Springer

Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- For **fax** submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Please return your proof together with the **permission to publish** confirmation.
- Remember to note the journal title, article number, and your name when sending your response via e-mail, fax or regular mail.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.

Please note

Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible**.

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].

If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us, if you would like to have these documents returned.

The **printed version** will follow in a forthcoming issue.

Fax to: +44 207 806 8278 or +44 870 762 8807 (UK)
or +91 44 4208 9499 (INDIA)

To: **Springer Correction Team**
6&7, 5th Street, Radhakrishnan Salai, Chennai, Tamil Nadu, India – 600004

Re: Genetica DOI:10.1007/s10709-008-9284-4
Divergent patterns of selection on the DAB and DXB MHC class II loci in Xiphophorus fishes

Authors: Kyle Summers · KellyE. Roney · Jack Silva · Gerald Capraro · BrandonJ. Cuthbertson · Steven Kazianis · GilG. Rosenthal · MichaelJ. Ryan · ThomasJ. McConnell

Permission to publish

I have checked the proofs of my article and

- ❑ I have no corrections. The article is ready to be published without changes.
- ❑ I have a few corrections. I am enclosing the following pages:
- ❑ I have made many corrections. Enclosed is the complete article.

Date / signature _____

Springer
the language of science

ELECTRONIC REPRINT ORDER FORM

After publication of your journal article, electronic (PDF) reprints may be purchased by arrangement with Springer and Aries Systems Corporation.

The PDF file you will receive will be protected with a copyright system called DocuRights®. Purchasing 50 reprints will enable you to redistribute the PDF file to up to 50 computers. You may distribute your allotted number of PDFs as you wish; for example, you may send it out via e-mail or post it to your website. You will be able to print five (5) copies of your article from each one of the PDF reprints.

Please type or print carefully. Fill out each item completely.

1. Your name: _____

   Your e-mail address: _____

   Your phone number: _____

   Your fax number: _____

2. Journal title (vol, iss, pp): _____

3. Article title: _____

4. Article author(s): _____

5. How many PDF reprints do you want? _____

6. Please refer to the pricing chart below to calculate the cost of your order.

Number of PDF reprints	Cost (in U.S. dollars)
50	$200
100	$275
150	$325
200	$350

NOTE: Prices shown apply only to orders submitted by individual article authors or editors. Commercial orders must be directed to the Publisher.

All orders must be prepaid. Payments must be made in one of the following forms:

- a check drawn on a U.S. bank
- an international money order
- Visa, MasterCard, or American Express (no other credit cards can be accepted)

PAYMENT (type or print carefully):

Amount of check enclosed: _____ (payable to Aries Systems Corporation)

VISA _____

MasterCard _____

American Express _____

Expiration date: _____ Signature: _____

Print and send this form with payment information to:

Aries Systems Corporation
200 Sutton Street
North Andover, Massachusetts 01845
Attn.: Electronic Reprints
— OR —
Fax this to Aries at: 978-975-3811

Your PDF reprint file will be sent to the above e-mail address. If you have any questions about your order, or if you need technical support, please contact: support@docurights.com

For subscriptions and to see all of our other products and services, visit the Springer website at:
http://www.springeronline.com

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Divergent patterns of selection on the *DAB* and *DXB* MHC class II loci in *Xiphophorus* fishes	
Article Sub-Title		
Article CopyRight - Year	Springer Science+Business Media B.V. 2008 (This will be the copyright line in the final PDF)	
Journal Name	Genetica	
Corresponding Author	Family Name	**Summers**
	Particle	
	Given Name	**Kyle**
	Suffix	
	Division	Department of Biology
	Organization	East Carolina University
	Address	27858, Greenville, NC, USA
	Email	summersk@ecu.edu
Author	Family Name	**Roney**
	Particle	
	Given Name	**Kelly E.**
	Suffix	
	Division	Department of Microbiology and Immunology
	Organization	University of North Carolina
	Address	Chapel Hill, NC, USA
	Email	
Author	Family Name	**Silva**
	Particle	**da**
	Given Name	**Jack**
	Suffix	
	Division	School of Molecular and Biomedical Science
	Organization	University of Adelaide
	Address	5005, Adelaide, SA, Australia
	Email	
Author	Family Name	**Capraro**
	Particle	
	Given Name	**Gerald**
	Suffix	
	Division	Department of Microbiology and Immunology, School of Medicine
	Organization	Wake Forest University
	Address	Winston-Salem, NC, USA
	Email	
Author	Family Name	**Cuthbertson**
	Particle	
	Given Name	**Brandon J.**
	Suffix	
	Division	

	Organization	NIH/NIEHS
	Address	Research Triangle Park, 27709-2233, Durham, NC, USA
	Email	
Author	Family Name	**Kazianis**
	Particle	
	Given Name	**Steven**
	Suffix	
	Division	
	Organization	Wistar Institute
	Address	19104, Philadelphia, PA, USA
	Email	
Author	Family Name	**Rosenthal**
	Particle	
	Given Name	**Gil G.**
	Suffix	
	Division	
	Organization	
	Address	3258 TAMU Biological Sciences Building West, 77843-3258, College Station, TX, USA
	Email	
Author	Family Name	**Ryan**
	Particle	
	Given Name	**Michael J.**
	Suffix	
	Division	Section of Integrative Biology
	Organization	University of Texas
	Address	78712, Austin, TX, USA
	Email	
Author	Family Name	**McConnell**
	Particle	
	Given Name	**Thomas J.**
	Suffix	
	Division	Department of Biology
	Organization	East Carolina University
	Address	27858, Greenville, NC, USA
	Email	

Schedule	Received	18 November 2007
	Revised	
	Accepted	9 June 2008

Abstract

Two MHC class II loci, *DAB* (a classical class II locus) and *DXB* (putatively a non-classical class II locus), were sequenced in samples of individuals from two populations of swordtail fish, *Xiphophorus multilineatus* and *X. pymaeus*. The *DAB* locus showed higher levels of genetic variation in the *B*1-encoding region, (putative binding region) than the *DXB* locus. We used two methods to investigate d_N/d_S ratios. The results from a maximum likelihood method based on phylogenetic relationships indicated positive selection on the *B*1 region of *DAB* (this method could not be used on *DXB*). Results from a coalescent-based method also showed evidence for positive selection in the *B*1 region of *DAB*, but only weak evidence for selection on the *DXB*. Further analyses indicated that recombination is an important source of variation in the *B*1 region

of *DAB*, but has a relatively small effect on *DXB*. Overall, our results were consistent with the hypothesis that the *DAB* locus is under positive selection driven by antagonistic coevolution, and that the *DXB* locus plays the role of a non-classical MHC II locus. We also used simulations to investigate the presence of an elevated synonymous substitution rate in the binding region. The simulations revealed that the elevated rate could be caused by an interaction between positive selection and codon bias.

Keywords (separated by '-')	Swordtail fish - Major histocompatibility complex - Positive selection - Non-classical MHC II locus
Footnote Information	**Electronic supplementary material** The online version of this article (doi:10.1007/s10709-008-9284-4) contains supplementary material, which is available to authorized users.

Journal:	10709
Article:	9284

Author Query Form

Springer
the language of science

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author,

During the preparation of your manuscript for typesetting, some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to spr_corrections1@sps.co.in

Disk use

In some instances we may be unable to process the electronic file of your article and/or artwork. In that case we have, for efficiency reasons, proceeded by using the hard copy of your manuscript. If this is the case the reasons are indicated below:

❑ Disk damaged ❑ Incompatible file format ❑ LaTeX file for non-LaTeX journal
❑ Virus infected ❑ Discrepancies between electronic file and (peer-reviewed, therefore definitive) hard copy
❑ Other: .....

We have proceeded as follows:

❑ Manuscript scanned ❑ Manuscript keyed in ❑ Artwork scanned
❑ Files only partly used (parts processed differently: ……..……..)

Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

❑ The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.
❑ *Uncited references*: This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.

Queries and/or remarks

Section/paragraph	Details required	Author's response
Body matter	The reference Hughes and Yeager (1998) is not cited in the text.	
Body matter	Please specify correct text or section for see below, see above, see text in body matter.	
Reference	The references Thompson et al. (1997), Yang (2001), Wilson (2006), Comeron and Kreitman (1998), Rozas et al. (1999) are not listed in the reference list. Please check.	
Reference	Please provide complete details for the reference Maddison and Maddison (2007).	

Genetica
DOI 10.1007/s10709-008-9284-4

Divergent patterns of selection on the *DAB* and *DXB* MHC class II loci in *Xiphophorus* fishes

Kyle Summers · Kelly E. Roney · Jack da Silva · Gerald Capraro · Brandon J. Cuthbertson · Steven Kazianis · G. Rosenthal · Michael J. Ryan · Thomas J. McConnell

Received: 18 November 2007 / Accepted: 9 June 2008

Abstract Two MHC class II loci, *DAB* (a classical class II locus) and *DXB* (putatively a non-classical class II locus), were sequenced in samples of individuals from two populations of swordtail fish, *Xiphophorus multilineatus* and *X. pymaeus*. The *DAB* locus showed higher levels of genetic variation in the *B*1-encoding region, (putative binding region) than the *DXB* locus. We used two methods to investigate d_N/d_S ratios. The results from a maximum likelihood method based on phylogenetic relationships indicated positive selection on the *B*1 region of *DAB* (this method could not be used on *DXB*). Results from a coalescent-based method also showed evidence for positive selection in the *B*1 region of *DAB*, but only weak evidence for selection on the *DXB*. Further analyses indicated that recombination is an important source of variation in the *B*1 region of *DAB*, but has a relatively small effect on *DXB*. Overall, our results were consistent with the hypothesis that the *DAB* locus is under positive selection driven by antagonistic coevolution, and that the *DXB* locus plays the role of a non-classical MHC II locus. We also used simulations to investigate the presence of an elevated synonymous substitution rate in the binding region. The simulations revealed that the elevated rate could be caused by an interaction between positive selection and codon bias.

Keywords Swordtail fish · Major histocompatibility complex · Positive selection · Non-classical MHC II locus

Electronic supplementary material The online version of this article (doi:10.1007/s10709-008-9284-4) contains supplementary material, which is available to authorized users.

K. Summers (✉) · T. J. McConnell
Department of Biology, East Carolina University, Greenville, NC 27858, USA
e-mail: summersk@ecu.edu

K. E. Roney
Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC, USA

J. da Silva
School of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005, Australia

G. Capraro
Department of Microbiology and Immunology, School of Medicine, Wake Forest University, Winston-Salem, NC, USA

B. J. Cuthbertson
NIH/NIEHS, Research Triangle Park, Durham, NC 27709-2233, USA

S. Kazianis
Wistar Institute, Philadelphia, PA 19104, USA

G. G. Rosenthal
3258 TAMU Biological Sciences Building West, College Station, TX 77843-3258, USA

M. J. Ryan
Section of Integrative Biology, University of Texas, Austin, TX 78712, USA

Introduction

Classical MHC class I and class II loci are responsible for presenting peptides to T cells. The interaction of the peptide-MHC class II complex with a T cell receptor (TCR) and its co-receptor CD4 is the crucial step in initiating an immune response (Wang and Reinherz 2002). The classical MHC genes are typically polygenic and polymorphic, and in fact are the most polymorphic gene system known in vertebrates (Hughes 1999). This variability is probably driven by

Journal : Large 10709 | Dispatch : 19-6-2008 | Pages : 11
Article No. : 9284 | LE | TYPESET
MS Code : GENE787 | CP | DISK

overdominance or negative frequency-dependence in the context of antagonistic parasite–host coevolution. This theory is supported by the observation that the most extensive variability found among MHC class II genes is found in the regions that encode the peptide binding region (the *B*1 region), and that variability is a result of positive selection (Hughes 1999, 2002). In contrast, nonclassical MHC class II genes encode proteins that perform functions other than antigen presentation to T helper cells, such as regulating the loading and unloading of peptide onto classical MHC molecules, and are not associated with high levels of polymorphism (Alfonso and Karlsson 2000).

Classical MHC class II proteins, after synthesis and transport into the ER of antigen presenting cells (APC), bind invariant chain (Ii), a protein chaperone (Cresswell 1996). The Ii guides MHC class II molecules out of the ER and targets them to MHC class II compartments (MIICs) where most peptide loading occurs (Neefjes 1999). Within the MIICs, Ii is degraded by proteases, leaving only a small fragment in the peptide binding pocket of MHC known as CLIP, or class II-associated Ii peptide (Robbins et al. 1996). In humans, CLIP is released from the CLIP-MHC class II complex either spontaneously or with the aid of nonclassical MHC class II HLA-DM heterodimer, which serves as a lysosomal chaperone (Alfonso and Karlsson 2000).

Another nonclassical MHC class II protein identified in humans, HLA-DO, is thought to associate with DM in B cells and inhibit the function of DM in endosomes (Kropshofer et al. 1998). Nonclassical class II genes are structurally similar to classical MHC class II genes in that they encode a leader sequence, $\alpha 1$, and $\alpha 2$ (or *B*1 and *B*2) domains, connecting peptide, a transmembrane segment, and a cytoplasmic tail. However, in exon 3, the most conserved exon, nonclassical MHC class II genes are more divergent from classical MHC class II genes than classical MHC class II genes are from each other (Servenius et al. 1987). Nonclassical MHC genes are generally much less polymorphic than classical MHC genes (Alfonso and Karlsson 2000).

MHC class II genes have been studied in many species of fish, but putative nonclassical MHC class II loci have only recently been identified (Roney et al. 2004). McConnell et al. (1998a) identified and characterized two highly divergent types of class II *B* loci, *DAB* and *DXB,* in fish of the genus *Xiphophorus*. The putative nonclassical class II locus, *DXB*, was identified largely based on the sequence divergence between *DXB* and *DAB*, and the presence of alternative RNA splicing found only in *DXB* transcripts (Roney et al. 2004). The presence of both *DAB* and *DXB* has only been confirmed in the *Xiphophorus* species *X. helleri*, *X. maculatus*, *X. multilineatus*, *X. pygmaeus*, and the guppy *Poecilia reticulata* (McConnell et al. 1998a, 1998b; Roney et al. 2004). *Xiphophorus* fishes have been used in research of the immune system well over fifty years (Kallman 1958, 1964). Backcrosses between certain strains of *helleri* and *maculatus* can result in fish that develop malignant melanoma, among other diseases (Anders et al. 1994; Kazianis et al. 2001). The availability of inbred strains and the initial characterization of MHC *DAB* and *DXB* genes in *Xiphophorus* make the genus an excellent system in which to study evolutionary dynamics of MHC class II genes.

In this study, we investigate the evolution of these genes in a comparative framework, using DNA sequences from several different populations of two different species of fish from the genus *Xiphophorus*: *X. multilineatus* and *X. pygmaeus*. In particular, we focus on the differences between *DXB* and *DAB* in terms of levels of polymorphism and patterns of molecular evolution.

The hypothesis that *DXB* represents an MHC locus equivalent to a nonclassical class II locus such as the HLA-DM locus in humans predicts that this locus will be under different selection pressures than *DAB*, which is a classic MHC locus involved in the binding of foreign peptides. The *DAB* locus should be intimately involved in the continual antagonistic coevolution that characterizes the interaction between immune system genes and parasite genes (Hughes 1999). Antagonistic coevolution drives diversifying selection, in which selection favors new protein variants in both the hosts and parasites (Hughes 1999). Therefore, we expect higher levels of polymorphism to occur in *DAB* relative to *DXB*, and we expect the action of diversifying selection to be stronger and more pervasive on *DAB* than on *DXB*.

We first briefly characterize levels of polymorphism in these two loci. We then investigate levels of trans-population and trans-species polymorphism in *DAB*, to determine the degree of ancient polymorphism present at this locus. To test the prediction of higher polymorphism and of stronger diversifying selection on the *DAB* locus than the *DXB* locus, we analyze the ratio of non-synonymous to synonymous substitution rates.

Materials and methods

Fishes

Xiphophorus multilineatus specimens were captured in the Rio Coy (21°45′0″ N, 98°57′25″ W) and the Arroyo Tambaque (21°41′6″ N, 99°2′30″ W), San Luis Potosi state, Mexico. *Xiphophorus pygmaeus* specimens were captured at two locations on the Rio Huichihayán 21°28′48.1″ N, 98°58′0″ W and 21°27′8.8″ N, 98°56′18.8″ W), San Luis Potosi state, Mexico, near the towns of Huichihyayán and La Y-Griega Vieja. Thirty-nine specimens were collected (25 *X. multilineatus* and 14 *X. pygmaeus*).

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

RNA isolation, cDNA preparation, and PCR conditions

Intestinal tissue RNA samples obtained from the 39 *Xiphophorus* fishes were obtained using the Trizol reagent protocol (Life Technology Inc., Gaithersburg, MD) and converted into single-stranded cDNA using the Gibco First Strand preamplification system with Oligo-dT (Life Technology Inc., Gaithersburg, MD). Single-stranded cDNA samples were then used to amplify MHC class II *DAB* and *DXB* transcripts using the identical primer sets and reaction conditions as previously described (Roney et al. 2004, McConnell et al. 1998a). Briefly, all *DAB* and *DXB* transcripts were isolated using a modified Polymerase Chain Reaction +1 (PCR +1) method to eliminate or minimize any possibility of chimeric PCR artifacts (Borriello and Krauter 1990; Hardee et al. 1995). *DXB* PCR +1 fragments were produced using primers TM341 (5′-ATCTCTGT TGCCAATCTAAGA-3′), TM328 (5′-ATGTGTAAAAG GCTAAATGAT-3′), and TM342 (5′-GAGAAGCTTAT-CTCTGTTGCCAATCTAAGA-3′) for the +1 step (*Hin*dIII site is underlined). *DAB* PCR +1 fragments were produced using primers TM396 (5′-GCTGGGCTGGCTGCTGGT-CAT-3′), TM398 (5′-GAAGCAGGAGGAACCAGAACC-3′), and TM399 (5′-AGAAAGCTTGCTGGGCTGGCTG CTGGTCAT -3′) for the +1 step. *DXB* transcripts were amplified by combining 50 ng of the single stranded cDNA sample, 1× Clontech Advantage cDNA Polymerase Mix (BD Biosciences, Palo Alto, CA), 1× cDNA PCR Reaction Buffer (BD Biosciences, Palo Alto, CA), 0.2 mM dNTP (BD Biosciences, Palo Alto, CA), 1 mM Primer TM328 and 0.2 mM primer TM341. This mixture was held at 94(C for 1 min, then cycled through 94(C for 1 min, 62(C for 1 min, and 68(C for two minutes for 35 times in a MJResearch PTC-200 Peltier Thermal Cycler (Waltham, MA) heated lid thermocycler. The "+1" cycle was performed as previously described (Roney et al. 2004), and the resulting DNA was ligated into pGEM-T Easy (Promega Corporation, Madison, WI). After electroporation of ligated plasmid into TOP 10 Ultracompetent Cells (Invitrogen, La Jolla, CA) via electroporation at 1.5 kV, 200 μF, and 25 F in a 0.1 cm cuvette, plasmid DNA that was isolated and linearized by the enzyme *Hind*III was selected for DNA sequencing, as linearization indicated that the clone contained a PCR +1 gene fragment.

DNA sequencing

From each individual fish, at least one PCR +1 positive plasmid clone from *DAB* and from *DXB* was sequenced, using Universal Forward, Universal Reverse, and gene specific sequence primers. Sequencing reactions followed the protocol recommended by Applied Biosystems (Foster City, CA) using 1 μl of primer, 2 μl Big Dye Terminator and 2 μl Big Dye Buffer (Applied Biosystems, Foster City, CA), and 200–500 ng of plasmid DNA for 26 cycles in an MJResearch PTC-200 Peltier Thermal Cycler (Waltham, MA). The reaction mixtures were processed on an ABI Prism 377 Sequencer (Foster City, CA). The resulting electropherograms were edited and assembled in AutoAssembler (Perkin-Elmer Applied Biosystems, Foster City, CA). Individual sequences were verified as *DXB* and *DAB* transcripts using the NCBI BLAST program (Altschul et al. 1990). None of the cDNA clones analyzed included primer-induced sequence. GenBank accession numbers for the sequences are as follows (accession numbers will be added upon acceptance of the manuscript).

In order to be conservative and avoid analyzing PCR and cloning artifacts, we filtered the *DAB* sequences as follows for the analyses of positive selection (the complete datasets for each locus were used for analyses of within-population polymorphism, because the results were very similar to the results from analysis of the culled datasets): First, we created a dataset consisting of unique haplotypes ($N = 37$ sequences). We then created a more restricted dataset ($N = 14$) consisting of only haplotypes that were independently replicated across individuals (which we will refer to as the "core dataset"). Given the extreme variability of the *DAB* sequences, it is virtually impossible that identical sequences would be present in two distinct individuals as the result of independent PCR or cloning artifacts. Hence the core dataset is highly unlikely to include any artificial haplotypes. For the analyses of positive selection, we analyzed both the unique haplotype and the core datasets. Also, multiple independent PCR amplifications and sequencing reactions were performed for *DAB* on a single fish specimen, resulting in eight near identical sequences. These eight sequences consisted of four identical sequences, three sequences that differed by a single nucleotide, and one sequence with two differing nucleotides. Therefore the many apparently distinct *DAB* alleles, differing from one another by 40–50 nucleotides, represent true alleles and are not the product of RT or *Taq* errors. Any ambiguous sites in differentiating alleles (a maximum of two sites per sequence) were assigned an "N" in analyses.

The *DXB* locus is much less polymorphic than the *DAB* locus (see below), and the variation among sequences is confined to small numbers of single nucleotide differences. Hence it was not possible to obtain a restricted (core) dataset with the method used for *DAB*. For *DXB*, we analyzed three datasets: the complete (original) dataset, a dataset consisting of only unique haplotypes, and a dataset consisting of the haplotypes of the same individuals as the restricted dataset for *DAB*. We note that the analysis of the full dataset for *DXB* is conservative with respect to our hypothesis, given that PCR and cloning artifacts should enhance variation among sequences.

Journal : Large 10709 | Dispatch : 19-6-2008 | Pages : 11
Article No. : 9284 | LE | TYPESET
MS Code : GENE787 | CP | DISK

Sequences used for the cross-species analysis for *DAB* were taken from GenBank: *Morone saxatilis* (GI: 501167), *Scophthalmus maximus* (GI: 62901681), *Stizostedion vitreum* (GI: 37724342), *Pagrus major* (GI: 37779051), *Oryzias latipes* (GI:7527374), *Poecilia reticulata* (GI:976097), *Xiphophorus maculatus* (GI: 2961096).

Sequence analysis

Sequences were aligned in Clustal X (Thompson et al. 1997), and checked by eye. Alignment was straightforward as there were very few gaps. Levels of polymorphism within populations and levels of divergence between populations were estimated in the program DnaSP (Rozas et al. 2003). Gene trees were estimated via parsimony and maximum likelihood in PAUP 4.0b10 (Swofford 2003). The coalescence of the most parsimonious gene trees within the species and populations tree was estimated with the coalescent simulation module in MESQUITE 2.0 (Maddison and Maddison 2007).

We used comparative sequence analysis methods to search for a signal of selection across individuals, populations and species. For cross-species analyses, we used maximum likelihood methods (implemented in the CODEML program of the PAML package (Yang 1997)) to investigate the ratio of nonsynonymous to synonymous substitution rates, or omega (ω). A value of omega greater than one indicates the action of positive selection (Yang and Bielawski 2000).

We carried out a cross-species comparison of ω for *DAB*, using sequences from a variety of other species obtained from the literature, as listed above (this could not be done for *DXB*, as this locus has been sequenced in only a few species as yet). For the cross-species comparisons, we used tree topologies that represented a combination (an informal supertree) of well-supported phylogenetic trees from the recent molecular systematics literature (see online supplement for details).

For the analysis of ω across species, we used the branch and codon-specific model MA implemented in CODEML (Zhang et al. 2005). The significance of evidence for positive selection was tested with a log-likelihood ratio test (LRT), and specific sites under positive selection were identified using a Bayesian method (the Bayes Empirical Bayes method).

The methods implemented in CODEML are designed to be used for the analysis of separate lineages in a phylogenetic framework (Yang 2001). This is not a problem for our cross-species analysis, but is a potential problem for the datasets containing multiple sequences from the same population and species. Previous research indicates that recombination can be extensive in MHC loci (Doxiadis et al. 2006). Recent simulation studies (Anisomova et al. 2003) indicate that substantial recombination among sequences can cause these methods to confuse recombination with positive selection. In this study, recombination among sequences within a population is a distinct possibility, as is recombination between populations, and even between species (the amount of hybridization is unknown).

To investigate the influence of recombination, we used the computer program LDhat (McVean et al. 2002). LDhat estimates recombination rate using the composite-likelihood method developed by Hudson (2001), but the model is extended by using a finite-sites model to estimate the likelihood of two-locus haplotypes under the coalescent (McVean et al. 2002). Under coalescent theory, one can estimate $\rho = 4N_e r$, where N_e is the effective population size and r is the recombination rate. The estimate of recombination is conditioned on theta (θ), which is estimated using a finite-series version of the Watterson estimator. This method is less subject to false positives because it takes into account the probability of recurrent substitutions, which can generate patterns of variation interpreted as evidence of recombination under some methods. Simulation studies indicate that the method implemented in LDhat provides accurate estimates of the relative contributions of point mutation and recombination to the observed sequence variation, even when both occur at high rates (Richman et al. 2003). The composite likelihood model implemented in LDhat does not take selection into account. However, recent simulation studies indicate that the method performs very well even in the presence of selection (Richman et al. 2003).

One caveat with respect to the data analyzed here is that the method implemented in LDhat assumes the data come from a single-locus, with intragenic recombination. Previous research (McConnell et al. 1998a) indicates that this is the case for *DXB* in *Xiphophorus*, but we cannot be sure that that is true for the *Xiphophorus DAB* data. Nevertheless, the method provides a useful first step in estimating the relative contributions of mutation and recombination in producing the sequence variation that characterizes the MHC II *DAB* gene region of *Xiphophorus*.

In order to address the evidence for positive selection on point mutations explicitly, we used a recently developed maximum likelihood method that does not utilize a phylogenetic framework, but rather uses the coalescent as a framework for the analysis of ω, implemented in the program omegaMap (Wilson and McVean 2006). The method estimates two focal parameters, the selection parameter (ω) and the recombination rate (ρ), as well as the transition-transversion ratio (κ) and the insertion/deletion rate (ϕ). Because there can be multiple phylogenetic trees along the sequence (due to recombination), the trees are treated as a nuisance parameter and the likelihood function is averaged across all possible trees and branch lengths. The addition of recombination makes the calculation of likelihoods highly

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

computationally intensive, so an approximation to the likelihood in the presence of recombination developed by Li and Stephens (2003) is used. This method uses a hidden Markov model to incorporate key properties of the proper likelihood while improving computational efficiency. The model of Nielsen and Yang (1998) is used to specify the transition rates among codons, with a modification that allows the addition of an insertion/deletion rate (Wilson and McVean 2006). The method employs a Bayesian framework, using a Markov Chain Monte Carlo approach to sample the likelihood surface and estimate posterior probabilities.

The model employs a flexible blocking structure implemented in the prior distribution on ω. This approach is based on the multiple change-point model of Green (1995) and developed by McVean et al. (2004). It allows estimation of variable recombination rates along a sequence. The blocking structure reduces that computational load and exploits the information available in the sequence more thoroughly.

Simulation studies and studies on sample datasets suggest that the method performs well, and that estimates of ω do not confound estimates of ρ or vice versa (Wilson and McVean 2006). The method requires that prior distributions be used for each of the parameters: For the initial runs, we used the following prior distributions and ranges or initial values: ω = inverse (0.01–100), ρ = inverse (0.01–100), μ = improper inverse (0.1), κ = improper inverse (3.0), ϕ = improper inverse (0.1), as suggested in the omegaMap documentation (Wilson 2006). Each combination of gene and species was run at least twice, and the posterior probabilities for each key parameter (ω, ρ, μ, κ, ϕ) were examined to insure that the results were consistent between the independent runs. We also ran each combination of gene and species using two different blocking structures for the prior distribution on ω: 30 codons and 5 codons. To determine the sensitivity of the analyses to the distribution and initial values chosen for the priors, we carried out two other runs for each locus using alternatives as follows: Prior A: exponential distribution, with mean 0.07 for μ, mean 3.0 for κ, mean 0.1 for ϕ, mean 1.0 for ω, and 0.1 for ρ; Prior B: a uniform prior distribution was used for μ and ρ, with range 0–10, an exponential ratio distribution for κ, with median equal to 1.0, an exponential prior distribution of ϕ with a mean of 1.0, and gamma distribution for ω, with a mean of 1.0 and a value of 2.0 for the shape parameter (Wilson and McVean 2006). For each analysis, we ran 500,000 generations. To summarize the data, we used the Summarize module of the omegaMap program, which summarizes the results from every 100th generation of the run. A burnin of 25,000 generations was used, and the data was visualized in Microsoft EXCEL.

We note that both intra- and inter-locus recombination are potentially important sources of adaptive variation in the MHC (Reusch and Langefors 2005). In some cases, recombination appears to play a dominant role in producing adaptive variation at the MHC (Richman et al. 2003). Hence, we used LDhat (McVean et al. 2002) to characterize the relative contributions of point mutation and recombination to variation in the MHC II *DAB* and *DXB* loci of both species, and we used omegaMap to investigate positive selection on point mutations, controlling for recombination.

Simulation of positive selection and codon bias

We wished to test whether the elevated d_S of the *B*1 region, as compared to the remainder, of *DAB* could be explained by a combination of positive selection on nonsynonymous sites due to selection at the protein level and positive selection on synonymous sites due to codon bias. In order to test this possibility, we took a simulation approach. The approach was to first infer the DNA sequence ancestral to all of the extant sequences used in this analysis and then evolve the ancestral sequence toward each of the extant sequences in a three-step process: (1) introduce the *minimum* number of nonsynonymous substitutions necessary to produce codons that code for the amino acids coded by the target extant sequence (not necessarily the codons of the extant sequence), thus simulating selection at the protein level; (2) introduce random synonymous substitutions at the rate estimated from pairwise comparisons of the non-*B*1 region (0.009 substitutions per site), which represents the background mutation rate; and (3) introduce any additional synonymous substitutions required to produce the most *preferred* codon of each codon family in the target sequence (not necessarily the codons of the extant sequence), thus simulating selection for preferred codons. This was conducted for the *B*1-coding and the non-*B*1-coding regions separately. The third step was used simply to determine whether codon selection is necessary to explain the elevated d_S in *B*1 that may otherwise arise because of positive selection on nonsynonymous sites alone (due to different codons being used to code for the same amino acid in different sequences). Because codon bias is expected to affect both regions equally, any difference in d_S due to codon selection between the two regions must be due to its interaction with positive selection on nonsynonymous sites in *B*1.

The ancestral sequence was inferred from 40 *DAB* sequences deemed to be unique alleles. These sequences were chosen by randomly eliminating all but one sequence from clusters, in a Neighbor-Joining tree, in which sequences diverged by a distance of <0.003 substitutions per site (the maximum intra-individual value). The ancestral sequence was then inferred from a Neighbor-Joining tree of the 40 remaining sequences using a Bayesian method implemented in PAML. The values of d_S and d_N

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

were estimated using the Nei-Gojobori method with pairwise sequence comparisons of Jukes-Cantor-corrected distances, using MEGA2 (Kumar et al. 2001).

Results

The total number of transcripts analyzed in the complete dataset were: 25 *X. multilineatus DXB*, 15 *X. pygmaeus DXB*, 50 *X. multilineatus DAB*, and 28 *X. pygmaeus DAB*. The *B*1 region through the cytoplasmic tail-encoding region (*CT*) of *DXB* is 702 nucleotides (234 amino acids) in length while *DAB* is smaller with 687 nucleotides (229 amino acids). We also used several reduced datasets in the analyses of selection (see above). *DAB* and *DXB* nucleotide transcripts and protein sequences from each species were aligned, and were 56 and 44.5% identical, respectively.

A standard measure of nucleotide diversity (Pi) showed *DAB* to be substantially more polymorphic than *DXB* (Table 1). In *DAB*, polymorphism was similar between species, and was clustered mostly within the first 300 nucleotides, which corresponds to the *B*1 region and the first 40 amino acids (out of 278) of the β2-encoding (*B*2) region. The observed polymorphism in *DXB* was not clustered into any particular domain. The small amount of variability found in *DXB* is mainly localized around three particular nucleotide positions (520, 525, 527) that are within the *B*2 region and consist of G–A transitions.

The average number of nonsynonymous substitutions per nonsynonymous site (d_N) for *DAB* was much higher in the *B*1 region than in the rest of the transcript (Table 2). The d_S for the *B*1 region of *X. multilineatus* was slightly higher than that of *X. pygmaeus*. *Xiphophorus pygmaeus* had a slightly higher d_N.

The d_S and d_N of *DXB B*1 region are very different from that of *DAB* (Table 2). The observed numbers of both synonymous and nonsynonymous substitutions per site (d_N and d_S) in the *B*1 was higher for *DAB* than for *DXB* (Table 2). The d_S of the *B*1 are small (0.006), and are actually less than that found in the remainder of the *DXB* transcript (0.016), versus the large d_S for the *B*1 of *DAB*. The d_N is higher than the d_S for *DXB B*1 (0.020 and 0.006), and higher than the d_N found in the remainder of the transcript (0.009). However, the d_N in the *B*1 region observed in *DAB* is much larger. *Xiphophorus multilineatus* had a slightly higher d_S than that of *X. pygmaeus* in the analysis of the *B*2 region through the cytoplasmic tail region.

The coalescent analysis revealed extensive lack of reciprocal monophyly among species (and populations), or trans-species polymorphism. Figure 1 illustrates this for one of the set of most parsimonious gene trees fitted into a "contained tree" (the hypothetical population and species tree), using the program MESQUITE (the results are similar no matter which tree of the set is used). Much of the polymorphism across these populations and species is ancestrally retained, indicating that analyses of selection should focus on all of the sequences as a single set subject to similar selection pressures (Richman et al. 2007).

The cross-specific analysis of ω at the *DAB* locus with CODEML yielded abundant evidence for positive selection, as has been found in previous analyses of the MHC in other taxa. The branch specific analysis (model MA) showed a very high value of ω: 321.7 at between 2 and 3% of the codon sites. The LRT for the comparison of MA with MA (fixed ω) was significant ($\chi^2 = 27.58$, $P < 0.05$). The Bayes Empirical Bayes (BEB) method detected significant probabilities of positive selection at two sites in the antigen-binding region, site 69 and 83. Surprisingly, significant probabilities of positive selection were also detected at three sites outside of the binding region: 208, 241 and 242.

Table 1 Levels of polymorphism and genetic divergence between populations of *X. multilineatus* and *X. pygmaeus*, for the *DAB* and the *DXB* locus, calculated in the program DnaSP (Rozas et al. 1999)

Species	Population	Locus	*N*	NHaps	HapDiv	NucDiv(Pi)
Ximu	Rio Coy	DAB	24	22	0.993	0.057
Ximu	A.T.	DAB	26	22	0.981	0.074
Ximu	Rio Coy	DXB	11	9	0.946	0.013
Ximu	A.T.	DXB	13	10	0.923	0.015
Xipy	Huich. 1	DAB	12	9	0.955	0.047
Xipy	Huich. 2	DAB	15	15	1.000	0.062
Xipy	Huich. 1	DXB	7	5	0.857	0.009
Xipy	Huich. 2	DXB	9	9	1.000	0.018

N, Number of sequences; NHaps, number of haplotypes; HapDiv, haplotype diversity; NucDiv (Pi), nucleotide diversity; Ximu, *X. multilineatus*; Xipy, *X. pygmaeus*; A.T., Arroyo Tambaque; Huich 1 & 2, Rio Huichihayan, site 1 and site 2

Table 2 Average number of substitutions in *DAB* and *DXB*: d_S (d_N) refer to the average number of synonymous (nonsynonymous) substitutions per synonymous (nonsynonymous) site

	DAB				*DXB*			
	d_S	S.E.	d_N	S.E.	d_S	S.E.	d_N	S.E.
*B*1	0.136	0.023	0.159	0.019	0.006	0.002	0.020	0.006
Ximu	0.157	0.021	0.137	0.024	0.005	0.003	0.020	0.005
Xipy	0.130	0.023	0.147	0.020	0.010	0.006	0.019	0.005
*B*2-CT	0.016	0.006	0.010	0.004	0.016	0.007	0.009	0.003
Ximu	0.015	0.006	0.009	0.003	0.019	0.007	0.007	0.003
Xipy	0.018	0.007	0.011	0.004	0.007	0.005	0.011	0.003

S.E., Standard error. *B*1 and *B*2-CT refer to specific regions (see text). Species averages follow the combined averages. Ximu, *X. multilineatus*; Xipy, *X. pygmaeus*

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

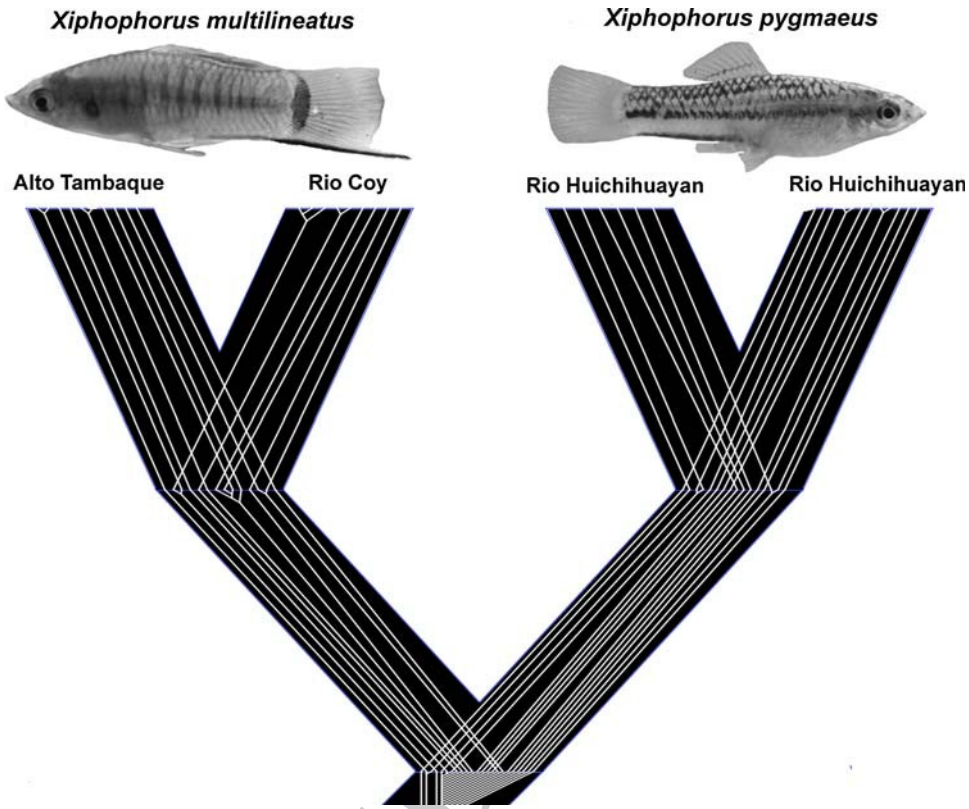

Fig. 1 A "contained tree", showing the coalescence of the haplotypes of the *DAB* gene tree within the population and species tree of *Xiphophorus* investigated in this study. The contained tree was created in the program MESQUITE 2.0 (Maddison and Maddison 2007)

Elevated probabilities of positive selection were also detected at a number of other sites in the binding region: Site (probability): 23 (0.83), 26 (0.91), 28 (0.91), 42 (0.77), 53 (0.79), 54 (0.68), 96 (0.60), 97 (0.54), and 99 (0.57).

The analyses with omegaMap also revealed evidence for selection on *DAB*. Figure 2f presents the results from the analysis of the core dataset (the analysis of the unique dataset produced virtually identical results). The point estimates indicate multiple regions in which ω is far above one. Posterior probabilities of positive selection were significant (above 95%) for each of these regions. The minimum estimates for the 95% highest posterior probability densities (HPDs) remained far above one in each region where selection was detected. Running the analyses with alternative priors did not affect the significance of the results for *DAB* (results not shown). For *DXB,* analysis of each dataset also showed evidence for positive selection on point mutations in the *B*1 region, but the estimated average levels were consistently far lower than for *DAB* (Fig. 3). Figure 3 shows the results for the core dataset for *DXB*, but the results were equivalent no matter which of the three *DXB* datasets was analysed. For *DXB*, regions of positive selection were not confined to the *B*1 region, but were also found in other parts of the sequence (Fig. 3). In contrast to the results from *DAB*, the minimum of the HPD was typically less than one, indicating that the evidence for positive selection on the *DXB* was weak. The posterior probabilities of positive selection for the estimates for the regions of *DXB* estimated to have an omega greater than one did exceed 95%, but, in contrast to *DAB*, this result was sensitive to the choice of prior. For example, under the exponential prior with the mean of ω set to 1.0, none of the regions showing positive selection had a posterior probability of 95% or above (results not shown). Hence, the signal of positive selection for the *DXB* locus was relatively weak, and strongly influenced by the prior.

Our analyses of recombination indicate that recombination does play an important role in producing and maintaining the variation seen in these MHC sequences. For *X. multilineatus*, the Watterson estimate of $4N_e\mu$ or θ (using the LDhat program) was equal to 43.09 (for *DAB*) and 10.44 (for *DXB*). For *X. pygmaeus*, the corresponding values for θ were 44.17 (*DAB*) and 8.14 (*DXB*). For *X. multilineatus*, the maximum likelihood estimate of $4N_e$r was 18 for *DAB* and 3 for *DXB*. For *X. pygmaeus*, the corresponding values for $4N_e$r were 44 (*DAB*) and 2 (*DXB*). In each case the evidence for recombination was highly significant using a likelihood permutation test ($P < 0.000$). The minimum number of recombination events, estimated by the method of Hudson and Kaplan (1985), was 28 for *X. multilineatus DAB*, 38 for *X. pygmaeus DAB*, 1 for *X. multilineatus DXB*, and 1 for *X. pygmaeus DXB*. This provides further evidence that levels of recombination in *DAB* are substantially higher than in *DXB*.

Journal : Large 10709 | Dispatch : 19-6-2008 | Pages : 11
Article No. : 9284 | LE | TYPESET
MS Code : GENE787 | CP | DISK

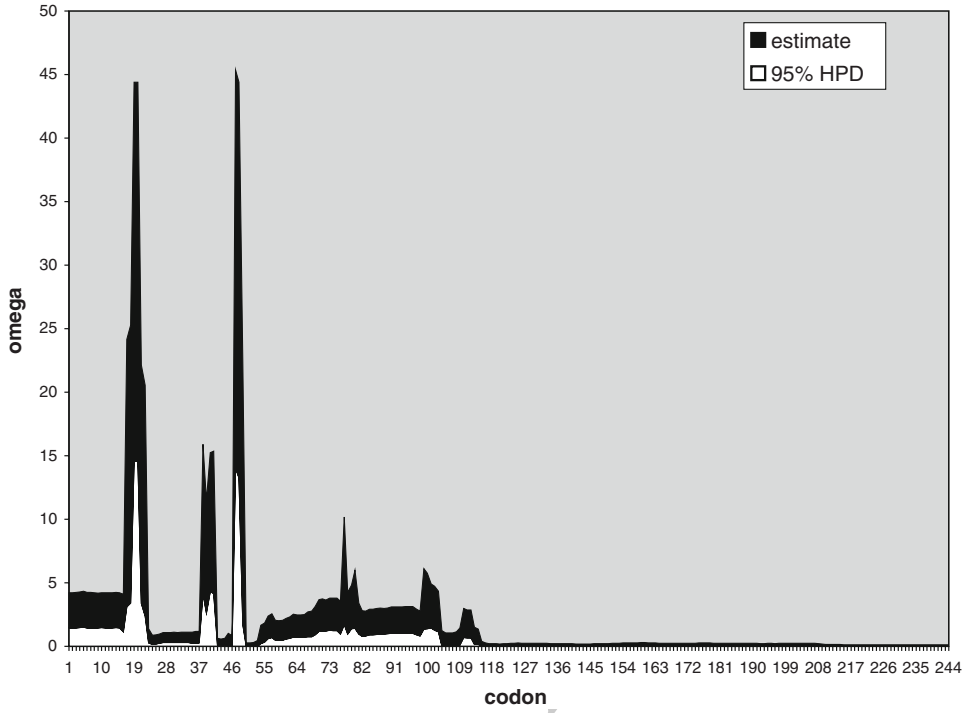

Fig. 2 Plot of omega versus codon position across the gene estimated from the omegaMap run for the *DAB* core dataset, using standard priors (see text). The markov chain monte carlo was run for 500,000 generations in each case, with aburnin of 25,000 generations. The point estimate for omega is shown in black, and the 95% minimum for the posterior probability density is shown in white. Figures show codon positions across the *X* axis, and estimates of omega on the *Y* axis

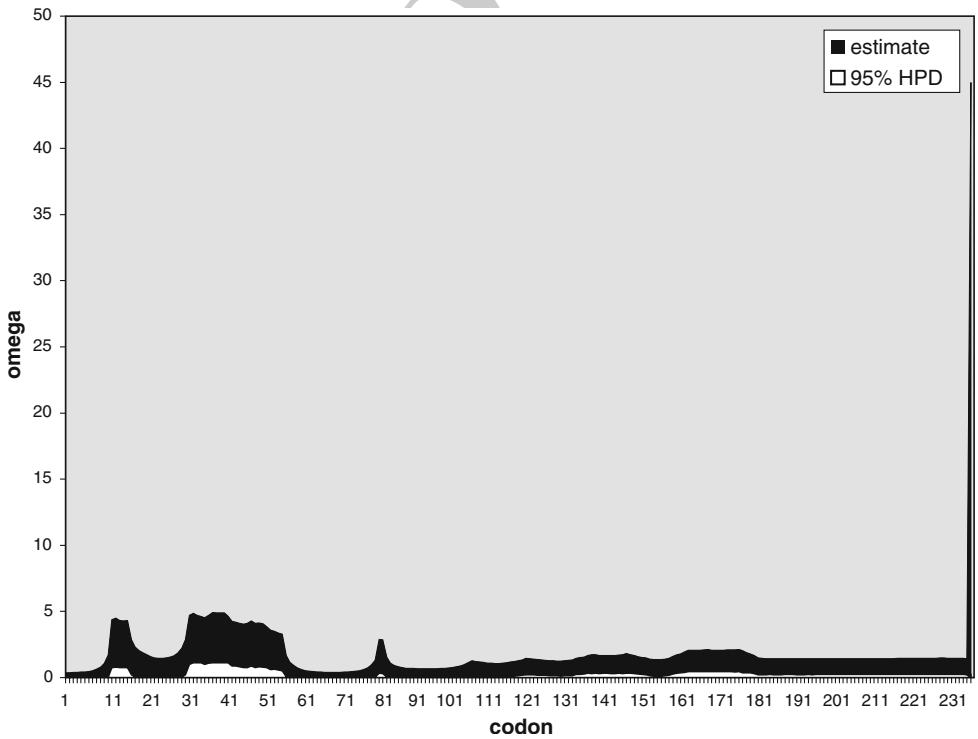

Fig. 3 The same as Fig. 2, except that the omega estimates for *DXB*, rather than *DAB*, are plotted

One unusual feature identified in this study is that both the nonsynonymous and the synonymous rates of substitution are highly elevated in the *B*1 region of *DAB*, relative to the remainder of the gene (*B*2-CT). One possible reason for the increase in synonymous substitutions is codon bias, which occurs when one or more triplet codes for the same amino acid are commonly used and other triplet codes for the same amino acid are rare. One explanation for codon bias is that in highly expressed genes, more abundant transfer RNAs (tRNAs) are used in proteins and less abundant tRNAs are thought to be eliminated by purifying selection (Nei and Kumar 2000). Another potential source of codon bias is

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

Table 3 Simulation of the combined effects of positive selection of nonsynonymous sites and selection of preferred codons on *DAB* regions

Simulation		Beta-1		Non-beta-1	
Step[a]	Replicate	d_S	d_N	d_S	d_N
1	1	0.033	0.181	0.004	0.013
	2	0.033	0.180	0.004	0.013
	3	0.032	0.181	0.003	0.014
2	1	0.050	0.181	0.018	0.013
	2	0.048	0.181	0.025	0.013
	3	0.049	0.181	0.021	0.014
3	1	0.152	0.191	0.013	0.016
	2	0.152	0.191	0.013	0.016
	3	0.152	0.191	0.013	0.016
Extant		0.153	0.183	0.018	0.013

[a] Simulation steps: (1) protein selection; (2) mutation; (3) preferred codon selection

mutational pressure to go from CG to AT or vice versa at silent first and third codon positions (Nei and Kumar 2000). We used simulation to test whether the elevated d_S seen in *DAB* could (in part) be caused by an interaction between positive selection and codon bias at this locus. Such an interaction could give rise to an elevated d_S if selection at the protein level does not necessarily produce the most preferred codon of a codon family, leading to subsequent selection for preferred codons and hence a higher d_S (Lipman and Wilbur 1984; DuMont et al. 2004; Comeron and Kreitman 1998).

The results show that the first step of the simulation (positive selection at the protein level) was sufficient to account for d_N observed for the *B*1 region, but that the third step (selection for preferred codons) was necessary to account for d_S in this region (Table 3). Similarly, the first step was sufficient to account for the observed d_N in the non-*B*1 region, but, in contrast, the second step (mutation) was sufficient to account for d_S in this region (Table 3). These results support the hypothesis of an interaction between codon selection and positive selection (which occurs for *B*1 only) being responsible for the elevated d_S observed for the *B*1 region.

Discussion

DAB and *DXB* are highly divergent MHC class II *B*-like genes. Both genes encode a *B*1 region, *B*2 region, a connecting peptide, transmembrane membrane, and cytoplasmic tail, yet are only about 45% identical. *DAB* is highly polymorphic, and is most variable in the region that corresponds to the encoded peptide binding pocket (*B*1). Classical MHC class II *B*-like genes also follow this pattern. In humans, classical MHC class II genes are more polymorphic than class I *A* genes, and have d_N values ranging from 0.10 to 0.20 in the *B*1 regions of the three classical MHC class II loci (HLA-*DR*, *DQ*, and *DP*) (Hughes and Yeager 1998a). These d_N values are comparable to the d_N value (0.16) observed for the *B*1 region of *DAB* in this study.

These extreme levels of polymorphism are thought to be maintained by overdominant selection or negative frequency dependent selection, and this selection is thought to be targeted specifically to the part of the gene that encodes the peptide binding region (Hughes and Yeager 1998a). Studies of MHC class II *B* polymorphism in a variety of species have supported the hypothesis that class II genes are under positive selection (Edwards et al. 1998; Hughes and Nei 1989; Hedrick et al. 2002). In fish, evidence for positive selection on MHC class II loci has been found in a number of fish, including danio, trout, several species of cichlid, and salmon (Figueroa et al. 2000; Graser et al. 1996; Miller and Withler 1996). In this study, the level of *DAB* polymorphism data is consistent with the hypothesis of overdominant or negative frequency-dependent selection, which suggests that the *DAB* encoded β-chain probably plays a classical role in *Xiphophorus*.

Our analysis of the coalescence of *DAB* haplotypes across populations and species of *Xiphophorus* indicates that both trans-population and trans-species polymorphism is extensive for this locus (Fig. 1). This is consistent with the persistent action of diversifying selection on this locus. The cross-specific analyses of positive selection using the branch- and codon-specific model (MA) in the CODEML program showed evidence for strong positive selection acting on the branch connecting all other species to *X. multilineatus* and *X. pygmaeus*, and the LRT was highly significant. The Bayes Empirical Bayes method indicated that there are a number of sites in the *B*1 region with significant posterior probabilities of positive selection. Analyses using omegaMap to estimate selection on point mutations independent of recombination also showed strong evidence of positive selection in the *B*1 region. The signal of selection was robust to variation in the mean and distribution of the prior on omega.

The results from the omegaMap analyses address positive selection on point mutations, and exclude the effects of recombination. Of course, recombination itself is likely an important mechanism for generating haplotype variation at this locus. In sexual species, recombination can rapidly generate new haplotypes by combining new protein variants that have arisen through separate mutation events within a single population. This is a crucial source of variation in sexually reproducing species. In fact, according to the red queen hypothesis developed by William D. Hamilton and others, immune system genes are likely to be key loci with respect to selection for the maintenance of sexual recombination itself (Hamilton et al. 1990). Several recent studies have identified recombination as the primary

Journal : Large 10709 Dispatch : 19-6-2008 Pages : 11
Article No. : 9284 LE TYPESET
MS Code : GENE787 CP DISK

mechanism producing variation in the peptide binding region of MHC class II sequences (e.g. Richman et al. 2003; Reusch and Langefors 2005). In our analyses using LDhat, we found evidence for extensive recombination in *DAB*. Hence, it is highly likely that recombination contributes to adaptive variation at this locus as well.

DAB and the rate of synonymous substitution

Our simulation of positive selection and codon bias in *DAB* shows that the elevated d_S of the *B*1 region compared to the non-*B*1 region could be explained by an interaction between these two forces. The argument is that positive selection produces new codons that are not necessarily the most preferred within the codon family and that subsequent selection for preferred codons elevates the synonymous substitution rate. However, our simulation does not provide the opportunity for genetic hitchhiking, in which synonymous mutations could spread to fixation due to their linkage to positively selected nonsynonymous mutations. Therefore, hitchhiking remains a possible alternative explanation. However, the observation of a high rate of recombination and strong codon bias in *DAB* supports the codon bias hypothesis over hitchhiking.

DXB and nonclassical MHC II loci

The results of the analyses of *DXB* sequences presented here and in previous work (McConnell et al. 1998a, 1998b; Roney et al. 2004) suggest that this locus may play a nonclassical role in the fish immune system. In human and mouse, the nonclassical MHC class II genes *DM* and *DO* have limited polymorphism that is not clustered in any one region of the gene (Alfonso and Karlsson 2000). In one study of HLA-*DOB* almost no polymorphism was observed (Naruse et al. 2002). This is thought to be a result of purifying selection on *DO* as a change in amino acid structure could be deleterious to the role of *DO* as a co-chaperon. *DM* also has very low levels of polymorphism in both mouse and human (Walter et al. 1996). *DXB* has very little polymorphism, and the few differences found are not confined to any particular gene region. *DXB* is also unique in MHC class II genes found in fishes in that it has at least two alternative transcripts (Roney et al. 2004).

The analyses of *DXB* using omegaMap indicated the presence of positive selection, but the signal was weak compared to *DAB*. The minimum of the 95% HPD interval was below one in most cases, and none of the regions estimated to show positive selection had significant posterior probabilities of positive selection under a conservative prior (i.e. exponential distribution with mean equal to 1.0). The signature of positive selection can persist for a very long time (Garrigan and Hedrick 2003), so it is possible that the signature detected in the analyses presented here came from positive selection on these sequences in the distant past. This suggests the possibility that the *DXB* currently serves a nonclassical role, but was originally derived from a duplication event involving a classical MHC II *B* locus.

The evidence presented here and in previous studies is consistent with the hypothesis that *DXB* plays a nonclassical role in *Xiphophorus* fishes. Nonclassical MHC class II *B*-like genes have not previously been identified as such in fishes. The nonclassical genes *DM* and *DO* are currently the focus of research to fully elucidate their function. Insight into a nonclassical MHC class II gene in fish will help characterize the evolution of nonclassical genes as well as the evolution the specific immune system.

Acknowledgements This research was supported by NSF grant #MCB 0110541 to Thomas McConnell. We thank Felix Breden and four anonymous reviewers for comments on the manuscript.

References

Alfonso C, Karlsson L (2000) Nonclassical MHC class II molecules. Annu Rev Immunol 18:113–142

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Anders A, Petry H, Fleming C et al (1994) Increasing melanoma incidence: putatively explainable by retrotransposons. Experimental contributions of the xiphophorine Gordon-Kosswig melanoma system. Pigment Cell Res 7:433–450

Anisomova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 168:1229–1236

Borriello F, Krauter KS (1990) Reactive site polymorphism in the murine protease inhibitor gene family is delineated using a modification of the PCR reaction (PCR +1). Nucleic Acids Res 18:5481–5487

Cresswell P (1996) Invariant chain structure and MHC class II function. Cell 84:505–507

Doxiadis GGM, van der Wiel MKH, Brok HPM et al (2006) Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a new world primate species. PNAS 103:5864–5868

Dumont VB, Fay JC, Calabrese PP et al (2004) DNA variability and divergence at the notch locus in *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. Genetics 167:171–185

Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of Agph-DAB1, an Mhc class II B gene in red-winged blackbirds (*Agelaius phoneniceus*). Mol Biol Evol 15:236–250

Figueroa F, Mayer WE, Sultmann H et al (2000) Mhc class II B gene evolution in East African cichlid fishes. Immunogenetics 51:556–575

Garrigan D, Hedrick PW (2003) Detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57:1707–1722

Graser R, O'HUigin C, Vincek V et al (1996) Trans-species polymorphism of class II Mhc loci in danio fishes. Immunogenetics 44:36–48

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732

Journal : Large 10709 | Dispatch : 19-6-2008 | Pages : 11
Article No. : 9284 | ☐ LE | ☐ TYPESET
MS Code : GENE787 | ☑ CP | ☑ DISK

Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaption to resist parasites. Proc Natl Acad Sci USA 87:3566–3573

Hardee JJ, Godwin U, Benedetto R et al (1995) Major histocompatibility complex class II A gene polymorphism in the striped bass. Immunogenetics 41:229–238

Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. Mol Ecol 11:1905–1913

Hudson RR (2001) Two-locus sampling distributions and their application. Genetics 159:1805–1817

Hudson RR, Kaplan N (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Hughes AL (1999) Adaptive evolution of genes and genomes. Oxford University Press, New York

Hughes AL (2002) Natural selection and the diversification of vertebrate immune effectors. Immunol Rev 190:161–168

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Hughes AL, Yeager M (1998a) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32:415–435

Hughes AL, Yeager M (1998b) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3:d509–d516

Kallman KD (1958) Genetics of fin transplantation in xiphophorin fishes. Ann NY Acad Sci 73:599–610

Kallman KD (1964) An estimate of the number of histocompatability loci in the teleost *Xiphophorus maculatus*. Genetics 50:583–595

Kazianis S, Gimenez-Conti I, Setlow RB et al (2001) MNU induction of neoplasia in a platyfish model. Lab Invest 81:1191–1198

Kropshofer H, Vogt AB, Thery C et al (1998) A role for HLA-DO as a co-chaperone of HLA-DM in peptide loading of MHC class II molecules. EMBO J 17:2971–2981

Kumar S, Tamura K, Jakobsen IB et al (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Li N, Stephens M (2003) Modelling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165:2213–2233

Lipman DJ, Wilbur WJ (1984) Interaction of silent and replacement changes in eukaryotic coding sequences. J Mol Evol 21:161–167

Maddison W, Maddison D (2007) MESQUITE version 2.0. Software for comparative analysis

McConnell TJ, Godwin UB, Norton SF et al (1998a) Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in Xiphophorus fishes. Genetics 149:1921–1934

McConnell TJ, Godwin UB, Cuthbertson BJ (1998b) Expressed major histocompatibility complex class II loci in fishes. Immunol Rev 166:294–300

McVean G, Awadalla P, Fearnhead PA (2002) Coalescent-based method for detecting and estimating recombination rates from gene sequences. Genetics 160:1231–1241

McVean G, Myers SR, Hunt S et al (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584

Miller KM, Withler RE (1996) Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. Immunogenetics 43:337–351

Naruse TK, Kawata H, Inoko H et al (2002) The HLA-DOB gene displays limited polymorphism with only one amino acid substitution. Tissue Antigens 59:512–519

Neefjes J (1999) CIIV, MIIC and other compartments for MHC class II loading. Eur J Immunol 29:1421–1425

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press

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

Reusch TBH, Langefors A (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. J Mol Evol 61:531–541

Richman AD, Herrera LG, Nash D et al (2003) Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. Genet Res Camb 82:89–99

Richman AD, Herrera G, Reynoso VH et al (2007) Evidence for balancing selection at the *DAB* locus in the axolotl, *Ambystoma mexicanum*. Int J Immunogenet 34:475–478

Robbins NF, Hammond C, Denzin LK et al (1996) Trafficking of major histocompatibility complex class II molecules through intracellular compartments containing HLA-DM. Hum Immunol 45:13–23

Roney KE, Cuthbertson BJ, Godwin UB et al (2004) Alternative splicing of major histocompability complex class II DXB transcripts in *Xiphophorus* fishes. Immunogenetics 56:462–466

Rozas J, Sanchez-Delbarrio J, Messeguer X et al (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Servenius B, Rask L, Peterson PA (1987) Class II genes of the human major histocompatibility complex. The DO beta gene is a divergent member of the class II beta gene family. J Biol Chem 262:8759–8766

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Walter W, Loos M, Maeurer MJ (1996) H2-M polymorphism in mice susceptible to collagen-induced arthritis involves the peptide binding groove. Immunogenetics 44:19–26

Wang JH, Reinherz EL (2002) Structural basis of T cell recognition of peptides bound to MHC molecules. Mol Immunol 38:1039–1049

Wilson DJ, McVean G (2006) Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172:1411–1425

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood CABIOS 13:555–556. http://abacus.gene.ucl.ac.uk/software/paml.html

Yang Z, Bielawski JP (2000) Statistical methods to detect molecular adaptation. Trends Ecol Evol 15:495–503

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479

Journal : Large 10709 | Dispatch : 19-6-2008 | Pages : 11
Article No. : 9284 | ☐ LE | ☐ TYPESET
MS Code : GENE787 | ☑ CP | ☑ DISK