Gerald A. Capraro, Jr. POLYMORPHISM OF THE MHC CLASS II *DAB* LOCUS IN THE SWORDTAILS *Xiphophorus multilineatus* AND *Xiphophorus pygmaeus*. (Under the direction of Dr. Thomas J. McConnell). Department of Biology, April 2002.

Major histocompatibility complex (MHC) class II αβ heterodimers on the surface of antigen presenting cells bind and present peptides to $CD4^+$ T cells (T_H). This is key in the activation of T_H cells to initiate a specific immune response. Polymorphism, the presence in a population of many alleles at a given locus, manifests itself, with respect to MHC genes, as non-synonymous nucleotide substitutions within regions encoding the peptide-binding domain of an MHC protein. This genetic variability within the MHC allows the organism to bind a wide repertoire of peptides. The *DAB* locus encodes a β-chain glycoprotein of the MHC class II heterodimer. We have cloned and sequenced cDNAs of *DAB* from two populations each of two species of swordtails, *Xiphophorus multilineatus* and *X. pygmaeus*, and compared them to determine the level of polymorphism. Analyses show nucleotide substitution patterns consistent with amino acid replacement at the putative peptide binding sites, as well as allelic lineages showing the persistence of these polymorphisms across species. Furthermore, phylogenetic analyses suggest that polymorphisms observed in these Xiphophorins are under positive selection. Evidence presented here also shows that the rates of synonymous substitution are higher within the peptide-binding region than elsewhere in the gene, similar to some mammalian species. Finally, the data observed within this study show evidence of at least two distinct *DAB* loci. Understanding the nature of the *DAB* locus will lead to a better understanding of the evolution of the *Xiphophorus* genus, as well as a clearer understanding of the gene duplication events that contributed to the evolution of the MHC in general.

POLYMORPHISM OF THE MHC CLASS II *DAB* LOCUS

IN SWORDTAILS *Xiphophorus multilineatus*

AND *Xiphophorus pygmaeus*

by

Gerald A. Capraro, Jr.

APPROVED BY:

DIRECTOR OF THESIS _____
Thomas J. McConnell, Ph. D.

COMMITTER MEMBER _____
John W. Stiller, Ph. D.

COMMITTER MEMBER _____
Kyle G. Summers, Ph. D.

COMMITTEE MEMBER _____
Kathryn M. Verbanac, Ph. D.

CHAIR OF THE DEPARTMENT OF BIOLOGY _____
Ronald J. Newton, Ph. D.

DEAN OF THE GRADUATE SCHOOL _____
Thomas L. Feldbush, Ph. D.

POLYMORPHISM OF THE MHC CLASS II *DAB* LOCUS IN SWORDTAILS *Xiphophorus multilineatus* AND *Xiphophorus pygmaeus*

A Thesis

Presented to

the Faculty of the Department of Biology

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Molecular Biology/Biotechnology

by

Gerald A. Capraro, Jr.

April 2002

Dedication

This work is dedicated to my wife, Jessica, my parents, Donna and Jerry Capraro, and my sister, Michelle, without whose support this thesis would never have been completed. I would also like to recognize the efforts of my Great Dane, Van Gogh. It was the look in his big dumb eyes that inspired me to keep going during the writing of this thesis. Finally, to Jason and to the men of C-section, Craig, John, and Tucci, your friendship and support will be forever appreciated.

Acknowledgments

I wish to acknowledge several people whose efforts were important in the development of my academic career. First and foremost, the members of my committee, Dr. Tom "T-Bone" McConnell, Dr. John Stiller, Dr. Kyle Summers, and Dr. Kathryn Verbanac, were invaluable in the direction of this thesis. I would also like to acknowledge the efforts of Ulla Godwin, who stressed the importance of positive controls in any experiment, and Denise Mayer, the sequencing goddess, who sacrificed the live chicken before each and every one of my sequencing runs. Many thanks also go to my undergraduate advisor, Dr. Steve Coggin, who convinced me to continue my studies in Biology, and to my high school biology teacher, Mrs. Gizowski, who introduced me to the study of the wonderful world of life. Finally, to Josh Pitzer, the captain of Team T-Bone, your insanity is only outdone by my own!

Table of Contents

List of Tables

List of Figures

Introduction

MHC and Specific Immunity

The major histocompatibility complex (MHC) is a set of genes that encode proteins involved in the acquired immune system. MHC molecules are responsible for recognizing and binding processed antigen for presentation to T cells for the purpose of initiating a specific immune response. MHC class I molecules typically function in cell-mediated immunity, where they present antigen to T_C cells. MHC class II molecules generally participate in both humoral and cell-mediated immunity, where they present antigen to T_H cells. These steps are key elements in the development of an immune response specific to the antigen.

Foreign and self protein antigens are processed by antigen-presenting cells (APCs) within the endocytic processing pathway. Antigens enter the cell via phagocytosis/endocytosis and are degraded into small peptides, which bind to MHC class II molecules within specialized vesicles called endosomes. These endosomes are then transported to the cell surface for presentation of the antigen to T_H cells via the MHC class II molecule. This is a key event in activating T_H cells, which can in turn activate B cells or T_C cells, thus initiating a specific immune response to the antigenic peptide.

MHC class II molecules are heterodimers composed of an α-chain and a β-chain, each of which is encoded by a distinct gene (*A* and *B*, respectively). Each chain of the $\alpha\beta$ heterodimer consists of two extracellular domains (α_1 and α_2, β_1 and β_2), a connecting

peptide, a transmembrane segment, and a cytoplasmic tail. Whereas the β_2-domain is part of the immunoglobulin superfamily and provides structural support to the molecule, the membrane distal domain (β_1) contains the peptide-binding region (PBR) of the MHC class II molecule (Brown *et al.* 1993). This structure, in conjunction with the PBR of the α_1 domain, is where the peptide is actually held in place through non-covalent interactions, particularly, hydrogen-bonding forces. The peptide-binding cleft is composed of antiparallel α helices on the sides and antiparallel β-pleated sheets on the bottom. It is within this cleft that the processed peptide associates with the MHC class II molecule and is presented to T_H cells.

MHC class II molecules typically bind peptides of 13-18 amino acid residues and hold the peptide at a roughly constant elevation above the floor of the peptide-binding cleft (Brown *et al.* 1993). Much like a foot-long hot dog being held within a hot dog bun, the PBR is open at both ends to allow longer peptides to extend outside of the PBR. Contact is made between the MHC class II molecule and the antigenic peptide through internal conserved motifs. Non-covalent interactions between the class II molecule and the backbone of the peptide are distributed throughout the PBR. The peptide itself tends to have a conserved core sequence of 7-10 amino acids, with an aromatic or hydrophobic residue at the amino terminus and three hydrophobic residues in the middle portion and at the carboxyl terminus. Brown and colleagues (1993) demonstrated in humans that codons within exon two are the key polymorphic sites that confer the ability to recognize and bind a wide repertoire of antigenic peptides on the MHC class II molecule.

Polymorphism

A polymorphism refers to a locus at which there are multiple alleles present. With respect to MHC molecules, most polymorphisms are nonsynonymous nucleotide substitutions within exon two. These substitutions result in key amino acid changes within the PBR and provide extensive variability (Hughes and Hughes 1995), thus allowing the MHC molecule to bind a wide array of foreign peptides. The evolutionary advantage of MHC polymorphisms is to allow an individual to survive that can successfully mount an immune response to an extensive variety of parasitic antigens.

Closely related to polymorphism is the phenomenon that the MHC is polygenic as well. Since the MHC is inherited as a complete haplotype, progeny of a given cross of parents will express MHC molecules from both the maternal and paternal MHC loci on the surface of a single cell. The presence of many loci, as well as multiple alleles at each locus, allows an organism to have many varieties of MHC class II molecules on the cell surface that can present a wide repertoire of processed antigenic peptides. This leads to a robust immune system that can be successful in mounting many different specific immune responses.

Codons of the PBR have been shown to be under direct positive selection. In most organisms the rate of non-synonymous nucleotide substitution (i.e., resulting in an amino acid change) per nonsynonymous site is greater than the rate of synonymous substitution per synonymous site within this region of the gene (Hughes and Nei 1989). It is just the opposite for all other regions of the gene; the rate of synonymous substitution

per synonymous site is greater than the rate of non-synonymous substitution per nonsynonymous site. Substitutions that enhance the ability of an MHC molecule to bind a specific antigen are positively selected and may be retained by members of an allelic lineage for long periods of time (Hughes and Yeager 1998). The literature is flush with examples of polymorphism at MHC loci, including in the endangered Przewalski's horse (Hedrick *et al.* 1999) primates (Kriener *et al.* 2000), and teleosts such as salmon (Kim *et al.* 1999), carp (van Erp *et al.* 1996), and striped bass (Walker and McConnell 1994). There are a few examples of organisms that demonstrate a lack of polymorphism at the MHC. Perhaps the most notorious of these is the East African cheetah (O'Brien *et al.* 1987), which is thought to have gone through two severe population bottlenecks, that severely limited the MHC variability of this Great Cat.

Fish MHC and *Xiphophorus*

The first MHC genes characterized in fish were isolated from the carp *Cyprinus carpio* (Hashimoto *et al.* 1990). Subsequently, MHC genes have been isolated from other fish species including striped bass, rainbow trout, zebrafish, catfish and guppy, as reviewed by Klein *et al* (1997). Teleosts were once thought to be the most anciently diverged class of vertebrates demonstrating acute allograft rejections, which are evidence for the presence of MHC gene products (Hashimoto *et al.* 1990). However, in the early 1990s it was discovered that cartilaginous fishes, which evolved earlier than teleosts, also

have MHC class II *A* genes (Kasahara *et al.* 1993) and *B* genes (Bartl and Weissman 1994).

Xiphophorus fishes (Teleostei: Poeciliidae) also express MHC genes and proteins (McConnell *et al.* 1998, Figueroa *et al.* 2001). These are livebearers that are native to the freshwater drainages of eastern Mexico and Central America. The genus *Xiphophorus* is categorized into three groups: the northern swordtails, the southern swordtails, and the platyfish, all of which are popular as aquarium fish. Swordtails display an elongate, colored structure (the sword) extending from the lower portion of the caudal fin (Basolo 1995). The coloration sometimes includes a stripe at the upper boundary (the upper stripe) but always includes a stripe at the lower boundary (the lower stripe). Swords are only displayed by members of the northern swordtails and southern swordtails, and length can vary from species to species. Generally, it is only presenting males; although, there are a few examples of elongated caudal fins in older or androgenized females (Basolo 1995). Platyfish species do not naturally grow a swordtail.

These fishes have also been the focus of much research into the development of tumors beginning with Gordon (1927), which led to the development of inbred lines (Kallman 1958). This, in turn, has led to the publication of extensive *Xiphophorus* gene maps (Morizot *et al.* 1979, Morizot *et al.* 1991).

McConnell and coworkers (1998b) described the existence of two class II *B* loci from *Xiphophorus* that mapped to separate linkage groups. The *DXB* locus, which maps to Linkage Group (LG) III, is paralogous with the *DAB* locus, which maps to the novel

linkage group LG U24; moreover, these two distinct loci both are expressed in *Xiphophorus* (McConnell *et al.* 1998b). The *DAB* locus described by these scientists has a high degree of genetic identity (98%) with *DAB* sequences isolated by Sato and coworkers in 1995 from the guppy.

MHC class II *DAB*

As discussed above, the *DAB* locus of *Xiphophorus helleri* encodes a β-chain of the MHC class II molecule (McConnell *et al.* 1998b). The β_1 domain comprises 89 amino acid residues, while the β_2 domain contains 94 amino acid residues. The 23 amino acid transmembrane segment is linked to the β_2 domain by a short connecting peptide of 10 amino acids. The intracellular portion of the β-chain is a 13 residue cytoplasmic tail. A leader peptide of 22 amino acid residues begins the β-chain protein. Though no studies have confirmed this in fish, the leader peptide is presumably cleaved as the newly synthesized β-chain is entering the endoplasmic reticulum.

Each domain of the β-chain is thought to be encoded by a separate exon, with the exception of the connecting peptide and transmembrane segments, which are encoded by one exon (McConnell *et al.* 1998a). Exon 1 encodes the leader sequence; exon 2 encodes the β_1 domain, and exon 3 encodes the β_2 domain. The connecting peptide and transmembrane region are encoded by exon 4, and the cytoplasmic tail is encoded by exon 5 (Figure 1). The peptide-binding regions of MHC class II molecules are encoded by the most polymorphic regions of their respective genes (Ono *et al.* 1993).

Allelic Lineages

Allelic lineages have been defined as groups of alleles at a given locus that share high nucleotide sequence homology and may be functionally identical. Phylogenetic dendrograms generated from sequence data of related alleles from different species often do not agree with taxonomic relationships. Instead, alleles from different species that are very similar in sequence will cluster together on an evolutionary tree. For example, consider a tree generated from a dataset composed of five nucleotide sequences from species A and five from species B. On the tree, if one sequence from species B clusters with the five from species A, this would represent a trans-species allelic lineage. The implication here is that the allele that gave rise to the sequences in this hypothetical cluster was already present in the common ancestor of these two species and has been maintained in both during their subsequent evolution. There have been published reports of such trans-species allelic lineages of MHC genes found in such organisms as teleosts and mammals.

One report by Sültmann *et al.* (1994) described the MHC class II *B* loci of zebrafish (*Brachydanio rerio*). This study showed a phylogeny that clustered the zebrafish and carp (*Cyprinus carpio*) class II *B* loci together apart from the class II loci of other vertebrates, including other teleosts. This suggests that all of the known zebrafish and carp class II *B* genes originated from a single ancestral gene that was different from the ancestors of the other allelic clusters from class II *B* loci.

Another example of allelic lineages comes from *Xiphophorus* (Figueroa *et al.* 2001). These researchers demonstrated allelic lineages of MHC class I sequences between various *Xiphophorus* species as well as between *Xiphophorus* and other teleosts. Evolutionary dendrograms generated in this study group MHC class I sequences of different species of *Xiphophorus* together (e.g., *X. montezumae* clusters with *X. helleri*). The tree also shows a clade in which the medaka (*Oryzias latipes*), the guppy (*Poecilia reticulata*), the cichlid fish (*Aulonocara hansbaenschi*), and the stickleback (*Gasterosteus aculeatus*) cluster with various *Xiphophorus* species. Each of these non-*Xiphophorus* fish is from a different taxonomic order (Beloniformes, Cyprinodontiformes, Perciformes, and Gasterosteiformes, respectively) yet their sequences cluster together with specific *Xiphophorus* allelic lineages. It is interesting to note that in this study the carp groups outside of all *Xiphophorus* sequences, whereas in the study by Sültmann *et al.* (1994) the carp clustered with *Xiphophorus*. This may indicate that MHC class I and class II alleles have separate evolutionary histories.

Hypothesis

Transcribed classical MHC class II *B* loci demonstrate polymorphisms in nearly all jawed vertebrates characterized to date. The hypothesis tested in this thesis was that *DAB* is polymorphic in *Xiphophorus*, and should show evidence of positive selection at the peptide-binding region. Also, as the presence of multiple loci has been described in nearly all organisms that demonstrate an MHC (or HLA in humans), we further hypothesized that *DAB* is represented by more than one locus in *Xiphophorus*. Understanding the nature of the *DAB* locus will lead to a better understanding of the evolution of the *Xiphophorus* genus, as well as a clearer understanding of the gene duplication events that contributed to the evolution of the MHC in general.

Materials and Methods

Amplification and Cloning of *DAB* Alleles

Wild-caught members of two species of *Xiphophorus* fishes were used to assess the polymorphic nature of the MHC class II *DAB* locus. Thirty-nine fish from two populations of *Xiphophorus multilineatus* and *Xiphophorus pygmaeus* were obtained by our collaborators and used to generate single stranded cDNA from mRNA of the gut tissues. Populations of *Xiphophorus multilineatus* were sampled from the Rio Coy and Arroyo Tambaque regions of eastern Mexico, and populations of *Xiphophorous pygmaeus* were sampled from the Huichihuayán and La Y-Griega regions. Table 1 describes each fish sample and the Global Positioning System (GPS) coordinates where each was caught. Collaborators (Luis Della Coletta, University of Texas – MD Anderson Cancer Center) used the Trizol Reagent (Gibco, Carlsbad, CA) to extract the RNA and the Gibco First Strand preamplification system using the oligo-dT strategy to generate first strand cDNA. This was followed by the use of RNAse H to produce single stranded cDNA.

The polymerase chain reaction (PCR) was then used to amplify full-length *DAB* sequences from cDNA templates. Each template was amplified one time, with the exception of *Xipy*14, which was used in ten separate PCR reactions. Primers TM 396A (5'–GCTGGGCTGGCTGCTGGTCAT–3') and TM 398A (5'–GAAGCAGGAGGAAC CAGAACC–3') were designed, respectively, based on the leader sequence and the 3'

untranslated region of the guppy (Sato *et al.* 1995). Using the PlatinumTaq Hi-fidelity polymerase enzyme (Invitrogen Corporation, Carlsbad, CA), a reaction mixture of 32.8 μL millipore water, 1 U PlatinumTaq HF, 5 μL 10x HiFi buffer, 1 μL 10 mM dNTP mix, 2 μL 50 mM $MgSO_4$, 4 μL TM 396A (1.5×10^{-3} mM), and 4 μL TM 398A (1.5×10^{-2} mM), and 50 ng cDNA template were subjected to thirty cycles of amplification (94°C for one minute, 60°C for one minute, and 68°C for two minutes) in a minicycler™ (MJ Research, Inc., Waltham, MA). The 5' primer (TM 396A) was used in a 10-fold less concentration than the 3' primer (TM 398A) because of the need for a limiting reagent for the next step, PCR+1 (Borriello and Krauter, 1990). At the end of the 30 cycles, each reaction was spiked with 10 μL of primer TM 399 (1×10^{-2} mM) and 1 U PlatinumTaq and subjected to an additional cycle of 94°C for 5 minutes, 55°C for 3 minutes, and 68°C for twenty minutes. The reasons for performing the PCR+1 step were two-fold. First, the technique limited cross-hybridization of DNA strands from more than one *DAB* locus and ensured that proper DNA strands annealed to each other. Second, the TM 399 primer (5'—AGAAAGCTTGCTGGGCTGGCTGCT—3') included a *Hind* III restriction site (underlined), which was used in further analysis following cloning of the transcript. PCR products then were analyzed by electrophoresing 3 μL of each reaction, mixed with 1 μL loading dye, through a 1% agarose/Tris-Borate-EDTA buffer (TBE) gel matrix at 60 mA for one hour and visualizing with ethidium bromide exposed to UV light.

PCR reactions yielding amplification of the proper size bands were cloned directly from the PCR mix into the pGem-T Easy cloning vector (Promega Corporation,

Madison, WI). Ligation reactions for each PCR product were set up with 2.5 μL of 2X Rapid Ligation Buffer, 25 ng pGem-T Easy vector, 1.5 U T4 DNA Ligase, 17.5 ng insert DNA, and deionized water to a final volume of 5 μL and allowed to incubate overnight in a 0.5 mL microcentrifuge tube at 4°C. These ligation reactions were then used to transform competent cells via electroporation.

Top 10 *Escherichia coli* cells (Invitrogen Corporation, Carlsbad, CA), in 40 μL aliquots were thawed on ice, and then 1-2 μL of the ligation reaction were gently swirled into the cells with a pipette tip. This mixture was then transferred to a cold Gene Pulser® electroporation cuvette (BioRad Laboratories, Hercules, CA) with a 0.1 cm electrode. The cuvette was transferred to the chilled safety chamber slide and seated between the contacts of the of the Gene Pulser apparatus (BioRad Laboratories, Hercules CA). The cells were electro-transformed at a capacitance setting of 25 μF, a Pulse Controller setting of 200 Ω, and a voltage setting of 1.8 V. Immediately following electroporation, 800 μL of room temperature SOC medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) were added to each cuvette to maximize the recovery of the transformants. The cell suspensions were then transferred to a 1.5-mL microcentrifuge tube and incubated at 37°C at 225 RPM for one hour. At the end of the incubation, 75 μL of each cell suspension were plated onto Luria-Bertani agar supplemented with X-Gal (80 μg/mL) and ampicillin (100 μg/mL) and incubated at 37°C overnight. Standard blue-white differentiation was used to assess the presence of an insert in the transforming plasmid.

Colony growth indicated uptake of the plasmid because resistance to ampicillin was encoded within the pGem-T Easy vector. Furthermore, the growth of white colonies demonstrated the insertion of DNA into the *lacZ* gene of the plasmid, interrupting the coding sequence of β-galactosidase. Cells that were transformed by a plasmid with no insert grew into blue colonies because they produced β-galactosidase, which reacted with the X-gal in the plates.

Plasmid Purification and Restriction Analysis of Putative *DAB* Clones

Positive clones were cultured in 10 mL Luria-Bertani broth supplemented with ampicillin (100 μg/mL) at 37°C at 225 RPM for a period of 10-12 hours and processed using the Concert Rapid Plasmid Miniprep System (Invitrogen Corporation, Carlsbad, CA) to purify the plasmids from these cells. To harvest the cells, 1.5 mL of culture were pelleted in an Eppendorf centrifuge 5415C (Westbury, NY) at 14000 RPM for 10 minutes, and the supernatant was decanted. This step was repeated to obtain 3 mL worth of cells that were then resuspended in 250 μL of Cell Suspension Buffer containing RNase A. Following the addition of 250 μL of Cell Lysis Solution the cells were allowed to incubate in a 37°C water bath for 20 minutes for optimal lysis. Lysis was terminated by the addition of 350 μL of Neutralization Buffer. This mixture was centrifuged at 14000 RPM for 10 minutes, and the supernatant was loaded onto a spin cartridge and placed into a 2-mL wash tube. This was centrifuged at 14000 RPM for 1 minute, and the flow-

through was discarded. A wash was performed with the addition of 500 μL of Optional Wash Buffer to the spin cartridge followed by centrifugation for one minute. A second wash was then performed with the addition of 700 μL of Wash Buffer containing ethanol followed by centrifugation for one minute. The flow-through was discarded, and the cartridge was centrifuged again to remove any residual wash buffer. The spin cartridge was placed into a 1.5-mL recovery tube, and the plasmids were eluted with 60 μL of 70°C millipore water and centrifuged at 14000 RPM for 2 minutes. Purified plasmids were then analyzed with restriction enzymes.

Restriction analyses using both *Eco*R I and *Hin*d III (Invitrogen Corporation, Carlsbad, CA) were performed on each sample to remove the DNA insert from the plasmid vector as a means of identifying useful plasmids for sequencing. *Eco*R I was used to separate the entire insert from the rest of the plasmid, while *Hin*d III was used to linearize the plasmid at the PCR+1 site. The *Eco*R I digest was performed by mixing 3 μL of the purified plasmid with 14 μL of millipore water, 2 μL of 10X React 3 buffer, and 10 U *Eco*R I. The *Hin*d III digest mixed 3 μL of the plasmid with 14 μL of millipore water, 2 μL of 10X React 2 buffer, and 10 U *Hin*d III. These reactions were incubated in a 37°C water bath overnight, and the results were observed via electrophoresis. Each reaction was mixed with 1 μL loading dye, and 7 μL of each sample were electrophoresed through a 0.9% agarose/TBE gel at 60 mA for one hour. Bands in the gel were visualized with ethidium bromide exposed to UV light.

DNA Sequencing and Analysis of *DAB* Plasmids

DNA sequencing was performed on recombinant plasmids containing a PCR+1 insert, as demonstrated by restriction digest analysis, by the dideoxynucleotide chain termination method (Sanger *et al.* 1977) using the fluorescently-labeled Big Dye Terminator sequencing reagent (Applied Biosystems, Inc., Foster City, CA). One clone was sequenced from each of the ten PCR reactions of *Xipy*14. Two clones were sequenced from each of the other fish samples. Universal M13 forward and reverse primers along with six internal gene specific primers (Life Technologies, Gaithersburg, MD) were used in the sequencing reactions. Table 2 shows the list of primers used in this study. Sequencing reactions were set up in a total volume of ten microliters with 2 μL of Big Dye Terminator mix, 200 ng of primer, 200 ng of plasmid template, and millipore water. These reactions were subjected to thirty cycles each of denaturation at 96°C, annealing at 50°C, and extension at 60°C. The DNA from each reaction was put into a mixture of 50μL of cold 95% ethanol and 3μL of cold 2M sodium acetate, pH 4.1, and allowed to precipitate overnight at -20°C. The precipitated DNA was pelleted by centrifugation (Eppendorf, Westbury, NY) at 4°C at a speed of 14000 RPM for thirty minutes. The supernatant was decanted, and the pellets were washed with 100μL of cold 70% ethanol; then the samples were centrifuged again, this time at room temperature, at a speed of 14000 RPM for five minutes. To prevent aspiration of the pellet the supernatant was drawn off by capillary action with a 200 μL pipette tip. The pellets were dried at 37°C for 15 minutes and resuspended in 5μL deionized formamide and 50mM EDTA

(5:1, v:v). The resuspended pellets were denatured (95°C) for 90 seconds, and they were immediately placed on ice. Sequencing reactions were then loaded onto an acrylamide sequencing gel made up of a mixture of Long Ranger acrylamide gel solution (BioWhittaker Molecular Applications, Rockland, ME), urea, TBE, and water, and polymerized by the actions of TEMED and a 10% solution of ammonium persulfate. The volume of sample loaded onto the gel was 1.5 µL, and the samples were electrophoresed for a period of seven hours. Sequences were analyzed using an Applied Biosystems 377 Automated DNA Sequencing System (Foster City, CA).

Resulting sequences were edited using the ABI Prism Sequencing Analysis software version 3.4.1 (Foster City, CA). Overlapping sequences for each clone served as a control for individual sequence reaction errors. Sequences were assembled with the Autoassembler program (Applied Biosystems, Inc., Foster City, CA). The BLAST (Altschul *et al.* 1990) program was used to compare the generated sequences with those in the GenBank database. The "blastn" program was used to search the nucleotide database and confirm the identity of each sequence based on known *DAB* sequences, and the "blastx" program was used to compare the nucleotide sequences, translated in all three reading frames, against the protein sequence database to verify that the cDNA sequences encoded DAB-like proteins. Sequences were aligned using the Genetics Computer Group set of sequence analysis programs (GCG, Devereux *et al.*, 1984) and the ClustalX program (Thompson *et al.*, 1997). The Molecular Evolutionary Genetics Analysis program version 2.1 (MEGA2, Kumar *et al.*, 1993) was then used for distance

analysis of the aligned sequences. Genetic p-distances were calculated using the Jukes-Cantor corrected distance method (1969), which is a simple model of nucleotide substitution that considers the possibility of multiple substitutions at a given site in a sequence and assumes that a nucleotide can be substituted by any other nucleotide with equal probability. The rates of synonymous nucleotide substitutions per synonymous site and nonsynonymous nucleotide substitutions per nonsynonymous site were also calculated using MEGA. Both MEGA and the program Phylogenetic Analysis Using Parsimony (PAUP*4.0b) (Swofford, 2000) were used to generate evolutionary trees based on simple Jukes-Cantor corrected distance data. Bootstrapping analysis of 500 replicates was used to assess the reliability of individual nodes on these trees.

Results

PCR, Cloning, and Sequencing of *DAB* Alleles

Gene fragments obtained from polymerase chain reaction using primers TM 396A and TM 398A were predicted to be approximately 700 base pairs (bp) in length based on the guppy sequence. This was confirmed upon electrophoresis of the PCR samples by comparison with a 100 bp ladder.

Subsequent sequencing found these products to be 777 nucleotides in length. After discounting the TM 399 primer site and the 3' UTR immediately following the stop codon, the resulting sequences had lengths of 735 nucleotides. The sequences spanned the length of the gene from 7 codons past the presumed start site through the stop codon (TGA). Each of the 86 sequences recovered had at least 90% identity to the *XimaDAB**01 sequence identified by McConnell *et al.* (1998b). When translated using the GCG program, each sequence encoded a protein of 244 amino acid residues similar (with greater than 95% identity) to known DAB proteins in public databases. Sequences were aligned using the GCG Pileup program, and the alignments for each population can be found in Figures 2, 3, 4, and 5.

Evolutionary Distance Calculations

The Jukes-Cantor corrected distance algorithm, used to assess genetic differences among *DAB* sequences, corrects for potential multiple nucleotide substitutions per site in a given sequence. Table 3 summarizes the results of distances analyses in MEGA. For

the entire dataset (i.e., 86 sequences) the average d_{JC} was determined to be 0.067 ± 0.006. This number is similar to the average d_{JC} calculated for each species when analyzed separately. The *Xiphophorus pygmaeus* sequences had an average d_{JC} of 0.063 ± 0.006, and the *Xiphophorus multilineatus* sequences had an average d_{JC} of 0.066 ± 0.006. The d_{JC} was then calculated for each population within each species. Both *X. multilineatus* populations showed similar values with the Rio Coy population having a mean distance of 0.060 ± 0.006, and the Arroyo Tambaque population having a mean distance of 0.063 ± 0.006. Conversely, the *X. pygmaeus* populations were significantly different from each other with the Huichihuayán population at 0.048 ± 0.005 and the La Y-griega Vieja population at 0.061 ± 0.006. This interpopulation difference may represent sampling error, since the Huichihuayán population was represented by only six fish, and the La Y-Griega Vieja sample was larger by two fish.

The average number of nucleotide differences between species was also assessed. *Xiphophorus pygmaeus* showed a mean of 180 nucleotide differences, while *X. multilineatus* showed 204. The majority of these changes occurred within Exon 2 (the β_1-encoding domain) with *X. pygmaeus* having an average of 144 and *X. multilineatus* having an average of 159. Comparison of average differences between populations of each species produced the same pattern as for the complete datasets. There was an average of 101 nucleotide differences within the sequences of the Huichihuayán population of *X. pygmaeus*, and 88 of these occurred in Exon 2. Within the La Y-griega Vieja population there was an average of 156 nucleotide differences, 126 of these occurring within Exon 2. An average of 144 nucleotide differences was identified in the

Rio Coy population of *X. multilineatus*, 117 of these found in Exon 2, and a mean of 178 differences was discovered in the Arroyo Tambaque population, with 149 in the β_1-encoding region.

Another important aspect of these *DAB* sequences is the presence of a six-site gap. This gap occurred within Exon 1 and only in 12 sequences from *X. multilineatus*; 11 of these were from the Rio Coy population, while just 1 came from the Arroyo Tambaque population. Eight of these sequences were sister clones (i.e., clones from the same fish) coming from *Ximu17*, *Ximu20*, *Ximu22*, and *Ximu26*. The remaining 4 sequences each came from a different fish (*Ximu19*, *Ximu21*, *Ximu23*, *Ximu32*), and each had a sister clone that did not demonstrate the presence of this gap. Although the gap retained the reading frame of these coding sequences, it resulted in the leader sequence being shortened by two amino acid residues. At this time it is unclear what effect the presence of this gap might have on cleavage of the leader sequence *in vivo*.

To assess the polymorphic nature of the MHC, one must take into account the number of nucleotide differences that result in amino acid changes in the encoded protein. Further, comparison was made between the rates of synonymous nucleotide changes per synonymous site (d_S) and nonsynonymous nucleotide changes per nonsynonymous site (d_N). These d_S vs. d_N values can be found in Table 4. Considering all coding DNA except Exon 2, *Xiphophorus pygmaeus* sequences were marked by a d_S (0.023 ± 0.008) that was larger than the d_N (0.009 ± 0.003). A similar result was observed for *X. multilineatus* sequences. The d_S (0.018 ± 0.007) was larger than the d_N (0.009 ± 0.003) for all portions of the sequence except Exon 2. When making these

calculations using Exon 2 alone, however, both *X. pygmaeus* and *X. multilineatus* showed d_N values that were larger than the d_S values. For *X. pygmaeus*, the d_N was 0.173±0.027, and the d_S was 0.128±0.028, and for *X. multilineatus*, the d_N was 0.180±0.027, and the d_S was 0.155±0.032. Furthermore, when analyzing these sequences as a single, large dataset, d_S was larger than d_N (0.021±0.007 and 0.009±0.003, respectively) for all portions of the sequence except Exon 2. Likewise, when considering Exon 2 alone, d_N for all 86 sequences exceeded d_S 0.184±0.027 to 0.147±0.028.

Similar patterns were observed within populations of each species. The d_S value for the Huichihuayán population outweighed the d_N value 0.011±0.004 to 0.005±0.002, and the values for the La Y-griega Vieja population were d_S =0.022±0.008 and d_N=0.008±0.003. For the Rio Coy population, the d_S value of 0.017±0.006 was larger than the d_N value of 0.010±0.004. The Arroyo Tambaque values were d_S=0.017±0.007 and d_N=0.008±0.003. Also, when considering Exon 2 alone, the d_N values were larger than the d_S values. Huichihuayán showed a d_N value of 0.132±0.021 and a d_S value of 0.125±0.033, while the La Y-griega Vieja showed a d_N value of 0.171±0.027 and a d_S value of 0.120±0.026. Finally, the Rio Coy population demonstrated a d_N value of 0.162±0.026 and a d_S value of 0.132±0.031, while the Arroyo Tambaque population had d_N and d_S values of 0.174±0.026 and 0.153±0.034, respectively.

Phylogenies

The programs MEGA and PAUP were used to reconstruct phylogenies for assessing evolutionary relationships among these sequences. Trees were generated with

the neighbor-joining method using Jukes-Cantor corrected distances. Figure 6 shows the phylogeny of all eighty-six sequences, plus *DAB* sequences from *Xiphophorus maculatus* (McConnell *et al.* 1998b) and *Poecilia reticulata* (Sato *et al.* 1995) for the purpose of comparison, and a sequence from the catfish *Ictalurus punctatus* for use as an outgroup. Due to the high level of similarity among these *Xiphophorus* sequences the tree shows a large number of polytomies (i.e., unresolved clades); however, the tree generated in this study does yield significant information about evolutionary relationships of *DAB* genes in these species, as evidenced by some high bootstrap values. Another complication of this phylogeny is that we cannot be sure that we are comparing DAB sequences from different alleles or from different loci. That said, the phylogeny in Figure 6 does yield some interesting information.

There were twenty cases of different clones from the same organism clustering together. In all other instances, sister clones clustered apart from one another. Also, *XimaDAB* and *PoreDAB* sequences fell along unresolved branches on the tree. Furthermore, there were twelve sequences that contained a six-site gap in the leader portion. These sequences were evenly split into two distinct clades on the tree. Finally, of the ten sequences from fish *Xipy*14 only eight of them clustered together with high bootstrap support, while the other two sequences (*Xipy*14.1 and *Xipy*14.2) clustered outside of this major clade.

With the exception of four cases, most of the clades on the tree in Figure 6 cluster members of the same species together. Much of the tree clusters *Ximu* sequences apart from clusters of *Xipy* sequences. The exceptions to this generalization are important

because they may be indications of the persistence of trans-species polymorphisms. This has also been referred to as the presence of allelic lineages. Potential allelic lineages are evident in Figure 6 where a sequence from one species clusters in the same clade with sequences from another species (designated by the blue star). An example of a potential allelic lineage is the clade of sister clones, *Xipy*8A, *Xipy*8B, *Xipy*9A, and *Xipy*9B, and in the middle of this clade lies a sequence from *Ximu*36K. Also, at the bottom of the tree is a clade almost entirely comprised of *Ximu* sequences; however, the branch attached to this clade contains sequences from *Xipy*1B, *Xipy*2J, and *Xipy*6A. As stated earlier, the implication here is that the allele that gave rise to the sequences in these clusters was present prior to their speciation and has been maintained since. Again it must be mentioned that another possibility for these sequences clustering together is that they are from similar loci, and not necessarily an allelic lineage.

Furthermore, it is interesting to note that in all but four cases, the species that cluster together on the tree do so according to the populations from which they were caught, excluding those examples of possible trans-species polymorphism. There are only five situations in which clades on the tree contain sequences from more than one population. The top cluster on the tree groups *Ximu*32D of the Arroyo Tambaque population with four other *Ximu* sequences from the Rio Coy population. In another clade, *Ximu*32E clusters with five *Ximu* sequences from the Rio Coy population. Another example of populations clustering together occurs in the clade of the sister clones, *Xipy*11B and *Xipy*11F of the La Y-Griega Vieja population, which group with three *Xipy* sequences, *Xipy*4A, *Xipy*4E, and *Xipy*5G, of the Huichihuayán population. Still another

example of this cross-population similarity occurs with *Ximu*38D, which falls in the middle of a clade containing the sister clones of *Ximu*25 and *Ximu*15. The final example is shown in a small grouping toward the bottom of the tree, which groups *Ximu*19G, from the Rip Coy population, with *Ximu*34E, from the Arroyo Tambaque population.

Finally, there were five instances where species grouped only with members of the same species. The top portion of the dendrogram contains four clades that each group 19 members of *Xiphophorus multilineatus* together. This group includes 8 pairs of sister clones from *Ximu*16, *Ximu*32, *Ximu*21, *Ximu*31, *Ximu*22, *Ximu*26, *Ximu*27, and *Ximu*28. The other three taxa in this group cluster in other regions of the tree, including *Ximu*36K, which clusters with four *Xiphophorus pygmaeus* taxa, again demonstrating possible trans-species polymorphism. There is a clade just beneath this major one that groups 5 members of *X. pygmaeus* together, four of which (*Xipy*11B, *Xipy*11F, *Xipy*4A, and *Xipy*4E) come from two fish. Also, 5 members of *X. multilineatus* group together in the upper-middle portion of the tree; again, four of these (*Ximu*25A, *Ximu*25G, *Ximu*15G, and *Ximu*15H) come from two different fish cDNA samples. A clade that contains 7 members of *X. pygmaeus* can also be found on the tree in Figure 6. There are four sequences in this group (*Xipy*12G, *Xipy*12H, *Xipy*3G, and *Xipy*3F) that come from two *X. pygmaeus* individuals. Finally, there are two major clades that group 9 members of *X. multilineatus* together. *Ximu*17F, *Ximu*17G, *Ximu*19B, *Ximu*19G, *Ximu*20G, and *Ximu*20J are three sets of sister sequences that group together in this clade. Two of the remaining three sequences in this clade (*Ximu*24E, *Ximu*23A, and *Ximu*34E) represent the

sister taxa of sequences from the major *X. multilineatus* clade at the top of the dendrogram.

The data from this phylogeny are consistent with the hypothesis that these species share allelic lineages, and that allelic diversity is maintained across the four populations represented in this study. It is also clear, however, that the populations within each species are quite similar to each other, on the order of only a relatively small number of nucleotide differences in each sequence. This creates a problem when trying to resolve a phylogeny such as the one presented here. Some of the sequences are so similar that their evolutionary relatedness cannot be precisely discerned. It would be prudent to perform further analyses on this phylogeny to confirm or deny our examples of possible allelic lineages.

Assessment of the Number of *DAB* Loci

To assess the number of possible *DAB*-encoding loci, one clone was sequenced from each of ten different PCR reactions of one sample of *Xiphophorus* cDNA (*Xipy*14). The phylogeny shown in Figure 7 represents the relationship among these ten sequences. The major clade of eight sequences is comprised of four identical sequences (*Xipy*14.7, *Xipy*14.8, *Xipy*14.9, and *Xipy*14.10) and four sequences that differ from the identical sequences by one (*Xipy*14.3, *Xipy*14.4, and *Xipy*14.6) or two (*Xipy*14.5) nucleotide substitutions, which occurred in exon 3. The two sequences that fall outside of this clade are *Xipy*14.1 and *Xipy*14.2, which differ from the identical sequences by 48 or 52 nucleotide substitutions. It is also noteworthy that these two sequences differ from each

other by 44 nucleotide substitutions. Nearly all (>90%) nucleotide substitutions within these sequences occurred in exon 2.

From these data, it is reasonable to conclude that *DAB* is encoded by more than one locus. However, the actual number of *DAB* loci is unclear. It is expected that if there were only one *DAB* locus, there would be two sequences at the most. If the nucleotide substitutions within the major clade of nearly identical sequences are, in fact, real, then the tree represents 7 different sequences and therefore at least four *DAB* loci. If, however, the nucleotide substitutions of this clade are merely PCR-induced mutations, either from the PCR enzyme used in amplification of these fragments (putative error rate of 1 in 1 million base pairs) or more likely from the reverse-transcriptase used during RT-PCR (putative error rate of 1 in 15,000 base pairs), there still are 3 clearly different sequences and therefore at least 2 loci. Although the polymerase used for PCR amplification was a high-fidelity enzyme, reverse-transcriptase is prone to error. Therefore, because these sequences were recovered from cDNA, the rigorous conclusion that can be drawn from these data is that *DAB* is encoded by at least two loci.

Discussion

Positive Selection

The fish in this study represent four populations of northern swordtails. The data indicate that the peptide-binding region of the *DAB* locus is under direct positive selection to maintain polymorphisms. This can be seen by comparing the rates of synonymous substitution per synonymous site and the rates of non-synonymous substitution per non-synonymous site both within and outside of exon 2. Positive selection is indicated when the rate of non-synonymous nucleotide substitution outweighs the rate of synonymous substitution (Hughes and Hughes 1995). This conclusion has been supported by such studies as those performed with class I MHC (Hughes and Nei 1988; Imanishi and Gojobori 1992), class II MHC (Hughes and Nei 1989; Schwaiger *et al.* 1994), the circumsporozoite antigen in *Plasmodium* (Hughes 1991), and the human influenza A virus (Fitch *et al.* 1991), among many others. Within exon 2, which encodes the peptide-binding region of the MHC class II DAB molecule, d_N is greater than d_S for these *Xiphophorus* sequences (Table 4). For all other regions of the gene, d_S is greater than d_N. Most non-synonymous substitutions are deleterious to protein structure or function and are quickly eliminated via purifying selection, whereas synonymous substitutions do not change the encoded amino acid, and generally are selectively neutral. However, if positive selection favors amino acid replacements in a certain region (e.g., the PBR), d_N can exceed d_S as we have shown here. In the case of MHC genes, polymorphisms that allow an organism to recognize and bind many different antigenic

peptides afford that organism a better fitness. Even though d_N values are not significantly higher than d_S values within exon 2, the fact that the proportion of d_N/d_S values within exon 2 are significantly higher than the proportion of d_N/d_S values outside of exon 2 indicates positive selection. There is a higher rate of evolution occurring at this region of the gene. Thus there is positive selection for increased variation at the amino acid level.

The persistence of polymorphisms for long periods of time is another indicator of positive selection acting on the MHC (Klein 1987; McConnell *et al.* 1988). Potential allelic lineages shown in Figure 6 demonstrate this, where a sequence from one species, *Ximu36K*, clusters with two sets of sister clones from the other species, *Xipy8A*, *Xipy8B*, *Xipy9A*, and *Xipy9B*. Another key example of an allelic lineage is the grouping of three *Xipy* sequences (*Xipy*1B, *Xipy*2J, and *Xipy*6A) with thirteen *Ximu* sequences from both populations. The implication here is that the allele that gave rise to these sequences preceded the appearance of these two species and has been maintained following their speciation. Trans-species polymorphism generally is believed to indicate positive selection working to maintain the advantage of polymorphism. The fact that this polymorphism has been maintained across species suggests that its function (i.e., binding peptide) confers upon the organism a level of fitness not achieved without it. One complicating factor of this analysis is that it is unclear whether this phylogeny contains paralogues, orthologues, or both. A rigorous phylogenetic analysis would require knowledge of *DAB* loci to distinguish whether apparent allelic lineages were actually due to the sampling of multiple loci. From these data, we cannot distinguish between sequences of different alleles and different *DAB* loci.

These analyses provide evidence that polymorphism of the MHC class II *DAB* locus is selectively maintained and that the selection maintaining this polymorphism is focused on the PBR-encoding region. Therefore, our evidence supports the hypothesis of Doherty and Zinkernagel (1975) that the main force driving the selection is the advantage of being able to bind a wider array of peptides and thus resist a variety of pathogens. Yet, the actual force driving the selection has not been confirmed.

Parasitism vs. Mate Selection

It is still unclear what mechanism maintains the polymorphism of the MHC, but the two favored hypotheses are parasite-host co-evolution and mate selection. The parasite-host co-evolution hypothesis is a relationship between the functionality of the MHC and the antigens against which it is working to protect. Perhaps the best example of this comes from the correlation of the human class I allele, HLA-B53 together with a particular class II haplotype, which conveys resistance to the malaria parasite *Plasmodium falciparum* in West Africa (Hill *et al.* 1991). This gene is under-represented in children suffering from severe malaria and is more common in the overall West Africa population than elsewhere in the world. These two factors suggest that selection by malaria has contributed to the increase in frequency of this protective allele in West Africa. Logically, it follows that the malaria parasite itself would also be under positive selection to evade the host immune system. So, it is a situation of continuing evolutionary conflict between the parasite attempting to evade the host immune system

and the host trying to escape its parasite. Increased genetic variability results from this ongoing battle.

Another example of parasitism increasing the genetic variability of a host immune system comes from Thursz *et al.* (1997). These researchers showed the level of heterozygosity at the *HLA-DRB1* and *HLA-DQB1* loci in a West African population was significantly lower among individuals with persistent hepatitis B virus (HBV) infection than among individuals who had cleared the HBV infection. Again, this indicates that a parasite was driving the increased genetic variability at MHC loci.

To date, it is unclear just what parasites co-exist with the Xiphophorins used in this study, although one might assume that since these are tropical fish, there might be a "host" of parasitic organisms within these fish. The fish examined were wild-caught members of natural populations in the streams of Mexico. It is therefore presumed that they share an environment with some forms of parasites. However, what those parasites may be is unclear at present. Dove and Ernst (1998) described a parasitic worm known as *Gyrodactylus bullatarudis* that infected *Xiphophorus helleri* in Australia. It is believed that this is the first record of a *G. bullatarudis* infection in wild *Xiphophorus* spp. However, it is unknown what type of infection this parasite causes in *Xiphophorus*; also, it is unclear if this organism infects *Xiphophorus* spp. in the streams of Mexico.

The other possibility for the maintenance of MHC polymorphisms is mate selection. Typically, the two explanations for mate choice are: 1.) to avoid interbreeding that would generate progeny with deleterious recessive alleles, and 2.) to increase genetic diversity to better deal with parasites. The latter concept tends to be invoked to explain

genetic variability at the MHC. Selection mediated by mating preferences indirectly promotes MHC polymorphisms. With this proposed mechanism, it is generally the female of a species that selects a mate with which to produce offspring. The resulting offspring may have a higher level of fitness than either of the parents, including increased variability at the MHC region. This concept is not unique to ectothermic vertebrates. It has also been shown that mice (Penn and Potts 1998) and humans (Ober *et al.* 1997) prefer mates that carry different MHC alleles.

Mate selection has been clearly demonstrated in swordtails. Female swordtails are known to exhibit a preference for males with long swords (Basolo 1990; Basolo 1995); theoretically, the longer the sword, the better genes the male will provide for the offspring. One exquisite study by Basolo (1990) showed that female platyfish (*Xiphophorus maculatus*) preferred conspecific males with artificial swords. Also, a study done by Morris and coworkers (1996) showed that large body size, not just sword length, was important for mate preference. Females of *X. pygmaeus* preferred larger males of *X. nigrensis* to the smaller males of their own species. As Burnet argued (1971), the mate selection hypothesis could account for the high levels of polymorphism in the vertebrate immune system. This might include the MHC class II loci of Xiphophorins.

d_S is Higher in Exon 2 Than Elsewhere in the Gene

One important question that arises from these data is: why is the rate of synonymous substitution per synonymous site larger in exon 2 than it is elsewhere in *DAB*? Since synonymous substitutions are selectively neutral, the rates of these types of

mutations should be relatively constant over the entirety of the gene. As demonstrated in Table 4, d_S values are higher in exon 2 for the *Xiphophorus* sequences in this study. One hypothesis put forth in the literature is the possibility of gene conversion.

In a series of recent papers, Ohta (1998; 1999) described the effect of gene conversion on increasing the polymorphism of the MHC. He used computer simulation studies to observe the MHC variability that arose as a result of gene conversion. Synonymous substitution increased as interlocus conversion became more frequent, possibly because this type of conversion supplies novel mutations to a locus thereby contributing to the high level of genetic variability, whereas, intralocus conversion simply reshuffles amino acids. His results complement the results seen in other non-primate mammals, specifically, that the rate of synonymous nucleotide substitution is higher within the PBR than outside of it. Organisms that exhibit this phenomenon include the mouse (Hughes and Nei 1988), bovine DRβ3 (Mikko and Anderson 1995), pig DQβ (Ohta 1995), and human DRβ (Bergstrom *et al.* 1998). It is believed that the research described in this thesis is the first report in fish of higher rates of synonymous substitution within the PBR than elsewhere in the gene.

It would seem that after positive selection maintains the polymorphisms within the PBR-encoding exon, gene conversion functions to homogenize non-PBR exons. This would have the effect of seemingly lowering the rates of synonymous substitution per synonymous site, since one gene was converted to another, thereby lessening the observed variability in these regions compared to exon 2.

A review by Martinsohn and coworkers (1999) provides the possible mechanism of gene conversion. A double-strand break-repair mechanism (DSBR) as originally described by Szostak et al. (1983) would be a logical probable mechanism for gene conversion, if this turned out to be the case in *Xiphophorus*. This model (Figure 8) generally functions by a double-strand break followed by exonucleolytic degradation at the 5' end, which leads to a double-strand gap with a 3' overhang of heteroduplex DNA (hDNA). One of the free ends then invades a region of homologous DNA in the donor duplex and acts as a primer for DNA synthesis, so that a D-loop is displaced in the donor and subsequently enlarged. The D-loop then anneals to the complementary sequences on the other side of the gap, priming another round of DNA synthesis from the other end. Once the gap is filled on both strands, ligation occurs forming two Holliday junctions. Resolution of these junctions in opposite orientations yields a crossover. Cleavage of the two Holliday junctions independently of each other results in half of all gene conversions being associated with crossovers. This model predicts conversion as a consequence of strand switching between loci, coupled with repair synthesis and mismatch repair of hDNA.

Because *Xipy* and *Ximu* sequences are similar, both within each species and to each other, it seems reasonable that gene conversion would occur in fish (possibly due to some kind of DSBR mechanism described above). Gene conversion, therefore, could lead to the polymorphism observed in exon 2 of the sequences in this study. Indeed, since the high degree of homology occurs outside of exon 2, these regions of the gene provide excellent sites of annealing to create a Holliday junction. Then, the mismatch

repair mechanism would correct nucleotides that were incorrectly paired. It has yet to be demonstrated, however, if this type of mechanism exists in *Xiphophorus*.

Studies also have shown that codon usage bias occurs in MHC genes in some mammals (McWeeney and Valdes 1999). These researchers argued that particular synonymous codons, usually those that flank PBR codons, could be "hitch-hiked" along with other polymorphic codons. The result of this would be a higher rate of synonymous nucleotide substitution within the PBR-encoding exon. Although in studies of microorganisms and *Drosophila* (Ikemura 1992; Powell and Moriyama 1997) it was shown that codon bias was due to translational selection, in the report by McWeeney and Valdes (1999) the codon bias was determined to be due to the high G+C content of the exons containing the PBR codons of humans and chimpanzees. This phenomenon has yet to be addressed in *Xiphophorus*.

Finally, a possible complication in the observation of higher d_S values in the PBR than elsewhere in the gene is the fact that these values were obtained when using the entirety of exon 2 as a dataset. Typical studies (e.g., Hughes 1999) calculating d_S values have only used the PBR-encoding codons. However, the specific codons responsible for encoding the PBR in teleosts have not yet been described experimentally, thus the entire exon 2 sequence was used, since it is known that this is the gene region responsible for encoding the MHC class II *DAB* PBR.

***DAB* Loci**

Finally, the question must be addressed as to how many *DAB*-encoding loci are present in *Xiphophorus*. These data indicate the presence of at least two loci. There are two possible explanations for the data derived from these *Xiphophorus* sequences. One, the polymerase enzyme (Platinum Taq HF) or the reverse-transcriptase (SuperScript II) produced a few random (i.e., artificial) nucleotide substitutions in four of the eight *Xipy* sequences in the major clade in Figure 6, and these sequences should be considered as identical to the other sequences in the clade. This would mean that there were only three different sequences identified from this fish sample. Since members of *Xiphophorus* are diploid, that would mean that there were at least two distinct loci, each encoding two possible sequences; we only identified three clearly unique sequences in *Xipy*14. The second possibility is that some or all of the observed nucleotide changes in the *Xipy* sequences from Figure 6 are real. If so, this would mean that there were actually up to seven different sequences identified from this fish. Again, since these fish are diploid, seven different sequences would indicate the presence of at least four distinct loci.

Another alternative explanation involves the enzyme that our collaborators used to generate the cDNA from these fish samples. It is possible that during RT-PCR, random mutations could have occurred, which would lead to the few substitutions observed in the ten *Xipy*14 sequences. If we assume that this is the case, then we can conservatively interpret the data as representing three different sequences.

One way to independently test the number of possible *DAB*-encoding loci would be to perform a Southern blot using a probe specific for the exon 3 (β_2 domain) region of

the gene. A Southern blot was performed by McConnell and coworkers (1998b) in their study of two MHC class II *B* loci, *DXB* and *DAB* in other *Xiphophorus* fishes. However, the probe used for *DAB* was a full-length fragment, and the number of loci was unclear. It would be prudent to reproduce this experiment. In fact, a new probe is currently being designed, and the Southern blot will be repeated in anticipation of publishing this research.

The full genomic sequence of the *Xiphophorus* MHC region has not been described, so it is still unclear how many possible loci are involved in encoding *DAB*. However, *DAB* has been mapped in *Xiphophorus* (McConnell *et al.* 1998b) to a newly assigned linkage group LG U24. Therefore, if there are multiple *DAB* loci they must be tightly linked within this group. As properties of the MHC have been found to be similar from ectothermic vertebrates to mammals, it is reasonable to assume that there will be more than one *DAB* locus in *Xiphophorus*, just as there are multiple loci for certain MHC genes in humans and mice.

Conclusions

A thorough population study would necessitate multiple gene markers, including mitochondrial DNA (mtDNA); however, this project focused on the evolution of the MHC within these populations of *Xiphophorus*. These studies have shown that the MHC class II *DAB* locus is polymorphic, and that positive selection is driving this polymorphism, as evidenced by the higher rates of synonymous and nonsynonymous nucleotide substitutions within the PBR and by the presence of potential allelic lineages for these *DAB* sequences. It is still unclear, however, whether positive selection is the effect of parasite-host co-evolution, or mate preference or both. We have further shown that the rates of synonymous nucleotide substitutions within the PBR are higher than elsewhere in the gene and have provided a possible mechanism for such observations. Finally, by comparing the sequences derived from ten independent PCR reactions of one fish, we have clearly demonstrated the presence of at least two *DAB* loci in *Xiphophorus pygmaeus*.

Table 1: A list of swordtail species, specific individual fish, and PCR+1 clones used in this study. Also included are the collection sites with Global Positioning System (GPS) coordinates.

Swordtail Species	Individuals	PCR+1 Clones	Collection Sites
Xiphophorus pygmaeus	*Xipy*1	*Xipy*1A,B	Huichihuayán (21°28'48.1"N 98°58'0"W)
	*Xipy*2	*Xipy*2F,J	
	*Xipy*3	*Xipy*3F,G	
	*Xipy*4	*Xipy*4A,E	
	*Xipy*5	*Xipy*5A,G	
	*Xipy*6	*Xipy*6A,B	
	*Xipy*7	*Xipy*7B,D	La Y-griega Vieja
	*Xipy*8	*Xipy*8A,B	(21°27'8.8"N 98°56'18.8"W)
	*Xipy*9	*Xipy*9A,B	
	*Xipy*10	*Xipy*10B,H	
	*Xipy*11	*Xipy*11B,F	
	*Xipy*12	*Xipy*12G,H	
	*Xipy*13	*Xipy*13D,E	
	*Xipy*14.1	*Xipy*14.1	
	*Xipy*14.2	*Xipy*14.2	
	*Xipy*14.3	*Xipy*14.3	
	*Xipy*14.4	*Xipy*14.4	
	*Xipy*14.5	*Xipy*14.5	
	*Xipy*14.6	*Xipy*14.6	
	*Xipy*14.7	*Xipy*14.7	
	*Xipy*14.8	*Xipy*14.8	
	*Xipy*14.9	*Xipy*14.9	
	*Xipy*14.10	*Xipy*14.10	
Xiphophorus multilineatus	*Ximu*15	*Ximu*15G,H	Rio Coy (21°45'0"N 98°57'25"W)
	*Ximu*16	*Ximu*16F,G	
	*Ximu*17	*Ximu*17F,G	
	*Ximu*18	*Ximu*18F,G	
	*Ximu*19	*Ximu*19B,G	
	*Ximu*20	*Ximu*20G,J	
	*Ximu*21	*Ximu*21B,D	
	*Ximu*22	*Ximu*22E,H	
	*Ximu*23	*Ximu*23A,F	
	*Ximu*24	*Ximu*24D,E	
	*Ximu*25	*Ximu*25A,G	
	*Ximu*26	*Ximu*26D,E	
	*Ximu*27	*Ximu*27A,B	Arroyo Tambaque (21°41'6"N 99°2'30"W)
	*Ximu*28	*Ximu*28B,E	
	*Ximu*29	*Ximu*29A,E	
	*Ximu*30	*Ximu*30B,D	
	*Ximu*31	*Ximu*31D,E	
	*Ximu*32	*Ximu*32D,E	
	*Ximu*33	*Ximu*33D,F	
	*Ximu*34	*Ximu*34D,E	
	*Ximu*35	*Ximu*35D,E	
	*Ximu*36	*Ximu*36K,L	
	*Ximu*37	*Ximu*37D,E	
	*Ximu*38	*Ximu*38D,F	
	*Ximu*39	*Ximu*39B,D	

Table 2: Oligonucleotide primers used in this study for PCR amplification or DNA sequencing and their locations. UTR-untranslated region, TM-transmembrane, CT-cytoplasmic tail, S-sense, A-antisense. Underlined portion of TM 399 indicates the *Hin*d III restriction site.

Primer Name	Sequence	Location
TM 396a	5′GCTGGGCTGGCTGCTGGTCAT 3′	Leader, codons 1-7, S
TM 398a	5′GAAGCAGGAGGAACCAGAACC 3′	3′UTR, 30bp beyond stop, A
TM 399	5′AGAAAGCTTGCTGGGCTGGCTGCTGGTCAT 3′	Leader, codons 3-7, S
M13/pUC F	5′CCCAGTCACGACGTTGTAAAACG 3′	pGEM-T Easy vector, S
M13/pUC R	5′AGCGGATAACAATTTCACACAGG 3′	pGEM-T Easy vector, A
TM 392	5′CCTCCATGCTGGTCTGCAGAGTT 3′	β_2, codons 122-129, S
TM 397	5′TCACTTCCTGTCCGTCTCTCA 3′	β_2, codons 142-148, A
TM 506	5′GGCTCAGCTGGTCGGAAC 3′	CT, codons 240-245, A
TM 507	5′GGGGCGTAGTGGAGAACAGTC 3′	β_2, codons 110-116, A
TM 518	5′ACATTTACAACAAGAAGGAGT 3′	β_1, codons 40-46, S
TM 525	5′AGCCAGAGACAAAATCAGACC 3′	TM, codons 220-226, A

Table 3: Average Jukes-Cantor corrected distance (d_{JC}) values for sequences, as well as the mean number (#) of nucleotide substitutions observed. Values in parentheses are the average number of nucleotide substitutions observed within exon 2 (β_1 domain). Values are given as one dataset as well as broken down into species and populations within each species. The number (n) of sequences in each dataset are also given.

Sample	n	d_{JC}	#
All 86 sequences	86	0.067±0.010	194 (130)
Xiphophorus pygmaeus	28	0.063±0.006	180 (144)
Huichihuayán population	12	0.048±0.005	101 (88)
La Y-griego Vieja population	16	0.061±0.006	156 (126)
Xiphophorus multilineatus	50	0.066±0.006	204 (159)
Rio Coy population	24	0.060±0.006	144 (117)
Arroyo Tambaque population	26	0.063±0.006	178 (149)
*Xipy*14 sequences	10	0.028±0.003	79 (67)

Table 4: Average d_S vs. d_N values for sequences. Values are given as one dataset as well as broken down into species and populations within each species. d_S and d_N values are given for all regions of the sequence except Exon 2 (All but Exon 2) and Exon 2 alone (Exon 2).

	All but Exon 2		Exon 2	
Sample	**d_S**	**d_N**	**d_S**	**d_N**
All 86 sequences	0.021±0.007	0.009±0.003	0.147±0.028	0.184±0.027
Xiphophorus pygmaeus	0.023±0.008	0.009±0.003	0.128±0.028	0.173±0.027
Huichihuayán population	0.011±0.004	0.005±0.002	0.125±0.033	0.132±0.021
La Y-griego Vieja population	0.022±0.008	0.008±0.003	0.120±0.026	0.171±0.027
Xiphophorus multilineatus	0.018±0.007	0.009±0.003	0.155±0.032	0.180±0.027
Rio Coy population	0.017±0.006	0.010±0.004	0.132±0.031	0.162±0.026
Arroyo Tambaque population	0.017±0.007	0.008±0.003	0.153±0.034	0.174±0.026
*Xipy*14 sequences	0.017±0.006	0.003±0.001	0.056±0.016	0.078±0.015

Figure 1: Schematic diagram of the MHC class II *DAB* locus. A. Genomic DNA, B. Processed mRNA, C. Encoded DAB protein on the cell surface. Abbreviations are: exon (E), intron (I), leader (L), connecting peptide (CP), transmembrane segment (TM), cytoplasmic tail (CT), and untranslated region (UTR).

A.


E1
I1
E2
I2
E3
I3
E4
I4
E5
5'
3'

B.


5'
UTR
L
Beta-1
Beta-2
CP TM CT
UTR
AAAAAAAA
3'

C.

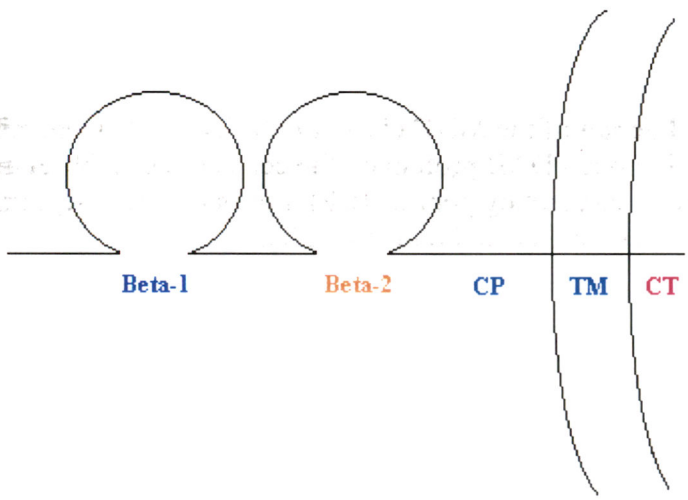
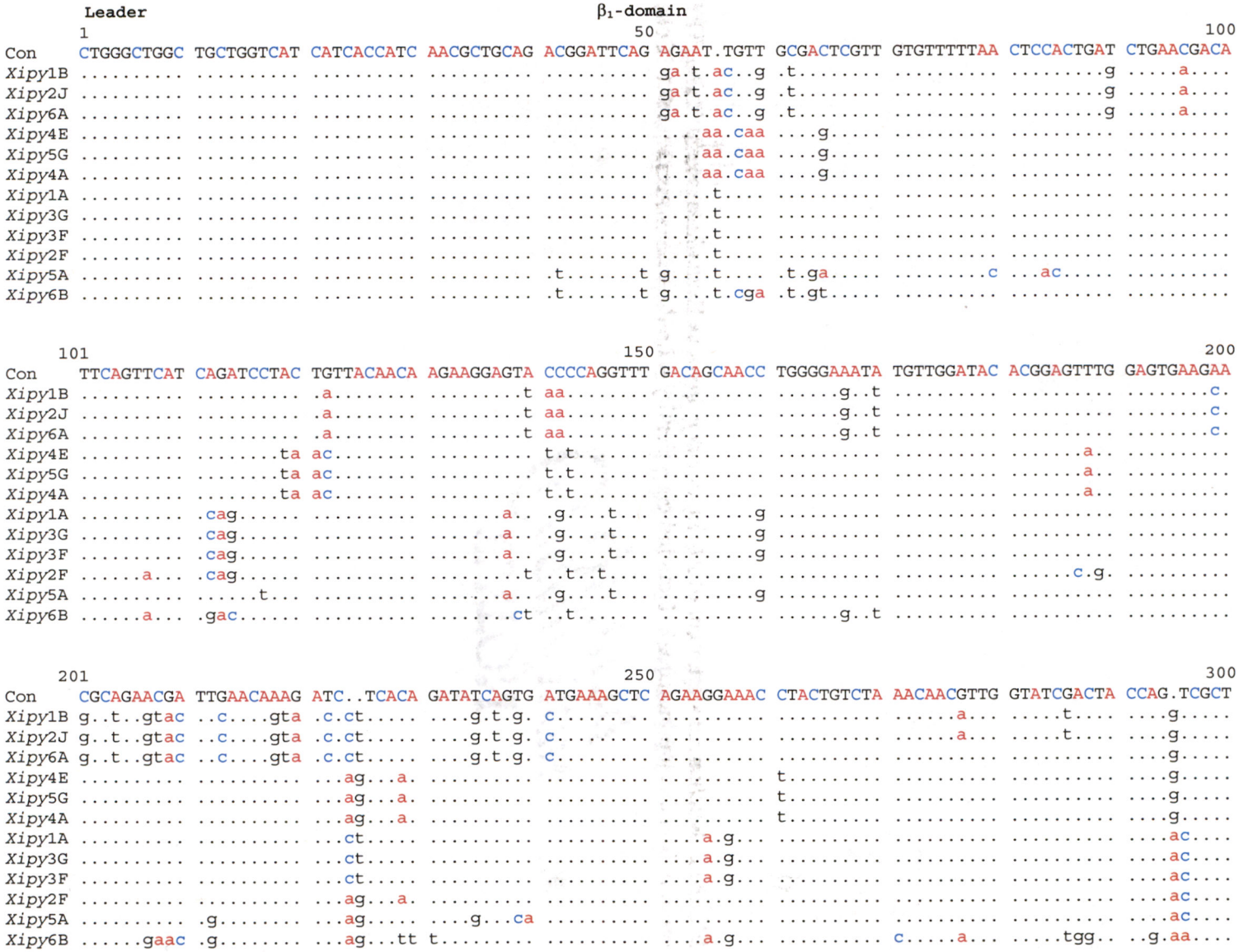
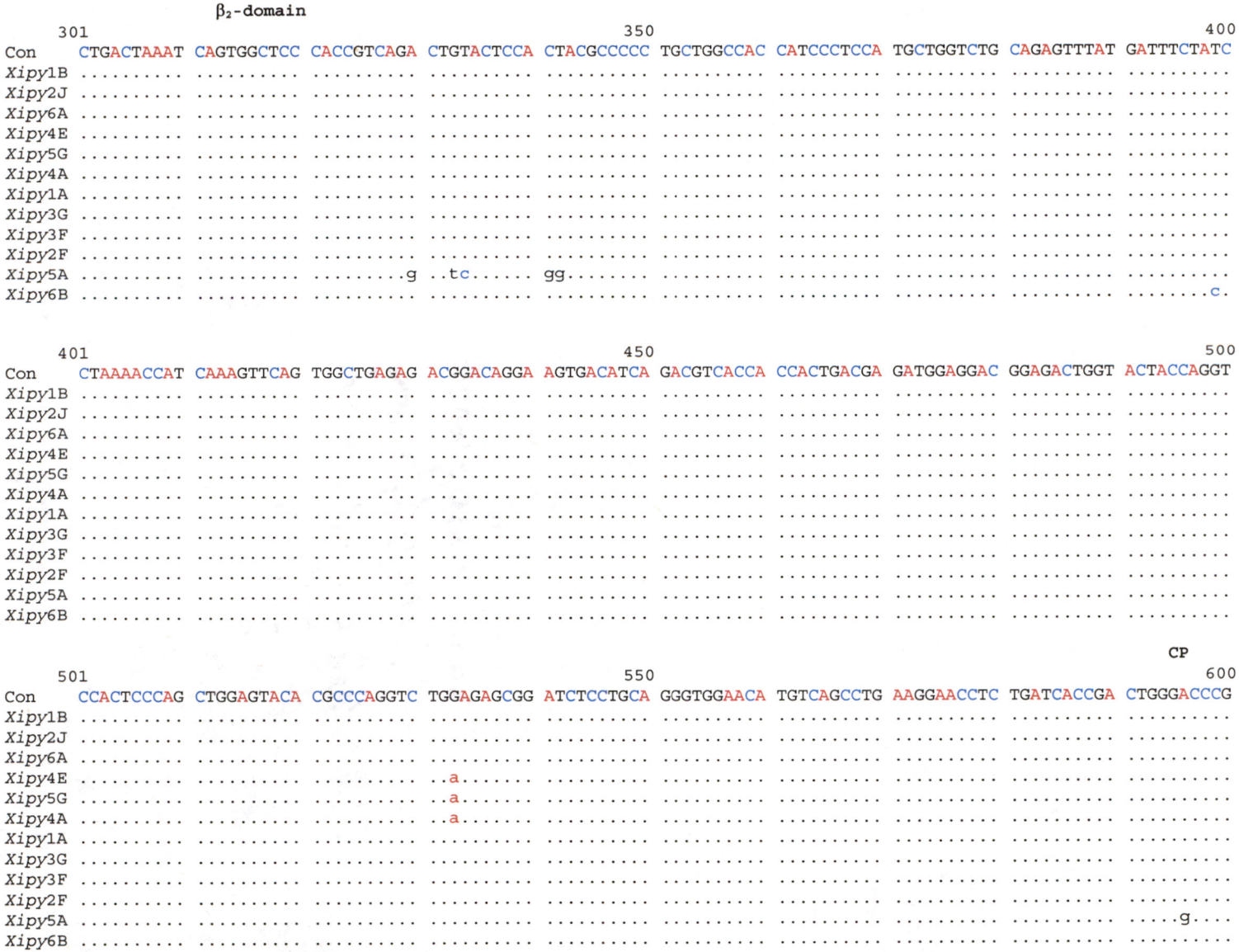

Beta-1
Beta-2
CP
TM
CT

Figure 2: Nucleotide alignment of *Xiphophorus pygmaeus* Huichihuyán population. The protein domains are indicated. Dots indicate identity with the consensus sequence (Con).


Leader
β1-domain
1
50
100
Con CTGGGCTGGC TGCTGGTCAT CATCACCATC AACGCTGCAG ACGGATTCAG AGAAT.TGTT GCGACTCGTT GTGTTTTTAA CTCCACTGAT CTGAACGACA
Xipy1B ga.t.ac..g .t..... ....g .....a....
Xipy2J ga.t.ac..g .t..... ....g .....a....
Xipy6A ga.t.ac..g .t..... ....g .....a....
Xipy4E ...aa.caa ....g..... ....
Xipy5G ...aa.caa ....g..... ....
Xipy4A ...aa.caa ....g..... ....
Xipy1A .....t.... ....
Xipy3G .....t.... ....
Xipy3F .....t.... ....
Xipy2F .....t.... ....
Xipy5A .t.....t g....t.... .t.ga..... .....c ...ac..... .....
Xipy6B .t.....t g....t.cga .t.gt..... ....
101
150
200
Con TTCAGTTCAT CAGATCCTAC TGTTACAACA AGAAGGAGTA CCCCAGGTTT GACAGCAACC TGGGGAAATA TGTTGGATAC ACGGAGTTTG GAGTGAAGAA
Xipy1B .....a..... .....t aa..... ....g..t ....c.
Xipy2J .....a..... .....t aa..... ....g..t ....c.
Xipy6A .....a..... .....t aa..... ....g..t ....c.
Xipy4E .....ta ac..... ..... t.t..... ....a.. .....
Xipy5G .....ta ac..... ..... t.t..... ....a.. .....
Xipy4A .....ta ac..... ..... t.t..... ....a.. .....
Xipy1A .cag..... ....a.. .g....t... .....g ....
Xipy3G .cag..... ....a.. .g....t... .....g ....
Xipy3F .cag..... ....a.. .g....t... .....g ....
Xipy2F .....a... .cag..... ....t ..t..t.... ....c.g. .....
Xipy5A .....t... ....a.. .g....t... .....g ....
Xipy6B .....a... .gac..... ....ct ..t..... ....g..t ....
201
250
300
Con CGCAGAACGA TTGAACAAAG ATC..TCACA GATATCAGTG ATGAAAGCTC AGAAGGAAAC CTACTGTCTA AACAACGTTG GTATCGACTA CCAG.TCGCT
Xipy1B g..t..gtac ..c....gta .c.ct..... ....g.t.g. c..... ....a... .....t.... ....g.....
Xipy2J g..t..gtac ..c....gta .c.ct..... ....g.t.g. c..... ....a... .....t.... ....g.....
Xipy6A g..t..gtac ..c....gta .c.ct..... ....g.t.g. c..... ....g.....
Xipy4E ....ag...a. .... t..... ....g.....
Xipy5G ....ag...a. .... t..... ....g.....
Xipy4A ....ag...a. .... t..... ....g.....
Xipy1A ....ct..... ....a.g... ....ac....
Xipy3G ....ct..... ....a.g... ....ac....
Xipy3F ....ct..... ....a.g... ....ac....
Xipy2F ....ag...a. ....ac....
Xipy5A ..... .g..... ...ag..... ....g...ca ....ac....
Xipy6B .....gaac .g..... ...ag...tt t..... ....a.g... ..... c.....a... .....tgg.. ..g.aa....


β2-domain
301
350
400
Con CTGACTAAAT CAGTGGCTCC CACCGTCAGA CTGTACTCCA CTACGCCCCC TGCTGGCCAC CATCCCTCCA TGCTGGTCTG CAGAGTTTAT GATTTCTATC
Xipy1B
Xipy2J
Xipy6A
Xipy4E
Xipy5G
Xipy4A
Xipy1A
Xipy3G
Xipy3F
Xipy2F
Xipy5A .....g ..tc..... gg.....
Xipy6B .....c.
401
450
500
Con CTAAAACCAT CAAAGTTCAG TGGCTGAGAG ACGGACAGGA AGTGACATCA GACGTCACCA CCACTGACGA GATGGAGGAC GGAGACTGGT ACTACCAGGT
Xipy1B
Xipy2J
Xipy6A
Xipy4E
Xipy5G
Xipy4A
Xipy1A
Xipy3G
Xipy3F
Xipy2F
Xipy5A
Xipy6B
CP
501
550
600
Con CCACTCCCAG CTGGAGTACA CGCCCAGGTC TGGAGAGCGG ATCTCCTGCA GGGTGGAACA TGTCAGCCTG AAGGAACCTC TGATCACCGA CTGGGACCCG
Xipy1B
Xipy2J
Xipy6A
Xipy4E ..a.....
Xipy5G ..a.....
Xipy4A ..a.....
Xipy1A
Xipy3G
Xipy3F
Xipy2F
Xipy5A .....g....
Xipy6B

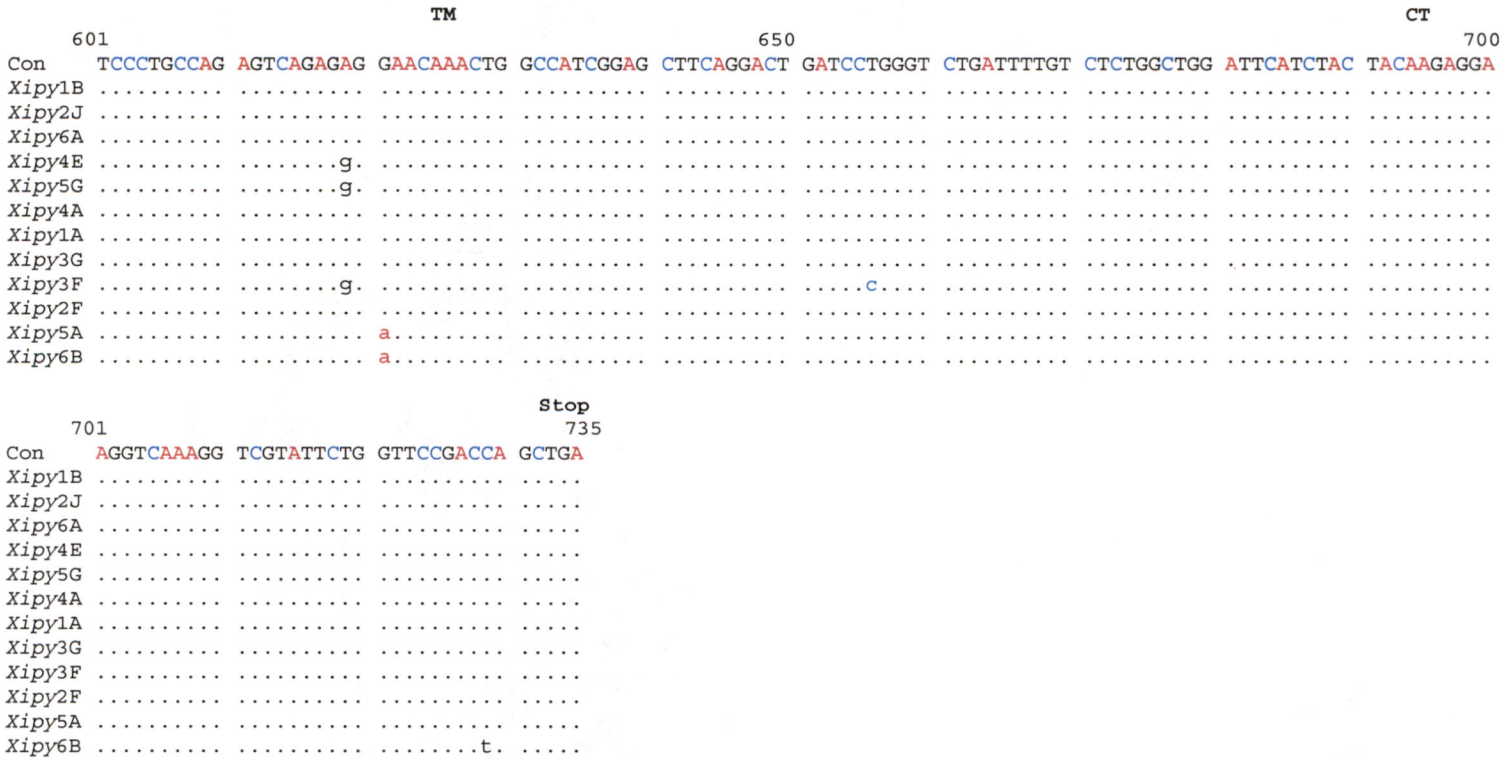
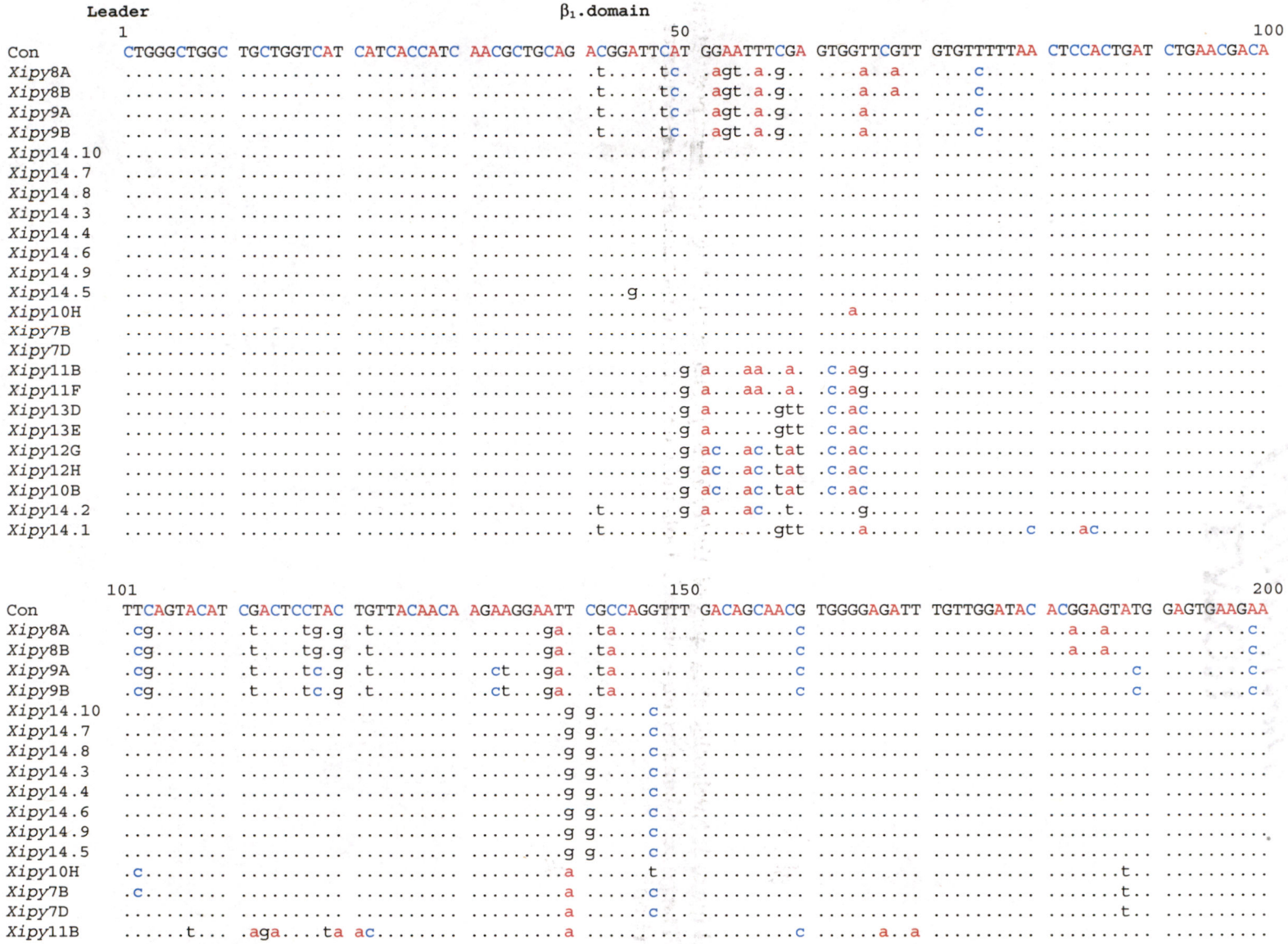
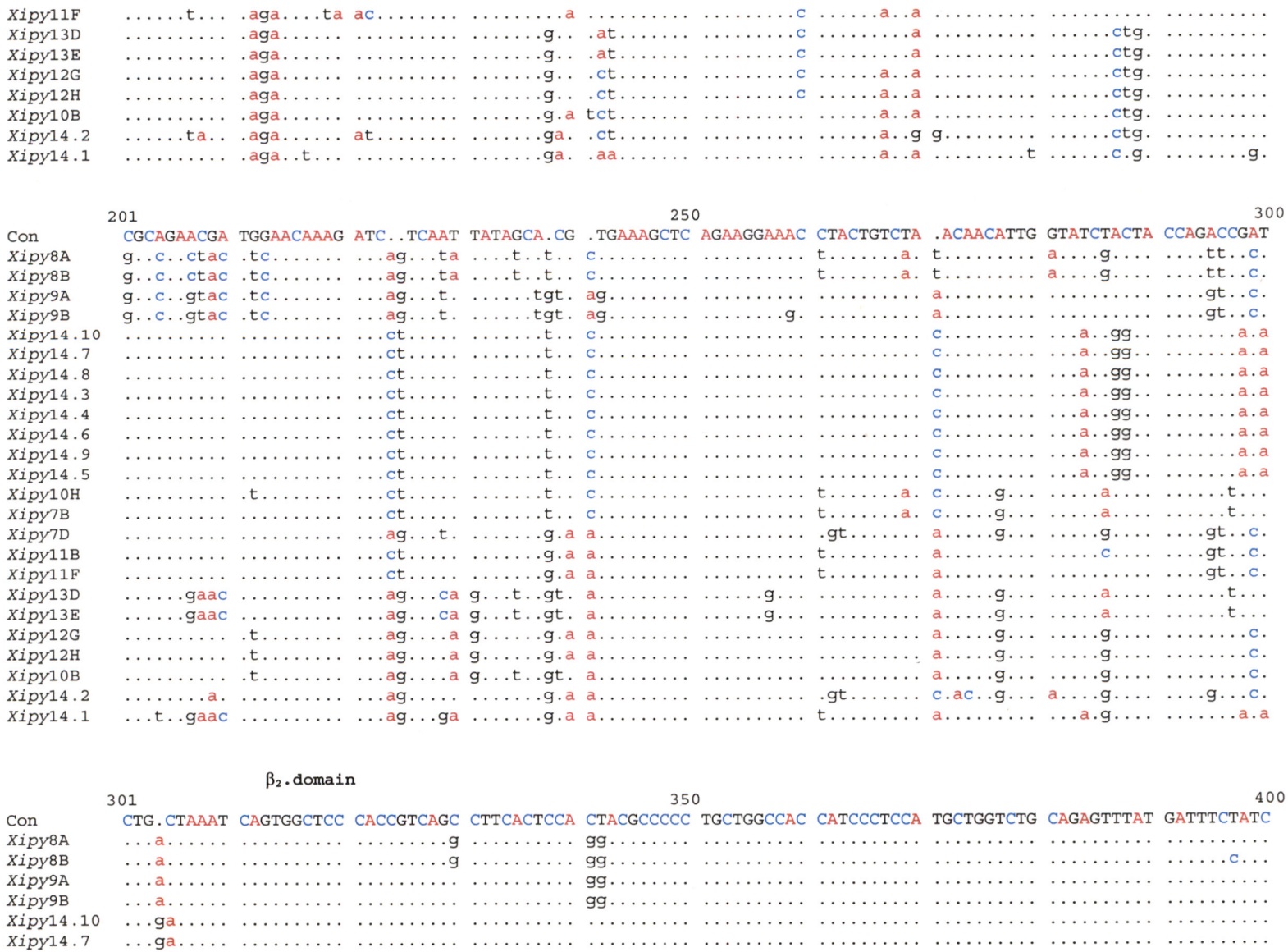
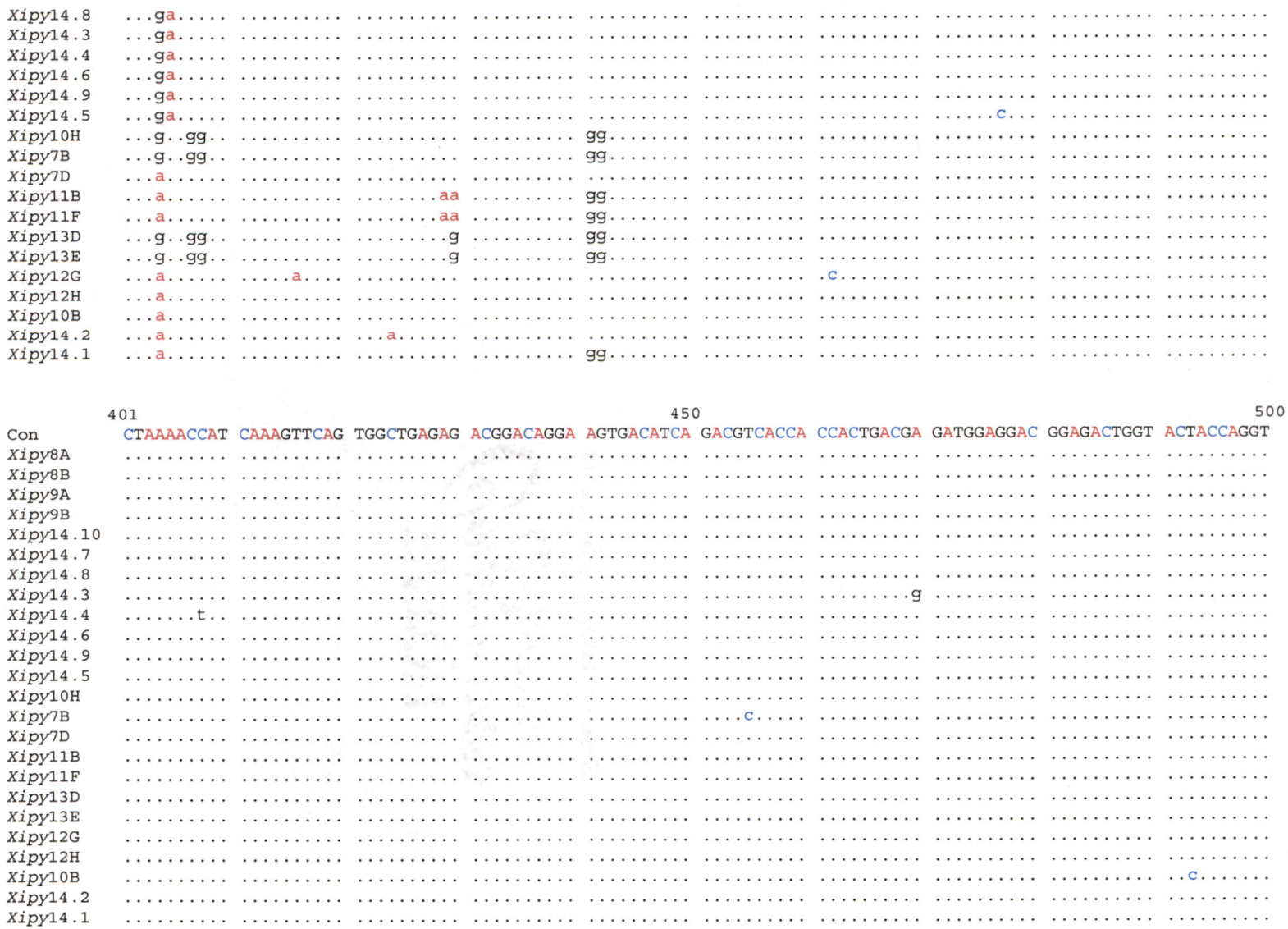
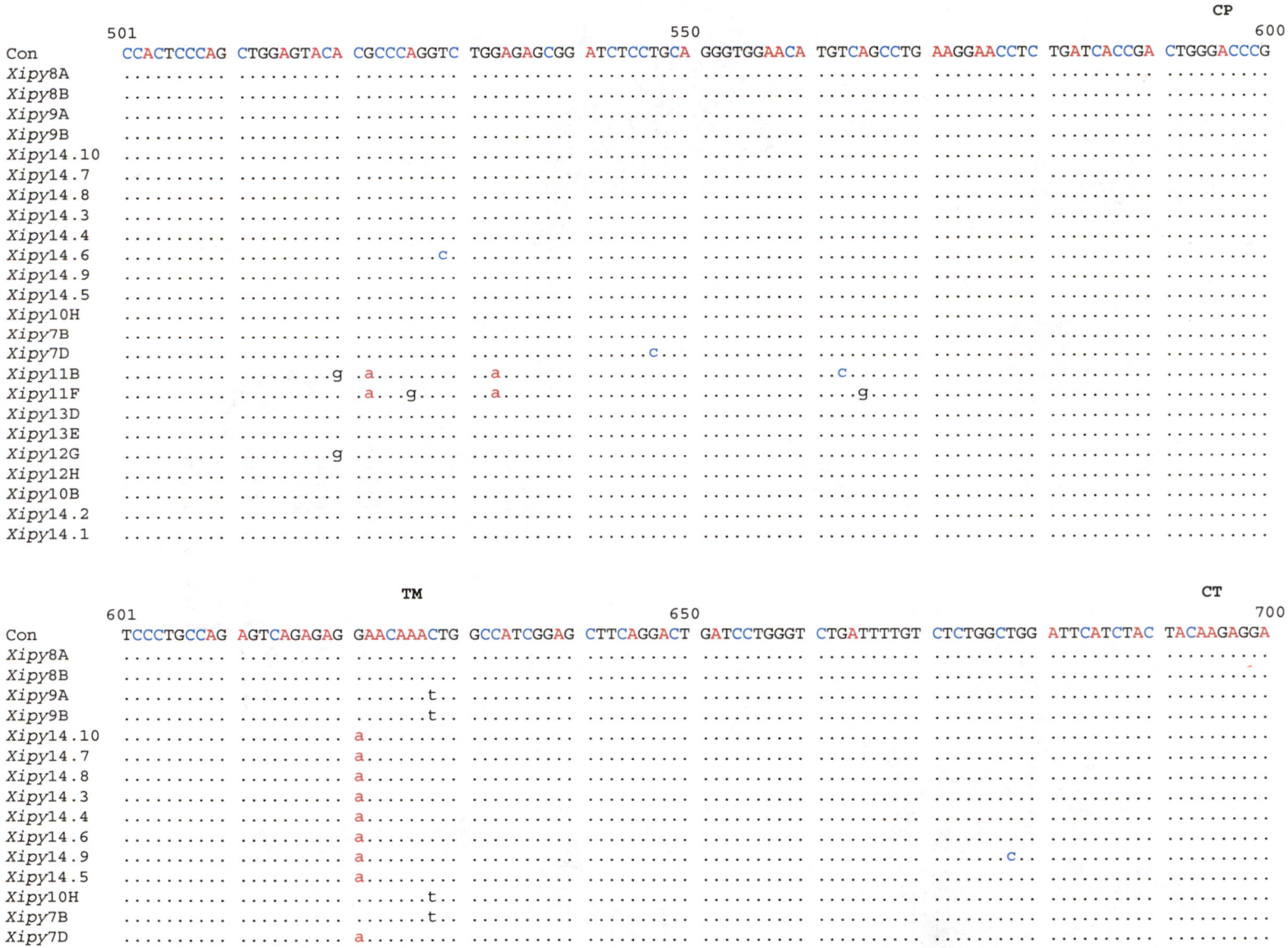

TM
CT
601
650
700
Con TCCCTGCCAG AGTCAGAGAG GAACAAACTG GCCATCGGAG CTTCAGGACT GATCCTGGGT CTGATTTTGT CTCTGGCTGG ATTCATCTAC TACAAGAGGA
Xipy1B
Xipy2J
Xipy6A
Xipy4E .....g.
Xipy5G .....g.
Xipy4A
Xipy1A
Xipy3G
Xipy3F .....g. .....c....
Xipy2F
Xipy5A a.....
Xipy6B a.....
Stop
701
735
Con AGGTCAAAGG TCGTATTCTG GTTCCGACCA GCTGA
Xipy1B
Xipy2J
Xipy6A
Xipy4E
Xipy5G
Xipy4A
Xipy1A
Xipy3G
Xipy3F
Xipy2F
Xipy5A
Xipy6B .....t.

Figure 3: Nucleotide alignment of *Xiphophorus pygmaeus* La Y-griega Vieja population. The protein domains are indicated. Dots indicate identity with the consensus sequence (Con).


Leader
β1.domain
1
50
100
Con CTGGGCTGGC TGCTGGTCAT CATCACCATC AACGCTGCAG ACGGATTCAT GGAATTTCGA GTGGTTCGTT GTGTTTTTAA CTCCACTGAT CTGAACGACA
Xipy8A .t.....tc. .agt.a.g.. ....a..a.. ....c.....
Xipy8B .t.....tc. .agt.a.g.. ....a..a.. ....c.....
Xipy9A .t.....tc. .agt.a.g.. ....a..... ....c.....
Xipy9B .t.....tc. .agt.a.g.. ....a..... ....c.....
Xipy14.10
Xipy14.7
Xipy14.8
Xipy14.3
Xipy14.4
Xipy14.6
Xipy14.9
Xipy14.5 ....g.....
Xipy10H ...a.....
Xipy7B
Xipy7D
Xipy11B .....g a...aa..a. .c.ag.....
Xipy11F .....g a...aa..a. .c.ag.....
Xipy13D .....g a......gtt .c.ac.....
Xipy13E .....g a......gtt .c.ac.....
Xipy12G .....g ac..ac.tat .c.ac.....
Xipy12H .....g ac..ac.tat .c.ac.....
Xipy10B .....g ac..ac.tat .c.ac.....
Xipy14.2 .t.....g a...ac..t. ....g.....
Xipy14.1 .t..... .....gtt ....a..... .....c ...ac.....
101
150
200
Con TTCAGTACAT CGACTCCTAC TGTTACAACA AGAAGGAATT CGCCAGGTTT GACAGCAACG TGGGGAGATT TGTTGGATAC ACGGAGTATG GAGTGAAGAA
Xipy8A .cg..... .t....tg.g .t..... .....ga. .ta.....c ....a..a.... .....c.
Xipy8B .cg..... .t....tg.g .t..... .....ga. .ta.....c ....a..a.... .....c.
Xipy9A .cg..... .t....tc.g .t..... ..ct...ga. .ta.....c ....c. .....c.
Xipy9B .cg..... .t....tc.g .t..... ..ct...ga. .ta.....c ....c. .....c.
Xipy14.10 .....g g.....c... .....
Xipy14.7 .....g g.....c... .....
Xipy14.8 .....g g.....c... .....
Xipy14.3 .....g g.....c... .....
Xipy14.4 .....g g.....c... .....
Xipy14.6 .....g g.....c... .....
Xipy14.9 .....g g.....c... .....
Xipy14.5 .....g g.....c... .....
Xipy10H .c..... .....a ......t... ....t..
Xipy7B .c..... .....a ......c... ....t..
Xipy7D ....a ......c... ....t..
Xipy11B ......t... .aga....ta ac..... .....a .....c ......a..a ....


Xipy11F ......t... .aga....ta ac..... .....a ....c .....a..a ....
Xipy13D ..... .aga..... ....g.. .at..... .....c .....a ....ctg. .....
Xipy13E ..... .aga..... ....g.. .at..... .....c .....a ....ctg. .....
Xipy12G ..... .aga..... ....g.. .ct..... .....c .....a..a ....ctg. .....
Xipy12H ..... .aga..... ....g.. .ct..... .....c .....a..a ....ctg. .....
Xipy10B ..... .aga..... ....g.a tct..... ....a..a ....ctg. .....
Xipy14.2 .....ta.. .aga..... at..... .....ga. .ct..... ....a..g g..... .....ctg. .....
Xipy14.1 ..... .aga..t... ....ga. .aa..... ....a..a .....t .....c.g. .....g.
201
250
300
Con CGCAGAACGA TGGAACAAAG ATC..TCAAT TATAGCA.CG .TGAAAGCTC AGAAGGAAAC CTACTGTCTA .ACAACATTG GTATCTACTA CCAGACCGAT
Xipy8A g..c..ctac .tc..... ...ag...ta ....t..t.. c..... ..... t.....a. t..... a....g.... ....tt..c.
Xipy8B g..c..ctac .tc..... ...ag...ta ....t..t.. c..... ..... t.....a. t..... a....g.... ....tt..c.
Xipy9A g..c..gtac .tc..... ...ag...t. .....tgt. ag..... .... a..... ....gt..c.
Xipy9B g..c..gtac .tc..... ...ag...t. .....tgt. ag..... .....g. ..... a..... ....gt..c.
Xipy14.10 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.7 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.8 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.3 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.4 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.6 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.9 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy14.5 ....ct..... .....t.. c..... .... c..... ...a..gg.. .....a.a
Xipy10H ..... .t..... ...ct..... .....t.. c..... ..... t.....a. c.....g... .....a.... .....t...
Xipy7B ....ct..... .....t.. c..... ..... t.....a. c.....g... .....a.... .....t...
Xipy7D ....ag...t. .....g.a a..... ..... .gt..... a.....g... .....g.... ....gt..c.
Xipy11B ....ct..... .....g.a a..... ..... t..... a..... .....c.... ....gt..c.
Xipy11F ....ct..... .....g.a a..... ..... t..... a..... ....gt..c.
Xipy13D ......gaac ....ag...ca g...t..gt. a..... .....g... ..... a.....g... .....a.... .....t...
Xipy13E ......gaac ....ag...ca g...t..gt. a..... .....g... ..... a.....g... .....a.... .....t...
Xipy12G ..... .t..... ...ag....a g.....g.a a..... .... a.....g... .....g.... .....c.
Xipy12H ..... .t..... ...ag....a g.....g.a a..... .... a.....g... .....g.... .....c.
Xipy10B ..... .t..... ...ag....a g...t..gt. a..... .... a.....g... .....g.... .....c.
Xipy14.2 .....a. ....ag..... .....g.a a..... ..... .gt..... c.ac..g... a....g.... ....g...c.
Xipy14.1 ...t..gaac ....ag...ga .....g.a a..... ..... t..... a..... ...a.g.... .....a.a
β2.domain
301
350
400
Con CTG.CTAAAT CAGTGGCTCC CACCGTCAGC CTTCACTCCA CTACGCCCCC TGCTGGCCAC CATCCCTCCA TGCTGGTCTG CAGAGTTTAT GATTTCTATC
Xipy8A ...a..... ....g ..... gg..... ....
Xipy8B ...a..... ....g ..... gg..... ....c...
Xipy9A ...a..... .... gg..... ....
Xipy9B ...a..... .... gg..... ....
Xipy14.10 ...ga..... ....
Xipy14.7 ...ga..... ....

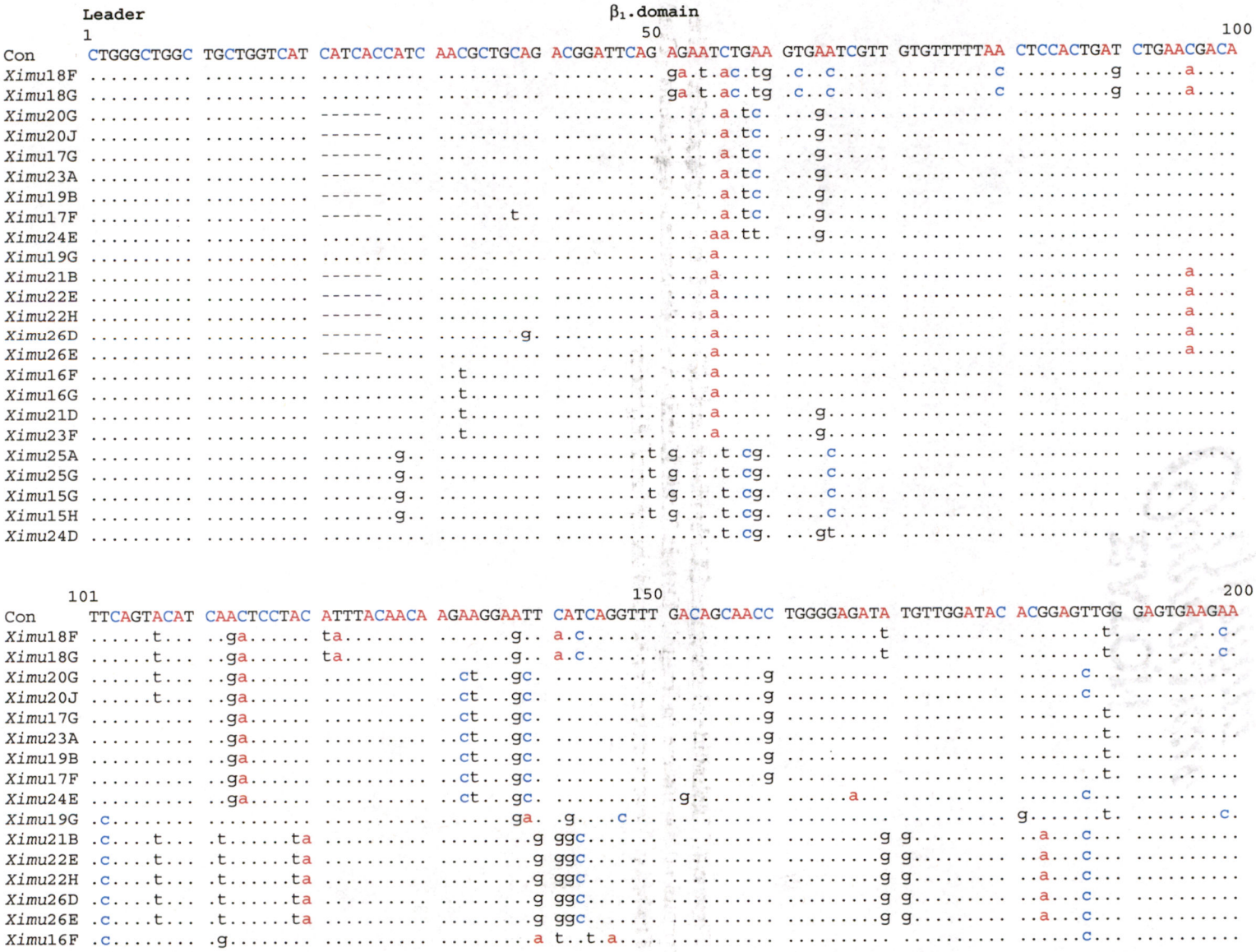

Xipy14.8 ...ga..... .....
Xipy14.3 ...ga..... .....
Xipy14.4 ...ga..... .....
Xipy14.6 ...ga..... .....
Xipy14.9 ...ga..... .....
Xipy14.5 ...ga..... .....
Xipy10H ...g..gg.. .....
Xipy7B ...g..gg.. .....
Xipy7D ...a..... .....
Xipy11B ...a..... .....
Xipy11F ...a..... .....
Xipy13D ...g..gg.. .....
Xipy13E ...g..gg.. .....
Xipy12G ...a..... .....
Xipy12H ...a..... .....
Xipy10B ...a..... .....
Xipy14.2 ...a..... .....
Xipy14.1 ...a..... .....
401
450
500
Con CTAAAACCAT CAAAGTTCAG TGGCTGAGAG ACGGACAGGA AGTGACATCA GACGTCACCA CCACTGACGA GATGGAGGAC GGAGACTGGT ACTACCAGGT
Xipy8A .....
Xipy8B .....
Xipy9A .....
Xipy9B .....
Xipy14.10 .....
Xipy14.7 .....
Xipy14.8 .....
Xipy14.3 .....
Xipy14.4 .....
Xipy14.6 .....
Xipy14.9 .....
Xipy14.5 .....
Xipy10H .....
Xipy7B .....
Xipy7D .....
Xipy11B .....
Xipy11F .....
Xipy13D .....
Xipy13E .....
Xipy12G .....
Xipy12H .....
Xipy10B .....
Xipy14.2 .....
Xipy14.1 .....

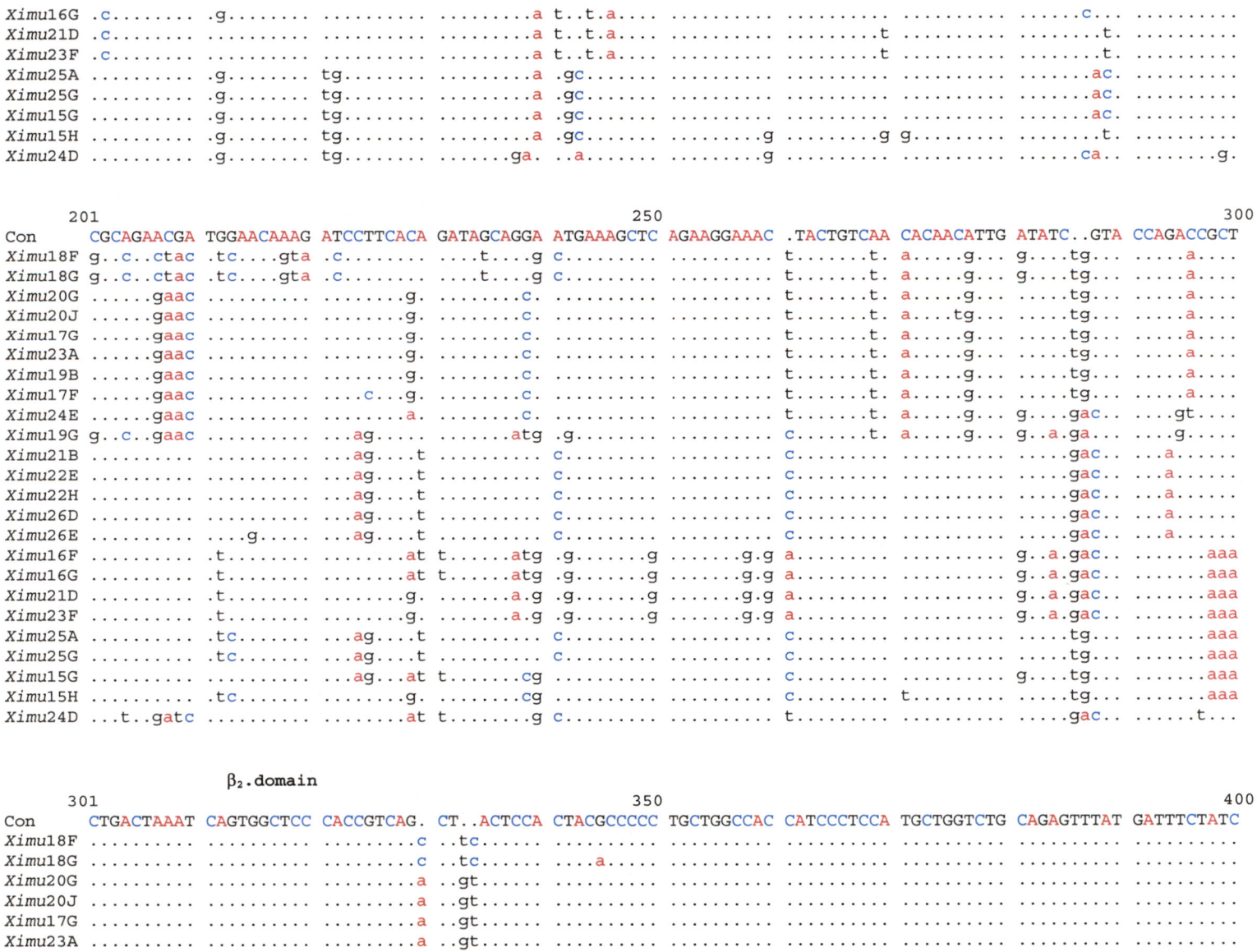

CP
501
550
600
Con CCACTCCCAG CTGGAGTACA CGCCCAGGTC TGGAGAGCGG ATCTCCTGCA GGGTGGAACA TGTCAGCCTG AAGGAACCTC TGATCACCGA CTGGGACCCG
Xipy8A
Xipy8B
Xipy9A
Xipy9B
Xipy14.10
Xipy14.7
Xipy14.8
Xipy14.3
Xipy14.4
Xipy14.6
Xipy14.9
Xipy14.5
Xipy10H
Xipy7B
Xipy7D
Xipy11B
Xipy11F
Xipy13D
Xipy13E
Xipy12G
Xipy12H
Xipy10B
Xipy14.2
Xipy14.1
TM
CT
601
650
700
Con TCCCTGCCAG AGTCAGAGAG GAACAAACTG GCCATCGGAG CTTCAGGACT GATCCTGGGT CTGATTTTGT CTCTGGCTGG ATTCATCTAC TACAAGAGGA
Xipy8A
Xipy8B
Xipy9A
Xipy9B
Xipy14.10
Xipy14.7
Xipy14.8
Xipy14.3
Xipy14.4
Xipy14.6
Xipy14.9
Xipy14.5
Xipy10H
Xipy7B
Xipy7D

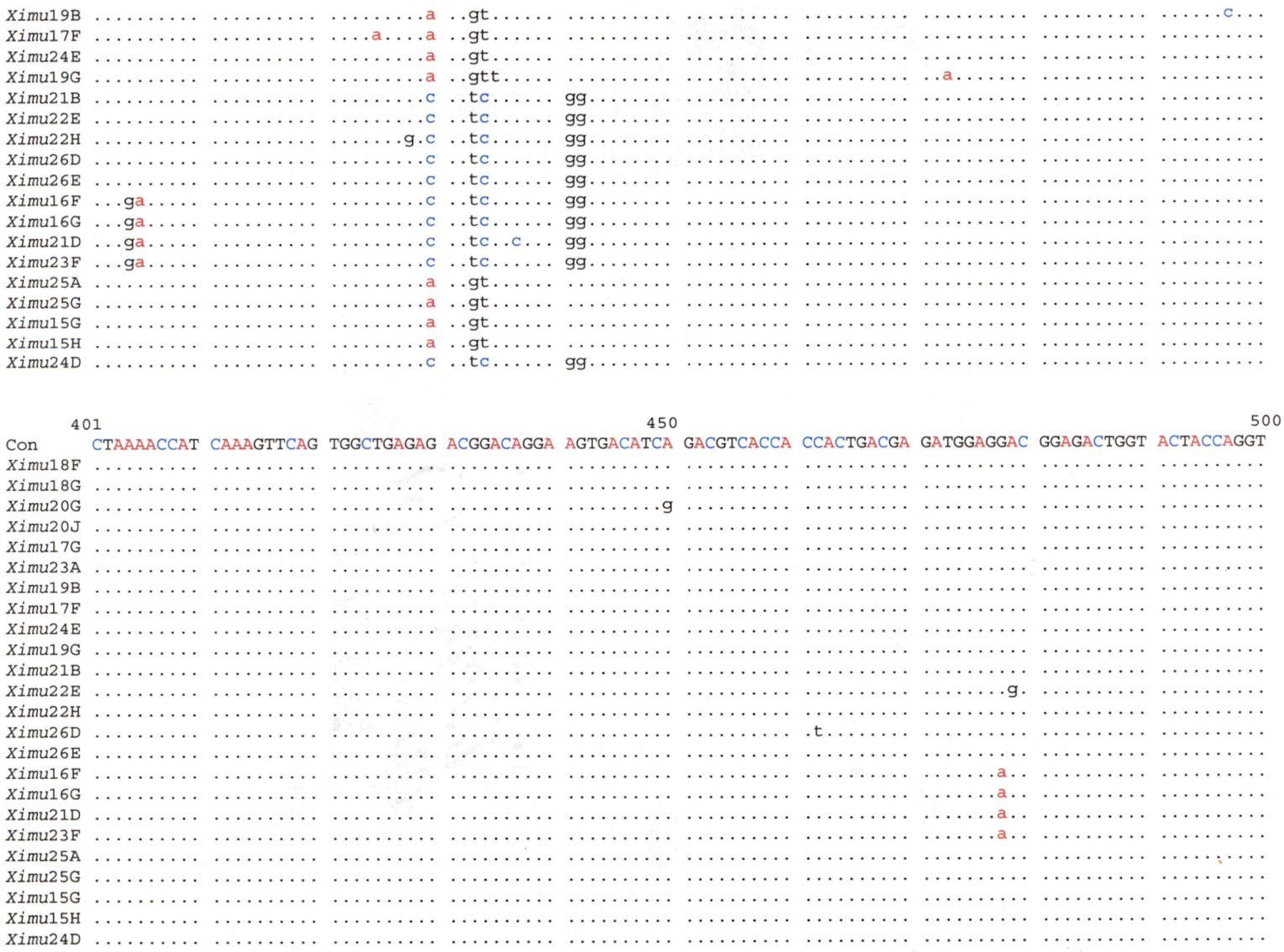
```
Xipy11B    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy11F    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy13D    .......... .......... .......... .......... .......... .......... c......... .......... .......... ..........
Xipy13E    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy12G    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy12H    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy10B    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy14.2   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Xipy14.1   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

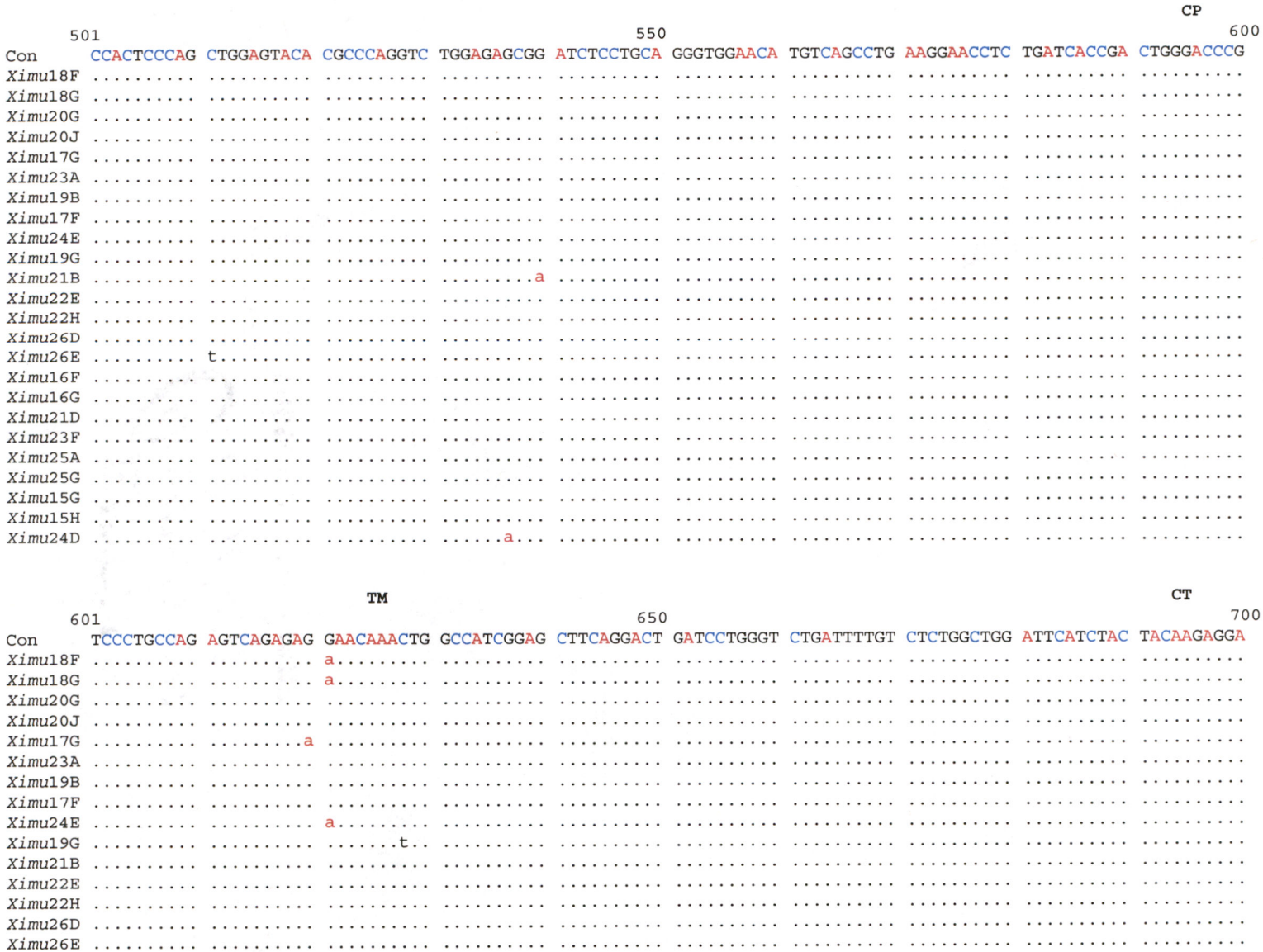
```
                                                 Stop
          701                                      735
Con        AGGTCAAAGG TCGTATTCTG GTTCCGACCA GCTGA
Xipy8A     .......... .......... .......... .....
Xipy8B     .......... .......... .......... .....
Xipy9A     .......... .......... .......... .....
Xipy9B     .......... .......... .......... .....
Xipy14.10  .......... .......... ........t. .....
Xipy14.7   .......... .......... ........t. .....
Xipy14.8   .......... .......... ........t. .....
Xipy14.3   .......... .......... ........t. .....
Xipy14.4   .......... .......... ........t. .....
Xipy14.6   .......... .......... ........t. .....
Xipy14.9   .......... .......... ........t. .....
Xipy14.5   .......... .......... ........t. .....
Xipy10H    .......... .......... .....a.... .....
Xipy7B     .......... .......... .....a.... .....
Xipy7D     .......... .......... ........t. .....
Xipy11B    .......... .......... .......... .....
Xipy11F    .......... .......... .......... .....
Xipy13D    .......... .......... .......... .....
Xipy13E    .......... .......... .......... .....
Xipy12G    .......... .......... .......... .....
Xipy12H    .......... .......... .......... .....
Xipy10B    .......... .......... .......... .....
Xipy14.2   .......... .......... .......... .....
Xipy14.1   .......... .......... .......... .....
```

Figure 4: Nucleotide alignment of *Xiphophorus multilineatus* Rio Coy population. The protein domains are indicated. Dots indicate identity with the consensus sequence (Con). Dashes represent indels introduced for optimal alignment.


Leader
β1.domain
1
50
100
Con CTGGGCTGGC TGCTGGTCAT CATCACCATC AACGCTGCAG ACGGATTCAG AGAATCTGAA GTGAATCGTT GTGTTTTTAA CTCCACTGAT CTGAACGACA
Ximu18F ga.t.ac.tg .c..c..... .....c .....g .....a....
Ximu18G ga.t.ac.tg .c..c..... .....c .....g .....a....
Ximu20G ------.... .....a.tc. ...g..... .....
Ximu20J ------.... .....a.tc. ...g..... .....
Ximu17G ------.... .....a.tc. ...g..... .....
Ximu23A ------.... .....a.tc. ...g..... .....
Ximu19B ------.... .....a.tc. ...g..... .....
Ximu17F ------.... .....t.. .....a.tc. ...g..... .....
Ximu24E .....aa.tt. ...g..... .....
Ximu19G .....a..... .....
Ximu21B ------.... .....a..... .....a....
Ximu22E ------.... .....a..... .....a....
Ximu22H ------.... .....a..... .....a....
Ximu26D ------.... .....g. .....a..... .....a....
Ximu26E ------.... .....a..... .....a....
Ximu16F .....t..... .....a.....
Ximu16G .....t..... .....a.....
Ximu21D .....t..... .....a..... ...g.....
Ximu23F .....t..... .....a..... ...g.....
Ximu25A .....g.. .....t g....t.cg. ....c.....
Ximu25G .....g.. .....t g....t.cg. ....c.....
Ximu15G .....g.. .....t g....t.cg. ....c.....
Ximu15H .....g.. .....t g....t.cg. ....c.....
Ximu24D .....t.cg. ...gt.....
101
150
200
Con TTCAGTACAT CAACTCCTAC ATTTACAACA AGAAGGAATT CATCAGGTTT GACAGCAACC TGGGGAGATA TGTTGGATAC ACGGAGTTGG GAGTGAAGAA
Ximu18F .....t... ..ga..... ta..... .....g.. a.c..... ....t ....t. .....c.
Ximu18G .....t... ..ga..... ta..... .....g.. a.c..... ....t ....t. .....c.
Ximu20G .....t... ..ga..... ....ct...gc. ....g ....c... .....
Ximu20J .....t... ..ga..... ....ct...gc. ....g ....c... .....
Ximu17G ....ga..... ....ct...gc. ....g ....t. .....
Ximu23A ....ga..... ....ct...gc. ....g ....t. .....
Ximu19B ....ga..... ....ct...gc. ....g ....t. .....
Ximu17F ....ga..... ....ct...gc. ....g ....t. .....
Ximu24E ....ga..... ....ct...gc. ..... .g..... .....a... ....c... .....
Ximu19G .c..... ....ga. .g....c... .... g.....t. .....c.
Ximu21B .c....t... .t.....ta ....g ggc..... ....g g..... ..a...c... .....
Ximu22E .c....t... .t.....ta ....g ggc..... ....g g..... ..a...c... .....
Ximu22H .c....t... .t.....ta ....g ggc..... ....g g..... ..a...c... .....
Ximu26D .c....t... .t.....ta ....g ggc..... ....g g..... ..a...c... .....
Ximu26E .c....t... .t.....ta ....g ggc..... ....g g..... ..a...c... .....
Ximu16F .c..... .g..... ....a t..t.a.... ....c... .....


Ximu16G .c..... .g..... ....a t..t.a.... ....c... .....
Ximu21D .c..... ....a t..t.a.... ....t ....t. .....
Ximu23F .c..... ....a t..t.a.... ....t ....t. .....
Ximu25A ..... .g..... tg..... .....a .gc..... ....ac. .....
Ximu25G ..... .g..... tg..... .....a .gc..... ....ac. .....
Ximu15G ..... .g..... tg..... .....a .gc..... ....ac. .....
Ximu15H ..... .g..... tg..... .....a .gc..... .....g .....g g..... .....t. .....
Ximu24D ..... .g..... tg..... .....ga. ..a..... .....g ....ca.. .....g.
201
250
300
Con CGCAGAACGA TGGAACAAAG ATCCTTCACA GATAGCAGGA ATGAAAGCTC AGAAGGAAAC .TACTGTCAA CACAACATTG ATATC..GTA CCAGACCGCT
Ximu18F g..c..ctac .tc....gta .c..... ....t....g c..... ..... t.....t. a.....g... g....tg... .....a....
Ximu18G g..c..ctac .tc....gta .c..... ....t....g c..... ..... t.....t. a.....g... g....tg... .....a....
Ximu20G ......gaac ....g. .....c. .... t.....t. a.....g... .....tg... .....a....
Ximu20J ......gaac ....g. .....c. .... t.....t. a....tg... .....tg... .....a....
Ximu17G ......gaac ....g. .....c. .... t.....t. a.....g... .....tg... .....a....
Ximu23A ......gaac ....g. .....c. .... t.....t. a.....g... .....tg... .....a....
Ximu19B ......gaac ....g. .....c. .... t.....t. a.....g... .....tg... .....a....
Ximu17F ......gaac ....c...g. .....c. .... t.....t. a.....g... .....tg... .....a....
Ximu24E ......gaac ....a. .....c. .... t.....t. a.....g... g....gac.. ....gt....
Ximu19G g..c..gaac ....ag..... .....atg .g..... ..... c.....t. a.....g... g..a.ga... ....g.....
Ximu21B ....ag....t ..... c..... ..... c..... ....gac.. ...a.....
Ximu22E ....ag....t ..... c..... ..... c..... ....gac.. ...a.....
Ximu22H ....ag....t ..... c..... ..... c..... ....gac.. ...a.....
Ximu26D ....ag....t ..... c..... ..... c..... ....gac.. ...a.....
Ximu26E ....g..... ...ag....t ..... c..... ..... c..... ....gac.. ...a.....
Ximu16F ..... .t..... .....at t......atg .g.....g .....g.g a..... ..... g..a.gac.. .....aaa
Ximu16G ..... .t..... .....at t......atg .g.....g .....g.g a..... ..... g..a.gac.. .....aaa
Ximu21D ..... .t..... .....g. .....a.g .g.....g .....g.g a..... ..... g..a.gac.. .....aaa
Ximu23F ..... .t..... .....g. .....a.g .g.....g .....g.g a..... ..... g..a.gac.. .....aaa
Ximu25A ..... .tc..... ...ag....t ..... c..... ..... c..... ....tg... .....aaa
Ximu25G ..... .tc..... ...ag....t ..... c..... ..... c..... ....tg... .....aaa
Ximu15G ....ag...at t.....cg .... c..... ..... g....tg... .....aaa
Ximu15H ..... .tc..... .....g. .....cg .... c..... t..... .....tg... .....aaa
Ximu24D ...t..gatc ....at t.....g c..... ..... t..... ....gac.. ......t...
β2.domain
301
350
400
Con CTGACTAAAT CAGTGGCTCC CACCGTCAG. CT..ACTCCA CTACGCCCCC TGCTGGCCAC CATCCCTCCA TGCTGGTCTG CAGAGTTTAT GATTTCTATC
Ximu18F ....c ..tc..... ....
Ximu18G ....c ..tc..... ....a..... ....
Ximu20G ....a ..gt..... ....
Ximu20J ....a ..gt..... ....
Ximu17G ....a ..gt..... ....
Ximu23A ....a ..gt..... ....

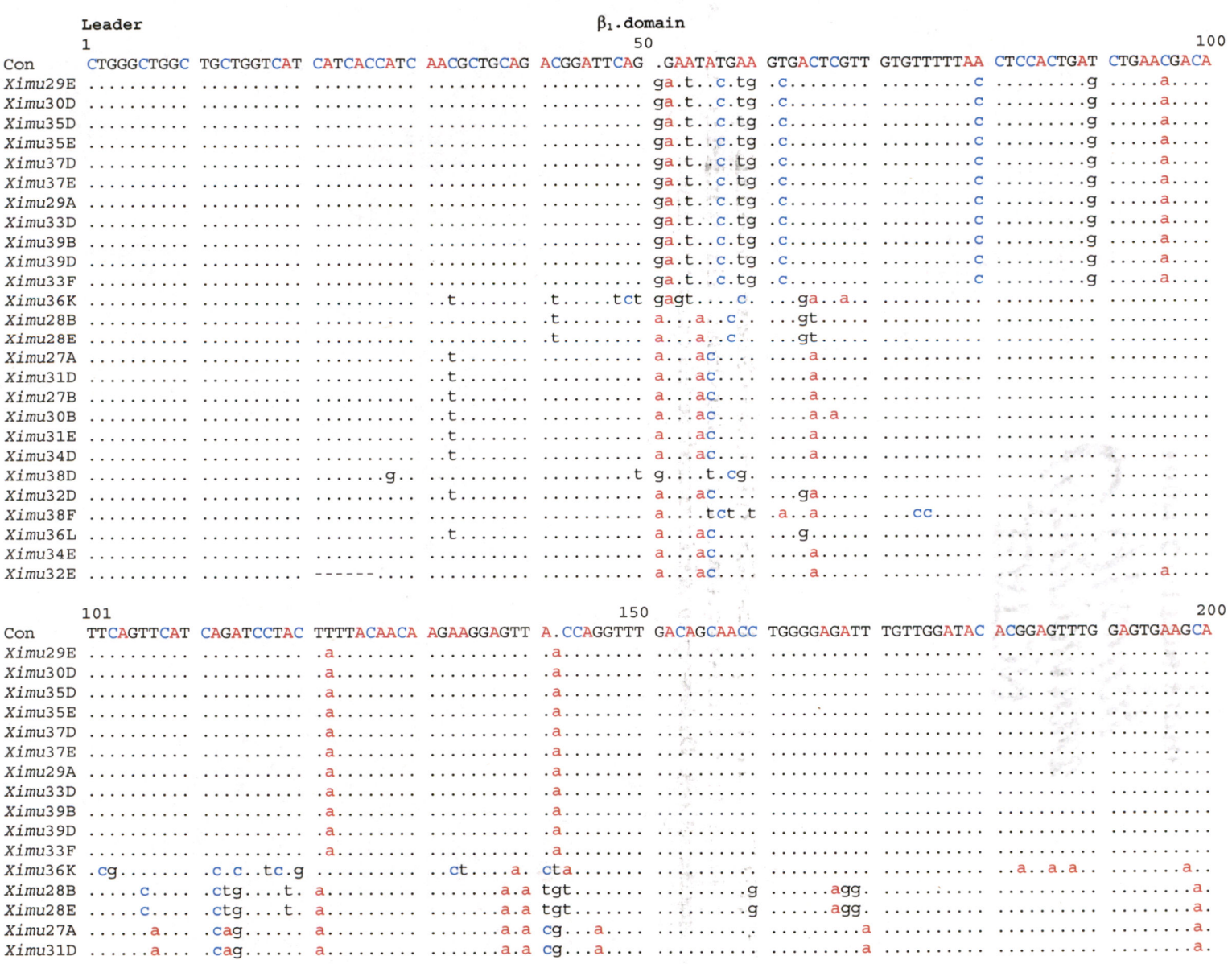

Ximu19B ....a ..gt..... ....c...
Ximu17F ....a....a ..gt..... ....
Ximu24E ....a ..gt..... ....
Ximu19G ....a ..gtt..... ....a..... ....
Ximu21B ....c ..tc..... gg..... ....
Ximu22E ....c ..tc..... gg..... ....
Ximu22H ....g.c ..tc..... gg..... ....
Ximu26D ....c ..tc..... gg..... ....
Ximu26E ....c ..tc..... gg..... ....
Ximu16F ...ga..... ....c ..tc..... gg..... ....
Ximu16G ...ga..... ....c ..tc..... gg..... ....
Ximu21D ...ga..... ....c ..tc..c... gg..... ....
Ximu23F ...ga..... ....c ..tc..... gg..... ....
Ximu25A ....a ..gt..... ....
Ximu25G ....a ..gt..... ....
Ximu15G ....a ..gt..... ....
Ximu15H ....a ..gt..... ....
Ximu24D ....c ..tc..... gg..... ....
401
450
500
Con CTAAAACCAT CAAAGTTCAG TGGCTGAGAG ACGGACAGGA AGTGACATCA GACGTCACCA CCACTGACGA GATGGAGGAC GGAGACTGGT ACTACCAGGT
Ximu18F ....
Ximu18G ....
Ximu20G ....g ....
Ximu20J ....
Ximu17G ....
Ximu23A ....
Ximu19B ....
Ximu17F ....
Ximu24E ....
Ximu19G ....
Ximu21B ....
Ximu22E ....g. .....
Ximu22H ....
Ximu26D .... .t..... ....
Ximu26E ....
Ximu16F ....a.. ....
Ximu16G ....a.. ....
Ximu21D ....a.. ....
Ximu23F ....a.. ....
Ximu25A ....
Ximu25G ....
Ximu15G ....
Ximu15H ....
Ximu24D ....

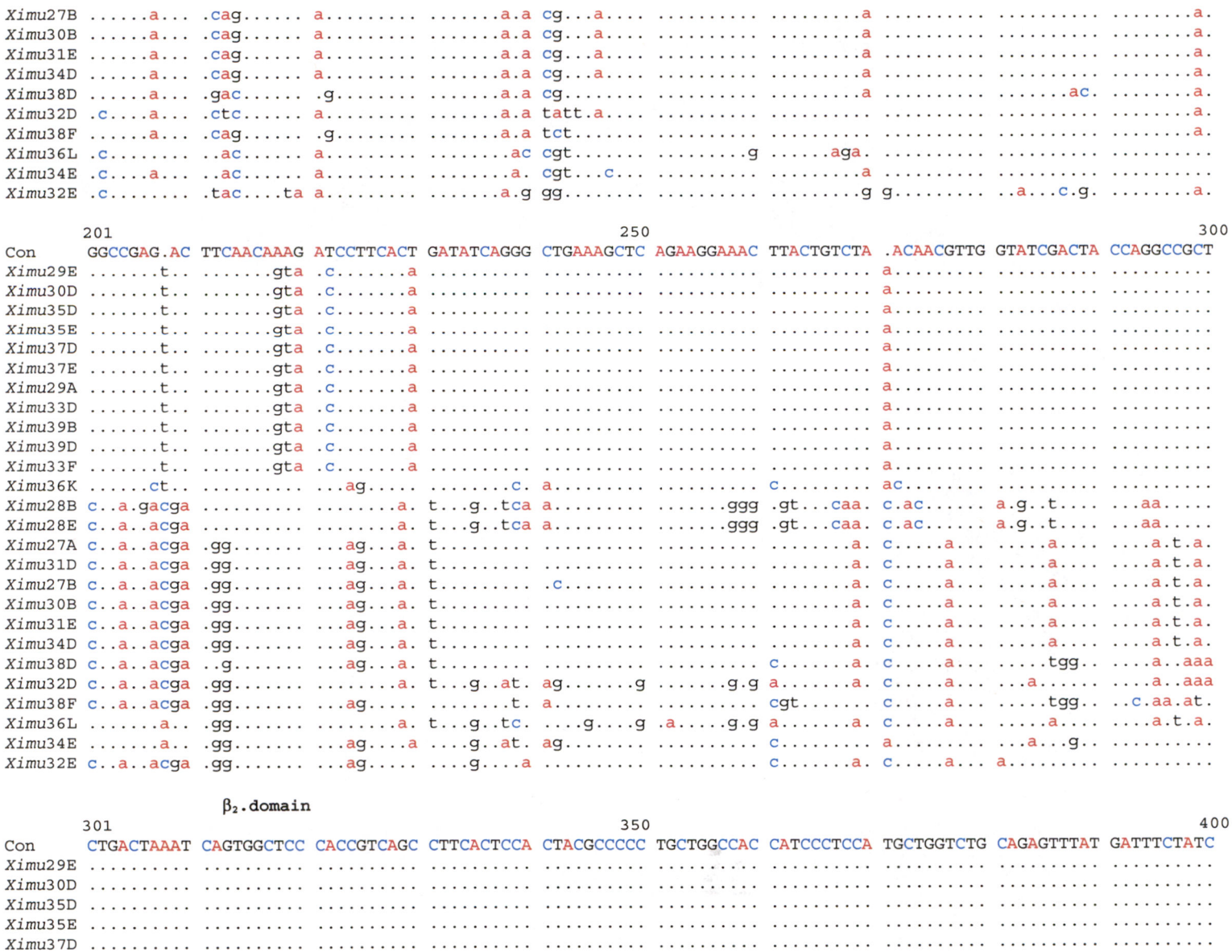

CP
501
550
600
Con CCACTCCCAG CTGGAGTACA CGCCCAGGTC TGGAGAGCGG ATCTCCTGCA GGGTGGAACA TGTCAGCCTG AAGGAACCTC TGATCACCGA CTGGGACCCG
Ximu18F
Ximu18G
Ximu20G
Ximu20J
Ximu17G
Ximu23A
Ximu19B
Ximu17F
Ximu24E
Ximu19G
Ximu21B .....a
Ximu22E
Ximu22H
Ximu26D
Ximu26E .....t
Ximu16F
Ximu16G
Ximu21D
Ximu23F
Ximu25A
Ximu25G
Ximu15G
Ximu15H
Ximu24D .....a...
TM
CT
601
650
700
Con TCCCTGCCAG AGTCAGAGAG GAACAAACTG GCCATCGGAG CTTCAGGACT GATCCTGGGT CTGATTTTGT CTCTGGCTGG ATTCATCTAC TACAAGAGGA
Ximu18F .....a.....
Ximu18G .....a.....
Ximu20G
Ximu20J
Ximu17G .....a
Ximu23A
Ximu19B
Ximu17F
Ximu24E .....a.....
Ximu19G .....t..
Ximu21B
Ximu22E
Ximu22H
Ximu26D
Ximu26E

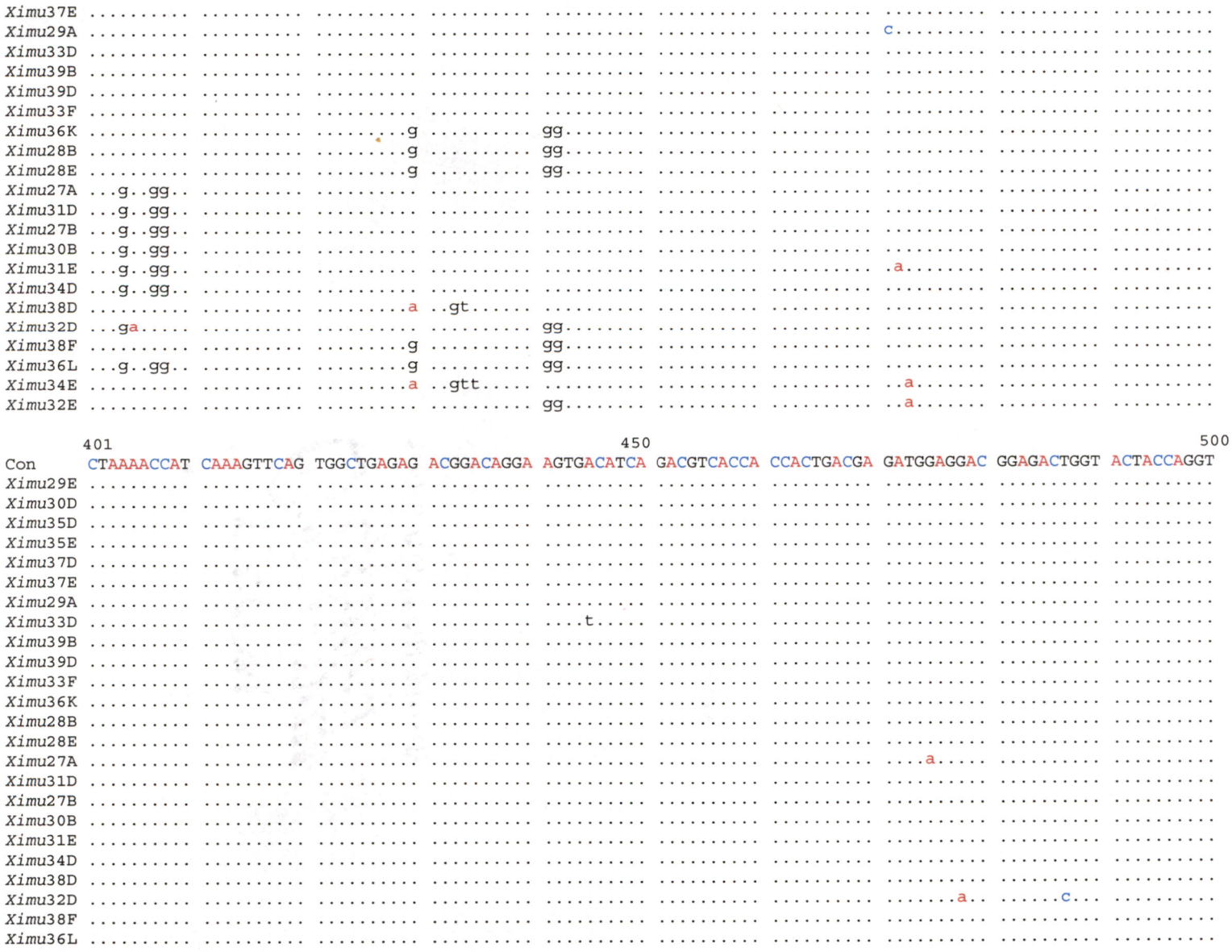
```
Ximu16F .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu16G .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu21D .......... ........g. .......... .......... .......... .......... .......... .......... .......... ..........
Ximu23F .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu25A .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu25G .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu15G .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu15H .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu24D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
                                                Stop
      701                                         735
Con      AGGTCAAAGG TCGTATTCTG GTTCCGACCA GCTGA
Ximu18F .......... .......... .......... .....
Ximu18G .......... .......... .......... .....
Ximu20G .......... .......... .......... .....
Ximu20J .......... .......... .......... .....
Ximu17G .......... .......... .......... .....
Ximu23A .......... .......... .......... .....
Ximu19B .......... .......... .......... .....
Ximu17F .......... .......... .......... .....
Ximu24E .......... .......... .......... .....
Ximu19G .......... .......... .......... .....
Ximu21B .......... .......... .......... .....
Ximu22E .......... .......... .......... .....
Ximu22H .......... .......... .......... .....
Ximu26D .......... .......... .......... .....
Ximu26E .......... .......... .......... .....
Ximu16F .......... .......... .......... .....
Ximu16G .......... .......... .......... .....
Ximu21D .......... .......... .......... .....
Ximu23F .......... .......... .......... .....
Ximu25A .......... .......... .......... .....
Ximu25G .......... .......... .......... .....
Ximu15G .......... .......... .......... .....
Ximu15H .......... .......... .......... .....
Ximu24D .......... .......... .......... .....
```

Figure 5: Nucleotide alignment of *Xiphophorus multilineatus* Arroyo Tambaque population. The protein domains are indicated. Dots indicate identity with the consensus sequence (Con). Dashes represent indels introduced for optimal alignment.


Leader
β1.domain
1
50
100
Con CTGGGCTGGC TGCTGGTCAT CATCACCATC AACGCTGCAG ACGGATTCAG .GAATATGAA GTGACTCGTT GTGTTTTTAA CTCCACTGAT CTGAACGACA
Ximu29E .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu30D .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu35D .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu35E .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu37D .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu37E .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu29A .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu33D .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu39B .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu39D .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu33F .... ga.t..c.tg .c..... .....c .....g .....a....
Ximu36K ....t..... .t.....tct gagt....c. ...ga..a.. ....
Ximu28B .... .t..... a...a..c.. ...gt..... ....
Ximu28E .... .t..... a...a..c.. ...gt..... ....
Ximu27A ....t..... ..... a...ac.... ....a..... ....
Ximu31D ....t..... ..... a...ac.... ....a..... ....
Ximu27B ....t..... ..... a...ac.... ....a..... ....
Ximu30B ....t..... ..... a...ac.... ....a.a... ....
Ximu31E ....t..... ..... a...ac.... ....a..... ....
Ximu34D ....t..... ..... a...ac.... ....a..... ....
Ximu38D ....g.. ....t g....t.cg. ....
Ximu32D ....t..... ..... a...ac.... ...ga..... ....
Ximu38F .... a....tct.t .a..a..... ...cc..... ....
Ximu36L ....t..... ..... a...ac.... ...g..... ....
Ximu34E .... a...ac.... ....a..... ....
Ximu32E .... ------.... .... a...ac.... ....a..... ....a....
101
150
200
Con TTCAGTTCAT CAGATCCTAC TTTTACAACA AGAAGGAGTT A.CCAGGTTT GACAGCAACC TGGGGAGATT TGTTGGATAC ACGGAGTTTG GAGTGAAGCA
Ximu29E .... .a..... ..... .a..... ....
Ximu30D .... .a..... ..... .a..... ....
Ximu35D .... .a..... ..... .a..... ....
Ximu35E .... .a..... ..... .a..... ....
Ximu37D .... .a..... ..... .a..... ....
Ximu37E .... .a..... ..... .a..... ....
Ximu29A .... .a..... ..... .a..... ....
Ximu33D .... .a..... ..... .a..... ....
Ximu39B .... .a..... ..... .a..... ....
Ximu39D .... .a..... ..... .a..... ....
Ximu33F .... .a..... ..... .a..... ....
Ximu36K .cg..... .c.c..tc.g ....ct....a. cta..... ....a..a.a.. .....a..
Ximu28B .....c.... .ctg....t. a..... .....a.a tgt..... .....g .....agg. ....a.
Ximu28E .....c.... .ctg....t. a..... .....a.a tgt..... .....g .....agg. ....a.
Ximu27A .....a.... .cag..... a..... .....a.a cg...a.... ....a ....a.
Ximu31D .....a.... .cag..... a..... .....a.a cg...a.... ....a ....a.


Ximu27B .....a... .cag..... a..... .....a.a cg...a.... ....a. ....a.
Ximu30B .....a... .cag..... a..... .....a.a cg...a.... ....a. ....a.
Ximu31E .....a... .cag..... a..... .....a.a cg...a.... ....a. ....a.
Ximu34D .....a... .cag..... a..... .....a.a cg...a.... ....a. ....a.
Ximu38D .....a... .gac..... .g..... .....a.a cg..... ....a. ....ac. .....a.
Ximu32D .c....a... .ctc..... a..... .....a.a tatt.a.... ....a.
Ximu38F .....a... .cag..... .g..... .....a.a tct..... ....a.
Ximu36L .c..... ..ac..... a..... .....ac cgt..... ....g .....aga. ....
Ximu34E .c....a... ..ac..... a..... .....a. cgt...c... ....a ....
Ximu32E .c..... .tac....ta a..... .....a.g gg..... ....g g..... ..a...c.g. .....a.
201
250
300
Con GGCCGAG.AC TTCAACAAAG ATCCTTCACT GATATCAGGG CTGAAAGCTC AGAAGGAAAC TTACTGTCTA .ACAACGTTG GTATCGACTA CCAGGCCGCT
Ximu29E .....t.. .....gta .c.....a .... a..... ....
Ximu30D .....t.. .....gta .c.....a .... a..... ....
Ximu35D .....t.. .....gta .c.....a .... a..... ....
Ximu35E .....t.. .....gta .c.....a .... a..... ....
Ximu37D .....t.. .....gta .c.....a .... a..... ....
Ximu37E .....t.. .....gta .c.....a .... a..... ....
Ximu29A .....t.. .....gta .c.....a .... a..... ....
Ximu33D .....t.. .....gta .c.....a .... a..... ....
Ximu39B .....t.. .....gta .c.....a .... a..... ....
Ximu39D .....t.. .....gta .c.....a .... a..... ....
Ximu33F .....t.. .....gta .c.....a .... a..... ....
Ximu36K .....ct.. ....ag..... .....c. a..... ..... c..... ac..... ....
Ximu28B c..a.gacga ....a. t...g..tca a..... .....ggg .gt...caa. c.ac..... a.g..t.... ...aa.....
Ximu28E c..a..acga ....a. t...g..tca a..... .....ggg .gt...caa. c.ac..... a.g..t.... ...aa.....
Ximu27A c..a..acga .gg..... ...ag...a. t..... ....a. c.....a... .....a.... ....a.t.a.
Ximu31D c..a..acga .gg..... ...ag...a. t..... ....a. c.....a... .....a.... ....a.t.a.
Ximu27B c..a..acga .gg..... ...ag...a. t..... .c..... ....a. c.....a... .....a.... ....a.t.a.
Ximu30B c..a..acga .gg..... ...ag...a. t..... ....a. c.....a... .....a.... ....a.t.a.
Ximu31E c..a..acga .gg..... ...ag...a. t..... ....a. c.....a... .....a.... ....a.t.a.
Ximu34D c..a..acga .gg..... ...ag...a. t..... ....a. c.....a... .....a.... ....a.t.a.
Ximu38D c..a..acga ..g..... ...ag...a. t..... .... c.....a. c.....a... .....tgg.. ....a..aaa
Ximu32D c..a..acga .gg..... .....a. t...g..at. ag.....g .....g.g a.....a. c.....a... ...a..... ....a..aaa
Ximu38F c..a..acga .gg..... ...ag..... .....t. a..... ..... cgt..... c.....a... .....tgg.. ..c.aa.at.
Ximu36L .....a.. .gg..... .....a. t...g..tc. ....g....g .a.....g.g a.....a. c.....a... .....a.... ....a.t.a.
Ximu34E .....a.. .gg..... ...ag....a ....g..at. ag..... ..... c..... a..... ...a...g.. .....
Ximu32E c..a..acga .gg..... ...ag..... ....g....a .... c.....a. c.....a... a..... .....
β2.domain
301
350
400
Con CTGACTAAAT CAGTGGCTCC CACCGTCAGC CTTCACTCCA CTACGCCCCC TGCTGGCCAC CATCCCTCCA TGCTGGTCTG CAGAGTTTAT GATTTCTATC
Ximu29E ....
Ximu30D ....
Ximu35D ....
Ximu35E ....
Ximu37D ....


Ximu37E
Ximu29A
Ximu33D
Ximu39B
Ximu39D
Ximu33F
Ximu36K
Ximu28B
Ximu28E
Ximu27A
Ximu31D
Ximu27B
Ximu30B
Ximu31E
Ximu34D
Ximu38D
Ximu32D
Ximu38F
Ximu36L
Ximu34E
Ximu32E
401
450
500
Con CTAAAACCAT CAAAGTTCAG TGGCTGAGAG ACGGACAGGA AGTGACATCA GACGTCACCA CCACTGACGA GATGGAGGAC GGAGACTGGT ACTACCAGGT
Ximu29E
Ximu30D
Ximu35D
Ximu35E
Ximu37D
Ximu37E
Ximu29A
Ximu33D
Ximu39B
Ximu39D
Ximu33F
Ximu36K
Ximu28B
Ximu28E
Ximu27A
Ximu31D
Ximu27B
Ximu30B
Ximu31E
Ximu34D
Ximu38D
Ximu32D
Ximu38F
Ximu36L

```
Ximu34E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu32E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

                                                                                                      CP
        501                                               550                                            600
Con     CCACTCCCAG CTGGAGTACA CGCCCAGGTC TGGAGAGCGG ATCTCCTGCA GGGTGGAACA TGTCAGCCTG AAGGAACCTC TGATCACCGA CTGGGACCCG
Ximu29E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu30D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu35D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu35E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu37D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu37E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu29A .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu33D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu39B .......... .......g.. .......... .......... .......... .......... .......... .......... .......... ..........
Ximu39D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu33F .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu36K .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu28B .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu28E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu27A .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu31D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu27B .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu30B .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu31E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu34D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu38D .......... .......... .......... .......... .......... .......... .......... .......... .......... .........c
Ximu32D .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu38F .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu36L .......... .......... .......... .......... .......... .......... .......... .......... .....g.... ..........
Ximu34E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Ximu32E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

                              TM                                                                     CT
        601                                               650                                            700
Con     TCCCTGCCAG AGTCAGAGAG GAACAAACTG GCCATCGGAG CTTCAGGACT GATCCTGGGT CTGATTTTGT CTCTGGCTGG ATTCATCTAC TACAAGAGGA
Ximu29E .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu30D .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu35D .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu35E .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu37D .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu37E .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu29A .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu33D .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu39B .......... .......... a......... .......... .......... .......... .......... .......... .......... ..........
Ximu39D .......... .........a a......... .......... .......... .......... .......... .......... .......... ..........
Ximu33F .......... .......... a......... .......... .......... .......... .c........ .c........ .......... ..........
Ximu36K .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
Ximu28B ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu28E ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu27A ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu31D ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu27B ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu30B ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu31E ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu34D ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu38D ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu32D ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu38F ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu36L ..........  ..........  ..........  ..........  ..........  ...t......  ..........  ..........  ..........  ..........
Ximu34E ..........  ..........  .......t..  ..........  ..........  ..........  ..........  ..........  ..........  ..........
Ximu32E ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........

                                                  Stop
        701                                         735
Con     AGGTCAAAGG TCGTATTCTG GTTCCGACCA GCTGA
Ximu29E ..........  ..........  ..........  .....
Ximu30D ..........  ..........  ..........  .....
Ximu35D ..........  ..........  ..........  .....
Ximu35E ..........  ..........  ..........  .....
Ximu37D ..........  ..........  ..........  .....
Ximu37E ..........  ..........  ..........  .....
Ximu29A ..........  ..........  ..........  .....
Ximu33D ..........  ..........  ..........  .....
Ximu39B ..........  ..........  ..........  .....
Ximu39D ..........  ..........  ..........  .....
Ximu33F ..........  ..........  ..........  .....
Ximu36K ..........  ..........  ..........  .....
Ximu28B ..........  ..........  ..........  .....
Ximu28E ..........  ..........  ..........  .....
Ximu27A ...c......  ..........  ..........  .....
Ximu31D ...c......  ..........  ..........  .....
Ximu27B ...c......  ..........  ..........  .....
Ximu30B ...c......  ..........  ..........  .....
Ximu31E ...c......  ..........  ..........  .....
Ximu34D ..........  ..........  ..........  .....
Ximu38D ..........  ..........  ..........  .....
Ximu32D ..........  ..........  ..........  .....
Ximu38F ..........  ..........  ..........  .....
Ximu36L ..........  ..........  .....a....  .....
Ximu34E ..........  ..........  ..........  .....
Ximu32E ...c......  ..........  ..........  .....
```

Figure 6: A phylogeny of *Xiphophorus DAB* sequences using the neighbor-joining tree construction method based on the Jukes-Cantor corrected distances for the entire sequence. Bootstrap values above 50% are listed on their corresponding branches. Blue stars (★) represent potential allelic lineages. Red stars (✦) represent sequences with a six-site gap in the leader sequence. Platyfish (*XimaDAB*) and guppy (*PoreDAB*) sequences were included for comparison purposes, and the channel catfish (*IcpuDAB*) sequence was included as an outgroup to root the tree.

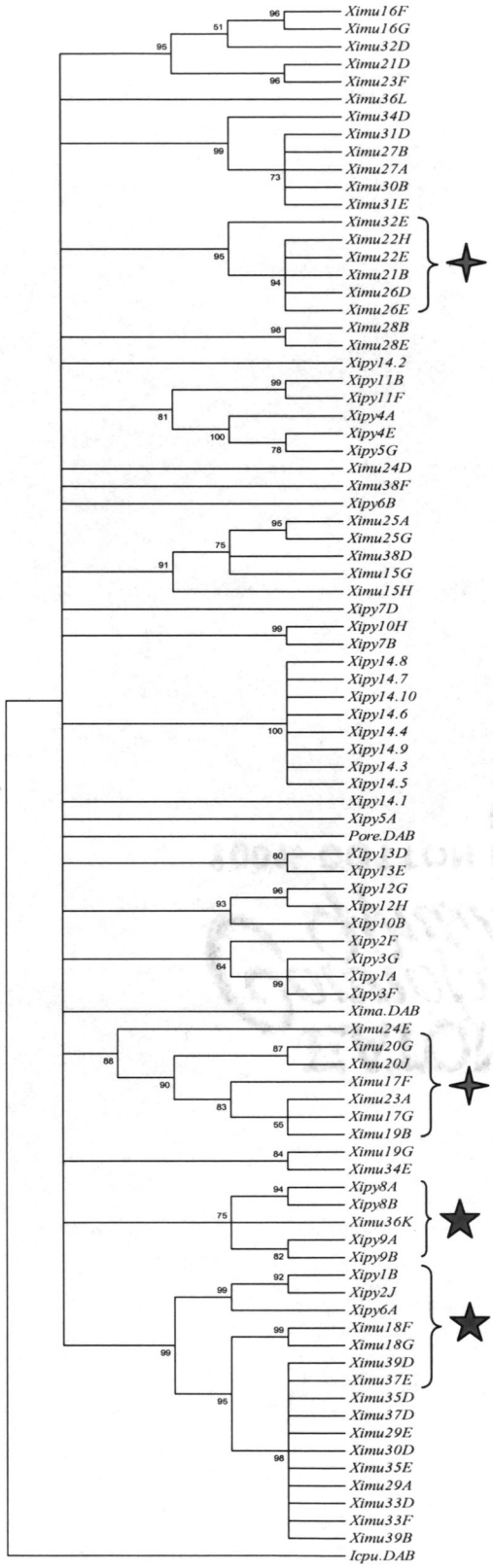

Ximu16F
Ximu16G
Ximu32D
Ximu21D
Ximu23F
Ximu36L
Ximu34D
Ximu31D
Ximu27B
Ximu27A
Ximu30B
Ximu31E
Ximu32E
Ximu22H
Ximu22E
Ximu21B
Ximu26D
Ximu26E
Ximu28B
Ximu28E
Xipy14.2
Xipy11B
Xipy11F
Xipy4A
Xipy4E
Xipy5G
Ximu24D
Ximu38F
Xipy6B
Ximu25A
Ximu25G
Ximu38D
Ximu15G
Ximu15H
Xipy7D
Xipy10H
Xipy7B
Xipy14.8
Xipy14.7
Xipy14.10
Xipy14.6
Xipy14.4
Xipy14.9
Xipy14.3
Xipy14.5
Xipy14.1
Xipy5A
Pore.DAB
Xipy13D
Xipy13E
Xipy12G
Xipy12H
Xipy10B
Xipy2F
Xipy3G
Xipy1A
Xipy3F
Xima.DAB
Ximu24E
Ximu20G
Ximu20J
Ximu17F
Ximu23A
Ximu17G
Ximu19B
Ximu19G
Ximu34E
Xipy8A
Xipy8B
Ximu36K
Xipy9A
Xipy9B
Xipy1B
Xipy2J
Xipy6A
Ximu18F
Ximu18G
Ximu39D
Ximu37E
Ximu35D
Ximu37D
Ximu29E
Ximu30D
Ximu35E
Ximu29A
Ximu33D
Ximu33F
Ximu39B
Icpu.DAB

Figure 7: A phylogeny of *Xipy*14 *DAB* sequences using the neighbor-joining tree construction method based on the Jukes-Cantor corrected distances for the entire sequence. Bootstrap values above 50% are listed on their corresponding branches. Asterisks (*) represent identical sequences.

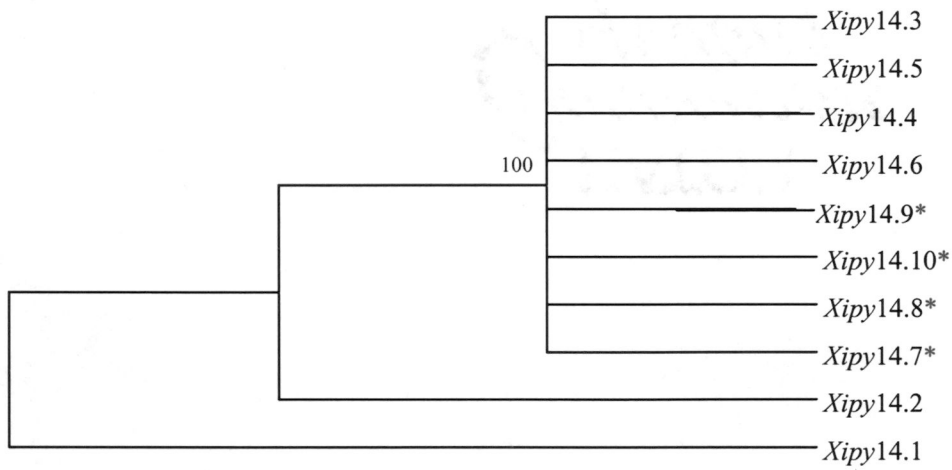
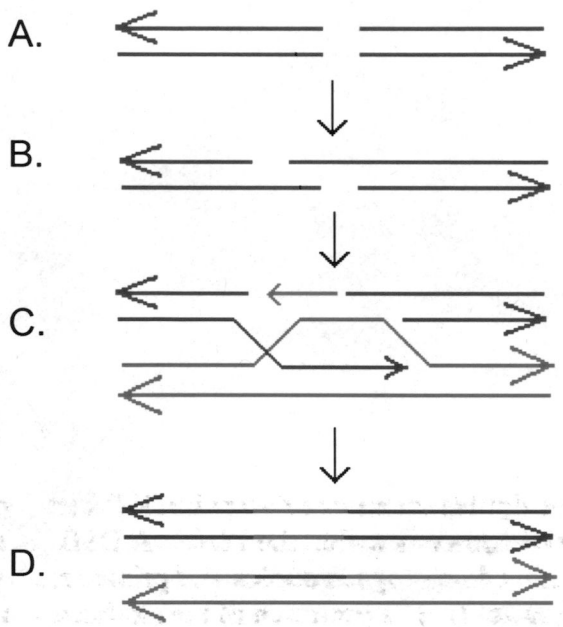

100
*Xipy*14.3
*Xipy*14.5
*Xipy*14.4
*Xipy*14.6
*Xipy*14.9*
*Xipy*14.10*
*Xipy*14.8*
*Xipy*14.7*
*Xipy*14.2
*Xipy*14.1

Figure 8: A schematic of the double-strand break-repair (DSBR) mechanism hypothesized to cause gene conversions within the MHC. A DSB (A.) is processed to two 3' ends (B.), which invade a homologous duplex and prime repair synthesis producing two Holliday junctions (C.). Resolution of the Holliday junction causes gene conversion as a consequence of strand switching between loci coupled with repair synthesis and mismatch repair of hDNA (D.). Adapted from Martinsohn *et al.* (1999).


A.
B.
C.
D.

References

Abbé, D.L., A. Belmaaza, F. Decary, and P. Chartrand, 1992 Elimination of heteroduplex artifacts when sequencing *HLA* genes amplified by polymerase chain reaction (PCR). Immunogenetics **35**: 395-397.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, 1990 Basic Local Alignment Search Tool. J. Mol. Biol. **215**: 403-410.

Bartl, S. and I.L. Weissman, 1994 Isolation and characterization of major histocompatibility complex class II *B* genes from the nurse shark. Proc. Natl. Acad. Sci. USA **91**: 262-266.

Basolo, A.L., 1990 Female preference predates the evolution of the sword in swordtail fish. Science **250**: 808-810.

Basolo, A.L., 1995 A further examination of a pre-existing bias favouring a sword in the genus *Xiphophorus*. Animal Behavior **50**: 365-375.

Beck, S. and J. Trowsdale, 1999 Sequence organisation of the class II region of the human MHC. Immunological Reviews **167**: 201-210.

Bergstrom, T.F., A. Josefsson, H.A. Erlich, and U. Gyllensten, 1998 Recent origin of *HLA-DRB1* alleles and implications for human evolution. Nat. Genet. **18**: 237-242.

Bickham, J.W., S. Sandhu, P.D.N. Hebert, L. Chikhi, and R. Athwal, 2000 Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. Mutation Research **463**: 33-51.

Borriello, F. and K. S. Krauter, 1990 Reactive site polymorphism in the murine protease inhibitor gene family in delineated using a modification of the PCR reaction (PCR+1). Nucleic Acids Research **18**: 5481-5487.

Brown, J.H., T.S. Jardetzky, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, and D.C. Wiley, 1993 Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature **364**: 33-39.

Burnet, F.M., 1971 "Self-recognition" in colonial marine forms and flowering plants in relation to the evolution of immunity. Nature **232**: 230-235.

Devereux, J., P. Haeberli, and O. Simithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Research **12**: 387-395.

Dixon, B., L.A.J. Nagelkerke, F.A. Sibbing, E. Egberts, and R.J.M. Stet, 1996 Evolution of MHC class II β chain-encoding genes in the Lake Tana barbell species flock (*Barbus intermedius* complex). Immunogenetics **44**: 419-431.

Dove, A.D.M. and I. Ernst, 1998 Concurrent invaders – four exotic species of Monogenea now established on exotic freshwater fishes in Australia. International Journal for Parasitology **28**: 1755-1764.

Ebert, D. and R. Lorenzi, 1994 Parasites and polymorphisms. Nature **369**: 705-706.

Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**: 368-376.

Felsenstein, J., 1988 Phylogenies from molecular sequences: inferences and reliability. Annu. Rev. Genet. **22**: 521-565.

Figueroa, F., W.E. Mayer, A. Sato, Z. Zaleska-Rutczynska, B. Hess, H. Tichy, and J. Klein, 2001 Mhc class I genes of swordtail fishes, *Xiphophorus*: variation in the number of loci and existence of ancient gene families. Immunogenetics **53**: 695-708.

Fitch, W.M. and E. Margoliash, 1967 Construction of phylogenetic trees. Science **155**: 279-284.

Fitch, W.M., J.M.E. Leiter, X. Li, and P. Palese, 1991 Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. USA **88**: 4270-4274.

Glamann, J., 1995 Complete coding sequence of rainbow trout *Mhc* II β-chain. Scand. Journal of Immunology **41**: 365-372.

Gordon, M., 1927 The genetics of a viviparous top-minnow *Platypoecilus*; the inheritance of two kinds of melanophores. Genetics **12**: 253-283.

Graser, R., C. O'hUigin, V. Vincek, A. Meyer, and J. Klein, 1996 Trans-species polymorphism of class II Mhc loci in danio fishes. Immunogenetics **44**: 36-48.

Grosberg, R.K. and M.W. Hart, 2000 Mate selection and the evolution of highly polymorphic self/nonself recognition genes. Science **289**: 2111-2114.

Hashimoto, K., T. Nakanishi, and Y. Kurosawa, 1990 Isolation of carp genes encoding major histocompatibility complex antigens. Proc. Nat. Acad. Sci. USA **87**: 6863-6867.

Hedrick, P.W., K. M. Parker, E.L. Miller, and P.S. Miller, 1999 Major histocompatibility complex variation in the endangered Przewalski's horse. Genetics **152**: 1701-1710.

Hill, A.V.S., C.E.M. Allsopp, D. Kwiatkowski, N.M. Anstey, P. Twumasi, P.A. Rowe, S. Bennett, D. Brewster, A.J. McMichael, and B.M. Greenwood, 1991 Common West African HLA antigens are associated with protection from severe malaria. Nature **352**: 595-600.

Högstrand, K. and J. Böhme, 1999 Gene conversion can create new MHC alleles. Immunological Reviews **167**: 305-317.

Hughes, A.L., 1991 Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics **127**: 345-353.

Hughes, A.L., 1999 Adaptive Evolution of Genes and Genomes, Oxford University Press, New York.

Hughes, A.L. and M.K. Hughes, 1995 Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics **42**: 233-243.

Hughes, A.L. and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominant selection. Nature **335**: 167-170.

Hughes, A.L. and M. Nei, 1989 Nucleotide substitutions at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**: 958-962.

Hughes, A.L. and M. Yeager, 1998 Natural selection and the evolutionary history of major histocompatibility complex loci. Frontiers in Bioscience **3**: 509-516.

Hughes, A.L. and M. Yeager, 1998 Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. **32**: 415-435.

Ikemura, T., 1992 Correlation between codon usage and tRNA content in microorganisms, in Transfer RNA in Protein Synthesis, edited by D.L. Hatfield, B.J. Lee, and R.M. Pirtle. CRC Press, Boca Raton.

Imanishi, T. and T. Gojobori, 1992 Patterns of nucleotide substitutions inferred from the phylogenies of the class I major histocompatibility complex genes. Journal of Molecular Evolution **35**: 196-204.

Jukes, T.H. and C.R. Cantor, 1969 Evolution of protein molecules, in Mammalian Protein Metabolism, edited by H.N. Munro. Academic Press, New York.

Kallman, K.D., 1958 Genetics of fin transplantation in Xiphophorin fishes. Annals New York Acad. Sci. **73**: 599-610.

Kangueane, P., M.K. Sakharkar, P.R. Kolatkar, and E.C. Ren, 2001 Towards the MHC combinatorics. Human Immunology **62**: 539-556.

Kasahara, M., E.C. McKinney, M.F. Flajnik, and T. Ishibashi, 1993 The evolutionary origin of the major histocompatibility complex: polymorphism of class II α-chain genes in the cartilaginous fish. Eur. J. Immunol. **23**: 2160-2165.

Kasahara, M., M.F. Flajnik, T. Ishibashi, and T. Natori, 1995 Evolution of the major histocompatibility complex: a current overview. Transplant Immunology **3**: 1-20.

Kasahara, M., J. Nakaya, Y. Satta, and N. Takahata, 1997 Chromosomal duplication and the emergence of the adaptive immune system. Trends in Genetics **13**: 90-92.

Kaufman, J., 1999 Co-evolving genes in MHC haplotypes: the "rule" for nonmammalian vertebrates? Immunogenetics **50**: 228-236.

Kim, T.J., K.M. Parker, and P.W. Hedrick, 1999 Major histocompatibility complex differentiation in the Sacramento River Chinook salmon. Genetics **151**: 1115-1122.

Klein, J., 1987 Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Human Immunology **19**(3): 155-162.

Klein, J., C. O'hUigin, F. Figueroa, W.E. Mayer, and D. Klein, 1993 Different modes of *Mhc* evolution in primates. Mol. Biol. Evol. **10**(1): 48-59.

Klein, J., D. Klein, F. Figueroa, A. Sato, and C. O'hUigin, 1997 Major Histocompatibility genes in the study of fish phylogeny, in Molecular Systematics of Fishes, edited by T.D. Kocher and C.A. Stepien. Academic Press, San Diego.

Kriener, K., C. O'hUigin, H. Tichy, and J. Klein, 2000 Convergent evolution of major histocompatibility complex molecules in humans and New World Monkeys. Immunogenetics **51**: 169-178.

Kumar, S., K. Tamura, and M. Nei, 1993 MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02, The Pennsylvania State University, University Park, PA.

Liskay, R.M., J.L. Stachelek, and A. Letsou, 1984 Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harbor Symp. Quant. Biol. **49**: 183-189.

Martinsohn, J.T., A.B. Sousa, L.A. Guethlein, and J.C. Howard, 1999 The gene conversion hypothesis of MHC evolution: a review. Immunogenetics **50**: 168-200.

McConnell, T.J., U.B. Godwin, and B.J. Cuthbertson, 1998a Expressed major histocompatibility complex class II loci in fishes. Immunological Reviews **166**: 294-300.

McConnell, T.J., U.B. Godwin, S.F. Norton, R.S. Narin, S. Kazianis, and D.C. Morizot, 1998b Identification and mapping of two divergent, unlinked major histocompatibility complex class II *B* genes in *Xiphophorus* fishes. Genetics **149**: 1921-1934.

McConnell, T.J., W.S. Talbot, R.A. McIndoe, and E.K. Wakeland, 1988 The origin of MHC class II gene polymorphism within the genus *Mus*. Nature **332**(6165): 651-654.

McWeeney, S.K. and A.M. Valdes, 1999 Codon usage bias and base composition in MHC genes in humans and common chimpanzees. Immunogenetics **49**: 272-279.

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

Mikko, S. and L. Andersson, 1995 Extensive *Mhc* class II diversity in African and European cattle. Immunogenetics **42**: 408-413.

Morizot, D.C., and M.J. Siciliano, 1979 Polymorphisms, linkage, and mapping of four enzyme loci in the fish genus *Xiphophorus* (Poeciliidae). Genetics **93**: 947-960.

Morizot, D.C., 1990 Use of fish gene maps to predict ancestral vertebrate genome organization, in Isozymes: Structure, Function and Use in Biology and Medicine, edited by Z.I. Ogita and C.L. Markert. Wiley-Liss, Inc., New York.

Morizot, D.C., S.A. Slaugenhaupt, K.D. Kallman, and A. Chakravarti, 1991 Genetic linkage map of fishes in the genus *Xiphophorus* (Teleostei: Poeciliidae). Genetics **127**: 399-410.

Morris, M.R., W.E. Wagner Jr., and M.J. Ryan, 1996 A negative correlation between trait and mate preference in *Xiphophorus pygmaeus*. Animal Behavior **52**: 1193-1203.

Morris, M.R., K. de Queiroz, and D.C. Morizot, 2001 Phylogenetic relationships among populations of northern swordtails (*Xiphophorus*) as inferred from allozyme data. *Copeia* **1**: 65-81.

Murray, B. W., S. Malik, and B.N. White, 1995 Sequence variation at the major histocompatibility complex locus DQ in Beluga Whales (*Delphinapterus leucas*). Mol. Biol. Evol. **12**(4): 582-593.

Nei, M., 1972 Genetic distance between populations. The American Naturalist **106**: 283-292.

Ober, C., L.R. Weitkamp, N. Cox, H. Dytch, D. Kostyu, and S. Elias, 1997 HLA and mate choice in humans. American Journal of Human Genetics **61**: 497-504.

O'Brien, S.J., D.E. Wildt, M. Bush, T.M. Caro, C. FitzGibbon, I. Aggundey, and R.E. Leakey, 1987 East African cheetahs: evidence for two population bottlenecks? Proc. Natl. Acad. Sci. USA **84**: 508-511.

Ohta, T., 1995 Gene conversion vs. point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. J. of Mol. Evol. **45**: 115-119.

Ohta, T., 1998 On the pattern of polymorphisms at major histocompatibility complex loci. Journal of Molecular Evolution **46**: 633-638.

Ohta, T., 1999 Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunological Reviews **167**: 319-325.

Ono, H., C. O'hUigin, V. Vincek, and J. Klein, 1993 Exon-intron organization of fish major histocompatibility complex class II *B* genes. Immunogenetics **38**: 223-234.

Penn, D. and W.K. Potts, 1998 Untrained mice discriminate MHC-determined odors. Physiology & Behavior **64**(3): 235-243.

Powell, J.R. and E.N. Moriyama, 1997 Evolution of codon usage bias in *Drosophila*. Proc. Natl. Acad. Sci. USA **94**: 7784-7790.

Powis, S.H., and D.E. Geraghty, 1995 What is the MHC? Immunology Today **16**: 466-468.

Quattro, J.M., J.C. Avise, and R.C. Vrijenhoek, 1992 An ancient clonal lineage in the fish genus *Poeciliopsis* (Atheriniformes: Poeciliidae). Proc. Natl. Acad. Sci. USA **89**: 348-352.

Rached, L.A., M.F. McDermott, and P. Pontarotti, 1999 The MHC big bang. Immunological Reviews **167**: 33-45.

Rzhetsky, A. and M. Nei, 1992 A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**: 945-967.

Sanger, F., S. Nicklen, and A.R. Coulson, 1977 DNA sequencing with chain terminating inhibitors. Proc. Nat. Acad. Sci. USA **74**: 5463-5467.

Saitou, N. and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.

Satta, Y., C. O'hUigin, N. Takahata, and J. Klein, 1993 The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc. Natl. Acad. Sci. USA **90**: 7480-7484.

Sato, A., F. Figueroa, C. O'hUigin, D.N. Reznick, and J. Klein, 1995 Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata*. Immunogenetics **43**: 38-49.

Schwaiger, F.W., E. Weyers, J. Buitkamp, A.J. Ede, A. Crawford, and J.T. Epplen, 1994 Interdependent MHCDR exon-plus-intron evolution in artiodactyls. Mol. Biol. Evol. **11**: 239-249.

She, J.X., S.A. Boehme, T.W. Wang, F. Bonhomme, and E.K. Wakeland, 1991 Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. Proc. Natl. Acad. Sci. USA **88**: 453-457.

Sültmann, H., W.E. Mayer, F. Figueroa, C. O'hUigin, and J. Klein, 1994 Organization of the Mhc class II *B* genes in the zebrafish (*Brachydanio rerio*). Genomics **23**: 1-14.

Suzuki, Y. and T. Gojobori, 2001 Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. Gene **276**: 83-87.

Swofford, D.L., 2000 PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4, Sinauer Associates, Sunderland, MA.

Szostak, J.W., T.L. Orr-Weaver, R.J. Rothstein, and F.W. Stahl, 1983 The double-strand break-repair model for recombination. Cell **33**: 25-35.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research **24**: 4876-4882.

Thursz, M.R., H.C. Thomas, B.M. Greenwood, and A.V.S. Hill, 1997 Heterozygote advantage for HLA class-II type in hepatitis B virus. Nat. Genet. **17**: 11-12.

van Erp, S.H.M., E. Egberts, and R.J.M. Stet, 1996 Characterization of class II *A* and *B* genes in a gynogenetic carp clone. Immunogenetics **44**: 192-202.

Walker, R.B. and T.J. McConnell, 1994 Polymorphism of the MHC *Mosa* class II β-chain encoding gene in striped bass (*Morone saxatilis*). Dev. Comp. Immunology **18**: 325-342.

Wiens, J.J., 1995 Polymorphic characters in phylogenetic systematics. Syst. Biol. **44**: 482-500.

Wiens, J.J., and M.R. Servedio, 1997 Accuracy of phylogenetic analysis including and excluding polymorphic characters. Syst. Biol. **46**: 332-345.

Wiens, J.J., 1999 Polymorphism in systematics and comparative biology. Annu. Rev. Ecol. Syst. **30**: 327-362.

Yuhki, N. and S.J. O'Brien, 1990 DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proc. Natl. Acad. Sci. **87**: 836-840.