RAPID EVOLUTION OF A GAMETE-RECOGNITION PROTEIN IN A HYBRID *MYTILUS* POPULATION

by

Stevan A. Springer

BSc, University of Northern British Columbia, 1998

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

In the Department of Biological Sciences

SIMON FRASER UNIVERSITY

August 2003

APPROVAL

Name: Stevan Anthony Springer

Degree: Master of Science

Title of Thesis:

Rapid evolution of a gamete-recognition protein in a hybrid *Mytilus* population.

Examining Committee:

Chair: Dr. L.M. Dill

Dr. B.J. Crespi, Professor, Senior Supervisor
Department of Biological Sciences, S.F.U.

Dr. F. Breden, Associate Professor
Department of Biological Sciences, S.F.U.

Dr. M.J. Smith, Professor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. E. Elle, Assistant Professor
Department of Biological Sciences, S.F.U.
Public Examiner

AUGUST 14, 2003
Date Approved

PARTIAL COPYRIGHT LICENCE

Title of Thesis:

Rapid Evolution of a Gamete-Recognition Protein in a Hybrid *Mytilus* Population

Author:

(Signature)

Stevan A. Springer

(Name)

August 14, 2003

(Date)

ABSTRACT

Behaviours are among the most variable and complex of phenotypes, but this does not imply that they have a complex genetic basis. Many interactions between biological units are mediated by direct physical contact between gene-products acting on behalf of each individual. These simple mechanisms can enact behaviours which are as subtle, strategic, and interesting as any, offering unique opportunities for the study of phenotypic evolution. These behaviours evolve in discrete units of quantifiable phenotypic effect, amino-acid substitutions, and the adaptive importance of specific changes can be identified chemically or computationally. Current evidence indicates that these traits are often the target of strong diversifying selection and therefore are important components of organismal fitness. The functional importance of the behavioural changes induced by these loci can be assessed in natural conditions where their consequences for organisms are real.

Sperm-egg fusion is a universal feature of sex. In every fertilization, contact between sperm and egg is mediated by gene-products expressed on the surface of each gamete. These "gamete-recognition" proteins can cause reproductive incompatibility of individuals or reproductive isolation between species. Since gamete-recognition proteins are expressed, and active, in all sexual organisms, studying their evolution can yield general insight into sexual selection and speciation. I examine the molecular evolution of a gamete-recognition protein in *Mytilus* blue mussels, broadcast-spawning marine invertebrates that hybridize in nature. Recent, human-mediated, secondary contact between two *Mytilus* species has resulted in the rapid adaptive evolution of a sperm protein that dissolves the outer carbohydrate layer of the egg. The adaptive amino-acid substitutions observed in sympatric *Mytilus* populations occur in functionally important parts of this protein, suggesting that they may modify sperm-egg compatibility.

This study is the first to genetically-characterize the adaptive evolution of a reproductive protein in a species that hybridizes in nature, and suggests that the adaptive evolution typical of gamete-recognition proteins may be of general importance in speciation.

ACKNOWLEDGMENTS

I thank Bernard J. Crespi for guiding me through this fledgling attempt at science, my completion of this work is a direct result of his patience, steely nerves and tactful supervision. Bernie provided his time and expertise selflessly, even after shaving his beard. Most importantly, he recruited and taught the excellent students who have been my colleagues.

I am very proud of my membership in BERG and the FAB-Lab, with whom I've shared some of my most memorable professional and un-professional activities. Acknowledgements are like wedding invitations, and so I'm forced to pass on this opportunity to embarrass you each for posterity, but rest assured, I will look back on these years fondly because of you.

I am grateful to my supervisory committee for their direction. Michael J. Smith provided well-timed advice that helped me keep this project and the aims of graduate school in perspective. Felix Breden instructed me in advanced photocopying technique, answered a million questions, and reminded me why I enjoy being a student. Elizabeth Elle shared her helpful pragmatic advice on surviving graduate school and greatly improved this thesis as my external examiner. Though he held no official authority over me, Arne Mooers deserves singular mention for allowing me access to his broad intellect and genuine excitement for all things interesting. Special thanks to Daniel Heath for giving me my start and continuing to foster my academic development.

I've enjoyed Trent Hoover's unique perspectives on science and pie, and Shelley Hoover's sensible counterpoints, one could not ask for better friends. Jabus Tyerman challenged me with his impressive knowledge of biology and shamed me into finishing with one astute figure. I could not have finished without Jill Sutton, who suffered many nights of reproductive isolation to see this thesis completed, my brother Simon Springer, who taught me the power of persistence, and my parents Anne and Del Springer whom I blame for everything. Thanks guys.

TABLE OF CONTENTS

Approval ii

Abstract iii

Acknowledgments iv

Table of Contents v

List of Figures and Tables vi

Preface vii

Chapter One: Brainless Behaviour and Adaptive Evolution 1

- Every Man for Himself, and the Devil Take the Hindmost 1
- The Goals of Evolutionary Research 2
- The Genetic Bases of Phenotypes 3
- Examples of Brainless Behaviour 6
- On the Origin of Species by Means of Natural Selection 8
- Post-zygotic Isolation 8
- Pre-zygotic Isolation 9
- Conclusion 10

Chapter Two: Rapid Evolution of a Gamete-Recognition Protein in a Hybrid *Mytilus* Population 12

- Abstract 12
- Methods 23
  - Tissue Collection, PCR Amplification and Sequencing 23
  - Data Analysis 24

Conclusion 27

Appendix: Nucleotide Variation in *Mytilus* Lysin-M7 Alleles 30

References 33

LIST OF FIGURES AND TABLES

Figure 1 - Global distribution of hybridizing *Mytilus* species. 14

Figure 2 - Phylogenetic relationships among Lysin-M7 alleles. 16

Figure 3 - Spatial locations of amino acid substitutions in the carbohydrate recognition domain of *Mytilus* Lysin-M7. 18

Table 1 - Lysin-M7 variation in *Mytilus* 19

Table 2 - Likelihood ratio tests of selection 25

PREFACE

"I am not sure what should be done now, since their work is, in essence, my thesis."

Stanley L. Miller to his supervisor Harold Urey after finding that his seminal work on the origins of life had been scooped (1953).

This thesis is composed of two chapters. The first chapter is an introduction to the molecules that mediate direct interactions between biological units and an exposition of their implications for the study of speciation and adaptation. Few programmes in evolutionary biology hold as much immediate and sweeping promise as study of the simple traits that arbitrate interactions between individuals. The second chapter is my attempt to justify this claim empirically, a case study of speciation in *Mytilus* blue mussels. I examine the molecular evolution of a sperm-surface protein that binds egg-surface carbohydrates allowing fertilization. This work is the first example of secondary contact and hybridization between two ecologically-distinct species resulting in positively selected divergence of a mate recognition trait whose genetic basis is known. Few studies provide such direct evidence that selection drives the divergence of reproductive traits prior to the evolution of reproductive isolation, and I believe this work is therefore of some general importance to evolutionary biology. I conclude with my thoughts on the future challenges and prospects of gamete-recognition research; forgive the indulgence, but this is a thesis after all.

CHAPTER ONE: BRAINLESS BEHAVIOUR AND ADAPTIVE EVOLUTION

Every Man for Himself, and the Devil Take the Hindmost

Behaviours are among the most multifaceted and variable of all phenotypic traits, and studies of behaviour often assume a complex genetic basis (Boake et al. 2002). In fact, even the most intricate behaviours have analogs at the molecular level; direct molecular interactions can produce behaviours with all the subtlety and strategy of their most complex counterparts without the need to invoke composite genetic mechanisms. The strategies employed by teams of individuals provide a useful analogy. Oval track bicycle racing is celebrated by its enthusiasts as a game of skill in which strategy wins more races than swiftness. Groups of riders cut through wind more efficiently than singletons, and racers must therefore negotiate membership in a group, or be left behind. One especially tactical contest, the Miss-and-Out race, begins with a group of riders circling the track; motivated by the removal of the last-place rider every lap. Single riders cannot distance themselves from the pack, and so alliances form and break as groups attempt to lead the charge and evade the chopping block. Eventually, the race is whittled to only two contestants, former friends and now bitter enemies, who must sprint the final lap to victory or defeat.

Possum and silverfish sperm face a similar dilemma. In these taxa sperm pairs swim more efficiently than loners (Moore and Moore 2002), but only one sperm can successfully fertilize an egg. Sperm that do not find a partner find themselves out of the running. Groups of sperm compete with other groups in a race to the egg, but individual sperm must decide when to abandon their partners for the final sprint to the fertilization line. Getting to the winner's circle requires that sperm make a number of "choices", settling on a partner, deciding when to defect and how best to stab the estranged ally in the back. What is most remarkable about these strategies, lauded for their complexity and subtlety by cycling devotees, is that they are employed by sperm without the benefit of a brain. Most interactions between biological entities involve complex negotiation, but very

few of these debates are under neural control. Indeed, in much of biological communication it is direct molecular interactions that do all of the talking. The simple genetic basis and direct relationship between genotype and phenotype that characterizes this class of behaviours make them attractive model systems for studies of phenotypic evolution and adaptation.

The Goals of Evolutionary Research

Evolutionary biologists seek the ultimate causes of phenotypic and genetic change. The primary goal of evolutionary research is then to understand the processes that act within populations and result in phenotypic evolution. Evolutionary processes come in two varieties, those that provide only mechanistic explanations for how a trait came to be the way it is, genetic drift for example, and those that additionally provide causal explanations, primarily natural selection. Evolutionary biologists also seek the means of phenotype production; knowledge of the relationship between genotype and phenotype is essential if we wish to understand the spectrum of variation available to evolutionary processes. A great deal is known about each research objective in isolation, but few studies integrate evolutionary processes acting at the genetic level with precisely determined means of phenotypic change and very few studies have identified contemporary molecular polymorphisms underlying phenotypic variation in quantitative traits with clear adaptive significance (Hollocher and Templeton 1994; Nachman et al. 2003).

The difficulty in uniting these research areas arises from the technical demands of each question; systems that are ideal for one type of study are often poorly suited to the other. Here I describe a class of phenotypes, traits that mediate interactions between biological units without the input of neural processes (termed brainless behaviours), which are tractable to detailed genotypic and phenotypic analyses of evolution using current technology. The proteins that directly mediate biological communication are attractive evolutionary research systems for two reasons. First, the genotype-phenotype relationship is simple relative to quantitative traits; phenotypic changes occur in discrete

units, amino acid substitutions, whose adaptive importance can often be biochemically assessed, or inferred from protein structure or mutagenesis. Second, the importance of current and historical evolutionary mechanisms can be estimated using appropriate models of molecular evolution. The prevalence of positive selection at these loci suggests that they often have large effects on organismal fitness. I begin by explaining each of these advantages in terms of their importance for studies of adaptation, and then provide examples of biologically important processes that might involve direct interaction between units, concluding with prospects for the study of speciation. My aim is not to describe the genetics of behaviour in general (Boake et al. 2002), rather I will focus on interactions between biological units that are mediated directly by gene-product interactions.

The Genetic Bases of Phenotypes

Finding the genetic basis of a phenotypic trait is not a trivial problem. Phenotypes are often a product of contributions from the environment and numerous genetic loci. Determining a trait's genetic basis involves first locating the underlying loci, and second evaluating their phenotypic effects. Many studies have mapped loci contributing to quantitative traits (Hawthorne and Via 2001), and theoretical frameworks for measuring selection on quantitative traits exist (Orr 1998; Turelli et al. 1988). These studies are important, but difficult and time consuming, and at best they yield indirect information on the connection between genotype and phenotype. Behavioural interactions involving direct contact between gene products minimally involve single loci, or pairs of loci, and because the phenotype-genotype connection is less complex than for quantitative traits, the genes that control brainless behaviours are often easier to genetically and functionally describe.

Traits with a simple, single-locus, genetic basis are more tractable to analyses of phenotypic evolution than quantitative traits. The evolution of single-locus traits occurs by amino-acid substitution which proceeds in discrete units of quantifiable phenotypic effect. Site-directed mutagenesis techniques can introduce specific substitutions into

these loci to recreate inferred ancestral proteins (Chang and Donoghue 2000), or examine the set of amino-acid residues which actively contribute to gene function (Morrison and Weiss 2001). In model genetic organisms, gene knockout studies (Queller et al. 2003) can examine the behavioural consequences of complete removal of the phenotype produced by a candidate gene, and expression of candidates in model genetic systems can be used to examine the phenotypic function of gene products independent of their genetic backgrounds. The study of direct behavioural interactions has several advantages over classical methods of studying behavioural evolution, it does not require the ability to make hybrids or backcrosses in the lab, nor does it involve the production and study of mutant phenotypes of questionable natural importance.

A variety of biochemical and computational techniques are used to locate the genes that directly mediate individual interactions. Current studies often involve screening candidate organs for genes of potential functional interest, for example, screens of the protein content of the sperm acrosome have been used to identify genes involved in sperm-egg interaction (Takagi et al. 1994). A similar logic underlies subtractive cDNA hybridizations that identify candidate genes expressed only in specified organs (Swanson et al. 2001b), physiological states (Thompson et al. 2003), or developmental stages (Hill et al. 2000). Genetic and chemical screening is a mainstay, but these techniques invariably require further biochemical characterization of the isolated proteins to determine their function and importance for organismal fitness. The predicted functional properties of a desired class of proteins can be used directly to identify their genetic basis. Yeast two-hybrid systems (Fields and Song 1989) use the expression of a reporter gene, triggered by extended contact of proteins attached to its translation apparatus, to identify interactions between pairs of proteins and could be employed in determining the genetic basis of brainless behaviours. Molecular green-beard candidates, genes whose biological function necessitates self interaction (Haig 1996) could be identified with a yeast two-hybrid technique by simply identifying gene products that bind to themselves. Yeast two-hybrid systems could also be used to identify interacting proteins from two different candidate organs. For example, genes expressed by the seminal glands of male

individuals could be screened against those expressed by the reproductive tract of females to identify pairs of interacting loci for molecular studies of sexual conflict and co-evolution.

Another way of identifying functionally-important genes is to estimate the evolutionary pressures leading to their divergence. Genes expressed in candidate organs can be screened *en masse* against the genome of a closely-related organism and a history of selection can be inferred from estimates of sequence evolution (Swanson et al. 2001b). These comparisons yield estimates of the importance of the candidate loci for organismal fitness. Similar computational techniques could also identify pairs of proteins that interact over evolutionary time. If change in one co-evolving locus directly selects for a compensatory change in its partner then the evolutionary rates of functional portions of the two genes are expected to show higher correlation than between two functionally unrelated genes. Inter-locus co-evolution, and an implied history of functionally important molecular interaction, should therefore be evident from phylogenetic analyses of co-evolutionary rates. Computational methods of detecting selection are targeted enough to assess the functional importance of single amino-acid substitutions (Yang and Swanson 2002) and can be incorporated into population genetic analyses, examining the selective importance of amino-acid polymorphism in natural conditions.

Selection is often assumed to be the driving force underlying phenotypic change, but studies of the molecular evolution of most genes indicate that selection operates mainly as a stabilizing force. Purifying selection constrains the evolution of the vast majority of genes; evidence for diversifying selection is not commonly found at the molecular level (Endo et al. 1996). The loci that control brainless behaviours are the major exception to this rule; pervasive adaptive evolution is a defining feature of the genes that directly mediate interactions between individuals. For example, proteins expressed in the outer cell membrane of *Plasmodium falciparum*, the micro-organism responsible for malaria, show excess polymorphism compared to those functioning in other parts of the cell (Volkman et al. 2002). The divergence of 11% of the genes expressed by the male accessory seminal gland of *Drosophila* are driven by positive selection (Swanson et al. 2001b); exceeding the genomic average by several orders of

magnitude (Endo et al. 1996). The striking prevalence and intensity of selection on brainless behaviours may exist either because tests of selection are not powerful enough to detect selection spread across the many loci that contribute to a typical quantitative trait, or because the simplified nature of the genotype-phenotype relationship and the strong and continued selection pressure exerted by co-evolutionary interactions concentrates the signal of selection at brainless behaviour loci. Regardless, the genes that directly mediate interactions between individuals are excellent systems to study the action of selection in shaping the evolution of traits whose phenotypic effects are relevant to organismal fitness.

Examples of Brainless Behaviour

Direct molecular interactions are a ubiquitous and commonly overlooked component of biological communication. An estimated three percent of genes in vertebrates are thought to encode adhesion molecules (Thiery 2003) and examples of direct molecular interactions can be found in many biological recognition processes. Some of the most interesting examples of brainless behaviour involve gene products that mediate aspects of development, host-parasite interactions and sex. Undiscovered molecular recognition processes could be important components of fertility disorders (Haig 1996), cancer (Thiery 2003), and the immunological diseases that result from improper immune activity such as multiple sclerosis. Interactions during cell-cell contact may ultimately also explain aspects of neural behaviour, for example, fetal alcohol syndrome is caused by ethanol-induced inhibition of neuron adhesion (Greenberg 2003; Wilkemeyer et al. 2003). Direct molecular interactions are already recognized as important components of virally induced diseases. The first functionally well-characterized example of positive selection came from the adaptive evolution of vertebrate MHC genes, and it is now clear that antibiotic design and treatment requires an evolutionary understanding of the molecular agents of pathogenicity (Hughes 2002).

Much of the evidence for positive selection in general comes from molecules that directly mediate conflicts of interest between biological units. In *Wolbachia*, surface proteins of parasitic strains evolve by positive selection and homologous proteins in mutualistic strains do not, suggesting that the adaptive evolution of these loci may be a mechanism of evading the host immune response (Jiggins et al. 2002). This type of antagonistic co-evolution is common to many biological interactions and appears to be a potent evolutionary force (Rice 1998). Conflicts of interest at the protein level are often exceedingly simple and offer excellent opportunities for biochemical characterizations of the phenotypic importance of adaptive genetic change. Transferrins and siderophores are proteins that compete, on behalf of their respective organisms, for limited iron reserves (Ford 2001). Since the amino-acid residues that bind iron have certainly been very efficiently optimized by selection, the co-evolutionary interaction between these two molecules likely involves a search for allosteric "pressure points" that cause these molecules to loosen their grip on an iron molecule. Combinatorial chemistry could be used to accelerate this co-evolutionary game in vitro to examine the course of adaptive phenotypic evolution. The products of parallel co-evolutionary experiments could also be competed against one another to examine the importance of contingency and constraint on molecular adaptation.

Conflicts of interest mediated by molecular interactions may also be indirectly important in shaping life history traits. For example, first-male and last-male sperm precedence are each associated with vastly different mating strategies (Birkhead and Møller 1998). In many cases, the factor that underlies sperm precedence and ultimately creates a selective advantage for life history tactics such as pre or post-copulatory mate guarding is a molecular interaction between male offensive and defensive seminal proteins, or between these proteins and the female reproductive tract (Rice 1998). Another important life history trait that might have a simple molecular impetus is the developmental stage at which organisms sequester their germ line. Variation in this trait may arise from alternate solutions to the conflict of interest that exists between germ line and somatic cells; analogous reproductive cheating behaviours have been observed to depend on a single homophilic cell adhesion protein in *Dictyostelium* slime molds (Crespi

and Springer 2003; Queller et al. 2003), suggesting that early solutions to this conflict may have been mechanistically simple. In species with doubly-uniparental mitochondrial inheritance, sex determination itself may depend on a simple recognition process; the inability to exclude male lineage mitochondria from the germ line appears to induce maleness in these taxa (Kenchington et al. 2002).

On the Origin of Species by Means of Natural Selection

Studies of speciation are intimately tied to behaviour, because an organism's behaviour is often a primary factor preventing the exchange of genetic material with closely-related taxa. The major goals of speciation research are therefore often met by studying behaviour of one kind or another. The genetic basis of most of the traits being studied is incompletely known (Hawthorne and Via 2001; Peichel et al. 2001) and even the most prominent work in this field is limited to inferring the historical action of selection from hybridization experiments which attempt to recreate ancestral phenotypes (Hatfield and Schluter 1996), or to cases of parallel evolution (Nosil et al. 2002; Rundle et al. 2000) which imply the action of natural selection. Direct molecular interactions are fundamental to many of the pre-zygotic and post-zygotic barriers that reproductively isolate species and studying the adaptive molecular evolution of these traits has the potential to establish the involvement of natural selection in the origin of species in general.

Post-zygotic Isolation

The major class of post-zygotic isolating barrier, Dobzhansky-Muller incompatibility (Turelli and Orr 2000), can result from direct molecular interaction, or lack thereof, between genes co-existing in the hybrids of incipient species. Studies of post-zygotic isolation commonly find reproductive isolation between taxa as a result of hybrid inviability or sterility, but the genetic underpinnings of these traits are incompletely known. Indeed, only a few studies, all in *Drosophila* (Barbash et al. 2003;

Presgraves 2003; Ting et al. 1998), have identified and genetically characterized loci involved in generating post-zygotic isolation between taxa. In each of the currently identified cases, reduced hybrid fitness is a consequence of disrupted protein-protein (Barbash et al. 2003; Presgraves 2003), or protein-DNA (Ting et al. 1998) interactions. This pattern, if general, suggests that an alternate strategy, which exploits the predicted functional properties of post-zygotic incompatibilities, might aid the search for the genetic underpinnings of post-zygotic isolation. Whole genome examinations of protein-protein and protein-DNA interactions are feasible with yeast two-hybrid systems (Walhout and Vidal 2001). Characterizing post-zygotic isolation with a yeast two-hybrid approach would involve proteome screens across species to identify candidate loci with the expected functional properties of a post-zygotic isolating mechanism, namely, pairs of homologous proteins that interact within each species, but do not interact across species (or vice-versa). This approach could potentially be applied to any pair of closely related taxa, extending information on the genetic bases of post-zygotic isolation beyond model genetic systems.

Pre-zygotic Isolation

Interactions between sperm and egg during fertilization are perhaps the best studied and most widely appreciated brainless behaviour. Gamete-recognition is a ubiquitous feature of sexual reproduction; every sperm-egg interaction involves molecules acting on behalf of each gamete (Palumbi 1992; Palumbi 1994). Studies of gamete recognition yield a number of consistent patterns; these loci are subject to strong positive diversifying selection that operates to drive divergence between species (Swanson and Vacquier 2002). Selection on gamete recognition proteins is particularly strong in taxa with broadly sympatric ranges (Hellberg 1998), or well developed blocks to polyspermy, which imply the interaction of single eggs with multiple sperm. Studies of the urchin sperm protein bindin have shown character displacement in sympatry (Palumbi et al. 1997), consistent with the operation of reinforcement and suggesting a role for the rapid divergence of reproductive proteins in speciation. However these loci create strong

barriers to fertilization between species and hybrids among taxa that are reproductively isolated by gamete interactions are rare in nature. Divergence at these loci could therefore have arisen after the evolution of reproductive isolation (Palumbi 1994).

Mytilus blue mussels are the exception to this rule. *Mytilus* share many of the features that typify model gamete recognition systems; species are reproductively isolated by gamete interactions (Bierne et al. 2002; Rawson et al. 2003), members of the genus are broadly sympatric (Hilbish et al. 2000), and blocks to polyspermy are well established (Togo and Morisawa 1997; Togo et al. 1995). However, reproductive isolation between three members of this genus is incomplete, and *Mytilus* species regularly produce hybrids in natural conditions. Studies of the factors that shape the evolution of gamete-recognition in *Mytilus* therefore have a direct bearing on speciation research. Chapter two describes an adaptive gamete-recognition locus polymorphism that exists within sympatric populations of *M. galloprovincialis*. Adaptive evolution of sperm Lysin-M7, the protein that lyses the egg vitelline envelope (Takagi et al. 1994), is evident in a sympatric hybridizing population that resulted from recent, human-mediated, secondary contact between allopatric *Mytilus* species (McDonald and Koehn 1988). This study offers a number of important opportunities for speciation research, because the selective factors that cause adaptive evolution within populations can be directly studied. For example, this is the first system in which the role of hybrid fitness in driving pre-zygotic isolation (reinforcement) can be directly studied on a reproductive trait with a known genetic basis.

Conclusion

Evolutionary studies of behaviour often focus on neural decision-making, but conflict and collaboration between biological units existed long before brains, and a great deal of organismal behaviour must therefore be mediated by direct molecular interactions. Studying these "brainless behaviours" offers a wealth of opportunity for evolutionary biologists. The phenotype-genotype relationship is simpler and the genetic basis easier to uncover than that of most quantitative behavioural traits. The current and

historical effects of selection can be rigorously determined using appropriate models of molecular evolution, and the role of selectively-driven changes in shaping the origin and maintenance of functional phenotypic variation can be directly measured. The phenotypic course of adaptive evolutionary change can be extensively evaluated in experimental conditions and, most importantly, the findings of these manipulations can be verified in natural conditions. The prevalence of divergent selection on the loci that control brainless behaviours indicates that they are major determinants of organismal fitness. These traits mediate interesting and important biological functions and their study is centrally important to a comprehensive understanding of phenotypic adaptation and evolution.

CHAPTER TWO: RAPID EVOLUTION OF A GAMETE-RECOGNITION PROTEIN IN A HYBRID *MYTILUS* POPULATION

ABSTRACT

Gamete fusion is a universal feature of sex. Gamete incompatibility can cause reproductive isolation between individuals and species. All sexual organisms express recognition molecules on their gamete surfaces that mediate male-female interactions, permitting or preventing fertilization. Gamete-recognition proteins typically evolve rapidly and adaptively, their divergence driven by unknown selection pressures. A role for this adaptive evolution in speciation is a tantalizing possibility, but natural hybrids between species with gametic isolation are rare and divergence may accumulate after reproductive isolation is complete. Here we report an adaptive gamete-recognition locus polymorphism within a hybridizing *Mytilus* population. Human-mediated introductions have allowed hybridisation between two ecologically-distinct, formerly allopatric *Mytilus* species. Secondary contact between *M. trossulus* and *M. galloprovincialis* is associated with adaptive evolution of Lysin-M7, a sperm protein that dissolves the egg vitelline envelope, allowing fertilization. These data represent a novel example of intraspecific selection in a contemporary geographic context on a reproductive trait whose genetic and functional basis is known. Additionally, they provide strong evidence that selection promotes extensive gamete-recognition divergence prior to, or coincident with, the evolution of reproductive isolation.

Mytilus blue mussels are globally-distributed broadcast-spawning marine bivalves. Three *Mytilus* species hybridise in nature: *M. trossulus* (T), native to the North Atlantic and Pacific, and the sister species *M. edulis* (E) and *M. galloprovincialis* (G), native to the Atlantic and Mediterranean respectively (Hilbish et al. 2000). Natural hybrids are produced in all contact zones; home range overlap and human-mediated introduction have created replicate hybrid-zones between *Mytilus* species (Fig. 1). Natural hybridisation between *M. edulis* and *M. trossulus* occurs in Atlantic Canada and the Baltic; *M. edulis* and *M. galloprovincialis* hybridise between Spain and the British Isles (Hilbish et al. 2000). The natural ranges of *M. galloprovincialis* and *M. trossulus* do not overlap; they are spatially separated, currently and historically (Wares and Cunningham 2001), by populations of *M. edulis*. Transport in shipping vessel ballast water (Geller et al. 1994) introduced *M. galloprovincialis* to the Pacific, creating secondary contact with *M. trossulus* and resulting in the first G/T hybrid zones, in California (Geller et al. 1994; McDonald and Koehn 1988; Rawson et al. 1999) and possibly Japan (Suchanek et al. 1997).

Mytilus species remain genetically distinct, despite extensive hybridisation, because of several pre- and post-zygotic reproductive isolating barriers. The two most closely related species, *M. edulis* and *M. galloprovincialis*, differ ecologically. The relative fitness of G/E hybrids depends on salinity, wave exposure, temperature and other exogenous factors; G/E larvae exhibit low viability even in lab conditions (Bierne et al. 2002). A similar suite of ecological differences are evident between the more distantly-related *M. edulis* and *M. trossulus* (Rawson et al. 2003; Toro et al. 2002). *M. trossulus* and *M. galloprovincialis* respond differently to temperature stress (Hofmann and Somero 1996) and G/T populations in California exhibit hybrid deficits and limited introgression despite extensive hybridization (Rawson et al. 1999) consistent with low fitness of G/T hybrids. Pre-zygotic isolation among sympatric *Mytilus* results from asynchronous spawning (Toro et al. 2002) and assortative gamete interaction, which is stronger between more distantly-related species (Bierne et al. 2002; Rawson et al. 2003). In Mediterranean *M. galloprovincialis*, sperm-egg binding and fertilization are reduced >50% in the presence of the specific carbohydrate that forms spikes in the vitelline coat surrounding

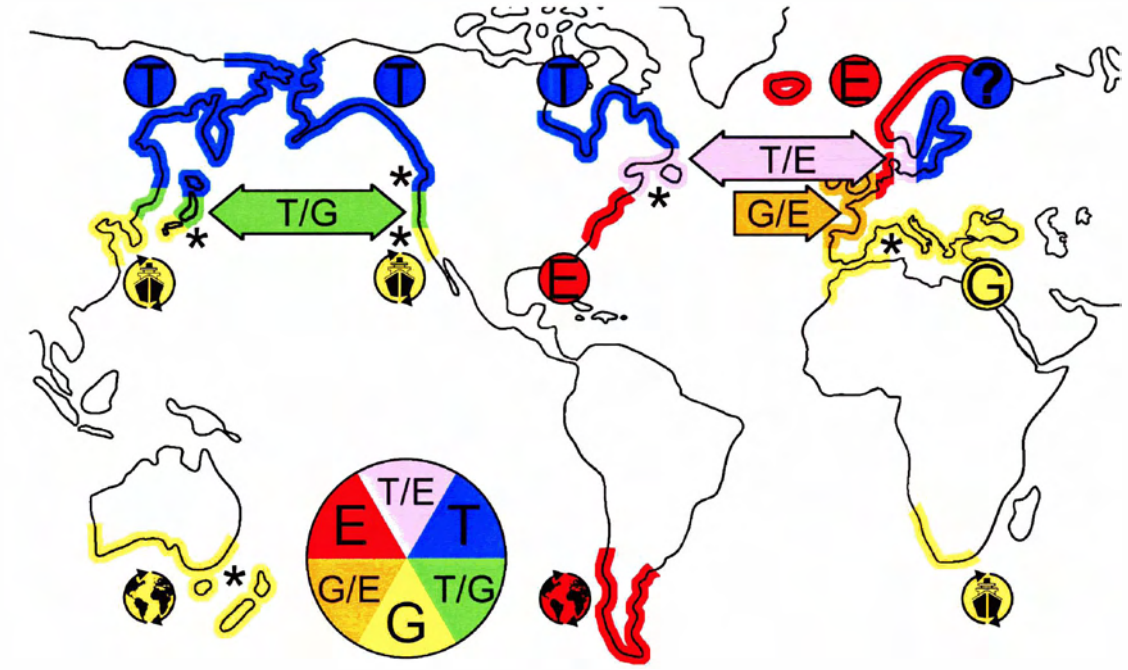


FIGURE 1 - GLOBAL DISTRIBUTION OF HYBRIDIZING *MYTILUS* SPECIES.

Primary colours represent pure allopatric populations, *M. galloprovincialis* (yellow), *M. edulis* (red), *M. trossulus* (blue). Hybrid populations are shown with the appropriate secondary colour. Circled symbols denote the origin of each population: letters are adjacent to populations in their natural range, globes represent populations transported during the Pleistocene transatlantic exchange (~1.5 mya), ships identify populations introduced by human activities. Hybridizing populations are highlighted by arrows, replicate G/T and T/E hybrid zones exist in the Pacific and Atlantic respectively. The origin of Baltic *M. trossulus* is unknown. Asterisks indicate sampling locations (methods).

the egg, but not in the presence of structurally similar sugars (Focarelli et al. 1991). Fertilization is not inhibited by this carbohydrate in Japanese *Mytilus* sp (Togo and Morisawa 1997), which suggests the presence of species or population-specific patterns of fertilization mediated by variation in carbohydrate recognition. Sperm-egg recognition is therefore an important component of pre-zygotic isolation in *Mytilus*.

Mytilus sperm lysins, proteins that bind and lyse the egg vitelline envelope facilitating sperm-egg fusion, are C-type lectins (Takagi et al. 1994), members of a functionally well-characterised family of carbohydrate-binding proteins (Drickamer and Fadden 2002). Selection drives divergence of at least one of these C-type lectins, sperm Lysin-M7, between *Mytilus* species (Riginos and McDonald 2003) but species-specific lysin-carbohydrate affinities have not yet been tested in this genus. Lysin-M7 alleles group into four distinct clades representing *M. trossulus*, *M. edulis*, *M. galloprovincialis* and an intraspecific radiation of *M. galloprovincialis* alleles (Fig. 2). Branch-specific maximum likelihood estimates (Yang 1997) of positive selection on the complete Lysin-M7 coding region detect adaptive evolution ($\omega>1$) among divergent *M. galloprovincialis* (G_D) alleles and between these alleles and typical *M. galloprovincialis* (G) alleles (Fig 2). Pairwise maximum likelihood estimates confirm this result, finding evidence of selection only in the G_D radiation; within clades (G_D, max $\omega=2.7621$; G, max $\omega=1.0391$; E, max $\omega=0.5749$; T, max $\omega=0.6798$) and between clades ($G_D|G$, max $\omega=3.8928$, $G_D|E$, max $\omega=1.006$; $G_D|T$, max $\omega=0.5459$; all other comparisons, max $\omega=0.4534$). Evidence for positive selection suggests that the amino-acid substitutions which occurred during divergence of the G_D clade alter the carbohydrate affinity of Lysin-M7.

C-Type lectins are active in immunity and their carbohydrate binding activity has been the subject of detailed functional investigations (Drickamer and Fadden 2002). C-type lectins are diagnostically identified by three disulphide-bridges that conserve a characteristic three-dimensional structure among gene-family members (Drickamer and Fadden 2002). Carbohydrate binding by C-Type lectins is calcium dependent, and structural mutagenesis studies have identified specific amino-acids that bind calcium and

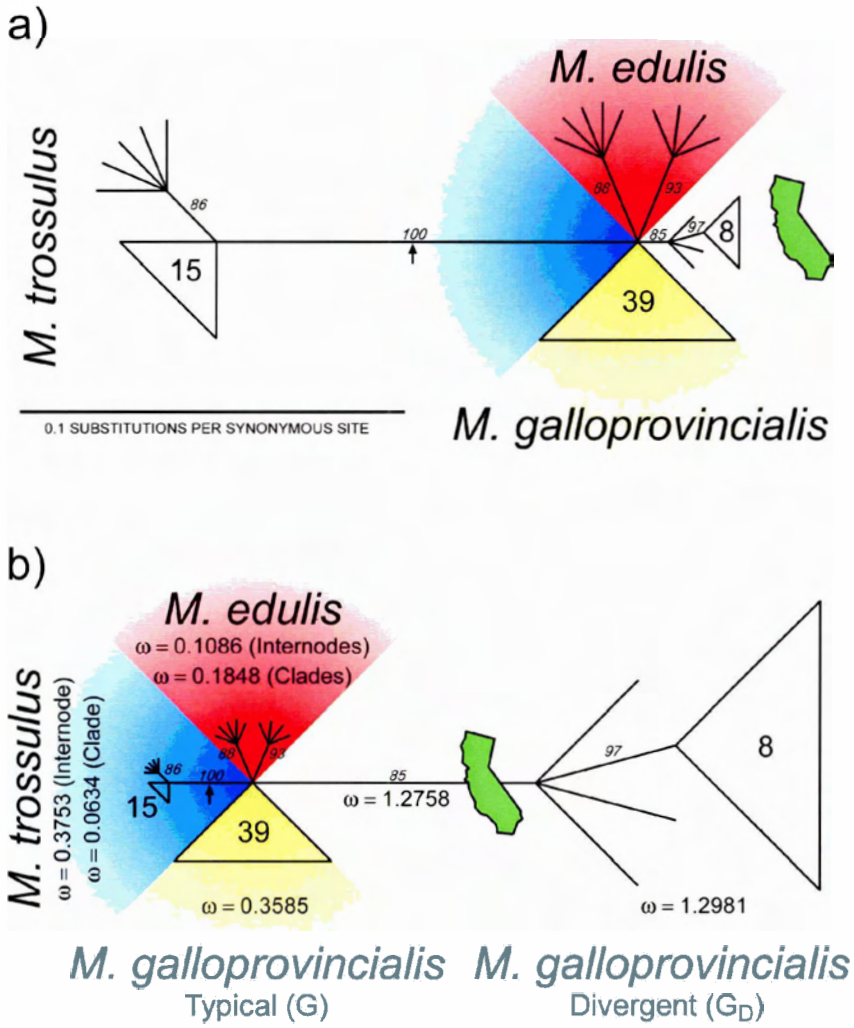


FIGURE 2 - PHYLOGENETIC RELATIONSHIPS AMONG *MYTILUS* LYSIN-M7 ALLELES.

a) Topology of relationships between unique Lysin-M7 alleles, branch lengths represent the rate of neutral evolution, number of substitutions per synonymous site. b) The same topology with branch lengths scaled to represent clade-specific estimates of selection. ω ratios exceed the neutral expectation (ω=1) in the divergent *M. galloprovincialis* alleles indicating that G_D is a product of intraspecific adaptive divergence. In both figures triangles enclose the indicated number of unique alleles. Bootstrap support va lues are italicised. The arrow indicates the position of the root, determined with Lysin-M6.

coordinate carbohydrate affinity and binding (Drickamer and Fadden 2002). Homology models (Lambert et al. 2002) predicting three-dimensional structures of the carbohydrate recognition domain of Lysin-M7 indicate that many amino acid substitutions which have accumulated between species and during the intraspecific G_D radiation occur in or near regions of functional importance (Fig. 3). Amino acid 144 is a particularly likely determinant of carbohydrate specificity, occurring two residues upstream of the functionally-conserved group of amino acids that bind one of the two calcium molecules (Drickamer and Fadden 2002). A substitution at residue 144 was previously identified as a target of selection in typical M. galloprovincialis alleles (Riginos and McDonald 2003) and this position is substituted a second time in derived products of the G_D radiation (Table 1). Ancestral G_D alleles are extant in Californian Mytilus populations, offering future opportunities to evaluate phenotypic evolution during the course of adaptive divergence.

Recombination contributes to the divergence of Lysin-M7; sequence comparisons (Hudson and Kaplan 1985) reveal a minimum of eight recombination events among Lysin-M7 alleles. Two diagnostic amino acid residues (18V, 144A) identify G_D alleles as *M. galloprovincialis*, but replacements at position 56 (S) and 180 (A) are typical of *M. edulis* and *M. trossulus* respectively, indicating that hybridization and recombination with lysins of these taxa, or convergent evolution, occurred during the divergence of G_D (Table 1). Three amino acid replacements (98Y, 141R, 144P) result from inter-locus gene conversion or convergence; these sites are identical to those of Lysin-M6 alleles and exist near functionally-relevant positions of the molecule (Fig. 3). Gamete-recognition loci apparently evolve via punctuated bursts of rapid adaptive evolution (Swanson and Vacquier 2002); when appropriate divergent selection pressures exist, new Lysin alleles may evolve from pre-existing polymorphism and functional variation, rather than mutating *de novo*. Gene conversion may generate new diversity very quickly (Nielsen et al. 2003); indeed, if hybridization between *M. galloprovincialis* and *M. trossulus* is the factor initiating adaptive Lysin-M7 divergence, the polymorphism composing the G_D radiation may have been assembled in as little as 100 years (Geller 1999; McDonald and Koehn 1988)

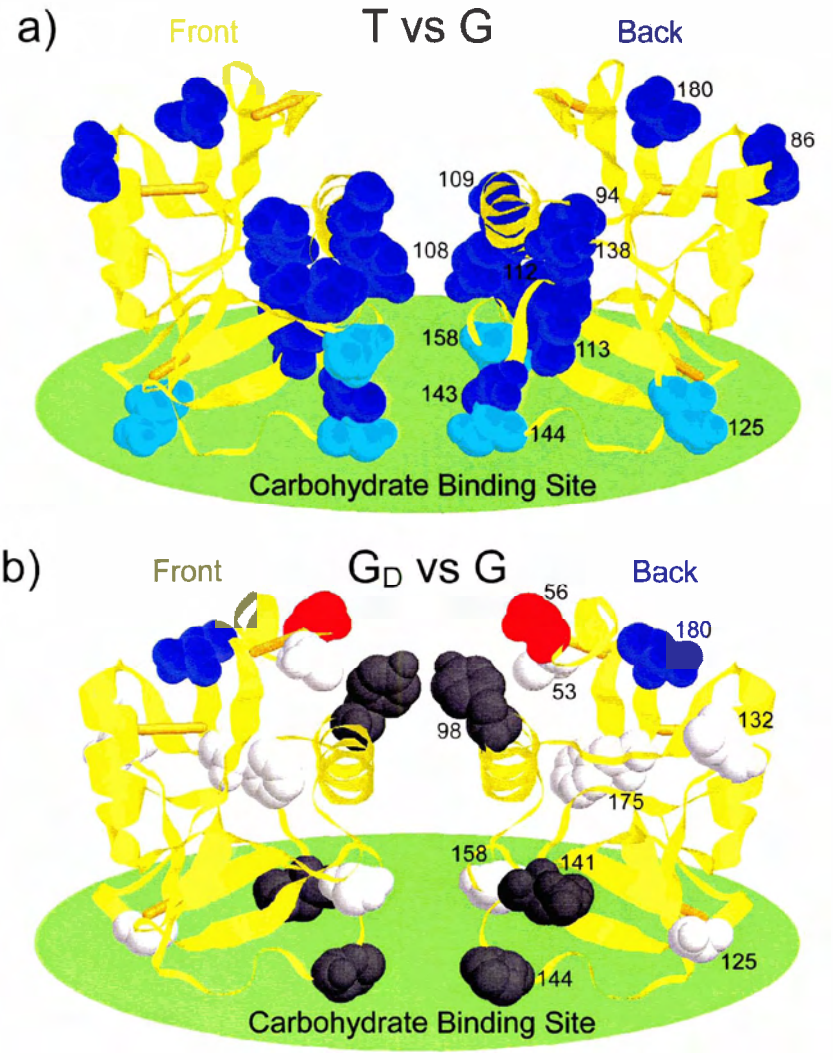
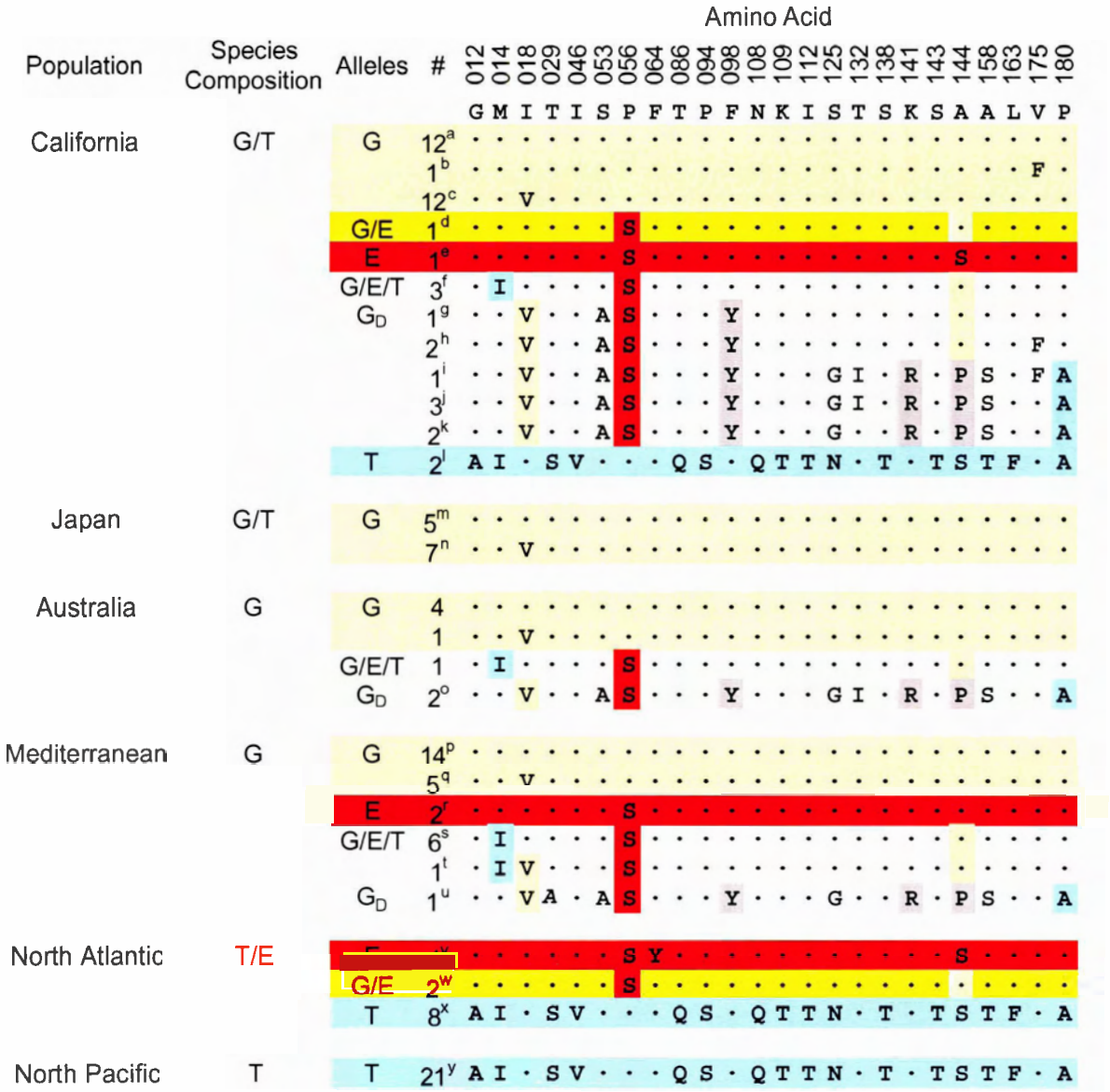


FIGURE 3 - SPATIAL LOCATIONS OF AMINO ACID SUBSTITUTIONS IN THE CARBOHYDRATE RECOGNITION DOMAIN OF *MYTILUS* LYSIN-M7.

a) Pairwise differences between T and G, blue residues are unique to M. trossulus; light blue (125, 144, 158) are substituted again in G_D. b) Pairwise differences between G_D and G. White residues (53, 125, 132, 158, 175) are unique to G_D alleles, Red (56), blue (180) and grey (98, 141, 144) are identical to M. edulis, M. trossulus and Lysin-M6 alleles respectively. Yellow ribbons are identical to M. galloprovincialis; yellow bars represent disulphide bridges. Numbers refer to the mature peptide; singletons are not shown (Table 1). In both comparisons amino acid substitutions occur in or near regions known to affect carbohydrate specificity.

TABLE 1 - LYSIN-M7 VARIATION IN *MYTILUS*



Lysin-M7 alleles found in each sampling location. Dots represent identity with consensus. Colouring conventions follow Figure 1. Grey shaded amino acids are identical to Lysin-M6. Details of alleles sampled from each sub-location, singleton amino acid replacements are in italics, dashes separate singletons from the same allele: [a]EK4,OP2,MB3(*75R*),SD3. [b]MB1. [c]EK2,MB1,OP2,RF2,SD5(*7K, 73N, 137S*). [d]SD1. [e]SD1. [f]EK2(*159R*),SD1. [g]MB1. [h]MB2. [i]MB1. [j]MB2(*21D-73E,35R-152A-159R*),RF1. [k]RF1, SD1. [l]MB1,OP1. [m]KO5(*96S*) [n]KO7(*5C,18L,48S,175A*). [o]AU2(*1S,29A*). [p]IK2(*4H*), RP7(*5D,82T,151R, 179R*),VE5(*59G-165N*). [q]IK2(*85R-172H*),VE3(*97A*). [r]IK1,RP1(*32L*). [s]IK4(*12D,72T*), RP1,VE1. [t]VE1. [u]RP1(*77R*). [v]NF4(*97A*). [w]NF2(*173K*). [x]NF8(*50K,113P*). [y]HA4(*165H*), OB17(*26K,130R,166K*).

Divergence of Lysin-M7 within *M. galloprovincialis* is associated with ongoing hybridization of *Mytilus* populations in California. G_D alleles are more abundant in Californian samples ($p=0.023$, random assignment of alleles to populations) and less abundant in Mediterranean samples ($p=0.044$) than expected by chance. The three G_D alleles found outside sympatry are derived products of the radiation sampled in California (Table 1). The polymorphism composing the G_D radiation may have originated in allopatry, maintained by intraspecific sexual selection (Palumbi 1998) and reaching high frequency as a result of selection pressures associated with secondary contact between *M. galloprovincialis* and *M. trossulus*. Alternately, the diversity of extant G_D alleles in California may indicate divergence in sympatry; G_D alleles may occur elsewhere at low frequency as a result of movement in ballast water, the same mechanism responsible for the initial (McDonald and Koehn 1988) and continued (Geller et al. 1994) introduction of *M. galloprovincialis* to the Pacific. Ballast introduction of *Mytilus* is pervasive, for example, 10% of individuals in Australia are products of contemporary introductions (Hilbish et al. 2000) and transport of Lysin-M7 alleles is evident in all but two of the populations sampled here (Table 1).

Californian populations of *M. galloprovincialis* appear to have arisen by direct introduction into sympatry with *M. trossulus* (Geller 1999). The rarity of species following such introductions could have important consequences for gamete-recognition evolution. Eggs of rare broadcast-spawning species are besieged by heterospecific sperm and may lack effective means of mate choice and polyspermy avoidance; sperm of rare species may be selected to efficiently fertilize the most frequently encountered, heterospecific, eggs. Two hypotheses proposed to explain the rapid evolution of reproductive proteins, sexual selection and reinforcement (Swanson and Vacquier 2002), explicitly predict increased gamete-recognition divergence in sympatry. These mechanisms make opposite predictions about the course of evolution following secondary contact via introduction; predicting phenotypic convergence and divergence respectively, of introduced recognition systems relative to residents.

Reinforcement hypotheses predict that reduced hybrid fitness will cause divergent selection on egg receptors to minimize heterospecific mating, producing an attendant evolutionary response by sperm and ultimately a new co-adapted mate recognition system (Marshall et al. 2002). Empirical studies of gamete-recognition are consistent with reinforcement (Geyer and Palumbi 2003); our data are the first to extend these results to naturally hybridizing taxa. Three of the four Californian G_D individuals for whom diploid Lysin-M7 genotypes were recovered are G_DG_D homozygotes, the remaining individual is G_DG; this pattern implies a modified male-female recognition process, but could also be explained by non-random opportunity for inter-clade sperm-egg interaction. A small proportion of *M. trossulus*, relative to *M. galloprovincialis*, were expressing Lysin-M7 during our survey of California (Table 1), suggesting that temporal isolation (Toro et al. 2002) contributes to pre-zygotic isolation in Californian *Mytilus*.

Sexual selection in the form of male-female conflict over fertilization, or male-male competition could optimise reproductive phenotypes within populations (Palumbi 1998). Following introduction, if post-zygotic isolation is insufficient to drive reinforcement, rare sperm and eggs may be convergently selected toward existing, population-specific, optima. Polyspermy blocks are active in *Mytilus* (Togo and Morisawa 1997) indicating that a given egg may interact with multiple sperm and that mechanisms invoking sexual conflict could plausibly operate (Palumbi 1998). Alternately, the reproductive function currently filled by Lysin-M7 may have originated as a costly signal of male immunocompetence. C-type lectins are agents of immunity in animals (Drickamer and Fadden 2002) and though Lysin-M7 appears to be expressed only in mature male gonad tissue, which suggests a strictly reproductive function, the involvement of Lysin-M7 in immunity and as a signal of male quality is an untested possibility.

Evolutionary studies of pre-zygotic isolation are advancing on two fronts. Ecological speciation studies link hybrid fitness in nature with divergent selection on reproductive traits whose genetic basis is largely unknown (Schluter 2001). Genetic studies of reproductive protein evolution have yielded substantial evidence of adaptive divergence, but these differences accumulate so rapidly that their importance in speciation is difficult to assess (Swanson and Vacquier 2002). Ideally, studies of

speciation would incorporate both approaches, analysing the causes of divergent selection on traits whose genetic and functional phenotypic basis is known. Studies of this kind are rare because catching the genes responsible for reproductive isolation in the act of divergence is uncommon. *Mytilus* hybrid zones represent the first system amenable to both ecological and molecular analyses of pre-zygotic isolation. The available evidence suggests that adaptive intraspecific divergence of sperm Lysin-M7 is associated with secondary contact and hybridization between ecologically-distinct *Mytilus* species, and occurred before complete reproductive isolation; these findings implicate the rapid evolution of reproductive proteins in speciation.

METHODS

Tissue Collection, PCR Amplification and Sequencing

Mature gonad tissue samples were collected (Ambion RNA*Later*) from sympatric and allopatric *Mytilus* populations (Fig.1), sample size at each sub-location is in brackets. **California** - sympatric G/T, January 2002: Eureka EK(8) 40°48.33'N 124°07.30'W, San Rafael RF(4) 37°:53:05'N 122°:26:35'W, Oyster Point OP(5) 37°:35:18N 122°:19:02W, Morro Bay MB(12) 35°:19:45'N 120°:49:57'W, San Diego SD(12) 32°:38:02'N 117°:06:32'W. **Japan** - sympatric G/T, April 2002: Kobe KO(12) 34°:40:44'N 135°:10:32'E. **Australia** - unknown G, March 2003: Sydney AU(8) 33°:52:29'S 151°:13:51'E. **Mediterranean** - allopatric G, February 2003: Iraklion Greece IK(9) 35°:20:33'N 25°:08:06'E, Rapallo Italy RP(10) 44°:20:55'N 9°:13:51'E, Venice Italy VE 45°:25:45'N 12°:19:40'E. **North Atlantic** - sympatric E/T, July 2001: Clarenville Newfoundland NF(14) 48°:09:00'N 53°:53:56'W. **North Pacific** - allopatric T, August 2000: Oyster Bay British Columbia OB(17) 49°:57:53'N 125°:12:23'W; April 2002 Haynes Alaska HA(4) 59°:14:23'N 135°:26:32'W.

Primers that amplify the 540 bp coding region of Lysin-M7 and flanking regions from cDNA (M7-143F, 5' GTC TTT TGT GCA GCA CTT ATT GT 3' and M7-874R 5' TTC ACT CTT CAC TTG TCA TGT G 3') were designed using published Lysin-M7 sequence (Genbank Accession D14731; Takagi et al. 1994): PCR amplifications of Lysin-M7 were performed on cDNA reverse transcribed (Amersham RT-PCR Beads) from the total RNA of mature male gonad tissue (Qiagen RNeasy). PCR product was cloned (Qiagen PCR Cloning^{plus}) and a minimum of two clones per allele were sequenced on both strands (Perkin Elmer Big Dye Terminator, ABI automated sequencer) using SP6 and T7 primers. A maximum of two alleles were recovered from any one individual indicating that these PCR primers amplify a single locus. Partial sequences are recoverable from genomic DNA and complete sequences are reproducible across RT-PCR reactions, clones and individuals and are therefore not the product of PCR template switching or RT-PCR artefacts.

Data Analysis

Phylogenetic relationships between Lysin-M7 alleles and topology confidence estimates were generated by Paup 4.0b10 (Swofford 2001); Neighbour joining, uncorrected p distance, 100,000 bootstrap replicates, 80% majority-rule consensus) and branch specific maximum likelihood estimates of non-synonymous (dN) and synonymous (dS) substitution rate ratios ($dN/dS = \omega$) and associated likelihood ratio test statistics (Table 2) were generated using PAML 3.13 (Yang 1997). Lysin-M7 shows evidence of recombination; a conservative unresolved phylogeny was therefore used to estimate clade specific ω ratios (Fig. 2), topology ((T), (E), G, (G_D)). Recombination causes the inference of two *M. edulis* sub-clades (Fig 2.); analysing them separately, together, or collapsing the E and G clades does not significantly change the ω ratio estimates for other clades. Pairwise maximum likelihood ω ratio estimates were also made with PAML (codon frequency=F3X4, κ estimated). Three dimensional structures and alignments of the most similar C-type lectin carbohydrate recognition domain with a known structure (human tetranectin, PDB: 1TN3, Percent identity: T 25.9%, P 27.3%) and M7-Lysin were generated using ESyPred3D (Lambert et al. 2002);) and visualised with Protein Explorer (Martz 2001).

An exact test of sample differentiation (Raymond and Rousset 1995) conducted in Arlequin version 2.000 (Schneider 2000) and based on the frequency of G and G_D alleles in California, Japan, Australia, and the Mediterranean detects significant differentiation between Californian and Mediterranean samples ($p= 0.03925$); all other comparisons showed no differentiation ($p>0.09$; 10,000 Markov chain, 1000 dememorization steps). The expected null frequency of G_D alleles in each sample was estimated from 14697 replicate data sets. A total of 88 (12 G_D and 76 G) alleles were randomly placed into the four locations according to sample size in the actual data set. The probability of observing as many (or as few) G_D alleles at each sampling location was estimated as the proportion of simulated data sets containing equal to or greater than (less than) the observed number. The minimum number of recombination events was estimated with DNAsp 3.53 (Hudson and Kaplan 1985; Rozas and Rozas 1999).

TABLE 2 - LIKELIHOOD RATIO TESTS OF SELECTION

A) Log likelihood and parameter estimates under specific models of evolution

Model	lnL	np	κ
A. 1 Ratio: $\omega T=\omega E=\omega G=\omega G_D$	-2355.3724	84	3.821
B. 2 Ratio $\omega T=\omega E=\omega G$, ωG_D	-2342.1839	85	3.824
C. 4 Ratio ωT, ωE, ωG, ωG_D	-2336.1642	87	3.827
D. 7 Ratio ωT, ωE, ωG, ωG_D*	-2330.1442	90	3.82
E. 17 Ratio ωT, ωE, ωG, Free ωG_D*	-2325.9541	100	3.82
F. Free Ratio Model	-2290.0248	165	3.822

The topology [((T),(E), G, (G_D))] is used

* Indicates that internode branches are considered separately from their respective clades.

lnL: log likelihood value

np: number of parameters of the model

κ : estimated transition/transversion rate ratio

B) Clade-specific ω ratio estimates under specific models of evolution

Model	ωT_C	ωT_I	ωE_C	ωE_I	ωG_C	$\omega G_{D\,C}$	$\omega G_{D\,I}$
A.	0.282	-	0.282	-	0.282	0.282	-
B.	0.208	-	0.208	-	0.208	1.282	-
C.	0.145	-	0.116	-	0.359	1.281	-
D.	0.063	0.375	0.109	0.185	0.359	1.276	1.2981
E.	0.063	0.375	0.109	0.185	0.359	free	1.2981
F.	free	free	free	free	free	free	free

The topology [((T),(E), G, (G_D))] is used

ωN_C: dn/ds rate ratio for the respective clade

ωN_I: dn/ds rate ratio for the respective internode

C) Likelihood ratio test statistics (χ^2 tests)

Models	Null	Alternate	2ΔlnL	d.f.	*P*
A vs B	$\omega T=\omega E=\omega G=\omega G_D$	$\omega T=\omega E=\omega G, \omega G_D$	26.376902	1	0.001‡
B vs C	$\omega T=\omega E=\omega G, \omega G_D$	$\omega T, \omega E, \omega G, \omega G_D$	12.039516	2	0.002†
C vs D	$\omega T, \omega E, \omega G, \omega G_D$	$\omega T, \omega E, \omega G, \omega G_D$*	12.040014	3	0.007†
D vs E	$\omega T, \omega E, \omega G, \omega G_D$*	$\omega T, \omega E, \omega G$, Free ωG_D*	8.380212	10	0.591
D vs F	$\omega T, \omega E, \omega G, \omega G_D$*	Free Ratio Model	80.238688	75	0.318

* Indicates that internode branches are considered separately from their respective clades

ωN: dn/ds rate ratio for the respective clade or internode

†: very significant, $p<0.01$

‡: extremely significant, $p<0.001$

CONCLUSION

Gamete-recognition loci offer the novel opportunity to study a truly general biological process. Fertilization is the common thread that unites all sexually reproducing organisms, and the conclusions drawn from studies of gamete-recognition are therefore of general importance to sexual selection and speciation. Sperm-egg binding and fusion is mediated by direct interactions between molecules which act on behalf of each gamete, permitting or preventing fertilization (Palumbi 1998). Molecular recognition processes are mechanistically simpler than other types of behavioural recognition (Palumbi 1998), allowing more detailed investigation of functional phenotypic evolution than is possible with most traits. Evidence for gametic isolation has also been found in taxa with internal fertilization (Marshall et al. 2002), indicating that gamete-recognition can override other mechanisms of mate choice. Rapid evolution driven by positive selection is evident in many of the reproductive proteins studied to date (Swanson and Vacquier 2002) suggesting that the phenotypes these loci produce are important contributors to organismal fitness. The selection pressures that cause this adaptive divergence are unknown, and it is likely that multiple processes drive the adaptive evolution of reproductive proteins. Resolving the relative importance of candidate selection pressures, and the circumstances under which they operate, is the most important challenge facing gamete-recognition researchers.

A number of selection pressures have been proposed to explain the divergence of gamete-recognition proteins (Swanson and Vacquier 2002), but the accumulating evidence suggests that two processes are particularly likely to motivate this adaptive evolution, reinforcement and sexual selection. The evolution of gamete-recognition loci is more rapid in taxa that are broadly sympatric, consistent with, but not conclusive evidence of, reinforcement (Geyer and Palumbi 2003). The study described in this thesis is the first to examine the evolution of gamete-recognition in naturally hybridizing taxa. Enhanced gamete-recognition divergence in sympatric *Mytilus* occurs before reproductive isolation is complete, suggesting that adaptive reproductive protein evolution may be important in speciation (Palumbi 1994; Swanson and Vacquier 2002).

Gamete-recognition loci are often (Palumbi 1999), but not always (Swanson et al. 2001a), typified by extensive intraspecific polymorphism, which appears to function in intraspecific mate recognition and sexual selection (Palumbi 1998; Palumbi 1999). Both the male and female components of gamete-recognition systems appear to evolve by positive selection (Galindo et al. 2003; Swanson et al. 2001c), suggesting that sexual selection may drive antagonistic co-evolution between the sexes. Processes of sexual conflict may be particularly important in broadcast-spawning taxa because the eggs of a single female can be fertilized by many different males (Palumbi 1998). The predictions of reinforcement and sexual selection regarding the maintenance of genetic polymorphism and the evolution of reproductive incompatibility via gamete incompatibility are not yet clear. Determining the roles of these two processes in shaping gamete-recognition divergence and polymorphism will ultimately require diploid models of sexual conflict and speciation, extending existing haploid models (Gavrilets 2000; Gavrilets and Waxman 2002), and models of reinforcement that take into account aspects of fertilization ecology (Marshall et al. 2002).

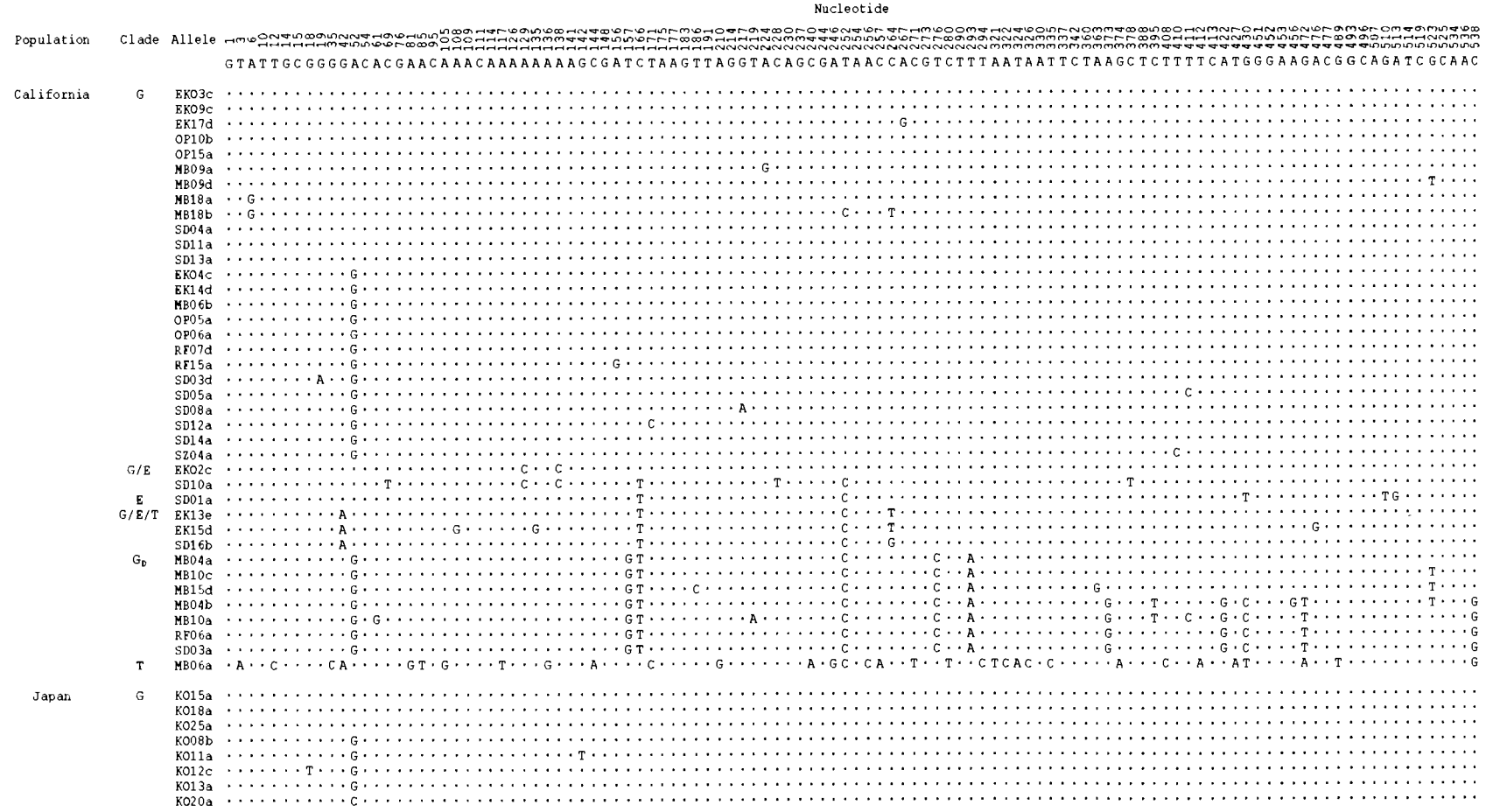
The ecology of fertilization is potentially an important component of gamete-recognition evolution and reproductive isolation. Broadcast-spawning species are particularly interesting systems for studying gamete-recognition evolution, because in these taxa few other mechanisms can lead to the evolution of pre-zygotic reproductive isolation. Much of the work on gamete-recognition comes from marine invertebrates, organisms whose fertilization ecology is complex, variable, and difficult to study observationally because fertilization occurs in the water column. Natural fertilization ecology may be an important, and underestimated (Palumbi 1992; Palumbi 1994), factor generating reproductive isolation between broadcast-spawning taxa. Assortative gamete interaction between broadcast-spawners is often asymmetrical. For example, the females of some species have particularly discriminating eggs (McCartney and Lessios 2002); this pattern may be a consequence of asymmetrical opportunity for heterospecific sperm-egg interaction created by the fertilization conditions of each species. Existing studies have shown conclusively that gamete interaction can lead to assortative mating between species in the lab (Bierne et al. 2002; McCartney and Lessios 2002; Metz et al. 1994),

future work needs to focus on inferring the ecological importance of gamete incompatibility from patterns of fertilization in natural conditions. Comparative studies of taxa in which males broadcast sperm and females retain a brood of eggs, such as *Acropora* corals (Miller and van Oppen 2003), or spawn eggs in enclosed packages such as *Montastrea* corals, may shed light on the importance of fertilization ecology and sexual selection in creating gamete incompatibility and pre-zygotic reproductive isolation between species.

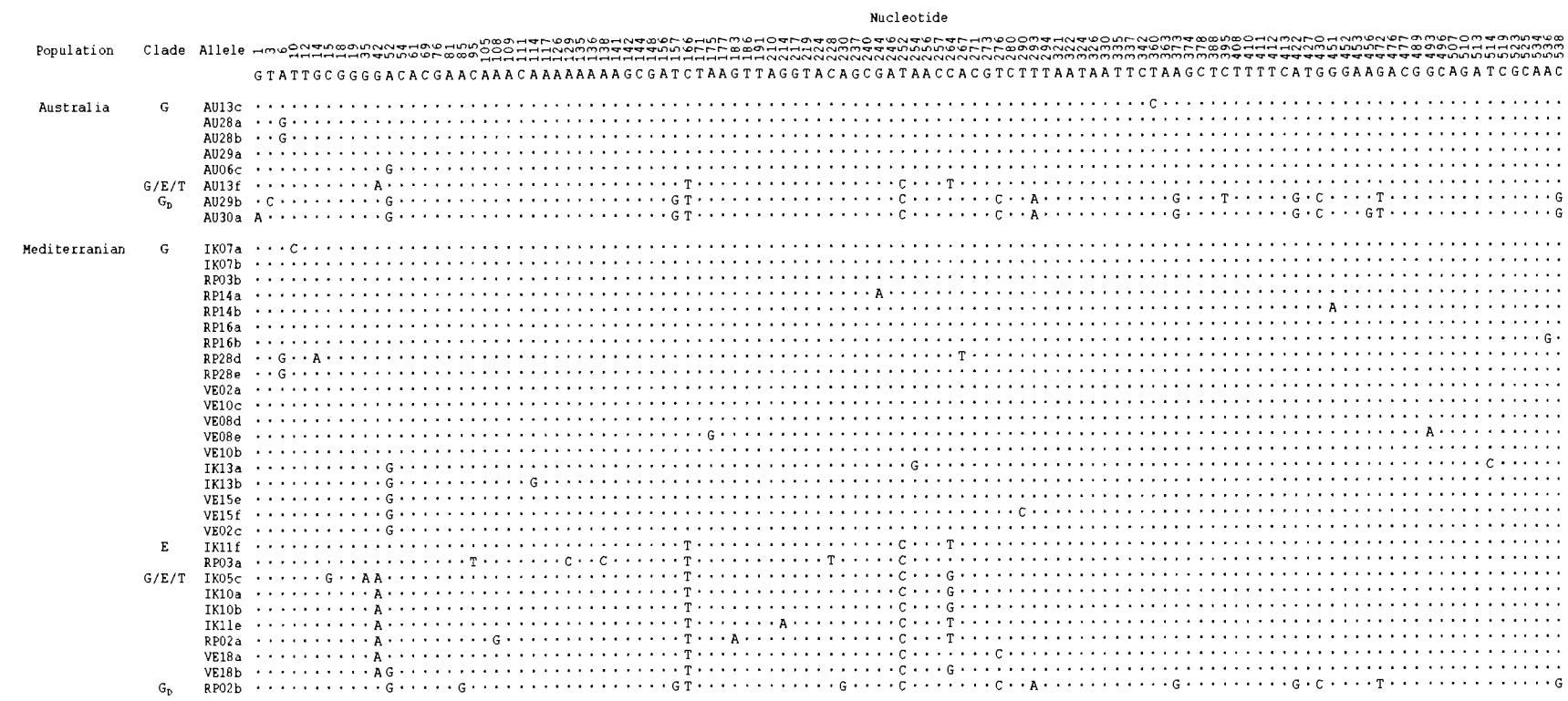
The most exciting aspect of gamete-recognition is the promise these systems hold for studies of molecular adaptation. Evidence for positive selection is being found on an increasing number of traits, but few studies take the next logical step and analyze the functional phenotypic importance of adaptive genetic change. As the genetic and functional phenotypic bases of traits underlying reproductive isolation become known (Presgraves et al. 2003; Swanson and Vacquier 2002), the current fascination with speciation will move very naturally to a focus on the architecture of adaptation. Studies of gamete-recognition are at the forefront of this transition for two reasons. First, the genetic basis of gamete-recognition is easier to uncover than that of most traits; these genes are often expressed only in reproductive tissues which greatly eases the discovery and screening of candidate loci (Swanson et al. 2001b). Second, the importance of phenotypic change at loci that directly mediate interactions between individuals is amenable to experimental phenotypic manipulation and directly measurable in terms of its importance for organismal fitness. Even if QTL studies can be narrowed down to specific genes (Presgraves et al. 2003), studies of the genetic bases of quantitative traits are likely to stumble at this point because the mechanisms of organismal phenotype production are shrouded in the interactions between many loci of small effect. Gamete-recognition systems are models for how simple molecular interactions can create biologically meaningful phenotypes and the prospects of these simple behavioural systems for studies of phenotypic adaptation are unmatched.

APPENDIX

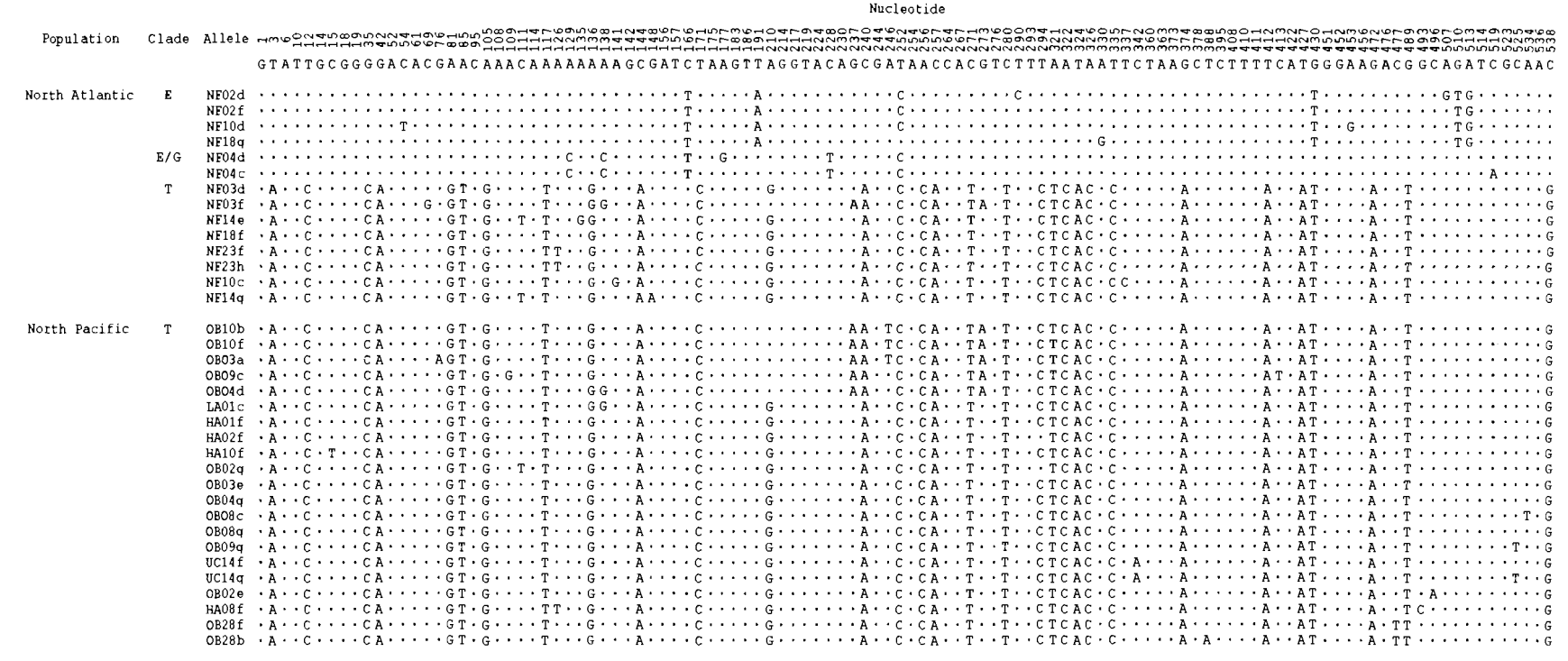
Nucleotide variation in *Mytilus* Lysin-M7 alleles



Nucleotide variation in *Mytilus* Lysin-M7 alleles (continued)



Nucleotide variation in *Mytilus* Lysin-M7 alleles (continued)



Lysin-M7 alleles found in each sampling location. Allele identifiers are coded, upper case letters denote the sampling location (see methods) and numbers denote the individual carrying a given allele; lower case letters indicate the clone from which a given allele was first recovered. Dots represent identity with the consensus sequence shown above each sub-table. Nucleotide positions are given relative to the coding region of the mature peptide.

REFERENCES

Barbash, D. A., D. F. Siino, A. M. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 100:5302-5307.

Bierne, N., P. David, P. Boudry, and F. Bonhomme. 2002. Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. Evolution 56:292-298.

Birkhead, T. R., and A. P. Møller. 1998. Sperm competition, sexual selection and different routes to fitness. Pp. 757-781 *in* T. R. Birkhead and A. P. Møller, eds. Sperm Competition and Sexual Selection. Academic Press, New York.

Boake, C. R. B., S. J. Arnold, F. Breden, L. M. Meffert, M. G. Ritchie, B. J. Taylor, J. B. Wolf, and A. J. Moore. 2002. Genetic tools for studying adaptation and the evolution of behavior. American Naturalist 160:S143-S159.

Chang, B. S. W., and M. J. Donoghue. 2000. Recreating ancestral proteins. Trends in Ecology & Evolution 15:109-114.

Crespi, B., and S. Springer. 2003. Social slime molds meet their match. Science 299:56-57.

Drickamer, K., and A. J. Fadden. 2002. Genomic analysis of C-type lectins. Pp. 59-72. Glycogenomics: The Impact of Genomics and Informatics on Glycobiology.

Endo, T., K. Ikeo, and T. Gojobori. 1996. Large-scale search for genes on which positive selection may operate. Molecular Biology and Evolution 13:685-690.

Fields, S., and O. K. Song. 1989. A novel genetic system to detect protein protein interactions. Nature 340:245-246.

Focarelli, R., D. Rosa, and F. Rosati. 1991. The vitelline coat spikes - a new peculiar structure of *Mytilus-galloprovincialis* eggs with a role in sperm-egg interaction. Molecular Reproduction and Development 28:143-149.

Ford, M. J. 2001. Molecular evolution of transferrin: Evidence for positive selection in salmonids. Molecular Biology and Evolution 18:639-647.

Galindo, B. E., V. D. Vacquier, and W. J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proceedings of the National Academy of Sciences of the United States of America 100:4639-4643.

Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403:886-889.

Gavrilets, S., and D. Waxman. 2002. Sympatric speciation by sexual conflict. Proceedings of the National Academy of Sciences of the United States of America 99:10533-10538.

Geller, J. B. 1999. Decline of a native mussel masked by sibling species invasion. Conservation Biology 13:661-664.

Geller, J. B., J. T. Carlton, and D. A. Powers. 1994. PCR based detection of mtdna haplotypes of native and invading mussels on the Northeastern Pacific coast - latitudinal pattern of invasion. Marine Biology 119:243-249.

Geyer, L. B., and S. R. Palumbi. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution 57:1049-1060.

Greenberg, D. A. 2003. Linking acquired neurodevelopmental disorders to defects in cell adhesion. PNAS 100:8043-8044.

Haig, D. 1996. Gestational drive and the green-bearded placenta. Proceedings of the National Academy of Sciences of the United States of America 93:6547-6551.

Hatfield, T., and D. Schluter. 1996. A test for sexual selection on hybrids of two sympatric sticklebacks. Evolution 50:2429-2434.

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904-907.

Hellberg, M. E. 1998. Sympatric sea shells along the sea's shore: The geography of speciation in the marine gastropod *Tegula*. Evolution 52:1311-1324.

Hilbish, T. J., A. Mullinax, S. I. Dolven, A. Meyer, R. K. Koehn, and P. D. Rawson. 2000. Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration. Marine Biology 136:69-77.

Hill, A. A., C. P. Hunter, B. T. Tsung, G. Tucker-Kellogg, and E. L. Brown. 2000. Genomic analysis of gene expression in *C. elegans*. Science 290:809-812.

Hofmann, G. E., and G. N. Somero. 1996. Interspecific variation in thermal denaturation of proteins in the congeneric mussels *Mytilus trossulus* and *M. galloprovincialis*: Evidence from the heat shock response and protein ubiquitination. Marine Biology 126:65-75.

Hollocher, H., and A. R. Templeton. 1994. The molecular through ecological genetics of abnormal abdomen in *Drosophila mercatorum* .6. The nonneutrality of the Y-chromosome Rdna polymorphism. Genetics 136:1373-1384.

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.

Hughes, A. L. 2002. Natural selection and the diversification of vertebrate immune effectors. Immunological Reviews 190:161-168.

Jiggins, F. M., G. D. D. Hurst, and Z. H. Yang. 2002. Host-symbiont conflicts: Positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Molecular Biology and Evolution 19:1341-1349.

Kenchington, E., B. MacDonald, L. Q. Cao, D. Tsagkarakis, and E. Zouros. 2002. Genetics of mother-dependent sex ratio in blue mussels (*Mytilus* spp.) and implications for doubly uniparental inheritance of mitochondrial DNA. Genetics 161:1579-1588.

Lambert, C., N. Leonard, X. De Bolle, and E. Depiereux. 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18:1250-1256.

Marshall, J. L., M. L. Arnold, and D. J. Howard. 2002. Reinforcement: the road not taken. Trends in Ecology & Evolution 17:558-563.

Martz, E. 2001. "Protein Explorer Software", http://proteinexplorer.org.

McCartney, M. A., and H. A. Lessios. 2002. Quantitative analysis of gametic incompatibility between closely related species of neotropical sea urchins. Biological Bulletin 202:166-181.

McDonald, J. H., and R. K. Koehn. 1988. The mussels *Mytilus galloprovincialis* and *Mytilus trossulus* on the Pacific coast of North-America. Marine Biology 99:111-118.

Metz, E. C., R. E. Kane, H. Yanagimachi, and S. R. Palumbi. 1994. Fertilization between closely-related sea-urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. Biological Bulletin 187:23-34.

Miller, D. J., and M. J. H. van Oppen. 2003. A 'fair go' for coral hybridization. Molecular Ecology 12:805-807.

Moore, T., and H. D. Moore. 2002. Marsupial sperm pairing: a case of 'sticky' green beards? Trends in Ecology & Evolution 17:112-113.

Morrison, K. L., and G. A. Weiss. 2001. Combinatorial alanine-scanning. Current Opinion in Chemical Biology 5:302-307.

Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proceedings of the National Academy of Sciences of the United States of America 100:5268-5273.

Nielsen, K. M., J. Kasper, M. Choi, T. Bedford, K. Kristiansen, D. F. Wirth, S. K. Volkman, E. R. Lozovsky, and D. L. Hartl. 2003. Gene conversion as a source of nucleotide diversity in *Plasmodium falciparum*. Molecular Biology and Evolution 20:726-734.

Nosil, P., B. J. Crespi, and C. P. Sandoval. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature 417:440-443.

Orr, H. A. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics 149:2099-2104.

Palumbi, S. R. 1992. Marine speciation on a small planet. Trends in Ecology & Evolution 7:114-118.

Palumbi, S. R. 1994. Genetic-divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics 25:547-572.

Palumbi, S. R. 1998. Species formation and the evolution of gamete recognition loci. Pp. 271-278 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford University Press, Oxford.

Palumbi, S. R. 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. Proceedings of the National Academy of Sciences of the United States of America 96:12632-12637.

Palumbi, S. R., G. Grabowsky, T. Duda, L. Geyer, and N. Tachino. 1997. Speciation and population genetic structure in tropical Pacific Sea urchins. Evolution 51:1506-1517.

Peichel, C. L., K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, and D. M. Kingsley. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414:901-905.

Presgraves, D. C. 2003. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. Genetics 163:955-972.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423:715-719.

Queller, D. C., E. Ponte, S. Bozzaro, and J. E. Strassmann. 2003. Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. Science 299:105-106.

Rawson, P. D., V. Agrawal, and T. J. Hilbish. 1999. Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. Marine Biology 134:201-211.

Rawson, P. D., C. Slaughter, and P. O. Yund. 2003. Patterns of gamete incompatibility between the blue mussels *Mytilus edulis* and *M. trossulus*. Marine Biology In Press

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

Rice, W. R. 1998. Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. Pp. 261-270 *in* D. J. Howard and S. H. Berlocher, eds. Endless Forms: Species and Speciation. Oxford University Press, Oxford.

Riginos, C., and J. H. McDonald. 2003. Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the mussel genus *Mytilus*. Molecular Biology and Evolution 20:200-207.

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

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287:306-308.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Schneider, S., Roessli, D., and Excoffier, L. 2000. Arlequin: A software for population genetics data analysis.Ver 2.000.Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Suchanek, T. H., J. B. Geller, B. R. Kreiser, and J. B. Mitton. 1997. Zoogeographic distributions of the sibling species *Mytilus galloprovincialis* and *M. trossulus* (Bivalvia: Mytilidae) and their hybrids in the north Pacific. Biological Bulletin 193:187-194.

Swanson, W. J., C. F. Aquadro, and V. D. Vacquier. 2001a. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. Molecular Biology and Evolution 18:376-383.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001b. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 98:7375-7379.

Swanson, W. J., and V. D. Vacquier. 2002. Reproductive protein evolution. Annual Review of Ecology and Systematics 33:161-179.

Swanson, W. J., Z. H. Zhang, M. F. Wolfner, and C. F. Aquadro. 2001c. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences of the United States of America 98:2509-2514.

Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony * and other methods. Ver. 4. Sinauer, Sunderland, MA.

Takagi, T., A. Nakamura, R. Deguchi, and K. Kyozuka. 1994. Isolation, characterization, and primary structure of 3 major proteins obtained from *Mytilus edulis* sperm. Journal of Biochemistry 116:598-605.

Thiery, J. P. 2003. Cell adhesion in cancer. Comptes Rendus Physique 4:289-304.

Thompson, G. J., Y. C. Crozier, and R. H. Crozier. 2003. Isolation and characterization of a termite transferrin gene up-regulated on infection. Insect Molecular Biology 12:1-7.

Ting, C. T., S. C. Tsaur, M. L. Wu, and C. I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501-1504.

Togo, T., and M. Morisawa. 1997. Aminopeptidase-like protease released from oocytes affects oocyte surfaces and suppresses the acrosome reaction in establishment of polyspermy block in oocytes of the mussel *Mytilus edulis*. Developmental Biology 182:219-227.

Togo, T., K. Osanai, and M. Morisawa. 1995. Existence of three mechanisms for blocking polyspermy in oocytes of the mussel *Mytilus edulis*. Biological Bulletin 189:330-339.

Toro, J. E., R. J. Thompson, and D. J. Innes. 2002. Reproductive isolation and reproductive output in two sympatric mussel species (*Mytilus edulis, M. trossulus*) and their hybrids from Newfoundland. Marine Biology 141:897-909.

Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. Evolution 42:1085-1089.

Turelli, M., and H. A. Orr. 2000. Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154:1663-1679.

Volkman, S. K., D. L. Hartl, D. F. Wirth, K. M. Nielsen, M. Choi, S. Batalov, Y. Y. Zhou, D. Plouffe, K. G. Le Roch, R. Abagyan, and E. A. Winzeler. 2002. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. Science 298:216-218.

Walhout, A. J. M., and M. Vidal. 2001. Protein interaction maps for model organisms. Nature Reviews Molecular Cell Biology 2:55-62.

Wares, J. P., and C. W. Cunningham. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. Evolution 55:2455-2469.

Wilkemeyer, M. F., S.-y. Chen, C. E. Menkari, D. E. Brenneman, K. K. Sulik, and M. E. Charness. 2003. Differential effects of ethanol antagonism and neuroprotection in peptide fragment NAPVSIPQ prevention of ethanol-induced developmental toxicity. PNAS 100:8543-8548.

Yang, Z. H. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556.

Yang, Z. H., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Molecular Biology and Evolution 19:49-57.