGTTGTGTGAGGG-3') was

positioned two bp downstream of the 3' end of the donor probe and was end-labeled with a 5' LC Red 640 and a 3' phosphate group (Synthegen, Houston, TX, USA). A cytosine located five bp from the 5' end of the acceptor probe produces a C:A heteroduplex when hybridized with SD-derived DNA and a C:G base pair with SC- and AB-derived DNA.

Melting curves were generated in a LightCycler™ (Roche, Mannheim, Germany) with equal concentrations of unpurified PCR-amplified DNA, 3.25 mM $MgCl_2$, and 300 nM of each probe in glass capillaries (Roche Molecular Biochemicals, Indianapolis, IN,USA). Each assay consisted of an initial denaturation step of 0 s at 95° C, then annealing at 45°C for 15 s followed by ramping of the reaction to 80° C, at rates between 0.2° C/s to 0.5° C/s. Fluorescence was detected continuously and the instantaneous rate of fluorescence change was calculated by taking the negative derivative of the melting curve. When the negative derivative is plotted versus temperature, samples will yield one or two distinct peaks depending on whether one or both SNP variants were present in the sample. In our assay, we found that this method readily distinguishes between the T189C SNP variants. In such melting peak profiles, these peaks correspond to the temperatures at which there is a maximum instantaneous rate of fluorescence loss due to melting of the acceptor probe from homo- and heteroduplexes. Control samples containing one SNP variant invariably yielded a single peak, while samples with both SNP variants yielded two distinct peaks.

We also established that the relative rate of fluorescence loss at the melting points of the acceptor probe from each SNP variant is linearly correlated with the

relative frequency of the SNP variants in the template pool (see Lyon et al. (2001) for a related method). This was determined by generating melting peak profiles of samples with known ratios of the T189C SNP variants. Genomic DNA from SD and SC populations was amplified and the PCR products combined in different ratios. Melting peak profiles of 5:1, 2:1, 1:1, 1:2, and 1:5 ratios yielded two distinct peaks for each ratio. We observed a tight linear correlation between these ratios (expressed as SD/(SD+SC)) and the ratios of the peak heights. This relationship facilitated quantification of the relative proportions of the SNP variants in a pool of amplified products. The ratios listed above were chosen because more skewed ratios frequently failed to yield two distinct melting peaks (not shown). In a few samples, two peaks were identified in the melting peak profile, but the inferred ratios of the SNP variants fell outside of the range of the standard curve. Ratios from these samples were calculated by extrapolating from the standard curve. This was done to indicate that both SNP variants were detected in such samples and to distinguish them from samples in which we found no evidence of both SNP variants. The frequency of the rarer SNP variant in these samples was inferred to be less than 1/6.

RESULTS

The polymorphisms derived from the source populations were clearly distinguished in the melting curve assay. The melting temperatures of the acceptor probe from the targets were 60-62 °C and approximately 67-69° C, for the SD (T:C heteroduplex) and SC/AB (G:C homoduplex) polymorphisms, respectively. This can be seen in a plot of the negative derivative of the melting curve ("melting peak profile", Fig. 1), where the peak at approximately 60° C corresponds to the SNP variant from the SD population and the peak at approximately 68° C corresponds to SNP variant from the SC population. The relative peak height (i.e., the relative instantaneous rate of fluorescence loss due to acceptor probe melting) from each polymorphism was linearly correlated to the frequency of the two SNP variants in a sample (Fig. 1). This relationship allowed us to determine the relative frequency of the 18S T189C SNP variants in PCR products amplified from genomic DNA or cDNA obtained from single hybrid copepods.

F_1 hybrids

We assayed a total of 39 F_1 hybrid males from two inter-population crosses (SD X AB and SD X SC) and their reciprocals. Genomic DNA and cDNA corresponding to the region of the 18S structural gene flanking the T189C SNP was amplified from each animal. The relative proportion of the T189C SNP variants in genomic DNA and the rRNA transcript pool was then inferred from the FRET melting-curve profile. If ribosomal genes are transcribed in a codominant fashion in

inter-population hybrids, we expect the ratio of population-specific 18S rRNA in the transcript pool to match the ratio of population-specific 18S rDNA in the genome.

The expectation of codominant transcription in F_1 hybrids was not met in either cross or their reciprocals. rRNA genes from each parental population were identified in genomic DNA of F_1 hybrids as expected (i.e., two peaks were present in melting curve plots of 18S rDNA amplified from genomic DNA, Figure 2A). However, analysis of amplified cDNA revealed a single peak that corresponded to either the AB or SC polymorphism in the majority of samples (Figure 2A). In a few cases, two peaks were identified in F_1 hybrids, which indicated the presence of SD transcripts. However, in all cases there was a strong bias towards the SNP variant from the AB or SC population.

Figure 3 contains plots of the ratios of population-specific rRNA and population-specific rDNA in F_1 hybrids, inferred from cDNA and genomic DNA respectively. In each plot, ratios of either rRNA or rDNA are expressed as SD/(SD+SC) or SD/(SD+AB). Each point represents rRNA and rDNA ratios from a single F_1 hybrid individual. The codominant expectation in these plots is indicated by a dotted line. Points above or below the line suggest dominance in rRNA gene expression.

With one exception, the rDNA ratio in F_1 hybrids was greater than zero and less than one. This was expected due to inheritance of rRNA genes from each parental population. Since ribosomal gene copy numbers may differ within or among populations (as yet undetermined), we do not necessarily expect a 1:1 ratio of rDNA in

the genomes of F_1 hybrids. In the cDNA pool, the majority of individuals have an inferred rRNA ratio of zero because SD-derived cDNA was not detected in most samples. Furthermore, this pattern was similar in the reciprocals of both crosses (Fig. 3A,3B). These results indicate that (1) there is no evidence of codominance, but rather a strong bias in the rRNA transcript pool in F_1 hybrids, (2) the dominance relationship of AB over SD and SC over SD is consistent among individuals, (3) and within a cross, the dominance relationship is independent of the population origin of the maternal parent.

F_2 hybrids

Sixteen F_2 adult hybrid males were assayed from the ABf X SDm cross and 20 were assayed from the SDm X SCf cross in the same fashion as the F_1 hybrids. In some F_2 hybrids the pattern of expression was indistinguishable from that observed in the F_1, while other F_2 hybrids showed notable differences from the F_1. In the former case, melting peak profiles from some F_2 hybrids showed a clear pattern of dominance. These individuals possessed rRNA genes from both populations, but rRNA from only one population was detected in the transcript pool. In these hybrids, the dominance relationship among populations was the same as observed in the F_1.

Some F_2 hybrids from the SD X SC cross showed a pattern suggestive of codominant expression. In these hybrids, rRNA genes and rRNA transcripts from both populations were present in approximately equal proportions as inferred from the melting peak profile (e.g., Fig. 2C). These hybrids have rRNA and rDNA ratios

greater than zero and less than one and fall along the diagonal line in Fig. 4B. There is no obvious relationship between rDNA genotype and codominant expression and hybrids with SD/(SD + SC) rDNA ratios ranging from less than 1/6 to greater than 5/6 may show codominance (Fig. 4B).

We identified 12 F_2 hybrids in the AB X SD cross and 2 in the SD X SC which were inferred to have rDNA from only one population. This could be attributable to recovery of a parental genotype in the F_2, but may also be due to a limitiation of the technique to detect low frequency (i.e., less than 1/6) SNP variants. Hybrids in which rRNA genes from only one population were identified were also found to express only the corresponding rRNA (Fig. 2B, 2D). In one SD X SC F_2 hybrid, we were unable to detect the presence of SC ribosomal genes, but we detected both SD and SC SNP variants in the cDNA. If SC genes were present at low frequency (e.g., as a result of segregation of two or more unlinked arrays, see discussion) then they may not have been detected in our assay.

DISCUSSION

The widespread occurrence of nucleolar dominance in inter-specific hybrids, but not non-hybrid organisms, indicates that uniparental silencing of ribosomal RNA is due to an emergent property of the hybrid nucleus. In theory, nucleolar dominance could be a functional response to cellular conditions not encountered in normal non-hybrid cells. If this is the case, transcriptional silencing of entire ribosomal RNA gene clusters could be a means of down-regulating rRNA transcription (Pikaard 2000ab). However, normal non-hybrid cells do not regulate transcription in this fashion (Grummt and Pikaard 2003). Therefore, it is unclear what property of the hybrid nucleus should cause hybrid cells to engage in such a large-scale alteration in the number of genes available for rRNA transcription.

Alternatively, nucleolar dominance may be a by-product of an interaction between rRNA gene arrays that differ in some fundamental property (e.g., the ability to sequester limiting sets of transcription factors). In *Xenopus* hybrids, *X. laevis* genes are transcriptionally dominant to *X. borealis* genes (Honjo and Reeder 1973). Dominance of the *X. laevis* cluster has been attributed to four-fold more enhancers in the *X. laevis* IGS and/or higher activity of the *X. laevis* spacer promotor (Reeder and Roan 1984, Caudy and Pikaard 2002). Either of these mechanisms could limit the availability of transcription factors to the underdominant cluster and result in silencing of *X. borealis* genes. This enhancer-imbalance model is attractive because it accounts for the absence of maternal effects on the pattern of dominance (Honjo and Reeder 1973), why nucleolar dominance is observed in hybrid nuclei (Blackler and Gecking

1972), and also provides a mechanism for the establishment of dominance early in development (Reeder and Roan 1984, Reeder 1985). Importantly, much of what is known about nucleolar dominance in *Xenopus* has come from competition assays between minigene constructs. When constructs with *X. laevis* regulatory elements are co-injected into oocytes with constructs containing *X. borealis* regulatory sequences, the pattern of transcripton mimics the dominance relationship observed in hybrid organisms (i.e., minigenes containing the *X. laevis* spacer are dominant to constructs with the *X. borealis* spacer). While these assays have clearly established that transcriptional dominance could be determined by competition for limiting transcription factors, definitive evidence for the action of enhancer-imbalance in a chromosomal context has not been established in hybrids of *Xenopus* or any other organism (Caudy and Pikaard 2002).

Is the *Xenopus* model consistent with observations from hybrid organisms? Tests of the enhancer-imbalance hypothesis have suggested that nucleolar dominance in plants is unlikely to be attributable to competitive differences in the ability of rRNA gene clusters to sequester transcription factors (Frieman et al. 1999, Chen et al. 1998, Pikaard 2000ab). In animals, there is insufficient data to determine whether the *Xenopus* model could provide a general explanation for nucleolar dominance in animal hybrids. F_1 hybrids between *D. melanogaster* and *D. simulans* show some similarities to the situation observed in *Xenopus*. For example, while *D. simulans* genes are silenced in female F_1 hybrids (Durica and Krider 1977), hybrid females with large deletions in the dominant *D. melanogaster* rRNA gene cluster express both *D.*

melanogaster and *D. simulans* genes (Goodrich-Young and Krider 1989). This is consistent with a prediction of the enhancer-imbalance model--when transcription factors are not limiting, the underdominant genes may be transcribed. However, demonstration of chromosomal position effects on the pattern of dominance (Durica and Krider 1978) suggests that enhancer-imbalance by itself is insufficient to explain the situation in *Drosophila* (Reeder 1985, Pikaard 2000a).

In animals and plants, the region of the IGS proximal to the promotor typically contains one or more sets of reiterated sequence elements. The functions of these repeats are unknown in most organisms, but structural similarities in the repeating units suggest broad conservation of function (Reeder 1990). In cases where these subrepeats have been assigned a function, they have been found to contain elements capable of enhancing transcription (Grimaldi and Di Nocera 1988, Pape et al. 1989, Kuhn et al. 1990). In a recent survey of rRNA gene variation in *T. californicus*, we found no repeated sequences in the IGS (Burton et al. 2005). Given the allure of the enhancer-imbalance model and the possible role of repeated enhancer elements in nucleolar dominance, we examined rRNA gene expression in inter-population hybrids to determine if nucleolar dominance can occur in the absence of IGS subrepeats. We found that F_1 hybrids from two inter-population crosses, exhibit uniparental silencing of rRNA transcription despite this unusual feature of the IGS. Consistent with other studies of nucleolar dominance, the pattern of dominance was highly repeatable among hybrid individuals from both crosses and was independent of the direction of the cross (Fig. 3A, 3B). This finding is significant because it suggests that nucleolar

dominance in other species may be established and maintained in the absence of IGS subrepeats and challenges the relevance of the enhancer-imbalance mechanism as a general model for the establishment of nucleolar dominance. This highlights additional questions about whether the mechanism that produces nucleolar dominance-like expression of *Xenopus* minigene constructs is responsible for nucleolar dominance in the chromosomal context of a hybrid nucleus (Caudy and Pikaard 2002).

The pattern of codominance observed in the F_2 hybrids from the SD X SC cross is also not consistent with predictions of the enhancer-imbalance model. Despite the absence of internal subrepeats, rRNA gene clusters from different populations could still differ in their ability to sequester limiting transcription factors (e.g., as a result of non-repetitive enhancers which differ in their affinity for transcription factors). A simple transcription factor titration model predicts that transcriptional dominance may be relaxed when there is a low dosage of the dominant array (e.g., Chen et al. 1998). In this situation, transcription factors may not be limiting and transcription of the underdominant cluster may occur. Because there are at least two unlinked clusters of rRNA genes in *T. californicus* (Burton et al. 2005), F_2 intercross hybrids in the SD X SC cross can differ in dosage of the dominant and under-dominant genes due to segregation in the F_1. Therefore, hybrids with low dosage of the dominant SC cluster may be expected to codominantly express their rRNA genes, but hybrids with high dosage are not. This was not observed. Rather codominant expression in hybrids appears to be independent of SC rRNA gene dosage and genotypes with high and low relative copy numbers of rRNA genes can

codominantly express their ribosomal RNA genes (Fig. 4B). This suggests that some feature other than dosage of the dominant array of genes and the corresponding availability of transcription factors is responsible for nucleolar dominance. This observation is comparable to results from hybrids of the allotetraploid *Arabidopsis suecica*. Chen et al. (1998) observed dominance of *A. arenosa* genes over *A. arabidopsis* in *A. suecica* allotetraploid hybrids. In backcross hybrids with a 3:1 ratio of under-dominant:dominant arrays, the dominance relationship was switched. Because the dominant genes were expected to efficiently sequester transcription factors even when outnumbered, codominance was expected to occur in these hybrids. Dominance reversal was unexpected, however. This also argues against a transcription factor titration model for nucleolar dominance (Chen et al. 1998, Pikaard 2000ab).

What alternative models could explain codominance in the F_2 of *T. californicus*? One possibility is that unlinked arrays of ribosomal RNA from the SC population could differ in their dominance relationship with SD genes. For example, consider a situation with two clusters of rRNA genes per haploid genome in which one SC array is dominant over SD and one SC array is codominant with SD. If an F_2 hybrid possesses a dominant SC array, then dominance may occur similar to the F_1. However, if the dominant SC array is not present due to segregation, then codominant expression may occur. Such differences in dominance among arrays within a population could be due to differences in the sequence of the regulatory region or

chromosomal position effects on the expression of rRNA genes (e.g., Durica and Krider 1978).

Although there has been considerable progress in characterizing the mechanisms that are responsible for the enforcement of nucleolar dominance (for review see Pikaard 2000ab, Grummt and Pikaard 2003), how nucleolar dominance is originally established remains unknown. In addition to the enhancer-imbalance model, a frequently cited alternative is the species-specific transcription factor hypothesis (Reeder 1985, Pikaard 2000a). This hypothesis emerged from two observations: (1) the RNA Pol I transcription apparatus is not interchangeable among distantly related species such as *D. melanogaster* and *D. virilis* (Kohorn and Rae 1983), which most recently shared a common ancestor approximately 39 MYBP (Russo et al. 1995) and, (2) nucleolar dominance in mouse-human somatic cell hybrids may be attributable to the loss of a species-specific factor (e.g., Miesfeld et al. 1984). However, a simple species-specific transcription factor model does not predict dominance in F_1 hybrids, because transcription factors from both progenitor species are presumably present. Furthermore, the absence of a maternal effect on nucleolar dominance implies that charging of the egg cytoplasm with factors necessary for rRNA transcription does not contribute to discrimination between dominant and under-dominant genes. Finally, a direct test of this hypothesis in *Brassica* found that the transcription apparatus between closely related species was inter-changeable (Frieman et al. 1999). Our results, together with observations from other species, may indicate that neither the species-specific transcription factor hypothesis nor the

enhancer-imbalance model are satisfactory explanations for the establishment of dominance.

Chen and Pikaard (1997) proposed a model that incorporates a transcription factor competition mechanism for the establishment of dominance and epigenetic chromatin modifications for the maintenance of dominance. In their model, competitive differences in the ability to sequester transcription factors initially discriminates between dominant and under-dominant rRNA genes. Inactive under-dominant genes are then epigenetically tagged (e.g., via cytosine methylation and histone deacetylation) which reinforces transcriptional silencing even when transcription factors are not limiting. While our data do not contradict this model, our findings indicate that discrimination between dominant and under-dominant genes must occur without repetitive elements in the IGS. It remains unknown whether differences in promotor strength or some other factor could contribute to competitive dominance of AB and SC rRNA genes over SD genes.

A recent report of allele specific, parental-origin-independent CpG methylation in humans may suggest an additional mechanism that could contribute to discrimination of dominant and under-dominant genes. Yamada et al. (2004) reported mono-allelic methylation at a single SNP in a CpG island ("CGI #130"). In all heterozygotes examined, one allele was methylated, but the other was not. This was attributed to dominance of one allele in its susceptibility to methylation. This mode of allele-specific methylation could function independent of the transcriptional status of the under-dominant array as proposed by Chen and Pikaard (1997). Unlike imprinted

alleles, whose methylation status depends on parental origin, this form of mono-allelic methylation is of interest to the study of rRNA gene silencing in hybrids because of the absence of maternal effects on nucleolar dominance. If a similar type of allele-specific methylation also occurs at rRNA gene loci, then dominance of rRNA gene expression could be specified by an inherently greater susceptibility of the silenced array to DNA methylation. Additional insight into mechanisms contributing to nucleolar dominance may be gained from future studies of *T. californicus* which should include evaluation of the methylation status of silenced rRNA genes.

ACKNOWLEDGEMENTS

We thank S. Harrison and C. Willett for helpful discussions and L. Chao for sharing ideas concerning codominance in F_2 hybrids. We also thank H. Huynh for assistance with laboratory crosses. We gratefully acknowledge the National Science Foundation for supporting this research.

REFERENCES

Ar-rushdi, A.H. 1963. The cytology of achiasmatic meiosis in the female *Tigriopus* (Copepoda). Chromosoma. 13:526

Blackler, A.W. and C.A. Gecking. 1972. Transmission of sex cells of one species through the body of a second species in the genus *Xenopus*. II. Interspecific matings. Developmental Biology. 27: 385-394.

Bernard, P.S., M.J. Lay, and C.T. Wittwer. 1998. Integrated amplification and detection of the C677T point mutation in the methylenetetrahydrofolate reductase gene by fluorescence resonance energy transfer and probe melting curves. Analytical Biochemisty. 255:101-107.

Burton, R. S. and M.W. Feldman. 1981. Population genetics of *Tigriopus californicus*. 2. Differentiation among neighboring populations. Evolution. 35:1192-1205.

Burton, R S. 1990. Hybrid breakdown in developmental time in the copepod *Tigriopus californicus*. Evolution. 44: 1814-1822.

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

Burton, R.S., E.C. Metz, J.M. Flowers and C.S. Willett. 2005. Unusual structure of ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal subrepeats. Gene. 344:105-113.

Caudy, A.A. and C.S. Pikaard. 2002. *Xenopus* ribosomal RNA gene intergenic spacer elements conferring transcriptional enhancement and nucleolar dominance-like competition in oocytes. J. Biol. Chemistry. 277: 31577-31584.

Chen, J. Z. and C. S. Pikaard. 1997. Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. Genes and Development. 11:2124-2136.

Chen, Z.J., L. Comai, and C.S. Pikaard. 1998. Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. Proc. Nat. Acad. Sci. USA. 95:14891-14896.

Durica, D.S. and H.M. Krider. 1977. Studies on the ribosomal RNA cistrons in interspecific *Drosophila* hybrids. I. Nucleolar dominance. Dev. Biol. 59:62-74.

Durica, D.S. and H.M. Krider. 1978. Studies on ribosomal RNA cistrons in inter-specific *Drosophila* hybrids. 2. Heterochromatic regions mediating nucleolar dominance. Genetics. 89:37-64.

Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution. 53: 1757-1768.

Edmands, S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. Molecular Ecology. 10: 1743-1750.

Edmands, S., H.V. Feaman, J.S. Harrison, and C.C. Timmerman. 2005. Genetic consequences of many gernations of hybridization between divergent copepod populations. J. Heredity. 96:114-123.

Frieman M., Chen Z.J., Saez-Vasquez J, Shen, L.A., C.S. Pikaard. 1999. RNA polymerase I transcription in a *Brassica* interspecific hybrid and its progenitors: Tests of transcription factor involvement in nucleolar dominance. Genetics. 152:451-460.

Ganz, H.H., and R.S. Burton. 1995. Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod *Tigriopus californicus*. Mar. Biol. 123:821-827.

Goodrich-Young, C. and H. M. Krider. 1989. Nucleolar dominance and replicative dominance in *Drosophila* interspecific hybrids. Genetics. 349-358.

Grimaldi, G., and P.P. Di Nocera. 1988. Multiple repeat units in *Drosophila melanogaster* ribosomal DNA spacer stimulate rRNA precursor transcription. Proc. Nat. Acad. Sci. USA. 85:5502-5506.

Grummt, I., Roth, E. and M. R. Paul. 1982. rRNA transcription *in vitro* is species-specific. Nature. 296: 173-174.

Grummt, I. 2003. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes and Development. 17:1691-1702.

Grummt, I. and C.S. Pikaard. 2003. Epigenetic silencing of RNA polymerase I transcription. Nat. Rev. Mol. Cell Biol. 4:641-649.

Honjo, T. and R.H. Reeder. 1973. Preferential transcription of *Xenopus laevis* ribosomal RNA in interspecies hybrids between *Xenopus laevis* and *Xenopus mulleri*. J. Mol. Biol. 80:217-228.

Kohorn, B.D. and P.M.M. Rae. 1982. Accurate transcription of truncated ribosomal DNA templates in a *Drosophila* cell-free system. Proc. Nat. Acad. Sci. USA. 79:1501-1505.

Kuhn, A., U. Deppert, and I. Grummt. 1990. A 140-base pair repetitive sequence element in the mouse ribosomal RNA gene spacer enhances transcription by RNA polymerase I in a cell-free system. Proc. Nat. Acad. Sci. USA. 87:7527-7531.

Lawrence, R.J., K. Earley, O. Pontes, M. Silva, Z.J. Chen, N. Neves, W. Viegas and C.S. Pikaard. 2004. A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. Molecular Cell. 13:599-609.

Long, E. O. and I. B. Dawid. 1980. Repeated genes in eukaryotes. Ann. Rev. Biochem. 49:727-764.

Lyon, E., A. Millson, M.C. Lowery, R. Woods, and C.T. Wittwer. 2001. Quantification of HER2/neu gene amplification by competitive PCR using fluorescent melting curve analysis. Clinical Chemistry. 47:844-851.

Miesfeld, R., B. Sollner-Webb, C. Croce, and N. Arnheim. 1984. The absence of a human-specific ribosomal DNA transcription factor leads to nucleolar dominance in mouse > human hybrid cells. Mol. Cell. Biol. 4:1306-1312.

Pape, L.K., J.J. Windle, E.B. Mougey and B. Sollnerwebb. 1989. The *Xenopus* ribosomal DNA 60 base pair and 81 base pair repeats are position-dependent enhancers that function at the establishment of the preinitiation complex – analysis *in vivo* and in an enhancer-responsive *in vitro* system. Mol. Cell. Biol. 9:5093-5104.

Pikaard, C. S. 2000a. The epigenetics of nucleolar dominance. Trends in Genetics. 16:495-500.

Pikaard, C. S. 2000b. Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids. Plant. Molec. Biol. 43:163-177.

Reeder, R.H. and J. G. Roan. 1984. The mechanism of nucleolar dominance in *Xenopus* hybrids. Cell. 38: 39-44.

Reeder, R.H. 1985. Mechanisms of nucleolar dominance in animals and plants. J. Cell Biol. 101:2013-2016.

Reeder, R.H. 1990. rRNA synthesis in the nucleolus. Trends in Genetics. 6: 390-394.

Russo, C.A.M., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol. Biol. Evol. 12:391-404.

Voordouw, M.J. and B.R. Anholt. 2002. Heritability of sex tendency in a harpacticoid copepod, *Tigriopus californicus*. Evolution. 56:1754-1763.

Yamada, Y., H. Watanabe, F. Miura, H. Soejima, M. Uchiyama, T. Iwasaka, T. Mukai, Y. Sakaki, and T. Ito. 2004. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Research. 14:247-266.

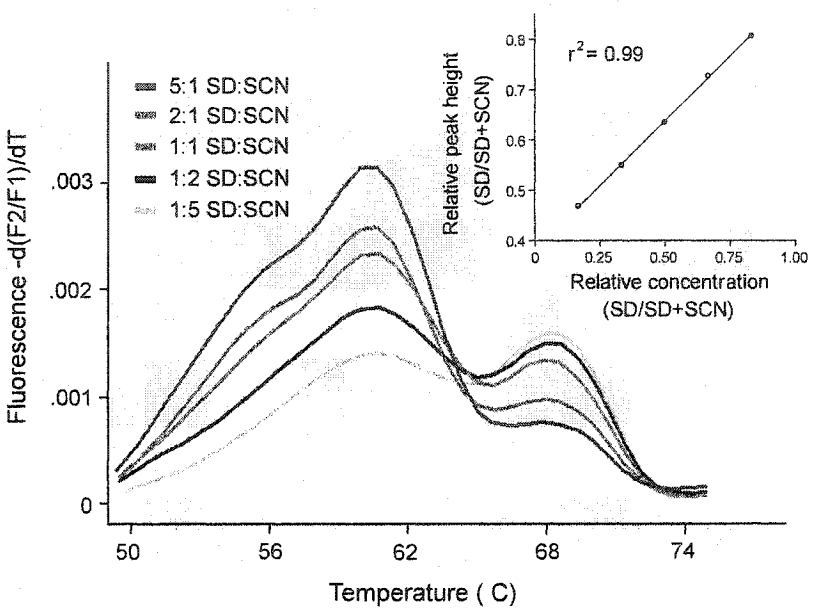


Figure 1. Melting peak profile with five different ratios of SD and SC DNA. Assays were conducted with the indicated ratios of PCR product amplified from each population. The inset is a standard curve derived from these peaks.

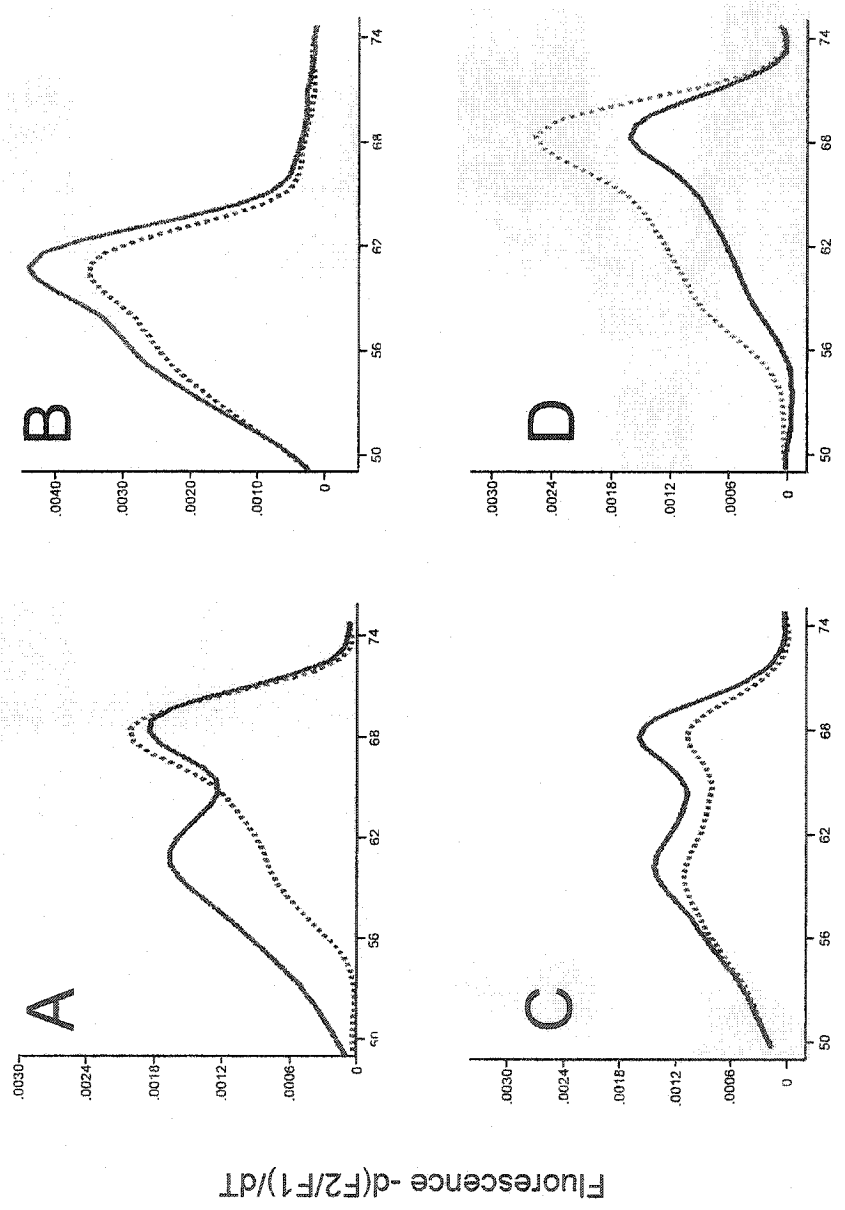


Figure 2. Melting peak profiles derived from PCR-amplified cDNA and genomic DNA from individual hybrid copepods. In each plot, the solid line represents genomic DNA and the hatched line represents cDNA. (A) ABf X SDm F_1 adult male. (B) ABf X SDm F_2 adult male. (C) SDm X SCf F_2 adult male. (D) ABf X SDm F_2 adult male.

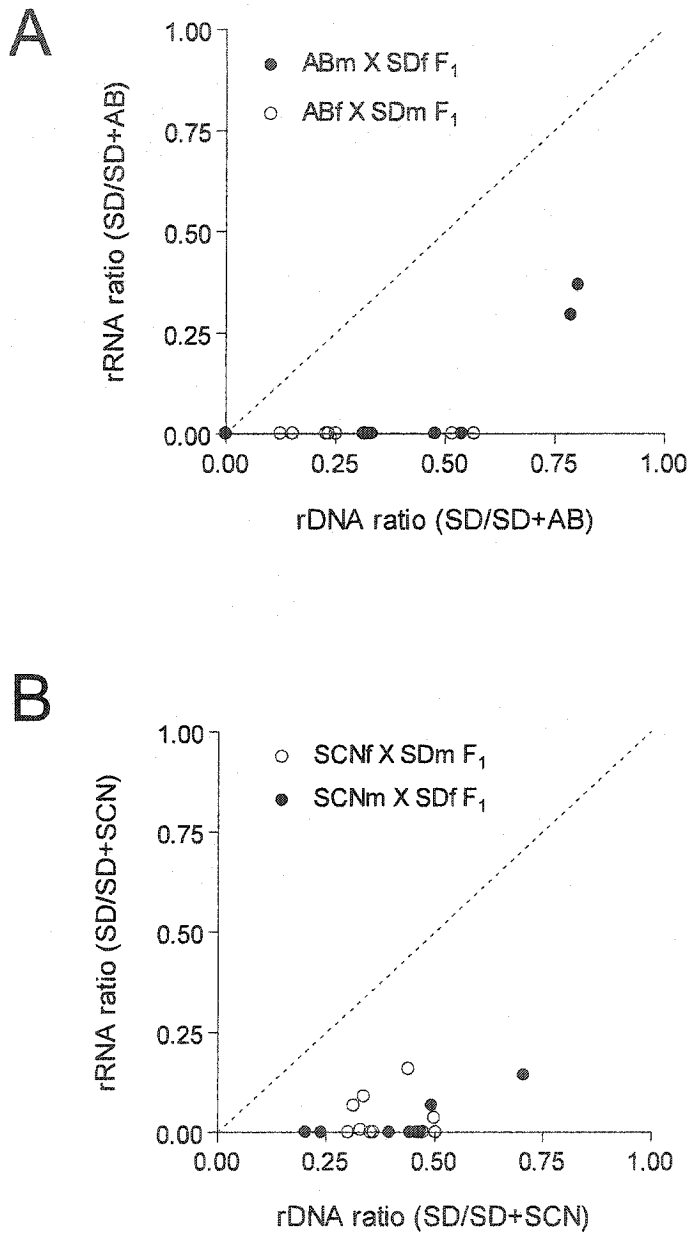


Figure 3. Plot of rRNA and rDNA ratios from F_1 inter-population hybrids. (A) AB X SD (B) SD X SC. The dotted line represents the expectation of codominant rRNA gene expression

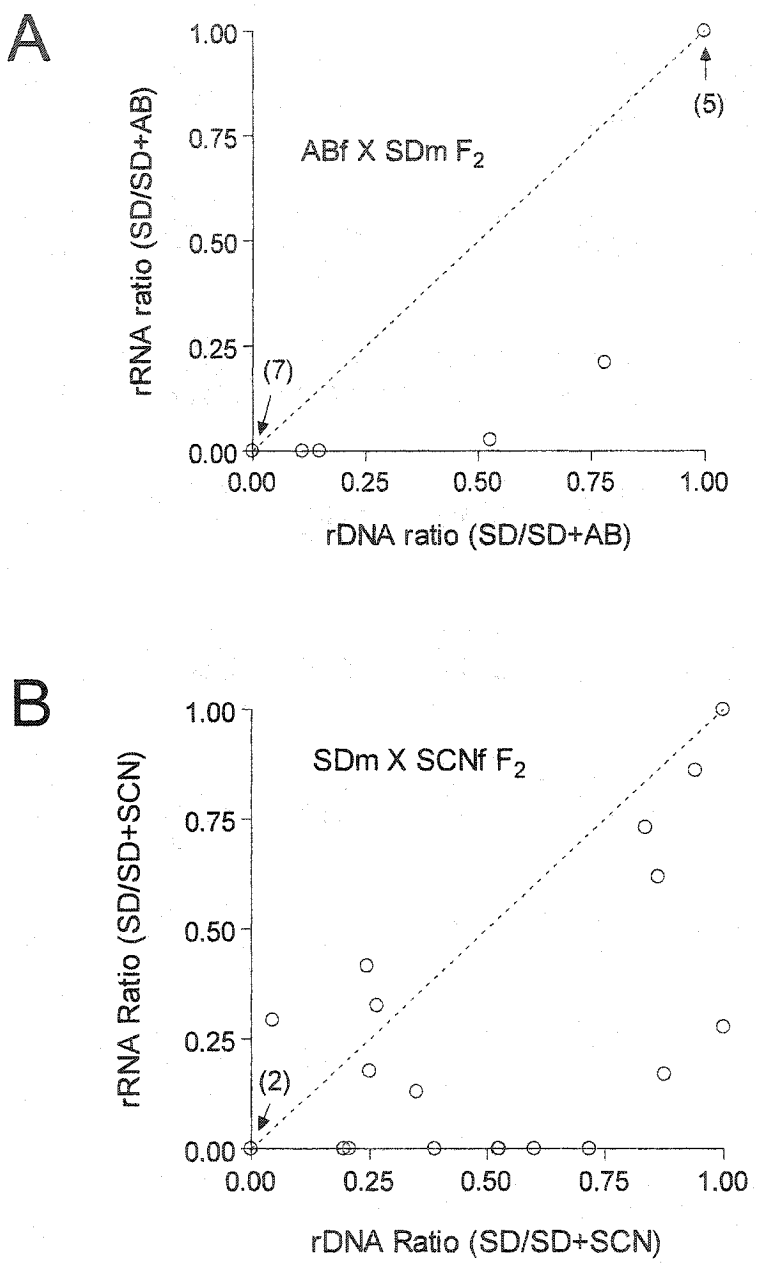


Figure 4. Plot of rRNA and rDNA ratios from F_2 inter-population hybrids. (A) ABf X SDm F_2 adult males (B) SCf X SDm F_2 adult males. The dotted line represents the expectation of codominant rRNA gene expression.

CHAPTER IV

Cyto-nuclear interactions and the evolution of mitochondrial transcription genes in *Tigriopus californicus*

ABSTRACT

Genes of the mitochondrial transcription apparatus are encoded in the nucleus and imported into the mitochondria where they regulate the transcription of mitochondrial DNA (mtDNA). Recent work on mitochondrial transcription in mouse and human indicates that functional incompatibilities have evolved between the mitochondrial transcription apparatus and the mitochondrial promoter over the course of mammalian evolution. Transcription assays with a recombinant transcription system have found that the mouse transcription apparatus is not capable of recognizing the human mitochondrial promoter. The gene responsible for this species-specificity of transcription is mitochondrial RNA polymerase. Similarities with the related T3/T7 RNA polymerase suggest that the species-specificity of mitochondrial transcription may have a simple genetic basis and therefore may evolve early in the process of species divergence. Here we characterize the mitochondrial RNA polymerase (*mtRPOL*) and a mitochondrial transcription factor *mtTFB1* from the marine copepod *Tigriopus californicus*. Populations of this species have diverged substantially at mtDNA loci and incompatibilities between the cytoplasm and nucleus are a well-documented feature of inter-population hybrids. We test for positive selection at *mtRPOL* and *mtTFB1* with sequence comparisons from divergent populations. Although these genes have diverged substantially we find no evidence that positive selection has contributed to their divergence. In addition, we show that *mtRPOL* and *mtTFB1* are tightly linked and that the linkage group marked by these

genes has strong viability effects in inter-population hybrids that are dependent on cytoplasmic background.

INTRODUCTION

Epistatic interactions between cytoplasmic and nuclear genes are an important class of interactions that impact hybrid fitness (Clark and Lyckegaard 1988, Hutter and Rand 1995, Breewuer and Werren 1995, Burke et al. 1998, Galloway and Fenster 2001, Willett and Burton 2001, James and Ballard 2003, Zeyl et al. 2005). Insight into the molecular basis of this class of interactions has come from studies of nuclear and organellar genes whose products must functionally interact. The most extensively studied examples include genes encoded in the nuclear and mitochondrial genomes whose products constitute the electron transport system (ETS, Blier et al. 2001). Proper ETS function therefore requires the coordinated expression of genes in both genomes and structural compatibility of their respective gene products. There is now substantial evidence from a number of experimental systems that indicate that functional incompatibilities between the cytoplasm and the nucleus are widespread among species that are no longer capable of exchanging genes (e.g., Osheroff 1983, Kenyon and Moraes 1997, Mckenzie et al. 2003). In some cases, however, functional incompatibilities between mitochondrial and nuclear genes may arise during the incipient stages of divergence (Edmands and Burton 1999, Burton and Rawson 2002, Sackton et al. 2003). In this situation, hybridization can produce genotypes in which incompatibilities between mitochondrial and nuclear genes may reduce hybrid fitness. A major goal, then, is to identify genes that contribute to reduced hybrid fitness and to determine the circumstances under which incompatibilities arise.

Studies of the genetic basis of cyto-nuclear interactions have been limited primarily to genes involved in the electron transport system. However, functional integration of nuclear and mitochondrial genomes extends well beyond this system of genes. Studies of mitochondrial transcription suggest that cyto-nuclear incompatibilities also evolve between the mitochondrial transcription apparatus and the mitochondrial promoter(s) (Fisher et al. 1989). For example, Gaspari et al. (2004) employed an *in vitro* recombinant transcription system to determine if the human mitochondrial transcription apparatus could functionally replace the mouse machinery and initiate transcription of the mouse light strand promoter (LSP). They found that the human apparatus was fully capable of transcribing the human LSP, but could not initiate transcription of the mouse promoter. The reverse experiment similarly indicated that the mouse apparatus was incapable of initiating transcription of the human promoter. This species-specificity of transcription of mtDNA indicates that functional incompatibilities between nuclear genes responsible for mitochondrial transcription and the mitochondrial promoter(s) have evolved during the course of mammalian evolution.

Several aspects of the mitochondrial transcription apparatus make this an intriguing and experimentally tractable system to evaluate how cyto-nuclear incompatibilities may evolve. The mitochondrial genome is transcribed by a dedicated, single-subunit RNA polymerase. This polymerase is encoded in the nuclear genome in most plants and animals and is unique among eukaryote RNA polymerases in that its origin can be traced to the RNA polymerase of bacteriophages T3 and T7

(Masters et al. 1987, Cermakian et al. 1997). The bacteriophage RNA polymerase is unusual among RNA polymerases in that it does not require additional factors for promoter-specific transcription (Cheetham et al. 1999). In eukaryotes, the mitochondrial transcription apparatus is somewhat more complex. In yeast, transcription requires a specificity factor, *MTF1,* and mitochondrial RNA polymerase, *Rpo41*. In mammals, the transcription factor *TFAM* and one of a pair of interchangeable factors, *TFB1M* or *TFB2M* are required by mammalian mitochondrial RNA polymerase (*POLRMT*) for transcription (Falkenberg et al. 2002). The presence of *TFAM*, *mtTFB1/TFB1M*, *mtTFB2/TFB2M* and mitochondrial RNA polymerase (*mtRPOL*) in the *Drosophila* genome (Rantanen et al. 2003), together with studies that support a role for *mtTFB2* in mitochondrial transcription in *Drosophila* (Matsushima et al. 2004, see also Matsushima et al. 2005), suggests that these components may generally comprise the mitochondrial transcription apparatus in animals.

An unusual feature of the mitochondrial transcription apparatus is that promoter recognition appears to be determined by the polymerase itself (Matsunaga and Jaehning 2004, Gaspari et al. 2004). In a factor-swapping experiment, Gaspari et al. (2004) showed that a recombinant system containing *mtRPOL/POLRMT*, *TFAM,* and *mtTFB2/TFB2M* from mouse could not transcribe the human LSP, but replacement of the mouse *mtRPOL/POLRMT* with the human polymerase resulted in successful transcription. This indicated that *mtRPOL/POLRMT* and not either of the auxiliary factors is responsible for the species-specificity of mitochondrial transcription. Since species-specificity of transcription is likely determined by direct

contacts made between *mtRPOL/POLRMT* and the promoter (Gaspari et al. 2004), this raises the possibility that the specificity of mitochondrial promoter recognition may have a simple genetic basis similar to that observed for T3 and T7 (Klement et al. 1990, Raskin et al. 1992). A possible consequence of this is that species-specificity of mitochondrial transcription may also have a simple basis and may evolve early in the process of divergence as a result of reciprocal changes in the *mtRPOL/POLRMT* and the mitochondrial promoter.

The copepod *Tigriopus californicus* has been a valuable model for investigating the incipient stages of the evolution of cyto-nuclear incompatibilities. The reason for this is that mitochondrial and nuclear gene functions are integrated within populations. Disruption of associations between mitochondria and the nucleus in hybrids has large effects on the function of the electron transport system (Edmands and Burton 1999, Rawson and Burton 2002). A primary reason for the integration of mitochondrial and nuclear gene function in this species is attributable to the historical allopatry of *T. californicus* populations and the large inter-population divergences observed among populations at mitochondrial (Burton 1998, Edmands 2001, Willett and Burton 2004) and in some cases nuclear gene loci (Rawson and Burton 2000, Willett and Burton 2004). For example, uncorrected population divergences at mitochondrial loci frequently exceed 20% and amino acid divergences often exceed 10% (Burton 1998, Edmands 2001, Willett and Burton 2004, unpublished). Therefore, functional incompatibilities between the cytoplasm and the nucleus in

hybrids are, in part, attributable to extensive inter-population divergences of interacting mitochondrial and nuclear gene loci.

Here we report on the cloning of *mtRPOL* and *mtTFB1* in *T. californicus*. We show that *mtRPOL* and *mtTFB1* are tightly linked and that the linkage group marked by these genes has strong viability effects in inter-population hybrids that are dependent on cytoplasmic background. We evaluate the extent of divergence in regions of *mtRPOL* that are likely responsible for promoter recognition and test for the action of natural selection on *mtRPOL* and *mtTFB1* genes by comparing sequences from *T. californicus* populations. We find that both genes have substantially diverged among populations, but find no evidence that positive selection has contributed to divergence at these loci.

MATERIALS AND METHODS

Sample collection and inter-population crosses

Samples were collected from locations throughout Baja California, Mexico and California, USA and maintained in laboratory cultures in 100% sea water (with minor fluctuations in salinity due to evaporation) at 20° C with a 12:12 light-dark cycle and small amounts of *Spirulina* flakes (aquarium fish food). With the exception of samples from three new sites at Punta Cabras, Baja California Norte, Mexico, San Simeon, and Bodega Head, collecting locals are the same as those in previous studies (e.g., Burton and Feldman 1981, Burton and Lee 1994, Ganz and Burton 1995, Burton 1998). Inter-population hybrids were obtained from reciprocal crosses between study populations from San Diego (SD), Abalone Cove (AB, Los Angeles County), and Santa Cruz (SC). Pair crosses were established following the method of Burton and Feldman (1981) by isolation of virgin females and mating them in mass culture with males from an alternate population. F_2 inter-cross hybrids were obtained by inter-crossing of F_1 hybrids in mass culture. Second generation hybrids were subsequently obtained by transferring fertilized F_1 females to new cultures for production of F_2 progeny. Because there is no recombination in *Tigriopus* females (Ar-rushdi 1964, Burton et al. 1981), linkage relationships were determined by genotyping the progeny of a backcross between F_1 SDf X ABm females to SD males.

Cloning of mitochondrial RNA polymerase (*mtRPOL*)

mtRPOL was cloned with a degenerate PCR-based approach. 1st strand cDNA synthesis was primed with an oligo(dT) primer (Invitrogen) with an extension at the 5' end to facilitate 3' RACE. Degenerate PCR was performed with a modified touchdown PCR protocol with forward (CCNCAYAAYATGGAYTTYMGNGG) and reverse (GCRTARTGYTGNARNCCRTTRCA) primers. Cycling conditions consisted of a 30 s 95° denaturation step, an initial annealing step of 30 s 55° C, and an extension step of 2 m 72° C. Every five cycles for the first 15 cycles the annealing temperature was stepped down by 5° C. The final 25 cycles were conducted with an annealing temperature of 48° C. The protocol yielded a single PCR product of approximately 405 bp which was gel purified and TA-cloned into the pCR 4-TOPO vector (Invitrogen). The 3' end of the gene was obtained by 3' RLM-RACE. The 5' end of the gene was obtained by multiple rounds of PCR walking by amplification from a Zap Express (Stratagene) cDNA library with gene-specific and vector primers.

Cloning of mitochondrial transcription factor *mtTFB1*

mtTFB1 was cloned with a similar degenerate PCR-based strategy. Primers were designed from an alignment of putative homologs in human, mouse, and *Drosophila* identified by Rantanen et al. (2003). Degenerate PCR was conducted with forward (GARGTIGGICCNGGNCCNGGNGG) and reverse (ACICNCCNACRTCNAC) primers over two rounds of amplification with oligo(dT)-primed cDNA as template. The first round of PCR consisted of a standard touchdown

protocol with cycling conditions of 95° 30 s, 52° 30 s, and 72° for 2 m. The annealing temperature was decreased by 1° C every two cycles for the first 14 cycles. The final 21 cycles were carried out with a 45° annealing step. A second round of PCR was performed with a 50° C annealing temperature for 30 cycles using 1.0 ul of the original PCR product as template. This protocol yielded a band of the expected size that was cloned and sequenced as described above. 5' and 3' ends of the gene were obtained by RLM-RACE (Invitrogen). Although the 5' end was identified in cDNA based upon a putative methionine start codon, we were unable to successfully amplify the 5' end of *mtTFB1* from genomic DNA with primers located in the putative 5' UTR. Subsequent PCR of genomic DNA utilized a primer anchored at the 5' end of exon 1.

DNA sequencing of population samples

mtRPOL sequences were obtained from 13 *T. californicus* populations including a partially reproductively isolated population from Playa Altamira, Baja California, Mexico (Ganz and Burton 1995). All sequences were obtained from genomic DNA isolated from single copepods. Sequences from Abalone Cove (AB1C), Royal Palms (RP1C), and Punta Baja (PBJ2A) were obtained by PCR-amplification of the entire structural gene with a proof-reading *Taq* polymerase (Triple Master, Eppendorf) followed by cloning of the PCR product with the TOPO-XL TA-cloning kit (Invitrogen). All other sequences were obtained by amplification of seven overlapping fragments from genomic DNA extracted from single copepods followed

by sequencing of both strands of each PCR product. *mtTFB1* sequences were collected from eight populations by amplifying three-overlapping fragments from individual animals followed by direct sequencing.

Genotyping of *mtTFB1* and *mtRPOL* in inter-population F_2 hybrids

Individual F_2 inter-cross or backcross hybrids were screened simultaneously for *mtTFB1* and *mtRPOL* genotypes with a multi-plexed PCR-RFLP approach. We identified restriction site polymorphisms in each gene that were fixed in each of the populations SD, AB, and SC and therefore could be used to unambiguously genotype alleles from each population. Both genes were amplified simultaneously with primers flanking the informative sites and then digested with a single enzyme that recognized the restriction site polymorphisms in each gene (Figure 1). For the SD X SC and SD X AB crosses, we simultaneously amplified *mtRPOL* and *mtTFB1* with primer pairs A/B and C/D, respectively (A: CGTGTTTATCCCATCCCT, B: CGAAGCTCTATACAAGCG, C: CTTGAAATACACCTTGAAGG, D: GAGCACGTCTTGGTCCTC). PCR products were subsequently digested with *HinF* I (Invitrogen) following the manufacturer's instructions. For the AB X SC cross, we amplified both genes with primer pairs E/F and G/H (E: GCGTATGCAATATGCCGAGG, F: TCTTTGGATAGCTCCATTCAAG, G: TTGGGGCGAGTTTAGGTCCG, H: GGCTGCCGTAGCTCCGCGTC) followed by digestion with *Msp* I (New England Biolabs). Cycling conditions for the multiplexed reactions with primers A-D consisted of 95° for 30° s, 50° for 1 m, and 72° for 2 m; E-

H consisted of 95° for 30 s, 50° for 30 s, and 72° for 1 min. Multiplexed reactions were all conducted by adding reactions directly to a hot block. Electrophoresis of digested PCR products was performed on 3% w/v agarose and visualized with ethidium bromide under UV light.

For the SDf X SCm cross, genotyping was also conducted on first stage naupliar larvae. F_2 nauplii were obtained by inter-crossing F_1 hybrids in mass culture and transferring fertilized females to a fresh petri dish. After three or four days, mature pink/red egg sacs were dissected from females and placed in a small volume of sea water. Once the nauplii had hatched, they were moved to filter paper by pipette and individually transferred with pins to 10 µl of lysis buffer with the aid of a dissecting microscope. Pins were used only once to prevent cross-contamination. Lysis preparations were incubated at 65° C for 1 hr followed by 85° C for 15 m and 2 µl used as template for PCR.

Tree reconstruction and tests of selection

DNA sequences were edited in Sequencher v. 4.2.2 (Gene Codes Corporation) and aligned using ClustalX (Thompson et al. 1997). Minor adjustments to the *mtTFB1* alignment were made by eye because of insertion/deletion polymorphism in introns. Inter-population distances were calculated in MEGA v. 2.1 (Kumar et al. 2001). Phylogenetic trees were constructed with maximum parsimony in PAUP* (Swofford 1998). Tests for selection were conducted using the random sites models of Nielsen and Yang (1998) as implemented by the CODEML algorithm in PAML v. 3.14 (Yang

1997). CODEML utilizes a codon-based model of substitution (Goldman and Yang 1994) to derive maximum likelihood estimates of model parameters (e.g., ω, the nonsynonymous/ synonymous rate ratio). The fit of hierarchically nested models with different constraints on ω can then be evaluated with log-likelihood ratio tests (LRTs). We tested for positive selection with LRTs of model M1a (nearly neutral) versus M2a (positive selection) and model M7 (nearly neutral) versus M8 (positive selection). The PAML v. 3.14 documentation provides details concerning how these models differ from their predecessors in previous versions of this package (http://abacus.gene.ucl.ac.uk/software/paml.html). Because *mtRPOL* contains multiple subdomains responsible for RNA polymerization and related functions we tested for selection on each of these domains independently. We defined each domain by identifying residues at the boundaries of each domain that are conserved from viruses to humans (Jeruzalmi and Steitz 1998). The domains identified include the *N*-terminal extension, fingers, palm, thumb, specificity loop, and the extended foot. We analyzed the following domains separately in our test of selection: *N*-terminal extension (mitochondrial targeting sequence and the *N*-terminal domain), fingers (fingers and the specificity loop), palm (palm and extended foot), and thumb domains. Finally, we tested for differences in rates of evolution among the *N*-terminal extension and the remainder of the *mtRPOL* gene (i.e., all domains required for RNA polymerase activity) using the fixed-site models of Yang and Swanson (2002, see also Pogson 2004).

RESULTS

Structure of *mtRPOL* and *mtTFB1*

The *T. californicus mtRPOL* consists of an open reading frame of 3,528 bp and 1,176 amino acids in the translated protein. Immediately downstream of the start codon is a putative 53 amino acid targeting peptide. This was predicted by the MITOPROT algorithm (Claros and Vincens 1996; http://us.expasy.org/tools/) to have an 83% probability of being a mitochondrial targeting sequence. Amino acid identities were 35.5%, 37.8%, and 33.0% between *T. californicus* and *Drosophila melanogaster*, *Anopheles gambiae*, and human mitochondrial RNA polymerases respectively. The *C*-terminal end of the *T. californicus* gene (see below), which contains domains responsible for RNA polymerase activity, showed considerably higher levels of conservation. Many residues that are conserved in all known eukaryotic mitochondrial RNA polymerases (Jeruzalmi and Steitz 1998) and have well-described functions in T3/T7 RNA polymerases are also conserved in *T. californicus*. This is reflected in higher identities at the *C*-terminal end (51.2%, 50.3%, and 50.4% between *T. californicus* and *D. melanogaster*, *A. gambiae* and human, respectively). Comparison of sequences from cDNA and genomic DNA revealed no introns in the *T. californicus mtRPOL*. The domain structure in *T. californicus* is illustrated in Figure 2.

The transcription factor *mtTFB1* in *T. californicus* consists of 365 amino acids and shares identitites of 36.3% and 38.9% with *D. melanogaster* (Rantanen et al. 2003, Matsushima et al. 2005) and human *mtTFB1/TFB1M* proteins, respectively. The

MITOPROT algorithm identified a 37 amino acid targeting signal at the *N*-terminus that had an estimated 94% probability of being a mitochondrial signal sequence. Comparison of cDNA and genomic DNA revealed two small introns (77 and 93 bp in the SD population) that contained small (i.e., 1-4 bp) insertion/deletions in inter-population comparisons. Our final alignment contained 1,268 bp and consisted of 88 bp of exon 1, exons 2 and 3, both introns and 22 bp of the 3'UTR. The structure of *mtTFB1* is illustrated in Figure 2.

mtRPOL and *mtTFB1* are linked

We capitalized on the absence of recombination in *T. californicus* females, to determine if *mtRPOL* and *mtTFB1* were linked. SDf X ABm F_1 hybrids were backcrossed to SD males and nine *SD/SD* double homozygotes and five double heterozygotes were recorded in the progeny. No recombinant genotypes were observed which indicates that *mtTFB1* and *mtRPOL* are linked ($P < 0.0001$). This result was confirmed in F_2 intercross hybrids (see below) in which we observed 12 recombinant genotypes in a total of 2,681 hybrids screened. This translates into a map distance of 0.45 centiMorgans (cM) between the two genes. We also tested for linkage between *mtRPOL/mtTFB1* and cytochrome *c* (*CYC*), a locus that has previously been found to have cytoplasmic effects on hybrid fitness (Willett and Burton 2001, 2003). Scoring of *mtRPOL/mtTFB1* and cytochrome *c* markers in ten backcross hybrids produced seven recombinant and three non-recombinant genotypes.

The presence of recombinant genotypes indicates that *mtRPOL/mtTFB1* and *CYC* are located on different chromosomes.

Reciprocal cross effects on segregation ratios

Previous studies of marker segregation have found large viability effects on marker genotypes in inter-population hybrids of *T. californicus* (Burton 1987, Willett and Burton 2001, 2003, Harrison and Edmands unpublished). The tight linkage of *mtTFB1* and *mtRPOL* and the hypothesized functional interactions of these genes with mtDNA motivated our screening of F_2 hybrids for cytoplasmic effects on segregation of this linkage group. Table 1 presents genotypic counts for adult F_2 inter-population hybrids for each of three crosses and their reciprocals. Table 2 contains the viabilities of each homozygote class for each cross assuming a viability of one for the heterozygote class. Deviations from the 1:2:1 Mendelian expectation for inheritance of a codominant marker were observed in the SDf X SCm ($P < 0.00001$). This was attributable to a significant deficit of maternally-derived *SD/SD* homozygotes in both males and females. Deviations from Mendelian expectations were not observed in either sex of the reciprocal cross. Results from the SD X SC cross were derived from four independent crossing experiments (i.e., two from each reciprocal) that yielded virtually identical results (Table 3). Because no difference was observed between males and females in either reciprocal, we pooled the sexes and conducted a test of the homogeneity of genotypic ratios between the SDf X SCm and SCf X SDm crosses.

This test indicated that genotypic counts were not homogeneous indicating a reciprocal cross effect on marker segregation ratios (χ^2=12.99, df = 2, P = 0.0015).

A similar pattern was observed for the AB X SC cross. In the ABf X SCm cross, genotypic ratios deviated from 1:2:1 in males ($P < 0.005$) and females ($P < 0.005$). In both sexes, the deviation appeared to be due to a deficit in the maternally-derived AB/AB homozygote class. In the reciprocal SCf X ABm cross, there was a marginally significant deviation from the Mendelian expectation attributable to an excess of heterozygotes in females ($P < 0.05$). A contingency chi-square test of homogeneity between ABf X SCm and SCf X ABm suggested a reciprocal cross effect ($\chi^2 = 11.49$, $df = 2$, $P < 0.005$) similar to the SD X SC cross.

In the AB X SD cross we observed marginally significant differences between the sexes. In both reciprocals, males deviated from 1:2:1, but females did not (Table 1). A contingency chi-square test of differences between the sexes suggested a marginally significant deviation from homogeneity in each reciprocal (Table 1). Because of the sex effect in this cross, we conducted a test of differences between the reciprocal crosses on males and females separately. Neither sex showed evidence of a reciprocal cross effect (males, $P = 0.186$; females, $P = 0.548$).

Segregation ratios in first stage naupliar larvae

Previous studies of *T. californicus* inter-population hybrids have demonstrated large viability effects during developmental stages between larval hatching and the adult stage (Edmands 1999, Willett and Burton 2001, S. Edmands personal

communication). To determine if segregation distortion observed in F_2 hybrids is attributable to post-hatching viability effects of the *mtRPOL/mtTFB1* linkage group (as opposed to pre-zygotic effects), we screened 229 first stage F_2 naupliar larvae from the SDf X SCm cross. In contrast to the adults from this cross (Table 4), genotypic ratios from nine different broods of nauplii did not deviate from the 1:2:1 expectation ($\chi^2 = 0.817$, $P = 0.665$, $df = 2$, Table 4). A test of homogeneity indicates that genotypic ratios differ between adults and nauplii for this cross ($\chi^2 = 11.03$, $df = 2$, $P = 0.004$, Contingency Chi-square test).

Population differentiation

Inter-population distances based upon 3.6 kb of *mtRPOL* and 1.2 kb of *mtTFB1* are presented in Table 5. Distances between populations at *mtRPOL* are approximately 1-2.5% among populations north of Punta Baja, Baja California, Mexico, while the partially reproductively isolated population from Playa Altamira in central Baja California is greater than 10% divergent from all populations to the north. At the *mtTFB1* locus, typical inter-population distances of 2-3% are somewhat higher than *mtRPOL* because of inclusion of introns, which consistently show higher levels of divergence than protein-coding sequences in inter-population comparisons in this species (Rawson and Burton 2000, Willett and Burton 2004). Amino acid polymorphism for *mtRPOL* and *mtTFB1* are presented in Table 6 and 7, respectively. The position of amino acid substitutions with respect to functional domains (see below) can be visualized in Figure 2. Hatch marks above and below the *mtRPOL* gene

represent amino acid substitutions between SD and SC and between SD and Playa Altamira, respectively.

Phylogenetic reconstructions are consistent with other published results (Figure 3). Both genes support regionally distributed clades, as well as distinct lineages within regions (Burton and Lee 1994, Burton 1998, Edmands 2001). In some cases, nodes not well supported in previous studies are better resolved in our parsimony reconstruction. For example, the *mtRPOL* tree supports a southern California clade and central California clade with AB and SD populations belonging to the southern California clade and SC population belonging to the central California clade. In addition, a population south of the biogeographic boundary at Point Conception from Carpinteria, Santa Barbara County, clusters with the southern California clade, while a population from San Simeon, San Luis Obispo county, north of Point Conception clusters with the central California clade.

Tests for selection

We were interested in determining if accelerated evolution in the mitochondrial genome (Willett and Burton 2004) contributes to parallel acceleration of mitochondrial transcription genes. We addressed this by conducting tests of positive selection on *mtRPOL* from 12 populations and on *mtTFB1* from eight populations. The analysis of *mtRPOL* was based on 3,510 bp and included all but six codons at the *C*-terminal end of the gene. A separate analysis included the divergent Playa Altamira population, but was restricted to 3,156 bp because of the inability to

obtain the ends of the gene from this population. Because results from this analysis were qualitatively similar to the analysis based on the longer sequence they are not discussed further.

We conducted analysis of the entire gene and independent analyses of each of four functional subdomains of *mtRPOL*. Results from comparisons between models M1a and M2a are presented in Table 8. More parameter-rich models (i.e., M7 and M8) yielded comparable results and did not provide a better fit to the data (not shown). Using the codon substitution model of Goldman and Yang (1994), we obtained an estimate of 0.1715 for the nonsynonymous/synonymous rate ratio, ω, for *mtRPOL*. Partitioning of the genes into four functional domains yielded ω values of 0.3312 for the *N*-terminal extension, 0.0425 for the fingers subdomain, 0.0959 for the thumb subdomain, and 0.0313 for the palm subdomain. Analysis of *mtTFB1* yielded an ω estimate of 0.1011 An LRT of the likelihoods of models M1a and M2a was used to test for positive selection on the entire gene and on each individual domain. In none of the analysis did model M2a provide a significantly better fit to the data than model M1a. Thus, we found no evidence of a class of sites with ω values > 1.

Previous studies have suggested that the *N*-terminal extension of *mtRPOL* may evolve differently from the *C*-terminus, which contains the domains responsible for RNA polymerase activity. We evaluated if this was also the case in *T. californicus* by testing for rate heterogeneity between the *N*-terminal extension and the remainder of the gene with the fixed site models of Yang and Swanson (2002). The hierarchical models (A-F) listed in Table 9 facilitate a test of whether parameter estimates from

each partition are the same. A model that allows for differences in overall rate (model B) did not provide a significantly better fit to a model incorporating a single rate matrix for both partitions (model A). Models that allowed for differences in ω and κ, the transition:transversion rate ratio, between the partitions (Models D and E) provided substantially better fits to the data than models in which partitions were constrained to have the same ω and κ (Table 9). Since estimates of κ did not vary between the partitions, this indicates that ω differs significantly between partitions.

DISCUSSION

The mitochondrial transcripton apparatus is dedicated exclusively to the transcription of mtDNA. In animals, the genes of the mitochondrial transcription machinery are encoded in the nucleus and imported into the mitochondrion, where they regulate the transcription of typically one or two mitochondrial promoters in the control region, or D-loop, of mtDNA (Jaehning 1993). The full complement of genes that constitute the mitochondrial transcription apparatus has only recently been described (Falkenberg et al. 2002), but this has already lead to interesting discoveries concerning the coordinated evolution of the mitochondrial transcription genes and mitochondrial promoter(s). Gaspari et al. (2004) demonstrated that the mouse apparatus was incapable of initiating transcription of the human LSP and the human system similarly failed to transcribe the mouse promoter. However, if the mouse transcription factors were coupled with human mitochondrial RNA polymerase successful transcription of the human LSP was accomplished. This indicated that coordinated changes in the mitochondrial promoter and mitochondrial RNA polymerase have evolved during the course of mammalian evolution such that the mouse polymerase is no longer capable of transcribing human mtDNA.

In *T. californicus*, mtDNA evolves at an accelerated rate presumably as a result of an exceptionally high mutation rate found in this (Willett and Burton 2004) and possibly other copepod species (Rocha-Olivares et al. 2001). Rapid evolution of mtDNA is apparent throughout the genome including non-coding regions (e.g., control region, unpublished) that harbor as yet uncharacterized promoter sequence(s). Given

the rapid rate of divergence mtDNA among populations (Willett and Burton 2004) and the extent of cytonuclear coadaptation in this species, we are interested if coordinated changes have evolved in nuclear-encoded mitochondrial transcription genes and the mitochondrial promoter similar to that observed in mammals.

Epistatic interactions between the cytoplasm and the nucleus that have large effects on fitness can be detected by comparing the viabilities of genotypes on different cytoplasmic backgrounds. Given that the cytoplasm is maternally transmitted, a natural hypothesis is that genotypes with alleles derived from the maternal population will have higher fitness than paternal genotypes (Breeuwer and Werren 1995, Burke et al. 1998, Willett and Burton 2001, Edmands et al. 2005). The reduced fitness of recombinant genotypes with paternally-derived alleles is expected in cases where nuclear genes have evolved positive associations with their native cytoplasms. Such interactions are apparent at the linkage group marked by the *CYC* locus in certain inter-population crosses of *T. californicus* (Willett and Burton 2001, 2003). Disruption of such associations in hybrid organisms can lead to unfavorable interactions between nuclear genes and the cytoplasm and may be detected in the form of reduced viability of affected genotypes.

We tested this basic expectation for the linkage group containing *mtTFB1* and *mtRPOL* by scoring F_2 hybrid genotypes in inter-population crosses. In the SD X SC and AB X SC crosses we detected significant viability effects in only one direction of each cross, consistent with a cytoplasmic effect on fitness. However, in neither cross did we observe the hypothesized reduced viability of paternal genotypes on a foreign

cytoplasmic background. Instead, we observed lower relative viabilities of homozygotes with maternally-derived alleles. For example, in replicate SDf X SCm crosses, we observed 50% fewer *SD/SD* genotypes than expected, while the *SC/SC* genotype was consistently found in the expected Mendelian proportions. The reciprocal cross showed no viability effects on either homozygote class. Similarly, in the ABf X SCm cross, the maternal *AB/AB* genotypic class showed a 50% reduction in viability, while the paternal *SC/SC* genotypic class showed no evidence of viability effects. The reciprocal ABm X SCf cross differed marginally from 1:2:1 because of a moderate excess of heterozygotes, but preferential reduction of the *AB/AB* genotypic class was not observed. These observations suggest that the effects we have observed are likely attributable to interactions between a component of the cytoplasm and the *mtTFB1/mtRPOL* marker locus, but the hypothesized negative interactions between paternal alleles on a foreign cytoplasmic background were not observed. Instead, recombinant genotypes (e.g., *SC/SC* on SD cytoplasm) had higher relative viability than non-recombinant genotypes (e.g., *SD/SD* on SD cytoplasm).

We determined if the effects we observed were attributable to pre- or post-zygotic effects by genotyping larvae from the SDf X SCm cross. In *T. californicus*, embryos are brooded by females in an external egg sac for a period of three to four days after fertilization. After this period, free-swimming naupliar larvae hatch and progress through six stages of development, prior to metamorphosis. We screened 229 first stage nauplii immediately after hatching and observed no deviations from the Mendelian expectation. This indicates that viability selection acts on *mtTFB1/mtRPOL*

genotypes at some point after this developmental stage and suggests that pre-zygotic effects such as meiotic drive are not responsible for the observed effects. Since these effects must occur at least 3 to 4 days after fertilization, this also suggests that the observed interactions are not attributable to maternal effects. This is because most maternal effect phenotypes, such as the maternal-effect embryonic lethals in *Tribolium,* are expressed early in embryogenesis (Beeman et al. 1992). Therefore, it is likely that the interactions we have observed are attributable to interactions between the nucleus and the cytoplasm that affect the viability of hybrids.

Interpretation of segregation data is constrained by the fact that the viability of each genotype is defined relative to another. However, results similar to those reported here have been interpreted in the context of beneficial interactions between heterospecific gene combinations. For example, in experimental F_2 *Iris* hybrids, Burke et al. (1998) reported that hybrids homozygous for *I. fulva* alleles at a nuclear marker locus had higher relative viability on a foreign *I. brevicaulis* cytoplasmic background than on the native *fulva* background. They suggested that heterospecific combinations increased hybrid fitness and may be adaptive in the context of a natural hybrid zone (Burke et al. 1998, Burke and Arnold 2001). Their interpretation is based on the potentially adaptive effects of beneficial epistasis in naturally-occurring *Iris* and *Helianthus* hybrids (Rieseberg et al. 1996, 1999, Burke and Arnold 2001). However, the adaptive effects of beneficial epistasis reported in these hybrids may differ significantly from the effects of hybridization in *T. californicus*. F_2 inter-population hybrids in *T. californicus* generally suffer from severe hybrid breakdown and hybrid

genotypes with fitness higher than parentals are rarely observed (e.g., Burton 1987, Burton 1990, Edmands 1999, Edmands et al. 2005). Therefore, it is questionable whether the observed effects (i.e., higher viabilities of recombinant genotypes) are attributable to favorable interactions between genes from different populations as proposed for *Iris* and *Helianthus* hybrids. Rather, we suggest that the negative pairwise interactions between maternal *mtTFB1/mtRPOL* homozygotes and the cytoplasm may be a consequence of higher order interactions. Interactions between the *mtTFB1/mtRPOL* linkage group that are benign on a non-hybrid genetic background may be modified in the context of a hybrid nucleus. In other words, the reduced viabilities of maternal homozygotes may be expressed only in the presence of paternal alleles at one or more nuclear loci that modify the interaction. Although relative viability data alone are equivocal, we suggest that the Burke et al. (1998) glass-half-full argument (increased fitness of heterospecific gene combinations) and our glass-half-empty explanation (decreased fitness of parental genotypes) may reflect fundamental differences in the genetic basis of hybrid fitness in these species.

Cytoplasmic effects observed for *mtRPOL/mtTFB1* genotypes are similar to effects observed for a linkage group marked by the *CYC* locus (Willett and Burton 2001). *CYC* is also expressed in mitochondria and functions in the transfer of electrons from Complex III to Complex IV in the electron transport chain. It is interesting that both linkage groups marked by genes expressed in mitochondria show cytoplasmic effects (Willett and Burton 2001, 2003). This also raises the question whether genes involved in cyto-nuclear interactions are widespread in the *T.*

californicus genome. Burton (1987) scored two allozyme markers, malic enzyme (*Me*) and glutamate-oxaloacetate transaminase (*Got-2*), in F_2 hybrids from the SD X SC cross and a cross between AB and a La Jolla, California population. Although strong viability effects were observed at the *Me* locus no reciprocal cross effects were observed. Willett and Burton (2001) found evidence of a reciprocal cross effect at a histone H1 marker, but a replicate experiment did not recover the same effect (Willett and Burton 2001). Finally, Edmands et al. (2005) reported no effect of cytoplasmic background on changes in allele frequencies at seven marker loci in a population cage experiment. These observations suggest that the cytoplasmic effects are not generally observed across the genome in this species and suggest that *mtRPOL*, *mtTFB1*, or another linked gene expressed in the mitochondria are likely responsible for the observed effects.

Tight linkage of *mtTFB1* and *mtRPOL* is interesting given their presumed functional relationship. Despite their close physical association of less than 0.5 cM, these genes likely mark a relatively large chromosomal region in our segregation study. Application of a typical recombination rate from flies and humans of 1.5 cM/Mb (Nachman 2002) suggests that these two genes could be 300 kb or more apart. It is intriguing, however, that functionally related genes are found in such close association in the genome as genes with similar functions are rarely clustered in eukaryotes. However, one of the more striking examples is the clustering of genes responsible for transcription, RNA processing, and mitochondrial genome maintenance in *Arabidopsis*. Elo et al. (2003) reported that at least 50 genes with

mitochondrial targeting capacity were clustered in a six Mb interval on *Arabidopsis* chromosome III. Most of these genes are involved in some component of mitochondrial RNA or DNA metabolism and included putative homologs to mitochondrial RNA polymerase and the mitochondrial transcription factor, *TFAM*. The reasons for this unusual arrangement are not understood (Elo et al. 2003). In *Drosophila*, *TFAM* and *mtTFB2* are tightly linked on chromosome arm 3R (<4 Mbp apart), but *mtTFB1* (3L) and *mtRPOL* (2L) are not linked. Although these observations are intriguing, future tests of the significance of clustering of genes involved in mitochondrial RNA metabolism will require further analysis of whole genomes.

Mitochondrial RNA polymerases assume an open right hand configuration typical of bacteriophage RNA polymerases. This three-dimensional structure is comprised of distinct subdomains including the fingers, thumb, and palm domains that have well-characterized functions in T3/T7 RNA polymerase. Mitochondrial RNA polymerases have an additional variable length *N*-terminal extension that is not found in bacteriophage RNA polymerase (Cermakian et al. 1997). The function of this domain has not been studied extensively, but work in yeast suggests dual roles in post-transcriptional RNA processing and maintenance of mitochondrial DNA (Wang and Shadel 1999, Rodeheffer and Shadel 2003).

In contrast to eubacterial RNA polymerases, mitochondrial RNA polymerase makes contact with specific nucleotides in the core promoter and is capable of promoter recognition and abortive transcription in the absence of auxiliary factors

(Matsunaga and Jaehning 2004, Gaspari et al. 2004). A primary role for promoter recognition by mitochondrial RNA polymerase is also supported by the demonstration that the polymerase itself is responsible for the species-specificity of transcription (Gaspari et al. 2004). Although the precise role of transcription factors in promoter utilization is unknown, it is likely that the mode of promoter recognition of mitochondrial RNA polymerase is similar to the well-described mechanism of T7 RNA polymerase. The RNA polymerases of bacteriophages T3 and T7 are highly specific and neither polymerase can initiate transcription from the heterologous promoter (Raskin et al. 1992). The specificity of promoter recognition in bacteriophages is determined by the specificity-loop, which forms a beta-hairpin that makes direct contacts with bases in the major groove of DNA (Cheetham et al. 1999). In fact, the specificity of transcription in bacteriophages appears to have a simple genetic basis. A single amino acid substitution in the specificity loop completely reverses the specificity of transcription (Raskin et al. 1992) as do single base changes in the promoter (Klement et al. 1990). Thus, the specificity loop is a candidate for the specificity of transcription in the mouse-human comparison and is also of interest in our comparison of *mtRPOL* sequences from *T. californicus* populations.

In general, we found only limited evidence for divergence in domains with RNA polymerase activity among populations (Table 6). We observed a single Ser/Thr substitution (position 1047, Table 6) in the specificity loop in our inter-population comparisons indicating that this region has not diverged substantially despite hypothesized interactions with rapidly evolving mtDNA. Outside of the conserved

RNA polymerase domains, levels of divergence are considerably higher. *dN/dS* estimated from *N*-terminal domain is 0.331, which contrasts with a *dN/dS* 0.046 from combined analysis of the three RNA polymerase domains (i.e., fingers, thumb and palm domains). Our analysis suggests that a relaxation of purifying selection and not positive selection is responsible for the different mode of evolution of these regions. In fact, despite moderate to high levels of amino acid divergence among populations, we found no evidence that positive selection has contributed to the divergence of either *mtRPOL* or *mtTFB1*. Analysis based on the codon-substitution model of Goldman and Yang (1994) are conservative, however, and only have the power to detect selection if mutations at individual sites occur recurrently and are carried to fixation by selection. Irregardless of whether adaptive evolution has contributed to diversification at these loci, it is possible that the extent of divergence may have consequences for the function of these genes in hybrids. Future work will focus on the potential functional significance of divergence of genes in the mitochondrial transcription apparatus of *T. californicus*.

ACKNOWLEDGEMENTS

We thank C. S. Willett, J. S. Harrison, and G. H. Pogson for helpful discussions. We also thank C. S. Willett, D. Peterson, J. S. Harrison and C. Ellison who helped with sample collection. We also thank J. S. Harrison who kindly provided the cDNA library used for obtaining the 5' end of *mtRPOL*. We thank L. Dingding, Sharon, T. Ballard, and H. Huynh for assistance with cultures and R. Byrne for laboratory assistance.

REFERENCES

Ar-rushdi, A.H. 1963. The cytology of achiasmatic meiosis in the female *Tigriopus* (Copepoda). Chromosoma. 13:526

Beeman, R. W., K. S. Friesen, and R. E. Denell. 1992. Maternal-effect selfish genes in flour beetles. Science. 256:89-92.

Blier, P. U., F. Dufresne, and R. S. Burton. 2001. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends in Genetics. 17:400-406.

Breeuwer, J. A. and J. H. Werren. 1995. Hybrid breakdown between two haplodiploid species: the role of nuclear and cytoplasmic genes. Evolution. 49:705-717.

Burke, J.M., T.J. Voss, and M.L. Arnold. 1998. Genetic interactions and natural selection in Louisiana iris hybrids. Evolution. 52:1304-1310.

Burke, J.M. and M.L. Arnold. 2001. Genetics and the fitness of hybrids. Ann. Rev. Genet. 35:31-52.

Burton, R.S. 1987. Differentiation and integration of the genome in populations of the marine copepod *Tigriopus californicus*. Evolution. 41:504-513.

Burton, R.S. 1990. Hybrid breakdown in developmental time in the copepod *Tigriopus californicus*. Evolution. 44(7): 1814-1822.

Burton R.S. 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. Evolution. 52:734-745.

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

Burton, R. S., M.W. Feldman, and S.G. Swisher. 1981. Linkage relationships among five enzyme-coding gene loci in the copepod *Tigriopus californicus*: A genetic confirmation of achiasmatic meiosis. Biochemical Genetics. 19:1237-1245.

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

Cheetham, G. M. T., D. Jeruzalmi, and T. A. Steitz. 1999. Structural basis for

initiation of transcription from an RNA polymerase-promoter complex. Nature. 399:80-83.

Clark, A.G. and E. M. S. Lyckegaard. 1988. Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mitochondrial DNA in *Drosophila melanogaster*. Genetics. 118:471-481.

Cermakian, N. T. M. Ikeda, P. Miramontes, B.F. Lang, M.W. Gray, and R. Cedergren. 1997. On the evolution of the single-subunit RNA polymerases. J. Mol. Evol. 45:671-681.

Claros, M.G. and P. Vincens. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. 241:779-786.

Edmands, S. 1999. Heterosis and outbreeding depression in inter-population crosses spanning a wide range of divergence. Evolution. 53:1757-1768.

Edmands, S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. Molecular Ecology. 10: 1743-1750.

Edmands, S. and R. S. Burton. 1999. Cytochrome c oxidase activitiy in interpopulation hybrids of a marine copepod: a test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. Evolution. 53:1972-1978.

Edmands, S., H. V. Feaman, J. S. Harrison, and C. C. Timmerman. 2005. Genetic consequences of many generations of hybridization between divergent copepod populations. J. Heredity. 96:114-123.

Elo, A., A. Lyznik, D. O. Gonzalez, S. D. Kachman and S. A. Mackenzie. 2003. Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the *Arabidopsis* genome. Plant Cell. 15:1619-1631.

Falkenberg, M., M. Gaspari, A. Rantanen, A. Trifunovic, N. G. Larsson, C. M. Gustafsson. 2002. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. Nature Genetics. 31:289-294.

Fisher, R. P., M. A. Parisi, and D. A. Clayton. 1989. Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. Genes Dev. 3:2202-2217.

Galloway, L.F. and C.B. Fenster. 2001. Nuclear and cytoplasmic contributions to intraspecfici divergence in an annual legume. Evolution. 55:488-497.

Ganz, H.H., and R.S. Burton. 1995. Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod *Tigriopus californicus*. Mar. Biol. 123:821-827.

Gaspari, M., M. Falkenberg, N. G. Larsson, C. M. Gustafsson. 2004. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. EMBO. 23:4606-4614.

Goldman, N. and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725-736.

Hutter, C.M. and D.M. Rand. 1995. Competition between mitochondrial haplotypes in distinct nuclear genetic environments: *Drosophila pseudoobscura* vs. *D. persimilis*. Genetics. 140:537-548.

Jaehning, J. A. 1993. Mitochondrial transcription: is a pattern emerging? Mol. Microbiol. 8:1-4.

James, A. C. and J. W. O. Ballard. 2003. Mitochondrial genotype affects fitness in *Drosophila simulans*. Genetics. 164:187-194.

Jeruzalmi, D. and T. A. Steitz. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO. 17:4101-4113.

Kenyon, L. and C. T. Moraes. 1997. Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitcohondrial cybrids. Proc. Natl. Acad. Sci. USA. 94:9131-9135.

Klement, J. F., M. B. Moorefield, E. Jorgenson, J. E. Brown, S. Risman, and W. T. McAllister. 1990. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. J. Mol. Biol. 215:21-29.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Bioinformatics Vol. 17. 12:1244-1245.

Masters, B.S., L. L. Stohl, and D. A. Clayton. 1987. Yeast mitochondrial RNA polymerase is homologous to those encoded by bateriophages T3 and T7. Cell. 51:89-99.

Matsunaga, M. and J. A. Jaehning. 2004. Intrinsic promoter recognition by a "core"

RNA polymerase. J. Biol. Chem. 279:44239-44242.

Matsushima, Y., D. Adan, R. Garesse, and L. S. Kaguni. 2005. *Drosophila* mitochondrial transcription factor B1 modulates translation but not transcription or DNA copy number in Schneider cells. J. Biol. Chem. 280:16815-16820.

Matsushima, Y., R. Garesse, and L. S. Kaguni. 2004. *Drosophila* mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in Schneider cells. J. Biol. Chem. 279:26900-26905.

McKenzie, M., M. Chiotis, C. A. Pinkert, and I. A. Trounce. 2003. Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome b and nuclear subunits of complex III. Mol. Biol. Evol. 20:1117-1124.

Nachman, M.W. 2002. Variation in recombination rate across the genome: evidence and implications. Curr. Opin. Genet. Devel. 12:657-663.

Nielsen, R. and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 148:929-936.

Pogson, G. H. and K. A. Mesa. 2004. Positive Darwinian selection at the Pantophysin (Pan I) locus in marine gadid fishes. Mol. Biol. Evol. 21:65-75.

Rand, D. M., A. G. Clark, and L. M. Kann. 2001. Sexually antagonistic cytonuclear fitness interactions in *Drosophila melanogaster*. Genetics. 159:173-187.

Rand, D. M., R.A. Haney, and A.J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol. Evol. 19:645-653.

Rantanen, A., M. Gaspari, M. Falkenberg, C. M. Gustafsson, and N. G. Larsson. 2003. Characterization of the mouse genes for mitochondrial transcription factors B1 and B2. Mammalian Genome. 14:1-6.

Raskin, C. A., G. Diaz, K. Joho, and W. T. McAllister. 1992. Substitution of a single T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. J. Mol. Biol. 228:506-515.

Rawson, P. D. and R. S. Burton. 2000. Isolation and characterization of cytochrome *c* from the marine copepod *Tigriopus californicus*. Gene. 248:15-22.

Rawson, P.D. and R.S. Burton. 2002. Functional coadaptation between cytochrome c

and cytochrome c oxidase within allopatric populations of a marine copepod. Proc. Natl. Acad. Sci. USA. 99:12955-12958.

Rieseberg, L. H., B. Sinervo, C. R. Linder, M. Ungerer, and D. M. Arias. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. Science. 272:741-745.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics. 152:713-727.

Rocha-Olivares, A., J. W. Fleeger, and D. W. Foltz. 2001. Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod. Mol. Biol. Evol. 18:1088-1102.

Rodeheffer, M.S. and G.S. Shadel. 2003. Multiple interactions involving the amino-terminal domain of yeast mtRNA polymerase determine the efficiency of mitochondrial protein synthesis. J. Biol. Chem. 278:18695-18701.

Sackton, T.B., R.A. Haney and D.M. Rand. 2003. Cytonuclear coadaptation in *Drosophila*: Disruption of cytochrome c oxidase activity in backcross genotypes. Evolution. 57:2315-2325.

Schmidt, T.R., W. Wu., M. Goodman, and L. Grossman. 2001. Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. Mol. Biol. Evol. 18:563-569.

Swofford, D. L., 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24: 4876-4882.

Wang, Y. and G. S. Shadel. 1999. Stability of the mitochondrial genomes requires an amino-terminal domain of yeast mitochondrial RNA polymerase. Proc. Nat. Acad. Sci. USA. 96:8046-8051.

Willett, C. S. and R. S. Burton. 2001. Viability of cytochrome c genotypes depends on cytoplasmic background in *Tigriopus californicus*. Evolution. 55:1592-1599.

Willett, C. S. and R. S. Burton. 2003. Environmental influences on epistatic interactions: viabilities of cytochrome c genotypes in interpopulation crosses. Evolution. 57:2286-2292.

Willett, C. S. and R. S. Burton. 2004. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. Mol. Biol. Evol. 21:443-453.

Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comp. Appl. Biosci. 13:555-556.

Yang, Z. and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among sites. Mol. Biol. Evol. 19:49-57.

Zeyl, C., B. Andreson, and E. Weninck. 2005. Nuclear-mitochondrial epistasis for fitness in *Sachharomyces cerevisiae*. Evolution. 59:910-914.

Table 1. *mtRPOL/mtTFB1* genotypic counts in F_2 inter-population hybrids with Chi-square goodness-of-fit to the Mendelian expectation. Recombinants between *mtRPOL* and *mtTFB1* ($n = 11$) are excluded.

		Genotypic counts			Chi-Square test of 1:2:1		Chi-Square test of males versus females	
Cross	Sex	*SD/SD*	*SD/AB*	*AB/AB*	χ^2	*P*	χ^2	*P*
SDf X ABm	Males	47	159	76	10.56	0.0051		
	Females	61	105	61	1.27	n.s.		
	Total	108	264	137	4.01	n.s.	8.661	0.0132
ABf X SDm	Males	14	59	16	9.54	.0085		
	Females	36	65	28	1.00	n.s.		
	Total	50	124	44	4.46	n.s.	6.109	.0471
		SD/SD	*SD/SC*	*SC/SC*				
SDf X SCm[a]	Males	23	103	46	12.87	.0016		
	Females	25	109	54	13.73	.001		
	Total	48	212	100	26.40	<.00001	.1824	0.9128
SCf X SDm[a]	Males	36	65	28	1.00	n.s.		
	Females	43	113	47	2.76	n.s.		
	Total	79	178	75	1.83	n.s.	1.982	0.3712

Table 1. continued.

Cross	Sex	Genotypic counts *AB/AB*	*AB/SC*	*SC/SC*	Chi-Square test of 1:2:1 χ^2	*P*	Chi-Square test of males versus females χ^2	*P*
ABf X SCm	Males	60	202	84	13.05	.0015		
	Females	46	208	98	27.00	<.00001		
	Total	106	410	182	37.87	<.00000001	2.962	0.2274
SCf X ABm	Males	50	98	31	5.65	.0594		
	Females	24	88	32	8.00	.0183		
	Total	74	186	63	8.18	.0167	5.966	.0506

[a]Results are pooled from two separate experimental replicates (see Table 2)

Table 2. Relative viabilities of *mtRPOL/mtTFB1* genotypes in F_2 inter-population hybrids relative to the heterozygote class. Relative viability estimates and their standard deviations (in parentheses) are calculated according to Haldane (1956) by setting the viability of heterozygote to one. Males and females are pooled when there was no evidence of sex differences in Table 1.

Cross	Cytoplasm origin	Sex	Relative viability of homozygote classes		
			SD/SD	*AB/AB*	*SC/SC*
SDf X ABm	SD	Males	0.588 (.08)	0.95 (.12)	-
		Females	1.151 (.18)	1.151 (.18)	-
ABf X SDm	AB	Males	0.467 (.12)	0.533 (.13)	-
		Females	1.091 (.22)	0.848 (.18)	-
SDf X SCm	SD	Total	0.451 (.06)	-	0.939 (.11)
SCf X SDm	SC	Total	0.883 (.11)	-	0.838 (.11)
ABf X SCm	AB	Total	-	0.516 (.05)	0.886 (.07)
SCf X ABm	SC	Total	-	0.791 (.10)	0.674 (.09)

Table 3. Test of homogeneity of *mtRPOL/mtTFB1* genotypic counts from replicate experimental crosses by contingency Chi-square. Recombinants between *mtRPOL* and *TFB1* are excluded.

			Genotypic counts[a]				
Cross	Replicate	Sex	*SD/SD*	*SD/SC*	*SC/SC*	χ^2	*P*
SDf X SCm	1	Male	11	50	23		
	2	Male	12	53	23	0.04	0.9813
	1	Female	14	60	33		
	2	Female	11	49	21	0.55	0.759
	1	Total	25	110	56		
	2	Total	23	102	44	0.48	0.7856
SCf X SDm	1	Male	18	37	20		
	2	Male	18	28	8	3.05	0.2175
	1	Female	20	55	18		
	2	Female	23	58	29	1.45	0.4843
	1	Total	38	92	38		
	2	Total	41	86	37	0.28	0.8688

[a]Pooled results from replicate crosses are presented in Table 1.

Table 4. *mtRPOL/mtTFB1* genotypic counts in nine broods of F_2 inter-population hybrid first stage nauplii from the SDf X SCm cross.

	Genotypic counts		
Brood	*SD/SD*	*SD/SC*	*SC/SC*
1	9	11	8
2	8	11	4
3	3	9	3
4	0	13	11
5	5	11	12
6[a]	5	13	2
7	12	18	12
8	7	7	3
9	5	19	8
Total[b]	54	112	63

[a]one recombinant genotype was observed

[b]$\chi^2_{1:2:1} = 0.817$, $P = 0.665$, $df = 2$

Table 5. Inter-population distances at *mtRPOL* and *mtTFB1*. Lower half matrix are Jukes-Cantor corrected distances from 3,563 bp of the *mtRPOL* including 51 bp of the 5'UTR. Distances in the upper half matrix are based upon 1,268 bp of *mtTFB1* including both introns and 22 bp of the 3'UTR.

	PA	PBJ	PC	PM	SD	LB	RP	AB	CRP	SS	SCN	PES	BH
PA[ab]	-	-	-	-	-	-	-	-	-	-	-	-	-
PBJ	0.118	-	-	-	-	-	-	-	-	-	-	-	-
PC	0.116	0.01	-	0.006	0.028	0.029	-	0.030	-	0.028	0.032	-	0.027
PM	0.115	0.015	0.011	-	0.028	0.029	-	0.032	-	0.029	0.033	-	0.029
SD	0.120	0.025	0.021	0.02	-	0.002	-	0.027	-	0.028	0.031	-	0.028
LB	0.120	0.023	0.019	0.018	0.005	-	-	0.028	-	0.029	0.032	-	0.029
RP	0.117	0.022	0.018	0.016	0.019	0.017	-	-	-	-	-	-	-
AB	0.115	0.02	0.015	0.014	0.017	0.014	0.007	-	-	0.032	0.031	-	0.029
CRP	0.115	0.022	0.017	0.016	0.018	0.017	0.013	0.011	-	-	-	-	-
SS	0.114	0.026	0.022	0.021	0.025	0.023	0.021	0.019	0.02	-	0.018	-	0.017
SCN	0.112	0.023	0.019	0.019	0.023	0.021	0.019	0.017	0.019	0.011	-	-	0.017
PES	0.113	0.024	0.019	0.02	0.024	0.022	0.02	0.018	0.019	0.012	0.003	-	-
BH	0.114	0.025	0.021	0.021	0.024	0.022	0.021	0.018	0.02	0.012	0.008	0.01	-

[a]See Figure 3 legend for population abbreviations
[b]Pairwise distances with Playa Altamira based on 3,158 bp

Table 6. Amino acid polymorphism at *mtRPOL* in *T. californicus* populations. Codons are numbered beginning with the methionine start codon. Polmorphisms include those observed at 1,170 codons in the global alignment (excludes six codons at the *C*-terminus). The Playa Altamira population has been excluded because of large amino acid divergence. Structural domains correspond to the following codon positions: mitochondrial targeting sequence = 1-53; *N*-terminal extension = 54-640; thumb = 641-738, palm = 739-885, 1,067-1,118; fingers = 886-1031, 1053-1066; specificity loop = 1,032-1,052; extended foot = 1,119-1,170. Population abbreviations are listed in the Figure 3 legend.

Sequence ID	Codon position[a]									
			11	1111222222	2222222233	3333333333	3334444445	5555566667	7888999001	111
	111123344	4456789934	4568000122	3334556903	3445666778	8992567882	3457817890	1135067492		
	3245950902	3823460647	8887246258	0257360427	9781146291	4235991046	0183458600	2205085767		
PBJ2A	RLGPIVLCPL	GSRQATVGLL	IKVSWQNTKN	ICSQTGQSIK	KSQSYPQTRI	PDVMNAISCI	VTLITVDMKT	YDKSQIGSWE		
PC1	.P...AV..V	...R.....	.....NA.N	.SA.....	NN.....PS.	....D...F.	M.....	.....		
PM1	.....V..V	.....E...	VQ...GNA.N	VSA.....Q	.N.....PSV	.H..DT.TF.	.....I..	.E.....		
SD5	.....V..V	.....QE..	...G.GNA.N	VSA....A..	NN.....PS.	....D..TF.	.....N.R.	.E.....		
LB1	..R...V..V	.....QE..	.....GNA.N	VSA....A..	NNH....PS.	....D..TF.	.....N.R.	.E..L.....		
RP1C	.....V.RV	.....E...	.....GSA.K	VSA..S..V.	NNH....PS.	..M.D..TF.	....S.N.RK	.EM.....R.		
AB2	....S.I.RV	.P....E..F	....CGNA.K	VSA.....V.	NNH....PS.	....D..TF.	...V..N.R.	.E.....		
CA1	.....VR.V	....EIE.M.	.....GNA.?	VSAR..H.V.	NNH..L.PS.	R...D..TF.	.....N.R.	.E....S...		
SS1B	.....V..V	A.....E...	..A..GNARN	VSA.....	NN..C..PS.	....D.VTF.	.R....N.R.	.E.P.L.T..		
SCN2	.....V..V	A.K...E...	..A..GDARN	VSA.A.....	NN.G..KPS.	....D.VTF.	.....N.R.	.E.P.L.T..		
PES1*	H..R..V..V	A.....E...	..A..GDARN	VSA.A.....	N?.G..KPS.	....D.VTF.	.....N.R.	FE.P.L.T..		
BH1*	.....V..V	A.....E...	..A..GDARN	VSA.A.....	NN.....PS.	...ID.VTFM	..Q..DN.R.	.E.P.L.T.K		

*PES1 and BH1 were polymorphic for Ala/Thr at position 253 and Ile/Thr at position 480

Table 7. Amino acid polymorphism at the *mtTFB1* locus. Codon positions are numbered beginning with the methionine start codon in the full length cDNA sequence from SD. Population abbreviations are listed in the Figure 3 legend.

	Codon position
Sample ID	122333 7889808156 9599515850
PC	LANQRFHYNA
PM50	..N.....N.
SD60	V.S.Q...KE
LB	V.S.Q...KE
AB16	V.GPP...K.
SS	VDS.Q...K.
SCN3	VDS.QY..K.
BH	V.S.QYQFK.

Table 8. Tests of positive selection at *mtTFB1* and *mtRPOL*. LRTs of the log-likelihood scores of models M1a and M2a were all nonsignificant.

Gene/Domain	Model[1]	l	Parameter Estimates[2]
mtTFB1	M0	-1881.81	$\omega = 0.1011$
	M1a	-1876.7	$p_0 = 0.897$, ($p_1 = 0.103$), $\omega_0 = 0.00$
	M2a	-1876.63	$p_0 = 0.912$, $p_1 = 0.00$ ($p_2 = 0.088$), $\omega_0 = 0.00$, $\omega_2 = 1.2428$
mtRPOL			
All domains	M0	-6639.54	$\omega = 0.1715$
	M1a	-6636.22	$p_0 = 0.903$, ($p_1 = 0.097$), $\omega_0 = 0.089$
	M2a	-6636.22	$p_0 = 0.903$, $p_1 = 0.048$, ($p_2 = 0.0481$), $\omega_0 = 0.0889$, $\omega_2 = 1.00$
N-terminal	M0	-3733.09	$\omega = 0.3312$
	M1a	-3731.3	$p_0 = 0.747$, ($p_1 = 0.253$), $\omega_0 = 0.115$
	M2a	-3731.3	$p_0 = 0.747$, $p_1 = 0.165$, ($p_2 = 0.088$), $\omega_0 = 0.1154$, $\omega_2 = 1.00$
Thumb	M0	-528.12	$\omega = 0.0959$
	M1a	-528.12	$p_0 = 1.00$, ($p_1 = 0.00$), $\omega_0 = .0959$
	M2a	-528.12	$p_0 = 1.00$, $p_1 = 0.00$, ($p_2 = 0.00$), $\omega_0 = 0.0959$, $\omega_2 = 1.00$
Palm	M0	-1352.77	$\omega = 0.0313$
	M1a	-1352.77	$p_0 = 1.00$, $p_1 = 0.00$, $\omega_0 = .0313$
	M2a	-1352.77	$p_0 = 1.00$, $p_1 = 0.00$, ($p_2 = 0.00$), $\omega_0 = 0.0313$, $\omega_2 = 1.00$
Fingers	M0	-955.23	$\omega = 0.0425$
	M1a	-954.71	$p_0 = 0.944$, $p_1 = 0.056$, $\omega_0 = 0.00$
	M2a	-954.71	$p_0 = 0.945$, $p_1 = 0.052$, ($p_2 = 0.003$), $\omega_0 = 0.00$, $\omega_2 = 1.0177$

[1]All analyses conducted using base frequencies estimated separately for each codon position (i.e., CodonFreq variable set to F3X4 in PAML)
[2]$\omega = dN/dS$, p = proportion of sites in the specified rate class

Table 9. Results from tests of heterogeneity in the pattern of substitution between the *N*-terminal and *C*-terminal domains (see text) of *mtRPOL* using the Fixed-Site models of Yang and Swanson (2002).

		Likelihood Ratio Tests			
Model	Log-likelihood	Comparison	2Δl	*df*	*P*
A (homogeneity of partitions)	-6639.54	A versus B	1.21	1	0.271
B (different rates)	-6638.93	B versus C	2.43	9	0.983
C (different rs^1 πs^2)	-6637.72	B versus D	42.06	2	7.35E-10
D (different *r*s, κ^3, ω)	-6617.90	C versus E	45.08	2	1.627E-10
E (different *r*s, κ, ω, πs)	-6615.18	E versus F	24.58	20	0.218
F (separate analyses)	-6602.89				

[1] *r* is the total rate
[2] π is the expected codon frequency table based upon the F3X4 means of estimating base frequencies
[3] κ is the transition/transversion rate

Figure 1. Restriction map for multiplexed PCR-RFLP genotyping of inter-population hybrids at *mtRPOL* and *mtTFB1*. Genotyping of F_2 hybrids was conducted by simultaneously amplifying both genes with primers flanking restriction sites that were fixed for alternate polymorphisms in each population. Hybrids from the SD X SC and SD X AB crosses were genotyped by amplifying with primers A, B, C, and D followed by digestion with *HinFI*. Hybrids from the AB X SC cross were genotyped by amplifying with primers E, F, G, and H followed by digestion with *MspI*.

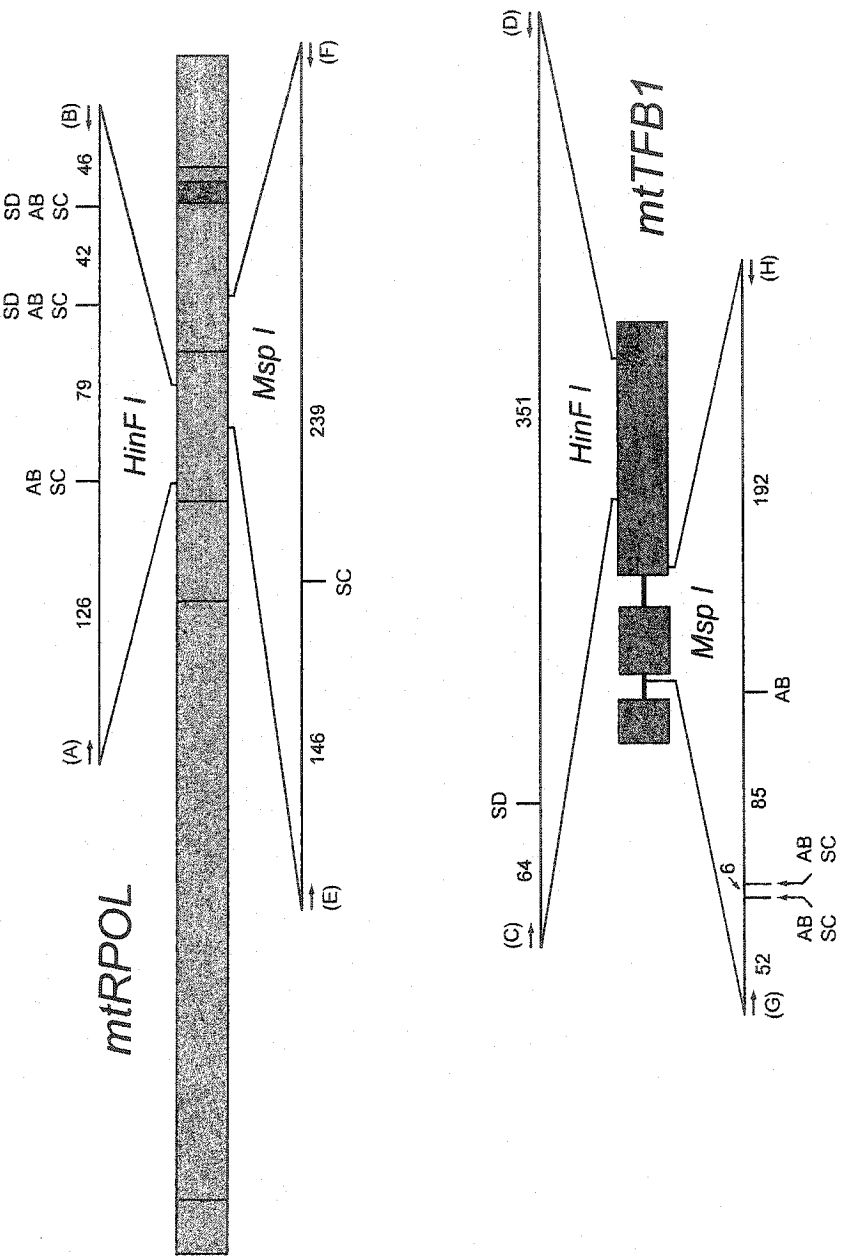


Figure 2. Structure and organization of *mtRPOL* and *mtTFB1* genes in *T. californicus*. Vertical lines above *mtRPOL* represent amino acid substitutions between the SD and SC populations. Lines below represent amino acid substitutions between the SD and Playa Altamira populations. Subdomains in *mtRPOL* were identified following Jeruzalmi and Steitz (1998).

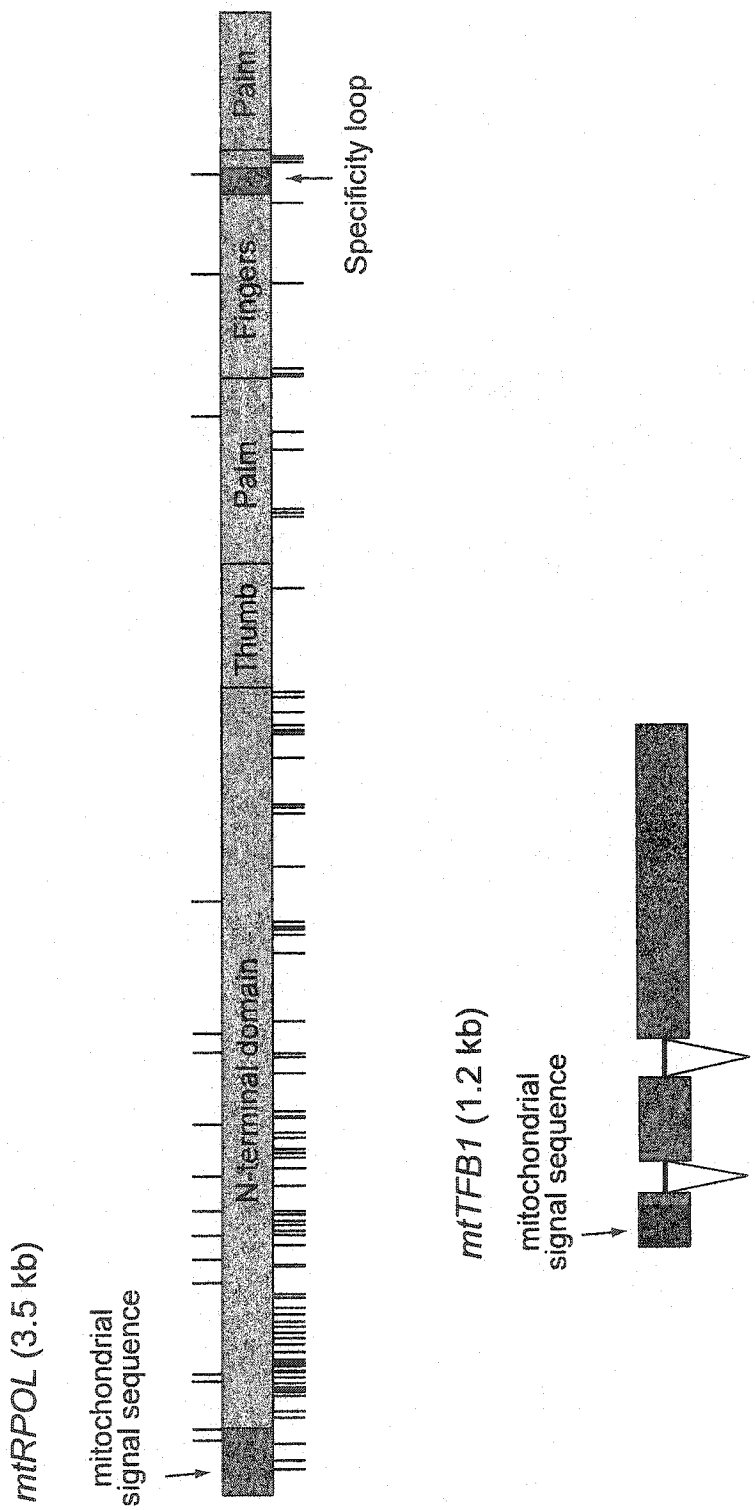
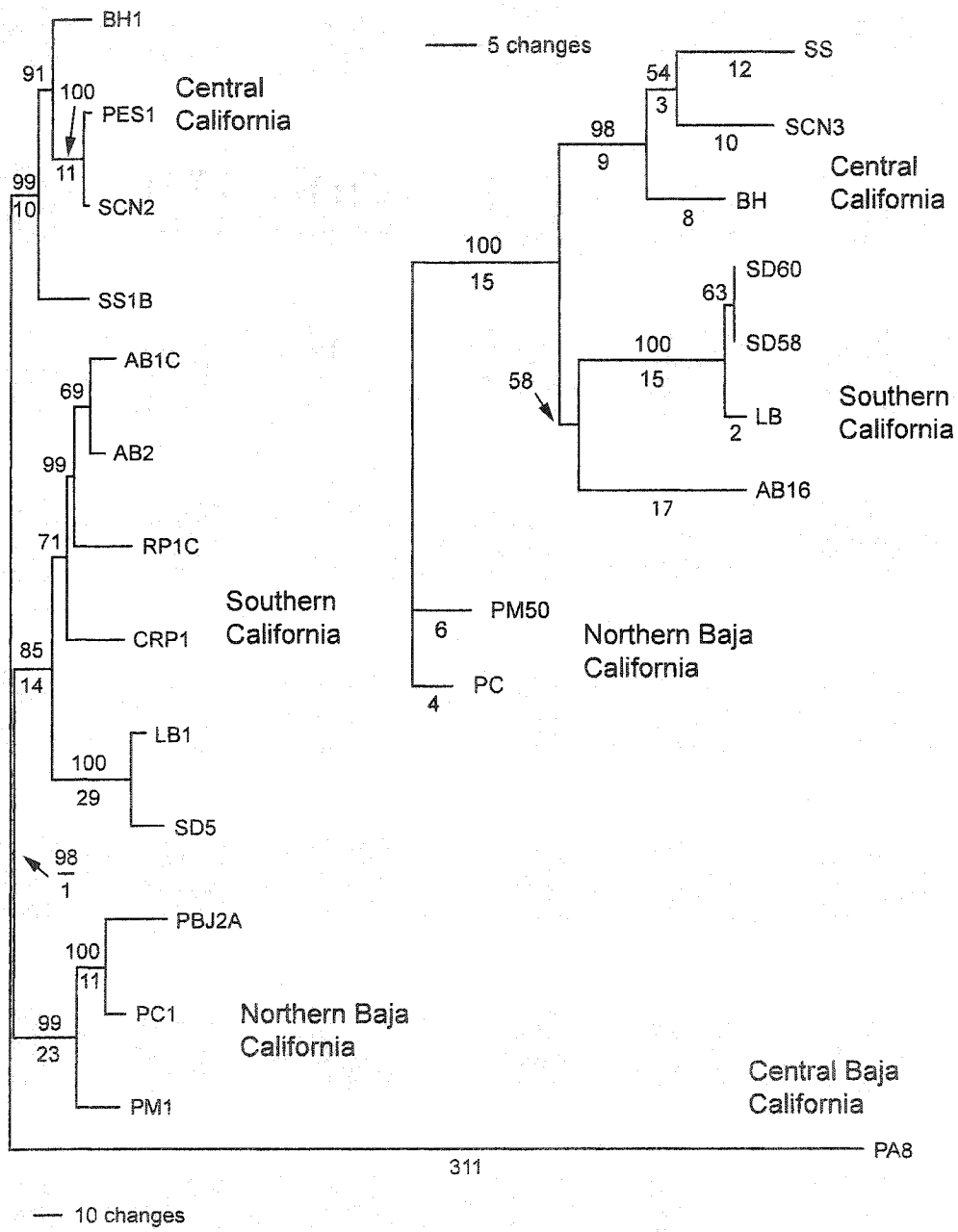


Figure 3. Phylogeny of *T. californicus* populations based on mitochondrial transcription genes. (A) Maximum parsimony tree (C.I. = .929, R.I. = .815) based upon 3,158 bp of the *mtRPOL* structural gene obtained using the branch-and-bound search algorithm. Values above the nodes are the bootstrap proportions from 1,000 replicates supporting each node. Below the line are the total number of steps along each branch. (B) One of three most parsimonious trees (C.I. = .923, R.I. = .889) based upon an alignment of 1,268 bp of the *mtTFB1* gene. Trees differed in their placement of the AB population (either as sister group to the SD/LB or central California clade) and the arrangement within the central California clade. Population abbreviations are: PA = Playa Altamira, PM = Punta Morro, PC = Punta Cabras, PBJ = Punta Baja, SD = San Diego (Ocean Beach Pier site), LB = Laguna Beach, CRP = Carpinteria State Beach, RP = Royal Palms, AB = Abalone Cove, SS = San Simeon, SCN = Santa Cruz (Swift Street site), PES = Pescadero, BH = Bodega Head.



APPENDIX A

The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus*

ABSTRACT

As a consequence of free-spawning in the unpredictable nearshore environment, marine species with large fecundities and high pre-reproductive mortality may be subject to extreme variance in reproductive success. If the unpredictability of the ocean results in only a small subset of the adult population contributing to each larval cohort, then reproduction may be viewed as a sweepstakes, with chance events determining which adults are successful each spawning season. Such a reproductive sweepstakes scenario may partially account for large reductions in effective population sizes relative to census population sizes in marine species. We evaluated two predictions of the sweepstakes reproductive success hypothesis by testing (1) whether sea urchin recruits contain reduced genetic variation relative to the adult population and (2) whether cohorts of sea urchin recruits are genetically differentiated. Mitochondrial DNA sequences were collected from 283 recently settled *Strongylocentrotus purpuratus* recruits from four annual cohorts spanning seven years in locations throughout California, USA. Observed haplotype numbers and haplotype diversities showed little evidence of reduced genetic variation in the recruits relative to the diversity estimated from a previously reported sample of 145 *S. purpuratus* adults. Different cohorts of recruits were in some cases mildly differentiated from each other. A computer simulation of sweepstakes recruitment indicates that our sampling strategy had sufficient statistical power to detect large variances in reproductive success.

INTRODUCTION

Marine species with high fecundity and high early mortality are susceptible to large variance in reproductive success. This has been attributed to the unpredictability of oceanographic conditions necessary for spawning, fertilization, larval development, and recruitment (Cushing 1990). Asynchronies between reproductive activity and suitable environmental conditions could lead to individual reproductive failure as a result of sperm limitation, variation in the availability of food for larvae, unpredictable nearshore oceanographic features, and predation. Each of these factors may have dramatic consequences on recruitment success and lead to reproductive failure by a significant fraction of the adult population. If such differential success results in few reproductively successful individuals, then reproduction in free-spawning species may be viewed as a sweepstakes, in which chance events determine which adults are successful each spawning season (Hedgecock 1994).

Variance in reproductive success influences genetic diversity through its effects on effective population size (N_e). In an ideal population, loss of variation due to genetic drift will be negligible if the population experiences Poisson or binomial variance in progeny number (Crow and Kimura 1970). In contrast, genetic drift may have important effects on equilibrium levels of variation, even if population sizes are large, when the variance in progeny number exceeds Poisson or binomial variance. Hedgecock (1994) argued that free-spawning marine species may have variances in progeny number much greater than Poisson or binomial variance as a result of enormous fecundities, high pre-reproductive mortality (i.e., type III survivorship), and

the unpredictability of the marine environment. If so, variance in reproductive success may have dramatic effects on N_e/N, especially if a species has small generation overlap (Gaggiotti and Vetter 1999). If large variance in reproductive success is common in marine organisms, then it may at least partially explain the frequently observed 10^2 to 10^5 reductions in estimates of N_e relative to census population sizes of marine species (Avise et al. 1988; Hedgecock 1994; Turner et al. 1999; Avise 2000).

Wide variation in reproductive success may have important consequences for life history evolution of marine species (Murphy 1968). Since larval mortality rates are high and the probability of recruitment is low, distributing reproductive efforts over many reproductive seasons (i.e., bet hedging) may be essential for successful reproduction in species that have a planktonic larval stage. This may provide long-lived individuals with a selective advantage over short-lived individuals, particularly if adult mortality rates are low relative to pre-reproductive mortality rates. This explanation for the evolution of longevity is attractive and predicts that large variance in reproductive success is a common feature of long-lived species with passively dispersing larvae.

If large variance in reproductive success is a general feature of reproduction in species with dispersive larval stages, then we expect this phenomenon to have two effects on the genetic composition of marine larvae. First, high variance in offspring number is predicted to result in the effective contribution of only a subset of the breeding population each year. If this is true, then marine larvae should reflect this effect and exhibit reduced genetic variation relative to the adult population

(Hedgecock 1994). Second, since environmental stochasticity is predicted to determine which individuals are successful each breeding season, different sets of individuals are predicted to be successful in different years. Random sampling error, then, should result in cohorts that are genetically differentiated over time (Li and Hedgecock 1998). Hence, if sweepstakes events are prominent features of recruitment in the sea, then their signatures may be detected by genetic sampling of larvae or recently settled recruits.

Sea urchins in the genus *Strongylocentrotus* are long-lived (Ebert 1967; Ebert et al. 1999), free-spawning species with early mortality, enormous fecundities (Kato and Schroeter 1985; Yakovlev 1987), long durations in the plankton (Strathmann 1978), and a strong dependence on appropriate environmental conditions for successful recruitment (Ebert et al. 1994; Wing et al. 1995; Miller and Emlet 1997; Morgan et al. 2000). Like many other free-spawning marine species, these life history characteristics make *S. purpuratus* susceptible to large variances in reproductive success. Genetic variation in age structured populations of *S. purpuratus* and a congener, *S. franciscanus*, have revealed patterns of variation that are consistent with predictions of the sweepstakes reproductive success hypothesis (Edmands et al. 1996; Moberg and Burton 2000). Yet, the generality of these observations, the mechanisms generating them, and the magnitude of the variance in progeny number are not understood.

Here we extend genetic analysis of *S. purpuratus* populations to newly settled recruits (from 1 to 14 days old) to test two predictions of the sweepstakes reproductive

success hypothesis. We evaluate whether genetic diversity is reduced in recently settled recruits relative to the adult population and determine whether different cohorts are genetically differentiated. Genetic variation at the cytochrome oxidase subunit I gene (COI) of the mitochondrial genome (mtDNA) was assessed for 283 purple urchin recruits from 16 recruitment events spanning seven years and seven locations along the California coast. Because mtDNA is maternally inherited, these sequences provide (1) a minimum estimate of the number of females that contributed to each cohort and (2) estimates of genetic diversity for comparison with 145 adult urchin sequences reported previously (Edmands et al. 1996).

MATERIALS AND METHODS

Urchin recruits were collected at weekly to bi-weekly intervals from settlement brushes from seven locations in California (Table 1). Individuals comprising a single annual cohort (i.e., individuals collected in the same year in the same general geographic area) were collected from locations spanning a minimum coastline distance of 14 km between Ocean Beach and Scripps Institution of Oceanography (SIO) in 2000 and a maximum of 337 km between Ocean Beach and Ellwood Pier in 1996. Specimens were identified as *S. purpuratus* based upon a lack of dorsal pedicellaria (Ebert et al. 1994) and stored in ethanol prior to molecular analysis. DNA was extracted from recruits by incubating samples for one hour at 65° C in 20 µl of lysis buffer (10 mM Tris pH 8.3, 50 mM KCl, 0.5% Tween 20, 200 µg/ml Proteinase K) (Hoelzel and Green 1992). Samples were then placed at 95° C for 15 minutes to denature the Proteinase K. After dilution with 80 µl of water, 2-5 µl was used as template for PCR. The canonical 358 base pairs of COI sequence correspond to positions 6491 to 6848 of the complete *S. purpuratus* mitochondrial genome (Jacobs et al. 1988) and were obtained by amplifying with any of four combinations of two forward primers, COIB 5'- CACAACTTTCTTTGACCCTG -3' (positions 6432 to 6451) or COIC (Edmands et al. 1996), and two reverse primers, COIA 5'- TGTATAGGCGTCTGGATAGTC -3' (positions 7128 to 7108) or COIJ (Edmands et al. 1996). The thermal cycling profile consisted of 30 s denaturation at 95° C, 30 s annealing at 50-55° C, and 30 s elongation at 72° C. Amplified products were purified with QIAquick PCR purification columns (Qiagen). One strand of the PCR products

was sequenced with Big Dye (Automated Biosystems Inc.) sequencing chemistry and run out on an ABI 373 automated sequencer. Sequences were obtained from 302 *S. purpuratus* recruits and edited with Sequencher 3.0 (Gene Codes Corp, Ann Arbor, MI). Nineteen sequences were removed from the dataset because of ambiguous nucleotides in polymorphic sites.

Statistical analysis

Population structure was assessed with the Analysis of Molecular Variance (AMOVA) algorithm and Exact Tests as implemented by Arlequin version 2.0 (Schneider et al. 2000). In the AMOVA analysis, F_{ST} analogs (ϕ_{ST}) were estimated from haplotype frequencies to evaluate both spatial and temporal variation in the genetic composition of sea urchin recruits. Statistical significance was assessed by generating 1024 replicate data sets by permutation and determining the proportion of occurrences with values greater than or equal to the observed ϕ_{ST}. Exact tests of spatial and temporal differentiation were used to test the hypothesis of equal haplotype frequencies among populations or cohorts (Raymond and Rousset 1995). In this analysis, statistical significance was assessed by determining the proportion of contingency tables that have an equal or lower probability of occurrence than the observed distribution of haplotypes as determined by 1000 iterations of a Markov Chain Monte Carlo algorithm (Schneider et al. 2000). Analyses of spatial genetic structure were conducted for the 283 sequences collected from recruits reported here. A second set of analyses were conducted on a combined dataset consisting of the

recruit DNA sequences and 145 juvenile (≤ 20 mm test diameter) and adult (> 20 mm test diameter) *S. purpuratus* DNA sequences reported in Edmands et al. (1996). These will be referred to as adults hereafter. These specimens were collected from ten intertidal or shallow subtidal locations in California, USA or from Baja California, Mexico between September 1994 and May 1995. The locations from which recruits were collected in the present study fell well within the range sampled by Edmands et al. (1996) except for the Fort Bragg and the Bodega Bay samples, which were collected north of the northernmost location. The gene region analyzed in the combined analysis consisted of 204 bp of the COI (positions 6491 to 6694) shared between both studies.

Differences in genetic diversity between adults and recruits were estimated by bootstrapping the adult data from Edmands et al. (1996). In this analysis, the adult sequences were truncated to include the 204 bp that were common to both studies. The entire adult dataset ($N = 145$) was then sampled with replacement to generate 10,000 replicate datasets from which a frequency distribution of haplotype diversities and haplotype numbers were generated. Because sample sizes varied among annual cohorts and individual recruitment events, this procedure was repeated for each of the experimental sample sizes. This removed any bias in the distributions of haplotype number and diversity that may be dependent upon sample size. One-tailed tests were conducted to evaluate whether there was a significant reduction in haplotype diversity or haplotype number in the recruits versus adult samples. Statistical significance was assessed by determining the proportion of values less than or equal to the observed

haplotype diversities and haplotype numbers from the corresponding 204 bp in the recruit dataset. A Bonferroni correction was applied where appropriate to account for multiple tests.

Simulation of sweepstakes variance in reproductive success

We conducted a series of simulations to determine heuristically the power of our sampling strategy to detect deviations from Poisson variance in reproductive success. We simulated populations of 1,000, 10,000 and 100,000 females with mtDNA haplotype frequencies equal to those in the truncated adult dataset from Edmands et al. (1996). The haplotype diversity of the simulated population was 0.804 and contained 39 haplotypes. We examined the effects of large deviations from Poisson variance in reproductive success by simulating a single episode of reproduction in each population. The probability of a female producing a given number of progeny, or failing completely, was gamma distributed, where

$$f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{x}{b}} \quad (1)$$

is the gamma distribution and Γ is the gamma function. Adjustment of parameters a and b allow for manipulation of two variables that contribute to absolute variance in reproductive success, (1) the proportion of the population that successfully breeds and (2) the number of progeny produced by successful breeders, respectively. These variables were allowed to assume values of different orders of magnitude, with a set to 10^{-1} to 10^{-4} and b set to 10^4 to 10^8. Under the parameters considered, this function approximates a negative logarithmic function that results in large variances in reproductive success where individuals have a high probability of reproductive failure and a low probability of achieving various levels of reproductive success.

For each simulated population experiencing gamma distributed progeny numbers we simulated a second population where the probability of producing a given number of progeny was Poisson distributed. The mean (and hence the variance) of the Poisson was set equal to the mean number of progeny produced by the paired population with gamma distributed progeny numbers. This facilitated a direct comparison of the parametric values of haplotype diversity and haplotype number in cohorts produced with Poisson or large deviations from Poisson (i.e., gamma) variance in reproductive success.

To estimate the power of our sampling strategy to detect reductions in genetic diversity in recruits, we randomly sampled 80 individuals (i.e., approximately the same number sequenced from each annual cohort) from each progeny pool produced by the paired populations. We repeated this 1,000 times to generate a pair of distributions of haplotype diversities and a pair of distributions of haplotype numbers. We interpreted the relative overlap of each pair of distributions as a measure of our ability to detect deviations from Poisson variance. The overlap was measured as the proportion of the distribution generated by resampling the progeny of females with gamma variance that falls below the lower 5% bound of the distribution generated by resampling the progeny of females with Poisson variance. Iterations where a large percentage of the former distribution was below the lower 5% bound of the latter were interpreted to mean that there was a high, yet unspecified, probability of our sampling strategy to detect large deviations from Poisson variance in progeny number (see simulation results below). Finally, we conducted 1,000 iterations for each of the

parameter sets that defined the gamma distribution, where each iteration consisted of paired populations with either Poisson or gamma variance in reproductive success.

RESULTS

Genetic Diversity

DNA sequencing of 358 bp from 283 *S. purpuratus* recruits yielded 59 polymorphic sites and 105 mtDNA haplotypes. The three most common haplotypes occurred at frequencies of 17.6%, 15.1%, and 12.4% in the entire dataset. No other haplotypes occurred at frequencies greater than 5%, and 73 haplotypes occurred only once. Nucleotide and haplotype diversity for the entire dataset were 0.00881 and 0.938, respectively. Divergence between pairs of haplotypes ranged from 0.28% to 3.1%.

Truncating the entire dataset to include the 204 bp common with the adult dataset (Edmands et al. 1996) yielded 50 haplotypes with nucleotide and haplotype diversities of 0.00985 and 0.771. The corresponding region in the adult dataset yielded 39 haplotypes with nucleotide and haplotype diversities of 0.01180 and 0.804. The combined dataset of 428 DNA sequences contained 73 haplotypes with nucleotide and haplotype diversities of 0.01069 and 0.782. Haplotype diversities and haplotype numbers of individual recruitment events and of annual cohorts (i.e., recruits collected in the same year) are presented in Tables 1 and 2 respectively.

Comparison of observed mtDNA diversity in recruits to expectations based upon bootstrap resampling of the DNA sequence data from adult sea urchins (Edmands et al. 1996) revealed no evidence for reduction in diversity of recently settled recruits relative to the adult population. No statistically significant reductions in either haplotype diversity or haplotype number were found in the recruit cohorts

(Table 2). Samples of from each recruitment event similarly revealed no significant reduction in haplotype diversities or haplotype numbers (Table 1) relative to adults after a Bonferroni correction for multiple tests.

Temporal Genetic Differentiation

Annual cohorts of recruits exhibited some evidence of genetic heterogeneity among years. Combined analysis of all cohorts revealed low levels of subdivision (ϕ_{ST} < 0.001, $P = 0.321$). However, pairwise analysis of cohorts indicated that recruits from 1996 were weakly differentiated from recruits from the 1999 cohort by AMOVA (ϕ_{ST} = 0.010, $P = 0.044$) and an Exact Test ($P = 0.040$). The 1998 cohort was also differentiated from the 1999 cohort ($\phi_{ST} = 0.009$) according to an AMOVA (P = 0.038) and an Exact Test ($P = 0.039$). Within site genetic heterogeneity was absent among years at Ocean Beach, SIO, Anacapa Island (maximum $\phi_{ST} = 0.006$). Statistical significance was found among some pairs of samples collected at Bodega Bay (maximum $\phi_{ST} = 0.047$, $P = 0.044$), but is probably best explained by sampling error due to small sample sizes rather than real genetic heterogeneity among recruits (Waples 1998).

Population Structure

Individual samples of recruits collected from different locations within the same season provided little evidence for heterogeneity within annual cohorts (ϕ_{ST} < 0.015). Pooling temporal samples collected from the same location similarly revealed

no evidence of genetic discontinuities in California. Analysis of hierarchical population structure with regional groupings based upon biogeographic regions including Bodega Bay and Fort Bragg (Oregonian Province), Anacapa Island, Paradise Cove, and Ellwood Pier (transition zone), and SIO and Ocean Beach (Californian Province) revealed no evidence of genetic subdivision among regions (F_{RT} = -0.003, P = 0.466). A comparison of northern California and southern California samples also suggested genetic homogeneity among regions (F_{RT} < 0.0001, P = 0.243). Other *a priori* regional groupings examined by Edmands et al. (1996) revealed no evidence of population structure. A combined analysis of the DNA sequences reported by Edmands et al. (1996) and the recruits reported here similarly revealed little evidence of genetic subdivision. In particular, an analysis of samples north and south of a possible genetic break in mtDNA between Laguna Beach and La Jolla (Edmands et al. 1996), revealed no evidence of population subdivision (F_{RT} = 0.001, P = 0.237).

Simulation results

The simulations indicate that the sampling strategy employed is suitable for detecting extreme variances in reproductive success. For example, one simulated population of 100,000 female urchins with gamma (parameter $a = 10^{-3}$, $b = 10^{7}$) variance resulted in 1,582 successful females, with mean and variance of progeny per female in the population of 9,932 and 1.42×10^{11}, respectively (Fig. 1, panel A). In contrast, every female in the paired population with Poisson variance was successful and produced between 9,400 and 10,400 progeny (Fig. 1, panel B). The parametric

haplotype diversities of the two progeny pools were 0.710 and 0.791, respectively. Accordingly, the resampled distribution of haplotype diversity from the progeny of females with large deviations from Poisson variance was reduced relative to the progeny of females that experienced Poisson variance, with 68.4% of the former distribution falling below the lower 5% bound of the latter (Fig. 2). This illustrates the capacity of our sampling strategy to detect large deviations from Poisson variance since the resampled distribution reflects the reduced haplotype diversity in the progeny pool of the female population with gamma variance in progeny number. This pattern was repeated throughout our simulations of extreme variances in reproductive success, but disappeared when less extreme variances were simulated.

Table 3 provides results from 1,000 iterations of each of four different pairs of populations with either gamma or Poisson distributed progeny numbers. These results illustrate (1) the effects that different gamma functions have on the number of successful females, the variance in reproductive success and the genetic diversity of the progeny and (2) how different variances influence the power of the sampling strategy to detect sweepstakes reproductive success. The sweepstakes events summarized in Table 3 consisted of between 142 and 69,489 successful females out of a population of 100,000 with absolute variances in reproductive success ranging between 9.29×10^{8} and 6.22×10^{12}. Genetic variation was reduced in the progeny of females with the most extreme gamma variances in reproductive success relative to the diversity of progeny of females with Poisson variance. However, when less extreme variances were simulated, all 39 haplotypes were always present in the

progeny pool and differences in haplotype diversities between progeny pools generated with different variances in progeny number were small or absent.

Resampled distributions of genetic diversity of the progeny pools generally reflected this pattern. When the number of successful females was low and variance in reproductive success was high in the population with gamma variance, the resampled distribution from the progeny of this population was frequently reduced relative to the resampled distribution from the progeny of the paired population with Poisson variance (e.g., Fig. 2). The differences between resampled distributions were small or absent, however, when the population with gamma distributed progeny number had less extreme variances and when the number of successful females was on the order of 1,000 or more. Although this appears to suggest that our power to detect sweepstakes was reduced when less extreme variances in reproductive success were simulated, it should be noted that the less extreme variances yielded only small differences in the parametric genetic diversities of the progeny of the paired populations (Table 3). The negligible reductions in genetic diversity when gamma variances were small is probably a consequence of simulating an adult population with low haplotype diversity in the short mtDNA fragment of 204 bp shared between this study and that of Edmands et al. (1996). Hence, it is clear that our sampling strategy is capable of detecting extreme sweepstakes events (Table 3), and it is likely that our sampling strategy is able to detect less extreme deviations from Poisson variance than demonstrated here.

DISCUSSION

Many marine invertebrate species have extremely high fecundities with individual females frequently producing between 10^6 and10^7 eggs. Not surprisingly, these species typically suffer high juvenile mortality as a consequence of unpredictable environmental variation. If the unpredictability associated with reproduction in the sea results in a small fraction of the adult population contributing to each annual cohort, then reproduction by many marine species may be viewed as a sweepstakes, in which chance events determine which adults are successful each spawning season (Hedgecock 1994). This hypothesized reduction in the effective number of breeders may have important consequences for the evolution of marine populations since genetic effective population sizes may be orders of magnitude lower than census population sizes. Although this hypothesis presents an attractive explanation for the lower than expected genetic diversity observed in marine populations (Avise et al. 1988; Hedgecock 1994; Turner et al. 1999; Avise 2000), there is little empirical evidence that individual females experience the large variances in reproductive success proposed by the sweepstakes hypothesis.

Genetic Signatures of Variance in Reproductive Success

Sweepstakes reproductive success is expected to leave a diagnostic signature on the genetic composition of marine larvae (Hedgecock 1994). First, if variance in progeny number is not Poisson distributed, then cohorts are predicted to exhibit reduced levels of genetic variation relative to the parental population. Our simulation

results suggest that this prediction is usually met, since large deviations from Poisson variance in reproductive success frequently generated progeny pools with haplotype diversities less than that of the population of reproducing females. The simulations also suggest that progeny pools generated by populations with large variance may have very different haplotype diversities than progeny pools generated by populations with Poisson variance.

In this study, we have focused on mtDNA haplotypes because all progeny of a given female will share the same haplotype. Our samples of natural populations of sea urchin recruits arriving at discrete habitat patches (settlement brushes) over known time intervals of 7-14 days), consistently have large numbers of haplotypes. These results clearly demonstrate that cohorts of recruits are rarely, if ever, composed of half-sibs, rather, they appear to represent the contributions of a large number of breeding females. Estimates of genetic diversity in each cohort are consistently high and are not reduced relative to estimates of diversity from the adult population. While the sampling strategy employed does not allow us to conclusively reject the sweepstakes recruitment hypothesis, it does suggest that extreme sweepstake events are probably not a common feature of recruitment in purple urchins.

A second prediction of the sweepstakes recruitment hypothesis is that annual cohorts of recruits should be differentiated. This prediction is based on the hypothesis that chance matching of reproductive activity with oceanographic conditions suitable for reproductive success will result in genetic drift, since different sets of adults are predicted to contribute to each annual cohort. For example, Li and Hedgecock (1998)

attributed genetic differentiation over time in larval cohorts of *Crassostrea gigas* to small groups of females from a single semi-isolated population that spawned at different times throughout the year. This result is notable because shifts in gene frequencies among samples are probably not attributable to gene flow since this population was largely isolated from other coastal populations (see also Ruzzante et al. 1996). Genetic differentiation among cohorts reported here, and elsewhere, indicate that marine larvae or recently settled recruits may frequently be temporally differentiated (Johnson and Black 1982; Gosling and Wilkins 1985; Moberg and Burton 2000) suggesting that different groups of spawning adults may contribute to different cohorts. As discussed above, this may result from the sweepstakes process within a single geographic population (e.g., Li and Hedgecock 1998), but it may also result from changes in patterns of dispersal between genetically differentiated populations (e.g., Kordos and Burton 1993). Since the two processes (drift within populations and gene flow between populations) occur simultaneously in most marine species with planktonic larval dispersal, distinguishing between allelic frequency changes due to genetic drift versus gene flow will often be impossible.

Previous empirical work addressing sweepstakes variance in marine populations has concentrated on examination of genetic subdivision among cohorts described above, kinship relationships among larvae (Avise and Shapiro 1986; Herbinger et al. 1997), or assessment of genetic drift in natural populations (Hedgecock 1994). Although an intuitively attractive hypothesis, empirical support for large variances in reproductive success is limited at best. Furthermore, when results

consistent with the sweepstakes recruitment hypothesis are found in natural populations, it is not clear whether variances in reproductive success, on the order experienced by the population, are large enough to substantially decrease N_e/N. Without estimates of the variance in progeny production it is difficult to assess how important variance in reproductive success has been relative to other demographic factors (e.g., historical fluctuations in population size) in depressing N_e/N of many marine species. Finally, although rarely discussed by population geneticists, low N_e/N ratios of marine species may simply be a consequence of life history schedule (Felsenstein 1971; Gaggiotti and Vetter 1999; O. Gaggiotti, pers. comm. 2002). For example, life history schedules of Pacific sardines and northern anchovy result in differences in total reproductive value and generation overlap that significantly impact the N_e of the two species (Gaggiotti and Vetter 1999). Further comparative studies assessing the effect of life history parameters on N_e may be useful in determining the cause(s) of reduced N_e/N in marine species.

Life History and the Evolution of Longevity

Life history and demographic features of marine free-spawning populations are hypothesized to have evolved, in part, as a result of the unpredictability of reproductive success (Murphy 1968; Ebert 1975). When pre-reproductive stages (e.g., planktonic larvae) experience high mortality rates, selection favors allocating proportionately more resources to growth and maintenance at the expense of reproduction. This switch in resource allocation increases the probability of

reproductive success by distributing reproductive efforts over many years (i.e., bet hedging). Hence, environments that predictably facilitate high pre-reproductive survival and low variance in reproductive success favor short life, while unpredictable environments with low pre-reproductive survival and high variance in reproductive success favor long life.

Urchins in the genus *Strongylocentrotus* are typically long-lived with lifespans in purple urchins exceeding ten years (Ebert 1967). If long life evolved in response to bet hedging, we expect *S. purpuratus* to experience large variance in progeny production. Although we found little evidence for "sweepstakes" variance in reproductive success (i.e. inter-familial variance *sensu* Hedgecock 1994), there is substantial evidence that recruitment is temporally variable throughout much of the range of *S. purpuratus* (Ebert 1983; Ebert et al. 1994). Annual settlement of larvae is consistently strong in southern California, but occurs only sporadically at locations north of Pt. Conception, California, with strong recruitment pulses being separated by as many as 20 years in Oregon (Ebert 1982). This temporal component of variance may strongly favor long-lived individuals of *S. purpuratus* as previously suggested for urchins in general (Ebert 1975). If large inter-familial, or "sweepstakes", variances are characteristic of some marine populations then they presumably would have a similar effect as temporal variance on the evolution of longevity since more opportunities for mating would provide long-lived individuals with a higher probability of reproductive success.

Conclusions

The proposal that sweepstakes variance in reproductive success may be a common feature of marine populations with substantial effects on N_e/N could provide a link between recruitment dynamics and the evolution of marine species. However, results from *Strongylocentrotus purpuratus* failed to provide convincing support for the sweepstakes reproductive success hypothesis within the detection limits of our sampling scheme. In the absence of substantial evidence for sweepstakes, here or elsewhere, the general significance of this phenomenon for marine populations remains in question.

ACKNOWLEDGMENTS

We thank C. English for helpful comments on the simulations and for providing much needed cpu time. We are grateful to H. S. Rapoport for writing a bootstrapping algorithm to test for genetic diversity differences between adults and recruits. For laboratory assistance, we thank T. Ballard, A. Chinen, R. E. Tornek, and H. Huynh. We thank J. Wolf for laboratory assistance and curation of our sea urchin settlement collection and L. Botsford and C. Lundquist for providing samples from Bodega Bay. We also thank O. E. Gaggiotti, C. Willett, M. A. Paine, E. Goetze, T. A. Ebert, J. P. Wares, J. D. Dixon, and an anonymous reviewer for contributing to earlier versions of this manuscript. This work was funded by a grant to R. S. B. 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-178 through the California Sea Grant College Program. The views expressed herein are those of the authors 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. Appendix A was published in the journal Evolution in 2002. Permission to include this manuscript in the dissertation of Jonathan M. Flowers was obtained from Annette Keeler, managing editor of Evolution.

LITERATURE CITED

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard Univ. Press, Cambridge, Massachusetts.

Avise, J. C., R. M. Ball, and Arnold. 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol. Biol. Evol. 5: 331-344.

Avise, J. C. and D. Y. Shapiro. 1986. Evaluating kinship of newly settled juveniles within social groups of the coral reef fish *Anthias squamipinnis*. Evolution. 40:1051-1059.

Crow, J. F. and M. Kimura. 1970. An introduction to population genetics theory. Burgess Publishing Co., Minneapolis, Minnesota.

Cushing, D. H. 1990. Plankton production and year class strength in fish populations: an update of the match / mismatch hypothesis. Adv. Mar. Biol. 26:249-293.

Ebert, T. A. 1967. Negative growth and longevity in the purple sea urchin *Strongylocentrotus purpuratus* (Stimpson). Science. 157:557-558.

Ebert, T. A. 1975. Growth and mortality of post-larval echinoids. Am. Zool. 15: 755-775.

Ebert, T. A. 1982. Longevity, life history, and relative body wall size in sea urchins. Ecol. Monogr. 52: 353-394.

Ebert, T. A. 1983. Recruitment in echinoderms. Pp. 169-203 *in*, M. Jangoux and J. M. Lawrence, eds. Echinoderm studies, A. A. Balkema Publisher, Netherlands.

Ebert, T.A., J. D. Dixon, S. C. Schoeter, P. E. Kalvass, N. T. Richmond, W. A. Bradbury, and D. A. Woodby. 1999. Growth and mortality of red sea urchins across a latitudinal gradient. Mar. Ecol. Prog. Ser. 190:189-209.

Ebert, T.A., S. C. Schroeter, J.D. Dixon, and P. Kalvass. 1994. Settlement patterns of red and purple sea urchins (*Strongylocentrotus franciscanus* and *S. purpuratus*) in California, USA. Mar. Ecol. Prog. Ser. 111:41-52.

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*. Mar. Biol.126:443-450.

Felsenstein, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. Genetics. 68:581-597.

Gaggiotti, O. E. and R. D. Vetter. 1999. Effect of life history strategy, environmental variability, and overexploitation on the genetic diversity of pelagic fish populations. In: Can. J. Fish. and Aquat. Sci. 56:1376-1388.

Gosling, E. M. and N. P. Wilkins. 1985. Genetics of settling cohorts of *Mytilus edulis*: preliminary observations. Aquaculture. 44:115-123.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? Pp. 122-134 *in* A. Beaumont, ed. Genetics and evolution of aquatic organisms. Chapman & Hall, London.

Herbinger, C. M., R. M. Doyle, C. T. Taggart, S. E. Lochmann, A. L. Brooker, J. M. Wright, and D. Cook. 1997. Family relationship and effective population size in a natural cohort of Atlantic cod (*Gadus morhua*) larvae. Can. J. Fish. Aquat. Sci. 54(suppl 1): 11-18.

Hoelzel, A. R. and A. Green. 1992. Analysis of population-level variation by sequencing PCR-amplified DNA. Pp. 159-187 *in* A. R. Hoelzel ed. Practical Approach Series: Molecular Genetic Analysis of Populations. Oxford University Press: New York, New York.

Jacobs, H. T., D. J. Elliot, V. B. Math, and A. Farquharson. 1988. Nucloetide sequence and gene organization of sea urchin mitochondrial DNA. J. Molec. Biol. 202:185-217.

Johnson, M. S., and R. Black. 1982. Chaotic genetic patchiness in an inter-tidal limpet, *Siphonaria* sp. Mar. Biol. 70:157-164.

Kato, S. and S.C. Schroeter. 1985. Biology of the red sea urchin, *Strongylocentrotus franciscanus*, and its fishery in California. Mar. Fish. Rev. 47:1-20.

Kordos, L. M. and R. S. Burton. 1993. Genetic differentiation of Texas Gulf-Coast populations of the blue-crab *Callinectes Sapidus*. Mar. Biol. 117:227-233.

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 variance in reproductive success. Can. J. Fish. Aquat. Sci. 55: 1025-1033.

Miller, B. A. and R. B. Emlet. 1997. Influence of nearshore hydrodynamics on larval

abundance and settlement of sea urchins *Strongylocentrotus franciscanus* and *S. purpuratus* in the Oregon upwelling zone. Mar. Ecol. Prog. Ser. 148:83-94.

Moberg P. E., and R. S. Burton. 2000. Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. Mar. Biol. 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. Fish. Oceanogr. 9:83-98.

Murphy, G. I. 1968. Pattern in life history and the environment. Am Nat. 102:391-403.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia Univ. Press, New York.

Raymond, M., and R. F. Rousset. 1995. An exact test for population differentiation. Evolution. 49:1280-1283.

Rozas, J. and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics. 15: 174-175.

Ruzzante, D.E., C. Taggart, and D. Cook. 1996. Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. Can J. Fish. Aquat. Sci. 53:2695-2705.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Version 2.0. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Strathmann, R. R. 1978. The length of pelagic period in echinoderms with feeding larvae from the northeastern Pacific. J. Exp. Mar. Biol. Ecol. 34:23-27.

Turner, T. F., L. R. Richardson, and J. R. Gold. 1999. Temporal genetic variation of mitochondrial DNA and the female effective population size of red drum (*Sciaenops ocellatus*) in the northern Gulf of Mexico. Mol. Ecol. 8:1223-1229.

Waples, R. S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. J. Heredity. 89:438-450.

Wing, S. R., J. L. Largier, L. W. Botsford, and J. F. Quinn. 1995. Settlement and transport of benthic invertebrates in an intermittent upwelling region. Limnol. Oceanogr. 40:197-213.

Yakovlev, A. N. 1987. Fecundity of the sea urchin *Strongylocentrotus intermedius* and methods for its estimation. Biologiya Morya (Vladivostok). 5:46-52.

Corresponding Editor: J. Merila

Table 1. Summary of *Strongylocentrotus purpuratus* recruit collections and haplotype diversity (h) and haplotype numbers from each sample based upon 358 bp of mtDNA sequence. Haplotype diversities and their standard deviations were calculated according to Nei (1987) with the DnaSP software package (Rozas and Rozas 1999).

Sample Location	Latitude (N)	Longitude (W)	Collection Period	N	$h \pm$ SD	Haplotype Number
Ocean Beach	32° 44.419'	117° 15.299'	July 3-17, 1996	18	0.8954 ± 0.0545	11
			April 7-14, 1999	10	0.9333 ± 0.0620	7
			April 12-19, 2000	13	0.9615 ± 0.0412	10
SIO	32° 52.491'	117° 15.220'	May 10-17, 1994	18	0.9608 ± 0.0301	13
			July 2-9, 1996	20	0.9421 ± 0.0295	12
			April 6-13, 1999	19	0.9766 ± 0.0267	16
			April 11-18, 2000	20	0.8895 ± 0.0494	11
Anacapa Island	34° 00.976'	119° 21.663'	July 2-16, 1996	20	0.9789 ± 0.0245	17
			April 6-20, 1999	20	0.9211 ± 0.0387	12
Paradise Cove	34° 01.159'	118° 47.192'	March 29 - April 29, 1999	21	0.9000 ± 0.0508	13
Ellwood Pier	34° 26.118'	119° 56.361'	July 9-23, 1996	19	0.9298 ± 0.0466	13
Bodega Bay	39° 19.208'	123° 05.055'	May 6-13, 1998	17	0.9779 ± 0.0267	14
			May 13-20, 1998	13	0.9872 ± 0.0354	12
			May 20 - June 2, 1998	17	0.8824 ± 0.0595	10
			June 23 – July 9, 1998	3	-	-
			July 9-17, 1998	16	0.8670 ± 0.0063	10
Fort Bragg	39° 20.764'	123° 49.396'	June 19-26, 1998	19	0.9708 ± 0.0273	15

Table 2. Genetic diversity of annual recruit cohorts of *Strongylocentrotus purpuratus* in California based on 358 bp of mtDNA sequence. The cohorts are pooled across multiple locations listed in Table 1. Haplotype diversities (h) and their standard deviations were calculated according to Nei (1987) with the DnaSP software package (Rozas and Rozas 1999).

Cohort*	N	$h \pm$ SD	Haplotype number
1996	77	0.943 ± 0.016	42
1998	85	0.949 ± 0.014	45
1999	70	0.919 ± 0.020	31
2000	33	0.924 ± 0.028	18

Table 3. Results from simulating reproduction in populations of 100,000 sea urchin females experiencing either Poisson or large deviations from Poisson variance in the number of progeny per female. One thousand episodes of reproduction were simulated for a population that experienced gamma variance and 1,000 episodes were simulated for a second population with Poisson variance in reproductive success (see methods for details). Simulations were run in pairs with the mean and variance of the Poisson defined by the mean number of progeny produced by the paired population experiencing gamma variance in progeny number. The number of successful females, the progeny per successful female, and the variance in reproductive success are mean values averaged over the 1,000 iterations from the populations with gamma variance in progeny number. Measures of genetic diversity (i.e., haplotype diversity and number) are average parameter values calculated from each progeny pool generated with females with either gamma or Poisson variance. To evaluate the power of the sampling strategy, a frequency distribution of haplotype diversities and numbers was generated for each of the 1,000 iterations by randomly sampling with replacement eighty individuals from both progeny pools. The proportion of the resampled datasets produced by sampling the gamma pool that are less than the lower 5% bound of the distribution generated by sampling the Poisson pool is reported for both measures of genetic diversity.

Successful females (mean ± SD)	Progeny per successful female (mean ± SD)	Variance in reproductive success (mean ± SD)	Haplotype diversity			Haplotype number		
			Gamma (mean ± SD)	Poisson (mean ± SD)	Proportion < 5% (mean ± SD)	Gamma (mean ± SD)	Poisson (mean ± SD)	Proportion < 5% ± SD
185 ± 13.72	$5.34 \times 10^6 \pm 1.70 \times 10^6$	$9.87 \times 10^{11} \pm 7.75 \times 10^{11*}$	0.725 ± 0.095	0.798 ± 0.001	0.433 ± 0.406	31.13 ± 2.39	39 ± 0	0.998 ± 0.015
1,608 ± 39.6	$6.2 \times 10^5 \pm 6.1 \times 10^4$	$1.00 \times 10^{11} \pm 2.49 \times 10^{10\dagger}$	0.791 ± 0.032	0.798 ± 0.001	0.116 ± 0.143	39 ± 0.032	39 ± 0	0.299 ± 0.188
13,009 ± 109	$7.69 \times 10^4 \pm 2.31 \times 10^3$	$9.99 \times 10^9 \pm 7.68 \times 10^{8\S}$	0.797 ± 0.010	0.798 ± 0.001	0.058 ± 0.034	39 ± 0	39 ± 0	0.045 + 0.023
68,989 ± 144	$1.45 \times 10^4 \pm 1.45 \times 10^2$	$1.00 \times 10^9 \pm 2.51 \times 10^{7\ddagger}$	0.798 ± 0.004	0.798 ± 0.001	0.049 ± 0.013	39 ± 0	39 ± 0	0.031 ± 0.009

*Parameters of the gamma distribution, $a = 0.0001$, $b = 10^8$
$^{\dagger}a = 0.001$, $b = 10^7$
$^{\S}a = 0.01$, $b = 10^6$
$^{\ddagger}a = 0.1$, $b = 10^5$

Fig. 1. Distribution of progeny numbers produced by two simulated populations of 100,000 sea urchin females with gamma ("sweepstakes") or Poisson distributed variances in reproductive success (see methods). (A) Gamma variance resulted in 1,582 successful females, with mean and variance of progeny per female in the population of 9,932 and 1.42×10^{11}, respectively. (B) Poisson variance, with the mean and variance of the distribution set to 9,932, resulted in each female producing between 9,400 and 10,400 progeny. In developing this histogram, frequencies of one (represented by the shortest bars) were adjusted to approximately 1.01 so they could be visualized on a log scale.

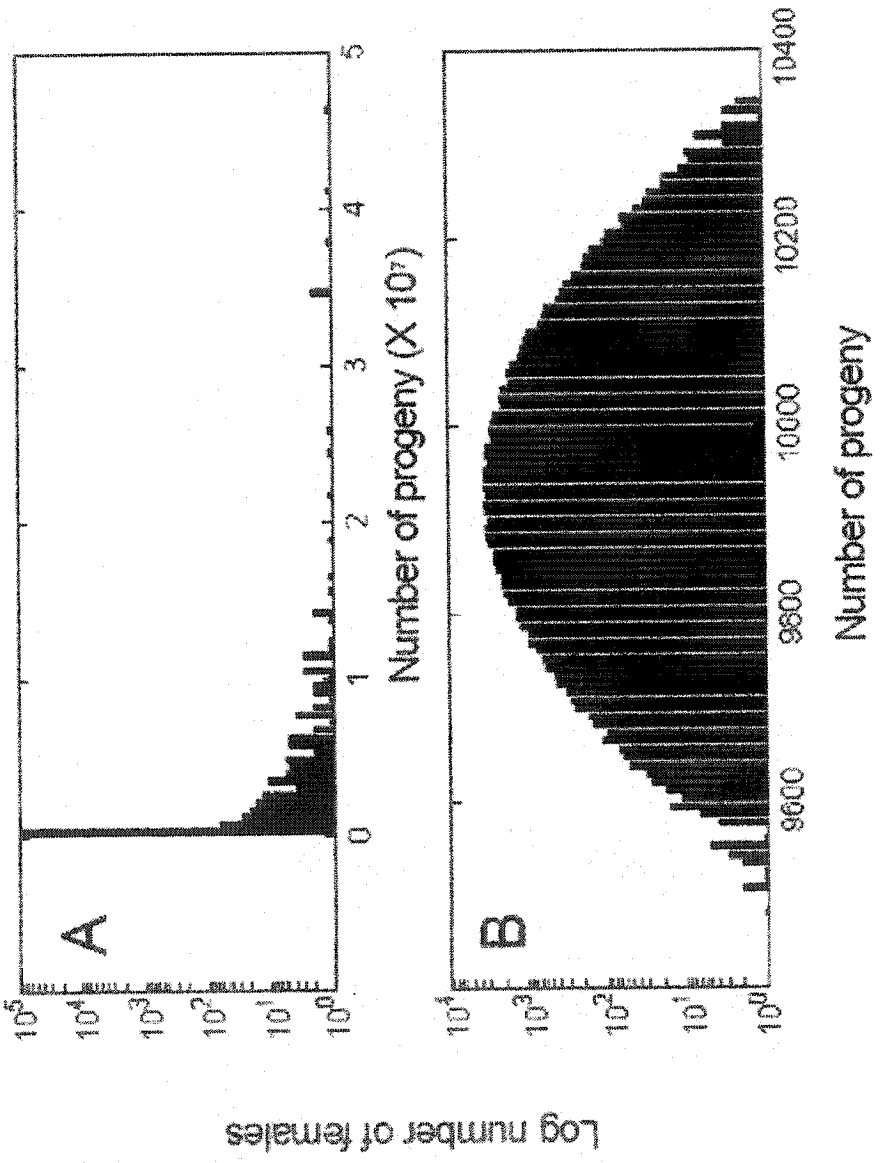
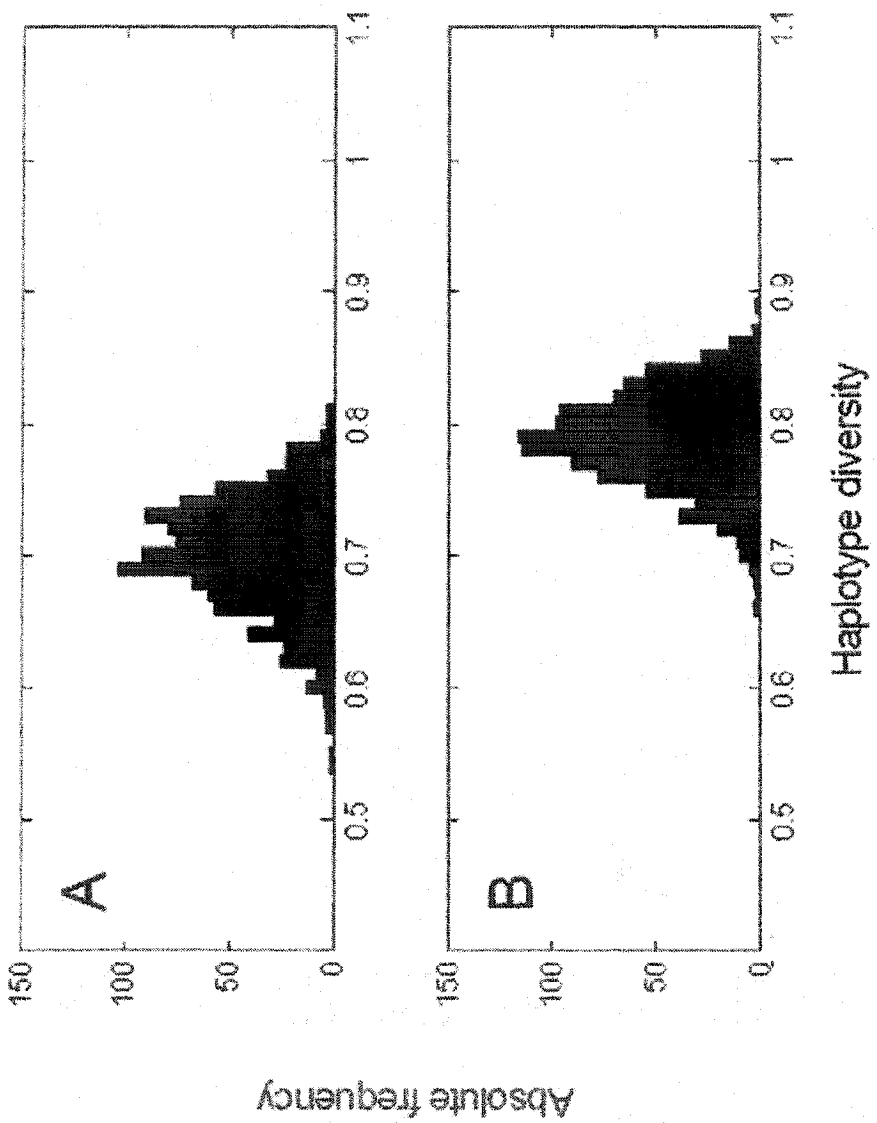


Fig. 2. Resampled distributions of haplotype diversities of progeny generated by simulating a single episode of reproduction by a population of 100,000 sea urchin females with either gamma (A) or Poisson (B) variance in reproductive success (see methods). Each distribution was generated by randomly sampling 80 individuals from the progeny pool and determining the haplotype diversity of each of the 1,000 resampled datasets. Parameter values of haplotype diversities of the two progeny pools were 0.710 for the progeny generated with gamma distributed progeny numbers and 0.791 for the progeny generated with Poisson distributed progeny numbers.



The text of Appendix A, in full, is a reprint of the material as it appears in Flowers, J. M., S. C. Schroeter, and R. S. Burton. 2002. The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. Evolution. 56:1445-1453. The dissertation author was the primary author. Stephen C. Schroeter provided animals for the study. Ronald S. Burton funded the research in its entirety and supervised the research. Permission to include this manuscript in the dissertation of Jonathan M. Flowers was obtained from Annette Keeler, managing editor of Evolution.