Evolution of a Genetic Incompatibility in the Genus *Xiphophorus*

Samuel V. Scarpino,*[1] Patrick J. Hunt,[1] Francisco J. Garcia-De-Leon,[2] Thomas E. Juenger,[1,3] Manfred Schartl,[4] and Mark Kirkpatrick[1]

[1]Section of Integrative Biology, University of Texas, Austin

[2]Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, S.C. La Paz, Mexico

[3]Institute for Cellular and Molecular Biology, University of Texas, Austin

[4]Physiological Chemistry, University of Würzburg, Biocenter, Am Hubland, Würzburg, Germany

***Corresponding author:** E-mail: scarpino@utexas.edu.

Associate editor: John Parsch

Abstract

Genetic incompatibilities are commonly observed between hybridizing species. Although this type of isolating mechanism has received considerable attention, we have few examples describing how genetic incompatibilities evolve. We investigated the evolution of two loci involved in a classic example of a Bateson–Dobzhansky–Muller (BDM) incompatibility in *Xiphophorus*, a genus of freshwater fishes from northern Central America. Hybrids develop a lethal melanoma due to the interaction of two loci, an oncogene and its repressor. We cloned and sequenced the putative repressor locus in 25 *Xiphophorus* species and an outgroup species, and determined the status of the oncogene in those species from the literature. Using phylogenetic analyses, we find evidence that a repeat region in the proximal promoter of the repressor is coevolving with the oncogene. The data support a hypothesis that departs from the standard BDM model: it appears the alleles that cause the incompatibilities have coevolved simultaneously within lineages, rather than in allopatric or temporal isolation.

Key words: **genetic incompatibility, phylogenetic analysis, coevolution, *Xiphophorus*.**

Article

Introduction

What forces drive the evolution of postzygotic isolating mechanisms? Despite the importance of this question to speciation, we in fact know quite little about how postzygotic isolation originates (Presgraves 2010). A key idea is the Bateson–Dobzhansky–Muller (BDM) hypothesis, which posits that two loci diverge in populations that are geographically isolated (Bateson 1909; Dobzhansky 1936; Muller 1942). On secondary contact, hybrids are produced that have novel combinations of alleles that reduce fitness. This hypothesis is appealing because it generates an adaptive valley between two populations without requiring either of them to cross the valley during their evolutionary history.

Several genetic incompatibility systems involving multiple interacting loci have now been discovered and interpreted as BDM incompatibilities (Coyne 1992; Presgraves et al. 2003; Presgraves 2010). There are, however, important gaps in our understanding of how these systems have evolved. To date there has been only a single test of the suggestion made by Bateson, Dobzhanksky, and Muller that incompatibilities arise by one substitution occurring in each of two allopatric populations. Contrary to that scenario, Cattani and Presgraves (2009) showed that incompatibilities between *Drosophila mauritiana* and the closely related *D. sechellia* and *D. simulans* arose by two or more substitutions in a single lineage that made it incompatible with the ancestral genotype. A second limitation to our current understanding regards how frequently different forces drive the evolution of incompatibilities. Incompatibilities can evolve as the result of divergent selection (Schluter and Conte 2009), genetic conflict between different parts the genome (Gavrilets 2003; Presgraves 2010), or mutation and random genetic drift (Gavrilets 2004). It is also uncertain whether loci that participate in BDM incompatibilities are typically fixed for the alternative alleles or are polymorphic (Cutter 2012). Finally, there is controversy over whether the genetic changes are most often coding, regulatory, or structural in nature (Hoekstra and Coyne 2007).

The genus *Xiphophorus* provides a remarkable opportunity to investigate how loci involved in a genetic incompatibility have evolved. Crosses between two *Xiphophorus* species provided the first example of a BDM incompatibility (Kosswig 1928; Gordon 1931). *Xiphophorus maculatus* and *X. hellerii* are sympatric, and they occasionally hybridize in nature (Kallman and Kazianis 2006; Meyer et al. 2006; Rosenthal and Garcia-De-Leon 2011; Kang et al. 2013). *Xiphophorus maculatus* individuals carry an oncogene and its repressor locus, both of which are absent or nonfunctional in *X. hellerii*. In experimental crosses, backcross hybrids segregate for both genes, with about one-quarter developing a spontaneous and lethal melanoma. Crossing experiments concluded that this hybrid melanoma is inherited as a two-locus, two-allele trait (Ahuja and Anders 1976). This genetic system, now known as the Gordon-Kosswig cross, has been studied extensively as a laboratory model for melanoma (Meierjohann and Schartl 2006). The same incompatibility system has been documented in

crosses between five additional pairs of *Xiphophorus* species (Schartl 2008). *Xiphophorus* remains one of the few systems in which the genetic basis of hybrid incompatibility is known (Coyne 1992; Wu and Ting 2004).

One of the interacting genes in the *Xiphophorus* incompatibility is an oncogene, the *Xiphophorus melanoma receptor kinase* (*xmrk*) gene (Wittbrodt et al. 1989). The *xmrk* gene is unique to *Xiphophorus*, and it arose by a duplication from an epidermal growth factor gene during the diversification of the genus (Weis and Schartl 1998). Critically, not all *Xiphophorus* species have a copy of the *xmrk* locus, and it has apparently been gained and lost multiple times (Schartl 2008). In *X. maculatus* and closely related species that carry *xmrk*, the locus maps to a recombining region of the sex chromosomes (Wittbrodt et al. 1989).

Although *xmrk* is known to cause melanoma, it also has effects that may at times be beneficial. Males that carry *xmrk* show increased pigmentation, a phenotype favored by females in mating (Fernandez and Morris 2008). These males are also larger and more aggressive (Fernandez 2010; Fernandez and Bowser 2010). Consequently, sexual selection may favor the presence of *xmrk*, at least in some lineages (Fernandez and Bowser 2010).

The second actor in the incompatibility is a gene whose phenotypic effect is to repress the melanoma. The repressor maps to a region of an autosome (linkage group V) that contains several genes (Kazianis et al. 1996, 1998). Three lines of evidence suggest the repressor is the gene *cdkn2a/b* (Kazianis et al. 1998). First, *cdkn2a/b* is homologous to the mammalian melanoma suppressor locus *CDKN2* (Kazianis et al. 1999). Variation in the promoter and coding regions of this gene are implicated in both the occurrence and progression of melanoma and other cancers (Merlo et al. 1995; Merbs and Sidransky 1999). Second, the two species involved in the Gordon-Kosswig cross, *X. maculatus* and *X. hellerii*, differ in both the coding and promoter regions of *cdkn2a/b* (Kazianis et al. 2000). Third, there is some evidence that *cdkn2a/b* expression in tumor cells in *Xiphophorus* is affected by these differences (Kazianis et al. 1999, 2004; Butler et al. 2007). In this article we refer to *cdkn2a/b* as the oncogene repressor, but definitive proof of that function using a transgenic construct with a loss-of-function mutant has not yet been obtained.

This situation motivated us to investigate whether *cdkn2a/b* has coevolved with *xmrk*. One plausible scenario, for example, was that *xmrk* might have spread by positive selection (Futreal et al. 2004; Fernandez and Morris 2008), which would then drive the evolution of *cdkn2a/b* to remediate its deleterious effects. To investigate the coevolution hypothesis, we gathered data on the promoter and coding region of *cdkn2a/b* across the genus *Xiphophorus*. We used these data to test for correlated evolution in the two genes. Specifically, we asked if the gain or loss of *xmrk* is associated with changes in either the coding or regulatory regions of *cdkn2a/b*.

We find evidence for rapid evolution caused by positive selection in the first exon of *cdkn2a/b*, but the pattern of evolution is not correlated with *xmrk* status (presence or absence). Features of the *cdkn2a/b* promoter, however, do correlate with *xmrk* status across the phylogeny. The length of a region in the proximal *cdkn2a/b* promoter is positively associated with the presence of *xmrk*. This region contains several repetitive elements, which we will refer to as the "promoter repeat region" and abbreviate as *PRR*; see Materials and Methods section for an exact definition of this region. We also find evidence of a correlation between the length of this region and *xmrk* status across three populations of *X. maculatus* that is consistent with the between-species pattern. These results suggest that the oncogene and its repressor have evolved together within lineages. Further, it seems they may coevolve simultaneously, rather than in spatial or temporal isolation as suggested by the BDM hypothesis.

Results

Positive Selection on the Coding Region of *cdkn2a/b*

The coding region of *cdkn2a/b* has experienced multiple bouts of positive selection. The best-fit model estimated using GABranch shows two selection regimes (fig. 1 and table 1). That model suggests that on 12 branches of the phylogeny there has been purifying selection ($d_N/d_S = 0$), while on the other 17 branches there has been positive selection ($d_N/d_S = 3.59$). These results give strong support for multiple bouts of amino acid substitution by positive selection.

We then tested whether patterns of evolution in the coding region of *cdkn2a/b* correlate with the presence of *xmrk*. First, we asked whether the presence of *xmrk* correlates with the mode of selection (positive or purifying) on terminal branches of the *cdkn2a/b* gene tree. That relationship is not significant ($\chi^2 = 2.34$, $P = 0.194$ as determined by

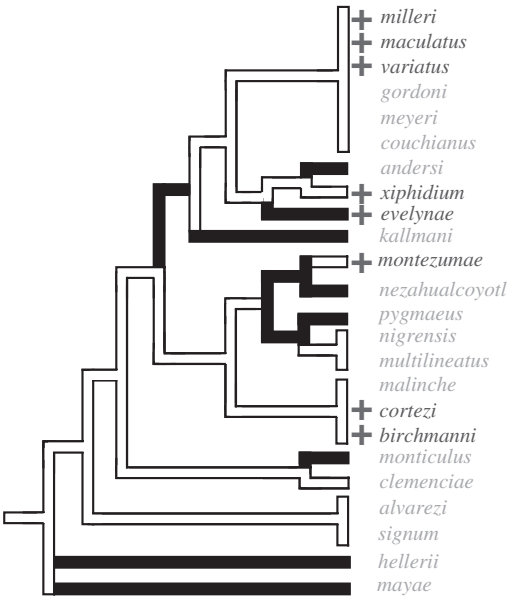

Fig. 1. Repeated bouts of positive evolution in the coding region of *cdkn2a/b*. The consensus gene tree for the first exon of *cdkn2a/b* is shown. Branches that are colored show statistically significant support for a d_N/d_S ratio that is >1. The tips are labeled by species, with plus signs and black letters indicating species that have *xmrk* and gray showing species that lack it.

Table 1. Results from GABranch.

No. of Regimes	ΔAIC	d_N/d_S	No. of Branches
1	−4.1	1.18	29
2	–	0	12
		3.59	17
3	−9.27	0	12
		0.79	1
		1,000.	16

NOTE.—Underlined values correspond to the best-fit model.

randomization). We then used stochastic character mapping to perform ancestral state reconstruction across the entire tree ([Bollback 2006], as implemented in the R package Diversitree v. 0.9-6 [FitzJohn 2012]). We did not see an association between positive selection and either the presence of *xmrk* or changes in its status (i.e., gain or loss) ($P = 0.99$). Last, we looked for a correlation between the rate of evolution of the coding region of *cdkn2a/b* and the presence of *xmrk*. We found no evidence of an association ($P = 0.2$, TraitRate [Mayrose and Otto 2010]).

The Promoter of *cdkn2a/b* Coevolves with *xmrk*

We found that the length of the *PRR* of the putative repressor gene *cdkn2a/b* evolves in a correlated way with *xmrk* status. Species with *xmrk* on average have a longer *PRR* than those without. This region was first identified in a comparison between the species in the Gordon-Kosswig cross: *X. maculatus* carries *xmrk* and has a substantially longer promoter than *X. hellerii*, which does not carry *xmrk* (Kazianis et al. 1999, 2004). Looking across the entire genus, we find that species with *xmrk* have significantly longer promoters than those without ($P = 0.007$, Wilcoxon Mann–Whitney rank sum test, $z = -2.64$). The results were also significant when *X. maculatus* is removed from the analysis ($P = 0.017$, Wilcoxon Mann–Whitney rank sum test, $z = -2.33$). That analysis does not account for potential phylogenetic dependencies, however, and so we next attempted to control for them.

Figure 2 shows the distribution of the *PRR* length and *xmrk* across the genus. The figure, which is based on just one of the phylogenies used in the following analyses, shows several cases that suggest when closely related species differ in their *xmrk* status, the species that carries *xmrk* tends to have a longer *PRR* at *cdkn2a/b*. Testing for the significance of that pattern is complicated by two issues: uncertainty in the phylogeny and introgression between species. Here we describe how we accounted for phylogenetic uncertainty, and we return to the issue of introgression in the Discussion section.

For a given phylogeny, we used BayesTrait Continuous to find the likelihood for a model in which evolution of *xmrk* and the *PRR* are correlated. We averaged those likelihoods over the 1,000 most likely phylogenies (accounting for >99% of the credible interval of phylogenetic space), weighting each value by the likelihood of the phylogeny. Using a dataset consisting of six loci to estimate the phylogenies, we reject the null hypothesis that the *PRR* length and the oncogene are evolving independently ($P = 0.0043$, $\chi^2 = 3.56$, df = 1, likelihood ratio test). We also reran the analysis with a recently published dataset that has 13 loci. The result again shows significant support for correlated evolution ($P = 0.026$, $\chi^2 = 2.01$, df = 1, likelihood ratio test).

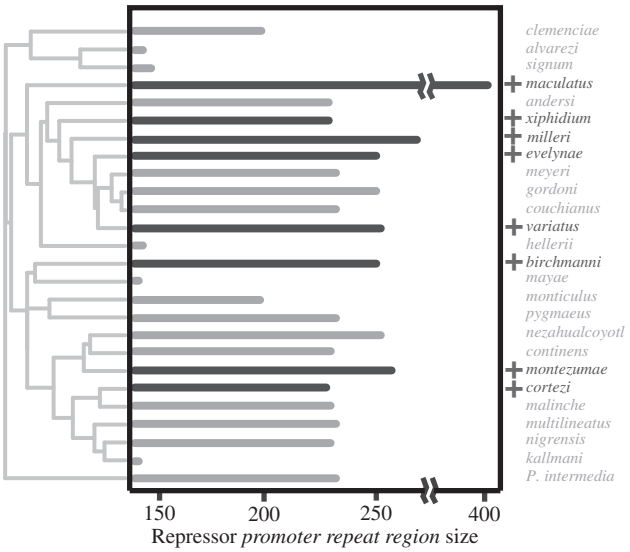

FIG. 2. The phylogenetic distribution of the size of the *PRR* for the putative repressor, *cdkn2a/b*. At left is a phylogenetic tree that has a high likelihood. The center box shows the *PRR* length. The species names at right contain a plus sign and are black for those that carry *xmrk* and are gray for those without *xmrk*.

We then asked whether these results might be driven by species with unusual promoters. The *PRR* in *X. maculatus* is much longer than that in any other species. We therefore removed *X. maculatus* from the data set and reran the analyses just described. There is again strong support for correlated evolution ($P = 0.0071$, $\chi^2 = 3.114$, df = 1, likelihood ratio test). We then replaced the actual *PRR* length in *X. maculatus* by the mean promoter length across all species in the genus, and again found significant support for correlated evolution ($P = 0.0048$, $\chi^2 = 3.45$, df = 1, likelihood ratio test).

Finally, we fit the BayesTrait Continuous model to three *Xiphophorus* phylogenies that were recently estimated from genomic-scale data (Cui et al. 2013a). Averaging the likelihood for the model across those three phylogenies, we find that support for the correlated evolution model is not quite significant ($P = 0.055$, $\chi^2 = 2.899$, df = 1, likelihood ratio test). Taken together, these results provide support for the hypothesis of correlated evolution of the oncogene (*xmrk*) and the *PRR* of its putative repressor (*cdkn2a/b*).

There are, however, caveats to that conclusion. First (and most obvious), the strength of our conclusion depends in part on which dataset is used to estimate the phylogeny. Second, there is evidence of introgression between species (Meyer et al. 2006; Schumer et al. 2012; Kang et al. 2013; Cui et al. 2013a). Consequently, even the true species phylogeny may not accurately reflect the evolutionary history of the two genes we are studying. Last, although there are differences in average promoter lengths, there are also individual species that do not fit the pattern. For example, the platyfish *X. xiphidium* carries the *xmrk* gene but has a shorter *PRR* (244 base pairs) than the non-*xmrk* bearing species *X. gordoni*

(251 bp), *X. meyeri* (247 bp), and *X. couchianus* (247 bp). We revisit these issues in the Discussion section.

Within-Species Polymorphism in *xmrk* Promoter Length of *cdkn2a/b*

The platyfish *X. maculatus* offers another opportunity to test the hypothesis of coevolution of *xmrk* and the *PRR* of *cdkn2a/b*. This species is polymorphic for the presence of *xmrk*, and there is significant variation between populations in its frequency (fig. 3). The coevolution hypothesis leads us to expect a positive correlation across populations in the length of the *cdkn2a/b* promoter and the frequency of *xmrk*.

We determined the lengths of the *PRR* in 107 individuals sampled from three populations. In two populations, *xmrk* is near fixation, while in the third population about 60% of individuals have *xmrk* (fig. 3). The population with the lower frequency of *xmrk* also has a significantly shorter mean promoter length ($P < 10^{-7}$, $F = 16.1$, df = 2, analysis of variance) (fig. 4).

In the population with the lower frequency of *xmrk* (San Juan), an individual's *xmrk* status is not correlated with the length of its *PRR* ($P > 0.05$, Welch *t*-test, $t = 0.007$, df = 36.0). That is, there is no evidence that the two loci are in linkage disequilibrium. But because they are on different chromosomes, however, detectable linkage disequilibrium would not be expected even with strong epistatic selection.

Since we only collected a single population with an intermediate frequency of *xmrk*, we were unable show a significant correlation between *xmrk* frequency and *cdkn2a/b PRR* length across populations. The probability that the pattern we found (the population with lowest *xmrk* frequency has the shortest mean *PRR* lengths) occurred by chance is 1/3. Nevertheless, these data add weight to our earlier conclusions about the coevolution of *xmrk* and *cdkn2a/b*.

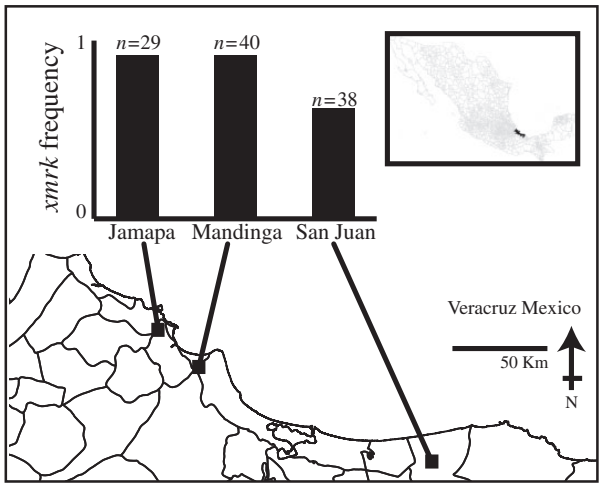

Fig. 3. The frequency of *xmrk* individuals in three *Xiphophorus maculatus* populations. The two northern populations, Jamapa, and Mandinga, had nearly 100% *xmrk* frequency, while the southern population of San Juan had a significantly lower observed frequency of *xmrk*. The Jamapa and Mandinga populations are likely from the same river drainage, while the San Juan population is from two drainages to the south.

Discussion

We have developed three lines of evidence that suggest how *xmrk*, an oncogene, may be coevolving with *cdkn2a/b*, its putative repressor, in the genus *Xiphophorus*. First, we find there have been repeated bouts of amino acid substitution in the first exon of *cdkn2a/b* apparently driven by positive selection. These bouts, however, do not appear to be correlated with *xmrk* status. Second, phylogenetic analyses suggest that there has been correlated evolution between the presence of *xmrk* and the length of the *PRR* of *cdkn2a/b*. Third, within one species that is polymorphic for *xmrk*, the *PRR* is longer on average in two populations with high *xmrk* frequencies than it is in a third population with an intermediate frequency of *xmrk*. Taken together, these results support the hypothesis that the *cdkn2a/b* promoter is coevolving with *xmrk*, the oncogene that it is thought to repress.

The evidence from our phylogenetic analyses, however, does have ambiguity. We find significant statistical support for coevolution when we account for phylogenetic uncertainty by averaging results over the 1,000 most likely phylogenetic trees estimated from two published datasets ($P < 0.03$ for one dataset and $P < 0.005$ for the other). On the other hand, the pattern is only significant at $P < 0.055$ when we use three specific phylogenies based on a much larger genomic dataset (Cui et al. 2013a). Even those three phylogenies, however, have uncertainty. For example, they vary slightly depending on the species against which the genome is assembled, and (like all phylogenetic estimates) they are based on an evolutionary model that may not be correct.

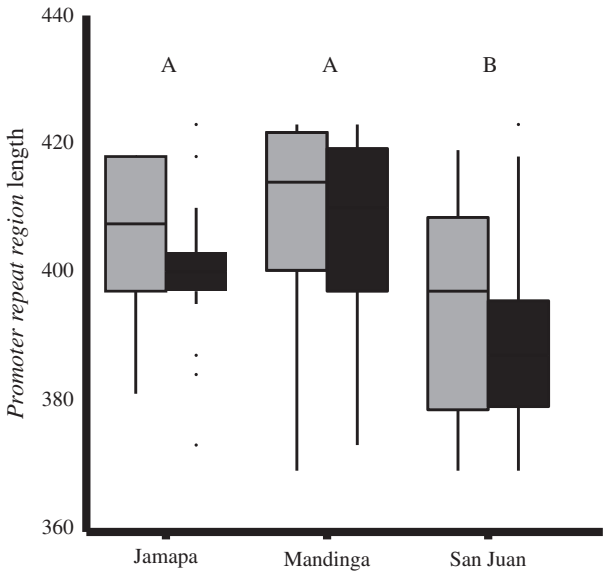

Fig. 4. Average *cdkn2a/b PRR* length correlates with the frequency of *xmrk* across three populations. Each population is separated into those individuals without *xmrk* (gray) and those with *xmrk* (black). The two populations labeled "A," Jamapa and Mandinga, had a significantly longer mean *PRR* length than the population labeled "B," San Juan ($P < 10^{-7}$, $F = 16.1$, df = 2, analysis of variance). In the San Juan population, an individual's *xmrk* status is not correlated with the length of its *PRR* ($t = 0.007$, df = 36.0, $P > 0.05$, Welch *t*-test).

An even more complicated issue arises when one considers there is strong evidence of introgression between species (Cui et al. 2013a). Consequently, the species phylogeny is likely to differ from the gene trees for *xmrk* and *cdkn2a/b*. Coevolution between the two loci has occurred in the context of their gene trees, not the species phylogenies. Unfortunately, estimates for the gene trees are far less certain than the (already uncertain) species phylogenies.

Given this imprecision, we favor the conclusions that are based on averages across many possible species phylogenies (in the hope that they capture major topological features of the gene trees) rather than relying on any single phylogeny (no matter how well it may be supported). We acknowledge, however, that those conclusions are subject to interpretation.

The oncogene *xmrk* and its repressor have long been cited as a classic example of a two-locus genetic incompatibility (Coyne 1992; but see Nei and Nozawa [2011] for an alternative view). Our results suggest that the *Xiphophorus* incompatibility has evolved in a different way than was originally proposed by Bateson (1909), Dobzhansky (1936), and Muller (1942). They suggested that in one population a substitution occurs at a first locus, say allele *A* replacing allele *a*, while in an allopatric population an independent substitution occurs at a second locus, say allele *B* replacing allele *b*. Secondary contact and hybridization then produce novel genotypes that carry the two derived alleles (*A* and *B*), and incompatibilities between them cause low fitness. A variant on the BDM hypothesis suggests that the substitution of allele *A* is followed at a later time by substitution in the same population of an allele at a second locus, say allele *C* replacing allele *c*. Following secondary contact, incompatibility then occurs if the derived allele *C* is incompatible with the ancestral allele *a* (Presgraves 2010). This "derived-ancestral" version of the BDM hypothesis is consistent with an analysis of incompatibilities between *D. mauritiana* and the closely related *D. sechellia* and *D. simulans* (Cattani and Presgraves 2009). Central to both versions of the BDM hypothesis is that low-fitness genotypes are never produced in the evolutionary history of either population before they have secondary contact.

Our phylogenetic analyses of coevolution of *xmrk* and *cdkn2a/b* suggest that both partners in the incompatibilities have evolved within single lineages, as envisioned in the derived-ancestral version of the BDM hypothesis. But in a departure from that hypothesis, the data further suggest that incompatible genotypes have been produced within single populations as the result of the "simultaneous" (rather than sequential) evolution of the two loci. We find that populations of *X. maculatus* are polymorphic at both loci, suggesting that the two genes coevolve and generate incompatible genotypes even in the absence of secondary contact.

What could cause *xmrk* to spread despite its deleterious effects? Fernandez and Morris (2008) and Fernandez and Bowser (2010) make the fascinating suggestion that *xmrk* has been favored by sexual selection that was sufficiently strong to offset the oncogene's negative effects on viability. A second possibility is that *xmrk* spread by hitchhiking with the closely linked macromelanophore locus, which has a function in kin recognition (Franck et al. 2001). A third hypothesis, also consistent with our phylogenetic analyses, is that evolution of *cdkn2a/b* drives evolution of the system. Its promoter might evolve by selection on a pleiotropic effect, by mutation pressure from the microsatellite motif, or simply by drift. Once an allele that acts as a melanoma repressor reaches an appreciable frequency, it enables *xmrk* to invade.

Under all three of these hypotheses, the fitness "valley" between the species caused by the incompatibilities in fact did not exist, at least when they first evolved. For example, under the Fernandez et al. hypothesis, the viability cost of the melanoma was offset by the reproductive advantage that *xmrk* conferred. It is certainly possible that the incompatibilities create a true fitness valley at present, for example because of changed pressures of sexual selection or the ecological environment. Further, all of these scenarios are made more plausible by the fact that the melanoma repressor has a dominant gene action. (Recall that melanomas do not appear until the backcross and later generations of the *maculatus* × *hellerii* cross.) That substantially decreases the negative selection against the incompatibility compared with cases where F_1 hybrids suffer an immediate fitness cost.

How could *xmrk* be gained and lost multiple times across the genus? Even if several independent losses by deletion seem plausible, parallel gains by duplication seem unlikely. Our phylogenetic analyses do not rule out the hypothesis that *xmrk* had only a single origin and was lost repeatedly. Another possibility is that multiple gains of *xmrk* have occurred by hybridization and introgression between species. Both phenomena are common in *Xiphophorus* (Meyer et al. 2006; Schumer et al. 2012; Kang et al. 2013; Cui et al. 2013a). Introgression could also explain some of the deviations from the positive correlation between *xmrk* and the *PRR* length of *cdkn2a/b*: for example, some of *Xiphophorus* species that depart from expected *PRR* lengths are known to hybridize in nature (Schartl 2008; Rosenthal and Garcia-De-Leon 2011; Cui et al. 2013a).

In this study, we considered *xmrk* to be a binary trait. Detailed molecular analysis, however, has revealed that there is additional variation associated with this locus (Weis and Schartl 1998; Schartl 2008; Regneri and Schartl 2011). In laboratory strains of *X. maculatus*, insertion/deletion polymorphism in the promoter and coding region of *xmrk* are associated with variation in melanoma prevalence, progression, and severity (Regneri and Schartl 2011). Thus variation at this locus is more complex than just its presence or absence. Our phylogenetic analyses also ignore within-species polymorphisms for the presence of *xmrk*. In fact, polymorphism is present in *X. maculatus* (Schartl 1990), *X. montezumae*, and all other *xmrk* carrying species studied to date (Schartl M, unpublished data). Our analyses neglect this polymorphism, and its evolutionary significance (if any) is unknown. However, our phylogenetic analyses are conservative with respect to polymorphism because any decrease in *xmrk* frequency will decrease its effect on the evolution of *cdkn2a/b*. No species scored as *xmrk* negative in our analyses (see fig. 1) has been found to carry a functional copy of the gene.

An important gap in our story is that *cdkn2a/b* has not been validated as the melanoma repressor using expression assays or transgenic constructs. There is, however, support for its role: in the Gordon-Kosswig cross, there are differences in *cdkn2a/b* expression between healthy individuals and those with *xmrk*-induced melanoma (Kazianis et al. 1999, 2004; Butler et al. 2007). Further, the cancer phenotype depends on whether an individual carries the *X. maculatus* or *X. hellerii* allele at *cdkn2a/b* (Butler et al. 2007). That observation, however, does not show whether it is the *PRR* or some other linked region that is responsible. Functional validation of the effects of the *PRR* is an important goal of future research.

How might variation at *cdkn2a/b* affect the melanoma? The length of the repressor's *PRR* could directly modulate expression of the *cdkn2a/b* protein, which then affects expression of *xmrk*. Alternatively, *cdkn2a/b* expression could act as a regulator (either upstream or downstream) at some other point in the melanoma pathway. This second scenario includes the case where *cdkn2a/b* affects expression of other closely linked genes that are inside the mapping region identified for the repressor (Schartl et al. 2013). The *PRR* of *cdkn2a/b* contains a series of predicted transcription factor binding sites. An important question is whether these length differences could in fact alter expression of the melanoma phenotype. That hypothesis is made plausible for one of the elements in the *PRR*, a GT microsatellite, by evidence from another complex disease, asthma. In humans, length variation in a GT repeat region of the *STAT6* promoter is associated with both progression and symptom severity (Gao et al. 2004). The addition of three bases in the GT repeat region is sufficient to explain symptom variation (Gao et al. 2004).

Our findings contribute further understanding to three general issues regarding the genetics of postzygotic isolation. Several evolutionary forces have been implicated in the evolution of postzygotic incompatibilities, including positive selection, neutral processes (drift and mutation), and genomic conflict (Presgraves 2010). Although the evidence is not definitive, the most plausible interpretation of the *Xiphophorus* system is that this incompatibility evolved by positive selection. Second, we find variation within species for the genetic incompatibility in the form of presence/absence polymorphisms for *xmrk* status and variable lengths of the *cdkn2a/b PRR*. Thus *Xiphophorus* provides another case of a genetically variable postzygotic incompatibility (Cutter 2012). Third, there is an ongoing discussion of whether postzygotic isolation typically results from evolution in coding, regulatory, or structural features of the genome (Hoekstra and Coyne 2007). In the case of *Xiphophorus*, evolution of the *cdkn2a/b PRR* is likely both structural and regulatory, whereas presence/absence variation at *xmrk* is an example of structural variation.

We have focused this article on the role that *xmrk* and its repressor may play in postzygotic isolation. The system offers rich opportunities to study other key problems in evolutionary genetics. These include the roles played by microsatellites in adaptation, by sexually antagonistic selection in the evolution of sex chromosomes, and by positive selection in the evolution of oncogenes.

Materials and Methods

Data Collection

We obtained genomic DNA from all species in the genus *Xiphophorus* (except *X. mixei*) and from a closely related outgroup species, *Priapella intermedia*. Tissue was taken from wild-caught individuals and from laboratory stocks that had been established from wild-caught fish. *Xiphophorus malinche* genomic DNA was generously provided by G. Rosenthal. We studied patterns of variation within *X. maculatus* using 107 individuals that we collected from three populations in Veracruz, Mexico, in 2008, 2009, and 2011. These populations are *jamapa*—Rio Jamapa (19° 0′49.90″N 96° 14′51.76″W), *mandinga*—Rio Jamapa (19° 0′46.97″N 96° 5′45.45″W), and *san juan*—Rio San Juan (18°20′3.65″N 95°27′36.95″W).

Data on *xmrk* status for all species was taken from Weis and Schartl (1998) and Schartl (2008), who used a variety of Southern blot and probe-based methods to detect *xmrk*. All species where *xmrk* is present are polymorphic for *xmrk*, with variable frequencies of the locus between populations (Schartl 1990; Schartl M, unpublished data). We coded all species where *xmrk* has been found at least once as having *xmrk*, which was conservative with respect to the phylogenetic analyses (see Discussion section).

We amplified, cloned, and sequenced the first 1000 bases of coding region of *cdkn2a/b* for all species. Primers spanning the first exon were Forward: ACG CCT GGT TCG GTT TTC CT and Reverse: GCC TTA TTC ACG GTT CTC AAT C. Polymerase chain reaction (PCR) conditions were initial denaturing time of 5 min at 94 °C, 40 cycles of 94 °C 30 s, 59 °C 30 s, 72 °C 30 s, and a final elongation step of 5 min at 72 °C. Cloning followed standard procedures for TOPO TA cloning with pCR 2.1-TOPO and TOP10 chemically competent cells, Invitrogen K4500-01. Sequences were deposited in GenBank under accession numbers KF002384–KF002407.

The promoters of *cdkn2a/b* in *X. maculatus* and *X. hellerii*, which are the species in the famous Gordon-Kosswig cross that first identified this two-locus incompatibility, have several differences (Kazianis et al. 1999). We focused on the most conspicuous features of this region: a length polymorphism containing a number of repetitive elements. The most striking is a GT microsatellite (Kazianis et al. 1999, 2004) that lies approximately 450 bases upstream from the start codon. This GT microsatellite is about 25 bases long in *X. hellerii* and about 170 bases long in *X. maculatus*. Because this GT repeat feature is a microsatellite, it may experience elevated mutation rates.

Pilot sequencing identified a highly variable repeat region in the proximal *cdkn2a/b* promoter that contains the GT microsatellite. (These sequences are deposited in GenBank under accession numbers KF002357–KF002383.) We refer to this section of the proximal promoter as the *PRR*. The boundaries of the region are defined by conserved nonrepetitive motifs that are given by the primer pair: Forward: ACA CTA AAT AGC CCT CTA CCA, Reverse: CAT AAA CAC CAG ACT GAA ACA C. To obtain precise estimates of its length,

we used a fragment analysis. We amplified the *PRR* using the PCR conditions described above but with an annealing temperature of 51.5 °C and using fluorescently labeled primers corresponding to the two sequences just stated. Fragment analysis was performed with standard protocols from the Institute for Molecular and Cell Biology's core facility at the University of Texas at Austin.

Models and Analyses

We analyzed the pattern of selection acting on the coding region of *cdkn2a/b* using GABranch (Pond and Frost 2004). We assumed the best-fit model of nucleotide evolution for our data, which was HKY85 (a model with two substitution rates). The method proceeds in three steps. First, we assume that along each branch the rates of nucleotide substitution are chosen from one of B sets of the two substitution rates. With $B = 1$, for example, a single set of two rates pertains to the entire tree, while with $B = 2$ there are two types of branches each with its own pair of values for the two rates. Second, maximum likelihood is used to estimate B (the number of substitution rate sets), the length of each branch on the gene tree, and which of the B sets of substitution rates pertain to each branch. Third, the fit of the model with different values of B is compared, and the optimal value of B is determined using the Akaike Information Criterion. (GABranch uses a version that includes a correction for small sample size, "AICc" [Akaike 1974; Sugiura 1978].) This analysis provides two kinds of information. First, it gives an estimate of d_N/d_S, the ratio of the synonymous to nonsynonymous rates of substitution, for the first exon of *cdkn2a/b* along each branch. Second, it determines whether this ratio varies across the gene tree.

To determine whether *xmrk* status was associated with bouts of positive selection, we first estimated ancestral states for branches and nodes and inferred the timing of gains and loses using stochastic character mapping. Stochastic character mapping was performed using the Diversitree package (v. 0.9-3) in R (v. 2.15.1) (FitzJohn 2012).

Next, we tested for correlated evolution between the *xmrk* status (its presence or absence) and the length of the *cdkn2a/b PRR*. The evolutionary hypothesis we tested has two components. The first is that *xmrk* has been gained and lost multiple times across the phylogeny. A special case of this model includes the scenario where *xmrk* had a single origin early in the evolution of the genus and was lost multiple times. The second part of our hypothesis is that the *cdkn2a/b PRR* is a continuous trait under stabilizing selection with an optimal length that differs depending on whether *xmrk* is present or absent. Within each of these two regimes (*xmrk* present vs. absent), changing selection pressures and random genetic drift might cause the promoter length to vary in time around the optimum.

Our statistical question therefore is whether the *PRR* lengths differ significantly when *xmrk* is present and when it is absent. We tested for that difference using BayesTrait Continuous (Pagel 1994). This model assumes that *xmrk*, a binary trait, evolves as Markov chain and that the *PRR* length, a continuous trait, evolves by Brownian motion with a constant variance. It calculates the likelihood of a correlation between average *PRR* length and *xmrk* status given a phylogeny. To control for phylogenetic nonindependence, the method uses a random effects model. We further evaluated the utility of BayesTrait Continuous using simulations of both the correlated and uncorrelated evolutionary model. The results demonstrate support for the ability of BayesTrait Continuous to favor the correct model of evolution used in the simulation.

We used three approaches for our phylogenetic analyses. First, we estimated phylogenies using sequences from five nuclear loci and one mitochondrial locus included in Meyer et al. (1994, 2006). Trees were sampled and their probabilities calculated using Mr Bayes (Ronquist et al. 2012). We obtained an overall likelihood by weighting each model likelihood by the probability for that tree (Huelsenbeck et al. 2000). The calculations in BayesTrait Continuous are computationally fast, which allows us to calculate the likelihood for both the correlated and uncorrelated model on each of the 1000 most likely trees. Second, we repeated the analysis using sequences from 11 nuclear and 2 mitochondrial loci included in Kang et al. (2013). All of the loci contained in Meyer et al. (1994, 2006), except one nuclear locus, were contained in the Kang et al. (2013) dataset. Finally, we used three recently published phylogenies constructed using genomic data (Cui et al. 2013a). Data for these phylogenies were obtained from Dryad (Cui et al. 2013b). Again, we fit both models, correlated and uncorrelated, to these trees.

Acknowledgments

The authors thank the Mexican federal government and the state of Veracruz for permission to collect samples (permit number DGOPA.00322.25011.-0100). Ron Walter and the *Xiphophorus* Genetic Stock Center at Texas State University provided useful information on *cdkn2a/b* and helpful suggestions during the initial phases of the research. The authors would also like to thank David Des Marais, Robin Hopkins, Rafael Guerrero, Janine Regneri, Sara Sawyer, and the reviewers for useful comments, discussion, and advice. This work was performed under IACUC protocols AUP-2001-00101 and 08071402 at UT Austin. This work was supported by the following grants: National Science Foundation DEB-1110526 to S.V.S and M.K., National Science Foundation DEB-0819901 to M.K., National Science Foundation DEB-0546316 to T.E.J., National Science Foundation Graduate Research Fellowship to S.V.S., and a graduate student research grant from the Ecology, Evolution, and Behavior Graduate Program at The University of Texas at Austin.

References

Ahuja M, Anders F. 1976. A genetic concept of the origin of cancer, based in part upon studies of neoplasms in fishes. *Prog Exp Tumor Res.* 20:380–397.

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automatic Control.* 119:716–723.

Bateson W. 1909. Heredity and variation in modern lights. In: Seward A, editor. Darwin and modern science. Cambridge: Cambridge University Press. p. 85–101.

Bollback J. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.

Butler A, Trono D, Beard R, Fraijo R, Nairn R. 2007. Melanoma susceptibility and cell cycle genes in *Xiphophorus* hybrids. *Mol Carcinog.* 46: 685–691.

Cattani MV, Presgraves DC. 2009. Genetics and lineage-specific evolution of a lethal hybrid incompatibility between *Drosophila mauritiana* and its sibling species. *Genetics* 181:1545–1555.

Coyne J. 1992. Genetics and speciation. *Nature* 355:511–515.

Cui R, Schumer M, Kruesi K, Walter RB, Andolfatto P, Rosenthal GG. 2013a. Phylogenomics reveals extensive reticulate evoution in *Xiphophorus* fishes. *Evolution* 67:2166–2179.

Cui R, Schumer M, Kruesi K, Walter RB, Andolfatto P, Rosenthal GG. 2013b. Data from: phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. Dryad Data Repository.

Cutter AD. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends Ecol Evol.* 27:209–218.

Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophilia pseudoobscura* hybrids. *Genetics* 21:113–135.

Fernandez A. 2010. A cancer-causing gene is positively correlated with male aggression in *Xiphophorus cortezi*. *J Evol Biol.* 23:386–396.

Fernandez A, Bowser P. 2010. Selection for a dominant oncogene and large male size as a risk factor for melanoma in the *Xiphophorus* animal model. *Mol Ecol.* 19:3114–3123.

Fernandez AA, Morris MR. 2008. Mate choice for more melanin as a mechanism to maintain a functional oncogene. *Proc Natl Acad Sci U S A.* 105:13503–13507.

FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol.* 3:1084–1092.

Franck D, Dikomey M, Schartl M. 2001. Selection and the maintenance of a colour pattern polymorphism in the green swordtail (*Xiphophorus helleri*). *Behaviour* 138:467–486.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. *Nat Rev Cancer.* 4:177–183.

Gao P, Heller N, Walker W, Chen C, Moller M, Plunkett B, Roberts M, Schleimer R, Hopkin J, Huang S. 2004. Variation in dinucleotide (GT) repeat sequence in the first exon of the *STAT6* gene is associated with atopic asthma and differentially regulates the promoter activity in vitro. *J Med Genet.* 41:535–539.

Gavrilets S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* 57:2197–2215.

Gavrilets S. 2004. Fitness landscapes and the origin of species. Princetion (NJ): Princetion University Press.

Gordon M. 1931. Hereditary basis of melanosis in hybrid fishes. *Am J Cancer.* 15:1495–1523.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.

Huelsenbeck JP, Rannala B, Masly J. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.

Kallman K, Kazianis S. 2006. The genus *Xiphophorus* in Mexico and Central America. *Zebrafish* 3:271–285.

Kang J, Schartl M, Meyer A. 2013. Comprehensive phylogenetic analyses of the genus *Xiphophorus*: hybrid origin of a swordtail fish, *Xiphophorus monticolus*. *BMC Evol Biol.* 13:25–44.

Kazianis S, Coletta LD, Morizot DC, Johnston DA, Osterndorff EA, Nairn RS. 2000. Overexpression of a fish *CDKN2* gene in a hereditary melanoma model. *Carcinogenesis* 21:599–605.

Kazianis S, Gutbrod H, Nairn RS, et al. (11 co-authors). 1998. Localization of a *CDKN2* gene in linkage group V of *Xiphophorus* fishes defines it as a candidate for the *DIFF* tumor suppressor. *Gene Chromosomes Canc.* 22:210–220.

Kazianis S, Khanolkar VA, Nairn RS, Rains JD, Trono D, Garcia R, Williams EL, Walter RB. 2004. Structural organization, mapping, characterization and evolutionary relationships of *CDKN2* gene family members in *Xiphophorus* fishes. *Comp Biochem Physiol C Toxicol Pharmacol.* 138:291–299.

Kazianis S, Morizot D, Della Coletta L, Johnston D, Woolcock B, Vielkind J, Nairn R. 1999. Comparative structure and characterization of a *CDKN2* gene in a *Xiphophorus* fish melanoma model. *Oncogene* 18:5088–5099.

Kazianis S, Morizot DC, McEntire BB, Nairn RS, Borowsky RL. 1996. Genetic mapping in *Xiphophorus* hybrid fish: assignment of 43 AP-PCR/RAPD and isozyme markers to multipoint linkage groups. *Genome Res.* 6:280–289.

Kosswig C. 1928. Uber Kreuzungen zwischen den Teleostiern Xiphophorus helleri und Platypoecilus maculatus. *Z Indukt Abstammungs- Vererbungsl.* 47:150–158.

Mayrose I, Otto SP. 2010. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Mol Biol Evol.* 28: 759–770.

Meierjohann S, Schartl M. 2006. From Mendelian to molecular genetics: the *Xiphophorus* melanoma model. *Trends Genet.* 22: 654–661.

Merbs SL, Sidransky D. 1999. Analysis of p16 (CDKN2/MTS-1/INK4A) alterations in primary sporadic uveal melanoma. *Invest Ophthalmol Vis Sci.* 40:779–783.

Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 1995. 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med.* 1:686–692.

Meyer A, Morrissey J, Schartl M. 1994. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature* 368:539–542.

Meyer A, Salzburger W, Schartl M. 2006. Hybrid origin of a swordtail species (Teleostei: *Xiphophorus clemenciae*) driven by sexual selection. *Mol Ecol.* 15:721–730.

Muller H. 1942. Isolating mechanisms, evolution and temperature. *Biol Symp.* 6:71–125.

Nei M, Nozawa M. 2011. Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. *Genome Biol Evol.* 3: 812–829.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc R Soc Lond B.* 255:37–45.

Pond SLK, Frost SDW. 2004. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol.* 22: 478–485.

Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nat Rev Genet.* 11:175–180.

Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between. *Nature* 423:715–719.

Regneri J, Schartl M. 2011. Expression regulation triggers oncogenicity of xmrk alleles in the *Xiphophorus* melanoma system. *Comp Biochem Physiol C Toxicol Pharmacol.* 155:71–80.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.

Rosenthal GG, Garcia-De-Leon FJ. 2011. Speciation and hybridization. In: Schlupp I, Pilastro A, Evans J, editors. Ecology and evolution of poeciliid fishes. Chicago (IL): University of Chicago Press. p. 109–119.

Schartl M. 1990. Homology of melanoma-inducing loci in the genus *Xiphophorus*. *Genetics* 126:1083–1091.

Schartl M. 2008. Evolution of Xmrk: an oncogene, but also a speciation gene? *Bioessays* 30:822–832.

Schartl M, Walter RB, Shen Y, et al. (19 co-authors). 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet.* 45: 567–572.

Schluter D, Conte G. 2009. Genetics and ecological speciation. *Proc Natl Acad Sci U S A.* 106:9955–9962.

Schumer M, Cui R, Boussau B, Walter RB, Rosenthal GG, Andolfatto P. 2012. An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. *Evoluion* 67:1155–1168.

Sugiura N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun Stat Theor Meth*. A7: 13–26.

Weis S, Schartl M. 1998. The macromelanophore locus and the melanoma oncogene Xmrk are separate genetic entities in the genome of *Xiphophorus*. *Genetics* 149:1909–1920.

Wittbrodt J, Adam D, Malitschek B, Mäueler W, Raulf F, Telling A, Robertson S, Schartl M. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. *Nature* 341:415–421.

Wu CI, Ting CT. 2004. Genes and speciation. *Nat Rev Genet*. 5: 114–122.