INTROGRESSIVE HYBRIDIZATION BETWEEN NATIVE AND INVASIVE CRAYFISH: A STUDY IN THE BLACKSTONE RIVER VALLEY

A Major Qualifying Project Report

Submitted to the Faculty of the WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the Degree of Bachelor of Science

by

Mallory McMurrough

&

Laura Saltzman

Date: April 30, 2009

Approved:

Professor Lauren Mathews, Major Advisor
Abstract

The Blackstone Valley Watershed provides a unique opportunity to study the phylogenetic and phylogeographic relationships between two related crayfish species, *Orconectes virilis* and *Orconectes quinebaugensis*. Crayfish were collected from various sites in the watershed from March to October of 2008. Individuals were analyzed for the mitochondrial gene cytochrome oxidase I, a microsatellite allele (Ov54), and pleopod and chelae morphology. The COI mitochondrial gene was found to have two haplotypes: Ov, associated with individuals subjectively identified as *Orconectes virilis*, and Oq, associated with individuals identified as *Orconectes quinebaugensis*. The Ov54 nuclear microsatellite locus yielded three genotypes, two homozygous, AA (A) and BB (B), and the heterozygous AB. It was found that the Ov54 genotype is dependent on COI haplotype (p<0.05); Ov54 genotype A is significantly associated with haplotype Ov; and that the interaction between Ov54 genotype and pleopod angle is significant; genotype A is associated with a larger pleopod angle (mean=149.7°). Additionally, the interaction between COI haplotype and top projection pleopod angle is significant; the mean pleopod angle of Ov was 148.8°, while the mean pleopod angle of Oq was 145.7°. A logistic regression was performed using Ov54 genotype, top projection pleopod angle and chelae length/width ratio as predictive variables for COI haplotype. It was found that the top projection pleopod angle and Ov54 genotype are significant (p=0.05) predictors of COI haplotype. These findings indicate that the populations of *O. virilis* and *O. quinebaugensis* are in the process of introgressing.

Acknowledgements

The authors gladly thank our advisor, Professor Lauren Mathews, as well as Professor Michael Buckholt, Professor Elizabeth Ryder, and Professor Joseph Petruccelli for their time, dedication, patience, and guidance in this endeavor. In addition, we would like to thank Amy Warren for all of the hard work and time she dedicated to helping with this project.
Table of Contents

Introduction..................................................................................................................................................................4

Materials and Methods..........................................................................................................................................16

Results..................................................................................................................................................................21

Discussion..............................................................................................................................................................26

References...............................................................................................................................................................30

Appendix.................................................................................................................................................................41
1. Introduction

Recent experimentation and analysis has confirmed a sister relationship between *O. quinebaugensis* and *O. virilis*, supported by both morphological and genetic data (Mathews et al., 2008). In this study, we examine a sample population of crayfish from both species for evidence of hybridization occurring in the Blackstone Valley Watershed. Hybrids occur when individuals from genetically distinct populations interbreed, resulting in the combination of genetic material from previously isolated gene pools (Harrison, 1993; Rhymer & Simberloff, 1996; Schwenk et al., 2008). Interspecific hybridization in crayfish is of special concern because 30% of North American species are threatened or endangered, and exist in small ranges, rendering populations more susceptible to genetic extinction through hybridization and introgression (Perry et al., 2001; Lodge et al., 2000). This study uses genetic and morphological data to gain insights into the phylogenetic relationships among *O. virilis* and *O. quinebaugensis*, and explores the possibility of extinction of the latter species through introgressive hybridization and genetic assimilation.

1.1 The Species Problem

The species concept continues to be one of the most widely debated issues in speciation. The majority of biologists concur that distinct evolutionary lineages exist among sexually reproducing organisms, and conduct their research accordingly. However, evolutionists conflict in their willingness, or lack thereof, to define these groups as "species," and how best to define them. Mayr (1942) approached the species problem with the articulation of his "biological species concept," which defined species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups." Since then, the number of species concepts has increased dramatically, with new concepts emerging yearly (Claridge et al., 1997; Howard & Berlocher, 1998; Wheeler & Meier, 2000; Hey, 2001; Brookfield, 2002; Coyne & Orr, 2004; Wilkins, 2006). The inability to define the word "species" in a way that applies to all naturally occurring organisms is the root of the "species problem." In general, there appear to be two primary approaches to the delineation of taxa as species and within each of these approaches there exist several justifiable variations. The first approach clusters organisms by structural similarities while the second methodology groups organisms together by their phylogenetic relationships (Kitcher, 1984; Coyne & Orr, 2004). Kitcher (1984) describes the species category as "heterogeneous," in that no single definition can meet the need of all biologists or apply to all groups of organisms. Consequently, any single definition of "species" is unavoidably subjective, and biologists have instead proposed a range of definitions; the definition a biologist employs is a pragmatic choice, influenced by the particularities of that biologist's research.

One species concept of particular interest to this study is the Genotypic Cluster Species Concept (GCSC) which defines a species as a morphologically or genetically discernible group of individuals that contains few or no hybrids when in contact with similar clusters (Mallet, 1995). A significant feature of the genotypic cluster is that species can be affected by the processes of gene flow, natural selection and
history, rather than being defined by them; it can accommodate some gene flow between species (clusters) as long as their integrity remains such that they can be distinguished (see Mallet, 1995 for detailed explanation). Genotypic clusters are identified by gaps that occur between groups of multilocus genotypes within a local area (Mallet, 1995, 2001), similar to the way in which Darwin's (1859) morphological cluster species are identified by morphological gaps. The genotypic cluster definition is related to the taxonomic practice inherent in the polytypic species. A polytypic species has two or more subspecies, considered to be discrete groups, that are clearly distinct from one another and do not generally interbreed (although there may be a relatively narrow hybridization zone), but which might interbreed freely if given the chance to do so (Shaffer & McKnight, 1994; Allendorf & Leary, 1988; Mallet, 1995). Genotypic cluster species are very similar to the practical taxonomic application of the biological species concept (Mayr, 1963). Most likely, genotypic cluster species will remain distinct entities from one another due to reproductive or ecological traits, and will ultimately attain genetic separation over time. The genotypic cluster definition permits numerous means of cluster maintenance and evolution while investigating the genetic results of the combination of processes. Because studies of biological and evolutionary processes are not pre-requisites of the investigation of speciation under the GCSC model (Mallet, 1995), using this definition alleviates issues of classification in taxonomy and conservation that are inherent in species models based on idealized evolutionary or biological concepts. The genotypic cluster species concept has proved useful in numerous taxonomic and evolutionary studies, as well as those related to conservation and speciation (Avise & Ball, 1990; Lanzaro et al., 1990; Narang et al., 1989; Mallet, 1989, Szymura & Barton, 1986; Yanchukov et al., 2006).

1.2 Hybridization and Hybrid Zones

The definition of hybridization varies slightly among scientists, but generally speaking, natural hybridization occurs when individuals from two populations, distinguishable on the basis of one or more characters (morphological and/or genetic) interbreed and produce offspring (Woodruff, 1973; Bigelow, 1965; Short, 1969). Hybrid zones are generally considered to exist as narrow regions in which genetically distinct populations come into contact, breed and produce these hybrids (Barton & Hewitt, 1985). However, the definition of the term varies greatly among researchers. Endler (1977) defines hybrid zones as “narrow belts (clines) with greatly increased variability in fitness and morphology compared to that expected from random mixing, separating distinct groups of relatively uniform sets of populations,” while Gartside et al.’s (1979) description depicts “interactions between populations which have diverged genetically so that interbreeding between individuals from the divergent populations results in progeny recognizably different from the parents.” The phenomenon of hybrid zones occur in a wide range of species. Littlejohn & Watson (1985) used an empirical and pragmatic approach to investigate hybrid zones and interactions between Australian frogs (the Geocrinia laevis complex and the Litoria ewingi complex). Searle (1991) examined the structure of hybrid zones occurring in subspecies of the West European house mouse (Mus musculus domesticus), and discovered staggered chromosomal clines. Baker (1981) studied chromosomes in hybrid zones of Peters tent-making bats (Uoderma bilobatum) and concluded that hybridization has a prolonged effect which produces chromosomal polymorphisms over a wide geographic range. Hewitt (1975) studied narrow sex-
chromosome hybrid zones in the grasshopper (*Podisma pedestris*) in the Southern French Alps, and proposed that such chromosomal races may frequently occur in relatively immobile species that are subject to isolation and expansion through major climactic and ecological changes. Stable hybrid zones have also been documented in iguanid lizards (*Sceloporus woodi* and *S. undulates undulates*) in Florida (Jackson, 1973); toucans (*Pteroglossus torquatus torquatus* and *P. sanguineus*) in Columbia (Haffer, 1967); warblers (*Dendroica coronate* and *D. auduboni*) in Alberta and British Columbia (Hubbard, 1969); pocket gophers (*Thomomys bottae* and *T. townsendii*) in California (Thaeler, 1968; and Hooded and Carrion Crows (*Corvus corone* and *C. cornix*) in Central Europe (Mayr, 1963).

Studies of interspecific hybridization have served as a major focus for evolutionary biologists. Although the fields of taxonomy and systematics often construe the occurrence of hybridization to be problematic, it has become a primary source of data for studies related to speciation and adaptation. Hybrids occur when individuals from what are believed to be genetically distinct populations interbreed, resulting in the combination of genetic material from previously isolated gene pools and subsequently, new genotypes (Harrison, 1993; Rhymer & Simberloff, 1996; Schwenk et al., 2008). Effects from genetic/evolutionary processes, such as recombination and natural selection, usually develop over extended periods of time; however, they are accelerated through hybridization, causing rapid and long-lasting changes among interbreeding lineages. The evolutionary change due to hybridization can occur within one generation, thereby exposing new gene combinations to natural selection. Furthermore, by shortening the length of time in which evolutionary processes occur, and by allowing the formation of new gene combinations, interspecific hybridization enables examinations of the processes of natural selection. The significance of hybridization and introgression has long been acknowledged by botanists (Arnold, 1997; Anderson & Hubricht, 1938; Anderson, 1948; Rieseberg & Wendel, 1993; Rensch, 1959). These views toward the importance of hybridization in plant evolution were summarized by Anderson and Stebbins (1954) who hypothesized that introgression and hybridization could transfer blocks of genes among stabilized, adapted groups, permitting rapid reshuffling of varying adaptations and complex modifier systems, thereby dramatically increasing levels of variation and allowing selection to act upon isolating blocks of genetic material instead of one or two alleles generated by mutation. Since the process permits evolution to proceed at a “maximum rate,” examples of interspecific hybridization have been described as “natural laboratories for evolutionary studies” (Hewitt, 1988), “windows on evolutionary process” (Harrison, 1990), and “an ecologically dependent behavioral phenomenon with genetic consequences” (Grant & Grant, 2008).

### 1.2 Introgression and Reinforcement

The term “introgression” was first defined by Anderson and Hubricht (1938) as a permeation of germ plasm from one species into another as a direct result of repeated backcrossing of hybrids to the parental species. However, this definition has since proven to be inaccurate since it restricts introgression to gene flow between species. Introgression can correctly describe gene exchange between any set of differentiated populations systems including, but not limited to, species, subspecies, and races (Rieseberg & Wendel, 1993). Therefore, introgression can be more accurately defined as a
permanent infiltration of genes from one set of distinct populations into another, i.e., the absorption of foreign alleles into a new, reproductively integrated population (Rieseberg & Wendel, 1993). Introggression has been portrayed as a primary evolutionary consequence of reproduction involving semifertile hybrids in a hybrid zone (Hardig et al., 2000), and has been known to transpire frequently among plants (Anderson, 1949; Arnold, 1997; Rieseberg & Wendel, 1993). Arnold et al (1991) examined introgression and hybrid speciation in Louisiana irises (Iris fulva, I. hexagona, and I. nelsonii), and found evidence supporting their hypothesis that localized and dispersed introgression were largely attributed to pollen transfer. Hardig et al (2000) molecularly and morphologically analyzed data from individuals in a hybrid zone between two species of willows (Salix sericea Marshall and S. eriocephala Michaux), and found inter- and intraspecific chloroplast diversity within the zone indicative of historic introgression and modern hybridization.

Historically, the role of introgression in the evolution of animals has not been widely recognized since the proportion of successful progeny segregating from hybrids is much lower for animals than plants (Stebbins, 1996). However, researchers have begun to realize that introgressive hybridization may play a larger role in the evolution of animals than previously believed. The strongest cases for introgression combine analysis of nuclear protein genes and taxaspecific maternally inherited mitochondrial DNA variation (Rieseberg & Wendel, 1993). Roques et al (2001) integrated microsatellite DNA data and multivariate statistical procedures to explore the dynamics of introgressive hybridization between two redfish (Sebastes fasciatus and S. mentella) in the North-west Atlantic. This study provides a rare example of broadscale introgressive hybridization in the ocean, and emphasizes the principal role this process plays in the genetic diversity, interspecific variations, and population structure among North-west Atlantic redfish. This investigation also suggests that selection promotes and maintains the observed pattern of introgression. Similar studies were conducted in populations of teleosts (Solea senegalensis, S. aegyptiaca, and S. vulgaris) (She et al, 1987); morphs of the tropical marina damselfish Acanthochromis polyacanthus (Planes & Doherty, 1997); and rockfishes (Sebastes auriculatus, S. caurinus, and S. maliger) (Seeb, 1998). Lewontin & Birch (1966) experimented with the Australian fruit fly (Dacus tryoni) and its closely related species D. humeralis. They hypothesized that the remarkable range expansion of D. tryoni occurred by means of an adaptation to extreme temperatures, with genetic variation obtained through introgression with D. humeralis. This possibility was tested by maintaining pure and hybrid populations from D. tryoni and F1 hybrids between the two species at a series of optimal and extreme temperatures. Hybrid lineages increased more rapidly at higher temperatures, indicating that introgression of alleles allowed for adaptation to extreme warmth. The importance of introgressive hybridization has also been described in several avian studies (Cade, 1983; Grant & Grant, 1992, 1994; Short, 1965, 1972).

There is usually a long period during lineage divergence when hybridization can lead to introgression. Most of the reproductive discontinuity between sister species is a result of assortative mating, and, during introgression, important speciation processes, such as reinforcement, can occur (Butlin, 1987; Howard, 1993; Coyne & Orr, 1996; Mallet, 2005). Generally speaking, reinforcement is the process by which prezygotic isolating barriers evolve in contact/hybrid zones as a response to a natural selection against hybridization, and typically results in reproductive character displacement (Howard, 1993;
However, interpretations of the process of reinforcement vary among researchers. Butlin (1987, 1989) contends that reinforcement should be applied only to situations in which fertile, but less fit hybrid offspring are produced. He argued that reproductive character displacement should refer to a similar process that transpires among two populations between which gene flow is prevented by pre-existing postzygotic barriers that prohibit any gene flow. Thus, by this definition, reinforcement acts a mechanism of speciation, while reproductive character displacement involves pre-existing species. Dobzhansky (1940) originally presented the reinforcement hypothesis, which stated that prezygotic isolating mechanisms can be selectively strengthened, or reinforced, along the edges of a hybrid zone, preventing hybridization and thus the production of inferior or inviable offspring. While the hypothesis was initially accepted, many biologists eventually became skeptical of its validity. One of the most influential arguments against reinforcement involves what are perceived as alternative, and frequently more likely, outcomes of secondary contact between two populations. If two populations that differ in size encounter one another, individuals of the more rare population will hybridize more frequently than individuals of the more common population. If the smaller population is significantly smaller, parents will encounter, and thus mate with, individuals of the opposite species more frequently than they do with individuals of their own species. This situation would result in extremely high selection intensities for the evolution of prezygotic isolation barriers. However, if the population in question is sufficiently small or has a low growth rate, then a high degree of "wasteful" reproductive effort on hybrid offspring could cause the extinction of that species before reinforcement could raise isolation barriers (Liou and Price, 1994). Other objections, along with counter-arguments by supporters of reinforcement, are reviewed in detail by Howard (1993).

Further theoretical analysis (Liou & Price, 1994), taxonomic surveys (Coyne & Orr, 1989, 1997; Howard, 1993), and numerous empirical studies (Gerhardt, 1994; Noor, 1995; Ryan et al., 1996; Saetre et al., 1997; Rundle & Schluter, 1998) have provided strong evidence supporting the reinforcement hypothesis. Noor (1995) determined that females derived from populations of *Drosophila pseudoobscura* sympatric with *D. persimilis* demonstrate a greater disinclination to mating with heterospecific males than females derived from allopatric populations, consistent with reinforcement. *F*₁ hybrid males are sterile, but hybrid females are fertile, allowing for gene exchange and selection against mismating. Butlin (1995) suggested that hybrid fitness levels might be low enough to halt gene exchange in these species, however a genetic study conducted by Wang *et al.* (1997) later showed low levels of gene flow (not just hybridization) between these two species. Further evidence of reinforcement was documented in species of flycatchers. Saetre *et al.* (1997) observed that pied flycatchers (*Ficedula hypoleuca*) are black and white in populations allopatric to the similarly black and white collared flycatchers (*F. albicollis*). However, pied flycatcher populations sympatric with collared flycatchers are brown in color. This color change is consistent with female preferences in the populations. The colors observed in allopatric populations seemed to be ancestral, indicating that the divergence in sympathy is a direct result of reinforcement. While conducting mating studies of two threespine stickleback (*Gasterosteus aculeatus*) morphs, Rundle & Schluter (1998) found evidence of reproductive character displacement. Some matings were discovered in the sticklebacks' natural habitat, and *F*₁ hybrids were found to suffer a foraging disadvantage relative to their parents. Evidence
such as this leads one to conclude that reinforcement does occur in nature, and is consequently an important mechanism in the foration of isolation barriers between diverged populations.

1.5 Extinction by Hybridization and Introgression

Several reviews have highlighted the innovative role that hybridization may play in adaptive evolution and speciation (Abbott, 1992; Arnold, 1997; Rieseberg, 1997; Grant & Grant, 1998; Ramsey & Schemske, 1998). Numerous effects of introgressive hybridization have been proposed, including increased genetic diversity, the origin and transfer of adaptations, the origin of ecotypes or species, the breakdown or reinforcement of isolating barriers, colonization and dispersal (Anderson, 1949; Potts & Reid, 1988; Rieseberg & Wendel, 1993). However, hybridization and introgression can have a reducing effect on native flora and fauna when nonindigenous species are introduced to a population: genetic extinction. The primary forces facilitating the process of extinction through introgressive hybridization are the introduction of anthropogenic species and subspecies, and habitat modification (Rhymer & Simberloff, 1996). Introduced species are a major cause of ecological breakdown and the loss of biological diversity worldwide (Elton, 1958; Lodge, 1993; Rhymer & Simberloff, 1996). Hybridization with nonindigenous species reduces the distinctness of resident species, and if a native taxon is rare or endangered, lacks a competitive advantage, or possesses weak reproductive barriers, hybridization can virtually eliminate them by genetic mixing or assimilation (Cade, 1983; Rieseberg, 1991; Ellstrand, 1992; Wolf et al, 2001; Perry et al, 2001). For example, the hybridization of the gray wolf (Canis lupus) and the coyote (C. latrans) played a major role in the near extinction of the red wolf (C. rufus) in North America through the incorporation of gray wolf and coyote genes into the already small red wolf gene pool (Goldman, 1944; McCarley, 1962; Nowak, 1979; Rhymer & Simberloff, 1996). Goodman et al (1999) genetically analyzed the hybridization of the native red deer (Cervus elaphus) in Argyll, Scotland, and the introduced Japanese sika deer (C. nippon). Using microsatellite markers and mitochondrial DNA, they predicted that a complete genetic and phenotypic introgression will likely take place as contact time and hybridization increases, much like the introgression that occurred between the two species in nearby Wicklow, Ireland (Harrington, 1973, 1979, 1982). Riley et al (1999) investigated the potential for hybridization between the rare California tiger salamander (Ambystoma californiense), native to central California, and A. tigrinum, a congener which had been deliberately introduced as fish bait. Genetic analysis using mitochondrial DNA and two nuclear loci revealed the existence of viable and fertile hybrids at all six sample sites, but only one of the sites contained more than 8% possibly pure A. californiense animals, implicating that hybridization poses a grave threat to the genetic purity of the native species. Numerous such cases of species threatened by hybridization have been documented worldwide, ranging from plants (Spartina foliosa, Ayres et al, 1999), to mammals (Mustela putorius, Davison et al, 1999; Canis simensis, Gotelli et al, 1994), to amphibians (Rana perezi, Arano et al, 1995; ), to fishes (Salmo trutta, Almodovar et al, 2001 and Wolf et al, 2001; Onorhynchus clarki, Allendorf & Waples, 1996; Coregonus lavaretus, Wilson & Brown, 1953; Hypomesus transpacificus, Trenham et al, 1998). Additional empirical cases are described by Rhymer & Simberloff (1996).
Hybridization and introgression with non-indigenous species seems to be widespread in freshwater ecosystems (Perry et al, 2001). More specifically, the effects of non-indigenous species, including hybridization, are often associated with the decline of crayfish populations. In 1989, Hobbs et al. summarized global crayfish introductions in a list of 20 crayfish species that had been introduced into new river drainages, states or continents and provided a detailed history of deliberate and accidental introductions. The number of introductions has undoubtedly increased since then. As a result, populations of native crayfishes have decreased dramatically over the last century, particularly in Europe (Lodge et al, 2000). However, North America contains over 75% of the world’s crayfish (30% of which are threatened or endangered) which exist in small ranges, making populations more susceptible to genetic extinction through hybridization and introgression (Perry et al, 2001; Lodge et al, 2000). A study by Perry et al (2001) was the first to genetically document hybridization between a resident and invading crayfish. They investigated the displacement of two resident crayfish species in northern Wisconsin, Orconectes propinquus and O. virilis by the introduced species O. rusticus (native to southwestern Ohio). While O. virilis does not appear to be hybridizing with either of the other two taxa, evidence indicates that hybridization and introgression are occurring between O. rusticus and O. propinquus, largely contributing to the displacement of the latter. Just as it did in Ohio (Jezerinac et al, 1995) and Illinois (Taylor and Redmer, 1996), the establishment of O. rusticus in the lakes of Northern Wisconsin reduced or eliminated O. virilis and O. propinquus within only a few years (Lodge et al, 1986; Olsen et al, 1991). O. virilis used to be the only common crayfish in the lakes and streams of northern Wisconsin and the upper peninsula of Michigan (Creaser, 1932). Now the taxon exists in only 44% of the lakes and 38% of the streams in this region (of 107 and 50 sites sampled, respectively) (Lodge et al, 2000). Wolf et al (2001) warns of this possibility, employing a simulation model to assess the risk of extinction through hybridization, suggesting that not only is hybridization a significant threat to both scarce and profuse species, but that it can lead to extinction in an exceptionally short period of time.

1.6 Phylogenetic and Morphological Characters of Orconectes virilis & Orconectes quinebaugensis

Although morphology-based taxonomy is vital to research in the field of biodiversity, nature’s complexity reaches far beyond what the human eye can perceive. Hidden, or cryptic, species are two or more distinct species incorrectly classified as a single species due to their extreme morphological similarities and the inability of biologists to differentiate between them (Beheregaray and Caccone, 2007; Bickford et al., 2007). Since speciation is not always associated with morphological change, the true number of biological species is liable to be much larger than the current number of ostensible species, the majority of which are defined solely on a morphological basis (Bickford et al., 2007). Although the concept has existed for hundreds of years, research of cryptic species has increased dramatically over the past two decades (Figure 1) (Bickford et al., 2007; Pfenninger and Schwenk, 2007). Innovations in the form of DNA sequencing through PCR, and its increasing availability, have given researchers the necessary tools to accurately and inexpensively identify and distinguish between these sibling species. The frequency with which cryptic species are revealed by means of DNA sequencing suggests that molecular analysis should routinely be incorporated in the research of alpha taxonomists in order to more accurately document biodiversity. Several recent reviews of the subject have
suggested that the distribution of cryptic species is non-random across taxa and biomes, which might crucially affect biodiversity assessments, macroecology, biogeography, and evolutionary theory (Bickford et al., 2007; Pfenninger and Schwenk, 2007; Beheregaray and Caccone, 2007).

![Graph showing the percent of publications in Zoological Record Plus (CSA) that refer to 'cryptic species' (circles) or 'sibling species' (triangles) in the title, abstract, or keywords has risen exponentially since the introduction of PCR (from Bickford et al., 2007).]

Regarded as a keystone species in the ecological communities in which they live, the existence of more than 540 species of crayfish has been documented in the world. These crustaceans, belonging to the Decapoda taxon, are divided among three families: Parastacidae, Astacidae, and Cambaridae. The freshwater crayfish family Cambaridae is the most diverse among the three currently recognized families. Cambaridae is comprised of nearly 390 species and subspecies among 12 genera and represents over 75% of all crayfish (Taylor et al., 2002). Approximately ninety-nine percent of Cambarid species are endemic to Central and North America, while the remaining one percent exists in Asia (Hobbs, 1988). Three predominant genera comprise North America’s freshwater crayfish population: Procambarus, Cambarus, and Orconectes.

*Orconectes* represents the third most diverse cambarid genus, with 11 subgenera, 81 species and 11 subspecies primarily occurring east of the continental divide (Hamr, 2002). Universally known as the virile crayfish, the species *Orconectes virilis* has a widespread distribution throughout North America, with populations ranging from the Canadian provinces of Quebec and Ontario, to the Midwestern states of Illinois and Kansas, and even as far south as Texas (Figure 2) (Hamr, 2002). The spread of *O. virilis* beyond its native habitat has been fostered by its popularity as food and baitfish, and the species is now widely considered invasive in many regions, including New England (The Global Invasive Species Database, 2005; Hamr, 2002.)
Crayfish of this species are primarily found in lakes, streams, rivers and ponds, and require shelter in the form of burrows, rocks, logs or heavy vegetation as safeguards against the threats of predation. During the cold months, *O. virilis* crayfish migrate into deeper water in order to survive the harsh winter conditions, and are thus dependent upon permanent bodies of water that do not freeze to the substrate at any point in time (Hamr, 2002). They have strong annual patterns of reproduction, with mating occurring in late summer and fall, and the laying of eggs taking place in the spring. Mature males oscillate between two morphological forms: form I (breeding) and form II (non-breeding). Males typically molt from form II to form I prior to the mating season, and return to form II at the conclusion of the season (Hobbs, 1991). Form I males are recognized by their large chelae and the calcified central projections of their gonopods, which are inserted into the female’s annulus during reproduction. The gonopods are the first pair of pleopods in male crayfish that are modified in sexual maturity, becoming relatively long, straight and thin (Hamr, 2002, Mathews et al., 2008). Female sexual maturity is determined by the existence of glair glands in uropods as well as eggs affixed to the pleopods. The shape of the male gonopods and female annulus ventralis were initially considered to be species specific, rendering hybridization in most species impossible (Hamr, 2002). However, as this study shows, the phenomenon of hybridization is more prevalent among freshwater crayfish populations than originally thought. Figure 3 displays the general anatomy of a male crayfish.
Figure 3. Dorsal and ventral views of general male crayfish (Hobbs, 1972)

Figure 4. Dorsal and ventral views of Orconectes virilis
Figure 5. Dorsal and ventral views of *Orconectes quinebaugensis*.

*O. virilis* is a medium to large crayfish, typically ranging from 10 to 12 centimeters in size, excluding the chelipeds and long antennae. The head and carapace of this species are often olive or reddish brown in color with dark brown spots dotting the midline of the abdominal segments. They possess yellow tubercles while their chelae and legs display a uniquely bluish tint (Figure 4) (Hamr, 2002; The Global Invasive Species Database, 2005). The average lifespan of *O. virilis* is approximately 3-4 years (Hamr, 2002; Weagle & Osburn, 1970).

*Orconectes quinebaugensis* (Figure 5) is a newly discovered species of crayfish found in southern New England. The species is named for the Quinebaug River which flows from eastern Connecticut through south-central Massachusetts, including the town of Sturbridge, Massachusetts which is the type locality for the species (Mathews et al., 2008). Its watershed encompasses approximately 850 square miles and extends into the state of Rhode Island (United States Geographical Society, 2007). *O. quinebaugensis* is typically found in small streams with rocky substrate and shallow ponds with muddy substrates, living in conjunction with or without *O. virilis* and other crayfish species. Mating season occurs between the months of August and December, during which form I males and sexually mature females with developed glair glands are prevalent. Eggs are fertilized and laid most frequently in April and May, followed by the summer months in which juveniles are predominantly seen (Mathews and Warren, 2008).

Recent experimentation and analysis has confirmed a sister relationship between *O. quinebaugensis* and *O. virilis*, supported by both morphological and genetic data. While these two species are the most closely related in the complex and morphologically similar in many ways (Figures 4 and 5), they strongly differ in the Form I male first pleopod morphology as well as in the shape of the dactyl of the chelipeds (Mathews et al., 2008). In *O. virilis*, both the central projection and mesial process are relatively straight while those of *O. quinebaugensis* are recurved to a roughly equal extent and account for a lesser percentage of the first pleopod’s overall size (Figure 6). The degree of curvature of the dactyl is much greater in *O. quinebaugensis*, and especially well-defined in Form I males (Mathews et al., 2008). The implementation of molecular genetics tools has revealed that the sister taxa are distinctly different at both mitochondrial and nuclear loci (Mathews et al., 2008). A useful tool in distinguishing *O.*
The occurrence of a single nucleotide insertion in the mitochondrial 16S rRNA gene. This insertion is entirely unique to *O. quinebaugensis* (Mathews *et al.*, 2008).

**Figure 6.** *Orconectes quinebaugensis*. A, Dorsal view of carapace; B, Dorsal view of right chela and carpus; C, Ischium of third pereopod; D, Lateral view of first pleopod of Form I male; E, Mesial view of first pleopod of Form I male; F, Lateral view of first pleopod of Form II male; G, Mesial view of first pleopod of Form II male; H, annulus ventralis; I, Lateral view of carapace (Mathews & Warren., 2008).

Targeting *O. virilis* and its sister taxon *O. quinebaugensis* for a more specific genetic and morphological investigation will allow for significant insights into the phylogenetic relationships between these two lineages, and provide vital information useful in distinguishing between them. A thorough examination of their respective genotypic and morphological characters can offer insight into the processes surrounding the two populations, including hybridization and introgression. These processes can have a profound effect on native crayfish populations, and can threaten the survival of the taxa as well as local and worldwide biodiversity.
2. Materials and Methods

2.1 Collection and Identification of Population Samples

We initially analyzed a total of 505 individuals, collected from 24 sites (23 Massachusetts sites, 1 Rhode Island site) in the Blackstone Valley watershed (Figure 7). Of the 505, there were 123 form I males from 18 of the 23 sites. Live samples of *O. virilis* and *O. quinebaugensis* were collected utilizing several methods: hand collection, dip nets, seines, and crayfish traps baited with raw chicken or salmon. The latitude, longitude, elevation, and total number of samples collected at each site were recorded. The crayfish were then transported live to the WPI biological laboratory in ~20 L water-filled buckets, where they were separated according to sex and morphological form. Females were identified by the annulus ventralis, while Form I males were distinguished from Form II males by the first pair of pleopods that are modified during sexual maturity. Specimens were then placed in conical tubes and labeled with the sex and/or morphological form, collection date and site number prior to being sacrificed and stored at -80°F for future DNA analysis.

![Figure 7. Map of collection sites in the Blackstone Valley watershed](image)
2.2 DNA Extraction

We followed the Solid Tissue Protocol in Gentra's Puregene: Genomic DNA Purification Kit (Qiagen, 2007). We extracted tissue from a detached leg of each specimen; approximately 5-10 mg of tissue was placed in 150 μL of Cell Lysis Solution along with 1.5 μL of Puregene Proteinase K. The samples were mixed by inverting and incubated at 55°C overnight. The Cell Lysis Solution disrupts the cell membranes, causing the cellular components to be released. The Puregene Proteinase K solution digests proteins, and removes inhibitors and contaminates that would otherwise affect the purity and quality of the DNA. Then 100 μL of Protein Precipitation Solution was added and vortexed at high speed for ~20 seconds. The sample was then centrifuged at 13,000 x g for 6 minutes in order to precipitate and form a pellet. Next, the supernatant was slowly added to 150 μL of isopropanol and mixed by inverting gently ~50 times to cause the DNA to precipitate. The samples were then centrifuged at 13,000 x g for 5 minutes to form a pellet. The supernatant was then discarded carefully to avoid removing the pellet. The pellet was gently washed with 150 μL of 70% ethanol and centrifuged again for 5 minutes at 13,000 x g to form another pellet. The supernatant was discarded and the pellet was allowed to dry for ~15 minutes. The pellet was re-suspended in 40 μL of DNA Hydration Solution by vortexing. The solution was incubated at 65°C overnight, then placed at room temperature to dissolve the DNA, and later stored at -20°C until needed. The quantity of the DNA was then checked using a 1% agarose gel electrophoresis, and dilutions were made to ~10 ng/μL for use in the polymerase chain reaction (PCR). When the DNA concentrations were too low or not present, we performed a second extraction for those samples and followed this protocol again. The quantity of DNA was estimated by running a comparison to 100 ng/μL of lambda bacterial DNA on the same gel.

2.2 Polymerase Chain Reaction (PCR)

The mitochondrial gene cytochrome oxidase I was amplified for each sample using the following primers: OrcoCOIF (5′- GTGTTAGTTACAGCCATGC-3′), OrcoCOIR (5′- CCAGACTCTGAACTACAAT-3′), OrcoCOIF3 (5′- AATGTGGTAGTTACAGCTC-3′) and OrcoCOIR2 (5′- GCCTCTTTTTACCAGACTC-3′). The following components were added to each of the 96 wells in the plates, for a total reaction volume of 20 μL per well: 2 μL 10x reaction buffer, 1.2 μL dNTPs (2.5 mM each nucleotide), 0.8 μL of each the fluorescently labeled forward primer (10 μM OrcoCOIF or OrcoCOIF3) and the reverse primer (10 μM OrcoCOIR or OrcoCOIR2), 0.1 μL Taq polymerase enzyme, 15 μL distilled water, and 1 μL of DNA. The PCR amplifications were performed in a thermocycler, using the program comprised of the following conditions: 40 cycles of 30 s at 95°C, 30 s at the annealing temperature of 52°C, and 60 s at 72°C, following an initial denaturation step of 2 min at 95°C. This program was followed by a final extension of 10 min at 72°C. A 2.0% agarose gel was run with the samples and negative controls, alongside a 100bp hyperladder as the molecular weight standard, in order to verify the presence of ample DNA in each reaction. The remaining DNA was frozen at -20°C.
2.3 DNA Sequencing

In order to sequence both strands of DNA, 2 wells had to be sequenced for each successful PCR well. Thus, 2 separate Microamp plates were used (forward and reverse) distinguished by the primer utilized in each plate (Orco COIF or OrcoCOIF3 and Orco COIR or OrcoCOIR2, respectively). The following components were combined for a total reaction volume of 10μL in each of the 96 wells: 6.15μL distilled water, 2μL 5x Buffer, 1μL primer (3.3μM Orco COIF or 3.3μM Orco COIR), 0.35μL Big Dye, and 1.5μL DNA. Plates were then placed in the thermocycler and run using the Mathews SEQ Program (4+ hours), comprised of 40 of the following cycles: 96°C for 5 seconds, 50°C for 5 seconds and 60°C for 4 minutes, and after the 40 cycles were complete, the cycler was run at 15°C for 15 minutes, and 4°C until removed. An ethanol cleanup of sequencing reactions (96-well plate format) was subsequently performed, and the products were then sent to the Biotechnology Resource Center at Cornell University for automated sequencing. Upon their return, sequences were assembled and aligned using the Seq-Man module of Lasergene v. 6.1, and alignments were manually refined in Bioedit 7.0.4.1.

2.5 Nuclear Microsatellite Data

Genotype data for a nuclear microsatellite-containing locus, Ov54, were obtained as part of another study and were provided to us (Mathews, personal communication. The locus yielded two alleles in the total data set that were ~100 base pairs different in size, and contained a (CA)_n repeat. The sequence of the two products was verified in 2-4 individuals for each allele in order to confirm that the two band sizes represented alleles of the same locus, and not a spurious PCR amplicon.

2.6 Morphological Measurements

Morphological measurements of the chelae and carapace of all Form I males were taken using calipers, and recorded to the nearest 0.1 mm. The chelae were measured for length and width, and the two measurements were divided to create a length/width ratio for each individual. Length was measured out to the end of the immovable finger, and height was measured at the widest part of the palm (Figure 8a). Pleopods from the right side of the individuals were collected and digitally photographed. The angles of the top and bottom projections (Figure 8, b and c) were measured using the TpsDIG2 digital measurement module (Rohlf, 2005).
2.7 Data Analysis

Two sets of statistical analyses were performed, the first using SPSS v17.0 and Microsoft Excel 2007, and the second using SAS v9.1.2. For each Form I individual, a set of data was collected, comprising carapace length, COI haplotype, Ov54 genotype, top projection pleopod angle, and chela length/width ratio. Individuals missing any of the 5 measurements (such as those lacking both claws and/or both pleopods) were eliminated, leaving a total sample group of 123 individuals. Each of the variables (carapace length excluded) was statistically compared with each of the remaining variables, so that all combinations of variables were examined. The overall Ov54 genotypes and the overall COI haplotypes of the sample group were compared for independence using a standard chi square test in Microsoft Excel. Following the initial chi square test, all combinations of COI haplotype and Ov54 genotypes A, AB, and B were tested for independence using individual 2x2 chi square tests, with a Bonferroni correction. Additionally, the frequency of Ov54 alleles A and B were compared to COI haplotype using a 2x2 chi square test.

In order to control for the effects of size on the top projection pleopod angle and chela length/width ratio, carapace length was statistically correlated in SPSS with each of the factors using bivariate correlation tests. Once the effects of size were determined, the Ov54 genotypes were compared individually with the top projection pleopod angles and chela length/width ratios using one-way ANOVAs in SPSS. Tukey’s post-hoc testing was performed following each ANOVA to isolate any dependent variables present. The COI haplotypes were compared with the top projection pleopod angles and chela length/width ratios using independent samples t-tests in SPSS. A bivariate correlation test in SPSS was used to statistically compare the top projection pleopod angles with the chelae.
length/width ratios. All variables were checked for normality and equality of variance prior to performing statistical analyses, and no violations were found.

In order to determine which variables were predictive of COI haplotype, and to develop a predictive model for future use, a logistic regression was performed using SAS. The logistic regression was performed using Ov54 genotype, top projection pleopod angle and chelae length/width ratio as predictive variables for COI haplotype. To account for the three possible variables (A, AB, and B) present in the Ov54 genotype data set, a simple contrast was performed. “Genotype Indicator 1” contrasted genotypes A and B, while “Genotype Indicator 2” contrasted genotypes AB and B. This allowed the model to account for predictive value imparted by any of the three variables.

A Hardy-Weinberg test of equilibrium was performed on the Ov54 genotype data set in Microsoft Excel using methods described in Futuyma (1998). Values for $p^2$, $2pq$ and $q^2$ were calculated using frequencies of A and B in the sample group. Expected values were calculated using the values for $p^2$, $2pq$ and $q^2$, and were compared to the observed values using a standard chi square test of independence.
3. Results

3.1 COI Haplotype and Ov54 Genotype Analysis

The COI mitochondrial gene was found to have two haplotypes: Ov, associated with individuals visually identified as *Orconectes virilis*, and Oq, associated with individuals visually identified as *Orconectes quinebaugensis*. The Ov54 nuclear microsatellite locus was found to have two distinct alleles, A and B, found both homozygously as AA (A) and BB (B), and heterozygously as AB. A chi squared analysis of COI haplotype and Ov54 genotype was significant ($\chi^2 = 7.58; \text{df}=2; p<0.025$). This means that Ov54 genotype is dependent on COI haplotype. Further analysis was performed, and it was found that there was no significant difference between the genotypes A and AB ($\chi^2 = 2.26; \text{df}=1$), and genotypes B and AB ($\chi^2 = 0.24; \text{df}=1$). This led us to conclude that genotype AB is not significantly associated with either COI haplotype. Genotypes A and B were compared, and it was found that there is a significant difference in association between A and B ($\chi^2 = 6.33; \text{df}=1; p<0.017$). Genotype A is significantly associated with COI haplotype Ov, and genotype B is significantly associated with COI haplotype Oq (figure 9a).

The frequency of the Ov54 alleles A and B was calculated, and compared to COI haplotype. It was found that the frequency of Ov54 allele is dependent on COI haplotype ($\chi^2 = 11.2; \text{df}=1; p<0.001$). Ov54 allele A is significantly associated with haplotype Ov and Ov54 allele B is significantly associated with haplotype Oq (figure 9b).
Figure 9: (a.) Ov54 genotype is dependent on COI haplotype (p<0.025); Ov54 genotype A is significantly associated with haplotype Ov and Ov54 genotype B is significantly associated with haplotype Oq. (b.) Frequency of Ov54 allele is dependent on COI haplotype (p<0.001), Ov54 allele A is significantly associated with haplotype Ov and Ov54 allele B is significantly associated with haplotype Oq.

3.2 Top Projection Pleopod Angle and Chelae Length/Width Ratio

To control for size effects, top projection pleopod angle and chelae length/width ratio were each examined for correlation with carapace length. Bivariate correlation tests showed that there was no significant correlation between carapace length and either top projection pleopod angle and chelae length/width ratio (p=0.148, and p=0.227, respectively; see appendix C, parts 2 & 3). We are able to definitively say that neither pleopod angle or chelae ratio are affected by the size of the individual the measurements were obtained from in our sample, and such, carapace length was not included as a factor in any subsequent analyses.

Once it was determined that size was not a confounding factor, the Ov54 genotype was compared with the top projection pleopod angle and chelae length/width ratio. The interaction between the Ov54 genotype and top projection pleopod angle was found to be significant (for statistical tables, see appendix C, part 4). Through post-hoc testing, it was found that Ov54 genotype A is significantly different from AB and B (p=0.039 and 0.033, respectively); and that genotype A is significantly associated with a larger pleopod angle (mean=149.7, figure 10).
The interaction between Ov54 genotype and pleopod angle is significant; Ov54 genotype A is significantly different from AB and B; A is associated with a larger pleopod angle (mean=149.7°). Error bars represent +/- 1 standard error.

The interaction between Ov54 genotype and chelae length/width ratio was found to not be significant (p=0.826). Chelae ratio of an individual is not dependent on its Ov54 genotype (for statistical tables, see appendix C, part 5).

The COI haplotype were compared with the top projection pleopod angle, and the interaction was found to be significant (p=0.011, statistical tables in appendix C, part 6). The mean pleopod angle for the Ov group was found to be 148.8°, while the mean pleopod angle of the Oq group was 145.7° (figure 11). This shows that Ov individuals are significantly associated with straighter pleopods (refer to Figure 8a in Methods section), while Oq individuals are associated with more curved pleopods (refer to Figure 8b in Methods section).
Figure 11: The interaction between COI haplotype and top projection pleopod angle is significant; the mean pleopod angle of Ov was 148.8°, while the mean pleopod angle of Oq was 145.7°.

The interaction between COI haplotype and chelae length/width ratio was compared, and it was found to be not significant (p=0.132). Chelae ratio of an individual is not dependent on its COI haplotype (for statistical tables, see appendix C, part 7).

The correlation between top projection pleopod angle and chelae length/width ratio was tested, and found to be not significant (p=0.571, statistical tables and graphs in appendix C, part 8). There is no correlation between pleopod angle and chelae ratio, and further, after performing several statistical tests, it can be said that chelae ratio is not related to any of the other variables examined.

3.3 Logistic Regression Model

A logistic regression model was created using Ov54 genotype, top projection pleopod angle and chelae length/width ratio as predictive variables for COI haplotype. The analysis yielded a significant model, as follows:
\[
\text{logit} \left( \frac{\text{COI haplotype } Oq}{\text{COI haplotype } Ov} \right) = 16.1287 - 0.7250 \left( \text{genotype indicator } 1 \right) + 0.1303 \left( \text{genotype indicator } 2 \right) - 0.0912 \left( \text{top projection pleopod angle} \right) - 9.4808 \left( \text{chelae l/w ratio} \right)
\]

By inserting in the appropriate values for each of the variables (see note and example in appendix C, part 9), one can calculate the log of the odds as to whether an individual is COI haplotype Oq (or 1-log of the odds for haplotype Ov). The logistic regression also determines which variables have significant predictive value. Each variable adds or subtracts a certain value from the final result, determining what the log of the odds of COI haplotype is. For example, the variable “chelae length/width ratio” is a number from 0-1. If the chela ratio was 0.44, plugging it into the “chelae l/w ratio” variable would yield -4.17. The magnitude to which a variable changes the final “log of the odds” is determined by how significantly predictive the variable is.

It was found that top projection pleopod angle and Ov54 genotype A are significant (p= 0.0501 and p=0.054, respectively) predictors of COI haplotype. This is in line with the previous analyses, which showed that Ov54 genotype A is significantly associated with COI haplotype Ov, and that the size of the top projection pleopod angle is significantly associated with Ov54 genotype A and COI haplotype. Ov54 genotypes B and AB were found to not be significant predictors of COI haplotype (p=0.72), as was chelae length/width ratio (p=0.124).

3.4 Hardy-Weinberg Equilibrium

Genotype distributions of Ov54 were examined for deviations from Hardy-Weinberg equilibrium for both our data and data from a separate project group conducting a related study (Becker, Mulhern & Perry, 2009: unpublished; see appendix D for sample calculation). It was found, through chi square analysis, that both were in extreme Hardy-Weinberg disequilibrium (our data set: \( \chi^2 = 30.4; \text{df}=1; p<0.0001 \); Becker et al.: \( \chi^2 = 98.8; \text{df}=1; p<0.0001 \)). Both of these indicate significant deviation from HWE, but the difference in chi-square values shows that the related study’s sample group is in greater disequilibrium.
4. Discussion

This project posed a unique problem: morphologically and genetically distinguishing between two nearly identical species, *Orconectes virilis* and *Orconectes quinebaugensis*. Through careful visual examination prior to the beginning of our project, we believed that the two species may differ morphologically in the curvature of the pleopod (figure 8a and 8b). This difference, while slight, allowed us to visually distinguish between those individuals we believed to be *O. virilis*, which appeared to have a straighter pleopod, and those individuals we believed to be *O. quinebaugensis*, which seemed to have a curved pleopod. This difference in pleopod angles was only a subjective observation at first. While examining dozens of Form I individuals, we found what appeared to be a range of pleopod angles, making it difficult to assign those individuals with intermediate pleopods to either species. We considered this as possible evidence indicative of hybridization between the two species, and we began to search for an objective, statistically testable way to distinguish between *O. virilis* and *O. quinebaugensis*.

Hybridization between sister taxa has been well documented in many species (Jackson, 1973; Hubbard, 1969; Haffer, 1967; Thaeler, 1968). Continued visual examination of individuals from various sites in the Blackstone River Valley yielded a full range of pleopod angles at every site where we believed both species to be present (personal observations). The cytochrome oxidase I (COI) mitochondrial locus of these individuals was examined using PCR, and was found to have two haplotypes: Ov and Oq. These haplotypes were also significantly associated with distinctive pleopod angles: the mean pleopod angle of Ov was 148.8°, while the mean pleopod angle of Oq was 145.7°. The dimorphism found at this locus, combined with pleopod angle data, suggested that these two taxa were, at one point, distinct species.

Analysis of the Ov54 nuclear microsatellite containing locus provided additional evidence that *O. virilis* and *O. quinebaugensis* once existed as two distinct species. The Ov54 nuclear microsatellite locus yielded three genotypes found both homozygously as AA (A) and BB (B), and heterozygously as AB. When compared to the COI haplotypes, it was found that Ov54 genotype A is significantly associated with haplotype Ov while Ov54 genotype B is associated with haplotype Oq. Our data were enriched by a simultaneous related study that examined the same variables of individuals from two sites believed to contain pure populations of *O. virilis* and *O. quinebaugensis* (R1 and M2, respectively) as those in our study: Ov54 microsatellite, COI haplotype, and morphology (Becker et al., 2009). Through the examination of the COI haplotypes and Ov54 genotypes, Becker et al. (2009) determined that the M2 population was not genetically pure. Mating experimentation by Becker et al. (2009) found that individuals from their study sites were significantly less likely to mate interspecifically, and this, in combination with their extremely dimorphic morphological data, led us to believe that hybridization was occurring less frequently in these two sites than in our sites in the Blackstone River Valley Watershed. Since sites containing genetically pure populations of *O. quinebaugensis* have not been found, and may not exist, we can instead consider sites R1 and M2 as “proxy-pure” sites. This, in turn, provides us with a genetic control for our largely hybridized sites. Taking this into consideration, the related study found that Ov54 genotype A is significantly associated with the proxy-pure *O. virilis* site (R1), and that Ov54 genotype B is significantly associated with the proxy-pure *O. quinebaugensis* site (M2). In addition, the study found very little evidence of genetic mixing in comparison to the populations examined in our
Our investigation determined that the haplotype Ov group comprised 26% Ov54 genotype B and 23% genotype AB, while the haplotype Oq group contained 21% Ov54 genotype A and 29% genotype AB. This combination of data leads us to believe that Ov54 genotype A and Ov54 genotype B were once fixed in O. virilis and O. quinebaugensis, respectively.

The Hardy-Weinberg Equilibrium (HWE) states that allele frequencies and genotype frequencies will remain at constant values in ideal populations only if the populations adhere to several assumptions, in which the population is infinite in size, sexually reproducing, randomly mating, not under selection, and consists of diploid individuals carrying 2N copies of each locus (Halliburton, 2004). HWE proposes that the population is in genetic equilibrium. If all of these assumptions are met, the allele and genotype frequencies will not change over time. The allelic and genotypic frequencies will remain the same from generation to generation until evolutionary forces, e.g. mutation, natural selection, and gene flow (migration), disturb this idealized state and cause deviations. These deviations can sway the allele and genotype frequencies within the populations. In testing for heterozygosity, HWE can reveal the presence of heterozygote excess (homozygote deficit) which occurs when there are fewer homozygotes than expected under HWE. Heterozygote deficiency is often caused by strong inbreeding or selection for or against particular alleles. Our microsatellite data indicate a significant heterozygote deficiency in the Ov54 locus relative to the expectations of Hardy-Weinberg equilibrium (p <<< 0.001). This deviation could be explained as a result of assortative mating. This study presents a case in which two lineages were, for a period of time, diverging, but have since come back into contact, halting the divergence. The previous incomplete reproductive isolation likely resulted in populations strongly biased toward homozygotes, AA (O. virilis) and BB (O. quinebaugensis). Hybridization must occur in order to produce heterozygotes; thus, if the two lineages are reproductively compatible, the population would eventually reach HWE. Becker et al. (2009) found that while the two lineages may be less likely to interbreed, they are apparently reproductively compatible. We believe that genotype AB is underrepresented because O. virilis and O. quinebaugensis are in the process of introgressing.

**Historical Context**

Recent experimentation and analysis has confirmed a sister relationship between O. quinebaugensis and O. virilis (Mathews et al., 2008). It is apparent that the two lineages descended from a common ancestor; however it is unclear when this divergence occurred. Using the assumption of a molecular clock, based on COI haplotype and 16s rRNA genes, Mathews (2008) found that O. virilis and O. quinebaugensis (then O. species novel A) diverged from a common ancestor approximately 1.8 to 2 million years ago. At this point, we believe that some major event physically isolated two populations of this common ancestor, with one population isolated to the upper Midwest of the United States, and the other isolated to New England. The most likely cause of this physical isolation was the descent of glaciers across North America, physically cutting off New England from the Midwest for much of the Pleistocene era. Previous investigations have lent support for the role of the Pleistocene glacial period in species diversification in North America (reviewed in Mathews, 2008).
During this roughly 2 million year period of physical isolation, the two groups of the ancestral species evolved to a state of incomplete reproductive isolation. This was most likely achieved through the changes in pleopod morphology currently observed in *O. virilis* and *O. quinebaugensis*. Previous research has suggested that male gonopod morphology is likely to be correlated with speciation, and that changes in gonopod morphology may affect divergence rates by reducing reproductive compatibility (Mathews et al., 2008). We believe that these two isolated populations, which we now know as *O. virilis* and *O. quinebaugensis*, evolved changes in their gonopod morphology over this period of time that reduced their reproductive compatibility with one another. This theory is bolstered by the data described by Becker et al. (2009); where their mating studies indicated that, given equal opportunity, the two species are significantly less likely to mate interspecifically, but are still capable of forming hybrids.

Additionally, we believe that during this period of isolation, the Ov54 microsatellite genotype A was fixed or nearly fixed in populations of *O. virilis*, and likewise, Ov54 genotype B was fixed or nearly fixed in populations of *O. quinebaugensis*. A comparison of the individuals in our study’s population to those of Becker et al. (2009) shows the same significant association between Ov54 genotypes and COI haplotype, and pleopod angle and COI haplotype. However, Becker et al.’s (2009) data shows a greater mean pleopod angles (Ov: 149.8°; Oq: 134.1°) and a lesser degree of incorporation of genotype A into Oq and genotype B into Ov. Our data shows a greater degree of genetic mixing occurring in our sampled populations, with genotype A and B more extensively incorporated into Oq and Ov, respectively. While both sets of data support our belief that interbreeding has been occurring between the two species for an extended period of time, during which both of the Ov54 genotypes have been incorporated into each of the species, the greater genetic mixing we observe in our data could be evidence of more frequent hybridization in the Blackston Valley Watershed as a whole. However, the question still remains, what process has caused this apparent genetic mixing in *O. virilis* and *O. quinebaugensis*?

The last global glacial maximum occurred about 20,000 years ago, and the glaciers have been receding ever since. Roughly 10,000 years ago, the major glacial sheet that was covering much of North America receded into Canada, allowing previously isolated areas to come back into contact. It is at this point we believe the two semi-reproductively isolated species, *O. virilis* and *O. quinebaugensis*, came into secondary contact. Due to their lack of complete reproductive isolation, the two lineages began to hybridize where their ranges overlapped. However, as a result of assortative mating, this process progressed slowly as they were less likely to interbreed (Becker et al., 2009). Due to the establishment of the region as a center of commerce, contact between *O. virilis* and *O. quinebaugensis* in the Blackstone Valley Watershed was most likely intensified in recent history by the anthropogenic introduction of additional *O. virilis*. Over time, interspecific hybridization between *O. virilis* and *O. quinebaugensis* resulted in the initiation of the introgression process. As hybridization continues to cause significant shifts of allele frequencies, the introgression process currently underway in the examined crayfish populations will eventually allow for the permanent incorporation of foreign alleles into a new, reproductively integrated population system. In other words, genotype B will be permanently incorporated into the genome of in *O. virilis*, and genotype A will be integrated into the genome of *O. quinebaugensis*. This will significantly reduce the distinctness of the latter taxa.
Consequently, *O. quinebaugensis* may be eliminated completely through genetic mixing, especially since it is the smaller population of the two.

We believe what is occurring in the Blackstone Valley Watershed today is evidence of a previous physical isolation that allowed *O. virilis* and *O. quinebaugensis* to speciate, accounting for the difference in pleopod morphology, followed by 10,000 years of hybridization leading to the introgression of Ov54 genotypes into both species.

Further research should examine the genotypic breakdown of individual geographic locations in the watershed as well as those in surrounding regions. It is currently unknown whether populations of *O. quinebaugensis* exist outside of Massachusetts and Rhode Island. If distant populations of *O. quinebaugensis* are identified, a comparison of their genotypes and morphology may provide additional insights into the evolutionary history of both species, especially if pure populations are located. Investigating the effects of physical barriers, such as dams, is also recommended as it may prove useful in determining the rate of gene flow and degree of physical isolation of populations on either side of the barrier. Understanding the effects of physical isolation on the rate of gene flow may provide additional insights into the genetic mixing currently observed between *O. virilis* and *O. quinebaugensis* in the watershed.
5. References


Pfenninger, M. & Schwenk, K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evolutionary Biology 7 (121).


References


W. Avault Jr), pp. 541±549. Louisiana State University, Division of Continuing Education, Baton Rouge, LA.


6. Appendix

A. Population Genetics and Phylogeography

Population genetics employs both mathematical theory and experiential studies in order to gain and understanding of genetic variation and the rates and dynamics of genetic changes within and among populations, including those that lead to the phenomena of adaptation and speciation. Variation in genes, as well as in environmental conditions, is translated by processes of development into deviations in phenotypic characters, such as morphological features (Futuyma, 2005). The genotype frequency is defined as the proportion of a population possessing a certain genotype (Futuyma, 2005). Any alteration of genotype frequencies in one generation will alter the frequencies of alleles carried by the population’s gametes through reproduction, which will in turn alter the genotype frequencies of subsequent generations. This alteration from generation to generation serves as the fundamental process of evolutionary change, and is effected by several different factors. The field of population genetics examines the allele frequency distribution and change under the influence of four evolutionary processes: natural selection, mutation, gene flow, and random genetic drift (Futuyma, 1998).

Natural selection occurs when some genotypic variants in a population benefit from a survival or reproduction advantage over others. Taking into account the Hardy-Weinberg principle, upon which nearly all the theory of population genetics of sexually reproducing organisms rests (Futuyma, 2005), the simplest population-genetic model of natural selection assumes one autosomal locus with two alleles, $A_1$ and $A_2$. The three diploid genotypes $A_1A_1$, $A_1A_2$ and $A_2A_2$ have different fitnesses, denoted by $w_{11}$, $w_{12}$ and $w_{22}$ respectively. These fitnesses are assumed to be constant across generations. A genotype’s fitness may be defined, in this context, as the average number of successful gametes that an organism of that genotype contributes to the next generation, which depends on how well the organism survives, how many matings it achieves, and how fertile it is. Unless $w_{11}$, $w_{12}$ and $w_{22}$ are all equal, then natural selection will occur, possibly leading the genetic composition of the population to change (Futuyma, 2005).

Mutation is the ultimate source of genetic variation, preventing populations from becoming genetically homogeneous in situations where they otherwise would. Even if one allele is selectively superior to all others at a given locus, it will not become fixed in the population; recurrent mutation will ensure that other alleles are present at low frequency, thus maintaining a degree of polymorphism (Hartl et al., 1997). A mutation is often referred to as an alteration of a gene from one form, or allele, to another, distinguished by a phenotypic effect(s) (Futuyma, 2005). However, in the realm of molecular genetics, a mutation is defined as an alteration of a DNA sequence independent of any phenotypic effect. A DNA sequence that differs from homologous sequences by at least one mutation is called a haplotype. Mutations occur most often during DNA replication (Futuyma, 2005).

Random genetic drift refers to the chance fluctuations in allele frequencies that arise in finite populations; it can be regarded as a type of ‘sampling error’ (Hartl et al., 1997). These fluctuations result in the replacement of old alleles by new ones. Chance factors will invariably affect organisms'
survival and reproductive success, thus causing gene frequencies to change. Genetic drift accounts for a large amount of the differentiation in DNA sequences among species (Futuyma, 2005).

Gene flow, or migration, into or out of a population is the fourth and final factor that can affect its genetic composition. Migrants may carry a high frequency of alleles that are different from those of the local population. When mating occurs, the local population’s genetic composition is altered. The evolutionary importance of migration stems from the fact that many species are composed of a number of distinct subpopulations, largely isolated from each other but connected by occasional migration (Futuyma, 1998). Migration between subpopulations gives rise to gene flow, which acts as a sort of ‘glue’, limiting the extent to which subpopulations can diverge from each other genetically (Futyama, 1998; Halliburton, 2004).

Phylogeography is characterized by the description and analysis of the processes that preside over the geographic distribution of genealogies, particularly within species and among directly related species. The spatial and geographical relationships of such genealogies can be analyzed in order to infer the evolutionary history of populations, subspecies and species (Futuyma, 2005). Of particular interest in this field is establishing the way in which distance and environmental barriers, historical processes and life histories shape the genetic structure of populations, and the effects they have upon gene flow (Balloux et al., 2002). These effects are linked to historical biogeographic events, migration patterns, life cycles, mating systems, and various other factors (Avise et al, 1987; Turner et al, 1996). Generally speaking, the geographical distributions of species are broader than an individual’s dispersal capability, rendering populations closer in proximity more genetically similar than those residing further from one another as a direct result of isolation caused by distance (Balloux et al., 2002). Similarly, organisms existing in the same watershed have more genetic similarities in common than those existing in separate watersheds. However, it is possible that physical barriers such as dams and waterfalls may interrupt the natural cycle of the species. The combination of these factors results in the creation of subpopulations. Population subdivision occurs when a large population partitions into smaller subpopulations due to some level of geographical isolation (Balloux et al., 2002). Consequently, the likelihood of an organism mating with an individual belonging to the same subpopulation is increased exponentially. This lack of gene flow in conjunction with mutations and evolutionary forces will inevitably result in a genetic divergence between populations (Halliburton, 2004). Differing rates of genetic divergence among taxa are usually attributed to discrepancies in life-history patterns, dispersal mechanisms and migration (Turner et al, 1996).

Phylogeography is examined on both large and small scales. For instance, the distribution of a species across the continent of North America and Europe can be studied in concurrence with the Pleistocene ice ages which reshaped the water of the portions of both continents (Mäkinen et al, 2007). A study of Gasterosteus aculeatus, a species of freshwater fish, more commonly known as the three-spined stickleback, revealed that the patterns of genetic variation between individuals correlated with glacial activity and the barriers they formed (Mäkinen et al, 2007). Another study used mitochondrial DNA to investigate the phylogenetic relationships between and within the Japanese clawed salamander, Onychodactylus japonicus, and its similar continental relative Onychodactylus Fischeri (Yoshikawa et al., 2008). Research suggests that both O. fischeri and O. japonicus comprise multiple cryptic species. The
estimated divergence times and available geohistorical data suggest that *O. japonicus* began to
differentiate in the Late Miocene sub-epoch, and that the pattern of genetic differentiation of this
species has been strongly influenced by climatic changes and geohistorical events such as volcanic
activity and mountain formation (Yoshikawa *et al.*, 2008). On a smaller geographic scale,
phylogeography was used to study the adaptation process and evolutionary forces acting on 21 natural
populations of *Biomphalaria pfeifferi*, a freshwater snail in southern Madagascar. The study suggests
that migration events along rivers were common, but migration events between watersheds were
extremely rare, which accounted for a high level of genetic differentiation observed among populations
(Charbonnel *et al.*, 2002). Similarly, a study of populations of the shovel-nosed salamander,
*Desmognathus marmoratus*, in the southern Appalachians was conducted in order to observe genetic
differentiation within and between river basins separated by the Eastern Continental Divide. Results
showed a high level of differentiation across the Eastern Continental Divide and a much lesser degree of
differentiation among subpopulations on either side (Jones *et al.*, 2006).

**B. Microsatellites**

Microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs), are
scattered profusely throughout the non-coding regions of the genomes of most plants and animals, and
display a high degree of polymorphism (Li *et al.*, 2002). Typically neutral and co-dominant,
microsatellite loci consist of reiterated short sequences (1-6bp in length) arranged tandemly at a specific
chromosomal location, and are employed as molecular markers (Avise, 2004).

The most common way to detect microsatellites is by designing PCR primers that are unique to one
locus in the genome and the base pair on either side of the repeated portion (Figure 12). Therefore, a
single pair of PCR primers will work for every individual in the species and produce different sized
products for each of the different length microsatellites.

![Figure 12. Detecting microsatellites from genomic DNA. Two PCR primers (forward and reverse gray arrows) are designed to flank the microsatellite region (from Davidson College, 2001).](image)

Microsatellites can be amplified for identification by using the polymerase chain reaction. First, a
primer(s) for the target sequence must be developed. This requires the construction of a genomic
library for the target species, which is subsequently screened for clones that contain microsatellite
repeats using a microsatellite probe. The inserts from those positive clones are then isolated and
sequenced, and information from unique sequences flanking each repeat region is used to synthesize
PCR primers (Avise, 2004; Freeland, 2005). After the primers are designed, they are used to amplify
microsatellite alleles in PCR reactions. DNA is repeatedly denatured at a high temperature to separate the double strand, and then cooled to allow annealing of primers and the extension of nucleotide sequences through the microsatellite (Avise, 2004). This process results in the production of ample DNA to be visible on agarose gels by means of gel electrophoresis.

The mutation rate of microsatellites is much greater than those of other sequences (approximately $10^{-4}$ and $10^{-10}$, respectively). This rapid mutation rate is attributed to slipped-strand mis-pairing during DNA replication (Freeland, 2005). While this characteristic may render them impractical for inferring evolutionary events in the relatively distant past, it makes them particularly useful for inferring relatively recent population genetic events due to the high level of polymorphism exhibited at each locus (Freeland, 2005; Kornfield and Smith, 2000). Taking into consideration the high degree of variability among microsatellites, they are also effective in genetically discriminating between individuals and populations (Van Oppen et al., 1998). Microsatellites are co-dominant; they analyze one locus at a time and allow the detection of both the heterozygous and homozygous condition of an individual (Nguyen et al., 2006). This co-dominancy allows researchers to make inferences relating to the time at which two populations diverged, as well as inferences about the gene flow within and among populations through the examination of differences in microsatellite lengths. These data proves extremely useful in subsequent phylogeographic studies (Yazdani, 2003; Selkoe et al., 2006). Thus, the excessive variability, co-dominant nature, and increasing availability of microsatellites have made them one of the most popular types of markers in population genetics.

Microsatellites occur in both the nuclear and cytoplasmic genomes, including animal mitochondrial DNA (mtDNA) (Lunt et al., 1998). Mitochondrial DNA has provided the first extensive and readily accessible data available to evolutionists in a form suitable for effective genealogical inference within a species. The rapid pace of mtDNA nucleotide substitution, combined with the maternal non-recombining mtDNA inheritance, presents advantages for phylogenetic analysis at the micro-evolutionary level. A nuclear gene system cannot easily compete (Avise et al., 1987).
C. Statistical Data & Charts

1. Chi Square Analysis of COI Haplotype Versus Ov54 Genotype

<table>
<thead>
<tr>
<th></th>
<th>Ov54 Genotype A</th>
<th>Ov54 Genotype AB</th>
<th>Ov54 Genotype B</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype Ov</td>
<td>50</td>
<td>23</td>
<td>26</td>
<td>99</td>
</tr>
<tr>
<td>Haplotype Oq</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Column Totals</td>
<td>55</td>
<td>30</td>
<td>38</td>
<td>123</td>
</tr>
</tbody>
</table>

| Observed | Expected | |O - E| |(O — E)^2| |(O — E)^2 / E|
|----------|----------|----------|--------|-----------|-----------------|-----------------|
| 50       | 44.3     | 5.7      | 32.49  | 0.733408578 |
| 5        | 10.7     | 5.7      | 32.49  | 3.036448598 |
| 23       | 24.1     | 1.1      | 1.21   | 0.050207469 |
| 7        | 5.9      | 1.1      | 1.21   | 0.205084746 |
| 26       | 30.6     | 4.6      | 21.16  | 0.691503268 |
| 12       | 7.4      | 4.6      | 21.16  | 2.859459459 |

chi square = \( 7.576112118 \)

7.58>7.378, genotype is dependent on haplotype to \( p<0.025 \)
### Observed vs. Expected Values

|        | Observed | Expected | \(|O - E| - 1/2\) | \((O - E)^2\) | \((O - E)^2/E\) |
|--------|----------|----------|-----------------|--------------|-----------------|
| 23     | 21.6     | 0.9      | 0.81            | 0.0375       |
| 7      | 8.4      | 0.9      | 0.81            | 0.09642857   |
| 26     | 27.4     | 0.9      | 0.81            | 0.02956204   |
| 12     | 10.6     | 0.9      | 0.81            | 0.07641509   |

\[ \text{chi square} = 0.23990571 \]

0.24 \(<\) 3.84, there is no significant difference between genotypes AB and B

---

### Haplotype Frequencies

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Ov54 Genotype AB</th>
<th>Ov54 Genotype A</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ov</td>
<td>23</td>
<td>50</td>
<td>73</td>
</tr>
<tr>
<td>Oq</td>
<td>7</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Column Totals</td>
<td>30</td>
<td>55</td>
<td>85</td>
</tr>
</tbody>
</table>

---

### Observed vs. Expected Values (continued)

|        | Observed | Expected | \(|O - E| - 1/2\) | \((O - E)^2\) | \((O - E)^2/E\) |
|--------|----------|----------|----------------|--------------|-----------------|
| 23     | 25.8     | 2.3      | 5.3            | 0.20542636   |
| 7      | 4.2      | 2.3      | 5.3            | 1.26190476   |
| 50     | 47.2     | 2.3      | 5.3            | 0.11228814   |
| 5      | 7.8      | 2.3      | 5.3            | 0.67948718   |

\[ \text{chi square} = 2.25910643 \]

2.26 \(<\) 3.84, there is no significant difference between genotypes AB and A

---

### Haplotype Frequencies (continued)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Ov54 Genotype A</th>
<th>Ov54 Genotype B</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ov</td>
<td>50</td>
<td>26</td>
<td>76</td>
</tr>
<tr>
<td>Oq</td>
<td>5</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Column Totals</td>
<td>55</td>
<td>38</td>
<td>93</td>
</tr>
</tbody>
</table>
Observed | Expected | |O -E| 1/2 | (O — E)² | (O — E)²/E
--- | --- | --- | --- | --- | ---
50 | 44.9 | 4.6 | 21.2 | 0.47216036
5 | 10.1 | 4.6 | 21.2 | 2.0990099
26 | 31.1 | 4.6 | 21.2 | 0.68167203
12 | 6.9 | 4.6 | 21.2 | 3.07246377

\[
\text{chi square} = 6.32530605
\]

Three tests performed: Bonferroni correction = alpha/3 = 0.05/3 = 0.017, chi square must be greater than the critical value of p = 0.017 to be significant.

critical value of p = 0.017 = 5.83

6.33 > 5.83, genotype A and B are significantly different, A is associated with Ov and B is associated with Oq

1a. Chi square of allele frequency and COI haplotype

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency of allele A</th>
<th>Frequency of allele B</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ov</td>
<td>123</td>
<td>75</td>
<td>198</td>
</tr>
<tr>
<td>Oq</td>
<td>17</td>
<td>31</td>
<td>48</td>
</tr>
</tbody>
</table>

| Column Totals | 140 | 106 | 246 |

| Observed | Expected | |O -E| | (O — E)² | (O — E)²/E |
| --- | --- | --- | --- | --- | --- |
| 123 | 112.7 | 10.3 | 106.1 | 0.94143744 |
| 17 | 27.3 | 10.3 | 106.1 | 3.88644689 |
| 75 | 85.3 | 10.3 | 106.1 | 1.24384525 |
| 31 | 20.7 | 10.3 | 106.1 | 5.12560386 |

\[
\text{Chi square} = 11.1973334
\]

11.2 > 10.83, frequency is dependent on haplotype to p = 0.001
2. Correlation between Carapace Length and Top Projection Pleopod Angle

<table>
<thead>
<tr>
<th></th>
<th>Correlations</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>top projection angle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top projection angle</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>-.131</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.148</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>123</td>
<td>123</td>
</tr>
<tr>
<td>Body Length (mm)</td>
<td>Pearson Correlation</td>
<td>-.131</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.148</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>123</td>
<td>123</td>
</tr>
</tbody>
</table>

Graphical Correlation
3. Correlation between Carapace Length and Chelae Length/Width Ratio

**Correlations**

<table>
<thead>
<tr>
<th>Body Length (mm) Pearson Correlation</th>
<th>Body Length (mm) Ratio</th>
<th>Body Length (mm)</th>
<th>Body Length (mm) Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Length (mm) Pearson Correlation</td>
<td>1</td>
<td>-.110</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>123</td>
<td>123</td>
<td></td>
</tr>
</tbody>
</table>

**Graphical Correlation**

![Graphical Correlation](image-url)
4. ANOVA of Ov54 Genotype Versus Top Projection Pleopod Angle

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>247.224</td>
<td>2</td>
<td>123.612</td>
<td>4.566</td>
<td>.012</td>
</tr>
<tr>
<td>Intercept</td>
<td>2527101.534</td>
<td>1</td>
<td>2527101.534</td>
<td>93352.340</td>
<td>.000</td>
</tr>
<tr>
<td>Ov54allele</td>
<td>247.224</td>
<td>2</td>
<td>123.612</td>
<td>4.566</td>
<td>.012</td>
</tr>
<tr>
<td>Error</td>
<td>3248.469</td>
<td>120</td>
<td>27.071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2703551.680</td>
<td>123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>3495.693</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .071 (Adjusted R Squared = .055)

Tukey’s Post-hoc Testing

Multiple Comparisons

top projection angle

Tukey HSD

<table>
<thead>
<tr>
<th>(J) Ov54 allele</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) Ov54 allele</td>
<td>(I-J)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>AB</td>
<td>2.926</td>
<td>1.1809</td>
<td>.039</td>
<td>.124</td>
<td>5.729</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.789</td>
<td>1.0975</td>
<td>.033</td>
<td>.184</td>
<td>5.394</td>
</tr>
<tr>
<td>AB</td>
<td>A</td>
<td>-2.926</td>
<td>1.1809</td>
<td>.039</td>
<td>-5.729</td>
<td>-.124</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>-.137</td>
<td>1.2707</td>
<td>.994</td>
<td>-3.153</td>
<td>2.878</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>-2.789</td>
<td>1.0975</td>
<td>.033</td>
<td>-5.394</td>
<td>-.184</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>.137</td>
<td>1.2707</td>
<td>.994</td>
<td>-2.878</td>
<td>3.153</td>
</tr>
</tbody>
</table>

Based on observed means.

The error term is Mean Square(Error) = 27.071.

*. The mean difference is significant at the .05 level.
5. ANOVA of Ov54 Genotype versus Chelae Length/Width Ratio

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>.001†</td>
<td>2</td>
<td>.000</td>
<td>.191</td>
<td>.826</td>
</tr>
<tr>
<td>Intercept</td>
<td>22.459</td>
<td>1</td>
<td>22.459</td>
<td>11544.370</td>
<td>.000</td>
</tr>
<tr>
<td>Ov54allele</td>
<td>.001</td>
<td>2</td>
<td>.000</td>
<td>.191</td>
<td>.826</td>
</tr>
<tr>
<td>Error</td>
<td>.233</td>
<td>120</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24.115</td>
<td>123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>.234</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .003 (Adjusted R Squared = -.013)

Graphical Representation of the ANOVA

![Graphical Representation of the ANOVA](image-url)
6. T-test of COI haplotype versus Pleopod Angle

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>top projection angle</td>
<td>Equal variances assumed</td>
<td>1.814</td>
</tr>
</tbody>
</table>

7. T-test of COI Haplotype versus Chelae Length/Width Ratio

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>ratio assumed</td>
<td>Equal variances assumed</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Correlation of Pleopod Angle and Chelae Ratio

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>top projection angle</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>top projection angle</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>123</td>
</tr>
<tr>
<td>ratio</td>
<td>Pearson Correlation</td>
<td>-0.052</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>123</td>
</tr>
</tbody>
</table>

![Scatter plot showing the correlation between Pleopod top projection angle and Chelae ratio.](chart.png)
9. Logistic Regression Model

Ov54 Contrasts:

For genotype A, insert “1” for genotype indicator 1, and “0” for genotype indicator 2.

For genotype AB, insert “0” for genotype indicator 1, and “1” for genotype indicator 2.

For genotype B, insert “-1” for genotype indicator 1, and “-1” for genotype indicator 2.

For Pleopod Angle:

Insert numerical angle for variable “pleopod projection angle”

For Chela Ratio:

Insert numerical ratio for variable “chelae length/width ratio”

Example:

Individual M6-1 has an Ov54 genotype of B, a top projection angle of 151 degrees, and a chela ration of 0.47.

For genotype B, insert “-1” for genotype indicator 1, and “-1” for genotype indicator 2.

Insert “151” for variable “pleopod projection angle”

Insert “0.47” for variable “chelae length/width ratio”

\[
\text{logit}(\text{COI}_{\text{haplotype } Oq}) = 16.1287 - 0.7250(\text{genotype indicator } 1) + 0.1303(\text{genotype indicator } 2) - 0.0912(\text{top projection pleopod angle}) - 9.4808(\text{chelae } l/w)
\]

\[
\text{logit}(\text{COI}_{\text{haplotype } Oq}) = 16.1287 - 0.7250(-1) + 0.1303(-1) - 0.0912(151) - 9.4808(0.47)
\]
**D. Hardy-Weinberg Equilibrium Calculations**

**Hardy-Weinberg Equilibrium**

Our Data:

Total Ov54:

\[ AA = 55 \]
\[ AB = 30 \]
\[ BB = 38 \]

\[ A = 55 + 30 + 38 = 140 \]
\[ B = 38 + 38 + 30 = 106 \]

Total = 246

Frequency of A = \( \frac{140}{246} = 0.57 = p \)

Frequency of B = \( \frac{106}{246} = 0.43 = q \)

\[ p^2 = (0.57)^2 = 0.325 \]

\[ 2pq = 2(0.57 \times 0.43) = 0.185 \]

\[ q^2 = (0.43)^2 \]

<table>
<thead>
<tr>
<th></th>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>AB</td>
<td>60</td>
<td>30</td>
</tr>
<tr>
<td>BB</td>
<td>23</td>
<td>38</td>
</tr>
</tbody>
</table>

| Observed | Expected | |O - E| | (O - E)^2 | | (O - E)^2 / E |
|----------|----------|----------------|-----------------|----------------|
| 55       | 40       | 15            | 225             | 5.625          |
| 30       | 60       | 30            | 900             | 15             |
| 38       | 23       | 15            | 225             | 9.782608696    |

\[ \text{chi square} = 30.4076087 \]

\[ p << < 0.001 \]