Oncogenic allelic interaction in *Xiphophorus* highlights hybrid incompatibility

Yuan Lu,a,b,1 Angel Sandovala,c, Sarah Vossa,c, Zhao Lai,b,c, Susanne Kneitzd, Will Boswellb, Mikki Boswella, Markita Savagea, Christi Walterc, Wes Warrena, Manfred Schartla,c, and Ronald Waltera,b1

*The Xiphophorus Genetic Stock Center, Texas State University, San Marcos, TX 78666; bDepartment of Molecular Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229; cGreehey Children’s Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229; dPhysiological Chemistry, Biozentrum, University of Würzburg, 97070 Würzburg, Germany; eDepartment of Cellular & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229; fBond Life Science Center, University of Missouri, Columbia, MO 65211; and gDevelopmental Biochemistry, Biozentrum, University of Würzburg, 97070 Würzburg, Germany*

Edited by Harmit S. Malik, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved October 9, 2020 (received for review May 28, 2020)

Mixing genomes of different species by hybridization can disrupt species-specific genetic interactions that were adapted and fixed within each species population. Such disruption can predispose the hybrids to abnormalities and disease that decrease the overall fitness of the hybrids and is therefore named as hybrid incompatibility. Interspecies hybridization between southern platyfish and green swordtails leads to lethal melanocyte tumorigenesis. This occurs in hybrids with tumor incidence following progeny ratio that is consistent with two-locus interaction, suggesting melanoma development is a result of negative epistasis. Such observations make *Xiphophorus* one of the only two vertebrate hybrid incompatibility examples in which interacting genes have been identified. One of the two interacting loci has been characterized as a mutant epidermal growth factor receptor. However, the other locus has not been identified despite over five decades of active research. Here we report the localization of the melanoma regulatory locus to a single gene, rab3d, which shows all expected features of the long-sought oncogene interacting locus. Our findings provide insights into the role of *egfr* regulation in regard to cancer etiology. Finally, they provide a molecular explainable example of hybrid incompatibility.

hybrid incompatibility | evolution | genetics | Bateson–Dobzhansky–Muller model | Xiphophorus

In the late 1920s, three investigators, Myron Gordon, Georg Haussler, and Kurt Kosswig, independently found that hybrids between two distant *Xiphophorus* fish species, *Xiphophorus maculatus* (southern platyfish), and *Xiphophorus hellerii* (green swordtail), develop spontaneous and lethal pigment cell tumors that were later determined to be melanoma (1–3). Since its establishment, this model system has been intensively studied to assess the underlying genetic contributions to tumor etiology. The development of hybridization-induced tumor has been viewed as a representation of the genome incompatibility hypothesis known as the Bateson–Dobzhansky–Muller (BDM) model (4–8). The BDM model states that negative epistatic interactions in hybrids serve as the molecular genetic mechanisms underlying genome incompatibility and is associated with problems in hybrid fitness. Although BDM incompatibility was identified in a few model organisms (9), *Xiphophorus* and mice represent the only vertebrate systems for which this model is established (10, 11). This *Xiphophorus* interspecies hybrid system, now termed the “Gordon–Kosswig–Anders (GKA) model,” was first described by Gordon and Kosswig in 1929s. This model employs crossing of *X. maculatus* and *X. hellerii* to produce F1 interspecies hybrids. *X. maculatus* exhibits a nevoid-like pigmentation pattern in its dorsal fin (spotted dorsal, *Sd*), while *X. hellerii* does not exhibit this trait. In the F1 hybrid, the *Sd* pigmentation pattern becomes expanded, with melanin pigmentation covering the entire dorsal fin due to melanocyte hyperplasia (12, 13). Backcrossing the F1 hybrid to the *X. hellerii* parent leads to three distinct phenotypes among the backcross (BC) progeny that follow Mendelian distributions: 25% of hybrids exhibit hyperplasia of pigmentation pattern as observed in the F1 hybrid, 25% exhibit lethal and invasive nodular exophytic melanoma, and the remaining 50% of progeny do not display a black pigmentation pattern (14) (Fig. 1). The hybridization-induced disease observed in *Xiphophorus* interspecies hybrid represents a type of genetic incompatibility. In 1950s, Anders argued this spontaneous tumorigenesis is due to segregation of two loci from *X. maculatus*; one was named *Tu* for “tumor” and another locus was named *R* for “repressor” or *Diff* for “differentiation” [hereafter referred to as *R(Diff)*]. These concepts led to what we now know as oncogenes and tumor suppressors (15, 16). In the late 1980s, it was shown the *Tu* gene encodes a mutant duplicate copy of *egfr*, and this gene was named as *Xiphophorus melanoma regulatory kinase* (*xmrk*) (17, 18). The *xmrk* oncogene is tightly linked to or part of the *Sd* locus and controls melanocyte proliferation.

In addition, the GKA model offers us a natural two-hit melanoma model wherein the oncogenic effect of *xmrk* can be fully eliminated by a regulatory locus that must have coevolved with *xmrk* (19). The *EGFR* gene is one leading oncogene of many human cancers (20). It is preproliferative and is an upstream activator of BRAF and NRAS signaling, which are the driver oncogenes in over 50% of all human melanomas (21). This evidence promotes the characterization of *R(Diff)* as having significant implications in cancer etiology. Therefore, identifying the *R(Diff)* gene will highlight the genetic interactions underlying

Significance

The Bateson–Dobzhansky–Muller (BDM) model describes negative epistatic interactions that occur between genes with a different evolutionary history to account for hybrid incompatibility and is a central theory explaining genetic mechanisms underlying speciation. Since the early 1900s when the BDM model was forwarded examples of BDM incompatibility have been described in only a few nonvertebrate cases. This study reports the only vertebrate system, with clearly defined interacting loci, that supports the BDM model. In addition, this study also poses that tumorigenesis serves as a novel mechanism that accounts for postzygotic isolation.


The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See online for related content such as Commentaries.

1To whom correspondence may be addressed. Email: y_l54@txstate.edu or rw12@txstate.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2010133117/-/DCSupplemental.

First published November 9, 2020.
the tumor induction due to hybrid incompatibility and in addition may forward novel molecular target(s) in regulation of EGFR function for human disease control.

However, the \( R(Diff) \) gene has not been identified. Previous effort to define the \( R(Diff) \) locus forwarded a chromosomal region that includes cdkn2ab, an ortholog of the human tumor suppressor genes CDKN2A and B, which is mutant in 10% of high-melanoma-risk families (22–27). In this study, we found cdkn2ab is not the \( R(Diff) \) gene but is tightly linked to it. More importantly, we have identified the long-hypothesized \( R(Diff) \) locus to a strong candidate gene, \(\text{rub3d}d\).

**Results**

**A Region on Chromosome 5 Determines Malignancy of Melanocytic Lesions.** Hybrids between \( X.\maculatus \) and \( X.\hellerii \) exhibit enhanced pigmentation in their dorsal fin (i.e., melanocyte hyperplasia) due to hemizygosity for the \( xmrk \) oncogene and heterozygosity of the \( R(Diff) \) locus [i.e., \( xmrk^{X.\maculatus} ; R(Diff)^{X.\maculatus/X.\hellerii} \)]. This phenotype and genotype are also present in \( \sim 25\% \) of BC hybrid progeny (Fig. 1). Due to meiotic recombination, successive backcrossing of such animals to \( X.\hellerii \) should stepwise reduce heterozygosity in the advanced BC progeny and finally result in a \( xmrk^{X.\maculatus} ; R(Diff)^{X.\maculatus/X.\hellerii} \) isogenic line (i.e., introgression). In this manner, advanced BC (BC
\( n \)) fish were produced in order to validate the candidate gene cdkn2ab on chromosome 5, 15.8 Mbp) as a locus carrying \( R(Diff) \). The parential paternal (i.e., interspecies hybrid) for each successive BC generation was genotyped for inheritance of \( xmrk^{X.\maculatus} ; cdkn2ab^{X.\maculatus/X.\hellerii} \) exhibiting the enhanced dorsal fin pigmentation was selected for a next round of backcrossing. \( \chi^2 \) tests were performed on genotyping data of each variant site collected from 90 BC
\( n \) hybrids that exhibited melanocyte hyperplasia to locate genes that displayed ancestral allele linkage disequilibrium and predominately showed a heterozygous inheritance pattern (SI Appendix, Figs. S4 and S5). Two genomic regions were found to correlate to the hyperplasia phenotype (Fig. 2): one on chromosome 21 which encompasses \( xmrk \), the melanoma driver oncogene that induces melanocyte proliferation, and a second region on chromosome 5 that corresponds to a previously mapped \( R(Diff) \) region of 5.8-Mbp region harboring the candidate \( cdkn2ab \) gene (Fig. 2B). An average loss of 50% heterozygous loci per BC generation is expected. Therefore, one expects to see a heterozygous region only accounting for an average of 5.5 and 2.7 Mbp of the 700-Mbp genome for BC
\( r \) and BC
\( n \) individuals, respectively. However, haploid maps produced from these advanced BC animals show the heterozygous content to be much higher than this expectation. In addition, heterozygous loci are predominantly surrounding the \( cdkn2ab \) region (SI Appendix, Figs. S3 and S6), suggesting selection of individuals that exhibited melanocyte hyperplasia and a genotype of \( cdkn2ab^{X.\maculatus/X.\hellerii} \) for further backcrossing coselected an adjacent locus on chromosome 5, rendering higher-than-expected heterozygosity in BC
\( n \) hybrids. These observations indicate that \( cdkn2ab \) itself is less likely to be \( R(Diff) \), while the coselected locus with \( cdkn2ab \) is the true \( R(Diff) \).

**Genetic Mapping of a Mutant EGFR Regulator Locus.** To locate the \( R(Diff) \) candidate gene(s) independent of artificial selection of genetic marker, we produced BC interspecies hybrid progeny of the GKA model, that is, \( X.\hellerii \) (Rio Sarabia) \( \times [X.\maculatus \ Jp163 \ A \times X.\hellerii \ (Rio \ Sarabia)] \) and performed targeted genomic sequencing on BC progeny that developed two distinct melanocyte phenotypes: benign hyperplasia and melanoma tumor. Association analyses between parental allele inheritance and melanocyte phenotypes was observed, and only heterozygous hybrids (i.e., \( xmrk^{X.\maculatus} ; cdkn2ab^{X.\maculatus/X.\hellerii} \)) exhibiting the enhanced dorsal fin pigmentation were selected for a next round of backcrossing. 

![Fig. 1. GKA model crossing scheme. The crossing scheme shows the Xiphophorus species used to produce F1 and BC1 interspecies hybrids. X. maculatus Jp163 A and X. hellerii are used to produce F1 hybrids artificially. The F1 hybrids are subsequently backcrossed to X. hellerii to produce BC hybrid progeny. BC hybrids exhibiting melanocyte hyperplasia and heterozygous R(Diff) are used as parents for next-generation BC.

Lu et al. PNAS | November 24, 2020 | vol. 117 | no. 47 | 29787
as expected (Fig. 3A). Genotypes of four linked polymorphic sites (10,582,852, 10,582,855, 10,582,868, and 10,582,870) on chromosome 5 are the most significantly correlated to melanocyte phenotypes, where individuals exhibiting benign pigment cell lesions inherited both parental alleles, while melanoma-bearing individuals only inherited the X. helleri alleles (Fig. 3B).

Haploid maps produced from all BC individuals supported the result of association analyses and forwarded a region (i.e., 10,515,844 to 10,617,563) that is free of chromosomal cross-overs in both cohorts (Fig. 3B). The R(Diff) locus is assigned to this 101.7-kbp region (Fig. 3A). As expected, this locus maps to the vicinity (5.2 Mbp) upstream of cdkn2ab.

![Image](https://via.placeholder.com/150)

**Fig. 2.** Genetic mapping of heterozygous loci in advanced BC hybrids. A total 90 BC hybrids (BC1 to BC90) that exhibited dorsal fin melanocyte benign hyperplasia and were produced by crossing cdkn2ab genotyped hybrid (cdkn2abX. hellerii) with X. helleri (cdkn2abX. maculatus). (A) Manhattan plot showing −log10 P value (χ2 test) across the genome. The y axis represents −log10 P value and the x axis represents amplicon chromosomal coordinates, which are labeled as red or blue. Only the −log10 P value of loci that exhibited higher X. maculatus allele frequency is plotted due to introgression. The light blue dashed line represents a P value of 0.01 that is suggestive of statistical significance. χ2 test P values were corrected using Bonferroni method across the genome-wide data. The red dashed line represents adjusted P value of 0.05 corresponding to 3.6 × 10−5. (B) A zoom-in view of the chromosome 5 10,029,746-15,895,617-bp region. This region is highly correlated to the pigmentation phenotype observed in BC hybrids.

The **rab3d Gene Is the Functional Carrier of the R(Diff) Function.** The 101.7-kbp R(Diff) locus encodes three gene models on the reverse strand: differential screening-selected gene aberrative in neuroblastoma (DAN) domain family member 5 (dnad5), tetraspanin-1 (tspan1), and ras-related protein Rab-3D (rab3d). Expression of a gene is the prerequisite for display of genetic function. To determine the gene expression pattern of these three candidate genes and to assess gene expression changes in parental nevus-like dorsal fin melanocyte spots, melanocyte hyperplasia, and melanoma tumor, transcription profiling was performed on these tissues. The rab3d gene is the only gene expressed among the three candidates in melanocyte spots of parental X. maculatus (three pools of dorsal fin spots), dorsal fin melanocyte hyperplasia (n = 22), and melanoma tumor (n = 22) of BC interspecies hybrids (Fig. 4). Therefore, rab3d serves as the only gene that can carry the R(Diff) function.

In addition, differential expression analysis of rab3d among the parental dorsal fin spots, interspecies hybrid melanocyte expansion, and in melanoma showed rab3d expressed at the same level between the melanocyte spots and hyperplastic pigmented cell lesions (adjusted P value = 0.81; Fig. 4), and expressed at a higher level in melanocyte hyperplasia (adjusted P value = 5.4 × 10−5) than in melanoma of BC fish.

The two parental rab3d alleles differ by an Asn residue in X. maculatus and Lys in X. helleri at the C-terminus tail downstream of the P-loop domain that harbors the guanosine triphosphatase (GTPase) activity (i.e., Asn/Lys-204; Fig. 5). Although this Asn/Lys site locates in the hypervariable C terminus of rab3d, comparative genomics showed that different from X. maculatus the Lys is conserved in human and 60% of all fish species analyzed (i.e., the next dominating allele is Arg) that include another xmrk-null *Xiphophorus* species, *Xiphophorus couchianus,* and other Poeciliidae fish, suggesting the *X. maculatus* allele is a genetic outlier (Fig. 5 and Dataset S1).

**Discussion**

The finding that rab3d is the only expressed gene located in the tumorigenesis-determining locus on chromosome 5 in *Xiphophorus* genome, forwarded by both genetic mapping and transcriptomics, concludes a search for the hypothetical locus R(Diff) that has been ongoing for over five decades (22, 25, 28-31). As a viviparous fish where embryonic development occurs within the female, the technical hurdles involved in transgenesis have not yet been overcome for *Xiphophorus*. Therefore, genetic manipulation...
of rab3d cannot be performed in the Xiphophorus system until our current development of Xiphophorus transgenesis is proven successful and efficient. Although a readily available system for such a test is currently not available, the negative epistasis between xmrk and rab3d hallmarks cancer, in addition to hybrid lethality (9) and sterility (32), an innovative mechanism for decreasing hybrid fitness, and reinforcing speciation.

Epistasis underlying human disease can be elucidated by investigating mechanisms that “evolutionary mutant models” developed to cope with similar mutations as in human disease and/or produce adaptive phenotypes that are similar to human disease (19, 33). X. maculatus is one such species where oncogenicity of a mutant EGFR (i.e., xmrk) is compromised by a regulatory allele R(Diff). EGFR is one of the most prevalent oncogenes exhibiting mutation and/or dysregulation in many varied human cancers (34–53). Despite over 40 y of effort in attempting to inhibit EGFR by blocking its kinase activity, and development of four generations of small molecules, or monoclonal antibodies, success in disease control is very limited to three cancer subtypes, that is, nonsmall cell lung cancers with kinase-activating mutations in EGFR (54–56), ~10% of metastatic colorectal cancers (57, 58), and a subcategory of advanced head and neck cancers (59, 60). Current methodologies to inhibit EGFR attempt to directly block the adenosine 5′-triphosphate (ATP) binding pocket or ligand binding pocket or by targeting acquired mutations that lead to acquired resistance have all turned out to be inefficient in providing promising therapeutic benefit. Therefore, it is of utmost importance to revolutionarily reattack the question of EGFR-associated cancer etiology and identify reliable next-generation treatment strategies that enable disease control with higher response rates and lower resistance. The xmrk, originated from a gene duplication event (61), encodes a mutant EGFR that autodimerizes and activates downstream proliferative pathways in a ligand-independent manner (62). The xmrk gene is a bona-fide oncogene because its ectopic expression leads to transformation and tumorigenesis of melanocytes in murine cells and medaka fish (63, 64). However, X. maculatus does not exhibit tumorigenesis, suggesting X. maculatus harbors a mechanism [i.e., R(Diff)] in suppressing the driver oncogene. Therefore, the identification of rab3d as gene
exhibiting R(Diff) regulatory function delineate mechanisms of how _X. maculatus_ counteracts the deleterious EGFR mutant in its genome. The molecular mechanism underlying _rab3d_ suppression of _xmrk_ can lead to innovative strategies in developing next-generation EGFR inhibitors.

The _rab3d_ genes showed both codon mutation between the two parental alleles and transcriptional differences between normal dorsal fin pigment cells, pigment cell hyperplasia, and melanoma tumor. These results suggest both structural and expression divergences between the two parental alleles of _rab3d_ are essential to elucidate molecular interactions we can learn in order to advance our knowledge in control of EGFR and associated disease. There are two questions to answer regarding _rab3d_ regulation of _xmrk_. First, what is the molecular nature of this _xmrk–rab3d_ interaction? _RAB3D_ is a ras-related small G protein with GTPase activity controlling exocytosis (65–67) and has been shown to regulate secretion of a broad range of molecules in different cell types (68–71). For example, _RAB3D_-dependent secretion of matrix metalloproteinase in macrophages is a prerequisite for macrophage recruitment to tumor cells (68), while secretion of the same molecule by tumor cells is a signal for tissue invasion and metastasis (72). We hypothesize that the _X. maculatus_ allele of _rab3d_ regulates cancer cell invasion by either mediating immune cell recruitment to tumor microenvironment or by hampering tumor cell secretion of molecules that facilitate metastasis. However, _RAB3D_ function in cancer has not been clearly characterized despite efforts made investigating its function in cancer cell proliferation and metastasis in vitro (73–83). Currently there is no understanding of the cell population within the _Xiphophorus_ tumor microenvironment (e.g., tumor cells, endothelial cells, cancer-associated fibroblasts, and immune cells) where expression of different parental _rab3d_ alleles may change the fundamental outcome of _xmrk_ expression (84, 85). _RAB3D_ is also involved in cell membrane–associated protein dynamics. It has been shown that human _RAB3D_ interacts with GOLM1 and selectively assist cytoplasmic EGFR recycling to cell membrane (83). Therefore, it is our second hypothesis that _rab3d_ regulates _xmrk_ function by directly controlling _xmrk_ protein turnover and cellular localization where it displays its full activity. Second, what functionality change is associated with the amino acid difference between the two parental alleles? The change of Asn204 in _X. maculatus_ to Lys204 in _X. hellerii_ is located in the C terminus of _RAB3D_, where proper modifications (i.e., methylation and geranylgeranylation) are required for subcellular localization (86, 87). The Lys (_X. hellerii_) and Asn (_X. maculatus_) exhibit both physical differences (i.e., Lys is charged and Asn is uncharged) and varied posttranslational modifications. This amino acid change may alter protein hydrophobicity, affect RAB3D subcellular localization, and hinder efficient transportation of secretion granules and eventually affect on-site and dynamics of the above mechanisms. In summary, the conclusion of epistasis underlying the interspecific hybridization-induced tumorigenesis provides insights into a strategy in countering detrimental conditions.

Overall, vertebrate organisms that support the molecular mechanism proposed by the BDM model is only limited to mice (32) and now _Xiphophorus_ fishes. The discovery of _xmrk–rab3d_ genetic interaction underlying spontaneous tumorigenesis in interspecies hybrids poses an example showing that hybrid-induced disease can act as a mechanism that reduces hybrid fitness. Characterizing mechanism of RAB3D functional regulation of EGFR can lead to development of innovative EGFR regulation strategy.

**Materials and Methods**

**Animal Model.** _X. maculatus Jp163 A, X. hellerii (Rio Sarabia),_ and first-generation _BC_ fish were used in this study were supplied by the _Xiphophorus_ Genetic Stock Center (https://www.xiphophorus.txstate.edu). _X. maculatus_ Jp163A strain female fish were artificially inseminated with sperm from male _X. hellerii_ (Rio Sarabia strain) to produce _F_1 interspecies hybrids. _F_1 hybrid males were then backcrossed to _X. hellerii_ females to generate the _BC_ 1 animal. At dissection, all fish were anesthetized in an ice bath and upon loss of gill movement were killed by cranial resection. Organs were dissected into RNAlater (Ambion Inc.) and kept at −80 °C until use. All _BC_ 1 fish were kept and samples taken in accordance with protocols approved by the Texas State University Institutional Animal Care and Use Committee (IACUC 201510711).

The advanced _BC_ interspecies hybrids (i.e., _BC_ through _BC_ 4) were produced in an independent series of crosses in the Biocenter of the University of Wurzburg, Germany. _F_ 2 _BC_ interspecies hybrids originated from the reciprocal cross: _X. maculatus_ Jp163A males were mated to _X. hellerii_ (Rio Lancerilla strain) females. The _F_ 1 hybrid females (_cdkn2abX. mac/X. hel_ genotype) were then successively backcrossed to _X. hellerii_ males to produce the advanced-generation of _BC_ hybrids for each generation of backcrossing, only the interspecies hybrid _BC_ fish that exhibited benign pigment cell hyperplasia and had the _cdkn2abX. mac/X. hel_ genotype were used to produce the next generation of _BC_ progeny. _F_ 1 to _BC_ 4 fish were collected from all _advanced_ _BC_ fish and stored in ethanol at 4 °C. All _advanced_ _BC_ fish were kept and samples taken in accordance with the applicable European Union and national German legislation governing animal experimentation. When needed, fish were killed by overanesthetization with MS222. These experiments were performed under authorization (568/300-187013/1) of the Veterinary Office of the District Government of Lower Franconia, Germany, in accordance with the German Animal Protection Law (TierSchG).

**DNA and RNA Isolation.** Fin clip, or muscular tissue, was digested by Protease K at room temperature for 1 h. The lysate was then transferred to 2.0-ml collection tubes. DNA isolation was performed by a QiAmp HT (Qiagen) automated biosample isolation system, with reagent contained in the QiAmp 96 DNA QiAcube HT Kit. The isolation system is equipped with a robotic arm with eight pipettes. Each pipette is able to pick and eject pipette tips, self-clean, and transfer liquids between wells/columns, or between master reservoirs and
wells/columns in standard 96-well plate formats. Each sample was independently maintained throughout the isolation process. Concentrations of DNA samples were measured using Qubit 2.0 fluorometer (Life Technologies) and adjusted for sequencing library preparation.

Dorsal fin spots, dorsal fin exhibiting benign hyperplasia, and melanoma tumors were excised from *X. maculatus* Jp163A, and BC interspecies hybrid. Tissue samples were homogenized in TRI-reagent (Sigma Inc.) followed by addition of 200 μL/mL chloroform, vigorously shaken, and subjected to centrifugation at 12,000 g for 5 min at 4 °C. Total RNA was further purified using a RNeasy mini RNA isolation kit (Qiagen). Column DNase digestion at 25 °C for 15 min removed residual DNA. Total RNA concentration was determined using a Agilent 2100 Bioanalyzer (Agilent Technologies) to confirm that RNA integrity number scores were above 8.0 prior to subsequent gene expression profiling.

**Genetic Variants Identification and Annotation.** To identify interspecies polymorphisms between the *X. maculatus* and *X. hellerii*, genomic DNAs of 4 *X. maculatus* and 4 *X. hellerii* were isolated. DNA samples were forwarded for genome shotgun sequencing library preparation using Illumina Nextera sequencing Library Prep Kit, followed by sequencing on HiSeq, 2000 (Illumina, Inc.) using 150-bp paired-end sequencing protocol. Raw sequencing reads were trimmed and filtered using a custom Perl script, and adapter sequences were removed from the sequencing reads. The reads were truncated based on similarity to library adaptor sequences using custom Perl scripts. Then, low-scoring sections of each read were removed, preserving the longest remaining sequencing read fragment (88). Filtered genome sequencing reads were mapped to the reference *X. maculatus* genome (GenBank assembly accession no. GCA_002775205.2) using Bowtie2 “head-to-head” mode to show mismatches between sequenced animals and reference genome (89). Alignment files were sorted using Samtools (90). Subsequently, pileup files were generated for each *X. maculatus* and *X. hellerii* sample, and variant calling was processed by both BCTools and VarScan for polymorphisms detection, with minimum variant locus coverage of 2 and a Phred score for variant detection of 0.05 for VarScan and variant genotyping call Phred score ≥20 for BCTools (90–92). Only the variants that were identified by both pipelines were forwarded for further analyses.

To localize fixed variants between the *X. maculatus* and *X. hellerii*, homozygous loci of *X. maculatus* were compared to those of *X. hellerii*. Such loci were identified if all *X. maculatus* were homozygous for one allele and all *X. hellerii* homozygous for the alternative allele.

These fixed species-specific genetic variants were functionally annotated using snpEff (93). A genome database was created using the *X. maculatus* genome sequence and annotation files (GenBank accession no. GCA_002775205.2). Each variant was queried to the genome database to determine if it was located in a genetic or intergenic region, and to determine what effect each variant may have on the peptide sequence structure.

**Amplicon Sequencing, Data Filter, and Genotyping.** Variants between *X. maculatus* and *X. hellerii* were used as references to design specific capture probes for targeted genomic sequencing. Variants with very high sequencing depth were removed due to the possibility of locating them in repetitive sequences. Sequencing probes were designed to amplify regions surrounding genetic variants. To genetically map candidate *R(Diff)* loci in a region (chromosome 5: 10,000,000 to 16,000,000) identified in a previous study (22), 406 sets of probe were designed to reach a resolution of 14.8 kbp within the 6-Mbp locus, for sex-determining regions (chromosome 21: 23,750,000 to 26,250,000; SI Appendix, Fig. S2 and Dataset S2), 101 sets or capture probes were designed for 24.8-kbp definition of genetic mapping; for the rest of the genome, 1,510 sets of probes were designed for genotyping and establish individual BC progeny haploid map at definition of 459 kbp (SI Appendix, Fig. S1 and Dataset S2). Therefore, a total of 2,017 probe sets were produced for amplicon sequencing. Amplicons were custom-made using Illumina Genotype Ne library preparation kit, with i7 and i5 indices incorporated into adaptor sequences added to each end of PCR products amplified by capture probes. Sequencing libraries were sequenced on Illumina MiSeq platform employing a 75-bp paired-end sequencing strategy (Illumina).

Sequencing adaptor contamination was first removed from raw sequenced reads using fastx toolkit, followed by trimming of low-quality sections of each sequencing read. Low-quality sequencing reads were further removed from sequencing result (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Processed sequencing reads were mapped to *X. maculatus* genome v5.0 (GenBank assembly accession no. GCA_002775205.2) using Bowtie2 (89). Mpileup files were made using samtools and genotyping was processed using both bcftools and VarScan (90–92). Genetic variant call and genotype were required to be supported by both pipelines for further analyses (i.e., BcTools: MAPQ ≥30, Phred score of genotype call ≥0, with alternative genotype call Phred score ≥20; VarScan: MAPQ ≥30, P-Value ≤0.05, depth ≥20). Herein “genotype” refers to inheritance of ancestral alleles, with “heterozygous” meaning that a locus exhibited genetic material from both ancestors (i.e., *X. maculatus* and *X. hellerii* for the parental mapping “A locus exhibits genetic material from only the recurrent ancestor” (i.e., *X. hellerii*). A haploid map was produced for each individual of BC progeny. To control the amplicon-sequencing-based genotyping result target specificity, only genotyping calls that locate less than 10 kbp away from previously identified loci were used in the mapping.

**Fig. 5.** Sequence alignment of rab3δ genes. Protein sequence comparison between *X. hellerii* rab3δ and *X. maculatus*, *X. couchianus*, and Poeciliid fish *P. mexicana*. Different amino acids between *X. maculatus* and *X. hellerii* are labeled in red and black, respectively. Functional domains are labeled with black lines underneath protein sequences, with thinner lines linked functional domain in tertiary structure. The only amino acid change between *X. maculatus* to other *Xiphophorus* is Lys-204 > Asn-204 mutation. The Lys is conserved, and only *X. couchianus* has a change to Asn. These fixed species-specific genetic variants were functionally annotated using snpEff (93). A genome database was created using the *X. maculatus* genome sequence and annotation files (GenBank accession no. GCA_002775205.2). Each variant was queried to the genome database to determine if it was located in a genetic or intergenic region, and to determine what effect each variant may have on the peptide sequence structure.

**Amplicon Sequencing, Data Filter, and Genotyping.** Variants between *X. maculatus* and *X. hellerii* were used as references to design specific capture probes for targeted genomic sequencing. Variants with very high sequencing depth were removed due to the possibility of locating them in repetitive sequences. Sequencing probes were designed to amplify regions surrounding genetic variants. To genetically map candidate *R(Diff)* loci in a region (chromosome 5: 10,000,000 to 16,000,000) identified in a previous study (22), 406 sets of probe were designed to reach a resolution of 14.8 kbp within the 6-Mbp locus, for sex-determining regions (chromosome 21: 23,750,000 to 26,250,000; SI Appendix, Fig. S2 and Dataset S2), 101 sets or capture probes were designed for 24.8-kbp definition of genetic mapping; for the rest of the genome, 1,510 sets of probes were designed for genotyping and establish individual BC progeny haploid map at definition of 459 kbp (SI Appendix, Fig. S1 and Dataset S2). Therefore, a total of 2,017 probe sets were produced for amplicon sequencing. Amplicons were custom-made using Illumina Genotype Ne library preparation kit, with i7 and i5 indices incorporated into adaptor sequences added to each end of PCR products amplified by capture probes. Sequencing libraries were sequenced on Illumina MiSeq platform employing a 75-bp paired-end sequencing strategy (Illumina).

Sequencing adaptor contamination was first removed from raw sequenced reads using fastx toolkit, followed by trimming of low-quality sections of each sequencing read. Low-quality sequencing reads were further removed from sequencing result (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Processed sequencing reads were mapped to *X. maculatus* genome v5.0 (GenBank assembly accession no. GCA_002775205.2) using Bowtie2 (89). Mpileup files were made using samtools and genotyping was processed using both bcftools and VarScan (90–92). Genetic variant call and genotype were required to be supported by both pipelines for further analyses (i.e., BcTools: MAPQ ≥30, Phred score of genotype call ≥0, with alternative genotype call Phred score ≥20; VarScan: MAPQ ≥30, P-Value ≤0.05, depth ≥20). Herein “genotype” refers to inheritance of ancestral alleles, with “heterozygous” meaning that a locus exhibited genetic material from both ancestors (i.e., *X. maculatus* and *X. hellerii* for the parental mapping “A locus exhibits genetic material from only the recurrent ancestor” (i.e., *X. hellerii*). A haploid map was produced for each individual of BC progeny. To control the amplicon-sequencing-based genotyping result target specificity, only genotyping calls that locate less than 10 kbp away from previously identified loci were used in the mapping.
75 bp from the designed polymorphic sites and also supported by at least another variant genotyping call within a 75-bp range were kept. Qualified genotyping calls were subsequently ordered by chromosome name and their chromosomal location to produce the haploid map. For each BC hybrid individual, percentage of heterozygous loci is calculated as Heterozygous % = (number of heterozygous loci)/(total number of genotyped loci). Because BC, were selected for a marker on chromosome 5 (i.e., cdkn2ab) and chromosome 21 (i.e., Sd), percentage of heterozygous was calculated using all genotyped loci, or loci that are outside of chromosomes 5 and 21 (SI Appendix, Fig. S3).

Calculation of Allele Frequency. Genotyped variants loci of all BC hybrids were combined together first to yield a data table with the rows as chromosomal coordinates and columns as hybrid individuals. Heterozygous, designated as “0/1,” is referred to inheritance of both X. maculatus and X. hellerii parental alleles, while homozygous, designated as “1/1,” is referred to inheritance of only X. hellerii alleles. The X. maculatus allele frequency for each locus is calculated as

\[ p_{\text{maculatus}} = \frac{\sum \text{heterozygous}}{2} \left( \frac{\sum \text{heterozygous}}{\sum \text{heterozygous}} + \sum \text{homozygous} \right) \]

Linkage Analyses. For advanced BC samples exhibiting benign melanocyte hyperplasia their genetic backgrounds are predominantly represented by the recurrent parental genome (X. hellerii). An average X. hellerii genome component per BC generation follows a rule determined by (1 − 0.5^−n), where n equals the BC generation. Therefore, every locus is expected to exhibit dominance of X. hellerii allele and therefore disequilibrium (SI Appendix, Fig. S4). Since our previous studies had determined that the R(Diff) locus is heterozygous within the BC hybrids exhibiting melanocyte benign hyperplasia, we only plotted the –log10 P values of loci where X. maculatus allele frequency is higher and assigned –log10 P values of X. hellerii dominated loci arbitrarily to 0, in order to visualize dominantly heterozygous loci within the BC hybrids (SI Appendix, Fig. S5). Because there is only one pigment cell phenotype of the BChybrids, numbers of heterozygous and homozygous individuals per variant site were used to form a one-dimensional contingency table and tested using a goodness-of-fit χ² test. χ² test P values were adjusted using Bonferroni correction.

For BC1, allele frequencies of heterozygous and homozygous individuals were counted for each pigmentation phenotype (i.e., melanoma and pigment cell hyperplasia), and numbers of each genotype per phenotype group (i.e., tumor or pigment cell hyperplasia) were used to form a contingency table and subsequently tested using a χ² contingency table test, with the null hypothesis that both their distributions follow random assortment. χ² test P values were adjusted using Bonferroni correction.

Gene Expression Profiling. RNA sequencing was performed upon sequencing libraries construction using the Illumina TruSeq messenger RNA (mRNA) library preparation kit (Illumina, Inc.). RNA libraries were sequenced as 125-bp paired-end fragments using an Illumina Hi-Seq. 2000 system (Illumina, Inc.). Two-tailed t tests were used to test if sizes of BC7 (n = 23) and BC1 (n = 66), respectively, χ² test P values were adjusted using Bonferroni correction.

Statistical Information. Goodness-of-fit and contingency table χ² tests were used to identify linkage disequilibrium for advanced BC (n = 90) and BC1 hybrids (n = 66), respectively. χ² test P values were adjusted using Bonferroni correction.

Data Availability. Sequencing data have been deposited in the NCBI Sequence Read Archive (BioProject number PRJNA610525, accession nos. SAMN14300088–SAMN14300223).
93. Cingolani et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92 (2012).