UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Interacciones parásito-hospedador y especiación en peces cíclidos de lagunas cratéricas neotropicales

Host-parasite interactions and speciation in neotropical crater lake cichlids

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Melinda Jeanet Hofmann

Directora

Marta Barluenga Badiola

Madrid, 2017

UNIVERSIDAD COMPLUTENSE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

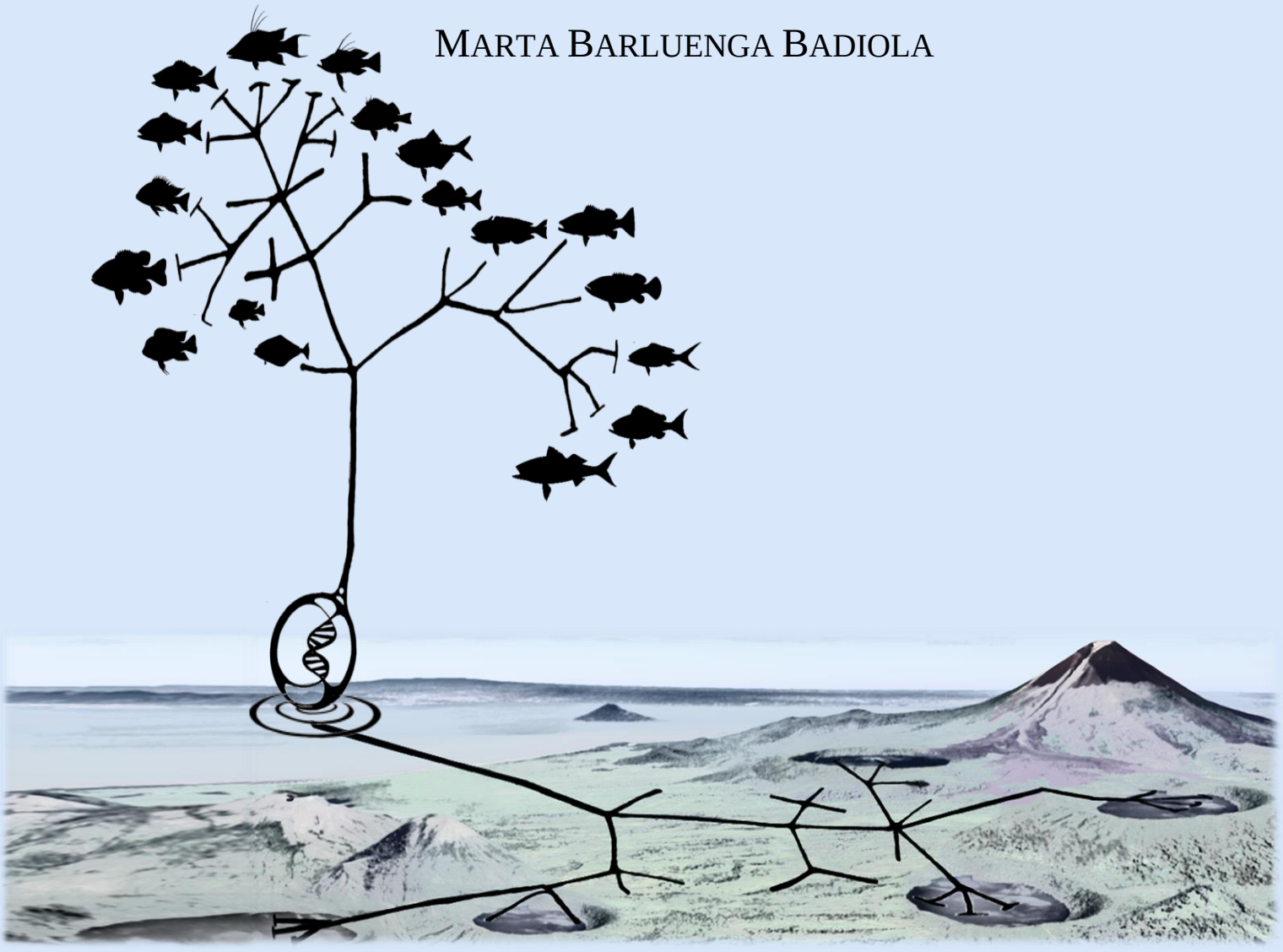
INTERACCIONES PARÁSITO-HOSPEDADOR Y ESPECIACIÓN EN PECES CICLÍDOS DE LAGUNAS CRATÉRICAS NEOTROPICALES

TESIS DOCTORAL

MELINDA J. HOFMANN

DIRIGIDO POR

MARTA BARLUENGA BADIOLA



INTERACCIONES PARÁSITO-HOSPEDADOR Y ESPECIACIÓN EN CÍCLIDOS DE LAGUNAS CRATÉRICAS NEOTROPICALES

HOST-PARASITE INTERACTIONS AND SPECIATION IN NEOTROPICAL CRATER LAKE CICHLIDS

UNIVERSIDAD
COMPLUTENSE
MADRID

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Melinda Jeanet Hofmann

Bajo la dirección de:

Marta Barluenga Badiola

Madrid, viernes 31 de marzo del 2017

TITULADO:

INTERACCIONES PARÁSITO-HOSPEDADOR Y ESPECIACIÓN EN CÍCLIDOS DE LAGUNAS CRATÉRICAS NEOTROPICALES

Fdo. La Doctoranda	Fdo. La Directora
Melinda Jeanet Hofmann	**Marta Barluenga Badiola**

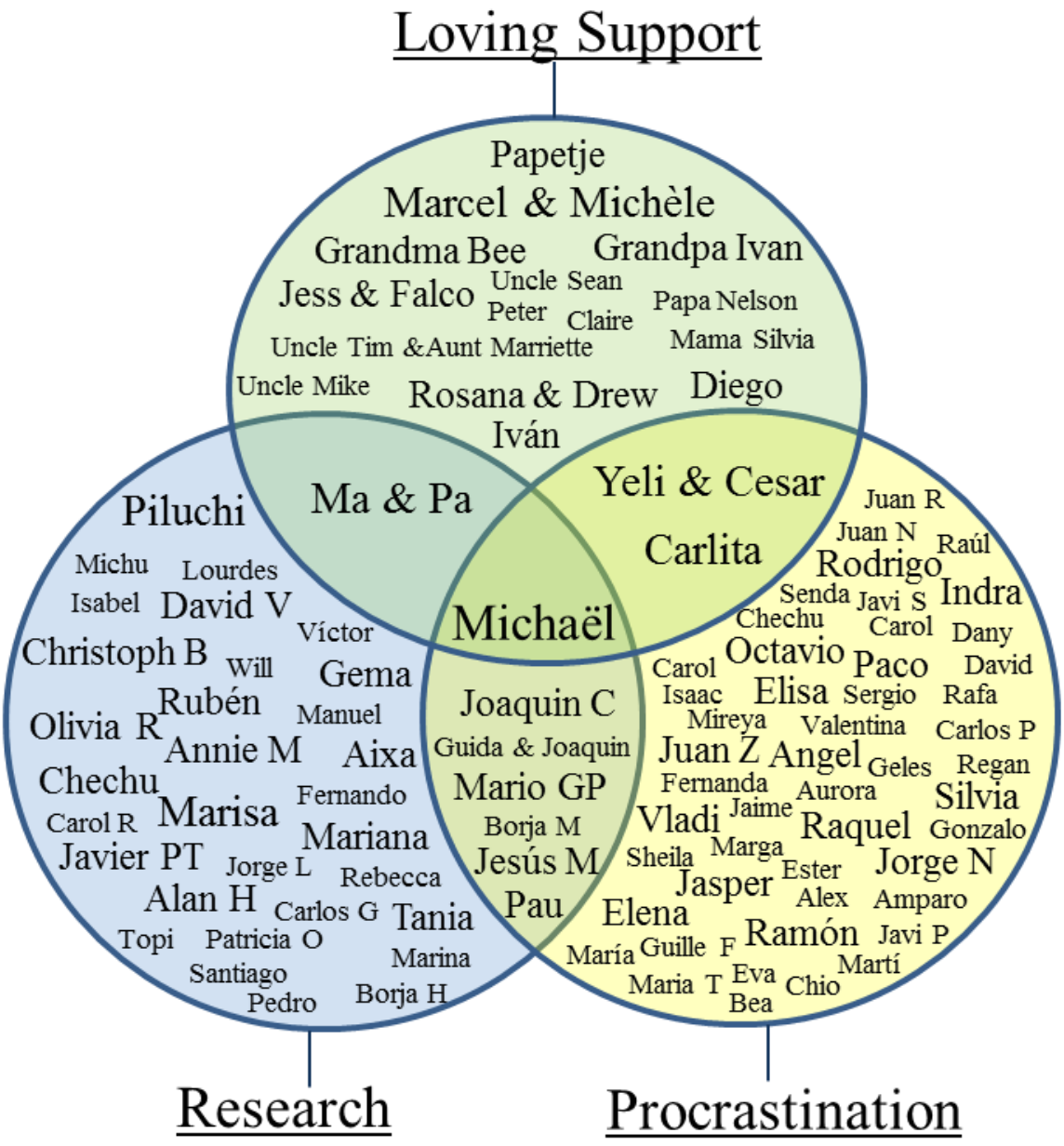
"Endless forms most beautiful and most wonderful have been, and are being, evolved."

- Charles Darwin

The Origin of Species 1859

CONTENTS

Acknowledgements


Loving Support
Papetje
Marcel & Michèle
Grandma Bee
Grandpa Ivan
Jess & Falco
Uncle Sean
Peter
Claire
Papa Nelson
Uncle Tim &Aunt Marriette
Mama Silvia
Uncle Mike
Rosana & Drew
Diego
Iván
Ma & Pa
Yeli & Cesar
Carlita
Michaël
Piluchi
Michu
Lourdes
Isabel
David V
Víctor
Christoph B
Will
Gema
Rubén
Manuel
Olivia R
Annie M
Aixa
Chechu
Fernando
Carol R
Marisa
Mariana
Javier PT
Jorge L
Rebecca
Alan H
Carlos G
Tania
Topi
Patricia O
Marina
Santiago
Pedro
Borja H
Joaquin C
Guida & Joaquin
Mario GP
Borja M
Jesús M
Pau
Juan R
Juan N
Raúl
Rodrigo
Senda
Javi S
Indra
Chechu
Carol
Dany
Octavio
Paco
David
Carol
Isaac
Elisa
Sergio
Rafa
Mireya
Valentina
Carlos P
Juan Z
Angel
Geles
Regan
Fernanda
Aurora
Silvia
Vladi
Jaime
Raquel
Gonzalo
Sheila
Marga
Ester
Jorge N
Jasper
Alex
Amparo
Elena
Ramón
Javi P
María
Guille
F
Martí
Maria T
Eva
Chio
Bea
Research
Procrastination

Funding

This work was funded by the Ministry of Economy and Competitiveness of the Government of Spain by a scholarship of Researcher Training Program FPI BES-2011-047645 granted to M. Hofmann, and by the state Program of Development of Scientific and Technical Research of Excellence, project CGL 2010-16103 awarded to M. Barluenga. This project was further enabled with the support of two German Science Foundation grants to C. Eizaguirre (DFG, EI841/4-1 and EI841/6-1), as well as two travel grants (BES-2011-047645 and EEBB-I-13-07490) awarded by the Ministry of Economy and Competitiveness of the Government of Spain to M. Hofmann.

Financiación

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad del Gobierno de España, con un beca del Programa de Formación de Personal Investigador FPI BES-2011-047645, y dos estancias de movilidad (BES-2011-047645 y EEBB-I-13-07490), concedidas a M. Hofmann. Además este trabajo se ha realizado en el marco de los proyectos del plan nacional (Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia) CGL 2010-16103 y CGL2013-42462-P, concedidos a M. Barluenga. Además, ha recibido financiación de la Fundación Alemana para la ciencia, a través de los proyectos DFG, EI841/4-1 y EI841/6-1 concedidos a C. Eizaguirre.

List of Abbreviations

AP Cell ..... **A**ntigen **P**resenting **C**ell
BAC ..... **B**acterial **A**rtificial **C**hromosome
B cell ..... **B** lymphocyte
BD ..... **B**enthic-**D**eep (species habitat)
bp ..... **B**ase **p**air(s) referring to nucleotides
BS ..... **B**enthic-**S**hallow (species habitat)
CD ..... **C**luster of **D**ifferentiation (cell surface molecules)
$CD4^+$ ..... **C**luster of **D**ifferentiation **4** (found on the surface of immune cells)
$CD8^+$ ..... **C**luster of **D**ifferentiation **8** (found on the surface of immune cells)
TCR ..... **T**-**C**ell **R**eceptor
cDNA ..... **c**omplementary **DNA** (synthesized from RNA)
CVA ..... **C**anonical **V**ariate **A**nalyses
CV ..... **C**anonical **V**ariates
^{13}C ..... Stable (non-radioactive) isotope of the chemical element **Carbon**
gDNA ..... genomic DNA
d_N ..... **N**onsynonymous substitution rate
DNA ..... **D**eoxyribo**n**ucleic **a**cid
d_S ..... **S**ynonymous substitution rate
GMQE ..... **G**lobal **M**odel **Q**uality **E**stimation
L ..... **L**imnetic (species habitat)
LP ..... **L**eading **P**eptide (determines the destination of newly formed proteins)
LPJ ..... **L**ower **P**haryngeal **J**aws (triangular bone with teeth used to process food)
LRT ..... **L**ikelihood **R**atio **T**est
M ..... **M**ahalanobis distance
Mb ..... **M**ega **b**ase pair(s) (106 base pairs)
MHC ..... **M**ajor **H**istocompatibility **C**omplex
mRNA ..... **m**essenger **R**NA
^{15}N ..... Stable (non-radioactive) isotope of the chemical element **Nitrogen**
PBR ..... **P**eptide **B**inding **R**egion
PC ..... **P**rincipal **C**omponents

PCA..... **P**rincipal **C**omponent **A**nalysis
PCR..... **P**olymerase **C**hain **R**eaction
PDB..... **P**rotein **D**ata **B**ank
RADseq.....**R**estriction **S**ite-**A**ssociated DNA **seq**uencing
SD .....**S**tandard **D**eviation
SMRT ..... **S**ingle **M**olecule **R**eal-**T**ime (sequencing technology)
SNP ..... **S**ingle **N**ucleotide **P**olymorphism
T cell ..... **T** lymphocyte
TSP..... **T**rans-**S**pecies **P**olymorphism

ABSTRACT

Since Darwin first proposed it in 1859, natural selection has been considered the main driver of diversification. Yet for the most part, speciation has been mainly viewed as a result of random processes, where genetic drift plays an important role. Only in the last decades has the consensus view of speciation shifted towards identifying the mechanisms responsible for the evolution of reproductive isolation, especially when gene flow is present during the process (*i.e.* sympatric speciation). This shift has resulted in a renewed recognition of the role of selection as a powerful cause of speciation, especially ecologically driven selection such as host-parasite interactions. Cichlid fish are the most species-rich families in vertebrates, distributed in rivers and lakes of Africa, Madagascar, India, and Southern and Central America. Cichlids represent a model system for the study of evolution due to their striking cases of adaptive radiation, morphological diversity, and frequent occurrence of evolutionary parallelism. Neotropical crater lake cichlids are especially ideal as they display one of the most convincing examples of sympatric speciation. The Major Histocompatibility Complex (MHC) is the key component of the adaptive immune system of all vertebrates and consists of the most polymorphic genes known to date. Due to this complexity, however, MHC remains to be characterized in many species including any Neotropical cichlid fish. In this work, we characterized the genes of MHC class IIB chain of the Midas cichlid species complex (*Amphilophus* cf. *citrinellus*). We designed new specific primers for PCR amplification, characterized the intron-exon conformation, and evaluated allele expression. We then created models of protein structure homology for each allele, and identified their key structural components. Using phylogenetic analysis, we identified one group of putative non-classical MHC and eleven groups of classical MHC which are under positive selection. We used Next Generation Sequencing to sequence the antigen binding site (exon 2) of MHCIIB and describe the population level variation among six Nicaraguan lakes. We report a striking case of parallel speciation in two of the volcanic crater lakes (Xiloá and Apoyo) where at least three ecologically and morphologically equivalent species of Midas cichlids have evolved independently in each of two lakes. Moreover, we conclude the speciation events follow the exact same succession in the two independent radiations. This case of parallel evolution involves overall body shape as well as other eco-morphological adaptations such as the structure of the lower pharyngeal jaws (LPJ) and diet; however, the pattern of MHC divergence is not as clear. Even though natural selection through ecological specialization to newly available environments is the likely force driving sequential parallel ecological speciation, the role of host-parasite interaction in this system is yet unclear. Overall, the Midas cichlid has one of the most diverse repertoires of MHC class IIB genes known, with extensive trans-species polymorphism, which could serve as a powerful tool to further elucidate the process of divergent radiations, colonization and speciation in sympatry.

RESUMEN

Desde que Darwin lo propuso por primera vez en 1859, la selección natural ha sido considerada como el principal motor de la diversificación. Sin embargo, a pesar de ello, la especiación se ha considerado como el resultado de procesos aleatorios, en los que la deriva genética juega un papel importante. Sólo en las últimas décadas se ha llegado al consenso de que hay que invertir los esfuerzos en la identificación de los mecanismos responsables de la evolución del aislamiento reproductivo, especialmente en casos donde el flujo de genes no tiene barreras físicas (i.e., la especiación simpátrica). Este cambio ha dado lugar a un renovado reconocimiento del papel de la selección como una poderosa fuerza de especiación, especialmente en la interacción parasito-hospedador. Los peces cíclidos son la familia más rica de vertebrados, se distribuyen en los ríos y lagos de África, Madagascar, India y sur y centro América. Los cíclidos representan un sistema ideal para el estudio de la evolución, debido a sus sorprendentes ejemplos de radiación adaptativa, su diversidad morfológica, y varias ocurrencias de paralelismo evolutivo. Los Cíclidos Neotropicales de lagunas crátericas son idóneos para estudiar los procesos de evolución ya que muestran uno de los ejemplos más convincentes de especiación simpátrica. El Complejo Mayor de Histocompatibilidad (MHC) es el componente clave del sistema inmune adaptativo de todos los vertebrados, y consiste en los genes más polimórficos conocidos. Debido a su complejidad el MHC aún no ha sido caracterizado en muchas especies, incluyendo los cíclidos Neotropicales. En este trabajo caracterizamos los genes del MHC de clase IIB del complejo del cíclido de Midas (*Amphilophus* cf. *citrinellus*). Para ello diseñamos cebadores específicos, y con ellos caracterizamos la conformación de intrones y exones, y la expresión alélica. Creamos modelos 3D de estructura molecular para cada alelo de MHC e identificamos sus claves componentes estructurales. Con análisis filogenéticos identificamos un grupo de alelos que parece no clásicos, y once grupos de MHC clásico. Con secuenciación masiva hicimos un análisis de MHC a nivel poblacional. Presentamos un caso de especiación en paralelo en dos lagunas cratéricas (Xiloá y Apoyo), donde al menos tres especies de cíclidos ecológica y morfológicamente equivalentes han evolucionado de forma independiente en cada una de ellas. Además, concluimos que los eventos de especiación siguen la misma sucesión en las dos radiaciones. En este caso de evolución en paralelo, se ven cambios en varias adaptaciones eco-morfológicas así como en su ecología. Sin embargo, el patrón de divergencia de MHC no es tan claro. A pesar de que la selección natural a través de la especialización ecológica es el mecanismo probable de especiación en paralelo, el papel de la interacción parásito-hospedador en este caso todavía no está claro. En general, los cíclidos de Midas tiene uno de los más diversos repertorios de genes de MHC II conocidos, y demuestran amplio polimorfismo trans-especie que podría servir como una herramienta poderosa para estudiar el proceso de radiaciones divergente, la colonización y especiación en simpatía.

GENERAL INTRODUCTION

Speciation

The diversity of life is one of the most striking aspects of our planet, with up to a trillion microbial species and several million eukaryotes it demonstrates an immense array in diversity and quantity of species (Mora et al. 2011). The origin and evolution of new species is one of the fundamental questions in biological research ever since Darwin proposed this idea (Darwin 1859). A central problem in biology is the elucidation of the mechanisms underlying the emergence of novel species. Ecological selection, the process by which barriers to gene flow evolve between populations as a result of divergent selection on ecological traits, has been increasingly emphasized in studies of speciation as a major driving mechanism (Schluter 2001). It is well established that ecologically relevant traits can evolve through natural selection, in particular through resource competition, followed by adaptation to different ecological niches, and ultimately by reproductive isolation (Arnegard et al. 2014). Theoretical models of ecological speciation propose resource partitioning as the most plausible initial step for speciation (Barluenga et al. 2006; Coyne 2007; Gavrilets et al. 2007). According to these models, resource partitioning could lead to assortative mating as byproduct and eventually evolution along separate trajectories until speciation is complete. Trying to determine which isolating barriers were important in a given speciation event is not easy since isolating barriers continue to accumulate after gene flow between taxa has stopped (Via 2009) To discriminate which barrier/s caused the isolation and which ones maintain the separation is not an easy task. Also over time, the relevance of different isolating barriers might change. Therefore, to fully understand the causes that isolate species it is ideal to focus on species that have only recently diverged.

Cichlid fish

Cichlid fishes are model species in evolutionary biology with over 2000 described species distributed across the tropics (Eschmeyer and Fong 2016). The extreme cichlid diversity in the Great African lakes in the rift valley has evolved through explosive speciation events, and is considered a textbook example of particularly fast adaptive radiations. Cichlids comprise some of the most diverse extant animal radiations, and provide a unique system to test predictions of speciation and adaptive radiation theory (e.g., Seehausen 2006; Takahashi and Koblmüller 2011; Burress 2015). Cichlids have been studied extensively in the context of evolution (Seehausen 2006), as they demonstrate repeated cases of radiation (Brawand et al. 2014), convincing cases of sympatric speciation (Schliewen and Klee 2004; Barluenga et al. 2006; Getz et al. 2016), and paralleled evolution (Manousaki et al. 2013). The Midas cichlid species complex inhabits several small crater lakes, as well as two large lakes in Nicaragua. Midas cichlids have recently colonized several crater lakes in Nicaragua where they have evolved, and are still evolving, into several

incipient radiations (Wilson et al. 2000; Barluenga and Meyer 2004; Kautt et al. 2016). This system offers a natural experiment that might aid to understanding the causes that isolate species. Cichlids inhabit multiple habitats where they are host to a large array of parasites (Vanhove et al. 2016), also making them ideal to study host-parasite interactions and speciation related to immune-gene related mate choice (Blais et al. 2007).

Host-parasite interaction and speciation

The antagonistic relationship between organisms, where one species is the parasite (bacteria, protozoans, metazoans, fungi or viruses that live at the expense of and benefits from the other species) and the other the host, can have a strong ecological effect on the evolution of both parties (Decaestecker et al. 2013). Parasitism is a common life strategy for many organisms, with estimates of up to 50% (Dobson et al. 2008) of existing organisms being parasitic. Unlike predators, parasites usually do not kill their host, are generally much smaller, and often live in or on their host for an extended period of time. Parasites can reduce the host's biological fitness by general pathology (Teifke et al. 2007), affect mate choice (Kennedy et al. 1987), produce reproductive hindrances (Galdino et al. 2014), or alter behavior (Barber and Wright 2006). In response to parasite pressure, hosts evolve to mount a better immune reaction to parasites, (Legendre et al. 2002). In this way hosts are subjected to parasite-mediated selection.

The role of MHC in the immune system

The immune system is a network of cells and molecular mediators which are essential in every step of all types of immune responses that work together to defend the body against attacks by "foreign" invaders such as parasites. The *innate immune system* involves white blood cells, such as macrophages and neutrophils which generically recognize parasites and destroy them by phagocytosis. This response affords a non-specific defense and does not change over time due to experiences (Janeway et al. 2005). On the other hand, the *adaptive immune system* has the capacity to recognize specific parasites and thereby responds with a defense that is adapted for that particular parasite and it establishes a memory that allows for a quicker response upon re-encounter. For the adaptive immune system to attack an invading parasite without harming the host it must be able to discriminate between the body's own tissue "self" and foreign tissue "non-self" derived from the parasite. This is done with the major histocompatibility complex (MHC), which are a set of cell surface molecules present in all jawed vertebrates (Hughes and Yeager 1998) and presumed absent in all non-vertebrates. The main function of classical MHC molecules is to bind to endogenous and exogenous self and non-self peptides and present them on the cell surface for recognition by T-cells (Trowsdale 1993). Non-classical MHC molecules have distinctive expression patterns and lack of polymorphism in the peptide binding region (Harstad et al. 2008), they are structurally

similar but functionally distinct from classical MHC molecules (Figueroa and Klein 1986). Classical MHC class I molecules are expressed on all nucleated cells, displaying peptides derived from inside the cell (endogenous). MHC class I present antigens, such as virus surface proteins or self-derived proteins that are processed into peptides by the proteasome. Peptides are then transported into the endoplasmic reticulum (ER), where they are loaded into the groove of the MHC class I complex. MHC class I complexes present antigens on the cell surface to cytotoxic $CD8^+$ T cells, T-cells that signal the infected cell to die via apoptosis while tolerating uninfected cells that display only normal self-peptides on their MHC molecules (Yewdell et al. 2003). Classical MHC class II molecules are usually only expressed on professional antigen-presenting cells which engulf intercellular (exogenous) antigens and process them into peptides by endolysosomal enzymes. These peptides are loaded into the groove of the MHC class II complex by displacing the class II-associated invariant chain peptide (CLIP). The loaded MHC class II molecule is then transported in an exocytic vesicle to be presented at the cell surface to T-cells which in response activates that cell to proliferate (see Figure 1). This sets off a cascade of immune defenses including the production of specific antibodies against that specific antigen (Trowsdale 1993). On the other hand, non-classical MHC class II genes are thought to serve as chaperones and accessory proteins assisting with folding, transport, and antigenic peptide loading (Yeager and Hughes 1999; Alfonso and Karlsson 2000; Neefjes et al. 2011).

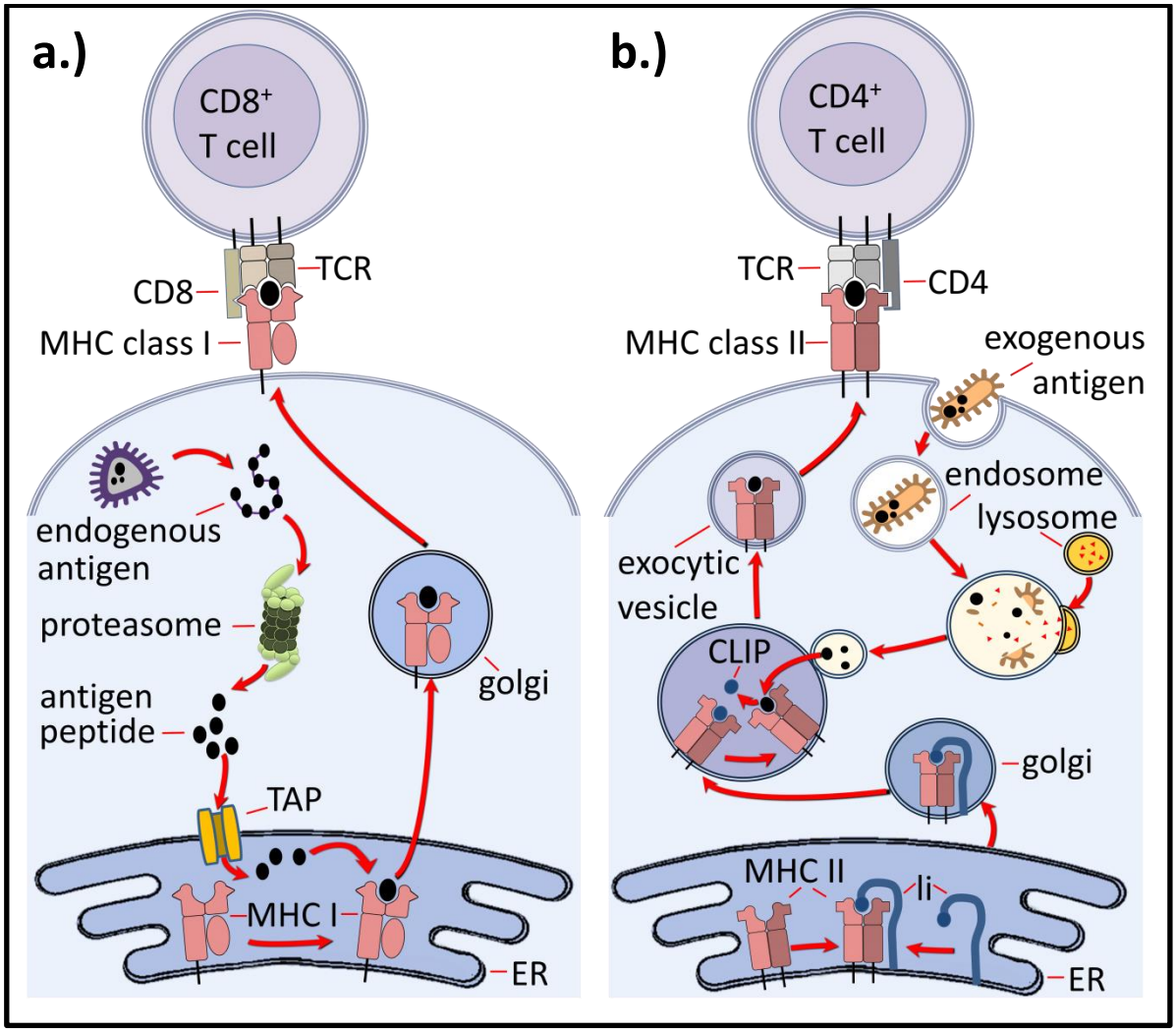


Figure 1: Figure modified from Kobayashi and van den Elsen (2012). The MHC class I and MHC class II antigen-presentation pathways. **a.)** MHC class I present intracellular (cytosolic) antigens on the cell surface to $CD8^+$ T cells. **b.)** MHC class II molecules are loaded with antigens from extracellular sources then transported in an exocytic vesicle to be presented at the cell surface to $CD4^+$ T cells for recognition.

MHC molecular structure

MHC class I molecules are composed of one glycoprotein α-chain that has three domains termed α_1, α_2, and α_3. The α_1-domain is attached to a β_2-microglobulin molecule (Figure 2). The α_3 is the transmembrane domain anchoring the MHC class I molecule to the cell surface membrane (Penn and Ilmonen 2005). Disulfide bonds (ss-bonds) give the α_2, and α_3-domains and the β_2-microglobulin rigidity. MHC class II is composed of two glycoprotein chains that form a heterodimeric structure, these are the α-chain and β-chain, each having two domains, termed α_1 and α_2, and β_1 and β_2 respectively (Penn and Ilmonen 2005).

Each chain has a transmembrane region that anchors the MHC class II molecule to the cell surface membrane (Figure 2). The β_1 and β_2 and α_2 are covalently bound by ss-bonds which give the molecule rigidity (Janeway et al. 2005). In both MHC class I and MHC class II the peptide binds to the peptide-binding groove aka the peptide-binding region to be presented to T-cells for recognition. In MHCII the peptide binding region is mainly formed by the α_1 and β_1 domain.

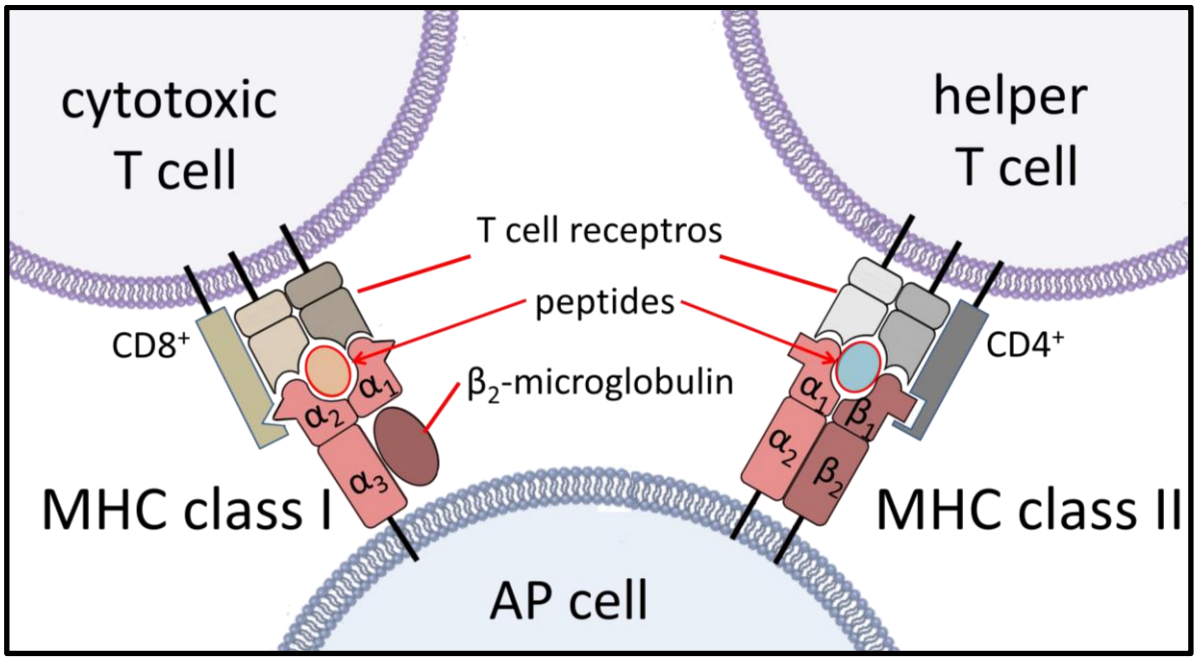


Figure 2: MHC class I molecule (left) composed of an α-chain that has three domains termed α_1, α_2, and α_3, and an associated β_2-microglobulin molecule. MHC class II molecule (right) composed of two chains, α and β, each having two domains, termed α_1 and α_2, and β_1 and β_2 respectively. The peptide binding region is located on the membrane-distal domain and is where the peptides (blue and orange circled in red) are held. The cluster of differentiation (CD) is a co-receptor that binds to MHC I ($CD8^+$, brown) or MHC II ($CD4^+$, dark gray) ensuring the correct immune response.

MHC polymorphism

MHC genes influence many important traits, including resistance to infectious diseases (Trowsdale 2011), mating preferences and speciation (Havlicek and Roberts 2009; Eizaguirre et al. 2009b). MHC genes are the most polymorphic genes in all vertebrates and the multiple varying copies found in each individual are co-dominantly expressed (Hviid et al. 1998). Classical MHC is so polymorphic that most individuals are likely to be heterozygous at each locus (Oliver et al. 2009). In any mating pair, for every MHC locus, four possible combinations of haplotypes can be found in the offspring; thus siblings are also likely to differ in the MHC alleles they express. In addition to the high numbers of alleles at the single locus, MHC has gone through many rounds of gene duplication (Kulski et al. 2002; Reusch et al. 2004; Bontrop 2006;

Bollmer et al. 2010) making it polygenic. Polygeny is the presence of several different related genes with similar functions that typically arise by gene duplication (Butler et al. 2011). Polymorphism and polygeny combine to produce the high diversity of classical MHC molecules seen both within an individual as well as in the entire population (Janeway et al. 2001).

Each MHC molecule can present only the peptides that match its particular peptide-binding region (color dots in Figure #3), and each individual has a set quantity of MHC alleles that bind to a limited range of peptides. However, at a population level, the enormous diversity of MHC molecules (double black line, figure #3) ensures that at least some individuals will have MHC variants that protect them against any parasite, this way no parasite can exterminate the whole population, even though it may be fatal in some or even most individuals. This ensures the survival of the population or a species despite the loss of some individuals to a particular parasite (Relle and Schwarting 2012).

It would seem advantageous for an individual to maximize coverage by expressing limitless different MHC molecules to resist as many parasites as possible (Woelfing et al. 2009). However, T-cells do not actually distinguish between self and non-self peptides. Self-tolerance of the adaptive immune system is ensured mainly by eliminating T-cell precursors with receptors specific to self-peptide-MHC complexes (Janeway et al. 2001). In a process called positive selection, T-cell receptors are tested for general reactivity to the individual's MHC molecules. Only T-cells with receptors that are functional in the context of the individual's MHC molecules pass this checkpoint. Next, during negative selection, T-cells, whose receptors react strongly with self-peptide-MHC complexes that might cause autoimmune diseases are eliminated before maturation (Janeway et al. 2001). Therefore, if an individual has limitless MHC molecules (high intra-individual MHC diversity) it may result in a severe depletion of its mature T-cell repertoire due to negative selection. This would mean that even though endless peptides are presented on copious MHC molecules, most will not be recognized by T-cells that initiate the immune response. This will ultimately reduce immunocompetence (Vidovic & Matzinger 1988; Lawlor et al. 1990) or cause an autoimmune disorder (Woelfing et al. 2009). These opposed requirements mean that the optimal MHC diversity is a delicate balance between self-tolerance and immune response.

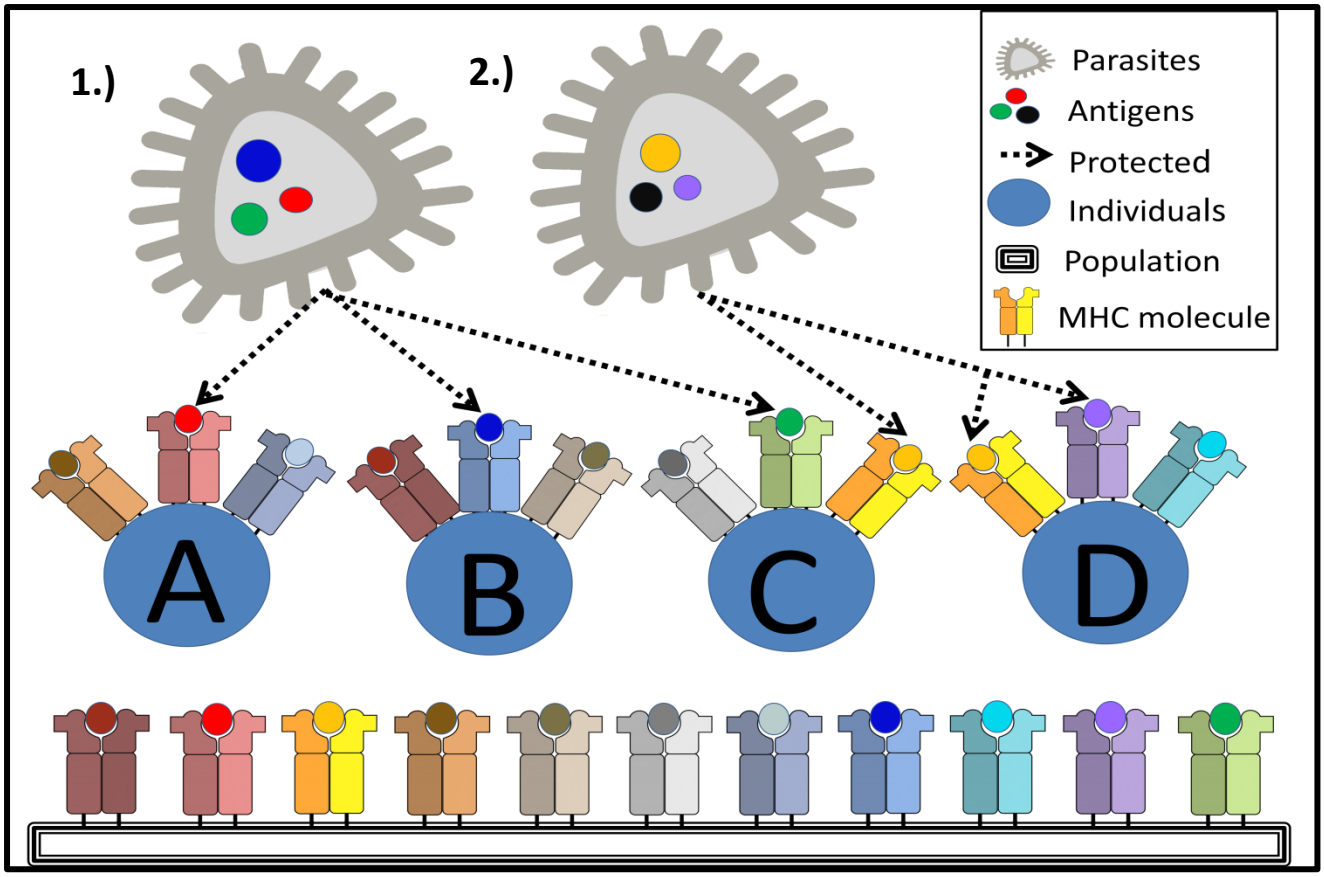


Figure #3: Protective effect of MHC polymorphism on populations (simplified scheme modified from Relle and Schwarting 2012). The two parasites (1, and 2), are depicted as having three antigens each (colored dots). The four blue circles (A, B, C, and D) represent individual hosts with a diversity of MHC molecules. The black double line represents the population with the total MHC diversity. The dotted lines indicate the MHC allele that offers protection from the parasite. In this situation, individual A, B, and C are protected by at least one MHC allele from Parasite 1, however individual D is vulnerable to this particular parasite. Individual A and B are vulnerable to parasite 2, while Individual C and D are protected.

MHC evolution and diversity

For most neutral polymorphic genes, the usual pattern is to have a very common allele and some rare variants. However, in the presence of balancing selection, several alleles are instead found at intermediate and similar allele frequencies (Ejsmond et al. 2010). Balancing selection actively maintains multiple alleles in the gene pool of a population at high frequencies, and for longer than expected from genetic drift alone, thus conserving genetic polymorphism. It is well accepted that MHC variation is maintained by balancing selection (Hedrick 1999), this being found in human populations (Meyer and Thomson 2001) as well as in other vertebrates (Ekblom et al. 2010; Consuegra et al. 2013; Niskanen et al. 2014). However,

the evolutionary mechanisms that act as driving forces of MHC diversity are still widely debated. A series of hypotheses have been proposed which could explain the maintenance of observed patterns of genetic variability at MHC loci in empirical studies (Spurgin and Richardson 2010). These hypotheses fall into two main categories: the disease-based mechanisms, and the reproductive-based mechanisms (Meyer and Thomson 2001; Bernatchez and Landry 2003; Piertney and Oliver 2005). They all rely on the notion of no single allele being absolutely most fit.

Overdominant selection hypothesis

This hypothesis is also referred to as the *heterozygote advantage hypothesis*. It was first proposed by Doherty and Zinkernagel (1975) and it assumes that all heterozygotes have equal and high fitness while all homozygotes have equal and low fitness. This entails that the phenotype of the heterozygote lies outside the phenotypic range of either homozygous parent (Klein and Figueroa 1986). However, the assumption of equal fitness for all heterozygotes (and all homozygotes) is not realistic and the overdominance model has been criticized as a non-realistic explanation of the driving force of balancing selection (De Boer et al. 2004). Moreover, this hypothesis has been expanded to consider that heterozygotes carrying more divergent allelic sequences have a greater ability to bind to a wider range of antigens, therefore resisting more pathogens (Takahata and Nei 1990).

Negative frequency-dependent selection also known as the *rare-allele advantage hypothesis*.

In this model, genotype fitness values are not fixed but change in proportion to allele frequencies in the population. It hypothesizes that common alleles are under greatest pathogenic pressure, driving positive selection of uncommon alleles (Takahata and Nei 1990; Tellier and Brown 2007). This means that parasite antigenicity will be selected to exploit the immune response of the most common host genotypes. This decreases the relative fitness of the common host genotypes and provides a selective advantage to new or rare MHC alleles. This leads to the cycling of fitness values of different genotypes in both hosts and parasite and results in the maintenance of the high genetic diversity of MHC alleles in the population. Theoretically, it has been shown that these cycling processes could maintain many alleles in the population over time (Borghans et al. 2004).

Fluctuating selection hypothesis

This hypothesis suggests spatial and temporal heterogeneity among parasites so that the selective advantage of different MHC genes fluctuates in time or space over the life time of individuals or over the geographical range of a population (Hill 1991). This model does not assume co-evolution between host and parasites as the determining factor for parasite fluctuations but allows for external factors to determine the distribution of parasites (Spurgin and Richardson 2010).

Moving target hypothesis

The *moving target hypothesis* by Penn & Potts (1999) and the *inbreeding avoidance hypothesis* (Potts and Wakeland 1990) assume that because MHC loci are often highly polymorphic, individuals that share MHC alleles are thus likely to be related. Considering this, they hypothesize that MHC-dependent mating preferences enhance parasite resistance in two ways. First, with MHC-disassortative mating, meaning that there is a reduction in the number of homozygotes and it is less likely that alleles are lost because of drift and therefore they can be maintained in the population for a longer time (Hedrick 1992). Mate preferences could increase the fitness of choosy parents because a disproportionate number of offspring would be MHC heterozygotes. Secondly, mating preferences provide a mechanism to 'keep up' the molecular arms race of co-evolution with parasites related to the genetic benefits of parasite resistance-based sexual selection favoring MHC polymorphism (Penn and Potts 1999).

Polymorphism

A gene is considered to be polymorphic if more than one allele occupies that gene's locus within a population. Most genes that are considered 'polymorphic' have only a few alleles that differ at few sites or have only single nucleotide polymorphisms (*e.g.* Wielinga et al. 2015). However, in the case of MHC, most species have many alleles that commonly differ by more than 20 substitutions (Rammensee et al. 1997). The high degree of polymorphism in MHC genes could be explained by a high mutation rate. However, Hughes and Nei (1988 and 1989) showed that non-synonymous (d_N) substitution rates (meaning nucleotide mutation that alters the amino acid sequence of a protein) only exceed the synonymous (d_S) substitution rates (substitution of one base for another such that the produced amino acid sequence is not modified) at the gene region coding for the peptide-binding region of the molecule. They also showed that the d_S value did not differ from other genes, thus reflecting a normal mutation rate. Hence, MHC polymorphism is specifically related to peptide binding and not to the genes in general. A significantly higher $d_N/d_S > 1$ ratio that indicates positive selection (on the codons/sites) requires a considerably long period of time of mutation and selection. Theoretically it is possible to achieve d_N/d_S ratio significance after several thousand generations if the population size is large and selection strong, however, it takes even longer for the d_N/d_S ratio to lose its acquired significance if selection is lost (Garrigan and Hedrick 2003). Therefore, using the d_N/d_S ratio as an estimator of positive selection reveals little about the current selection pressure at the population level, although it provides evidence that genes of a certain species have been under positive selection during a considerable part of its history or even predating the species formation. This supports the theory that MHC alleles likely diverged long ago, meaning that the diversity of MHC alleles in some species is ancient, even predating the origin of the species itself. This phenomenon is termed *trans-species polymorphisms* (TSP).

Trans-species polymorphism

TSP is the occurrence of shared orthologous allelic lineages among different species (Klein et al. 1998; Ottová et al. 2005). TSP was first described by Klein (1987) referring to alleles that are maintained over long evolutionary times beyond species formation and thus are shared among species. TSP implies that a lot of the variation that we see in MHC genes today is derived from ancestral species and not generated in each species by convergent evolution following speciation (Klein 1987; Figueroa et al. 1988; Lenz et al. 2013). MHC alleles will consequently cluster according to allelic lineages and not according to species phylogeny. The trans-species evolution theory is demonstrated by unexpectedly large genetic distances between MHC alleles of one species, that is the number of nucleotide differences between them is often greater than expected for alleles within a species (Klein et al. 1998).

OBJECTIVES

In this work, I aim to understand alternative mechanisms underlying rapid speciation. Different environmental pressures have been proposed to contribute to the rapid diversification of populations. Here, I evaluated the hypothesis that host-parasite interaction triggers sympatric speciation in crater lake cichlids. To this end, I characterize the genetic divergence in the genes responsible for the initiation of the adaptive immune response in vertebrates, MHC, in several incipient species, sympatric and allopatric, belonging to the Neotropical Midas cichlid complex.

Specific research aims

The first objective of this study was to characterize MHC genes in the Midas cichlid species complex. These genes have never been characterized in any Neotropical cichlids before, and therefore a *de novo* approach had to be taken. In **Chapter I,** I develop primers specific for the amplification of MHC class IIB genes in the Midas cichlid. I elucidate the structure of these genes, the number of introns and exons, and the variability of alleles for this species complex. I used phylogenetic analyses to establish the relationship between alleles and estimate the selection pressures they are subjected to.

The second objective of this work was to describe the diversity of MHC class IIB alleles at a population level in several incipient species of the Midas complex, living in allopatry (in different lakes), and in sympatry (within the same lake). **Chapter II,** describes population-level sequencing of MHC class IIB alleles of the Midas cichlid species complex in Nicaragua applying Next Generation methods. The goal was to determine if all species had the same diversity of alleles, if they shared alleles according to shared geographic habitat, or shared trophic habits, and if there are alleles that are species-specific or lake-specific.

The third objective of this thesis was to determine if MHC has evolved in parallel in two different lakes . To this end in **Chapter III** I do an in-depth analysis of the alleles of three species that occupy the same zones in each of these lakes, the littoral-benthic, the deep-benthic, and the open-limnetic zones of lake Apoyo, and Xiloá. These lakes display repeated evolution of trios of species, which are morphologically equivalent in pairs, ecologically equivalent in pairs, but genetically closer to each other within lakes despite reproductive isolation

.

CHAPTER I

Molecular characterization of MHC class IIB genes of sympatric Neotropical cichlids

Published in BMC Genetics 18:15. 2017

Abstract

Background: The Major Histocompatibility Complex (MHC) is a key component of the adaptive immune system of all vertebrates and consists of the most polymorphic genes known to date. Due to this complexity, however, MHC remains to be characterized in many species including any Neotropical cichlid fish. Neotropical crater lake cichlids are ideal models to study evolutionary processes as they display one of the most convincing examples of sympatric and repeated parallel radiation events within and among isolated crater lakes.

Results: Here, we characterized the genes of MHC class IIB chain of the Midas cichlid species complex (*Amphilophus* cf. *citrinellus*) including fish from five lakes in Nicaragua. We designed 19 new specific primers anchored in a stepwise fashion in order to detect all alleles present. We obtained 866 genomic DNA (gDNA) sequences from thirteen individuals and 756 additional sequences from complementary DNA (cDNA) of seven of those individuals. We identified 69 distinct alleles with up to 25 alleles per individual. We also found considerable intron length variation and mismatches of alleles detected in cDNA and gDNA suggesting that some loci have undergone pseudogenization. Lastly, we created a model of protein structure homology for each allele and identified their key structural components.

Conclusions: Overall, the Midas cichlid has one of the most diverse repertoires of MHC class IIB genes known, which could serve as a powerful tool to elucidate the process of divergent radiations, colonization and speciation in sympatry.

Introduction

The Major Histocompatibility Complex (MHC) is a key component of the adaptive immune system of all jawed vertebrates (Dixon et al. 1995; Pastoret et al. 1998). The function of the MHC molecules is to present short self and non-self peptides often derived from parasites and parasites for recognition by T-lymphocytes (Janeway et al. 2005). This sets off the cascade of targeted immune defenses against those specific parasites and parasites. MHC also plays a role in establishing a memory to rapidly eliminate those agents in case of future encounters (Janeway et al. 2005). MHC molecules are encoded by the most polymorphic genes in all jawed vertebrates, and most species have a different number of loci that are co-dominantly expressed [*e.g.* 4,5]. There are two classical antigen-presenting MHC molecules, MHC class I, that is expressed on all nucleated cells and elicits a response against intracellular parasites, and MHC class II, that is only expressed on antigen-presenting cells (macrophages, B-cells, and dendritic cells), which actively engulf and process inter-cellular parasites (Janeway et al. 2005). Here, we focus on MHC class II which is composed of two chains, α, and β, which together form the peptide-binding groove (Madden 1995). The peptide-binding region of the β chain is the most polymorphic and hence the most studied region of the MHC. In general, the MHC IIB region is divided into 5 to 6 exons with varying intron lengths depending on species and haplotypes (Ono et al. 1993b; Schwaiger et al. 1993; Reusch and Langefors 2005).

The highly polymorphic multigene nature of MHC causes some technical difficulties when trying to simultaneously detect all alleles, particularly those that are rare in the target population. Cloning and Sanger sequencing have associated PCR-based errors and PCR amplification biases (Lenz and Becker 2008; Cummings et al. 2010; Burri et al. 2014), making accurate amplification a laborious and costly process. Next generation sequencing technologies have, to some extent, facilitated population level studies of MHC, although those new techniques tend to overestimate allelic diversity (Babik 2010; Lighten et al. 2014b). Overcoming those challenges allows the use of MHC as a powerful tool to study biodiversity (Sommer 2005; Eizaguirre et al. 2009a), disease dynamics (Trowsdale 2011), evolutionary processes (Klein et al. 2007; Sepil et al. 2013a), and even to estimate the number of founders of a population (Vincek et al. 1997).

Cichlid fish are excellent model systems to study evolutionary processes since they demonstrate some of the most extreme examples of explosive adaptive radiations [*e.g.* 29–32]. They are some of the most species-rich families of freshwater fishes worldwide, and their hotspots of diversification are the great lakes of East Africa. They are also present in Central and South America (Barlow and Munsey 1976; Kullander 2003). Particularly, the Neotropical Midas cichlid species complex (*Amphilophus spp.*) is a valuable model system for the study of recent speciation (Barluenga and Meyer 2004; Gavrilets et al.

2007; Elmer et al. 2010b). This group not only comprises one of the most compelling examples of sympatric speciation (Barluenga et al. 2006), but also recent independent colonization events and *in situ* rapid diversification (Barluenga and Meyer 2010), which makes it an excellent natural experiment of adaptation and incipient speciation (Elmer et al. 2010b).

Many studies have attributed cichlid´s rapid speciation events to various factors, including phenotypic plasticity (Muschick et al. 2011), reproductive behavior and local adaptation (Kornfield and Smith 2000; Thibert-Plante and Gavrilets 2013), and even genomic processes (Elmer et al. 2010a; Kautt et al. 2012; Santos et al. 2014). It has been suggested that the mechanism of adaptive speciation in general, and in sympatry in particular, may result from a pleiotropic role of the MHC in co-evolutionary dynamics of local host-parasites and odor-mediated mate choice ultimately leading to reproductive isolation (Summers et al. 2003; Blais et al. 2007; Eizaguirre et al. 2009a). Here, we characterized the β chain of the MHC class II in the Neotropical Midas cichlid species complex to establish the baseline for evaluating the role of parasites and immune system in sympatric speciation.

A striking characteristic of MHC polymorphism is the occurrence of similar alleles in related species, known as trans-species polymorphism (TSP) (Klein 1980). This similarity might have arisen by convergence (Xu et al. 2008; Lenz et al. 2013), although a more commonly accepted idea is that this polymorphism is maintained, mostly by balancing selection, beyond the species formation (Klein et al. 1998). This polymorphism transmitted through several speciation phases can be a useful tool to study speciation itself (Klein et al. 1998). TSP has been found to occur across related species of reptiles (Bryja et al. 2006; Stiebens et al. 2013), amphibians (Kiemnec-Tyburczy et al. 2010), mammals (Eimes et al. 2015), birds (Kikkawa et al. 2009), and fish (Ottová et al. 2005). In this study, we also characterize events of TSP.

There is some knowledge about MHC diversity patterns of cichlid fish (Klein et al. 1993; Málaga-Trillo et al. 1998; Figueroa et al. 2000; Murray et al. 2000; Blais et al. 2007), but this comes exclusively from African species. Some studies have focused on the diversity of MHC class I in cichlids from Lake Victoria, finding many common alleles across species (Sato et al. 1997). Other studies have found a high diversity of MHC class IIB alleles in different species of Lake Malawi cichlids (Klein et al. 1993; Ono et al. 1993a). A population genetic analysis on MHC class IIB of Lake Malawi cichlids even suggested that adaptive divergence at this locus could be linked to speciation in cichlids (Summers et al. 2003).

Old and New World cichlids have been geographically separated for a very long time (Albert and Reis 2011; Friedman et al. 2013; Matschiner et al. 2016), MHC evolution in Neotropical cichlids is likely to have followed its own evolutionary trajectory. Therefore, MHC has to be characterized *de novo* in the Midas cichlid in order to understand its role in their adaptation and speciation. We first sequenced exons 1-6 of the MHC IIB and described intron and exon conformation as well as most intron length variability.

Then we used both genomic (gDNA) and expressed transcripts (mRNA) to characterize the allelic diversity existing in exon 2 – that which encodes for the peptide binding groove. We then tested for various modes of evolution of the MHC and modeled the tertiary structure of each detected allele to identify the structural components of the MHC molecules.

Methods

Sampling, DNA and RNA isolation

Sampling of Midas cichlid fish took place in several Nicaraguan lakes (Fig. 1). Adult fish were captured using gill nets (collection permit number 001-012012), anesthetized with MS 222 following standard procedures and euthanized on ice before processing. Fin tissues of 13 randomly selected individuals were preserved in 100 % ethanol at 4ºC. Those 13 samples represent a good portion of the diversity of this species complex (Fig. 1, Supplementary Table 1). Additional spleen tissue samples of 7 of these individuals were preserved in RNAlater® (Qiagen, Hilden, Germany) and stored at -80ºC.
Total genomic DNA (gDNA) was extracted using DNeasy spin columns for Blood and Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol, with the addition of RNAse. DNA was quantified using Nanodrop 1000 (ThermoFisher Scientific, Bonn) and standardized to a concentration of 20 ng / µl. RNA was extracted with Invitrap Spin tissue RNA mini kit® (Berlin, Germany) and the reverse transcription performed with the QuantiTect® Reverse Transcription kit (Qiagen, Hilden, Germany).

Figure I.1. Map of the Pacific coast of Nicaragua (Central America) with the great lakes and several crater lakes where samples were collected. Samples belonged to the four species *Amphilophus citrinellus, A. labiatus, A. amarillo* and *A. xiloaensis*.

Primers Design

Firstly, to characterize the allelic diversity in the exon 2 of the MHC IIB gene in the Midas cichlid, we retrieved MHC sequences from GenBank choosing different sequences from related fish species with well-characterized MHC genes. We aligned the sequences (see Supplementary Table 2) and designed a reverse primer (MHC_Rev3 "GATCTGTTTGGGGTAGAAGTCG") located in the most conserved region in the middle of exon 3. We did a PCR by pairing this reverse primer with two published forward primers designed for sticklebacks located in conserved upstream regions of exon 2 (SatoGa11_mod1[55], GAIIEx2startF [56]). Using the resulting sequences we designed 14 additional primers in a stepwise manner. We designed new primers in adjacent regions of the sequence, considered sets of new amplicons, and designed additional primers on the new sequences maximizing allele amplification (Table 1). We paired all forward and reverse primers. Additionally, we specifically designed 4 primers (AcMHCIIF3, AcMHCIIF4, AcMHCIIF5, and AcMHCIIF6) to discriminate a group of particularly abundant alleles that were not very variable (later referred to as Group I), which could hinder the detection of rarer alleles.

Table I.1. Primer combinations used for amplification of MHC IIB

Forward Primer	Reverse Primer	T_A	Fragment
SatoGa11_mod1	**MHC Rev3**	59°C	E2 - E3
AACTCCACKGAKCTGAAGRAC	GATCTGTTTGGGGTAGAAGTCG		
GAIIEx2startF	**MHC Rev3**	55°C	E2 - E3
GTCTTTAACTCCACGGAGCTGAAGG	GATCTGTTTGGGGTAGAAGTCG		
GAIIEx2startF	**AcMHCIIBR1**	59°C	E2 - E3
GTCTTTAACTCCACGGAGCTGAAGG	GGRGTGAAGTCTGACTRATGG		
AaMHCIIBE1F1b	**MHC Rev3**	56°C	E1 - E3
ATGGCTYCATCCTTYMTC	GATCTGTTTGGGGTAGAAGTCG		
AaMHCIIBE1F1b	**AcMHCIIBR4**	59°C	E1 - E5
ATGGCTYCATCCTTYMTC	ACCCAGGATCAGTCCTGAGG		
AcMHCIIBF2 (excludes group I)	**AcMHCIIBR3** (excludes gop I**)**	64°C	E2 - E2/I2
TTAACTCCACKGAGCTGAASGACA	GAYGATGAAYCATAACTCACCTGAT		
AcMHCIIBF3 (only group I)	**AcMHCIIBR4**	46°C	E2/I2 - E5
TCAGGTGAGTYATGDTTCATC	ACCCAGGATCAGTCCTGAGG		
AcMHCIIBF3 (only group I)	**AcMHCIIBR5**	44°C	E2/I2 - R5
TCAGGTGAGTYATGDTTCATC	TTCCTCTTGTAGTAGATGAATCC		
AcMHCIIBF4 (excludes goup I)	**AcMHCIIBR4**	54°C	E2/I2 - E5
TCAGGTGAGTCTGTTTCTGTG	ACCCAGGATCAGTCCTGAGG		
AcMHCIIBF5 (excludes group I)	**MHC Rev3**	55°C	E2 - E3
CCACKGAGCTGAASGACATSGAG	GATCTGTTTGGGGTAGAAGTCG		
AcMHCIIBF5 (excludes group I)	**AcMHCIIBR4**	59°C	E2 - E5
CCACKGAGCTGAASGACATSGAG	ACCCAGGATCAGTCCTGAGG		
AcMHCIIBF7	**AcMHCIIBR9**	50°C	E2 - E2
CGAGTACGTTCGATCTTTGTATTGC	DCTGATTTAGTCAGAGCAGTCT		
AcMHCIIBF8	**AcMHCIIBR9**	55°C	E2 - E2
CGAGTWCATCARVTCTTACTAYTWC	DCTGATTTAGTCAGAGCAGTCT		
AcMHCIIBF9	**AcMHCIIBR8**	NA	I2 - I2
GAAACCTGTTCACAGCAGTCCCTC	CATGTGCTACATGCAACATATCA		
AcMHCIIBF6	**MHC Rev3**	55°C	E2 - E3
CCACTGAGCTGAASGACATSGAG	GATCTGTTTGGGGTAGAAGTCG		

Fragment indicates the region of the gene amplified: 'E' exon, 'I' intron, '/' exon-intron boundary. The primers AcMHCIIBF9 and AcMHCIIBR8 were used only to sequence through clones with long intron 2.

PCR Amplification and gel extraction

PCR amplifications were performed following the recommendations of Lenz and Becker (Lenz and Becker 2008) in order to reduce PCR artifacts, common in the amplification of multigene families. We used Taq Polymerase with no proof-reading capacity, extended elongation times, excess of primers to avoid incomplete amplicons acting as heteroduplex primers, and duplicated reactions. However, we did not reduce PCR cycles or do a reconditioning PCR since under those recommendations our bands were too weak for cloning. Each amplification was done in two independent reactions, each consisting of 2 µl 10x Dreamtaq Buffer, 1 µl dNTP's (10mM), 2 µl of each primer (5 pmol / µl), 0.2 µl Taq Polymerase (Dreamtaq®) and 2 µl of template DNA in a total volume of 20 µl. The thermal profile started with an initial denaturation step of 95°C for 3 min, followed by 30 cycles of denaturation at 94°C for 30 sec,

annealing at specific temperature for each primer pair (ranging from 44 - 64°C, Table 1) for 1 min, elongation at 72°C for 1 min, with a final elongation at 72°C for 10 min. The PCR reactions for each individual and primer pair were then pooled and loaded into a 2 % agarose gel and run at 40 V for 4 h. Gels were then stained with Ethidium bromide to visualize bands. The bands of interest were cut and extracted with NucleoSpin Gel and PCR Clean-up® (MACHEREY-NAGEL GmbH & Co. KG) for further cloning.

Cloning

PCR amplicons were cloned with the Qiagen PCR cloning Kit® (Qiagen, Hilden, Germany). Cloning followed the protocol described in Bracamonte *et al*. (Bracamonte et al. 2015). For clone screening, 1 µl of the denatured clones was used as a template for a PCR with the universal M13 primers: M13_Funi (5´ACGACGTTGTAAAACGACGGCCAG 3´), and M13_RP15 (5´TTCACACAGGAAACAGCTATGACC 3´). The reaction had a final volume of 10 µl, included 1 µl 10x Dreamtaq Buffer, 0.5 µl dNTPs (10 mM), 1µl of each primer (5 pmol / µl), 0.1 µl Taq Polymerase (Dreamtaq®), and ran using the following thermal profile: initial denaturing step at 95°C for 1 min, followed by 25 cycles of denaturing at 96°C for 10 sec, annealing at 50°C for 10 sec, elongation at 72°C for 1 min, with a final elongation at 72°C for 7 min. Two µl of this PCR product were then loaded in a 1 % agarose gel and run for 30 min at 90 V, and ultimately stained with Ethidium bromide to visualize bands. We sequenced the clones that were positive for the amplicon. We sequenced between 16 and 48 clones per amplification in order to detect rare alleles.

Sequencing

Cycle sequencing was done using the Big Dye Terminator v3.1 using the Cycle Sequencing Kit (Life Technologies) following the manufacturer's protocol scaled to 10 µl total reaction volume per sample. The product was then purified using 50 µl BigDye X Terminator™ Purification Kit® mix (Life technologies, Thermo Fisher Scientific Inc). Sequencing took place on an ABI 3730 Genetic Analyzer (Life Technologies). Even though MHC sequence variants may stem from different loci and therefore may be paralogs, we will refer to them as alleles.

Identifying and naming alleles

The taxonomic status of the species within the Midas cichlid complex is under considerable debate. Although some species have been described recently within this species complex, due to their very recent history (<50,000 years) we name our alleles under the generic name "Amci" for *Amphilophus* cf. *citrinellus*. Sequences were aligned using CODON-CODE ALIGNER® (Codoncode Corporation 2002-2009)

and analyzed with BIOEDIT v1 (Hall 1999). To avoid sequence artefacts, generally alleles were only considered true alleles if they were amplified in at least two independent PCR and cloning events (see Results for exceptions on this). We manually inspected the alignment and removed all sequences that were identified as chimeras by being partially identical to one allele group, and partially identical to another allele group and only existing in one PCR product. Sequences identified as chimeras were removed from all further analyses. A consensus was created from all identical sequences of different lengths and this was hereafter considered an allele. All alleles were aligned with well annotated published sequences (Lenz et al. 2009a) and checked for stop codons within exons to further rule out the presence of pseudogenes. Since there is no complete genome for this species, identifying the family and locus that each allele belonged to is not currently possible. Alleles were named following the allele nomenclature guidelines for MHC established by Klein *et al* (Klein et al. 1990). We looked for tandem repeats in all alleles using Tandem Repeat Finder v4.09 (Benson 1999).

We did a blast search of all alleles on the *Amphilophus citrinellus* draft genome shotgun sequencing project (BioProject PRJEB6974) (Elmer et al. 2014) to confirm the exon structure of the MHC IIB alleles we found and to confirm intron length variability.

Alleles were assigned to groups according to estimates of evolutionary divergence between sequences. Analyses were conducted using the Maximum Composite Likelihood model (Tamura et al. 2004) in MEGA v5.2 (Tamura et al. 2011). Codon positions included were 1^{st}, 2^{nd}, 3^{rd}, and noncoding. All positions containing gaps and missing data were eliminated. The final dataset had a total of 201 positions. We validated the resulting groups with a randomization analysis in R (Bunn and Korpela 2016). The mean pairwise distances within each group were compared to a null distribution generated by a random selection of equivalent alleles repeated 999 times to calculate the p-value. The results of this analysis were contrasted with the phylogenetic reconstruction of the allele relationships.

Phylogenetic and statistical analyses

In order to determine the phylogenetic relationships between alleles, we used exon 2 and 3 to construct a phylogenetic tree using Bayesian inference with BEAST v2.0 (Bouckaert et al. 2014). We found the most appropriate substitution model (HKY+G) and partitioning scheme according to the Bayesian Information Criterion implemented in PARTITIONFINDER (Lanfear et al. 2014). We specified the parameters in BEAUti v2.0 (Bouckaert et al. 2014) for the BEAST-run, and the MCMCs were run for 10^9 generations sampling every 100,000 trees. We used a strict clock model and a Yule speciation prior. We inspected the traces for convergence with TRACER 1.5 (http://tree.bio.ed.ac.uk/software/tracer/) and checked that they were higher than 200. The 10,000 resulting trees were summarized with TREEANNOTATOR v2.1.2 (http://beast.bio.ed.ac.uk/treeannotator) applying a 10 % burn-in. We depicted the phylogeny with the

corresponding posterior values of each node with FIGTREE v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

To further analyze the relationship between the alleles we constructed a neighbor-net network with SPLITSTREE4 (Huson and Bryant 2006). We calibrated the network by calculating the best fit model using MEGA and estimated the probability of invariable sites.

To evaluate TSP we gathered previously published sequences from several fish species. We selected 11 sequences of MHC IIB of a well-studied African cichlid, Nile tilapia (*Oreochromis niloticus*), from Sato *et al*. (Sato et al. 2012), and 9 sequences from a search at NCBI´s Genbank database with BLASTx from other cichlids and two other fish families: Sciaenidae and Sebastidae which had alleles closest to cichlid ones (Supplementary Table 3). We constructed a phylogeny with all the alleles (69 from the Midas cichlid and 20 from other species) using the same methodology and parameters as described above.

Tests of selection

In order to elucidate on the evolutionary history of MHC II B of Midas cichlid, to determine if different selective pressures have shaped the different domains of the molecule, and if selection has acted differently on the different groups of alleles, we performed selection tests, both by domain and on the entire sequence, as implemented in MEGA. We acknowledge the limitations of these tests given the fact that that we cannot allocate alleles to specific loci. Selection tests were performed on the groups of alleles established by sequence divergence. We tested separately the alleles of Group I since they appear to follow a different evolutionary pathway. We calculated rates of substitution (d_N and d_S) and tested for overall positive or purifying selection with a Z-test of selection for each domain separately. For alleles in Group I we applied the Pamilo-Bianchi-Li method with Kimura 2-parameter correction, and for the rest of alleles, we applied the Nei-Gojobori method with Jukes-Cantor correction, in accordance with their corresponding best-fit substitution model. Significance levels were estimated with 10,000 bootstrap replicates. We used the Nei-Gojobori method for calculating the "absolute" number of synonymous and non-synonymous sites since the Pamilo-Bianchi-Li method is not available for this test. We do not have the full sequence coverage for all alleles, and therefore we used a pairwise method that compares two sequences at a time and then averages over all possible comparisons. With this method, we could compare all 69 allele despite some not having complete sequences.

To further evaluate the mode of evolution of each group of alleles, to identify possible peptide binding sites, and to identify potentially non-classical MHC alleles, we used site-specific tests of selection focusing on exons 2 and 3 that are the exons for which we have full sequence coverage for most alleles. We tested site-specific positive selection within each of the four groups of alleles, and across all alleles excluding Group I, due to its likely different evolutionary history. Site-specific selection was tested using

CodeML in PAML v4.7 (Yang 2007), assuming different ω parameters among codons with no *a priori* knowledge of which class of selection (neutral, purifying, or positive) a given codon belongs to. We estimated parameters under five different codon substitution models: Beta models M7 (no positive selection), and M8 (positive selection) M8a (no positive selection and $\omega = 1$), and models M1a (nearly neutral) and M2a (positive selection) (Yang 2007). A likelihood ratio test (LRT) was performed to compare the fit of the models with and without selection. Statistical significance was determined by comparing twice the difference of log-likelihood scores ($2\Delta lnL$) to the X^2 distribution with degrees equal to the difference in the number of parameters between the models to be compared.

Protein tertiary structure homology models

To characterize the tertiary structure of each allele, and determine if they have the proper characteristics to allow them to fold into a potentially functional MHC molecule we built protein homology models using the Swiss-Model workspace v8.05 (Arnold et al. 2006). This method identifies structural templates from a protein data bank, aligns the target sequence to a template structure, builds a 3D model, and evaluates the quality of the model. The parameters considered were: *overall model quality* (QMEAN4) in which less negative values indicate more reliable homology models and *global model quality estimation* (GMQE) in which the higher numbers indicate the more reliable models. We also located the four conserved cysteine residues that are essential for the stability of the β_1 and β_2 domains of MHC IIB and analyzed the structure of the N- and C-terminal areas of the protein.

Results

Cloning, sequencing and MHC organization

For a total of 13 individuals, we obtained 867 gDNA and 756 cDNA sequences, all belonging to the MHC IIB. These sequences revealed 69 distinct alleles (GenBank accession numbers: KY039442-KY039474 and KY354964-KY355011). Three of these alleles (DXB*05, DXB*1002, and DXB*13) were only represented by one single sequence while one was only represented by 2 sequences (DXB*0402), and would remain to be confirmed by an additional PCR amplification. We found an average of 12.5 (SD = 6.1) alleles per individual and a maximum of 25 (Table 2 and Fig. 2). We identified six exons and five introns, but we did not sequence all exons for all alleles (Table 2). We characterized the gene from exon 1 through exon 6 focusing efforts on exon 2 where we expect most variability (Fig.3). For fifteen alleles we amplified exon 1 and found that it has 55 bp, and includes the start codon "ATG" that codes for a methionine residue at the N-terminus. Exon 1 mainly codes for the signal peptide formed by 16-20 amino acids. Exon 2 is the most polymorphic of all exons with up to ten amino acids per site (Supplementary Fig. 1) and consists of 273 bp, of which we sequenced at least 231 bp for all 69 alleles. Exon 2 encodes the beta sheets and alpha helixes of the β_1 domain of the mature MHC class II molecule. For most alleles, we amplified intron 2 which revealed extreme length variation ranging from 200 to 2500 bp (Fig. 3). Alleles with long intron 2 contained three tandem repeats, a 42mer repeated 6.4 times, a 21mer repeated 12.6-18.6 times, and a 127mer repeated twice which contributed to their length. For exon 3 we amplified from 99 bp to 214 bp of most alleles, exon 3 is 214 bp long and is enriched in the amino acid proline, aiding in the formation of beta turns that are essential to the structure of the β_2 domain. We obtained partial sequences of intron 3, for two alleles, and a complete sequence for one allele that was 405 bp long. We obtained sequences of exon 4 for ten alleles that were 69 bp. For nine alleles we sequenced 64 bp of exon 5 which is expected to have 114 bp. Two of them (DXB*060101 and DXB*07) have additional 12 bp on the 5´end of exon 5. Introns 4 and 5 were not sequenced for any allele because of their known low variability (Dixon et al. 1995). We generated only partial sequences of exon 6, which we cannot assign to individual alleles as they do not span any allele-specific polymorphisms due to the low variability of this region. The summary of the length in bp of each sequenced intron and exon in shown in Supplementary Table 4 with the reference of the sequence of *Oreochromis niloticus*.

Figure I. 2. Alignment of amino acid sequences of MHC IIB. A majority consensus sequence was made for comparison. '•' represent identical amino acids, '-' represent gaps or introns that were not sequenced. Cysteine residues are outlined in red boxes.

Table I. 2. List of all of MHC IIB alleles found in the Midas cichlid

Samples	AM05	AM06	AH65	AL11	AL64	AQ72	27D3		25D8		27G4	25A2		25D1		25B2		27E7		Total Secuenses
Template	gDNA	gDNA	gDNA	gDNA	gDNA	gDNA	gDNA	cDNA	gDNA	cDNA	cDNA	gDNA	cDNA	gDNA	cDNA	gDNA	cDNA	gDNA	cDNA	
Allele																				
Amci-DXB*000101	12	13	69	3		21	46													**164**
Amci-DXB*000102	71	10	49		13	77	4													**224**
Amci-DXB*000103	1					1														**2**
Amci-DXB*000104	1			1																**2**
Amci-DXB*000105	1		1																	**2**
Amci-DXB*000106		1				2														**3**
Amci-DXB*000107			1			1														**2**
Amci-DXB*000108			1			1														**2**
Amci-DXB*0002	1				1															**2**
Amci-DXB*0003	1		1									1								**3**
Amci-DXB*0004			1	1		1	13													**16**
Amci-DXB*0005							2													**2**
Amci-DXB*0006			2			1	2													**5**
Amci-DXB*0007	1				1															**2**
Amci-DXB*0008	2	1			1															**4**
Amci-DXB*0009			1			1														**2**
Amci-DXB*0010				1		1														**2**
Amci-DXB*0011		1	1																	**2**
Amci-DXB*0012			1			1	1													**3**
Amci-DXB*0013			1	1																**2**
Amci-DXB*0101			2	2		3	1		2											**10**
Amci-DXB*0102	5		2	2	1	9														**19**
Amci-DXB*0201							39					1				4				**44**
Amci-DXB*0202							17					14				26	1			**58**
Amci-DXB*0203									6			6								**12**
Amci-DXB*03												26				1				**27**
Amci-DXB*040101	9	1	4		2	13	7		4			9				10		2		**61**
Amci-DXB*040102	5	1	5	3	1						17				3		8			**43**
Amci-DXB*040103	1					1														**2**
Amci-DXB*0402	1					1														**2**
Amci-DXB*040301				3	2	1		1		25					3		7			**42**
Amci-DXB*040302			1			1														**2**
Amci-DXB*040303	2					16	13													**31**
Amci-DXB*040304				3	8		18													**29**
Amci-DXB*040305												4								4
Amci-DXB*0404	2					1														**3**
Amci-DXB*0405							3	26				1			16		20			**66**
Amci-DXB*0406								1							1					**2**
Amci-DXB*05			1																	**1**
Amci-DXB*060101								1					43	1	31	2	42		16	**136**
Amci-DXB*060102													1		1					**2**
Amci-DXB*0602													1		6					**7**
Amci-DXB*0603										5									3	**8**
Amci-DXB*0604		1													1				16	**18**
Amci-DXB*07													2							**2**
Amci-DXB*08														1	40					**41**
Amci-DXB*09										50									9	**59**
Amci-DXB*1001															1				25	**26**
Amci-DXB*1002															1					**1**
Amci-DXB*110101							31	42												**73**
Amci-DXB*110102								2												**2**
Amci-DXB*1102								8												**8**
Amci-DXB*12										1					19					**20**
Amci-DXB*13											1									**1**
Amci-DXB*14																			38	**38**
Amci-DXB*15										16										**16**
Amci-DXB*16							4								34					**38**
Amci-DXB*17							4	18												**22**
Amci-DXB*18												17	44			23	50			**134**
Amci-DXB*19																	9			**9**
Amci-DXB*20																			2	**2**
Amci-DXB*21															2					**2**
Amci-DXB*2201							2	23												**25**
Amci-DXB*2202								2												**2**
Amci-DXB*23								2												**2**
Amci-DXB*2401								13												**13**
Amci-DXB*25												1				3				**4**
Amci-DXB*26										6										**6**
Amci-DXB*27	2																			**2**
Total secuences	**118**	**29**	**144**	**20**	**30**	**154**	**207**	**139**	**12**	**103**	**18**	**80**	**91**	**2**	**159**	**69**	**137**	**2**	**109**	**1623**
Total Alleles	17	8	18	10	9	20	25		9		2	14		14		11		8		69

Alleles found in 13 Midas cichlid samples (listed on the top) showing number of copies recovered per allele amplified in gDNA and cDNA in each individual. All alleles except DXB*05, DXB*1002, DXB*0402, and DXB*13 were confirmed with at least two independent PCR reactions. At the bottom, we summarize a total number of sequences and the total number of alleles obtained.

To further corroborate our results, we did a BLAST search on the *A. citrinellus* shotgun genome project (BioProject PRJEB6974) and recovered three complete MHC IIB genes from contig 1595 position 54549-67334 (CCOE01001596.1), contig 1079 position 91043-98760 (CCOE01001080.1), and unplaced scaffold 125 position 28791-32030 (CCOE01001892.1), as well as several incomplete MHC IIB sequences. The three complete sequences contained all 6 exons, with the same length we describe. This data confirms that the length of exon 5 is 114 bp (of which we only retrieved partial sequences), but only contained one of the two variants we found (the short variant). Intron length variability was also represented in the genome sequences. We did not find in this data all intron variants we found in our sequences (*e.g.* we found intron 2 with 702 bp and 1.5 kb, but not with 248 bp), but found new intron length variants (*e.g.* intron 2 of contig 1595 is 10 kb). This data must be taken with caution since the shotgun sequences have not been confirmed with PCR and sequencing standard protocols employed when working with MHC. Nevertheless, this data provides valuable corroboration about intron lengths that is otherwise unavailable at this time.

Although in general alleles are defined as alternate sequences of a single locus, since there is no complete reference genome for this species, there is no information on how many MHC loci to expect, or how variable they may be. It must be noted that our alleles cannot be assigned to specific loci at this stage.

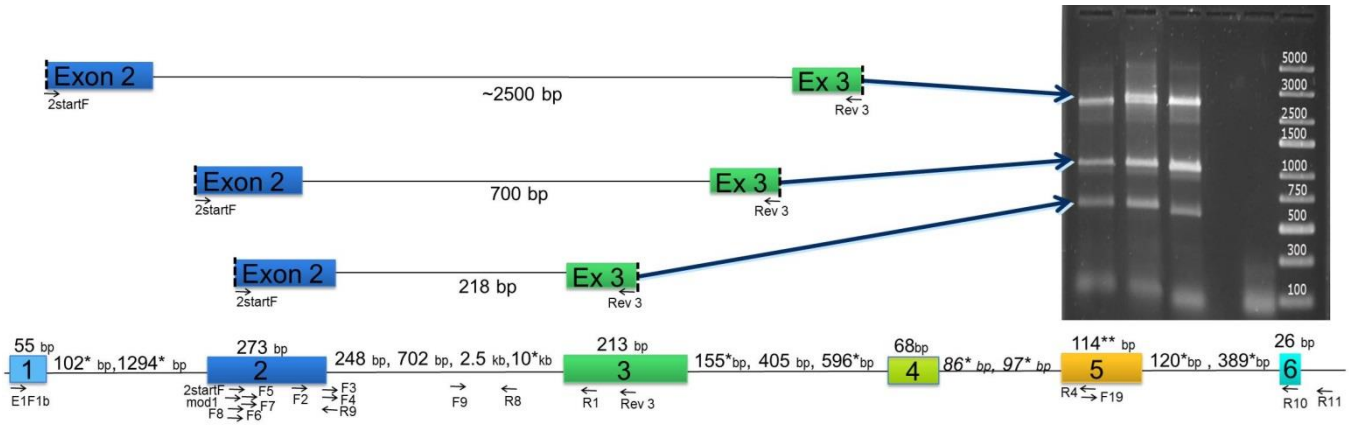


Figure I.3. Representation of intron-exon organization of MHC IIB sequences showing length of fragments and position of primers. The gel shows variation in the length of intron 2 (showing exon 2 / intron 2 / exon 3) for three individuals. * indicates lengths observed in sequences of the *Amphilophus citrinellus* draft genome shotgun sequencing project and ** indicate the exon that we found to be 12 bp longer in allele DXB*060101 and DXB*07.

Grouping of alleles and Phylogenetic analyses

We constructed a phylogeny with all 69 alleles recovered from exons 2 and 3 (Supplementary Fig. 2). We also constructed a neighbor-net network (Fig. 4). According to estimates of evolutionary divergence between sequences, alleles grouped into five major groups (Supplementary Fig. 3). When plotting these groups into the network one of these groups was not consistent and we left them as ungrouped (14 alleles). The remaining alleles clustered into four groups (I, II, III, IV). The results of the randomization analysis validating the groups are shown in Supplementary Table 5. Group I grouped 20 alleles, which were only found in gDNA. These alleles showed very low polymorphism and might represent non-classical MHC genes or MHC pseudogenes (see selection analyses section). Group II was further divided into two sub-groups, II-a and II-b. Group II-a assembled 13 alleles, 9 found in gDNA, 1 found in cDNA, and 3 found in both cDNA and gDNA. Group II-b had 6 alleles, 5 found in gDNA and 1 found in both gDNA and cDNA. Group III assembled 6 alleles, 5 found in cDNA and 1 found in both gDNA and cDNA. Ten alleles were included in Group IV, 7 found in cDNA, and 3 in both gDNA and cDNA. From the 14 ungrouped alleles, 12 were found in cDNA, 1 in gDNA, and 1 in both cDNA and gDNA.

Taken together the Midas cichlid MHC IIB alleles belonged to two types, those alleles not found expressed in the tissue examined in this study (alleles in Group I), and those found in both gDNA and cDNA (alleles in the rest of the groups).

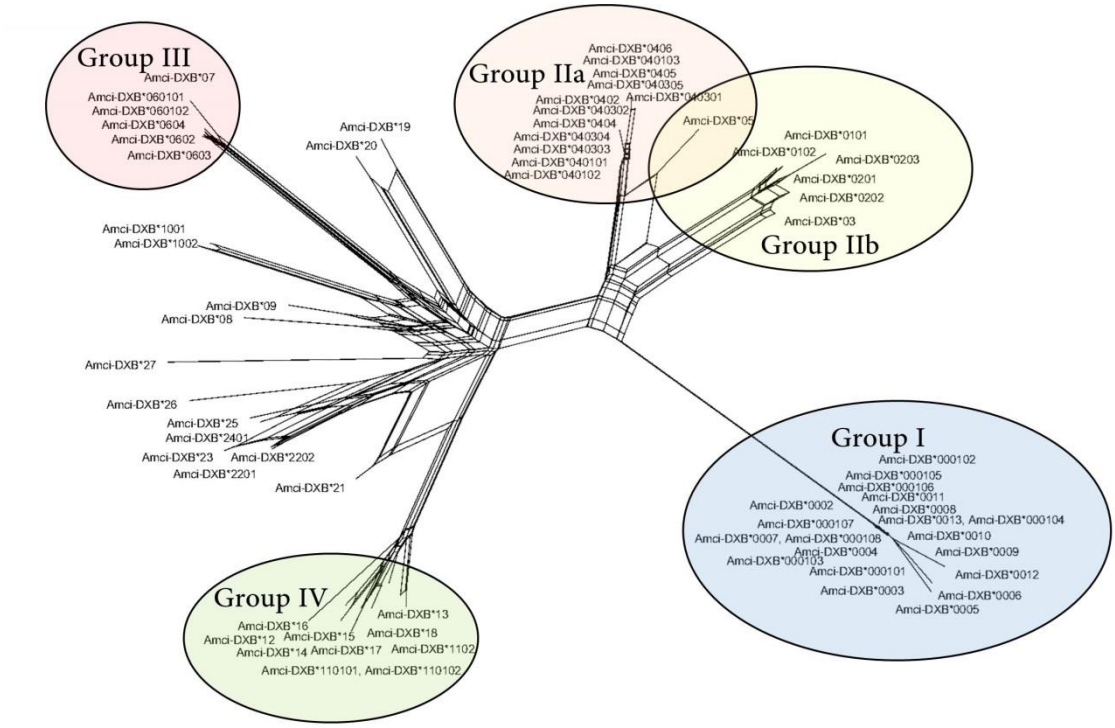


Figure I.4. Neighbor-net network based on exons 2 and 3 of MHC IIB allele relationships. Groups of alleles found with estimates of evolutionary divergence between sequences are shown.

Trans-species polymorphism

We reconstructed the phylogenetic relationships of exon 2 MHC IIB of all alleles we retrieved in the Midas cichlid, and 20 alleles of several other fish species of cichlids and non-cichlids (*Sebastes caurinus*, *Sebastes maliger*, *Miichthys miiuy*, *Pundamilia nyererei*, *Haplochromis sp. 'rockkribensis'*, *Haplochromis xenognathus*, and *Oreochromis niloticus* (see Supplementary Table 3). All of these alleles clustered within the Midas cichlid alleles providing evidence for trans-species polymorphism. Nile tilapia (*Oreochromis niloticus*), the species for which there exists most MHC information, had alleles distributed throughout the tree (Fig. 5). One of these alleles clustered with our Group I (DJB accession # AB677258.1). Several alleles from different species clustered with alleles in Group IV, and also with the ungrouped alleles.

Test of Selection

The analyses of site-specific selection revealed that alleles in groups II, III and IV had sites that fitted best the model with positive selection (Group II, $p = 0.02$; groups III and IV, $p < 0.001$), while alleles in Group I did not ($p = 0.10$). Alleles in Group I did not have any synonymous substitutions, therefore d_N / d_S ratios could not be calculated. We tested all alleles including those in Group I, and found 19 positively selected sites, 13 of which were inferred at 99 % level (Table 3). When we excluded alleles in Group I, 19 and 20 sites were inferred to be positively selected for models M2a and M8 respectively, 15 and 16 of them were

inferred at 99 % level. However, only 4 sites overlapped with those in the previous analysis. For the groups II, III and IV, we found between 1 and 9 sites to be under positive selection (Table 3). However, these sites were not shared between groups whether or not Group I was included, indicating that the groups might have followed different evolutionary trajectories.

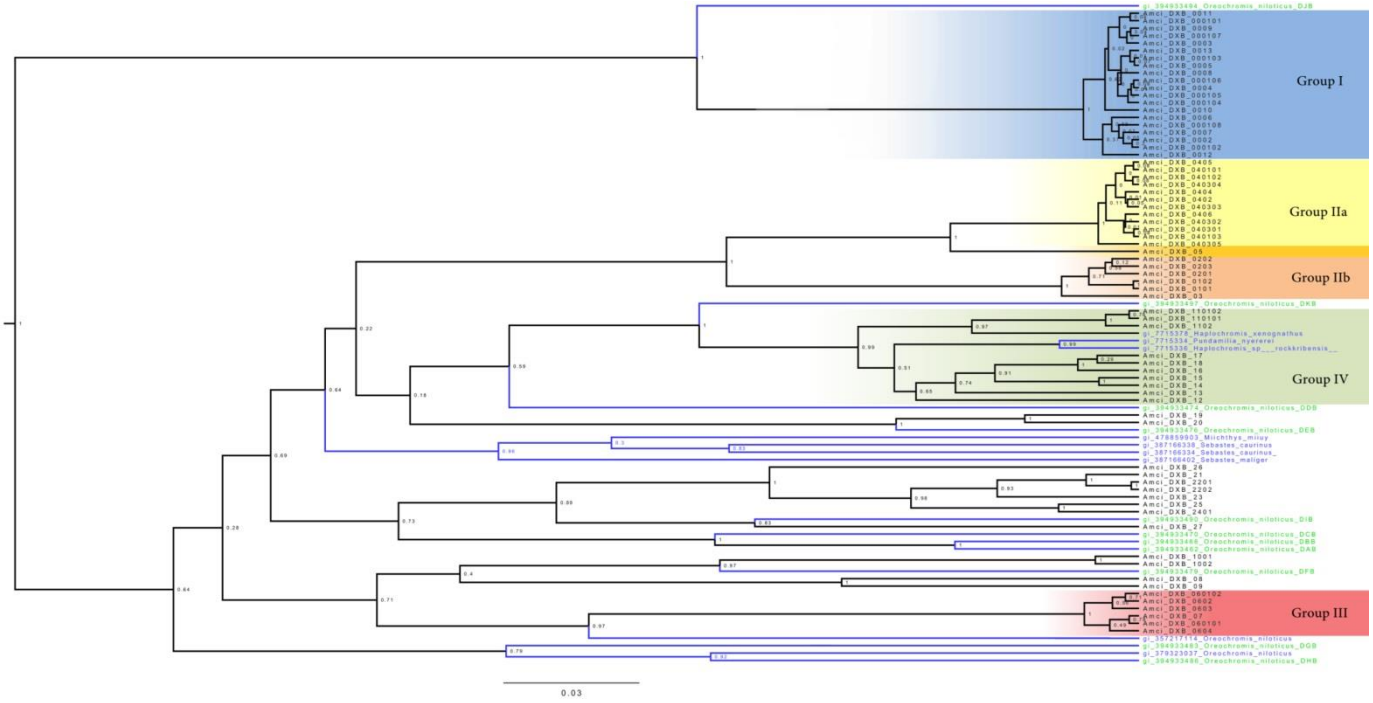


Figure I.5. Phylogenetic tree showing trans species polymorphisim of exons 2 and 3 of all MHC IIB alleles found in the Midas cichlid (in black) and 20 alleles from other species (in blue), with posterior probabilities. Alleles are grouped according to how they cluster in Fig. 4.

Table I.3. Summary of likelihood ratio tests for site-specific positive selection of MHC IIB genes comparing groups of alleles

Allele Group	Models compared	LRT (2Δl)	p-value	Estimate for ω >1	Proportion of PSS	PSS
All groups	M1a versus M2a	81.738836	1.09E-06	3.71226	0.34292	**1**, **2**, 7, 9, **10**, **32**, 33, **37**, **38**, **39**, 41, **42**, **43**, 46, **57**, **58**, **60**, 61, **62**
All groups	M7 versus M8	84.941740	6.61E-07	3.55150	0.35599	**1**, **2**, **7**, 9, **10**, **32**, **33**, **37**, **38**, **39**, **41**, **42**, **43**, **46**, **57**, **58**, **60**, 61, **62**
All groups	M8a versus M8	73.238718	6.12E-07			
All except I	M1a versus M2a	83.327546	1.16E-06	3.89788	0.32156	**8**, **9**, **16**, **17**, 32, 39, **40**, **44**, **45**, **46**, 48, **49**, **50**, **53**, **64**, **65**, **67**, 68, **69**
All except I	M7 versus M8	80.631206	6.74E-07	3.72628	0.33118	**8**, **9**, 14, **16**, **17**, 32, **39**, **40**, **44**, **45**, **46**, 48, **49**, **50**, **53**, **64**, **65**, **67**, 68, **69**
All except I	M8a versus M8	74.046658	8.90E-07			
Group I only	M1a versus M2a	4.594664	0.10			
Group I only	M7 versus M8	4.594238	0.10			
Group I only	M8a versus M8	4.594000	0.10			
Group II only	M1a versus M2a	7.702834	0.021	3.16267	0.48401	49, 65
Group II only	M7 versus M8	7.706184	0.021	3.16268	0.48401	13, 49, **65**
Group II only	M8a versus M8	7.702818	0.021			
Group III only	M1a versus M2a	14.605000	0.0007	31.27408	0.03609	**73**
Group III only	M7 versus M8	14.663852	0.0007	31.37037	0.03603	**73**
Group III only	M8a versus M8	14.604918	0.0007			
Group IV only	M1a versus M2a	20.610518	0.00003	7.27232	0.15698	11, **48**, **55**, 67, **69**, **106**
Group IV only	M7 versus M8	20.981230	0.00003	7.22462	0.15872	**11**, 42, **48**, **55**, 66, 67, **69**, 77, **106**
Group IV only	M8a versus M8	20.620836	0.00003			

The LRT models compared by the likelihood ratio test with five codon substitution models: Beta models M7 (no positive selection), and M8 (positive selection) M8a (no positive selection and $\omega = 1$), and models M1a (nearly neutral) and M2a (positive selection). The $2\Delta l = 2$ (lb - la), $\omega = d_N / d_S$, positively selected sites (PSS) are inferred by empirical Bayesian posterior probabilities. PSS in bold are inferred at 99 %. The tests of overall selection among all alleles showed that the entire sequence (LP, β_1 and β_2 domains) of all alleles is under purifying selection ($p = 0.021$), and this pattern remains when the alleles of Group I are excluded ($p = 0.005$) (Table 4). Nine alleles were excluded from the analysis of the β_2 domain because we did not obtain sequences of this domain for those alleles. The β_2 domain shows purifying selection for all alleles ($p < 0.001$), as well as when excluding Group I ($p = 0.001$). However, neither the β_1 domain nor the entire sequence showed signs of overall positive selection ($p = 1.00$) excluding Group I ($p = 0.001$) (Table 4).

Table I.4. Tests of overall selection and selection by domain.

Domain	Alleles	length (AAs)	d_N	d_N Number	d_S	d_S Number	d_N/d_S	*P*- value (purifying)	*P*- value (positive)
LP, β_1 & β_2	Group I-IV and ungrouped (N69)	228	0.221 (0.024)	44.488 (4.998)	0.327 (0.053)	18.525 (2.234)	0.676	**0.021**	1.000
β_1	Group I-IV and ungrouped (N69)	91	0.254 (0.032)	37.703 (3.988)	0.301 (0.059)	12.807 (1.679)	0.844	0.219	1.000
β_2	Group I-IV and ungrouped (N60)	93	0.129 (0.029)	8.502 (2.154)	0.459 (0.101)	7.238 (1.585)	0.282	**<0.001**	1.000
β_1 & β_2	Group I (N20)	113	0.008 (0.003)	1.616 (0.521)	0.000 (0.000)	0.000 (0.000)	--	--	--
LP, β_1 & β_2	Group II-IV and ungrouped (N49)	229	0.193 (0.022)	41.637 (4.845)	0.306 (0.044)	18.101 (2.142)	0.633	**0.005**	1.000
β_1	Group II-IV and ungrouped (N49)	91	0.216 (0.028)	33.987 (3.776)	0.270 (0.049)	11.568 (1.579)	0.800	0.131	1.000
β_2	Group II-IV and ungrouped (N40)	94	0.138 (0.034)	9.378 (2.293)	0.463 (0.114)	8.030 (1.679)	0.298	**0.001**	1.000

Tests of overall selection on the entire length of sequences, and on the different domains were performed. The leading peptide LP, and the two domains β_1 and β_2 comprise the entire known sequence, despite not all alleles have full coverage of both domains and the LP. The amino acid lengths for each test are given as the number of sites that were included in the test. Three alleles (0005, 0006 and 0012) from Group I,

and six alleles (0203, 040103, 0406, 05, 25, 27) from Groups II-IV were excluded for the test of selection on the β2 domain, as they lack sequence information for this domain. $d_{N,}$ number of non-synonymous substitution per non-synonymous site; $d_{S,}$ number of synonymous substitutions per synonymous site; standard error (SE). Significant p-values (at 99 % confidence level) are shown bold.

Protein Structure Homology Models

We built protein homology models for all alleles to characterize their tertiary structure, and to determine if they can fold into a potentially functional MHC molecule. The QMEAN4 ranged between -1.72 and -4.32, indicating generally good quality of the proposed models (see Supplementary Table 6 for details). GMQE score ranged between 0.61 and 0.75, indicating an overall good quality for most models (Supplementary Table 6).

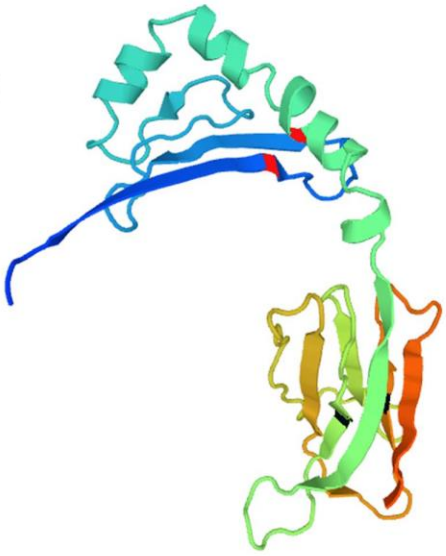
We located the cysteine residues in the 3D structure of alleles to see if they were in the correct position to make structural disulfide bonds. All alleles had an unpaired cysteine at position 7 of the leader peptide, and the four expected cysteine residues at positions 29, 94, and 132, 188, which pair covalently to form disulfide bonds that increase the stability of the β_1 and β_2 subunits respectively (Fig. 6a). All alleles in Group III had an additional unpaired cysteine at position 47 (Fig. 6b), and all alleles in Group II-b had an additional cysteine at position 98. Alleles of Group I showed no notable anomalies in their 3D structure, but allele DXB*0007 presented an unpaired cysteine at position 47 similarly to alleles of Group III. The N-terminal area of MHC IIB protein included an alpha helical region and a beta sheet of four strands in an antiparallel orientation. It also showed that the C-terminal area mainly has a beta-fold structure and is characterized by an immunoglobulin-like beta-sandwich made of two anti-parallel sheets. Interestingly, our work revealed that all 3-D models were similar among all groups of alleles, with the exception of the unpaired cysteine positions.

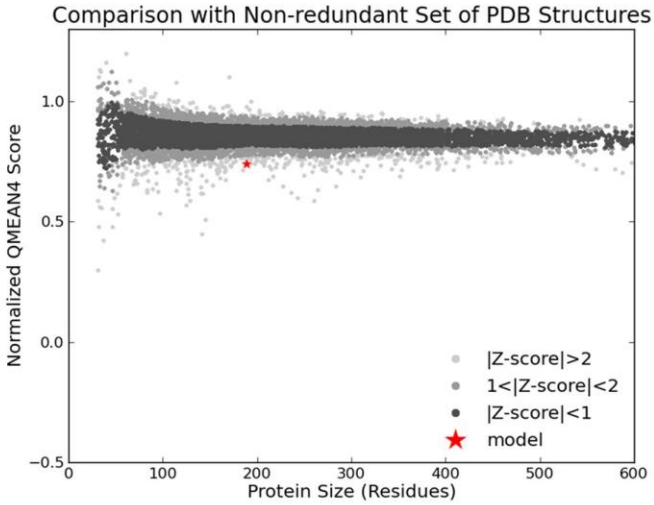
Discussion

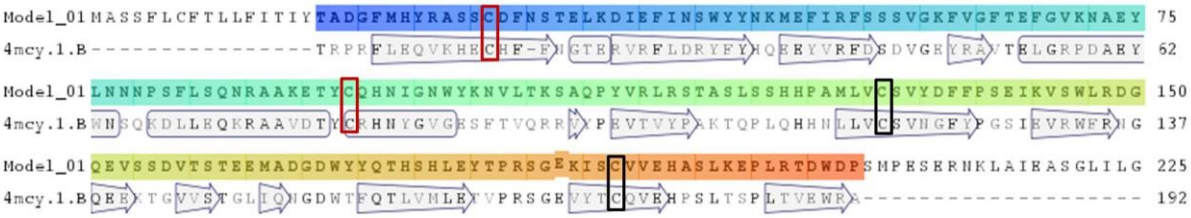
We found a total of 69 alleles of MHC IIB exons 2 and 3 in 13 Midas cichlid individuals. This represents very high allelic diversity in this species despite the small sample size tested. Individual Midas cichlids harbor a large number of alleles, with a maximum of 25 per individual, and an average of 12.5 (SD = 6.1). Together, this implies that the Midas cichlid has at least 13 distinct MHC IIB loci, although this may be an underestimation as we cannot detect allele sharing between loci, and we may have insufficient sequences for some alleles. In other Old World cichlid species, up to 17 polymorphic loci have been found, and between 1-13 alleles per individual (Málaga-Trillo et al. 1998). Hence, our characterization of Neotropical cichlids revealed comparable structure and diversity between Old and New World cichlids.

The Large variation in number of MHC loci, the total number of alleles at the population level, and heterozygosity at the individual level, have been reported among fish taxa. For example, pipefish (*Syngnathus typhle*) and Atlantic cod (*Gadus morhua*) have lost all MHC class II loci (Star et al. 2011; Haase et al. 2013), and appear to compensate for it by larger diversification of MHC class I (Star and Jentoft 2012). Seahorse (*Hippocampus abdominalis)* has a single MHC IIB locus, and a maximum of two alleles per individual (Bahr and Wilson 2011). Spotted halibut (*Verasper variegatus*) and Japanese flounder (*Paralichthys olivaceus)* have up to five alleles per individual (Xu et al. 2010; Li et al. 2011), and guppy (*Poecilia reticulata*) and eel (*Anguilla anguilla*) have up to six alleles (Lighten et al. 2014a; Bracamonte et al. 2015), which suggests that they all have at least 3 distinct loci. Stickleback (*Gasterosteus aculeatus*), a species with thoroughly studied MHC, has between 3 and 6 MHC class IIB loci depending on the population of origin (Sato et al. 1998; Lenz et al. 2009b). However, cichlid MHC IIB alleles are more numerous than those of any other fish that has been studied. This may have contributed to, or be a result of their great diversification.

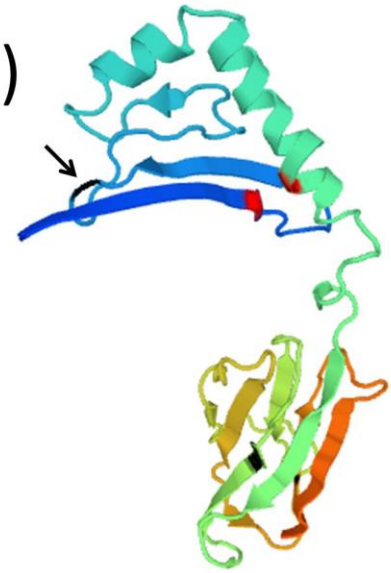
a.)

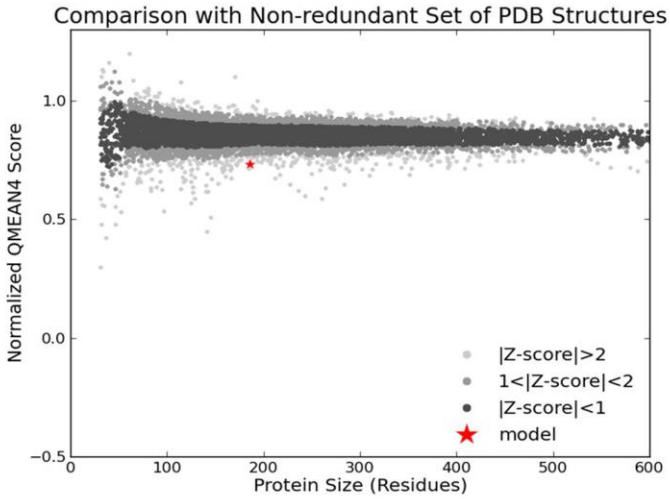



Comparison with Non-redundant Set of PDB Structures
Normalized QMEAN4 Score
Protein Size (Residues)
|Z-score|>2
1<|Z-score|<2
|Z-score|<1
model


Model_01
4mcy.1.B

b.)




Comparison with Non-redundant Set of PDB Structures
Normalized QMEAN4 Score
Protein Size (Residues)
|Z-score|>2
1<|Z-score|<2
|Z-score|<1
model

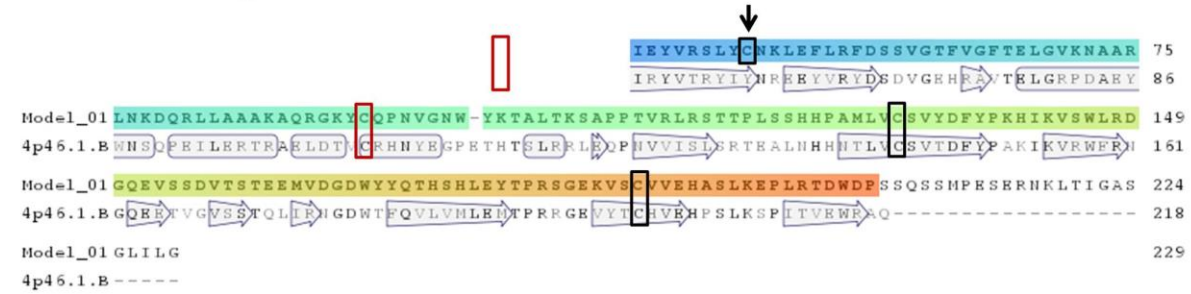

Model_01
4p46.1.B

Figure I.6. Models of tertiary structure of MHC IIB sequences, where red boxes represent cysteine amino acids forming the disulfide bond in the β_1 domain, and black boxes represent cysteine residuals that form the disulfide bonds of the β_2 domain. **a)** Model of allele DXB*2202; **b)** Model of allele DXB*060101, with an unpaired cysteine at position 47 (indicated with a black arrow). The graphs show how each model (red star) compares to a non-redundant set of Protein data bank (PDB) structures, indicating the quality of the model compared to molecules of the same size as a value equivalent to a z-score test.

Because MHC genes are directly responding to local parasite pressure (Spurgin and Richardson 2010; Milinski 2014) they may encode for a magic trait contributing to local adaptation and ultimately to ecological speciation (Eizaguirre et al. 2009a; Eizaguirre et al. 2012). Determining the contemporary MHC diversity is, however, challenging due to its multigene nature and fast turnover, often affecting conserved regions that are typically used for setting up primers for amplification (Wegner 2009). Neotropical cichlids are model species for speciation in sympatry (Barluenga et al. 2006), but their MHC genes had not been characterized to date. The nearest relatives with well characterized MHC class II genes are African cichlids (Figueroa et al. 2000; Sato et al. 2012; Pang et al. 2013), from which they split 93 MY ago (Albert and Reis 2011). We, therefore, carried out this study to establish reference sequences in order to obtain specific primers that would amplify a comprehensive diversity of MHC IIB genes for future population-based studies. Like most other fish, the sequences we obtained of the Midas cichlid MHC IIB genes are composed of 6 exons (Ono et al. 1993b; Pang et al. 2013). We did not obtain the sequence of all exons for all alleles, but for those we have the complete sequence, there are always 6 exons. Published sequences of these genes in other cichlids show the same intron-exon organization (Ono et al. 1993b; Pang et al. 2013). This is the most standard structure in fish, even though species with 5 exons and functional MHC have also been described (*e.g.*, sea bream *Chrysophrys major* (Chen et al. 2006) or Japanese flounder *Paralichthys olivaceus* (Zhang et al. 2006)). Variation in exon number is due to either exon fusion (Dijkstra et al. 2007), or exon splitting (Ono et al. 1993b; Figueroa et al. 1995).
We did, however, find considerable variation in the length of intron 2, which ranged from 239 bp to 2.5 kb in the sequences we obtained, and 10 kb in the genome sequences. This intron length variation is likely the reason why we were unable to amplify some alleles that were obtained from cDNA when using the same primers spanning introns in gDNA. Length variation in intron 2 has been reported in several species including other cichlids (Málaga-Trillo et al. 1998; Xu et al. 2011; Jiang et al. 2015). The tandem repeats found in intron 2 contribute to the increased length of the long intron. Reusch and Langefors (Reusch and Langefors 2005) reported a 10-mer repeat in intron 2 of three-spined sticklebacks, responsible for important changes in sequence length, demonstrating that this mode of intron evolution can happen in other fish species. The genome sequences also revealed length variation in all other introns, most notably in intron 3 that varies between 155 bp to 5 kb.

Within the 69 alleles found in the Midas cichlid, we distinguished different groups. One group of alleles (Group I) resembled a pattern of non-classical MHC IIB genes (Grimholt 2016). These alleles showed low variability, are apparently not expressed, and none of their positions seemed to have evolved under positive selection according to our selection analyses. A pattern of low polymorphism is typical of non-classical MHC IIB genes since their primary function is to assist in loading the antigenic peptides onto classic MHC class II molecules (Kropshofer et al. 1999). Because non-classical MHC molecules do not have to bind to a wide array of antigen peptides, the sequence between the different alleles does not follow the typical "patchwork" pattern of classic MHC II alleles, especially in the peptide binding region (Kropshofer et al. 1999). Non-classical MHC IIB genes similar to those in Group I of the Midas cichlid have also been described in Atlantic Salmon (*Salmo salar*; Harstad et al. 2008). All other groups of alleles and the ungrouped alleles showed classic MHC class II gene patterns. They all displayed sites subjected to strong positive selection, suggesting that they might have evolved under strong parasite-mediated balancing selection (Takahata et al. 1992; Xu et al. 2016). However, the results of the selection tests have to be taken with caution since we cannot allocate alleles to specific loci and alleles within groups could at potentially belong to different evolutionary lineages. This might also have an influence on the overall selection tests for which all groups were combined. It might explain some of the discrepancies between strong positive selection in site-specific tests and the absence of positive selection in the overall tests. Antagonistic coevolution between hosts and parasites is recognized as a powerful force capable of driving rapid evolutionary changes, which might significantly contribute to biodiversity [*e.g.*, 91]. In fish, MHC frequency shifts of resistance alleles have been observed as a response to local parasite-mediated selection (Eizaguirre et al. 2012). Combined with MHC-based mate choice reported in almost all jawed vertebrates (Milinski 2014), host-parasite interaction through MHC genes has been suggested to contribute to speciation, even in sympatry (Blais et al. 2007; Eizaguirre et al. 2009b; Eizaguirre and Lenz 2010).
The phylogenetic tree of all 69 alleles plus 20 of other fish species, displayed a pattern that strongly supports trans-species polymorphism. Some alleles of the Midas cichlid seem to be more closely related to alleles of other species than to other alleles of the same individual. For example, alleles of Group I are more closely linked to Nile tilapia DJB allele (DJB accession # AB677258.1) than to any other Midas MHC IIB allele. Similarly, allele DXB*27 of the Midas cichlid is closely related to allele DIB of Nile tilapia, indicating homology and hence TSP. TSP is a common pattern of the MHC and has been observed in many taxonomic groups (reptiles (Bryja et al. 2006; Stiebens et al. 2013), amphibians (Kiemnec-Tyburczy et al. 2010), mammals (Eimes et al. 2015), birds (Kikkawa et al. 2009), and fish (Ottová et al. 2005; Lenz et al. 2013)). TSP is evident throughout the phylogenetic tree and seemed to be most common with alleles of Group IV and ungrouped alleles.

The tertiary structure models showed that similar to the Nile tilapia (Zhou et al. 2013), the Midas cichlid MHC IIB sequence has all the necessary features for the molecule to be functional, including two pairs of cysteine residues. The biological function of unpaired cysteine residues in the MHC molecules remains unknown. It has however been suggested that they could play a role in the formation of exosomal dimers (Lynch et al. 2009). We found two groups of alleles with extra unpaired cysteines (groups II-b and III), but nothing noteworthy was found in the structure of these alleles. Future studies focusing on the tertiary structure of MHC molecules should focus on determining the function of unpaired cysteines, to further reveal their contribution to immunity specifically, and species' evolution in general.

Despite considerable sequencing effort, we were not able to find all alleles in both cDNA and gDNA. Alleles in Group I were never detected in cDNA, which makes us think they might be putative non-classical or even pseudogenes. On the other hand, we had difficulties in amplifying alleles from Group III in gDNA while they were readily obtained in cDNA. We only succeeded in amplifying these alleles in gDNA by using primers that excluded intron 2. Intron 2 is, therefore, likely causing sequencing difficulties due to particularly long sequences or rich GC content. Another explanation for these difficulties might be alternative splicing, which is known to occur in MHC (Laurens et al. 2001). Indeed, in salamanders, over 20 % of the transcripts can be alternatively spliced, with variation in different organs, see Bulut *et al.* (Bulut et al. 2008). As the alleles discovered here seem to be functional and variable, and they may be contributing to the dynamic response of MHC to parasite selection (Eizaguirre and Lenz 2010).

Conclusions

Taken altogether, MHC IIB genes in the Midas cichlid showed enormous richness in allele diversity and copy number. This diversity is larger than that described in most other fish and is only comparable to that found in other cichlids. Our findings promise great potential in studying the processes of evolution and speciation in this model system and should be further studied at the ecotype, population and species levels to elucidate the role that parasites may play in sympatric speciation.

CHAPTER II

MHC diversity of the Midas cichlid in the Nicaraguan lakes

Manuscript submitted

Abstract

Background: The immense diversity of cichlids and their ability to proliferate in new environments has made them a model system for studying evolution. The Midas cichlid species complex inhabits several small crater lakes and two large tectonic lakes in Central America. Midas cichlids also exhibit one of the most convincing cases of sympatric speciation, many studies have been done to determine the mechanisms that are driving divergence in this system, they have focused on morphology, genetics and genomics as well as ecological interactions, however none so far have studied host-parasite interactions as a potential mechanism for divergence and speciation. Parasite-mediated local adaptation has been identifies as one of the mechanisms that contribute to population divergence. The Major Histocompatibility Complex (MHC) class II genes are a multi-gene family that is responsible for recognizing invading parasites and triggering the immune system to combat them.. Recent innovations in sequencing technologies facilitate the sequencing of multigene families like MHC at a population level at a fraction of the cost and effort of traditional cloning.

Results: Here we characterize the MHC profiles of 8 species of Midas cichlid fish across six different lakes. We sampled 287 individual and got over 4 million reads that belonged to 147 alleles. Bayesian inference grouped alleles into 6 different groups, which we consider supertypes that may correspond to functionally equivalent groups. We inferred divergent selection by testing for selection pressures at putative antigen binding sites. Each species within each lake displayed a signature MHC pattern of distinctive alleles. Lake Xiloá and Lake Asososca León were the most distinct, and at the species level *A. sagittae* had the most distinct repertoire of alleles. Individuals in all studied populations had an average of 7 alleles, belonging to 6 supertypes.

Conclusions: This study shows that there is great diversity in MHC allele repertoire among the Midas cichlid populations in the different Nicaraguan lakes, suggesting that they could have important adaptive

value. With this we provide the first evidence of adaptive divergence of MHC in the Midas cichlid species complex.

Introduction

The process and mechanisms of speciation have received substantial interest within the last two decades (Sugden 2000; Wu 2001; Beaumont 2005; Rundle and Nosil 2005; Maan and Seehausen 2011; Nosil 2012; Butlin et al. 2012; Bernardi 2013). Recent innovations in sequencing technologies offer an unprecedented opportunity to study the genetics behind mechanisms involved in ecological speciation at a population level (Wu and Ting 2004; Chevin et al. 2014; Seehausen et al. 2014). There is growing evidence that ecological selection on traits such as feeding morphology, environmental tolerance, or host-parasite interactions, plays an important role in the divergence of populations (e.g., Schluter 2001; Eizaguirre et al. 2009a). One of the mechanisms that may be contributing to population divergence is parasite-mediated local adaptation. The key genetic components that are under selection in parasite-mediated selection are the genes of the Major Histocompatibility Complex (MHC). Specific MHC alleles have been found to be associated with long term survival (Bateson et al. 2016), olfactory based mate choice (Evans et al. 2012), and even reproductive success (Sepil et al. 2013b), making it a prime target for studying adaptation and speciation at the population level .The Great Lakes region in the western zone of Nicaragua (Figure II.1), provide an ideal setting to study speciation. The numerous crater lakes of this area provide a 'natural experiment' for studying the process of recent population divergence. These lakes were formed as a result of volcanic activity dominant in the region between the late Tertiary and early Quaternary eras (Bussing and Rica 1976).

Lake Managua (also called Xolotlán) is estimated to be > 500,000 years old. It is the smaller of the two Nicaraguan Great Lakes. The watershed is six fold the lake area, which is approximately 1,042 km² (58.4 km long and 32.7 km wide) with a mean depth of 7.8 m and maximum depth of 26 m. It experiences seasonal fluctuation of water level, with considerable flooding during punctual severe hurricanes. Raw sewage from over one million people settled around the southern shore as well as industrial effluents from some 300 small enterprises also find their way into the lake water. The most striking features of this lake are its shallowness in relation to its great extension, and the turbidity of the water since there is deep mixing of the entire water column occurring thought the year (Montenegro-Guillen 1991). The water of this lake is alkaline (541.75 $Mg.I^{-1}$), has a homogeneous distribution and usually high concentrations of dissolved oxygen throughout the water column despite high influx of organic matter (Montenegro-Guillen 1991). Its unstable sediments do not allow the establishment of a diverse and abundant benthic community (Bijlmakers and Sobalvarro 1988).

Lake Nicaragua (also known as Cocibolca) is a tectonic lake located adjacent to Lake Managua. Like Lake Managua, it is also estimated to be > 500,000 years old. It is the largest lake in Central America, and the 19th largest lake in the world, with a surface area of around 8,000.00 km^2, mean depth of 13m, and

max depth of 45 m. This lake is somewhat contaminated, as most sewage in the catchment area is drained to the lake and solid waste is placed untreated in landfills which can leach into the lake. However, unlike L. Managua, L. Nicaragua is considered to have only minor pollution from heavy metals, except for a few pollution hotspots (Scheibye et al. 2014). Sizeable waves driven by the easterly winds blowing west to the Pacific Ocean cause the fine sediment to mix and the water to become murky. Lake Nicaragua has two connected volcanic islands, Ometepe and Zapatera, as well as an archipelago, Islas Solentiname, in the south. Motor boats and ferry's are the main mode of transportation in these islands. Lake Nicaragua is seasonally connected to Lake Managua threw the R. Tipitapa and drain into the Atlantic Ocean through the River San Juan.

Of the four crater lakes we sampled for this study, **crater Lake Asososca León** is the smallest (surface 0.81 km^2) and shallowest with an average depth of 17.2 m and maximum depth of 35 m. It's age is unknown, although genetic data suggest that it is several tens of thousands of years old (Barluenga and Meyer 2010), it is relatively clean with well-preserved vegetation on its caldera slopes. As with the other three crater lakes in this study, motor boats are not permitted due to contamination concerns. The youngest of the lakes is **crater Lake Xiloá**, with an estimated age of 10,000 years, and a surface area of 3.75 km^2 with average depth of 60 m and a maximum depth of 88.5 m. In a comparative chemical test, Lake Xiloá presented many fold more sodium than any other lake tested (Barlow et al. 1976). **Crater Lake Masaya** is relatively old, dated between 25000-100000 years old, and has a surface area twice as big as crater L. Xiloá (8.38 km^2), with similar depth (average depth 40.7 m, maximum depth 72.5 m; (Bice 1985; Kutterolf et al. 2007; Pérez et al. 2009). Parts of L. Masaya are littered with garbage that washes into it from the nearby settlement and accumulates due to it having no water outlet (see FigII.1). It is the most eutrophic of the crater lakes having continual blooms of planktonic algae due to its high concentration of soluble nitrogen. This usually limits visibility to one meter or less (Barlow et al. 1976). **Crate Lake Apoyo** is by far the largest (surface area ~21.1 km^2) and deepest of the Nicaraguan crater lakes, with an average depth of 142 m and a maximum depth estimated of > 200 m, and has been aged by geologists at < 23,000 years old (Kutterolf et al. 2007). This is the most oligotrophic lake, which in contrast to crater L. Masaya, shows virtually no soluble nitrate or nitrites. This lakes has little or no vegetation or sediment (Barlow et al. 1976). Although crater L. Apoyo is a popular tourist destination, it is also a reservoir for clean drinking water for the adjacent populations, and it is generally well managed (Guillén Bolaños 2006). All four of these crater lakes are very well isolated from other bodies of water (Cole 1976; Waid et al. 1999), making them akin to islands to the fish that inhabit them.

Figure II.1. Three contrasting examples of crater lake environments. On the left, an image of the shore line of crater L. Masaya littered in trash and human waste with areas of mild eutrophication. In the middle the shoreline of crater L. Asososca León covered with vegetation. On the right the shore of crater L. Apoyo covered in sand and stones showing relatively clear water.

The Midas cichlid species complex in the Nicaraguan lakes is an ideal model system for the study of fast speciation. Several crater lakes have been recently colonized by fish, and small scale radiations of the Midas cichlid have taken place confined to the lakes evolving several sympatric sister species (e.g., McKaye et al. 2002; Barluenga et al. 2006, ref elmer). Many species pertaining to this species complex have been described over the years, and in some lakes there are morphotypes that are yet to be described as new species. *Amphilophus citrinellus* (Günther 1864) is a widespread species considered to be the ancestral form, typically found in the great lakes Managua and Nicaragua; *A. labiatus* (Günther 1864) is only distributed in the great lakes, although thick-lipped morphs exist in several crater lakes (e.g., McKaye et al. 2002; Elmer et al. 2010b). Within crater L. Apoyo several endemic species have been described including *A. astorquii*, *A. chancho* (Stauffer et al. 2008), and *A. zaliosus* (Barlow and Munsey 1976). Also in crater L. Xiloá several endemic species have been described, including *A. amarillo*, *A. sagittae*, and *A. xiloaensis*. The other two crater lakes studied, Masaya and Asososca León, have not been studied in depth yet, and for clarity we classify all individuals collected in Masaya as *A.* sp1. *aff citrinellus* and *A.* sp1. *aff labiatus* (if the fish presented enlarged lips), and all individuals collected in crater L. Asososca León as *A.* sp2. *aff citrinellus* (in this lake there are two distinguishable morphs, a benthic-like high bodied morph abundant near the shore, and a limnetic-like slender bodied morph present in the deeper zone).
The Midas cichlid complex is a very recent radiation. The newly formed species are considered early speciation stages (Wilson et al. 2000), and it is known that they sporadically hybridize (Geiger et al. 2010). By studying the isolating mechanisms that keep these incipient morphotypes apart we can learn something about the multiple barriers preventing extensive gene flow among them.
Newly formed cichlid species remain isolated in some cases by pre-zygotic behavioral mechanisms (McKaye 1986; Rogers and Barlow 1991; Rogers 1995; van Staaden and Smith 2011; Selz et al. 2014). Female mating preferences might use visual and olfactory cues to choose males (Elmer et al. 2009; Selz et al. 2014). Major Histocompatibility Complex (MHC) genes are thought to influence olfaction and indirectly mating preferences.

Incipient species in the Midas cichlid complex are known to differ in the habitats they use (Barluenga et al. 2006; Barluenga and Meyer 2010; Kautt et al. 2016). Those habitats are exposed to different parasite communities, and this environmental pressure might drive a host race affecting immune genes. MHC variants might be linked to specific habitats, and enforce the speciation process. Species of the Midas Cichlid have a very diverse array of MHC class B alleles (Hofmann et al. 2017). Insights about the levels of diversity of MHC in different Midas cichlid populations within and among several lakes may help to better understand the speciation processes acting on the Midas Cichlid. Here we characterize the diversity of MHC, the genes responsible for triggering the adaptive immune system, in several populations of the

Midas cichlid living six Nicaraguan lakes. We evaluated the hypothesis that adaptation to local environments will have direct effects on the array of MHC alleles, and this will ultimately affect population recognition, and eventually be linked to the speciation mechanisms.

Methods

Sampling

Sampling was carried out over four weeks in December during two consecutive years, 2009 and 2010. Samples were collected at six locations: the two great lakes, Managua and Nicaragua, and four crater lakes, Asososca León, Xiloá, Masaya, and Apoyo. Fish were caught with gill nets and anesthetized with Tricaine mesylate (MS-222). A picture was taken for each individual in a standardized position for later identification and morphometric analyses, and fish were measured and weighted. Fish were euthanized on ice and fin clips were taken and preserved in 100 % ethanol for DNA analyses. Fish were dissected and sexed, and pharyngeal jaws were excised for species identification. Internal organs were preserved in RNA latter for expression analyses.

DNA extraction

Total genomic DNA was extracted using DNeasy spin columns for Blood and Tissue Kit® (Qiagen, Hilden, Germany), according to the manufacturer's protocol, with the addition of RNAse. DNA was quantified using a Nanodrop 1000 (ThermoFisher Scientific, Bonn), then standardized to a concentration of 20 ng / µl. And re-quantified using fluorometric quantitation by Q-bit (ThermoFisher Scientific, Bonn). Samples were run in a 1 % agarose gel at 70 v for 30 min, to evaluate DNA quality. All samples were confirmed to be of good quality before proceeding.

Figure II.2. Map of the area studied showing the geographic position of the two tectonic large lakes - Managua and Nicaragua- and the four crater lakes sampled, Asososca León, Xiloá, Masaya, and Apoyo.

Table II.1. List of the individuals used in this study per lake and species/morphotype, each associated to its preferred habitat and body morphotype.

Lake	Habitat zone	Morphotype	Individuals
Managua	**Tectonic Lake**		**24**
A. citrinellus	shallow unspecific	High-bodied	20
A. labiatus	shallow Rocky	Big lipped	4
Nicaragua	**Tectonic Lake**		**49**
A. citrinellus	shallow unspecific	High-bodied	24
A. labiatus	shallow Rocky	Big lipped	25
Apoyo	**Crater lake**		**62**
A. astorquii	litoral benthic	High-bodied	20
A. chancho	Profundal benthic	High-bodied	21
A. zaliosus	Limnetic	Long-bodied	21
Xiloa	**Crater lake**		**64**
A. amarillo	litoral benthic	High-bodied	22
A. xiloaensis	Profundal benthic	High-bodied	13
A. sagittae	Limnetic	Long-bodied	29
Masaya	**Crater lake**		**45**
A. sp. 1 *aff. citrinellus*	Sandy or muddy	High-bodied	27
A. sp. 1 *aff. citrinellus*	Rocky	Big lipped	18
Asososca León	**Crater lake**		**43**
A. sp. 2 *aff. citrinellus*	litoral benthic	High-bodied	23
A. sp. 2 *aff. citrinellus*	Limnetic	Long-bodied	20
		Total	**287**

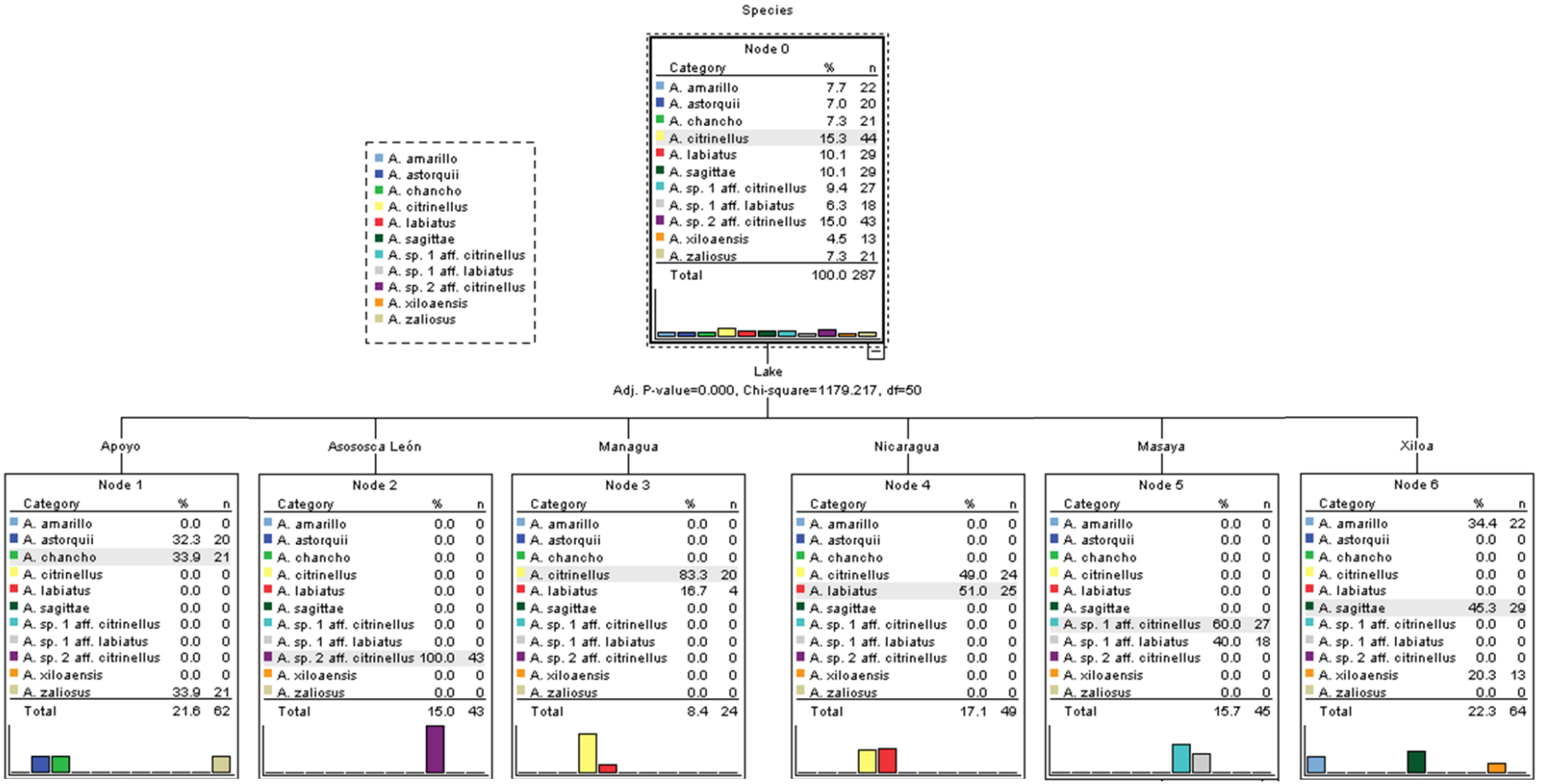


Figure II.3. Description of the composition of species/morphotypes per lake included in the analysis. The number of individuals and proportion per species/morphotype is given.

Primers design

We set a forward primer of 33 bp in a very conserved area of exon 2, downstream from the intron 1 exon 2 boundary (AcMHCIIBF5, 5´CCACKGAGCTGAASGACATSGAG 3´). After several tests (Hofmann et al. 2017) this primer showed to be very efficient in amplifying all previously identified alleles, excluding specifically those putatively non-classical. To set a reverse primer we made an alignment of all previously characterized 69 MHC alleles (Hofmann et al. 2017). Since there was no area conserved enough to place a single reverse primer neither downstream exon 2, nor in the exon 2 - intron 2 border, we established two different primers (AcMHCIIBR10, 5´GCAGTAYNTCYCCYTCTGAG 3´ and AcMHCIIBR11 5´GCAGWMTSTCTCCTTYKCAG 3´), which recovered the majority of all previously identified alleles, with the exception of DXB*14 and DXB*15 that presented three nucleotide differences at the primer binding cite.

Massive Sequencing

We performed a first round of PCRs using the MHC-specific primers, tailed with 20 bp derived from standard Illumina adaptors. In a second round of PCRs we included a set of 96 dual index combinations of full-length Illumina adaptors onto the amplicons. The final PCR was done using standard Illumina PCR primers in emulsion on pooled second round PCRs to achieve a high yield of amplicons while at the same time minimizing artifacts like chimeric amplicons. The PCRs included 10-80 ng of DNA extract, 15 pmol of each forward and reverse Primer in 20 µl Volume of 1 x MyTaq buffer containing 1.5 Units MyTaq Polymerase (Bioline) and 2 µl of BioStabII PCR Enhancer (Sigma). PCRs were carried out for 20 cycles using the following parameters: 2 min 96°C predenaturation; 96°C for 15 sec, 50°C for 30 sec, 70°C for 60 sec. Primers from first round amplifications are removed by standard ExoI digestion (NEB) for the second round amplification, the PCRs included 2 µl of ExoI digested first round PCR product, 10 pmol of each forward and reverse primer with adaptor including 8 bp Index in 20 µl volume of 1 x MyTaq buffer containing 1.5 Units MyTaq Polymerase (Bioline) and 2 µl of BioStabII PCR Enhancer (Sigma). PCRs were carried out for 3 cycles using the following parameters: 1 min 96°C predenaturation; 96°C for 15 sec, 50°C for 30 sec, 70°C for 60 sec, followed by 12 cycles of 96°C for 15 sec, 70°C for 90 sec. For pooling and size selection, 5 µl each of 96 second round PCR products were pooled and purified using 1 Vol AgencourtXP beads (Agilent). An emulsion PCR was then set up using buffers and enzymes from the emPCR Kit (Roche 454). The Micellula DNA Emulsion & Purification Kit (EURx) supplied the oil-surfactant mixture for creating the emulsion as well as columns and buffers for post-emPCR purification. PCR water phase: 5 µl amplicon pool, 20 µl 5 x Amp Mix, 40 µl emPCR Additive, 20 µl BiostabII PCR Enhancer (Sigma), 4 µl BSA (10mg/ml, NEB), 4 µl TruSeq primer (50 pM / µl each), 5 µl emPCR

enzyme Mix, 0.2 µl Ppiase. Oil surfactant phase consisting of 440 µl Emulsion Component T-1, 40 µl Emulsion Component T-2, 120 µl Emulsion Component T-3. 100 µl PCR water phase and 600 µl oil surfactant phase were combined in a 1.5 ml reaction tube and vortexed for 5 min to generate a stable emulsion. Aliquots of 100µl were distributed for PCR. PCR conditions were: 1 min at 93°C followed by 15 cycles 93°C for 30 sec; 58°C for 30 sec; 68°C during 1min. The breaking of the emulsion and subsequent DNA purification were done according to the Micellula DNA Emulsion & Purification Kit manual (EURx). The final library purification was done by size selection on LMP agarose gel. The PCR products were then sequenced using Illumina Miseq 5 million read run at LGC Genomics (Berlin, Germany).

Data pre-processing workflow

Demultiplexing of all libraries for each sequencing lane was done using the bcl2fastq 1.8.4 software (Illumina Inc. 2017©). One or two mismatches (further on 'N') were allowed in the barcode reads when the barcode distances between all libraries on the lane allowed for it. Reads were sorted by the amplicon inline barcodes with one mismatch allowed per barcode, the barcode sequence was clipped from the sequence after sorting. Reads with missing barcodes, one-sided barcodes, or conflicting barcode pairs were discarded. All reads with a final length of < 100 bases were discarded as they are likely remnants of sequence adapters. We removed all reads containing more than one N, and we trimmed the reads at the 3´end to get a minimum average Phred quality score of 20 over a window of 10 bases. For primer detection and clipping, three mismatches were allowed per primer and pairs of primers. All sequences that did not contain both primers (F-R or R-F) were discarded, and if primer-dimmers were detected, the outer copies were clipped from the sequence. The sequences were forward-reverse oriented after removing the primer sequences, the forward and reverse read pairs were then combined using BBmerge 34.48 (http://bbmap.sourceforge.net/).

Allele calling

For unique sequence clustering and erroneous sequence filtering we used a web server of the Amplicon Sequencing Analysis Tools 'AmpliSAT' (http://evobiolab.biol.amu.edu.pl/amplisat/index.php) that were designed for multi-gene families such as MHC (Sebastian et al. 2016). It determines whether a given cluster is a true allele or an artifact based on within-sample frequency and amplicon depth, and discards sequences that are chimeras of other major sequences. We first used the AmpliCHECK tool in order to optimize parameters for accurate allele calling. Then since AmpliSAS has a limit of 5000 reads per individual and our total filtered and combined reads averaged 14,500 reads per individual, and some individuals had up to sixty-one thousand reads, we did a random selection of 5000 reads per sample, five

different times. We then separately ran these five replicate in AmpliSAS with the default parameters, that are optimized for Illumina sequencing technology, except that we set max alleles per individual at 50, and minimum amplicon depth to 1000. We used the AmpliCOMPARE tool to compare the results of the five replicates. Any alleles that did not appear in a given individual in at least three of the five replicates were discarded. We numbered the alleles following the sequence a1 to a146 in order of abundance in a descendent manner. We finally did a BLASTx search on NCBI´s Genbank database to confirm that all alleles matched MHC.

Phylogenetic reconstruction and network analysis

To determine the phylogenetic relationships between alleles, we constructed a phylogenetic tree using Bayesian inference with BEAST v2.0 (Bouckaert et al. 2014). We found the that most appropriate substitution model was HKY+G, and the appropriate partitioning scheme according to the Bayesian Information Criterion implemented in PARTITIONFINDER V2 (Lanfear et al. 2014). We specified the parameters in BEAUti v2.0 (Bouckaert et al. 2014) for the BEAST-run, and the MCMCs were run for 10^9 generations sampling every 100,000 trees. We used a strict clock model and a Yule speciation prior. We inspected the traces for convergence with TRACER V1.5 (http://tree.bio.ed.ac.uk/software/tracer/) and checked that they were higher than 200. The 10,000 resulting trees were summarized with TREEANNOTATOR v2.1.2 (http://beast.bio.ed.ac.uk/treeannotator) applying a 10 % burn-in. We depicted the phylogeny with the corresponding posterior values of each node with FIGTREE v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). We also reconstructed the phylogeny of the MHC alleles including our own previously published allele sequences for this species (Hofmann et al. 2017).

To further analyze the relationship between the alleles we constructed a neighbor-net network with SPLITSTREE4 (Huson and Bryant 2006) using a pairwise distance matrix constructed with the best-fit nucleotide distance model computed in MEGA V7 (Kumar et al. 2016). The Network analysis was performed also for the two sets of data, the alleles derived from this study, and the previously described alleles for the Midas cichlid. To establish allele connectivity and cluster alleles by similarity, which may indicate functional allele groups, we also built a fully connected network with the same distance matrix of the 146 alleles. Each link in the network joining pairs of alleles and was labeled with the corresponding nucleotide distance. The links with the largest genetic distance were removed in a stepwise manner until the structure of the network fragmented into small clusters at the percolation threshold (see Stauffer and Aharony 1994; as implemented in Rozenfeld et al. 2008). Once we conducted the percolation analysis, we transformed the link weight to similarity values between 0-1 and used infomap (Rosvall and Bergstrom 2008) to detect network communities (groups of alleles that are more similar among themselves then

others). Infomap was run 1000 times as implemented in the R pakage Igraph (Csárdi and Nepusz 2006) and then we selected the partition with the shortest code length.

Test of selection

To infer the putative ABS of the 146 MHC IIB alleles, and to determine if selection has acted differently on the different groups of alleles, we did a site-specific positive selection test within each of the four groups of alleles, and across all alleles. This was implemented using CodeML in PAML v4.7 (Yang 2007), assuming different ω parameters among codons with no *a priori* knowledge of which class of selection (neutral, purifying, or positive) a given codon belongs to. We estimated parameters under five different codon substitution models: Beta models M7 (no positive selection), and M8 (positive selection) M8a (no positive selection and $\omega = 1$), and models M1a (nearly neutral) and M2a (positive selection) (Yang 2007). Then we performed a likelihood ratio test (LRT) to compare the best-fit models with and without selection. Statistical significance was determined by twice comparing the difference of the log-likelihood scores ($2\Delta\ln L$) to the X^2 distribution with degrees equal to the difference in the number of parameters between the models to be compared. To visualize the peptide and sequence motif conservation and variation at ABS´s we used a calculation of block entropy of the 146 aligned alleles as described in (Olsen et al. 2013).

Statistical Analyses

For all lakes and species samples, we calculated overall polymorphism of the alleles, and the minimum, maximum, and mean number of alleles per individual, as well as the nucleotide diversity per lake. We did a test of equality of group means, and calculated the Wilks' Lambda score for each allele. We also calculated Fisher´s linear discriminate function coefficients to determine if there is a pattern in the classification of alleles by lake. For each of the allele supertype, we did a Chi-square test to determine whether there is a significant difference in association between the allele groups and the lake is present in. We also reported the allele count and frequency of each allele group per lake.

Results

Massive sequencing of MHC PCR amplicons for 287 individuals with Illumina MiSeq resulted in 4 million reads. After cleaning, clipping and preliminary quality filtering we obtained a total of 2,890 clusters. After filtering the clusters of sequences using strict parameters that avoid intrinsic sequencing errors common to multilocus genotyping, these clusters collapsed in 146 distinct alleles. Alleles were named according to their abundance in the samples in descendent order (a1-a146). One of these alleles corresponded to non-classical MHC alleles, despite using primers that specifically excluded this group of alleles.

Table II.2. Descriptive information about samples used, and diversity values found per individual by lake. (SD. standard deviations)

	Nicaragua	Managua	A. León	Apoyo	Xiloá	Masaya	Total
No. individuals	49	24	43	62	64	45	**287**
No. alleles H	61	61	27	44	65	47	**146**
Average No. alleles	6.7 (3-10)	7.5 (6-10)	6.0 (4-8)	6.5 (4-10)	7.6 (5-10)	6.8 (4-9)	**6.8 (3-10)**
Average No. MHC supertypes	5.7 (3-8)	6.0 (5-8)	5.4 (4-7)	5.4 (4-8)	6.3 (5-9)	5.7 (4-9)	**5.7 (3-9)**
Variable sites S	93	92	90	91	95	97	**98**
Nucleotide diversity per site Pi	0.221	0.213	0.231	0.213	0.22	0.22	**0.20**

Alleles were 142 bp. We found some variation in exon length, with two alleles ‘a90’ and ‘a143’ with 3 and 1 nucleotide less respectively. There were 98 variable sites, and nucleotide diversity per site per lake ranged between 0.231 (in crater L. Asososca León) and 0.213 (in both lakes Apoyo and Managua) (Table II.2). Eight alleles were present in all 6 lakes and 75 alleles were exclusive to single lakes (Fig. II.4). One allele (a1) was present in all 287 sequenced individuals. Average number of alleles per individual was 6.8, with a minimum of 3 and a maximum of 10 alleles per individual (Table II.2). Crater L. Asososca León had the lowest total number of alleles (27), and the lowest number of alleles per individual (6), while lakes Xiloá, Managua and Nicaragua had the largest total number of alleles (Table II.2). There were no differences among lakes in average number of alleles or average number of supertypes per individual (P = 0.99).

Phylogenetic reconstruction

The phylogenetic relationships of the 146 MHC alleles with the addition of the 69 previously characterized alleles (see Hofmann et al. 2017), revealed 12 well supported clusters. Clusters were named in correlative order according to size. Five of these clusters were already recovered in previous work. The largest group was composed of 55 alleles, and four groups were composed of just one allele. The allele in Group 11 corresponded to a putative non-classical allele (according to Hofmann et al. 2017). Eight alleles were present in all lakes, and 6 alleles present in all species, showing that TSP is widespread within the Midas species complex. About half (75) of the alleles were unique to a single species. All 12 clades are well supported, the nodes that composed the phylogenetic groups had high posterior probability values (0.97-1), except Group 2 that were split by Group 8. However, most of the other ancestral nodes do not have statistically significant support values, this is likely due to recombination events between the different MHC allelic lineages witch is common in MHC (Nei et al. 1997).

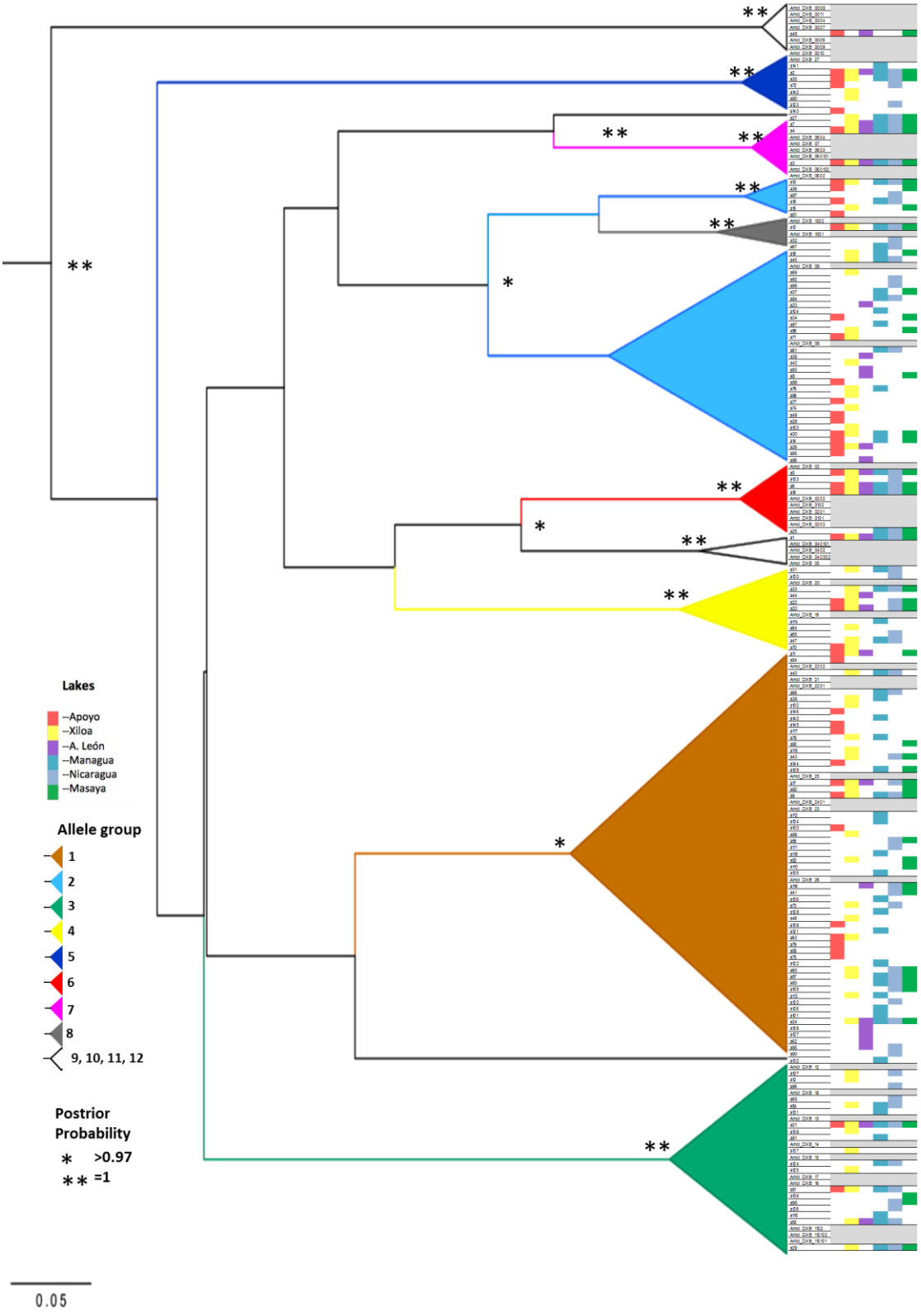


Figure II.4. Phylogenetic reconstruction of all 146 alleles recovered in this study plus the 96 alleles of chapter 1. ** Indicates posterior probability =1, and * indicates posterior probability > 0.97. Colored squares reflect presence / absence of each allele in each lake. Alleles from chapter I (Hofmann et al. 2017) are shown in gray.

Test of Selection

The analyses of site-specific selection revealed that alleles of all supertypes combined found sites that fitted best the model that allow for positive selection (p <<<0.0001), The test could not be done on all groups separately due to small number of alleles per supertype. We tested all alleles together and found 9 positively selected sites that were found comparing both models (M2a vs M1a, and M8 vs M7), an additional 4 sites were found only by M8 vs M7, most of which were inferred at 99 % level (shown in bold in Table II.#). We did not test for selection on supertypes or groups separately as some are formed by too few sequences to perform this type of test. The sequence logo plot of the residues frequencies' per site visually displayed the aminoacid frequencies per site, where site 4, 11, 14, 17, 20, 23, 24, and 29 are invariant conserved sites that are shared with all alleles. Sites 13 and 42 are the most variable site by far and are under positive selection.

Table II.3. Summary of likelihood ratio tests for site-specific positive selection of MHC IIB genes comparing all alleles

model comparison	2Δℓ	p value	proportion of PSS	ω > 1	PSS
M2a vs M1a	175.484938	<<<0.0001	0.28230	4.02553	**1**, **3**, **13**, **22**, 28, **35**, **42**, **45**, **46**
M8 vs M7	145.409038	<<<0.0001	0.29293	3.88988	**1**, **3**, **5**, 10, **12**, **13**, **22**, **28**, **35**, 40, **42**, **45**, **46**
M8 vs M8a	127.622730	<<<0.0001			

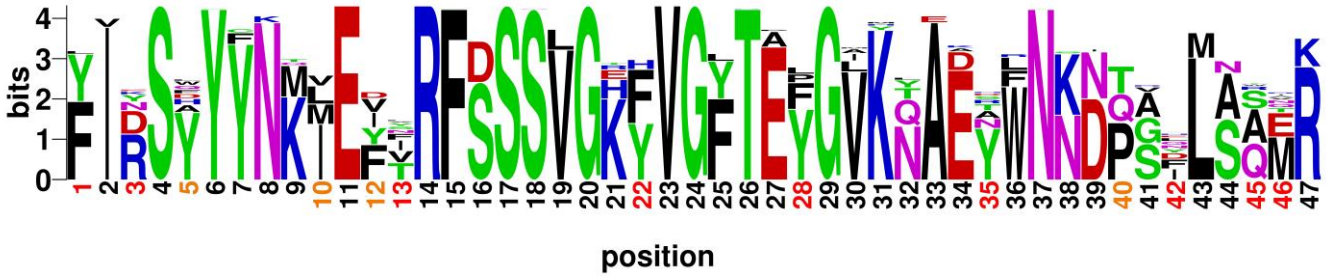


Figure II.5. Sequence logo plot of the frequency of the 47 aminoacid residues of all 146 sequences computed using WebLogo, The height of the logo element represents the sequence motifs log-transformed frequency displayed as bits of information. Low-frequency motifs are not displayed due to their heights being below useful resolution. Positions in orange indicate sites that are under positive selection only when comparing M8 vs M7, sites that are in red are under positive selection with both model comparisons. A logo plot for each supertype separately is shown in Sup. Fig II.3.

Definition of MHC supertypes

To attempt to identify functional groups of alleles, amino acid sites under positive selection as identified above were used to support the cluster analysis and to define Supertypes (defined in Doytchinova and Flower 2005) according to their genetic similarity and shared ABS inferred by positive selection. However, as we are not able to assign each allele to specific loci we take caution in the prediction antigen binding sites so we base our grouping on overall genetic disimilarity. To this end we also performed a neighbor-net Network analysis using a pairwise distance matrix constructed with the best-fit nucleotide distance model (Fig. II.6). Allele connectivity was furhter established with a percolation analysis which grouped the most similar alleles by linking all 146 alleles with their pairwize distance value, then it removes the most distant connections in a pairwize manner untill percolation threshold of genetic distance at which the alleles collapsed into discrete groups was 0.126 (Sup. Fig. II.2).

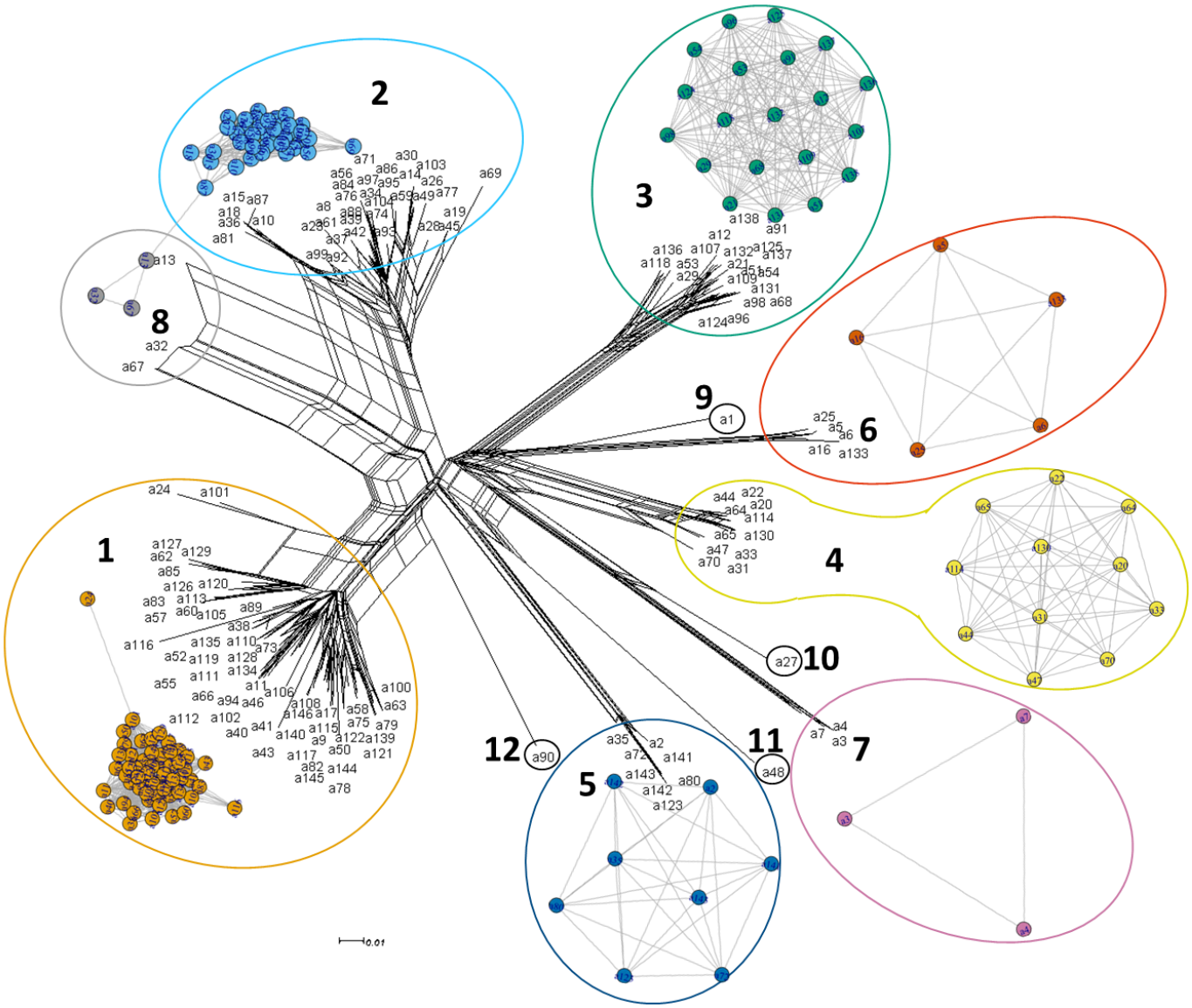


Figure II.6. Neighbor net network analysis and the percolation threshold showing which clusters of alleles corresponded to MHC supertypes.***** indicate non-classical MHC.

According to this analysis we can clasify all MHC alleles in 12 distinct goups, 11 of which we asign to supertypes. We explude group 11 from the supertipes, as we have previusly determined that it belongs to non-classical MHC (Hofmann et al. 2017). In comparison with the phylogenetic analysis this analysis additionally separated groups 2 and 8 undifferentiated in the previous result. The largest supertype (1) was formed by 55 alleles. Two supertypes were formed by 3 alleles (7 and 8). Three supertypes were formed by a single allele. One of them (Group 9, a1) was present in every single individual analyzed. This allele can be considered as a 'MHC Core allele'. The other 2 groups fomed by a single allele were generally rare in all lakes, and Group 12 was exclusive of L. Nicaragua. All individuals in all lakes (with 2 exceptions,

one individual in L Nicaragua and another in crater L. Masaya) had at least one allele from supertype 5 (with 8 alleles), which we consider a 'MHC Core group'.

As mentioned before, average number of alleles per individual was between 6 and 7. Average number of MHC supertypes per individual was 5.7, and maximum number of alleles per supertype was 3.

Lake specific MHC alleles

To evaluate if MHC allele composition could distinguish the populations of the Midas cichlid in each lake we performed Fisher's linear discriminant functions. This analysis showed that the populations on crater lakes Xiloá and Asososca León were very distinct from the rest in terms of total allele composition. The list of the most discriminant alleles together with the classification function coefficient that is most important in each lake is shown in Sup. Table II.2. Descriptive diversity values per lake are shown in Table II.2.

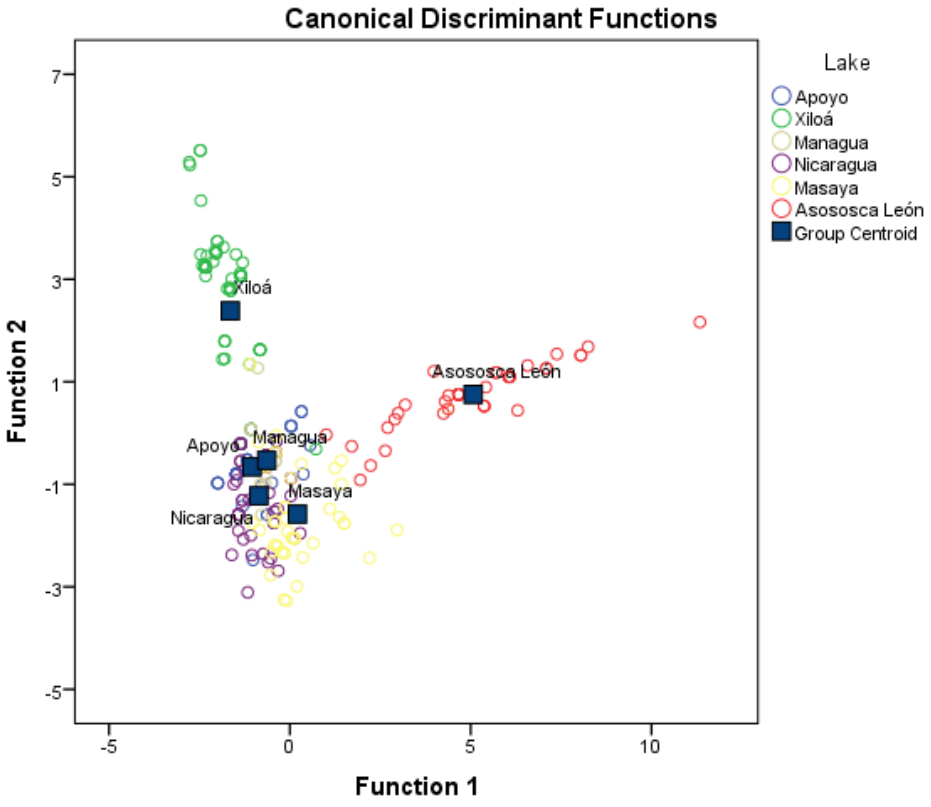


Figure II.7. Fisher's linear discriminant functions. Each color shows a different lake, and the blue square indicates the lake centroid.

For each of the allele supertypes, and non-classical group 11that we determined by the network analysis, we evaluated if there were differences in their abundances per lake. Figure II.8 show the proportion of presence/absence or alleles in each of the identified groups per lake. Chi-square tests were performed to evaluate statistical differences among groups (Sup. Tables II .5). Chi-Square tests were significant for all

allele groups except groups 1 and 5. A Chi-square test was not done on group 9 as it represents a constant with all individuals of all lakes having the allele present.

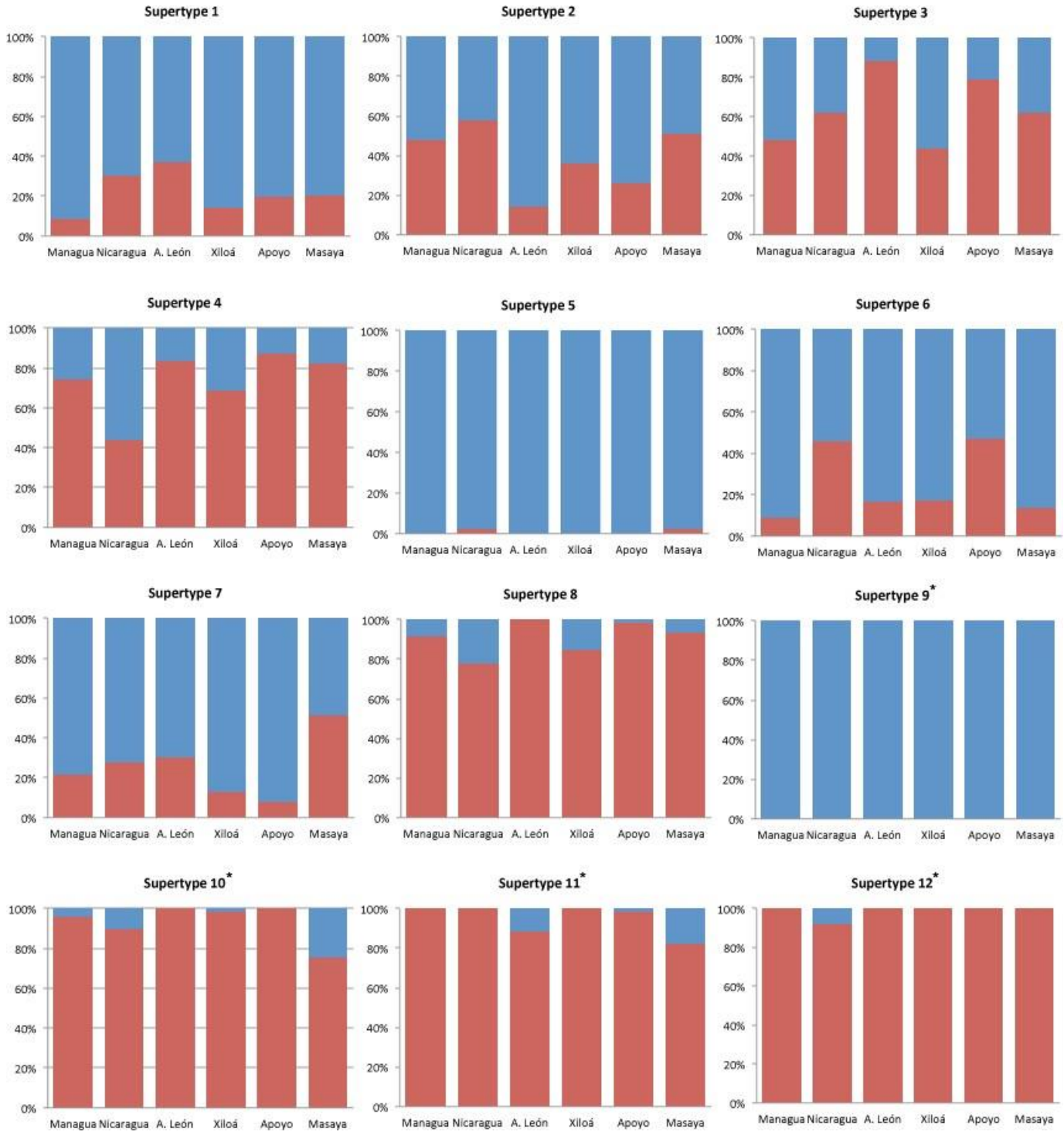


Figure II.8 proportion of individuals per lake with alleles from each allele cluster (supertype). Individuals with at least one allele present in allele group are in blue, and Individuals with no alleles in the allele group are red. * Indicate clusters formed by a single allele, Group 11 is non-classic MHC.

Discussion

This study aimed to test the idea that host-parasite interaction might facilitate the isolation of incipient species. Populations or eco-morphotypes of the Midas cichlid inhabiting diverging habitats will encounter different parasite communities, and this will induce changes in the host. We bid to address this question by studying population variation at the most polymorphic segment of the immune genes, MHC class IIB exon 2, responsible for triggering and mediating immune reaction that are parasite specific.
We found great variation in this sequence across all fish analyzed. We identified a total of 146 distinct alleles. Adding to our previous knowledge on this species (Hofmann et al. 2017) we can confirm 161 unique alleles within the Midas cichlid species complex. Fifty-five of the alleles recovered in Chapter 1 with cloning and Sanger sequencing on 13 individuals were common to the Illumina approach at the population level
We found some variation in sequence length Allele sequences were 142 bases long, with the exception of allele a90, which had three bases less, and allele a123, which had one base pair less. Classical MHC alleles have been reported before with three bp sequence length polymorphisms, the Atwater's prairie-chicken (*Tympanuchus cupido attwateri*) has well characterized alleles with 3 nucleotide extra from position 40 in exon 3 (Bateson et al. 2016). However, allele a123 is more likely a pseudogene and not a PCR artifact, as it appeared independently in more than one individual.
With massive sequencing we found a maximum of 10 alleles per individual, and an average of 6.8. In a previous study we found a maximum of 26 alleles per individual (implying at least 13 distinct loci), but the method used here is much more stringent, and therefore lower individual diversity was expected with the following considerations. In our previous study we sequenced 16 non-classical alleles, here, due to specifically designed primers we were able to exclude sequencing all but 1 of these non-classical alleles. The main reason however of finding fewer alleles in this study then we previously found is that many similar alleles have no nucleotide variation within the 142 fragment that we sequenced. For example, supertype 9 with a single ubiquitous allele, corresponds to Group 4b in Hofmann et al. (2107) formed by 12 alleles, which vary at in the downstream end of exon 2 as well as the intron, making it impossible to distinguish between these 12 alleles in this study as we have not sequenced the variable parts.

MHC Supertypes

The number of MHC alleles that an individual holds depends on the number of gene copies that characterizes the species. This is directly related to the capacity for detection parasites, but having several copies of MHC genes that are very similar does not add significant immunity against a wider set of

parasites. However, a growing body of evidence across taxa highlights the importance of grouping MHC alleles into supertypes (often defined based on functional properties) in order to classify alleles functionally into distinct functional groups (Trachtenberg et al. 2003; Huchard et al. 2010) With this reasoning we concentrated alleles according to their genetic similarity and found 11 supertypes excluding non-classical MHC (Fig. II.6). MHC supertypes were very uneven in the number of alleles that they contained. While some groups were very large, with more than 50 alleles, 3 of them only had a single allele (see Supplementary table II.2). Some supertypes were very frequent in all population, while some others were rare. A Chi-square test showed distribution of supertypes between lakes was significant for all supertypes except 1 and 5, which is likely due to them being present in high proportions in all lakes. All but one of the supertypes with a single allele were very rare in all populations. Supertype 12 was only found in L. Nicaragua (in both *A. citrinellus* and A. *labiatus*). Supertype 11 was never found in the large lakes, but appeared in several crater lakes (with the exception of Xiloá). Supertype 10 is absent in crater lakes Apoyo and Asososca León, however supertype 9 was not only present in all lakes and species, it was found in all individuals sequences, making it one of the core MHC alleles of Midas cichlids.

Core MHC supertypes

Some MHC supertypes appear to be critical for immune defense in the Midas cichlid while others were only occasionally found. MHC supertypes 5 and 9 are present in every single individual suggesting that they are essential alleles. These MHC Core supertypes have been maintained across multiple colonization incidents and demographic events. This signifies that they are likely essential to the survival of individuals in all lakes and habitats, and hence have been maintained throughout the speciation process further supporting TSP. The test of selection revealed that there are sites that are sites that are under positive selection, inferring that they are crucial in binding a specific range of antigens. The crucial sites are not only under selection, but they present aminoacids with very distinct and opposing properties.

Lake-specific MHC signature

Populations of the Midas cichlid in different Nicaraguan lakes clearly differed in their MHC signature. The crater lakes had typically lower MHC diversity, while the large lakes, and surprisingly also crater L. Xiloá had larger diversity. Supertype diversity per lake ranged between 5.4 in crater lakes Apoyo and Asososca León and 6.3 in crater L. Xiloá (min. 3, max 9). Interestingly, the genetic MHC fingerprint of each lake population was clearly differentiated (Fig. II.7). The most differentiated populations were those in crater lakes Xiloá and Asososca León. The other populations had a more similar allele repertoire. This is a very interesting result that shows that selective pressures in the different lake habitats do shape the immune response in the Midas cichlid. Information about the specific parasite communities in each lake

will aid our understanding of the relative effects of stochastic demographic processes in the MHC genetic fingerprint of populations. In any case, either differential selective pressures or stochastic effects provide a perfect setting for differentiation among populations in immune response, which could directly have consequences in mate recognition and species barriers.

Seventy tree alleles are species specific, however most of them are quite rare meaning that they could be present in the other alleles but a much larger number of individuals would have to be sequenced. Some alleles however are species or lake specific despite being present in a high proportion of individuals. Allele a8, and a23were found in 26, and 10 samples of *A.* sp. 2 *aff. citrinellus* despite not being present in any other species or lake. Allele a12 was unique to Lake Xiloá where it was found in 22 samples all belonging to *A. sagittae.* Total number of alleles per lake varied from 65 in Xiloá to 27 in Asososca León, however the average allele per individual, or supertype per individual, did not vary significantly between lakes.

Conclusions

This study shows that there is great diversity in MHC allele repertoire among the Midas cichlid populations in the different Nicaraguan lakes. There are some alleles that are present in every lake population, and even in every single individual, suggesting that they could have important adaptive value. However, there is a large body of MHC alleles exclusive of each population, indicating a important degree of local adaptation. Each Midas cichlid population has a different MHC signature. These results prove that the alternative habitats to which the Midas cichlid populations are exposed in the different lakes pose different selective pressures, which coupled with demographic events, confer them with a different repertoire of MHC alleles

Individuals in all studied populations had an average of 7 alleles, belonging to 6 supertypes or functional groups. Optimal MHC supertype per individual could diverge among lake populations, ranging from 5.4 in crater lakes Apoyo and Asososca León to 6.3 in crater lake Xiloá. Although total alleles per lake varied significantly, the alleles per individual did not.

CHAPTER III

Parallel evolution and MHC in crater lake cichlids

Manuscript in preparation

Abstract

Background: Parallel evolution in African cichlids is a classic textbook example with remarkably similar trophic morphologies and coloration.

Results: Here we report a striking case of parallel speciation in two small volcanic crater lakes in Nicaragua, Central America, over a period of no more than just a few thousand years. At least three ecologically and morphologically equivalent species of Midas cichlids (*Amphilophus* spp.) have evolved independently in each of two crater lakes Apoyo and Xiloá. Moreover, the speciation events follow the exact same succession in the two independent radiations, first giving rise to a limnetic species, and then splitting up the ancestral benthic form into a benthic species living in shallow water and a benthic species adapted to deeper zones. Parallel evolution not only involves overall body shape but also other eco-morphological adaptations such as the structure of the pharyngeal jaws, however the role of parasite-mediated selection is unclear, despite distinct MHC profiles. Stable isotope and stomach content analyses confirm that these species-pairs occupy equivalent trophic niches in these two crater lakes. Natural selection through ecological specialization to newly available environments is the likely force driving sequential parallel ecological speciation.

Conclusions: Speciation in two flocks of Midas cichlids in two independent crater lakes appears to have occurred mediated by natural selection through ecological specialization. Adaptation to novel available environments was associated to equivalent phenotypic adaptations in the two systems, and also divergent variation in the immune response measured as MHC variation. However, possible parallelism in parasite mediated selection in similar environments is still ambiguous and needs further study.

Introduction

Evolution sometimes repeats itself producing the same phenotypic adaptations in independent lineages exposed to similar environmental conditions (Kocher *et al.* 1993; Schluter & McPhail 1993; Schluter & Nagel 1995; Losos *et al.* 1998; Rundle *et al.* 2000; Nosil *et al.* 2002; Schluter & Conte 2009). These striking examples of parallel evolution emphasize the central role of natural selection in speciation (Schluter & Nagel 1995; Schluter 2009), but also support the notion of a deterministic fate of evolution (Gould 2002; Brakefield 2006). A classic textbook example of parallel evolution are cichlid fishes of the East African Great Lakes Tanganyika and Malawi, where remarkably similar trophic morphologies and coloration types have evolved from different founder lineages in a period of a few million years (Kocher *et al.* 1993; Salzburger 2008).

The parallel evolution of similar traits in independent lineages exposed to similar environmental settings implies that natural selection is a key factor causing phenotypic evolution (Schluter & Nagel 1995; Rundle *et al.* 2000; Schluter 2009). At the same time, the recurrent parallel evolution of arrays of morphologically and ecologically equivalent species is indicative of a tendency of natural selection to come up with similar phenotypic solutions when confronted with the same ecological problems: the same ecomorphotypes of *Anolis* lizards have emerged in replicate on isolated Caribbean islands (Losos *et al.* 1998); reduction of armour plates and transition to benthic body shape have occurred repeatedly in threespine sticklebacks (Cresko *et al.* 2004; Colosimo *et al.* 2005; Shapiro *et al.* 2006); the same mimetic warning wing coloration has evolved in parallel in different lineages of *Heliconius* butterflies (Kronforst *et al.* 2006); and equivalent trophic morphologies, body forms, pharyngeal jaws (a second set of jaws in the pharynx of cichlids that is used to process food), and colour patterns have originated independently in adaptive radiations of cichlid fishes in East Africa (Kocher *et al.* 1993; Salzburger 2008).

To maximize survival and reproduction in their local environment, organisms adapt to resist their local parasite communities. Locally adapted organisms are expected to have higher fitness in their resident environment compared with migrants from adjacent populations (Kawecki and Ebert 2004; Eizaguirre et al. 2012). Hence, parasite-mediated natural selection can play a crucial role in immunogenetic adaptation in the hosts as well as local adaptation of parasite infectivity and virulence (Kawecki and Ebert 2004). The Major Histocompatibility complex (MHC) is a highly polymorphic multi-gene family that is a part of the adaptive immune system. Its key function is to respond to pathogen infections in a targeted way (Janeway et al. 2005). MHC variation contributes to reproductive isolation and its role in mate choice and assortative mating has been demonstrated in many species, including three-spined sticklebacks (Aeschlimann et al. 2003; Milinski 2006; Eizaguirre et al. 2010), salmons (Dionne et al. 2007; Evans et al.

2012), rodents (Harf and Sommer 2005), and birds (Rymešová et al. 2017). Due to the dual role of MHC in fitness and mate choice, this trait has been considered a 'magic trait' *sensu* Gavrilets (2004) (Eizaguirre et al. 2009). This trait is under divergent selection and pleiotropically may affect mating decisions (Maan and Seehausen 2011; Thibert-Plante and Gavrilets 2013). Cichlids have one of the most diverse sets of MHC genes (Málaga-Trillo et al. 1998; Hablützel et al. 2014; Hofmann et al. 2017), which have been linked to parasite mediated mate selection and even reproductive isolation (Blais et al. 2007). The great diversity of MHC in many species, including cichlids, has been attributed to Trans-species polymorphism (TSP, Klein et al. 2007; Hofmann et al. 2017), a phenomenon where some alleles are passed from ancestral to descendant species with little or no change. Parasite composition and MHC allele variation has been found to evolve in a paralleled pattern in two morphs of three-spined stickleback *Gasterosteus aculeatus* (Natsopoulou et al. 2012). They found that MHC allele diversity was not consistent with diversity reported at neutral markers, but that the most common number of alleles in each morph was associated with lower parasite infection levels.

The Midas cichlid species complex (*Amphilophus* spp.) in Nicaragua, Central America, is an eco-morphologically diverse and dynamic biological system (Barlow 1976; Elmer *et al.* 2010a), which has proved to be a powerful model system for the study of ongoing speciation (Barluenga *et al.* 2006; Gavrilets *et al.* 2007; Barluenga & Meyer 2010; Elmer *et al.* 2010b). The distribution range of the Midas cichlid assemblage is located in the Central American portion of the 'Pacific Ring of Fire', and is characterized by the existence of two relatively large, old, and shallow tectonic lakes, Managua and Nicaragua, as well as several small, recently emerged, and comparatively deep crater lakes of volcanic origin (see Fig. I.1 in Chapter 1). Colonization of newly formed, hence empty, crater lakes and subsequent radiation of Midas cichlids has occurred repeatedly in the last few thousand years (McKaye *et al.* 2002; Barluenga & Meyer 2004; Barluenga *et al.* 2006; Barluenga & Meyer 2010; Elmer *et al.* 2010b), creating an ideal set-up to study parallel adaptive radiation. These lakes are ecologically distinct from the large lakes Nicaragua and Managua in that they are oligotrophic, with less abundance of phytoplankton and invertebrates, and have a more impoverished fish fauna, which potentially relaxes interspecific competitive interactions and predation pressure (Barlow 1976). Most importantly, the two crater lakes have a new ecological dimension, depth, which is absent in the large Nicaraguan lakes (average depth: L. Managua, 12 m; L. Nicaragua, 17 m; L. Apoyo, 142 m; L. Xiloá, 60 m). Ecological speciation is the likely mechanism accounting for the independent origin of new species of Midas cichlids in crater lakes Apoyo and Xiloá (Barluenga *et al.* 2006; Elmer *et al.* 2010b). However, the role of parasite-mediated selection on immune genes has not been studied. Since the crater lake species differ in eco-morphological traits used to

exploit the crater lake specific niches and resources, ecological specialization to new habitats and food types appears to be the driving force in the Midas radiations in lakes Apoyo and Xiloá.
Here we focused on the two Nicaraguan crater lakes with the eco-morphologically most diverse set of Midas cichlids, lakes Apoyo and Xiloá. Albeit slightly different in a few parameters (size: 5.5 *versus* 3 km diameter, max. depth: >200 *versus* 88.5 m, presumed age of origin: < 23.000 *versus* < 10.000 years), the two lakes resemble each other with respect to their volcanic origin, circular shapes, depth profiles, oligotrophic waters, and distribution of habitats suitable to cichlid fishes. Their Midas cichlid faunas resemble each other, too: Lakes Apoyo and Xiloá each contain at least three endemic *Amphilophus* species (McKaye *et al.* 2002; Stauffer & McKaye 2002; Stauffer *et al.* 2008), a benthic species living in the shallow area (BS), a benthic species inhabiting the deeper waters (BD), and a limnetic species (L) living in the open column (Fig. III.1 and Fig. III.2). Genetic analyses based on DNA sequences of the mitochondrial control region and several nuclear loci corroborate the genetic distinctiveness of species between the two lakes (Barluenga & Meyer 2010). Behavioural studies both in the field and under laboratory conditions report strong assortative mating between the species in each lake, on the basis of prezygotic barriers through differences in courtship behaviour in crater L. Apoyo (Baylis 1976; Oldfield 2009) and spatial segregation of breeding sites in Lake Xiloá (McKaye *et al.* 2002).

Here we uncover the role of ecological specialization, morphological traits, and parasite-mediated selection in the parallel evolution of very recent radiations of the Midas cichlid within small isolated crater lakes in the Neotropics. To this end we examined the morphology, ecology and genomic differentiation of the small-scale radiation of the Midas cichlid in the crater lakes Xiloá and Apoyo. In addition we analysed variation of MHC genes. We used geometric morphometric methods on body shape and lower pharyngeal jaw –LPJ, analyses of stomach contents and stable isotope analyses, Rad-tag sequencing, and My-seq Illumina sequencing of MHC.

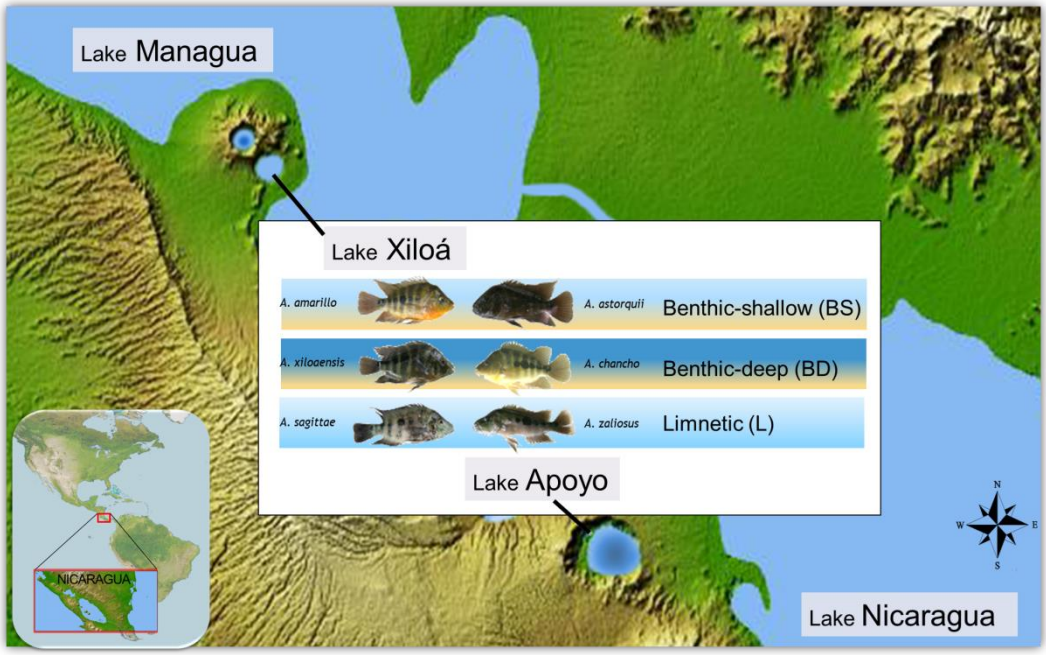


Figure III.1. Map of the two crater lakes Apoyo and Xiloá, with the three parallel species pairs.

Methods

Sampling and DNA extraction

We collected fish in two crater lakes, Apoyo and Xiloá, for four weeks during the breeding season in 2009 and 2010. We collected 126 adult individuals (62 in crater L. Apoyo and 64 in crater L. Xiloá) using gill nets. Fish were anesthetized with MS 222 following standard procedures and euthanized on ice before processing. A picture was taken for each individual in a standardized position for later identification and morphometric analyses, and fish were measured and weighted. We collected the entire digestive tract, and the lower pharyngeal jaw apparatus in 95% ethanol to later analyze gut content and jaw morphology in the laboratory. For stable isotope analysis from each individual we collected a piece of muscle from between the operculum and the dorsal fin, we preserved it in 95 % ethanol until further process. For molecular analysis we also collected fin tissues and preserved in in 100 % ethanol at 4ºC.

Total genomic DNA was extracted using DNeasy spin columns for Blood and Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol, with the addition of RNAse. DNA was quantified using Nanodrop 1000 (ThermoFisher Scientific, Bonn), and standardized to a concentration of 20 ng / µl. Then re-quantified using fluorometric quantitation by Q-bit (ThermoFisher Scientific, Bonn). Samples were run in a 1 % agarose gel at 70 v for 30 min to evaluate DNA quality. All samples were confirmed to be of good quality before proceeding.

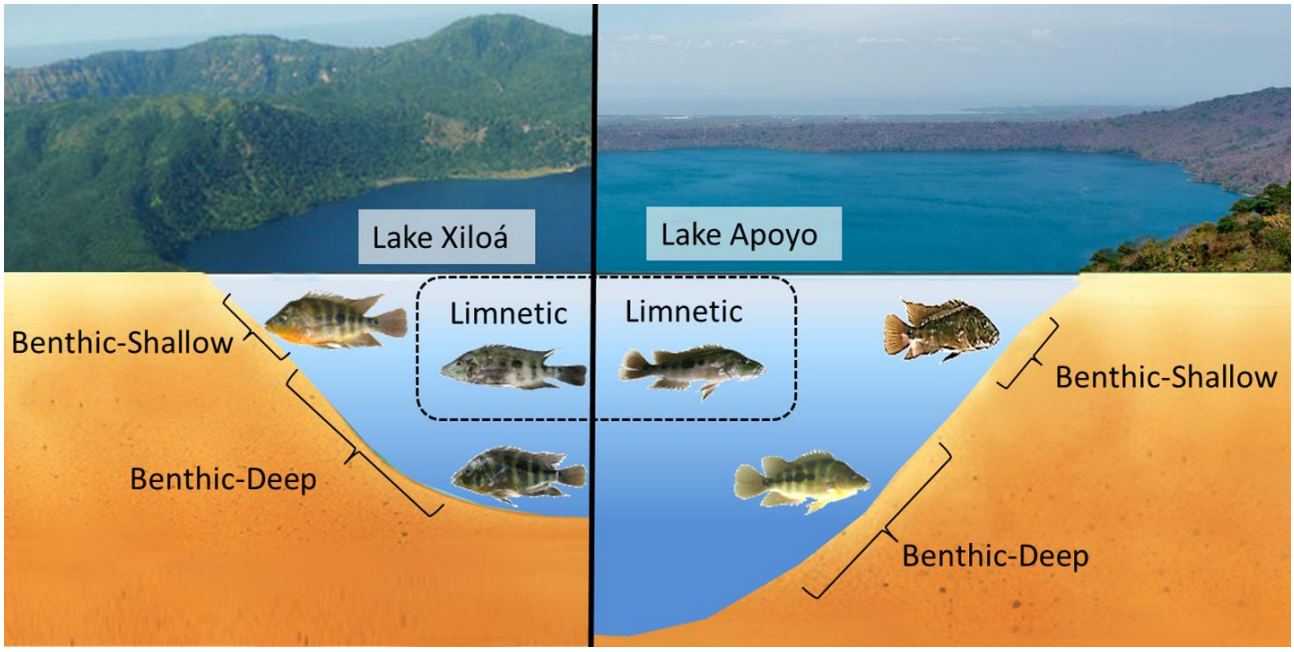


Figure III.2. Profile of the two crater lakes studied and distribution of the three parallel species.

Geometric morphometric analyses

We analysed variation in the shape of the fish and of a trophic structure, the LPJ. We cleaned all muscle and soft tissue from the LPJ, and scanned them to obtain digital images. We used the software TPSUTIL v1.56 (Rohlf, 2004) to transform the jaw scans and fish images taken in the field from .jpg into .tps format. To quantify body shape variation across individuals on the digital images of the fish, we digitized 15 homologous landmarks (Fig. III.2A) using TPSDIG v2.17 (Rohlf, 2001). Similarly, we placed 8 landmarks on the digital images of the LPJ to quantify shape variation in this trophic structure (Fig. III.2B). Landmark coordinates were exported and analysed using the software MORPHOJ v1.06b (Klingenberg, 2011). First, we performed a Procrustes superimposition (Dryden & Mardia, 1998) to extract shape coordinates for further analyses, in which the configuration of the landmarks for each specimen was scaled to a unit centroid size, translated to a common position and rotated to minimize Procrustes distances between all landmark configurations (Dryden & Mardia, 1998; Rohlf, 1999; Zelditch et al., 2004). We found a significant effect of size on body shape (3.16 % of variation predicted by size, $p < 0.001$), but not on jaw shape (0.38 % variation predicted by size, $p = 0.55$). Therefore, we performed a size correction in body shape analyses to account for allometric effects (Loy et al. 1996; Reis et al., 1998; Klingenberg, 2003) using a multivariate regression of individual Procrustes coordinates and size (using the centroid size), with a permutation test with 10,000 iterations. We performed Principal Component Analyses (PCA) in MORPHOJ for both body and jaw shape. This analysis sorted the individuals based on axes of maximum differentiation or principal components (PCs) according to the landmark coordinates. The PCA simplifies the variation patterns of the different morphometric variables replacing them with

PCs compiling multiple dimensions of variation (Zelditch *et al*., 2004). In order to statistically compare the established groups of body and jaw types we performed Canonical Variate Analyses (CVA) also in MORPHOJ with 10,000 iterations. This analysis reduces the variation within each group and quantifies the differences among them statistically. In this analysis the relative positions of the landmarks are grouped and processed in canonical variates (CVs) (Zelditch et al., 2004). The distance among groups was measured as Mahalanobis distance (M) with an associated p-value.

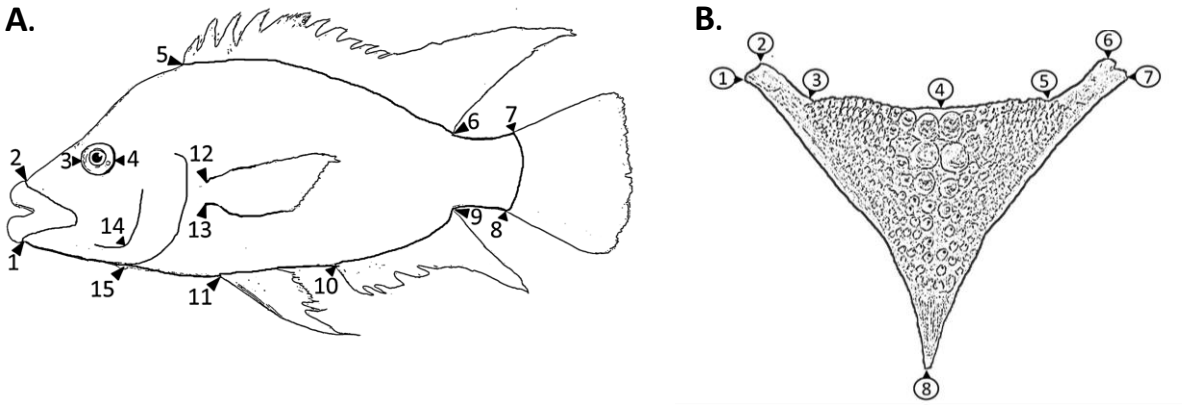


Figure III.2. Position of landmarks used for morphometric. **A.)** 15 homologous landmarks for body shape. **B.)** 8 homologous landmarks for LPJ.

Ecological analyses

The digestive tracts were dissected using a stereomicroscope, and the stomach and gut contents were assorted to five categories (see Sup. Fig. III.3). The volume of each category present in each individual was estimated, and used to perform analysis of similarity (ANOSIM) (Clarke 1993) of diet as implemented in the R package, and to calculate Schoener's index of diet overlap (see analysis in Barluenga *et al.* (2006) **(**values greater than 0.6 generally indicate biologically significant overlap).
An alternative quantification of the food spectrum consumed is to measuring rare stable isotopes (^{13}C and ^{15}N) accumulated in the fishes' muscle, for this analysis, all instruments were thoroughly cleaned with soap and water, then wiped down with hexanes to remove any organic contaminants and acetone to remove any humidity. The working area was covered with aluminium foil that was wiped clean with hexane before tissue were cut up in small pieces, placed in clean aluminium weigh boats and desiccated in a dryer oven at 60°C overnight. Once the samples were dry, they were individually ground into a fine powder with a clean pestle and mortar and 1 mg of it was weighed and placed in an aluminium tin capsule for analysis in the UC Davis stable isotope facility (California, USA). Stable isotope analysis was used to determine $\delta^{13}C$ and $\delta^{15}N$ of muscle tissue as described by Harrod *et al*. (2005). To account for ^{13}C depletion in lipid containing samples a revised model by Kiljunen *et al*. (2006) was employed. Isotopic

differences among species were estimated with non-parametric MANOVA (Anderson 2001) (Euclidean distance; 10 000 iterations) in PAST v1.75b (http://folk.uio.no/ohammer/past/). Significance levels were Bonferroni corrected.

MHC Genotyping

Amplification of MHC followed the same procedure as in Chapter 2. In brief one forward (AcMHCIIBF5) and two reverse (AcMHCIIBR10 and AcMHCIIBR11) primers were used to generate a library in a three step PCR. In a first round of PCRs we used adapter-tailed MHC-specific primers; in a second round PCRs included a set of 96 dual index combinations of full-length Illumina adaptors onto the amplicons; and the final PCR was done using standard Illumina PCR primers in emulsion on pooled second round PCRs to achieve a high yield of amplicons while at the same time minimizing artifacts like chimeric amplicons. The PCR conditions are detailed in Chapter II.

The final library purification was done by size selection on LMP Agarose gel. The PCR products were then sequenced using Illumina Miseq at LGC Genomics (Berlin, Germany). The library was demultiplexed using the bcl2fastq v1.8.4 software (Illumina Inc. 2017©). Sequence clustering and allele calling was done using the Amplicon Sequencing Analysis Tools "AmpliSAT" that were designed for multy-gene families such as MHC (Sebastian et al. 2016) .We first used the AmpliCHECK tool in order to optimize parameters for accurate allele calling. Then in AmpliSAS we ran 5 replicates of a maximum of 5000 randomly selected reads per individual with the default parameters, that were optimized for illumine sequencing technology, except, max alleles per individual was set at 50, and minimum amplicon depth was set to 1000. We used the AmpliCOMPARE tool to compare the results of the five replicates. Any alleles that did not appear in a given individual in at least three of the five replicates were discarded. We finally did a BLASTx search on NCBI´s Genbank database to confirm that all alleles were indeed MHC.

RAD sequencing

To obtain genome-wide genetic markers we did RAD-Tag massive sequencing. We prepared libraries of individually barcoded, restriction site-associated DNA sequencing (RAD seq) (Baird et al. 2008) following the protocol implemented in Roesti et al. (2012). For the restriction endonuclease digest we used 44.5 ul DNA (at 20 ng / µl), 5 µl NEB Buffer 4 (at 10 %) and 0.5 µl HF Sbf1 in a 50 µl reaction volume. We vortexed for 2 sec, centrifuge at 5000 rpm for 3-5 sec, and incubated at 37°C for 65 min then 65°C for 20 min at 350 rpm. For the P1 adapter ligation we used 1.0 ul NEB Buffer 2 (at 10 %), 5.0 ul P1 adapter, 0.6 ul rATP, 0.5 ul, 2.9 ul ddH20, and T4 DNA-Ligase. We vortex for 2 sec, centrifuge at 5000 rpm for 3-5 sec, and incubated at room temperature for 45 min, then at 65°C for 20 min at 350 rpm. To

multiplex samples we pooled 42 samples per library into 7 libraries (these libraries contain more samples than those included in this study). We then sheared the libraries to ~500 bp fragments using sonication and cleaned fragments with MinElute clean-up (ThermoFisher Scientific, Bonn) and size selected on an agarose gel using the MinElute Gel Purification kit (ThermoFisher Scientific, Bonn). We blunted the overhangs by adding 2.5 ul dNTP mix and 1.0 ul Blunt Enzyme mix and purified libraries with QIAquick spin columns (Qiagen, Hilden, Germany) following manufacture´s protocol. To prepare the DNA fragments for ligation to the P2 Illumina adapters, we then added a "A" overhang to the 3`-end, each library was then purified with QIAquick spin columns (Qiagen, Hilden, Germany) and eluted in 52 ul elution buffer. For the sequence amplification, we combined 25.5 ul Phusion HF Master Mix (ThermoFisher Scientific, Bonn), 1.2 ul forward Solexa amplification primer (10 uM), 1.2 ul Solexa amplification primer (10 uM), and 5.0 ul RAD library templet from previous step. Cycling conditions were: initial denaturing at 98°C for 30 sec, followed by 17 cycles of 98°C for 10 sec, 65°C for 30 sec, and 72°C for 30 sec, with a final elongation of 72°C for 5 min. The PCR product for each library was then run on a 1% agarose gel at 100 V for 50 min, then the bands were excised and cleaned with MinElute Gel Purification kit (ThermoFisher Scientific, Bonn) following the manufacture´s protocol. Finally the libraries were sequenced with 76 cycles in 7 separate lanes on an Illumina Genome Analyzer IIx. Bioinformatic analysis was done following the pipeline described in Roesti et al. (2012).

Genetic structure and Clustering analyses with SNPs

We applied a Bayesian model-based clustering algorithm (Pritchard et al. 2000) as implemented in STRUCTURE v2.3.3 (http://pritch.bsd.uchicago.edu/structure.html). We used the admixture model and the number of ancestral clusters, *K*, was determined by comparing log-likelihoods in multiple runs for values of *K* between one and five. Each run consisted of 1,000,000 iterations with a burn-in period of 50,000.

MHC Genotyping

Amplification of MHC followed the same procedure as in Chapter 2. In brief one forward (AcMHCIIBF5) and two reverse (AcMHCIIBR10 and AcMHCIIBR11) primers were used to generate a library in a three step PCR. In a first round of PCRs we used adapter-tailed MHC-specific primers; in a second round PCRs included a set of 96 dual index combinations of full-length Illumina adaptors onto the amplicons; and the final PCR was done using standard Illumina PCR primers in emulsion on pooled second round PCRs to achieve a high yield of amplicons while at the same time minimizing artifacts like chimeric amplicons. The PCR conditions are detailed in chapter II.

The final library purification was done by size selection on LMP Agarose gel. The PCR products were then sequenced using Illumina Miseq at LGC Genomics (Berlin, Germany). The library was

demultiplexed using the BCL2FASTQ v1.8.4 software (Illumina Inc. 2017©). Sequence clustering and allele calling was done using the Amplicon Sequencing Analysis Tools "AmpliSAT" that were designed for multy-gene families such as MHC (Sebastian et al. 2016) .We first used the AmpliCHECK tool in order to optimize parameters for accurate allele calling. Then in AmpliSAS we ran 5 replicates of a maximum of 5000 randomly selected reads per individual with the default parameters, that were optimized for illumine sequencing technology, except, max alleles per individual was set at 50, and minimum amplicon depth was set to 1000. We used the AmpliCOMPARE tool to compare the results of the five replicates. Any alleles that did not appear in a given individual in at least three of the five replicates were discarded. We finally did a BLASTx search on NCBI´s Genbank database to confirm that all alleles were indeed MHC.

Phylogenetic analysis of MHC alleles

To determine the phylogenetic relationships between alleles, we constructed a phylogenetic tree using Bayesian inference with BEAST v2.0 (Bouckaert et al. 2014). We found that the most appropriate substitution model and partitioning scheme, and constructed a phylogenetic tree using BEAST (for details see chapter II), We depicted the phylogeny with the corresponding posterior values of each node with FIGTREE v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). To further analyze the relationship between the alleles we used the Network from Chapter II that placed each allele in contest of the global allele population. As we are not able to test for positive selected sites to determine allele functionality within groups due to the lack of allele lineage or loci information, we considered each group a supertype based high overall genetic nucleotide and amino acid differences between them.

Analysis of MHC variation

To determine if MHC allele frequencies segregated species, we did a canonical discriminant function test, and calculated Wilks´Lamda and Chi-square among species and lakes. To identify those alleles that can discriminate the benthic (benthic-deep, and benthic-shallow) and limnetic habitats, we did a logistic regression test. Finally we did canonical discriminant function grouping the shallow-benthics (*A. amarillo* with *A. astorquii*), deep-benthics (*A. xiloaensis* and *A. chancho*), and limnetics (*A. sagittae* and *A. zaliosus*) of both lakes to test for parallelism. We also did a canonical discriminant function test within each lake to determine if there are alleles that distinguish species,.

Results

Geometric morphometric analyses

Body shape variation. The first two principal components (PC) of the analysis of body shape in the species of crater L. Apoyo explained 59.91 % of the variance among individuals (PC1 = 44.38 %, PC2 = 15.53 %), and the first two PCs in the analysis of the species in crater l. Xiloá explained 49.3 % of the variation (PC1 = 26.2 %; PC2 =23.1 %). The cluster analysis based on these PCs identified three body shape clusters in the two lakes. In both cases one group with elongated body (limnetic, L) was differentiated from two the two other groups of deeper-bodied fish (benthic-like, B). Within the benthic species differences were subtler, with one species (the deeper one, BD) having rounder bodies and smaller heads, while the other form (shallower, BS) had thinner bodies and longer heads). Differences among species were significant according to the CVA analysis ($p < 0.001$).

LPJ variation. The analysis PCA of LPJ variation revealed that most variation was explained by the first two PCs in both lakes, explaining XXX in crater L. Apoyo and 93.3 % in crater L. Xiloá (PC1 = 81.9 %, PC2 = 11.4 %). This analysis differentiated two main groups, one formed by slender jaws with small pointed teeth (papilliform) and another with bulkier jaws with large rounded teeth (molariform). Limnetic fish in both lakes always had papiliform teeth, as well as benthic deep species. Benthic shallow species always had molariform LPJ. According to the CVA, not only molariform and papilliform LPJ differed, but also papilliform LPJ between limnetics and deep benthic within lakes (Apoyo: papilliform vs molariform*A. zaliosus* and *A. chancho,* $M = 1.41$, $p < 0.01$; Xiloá: papilliform vs molariform, $M = 6.63$, $p < 0.01$, *A. sagittae* and *A. xiloaensis,* $M = 1.41$, $p < 0.01$).

Ecological analyses

We found differences in diet among species (Table XX). BS species always had snails in their guts. BD and L species always had fish remnants in their guts. Despite their relatively similar diets BD and L species differed in the amount of insects ingested. Stable isotope analyses also differentiated the three species in both lakes (Fig. III.3). In both cases, BS species had lower ^{15}N levels, suggesting that this species is placed lower in the trophic chain. BD and L species overlapped in their stable isotope signature, and showed higher levels of ^{15}N, indicating higher position in the trophic chain.

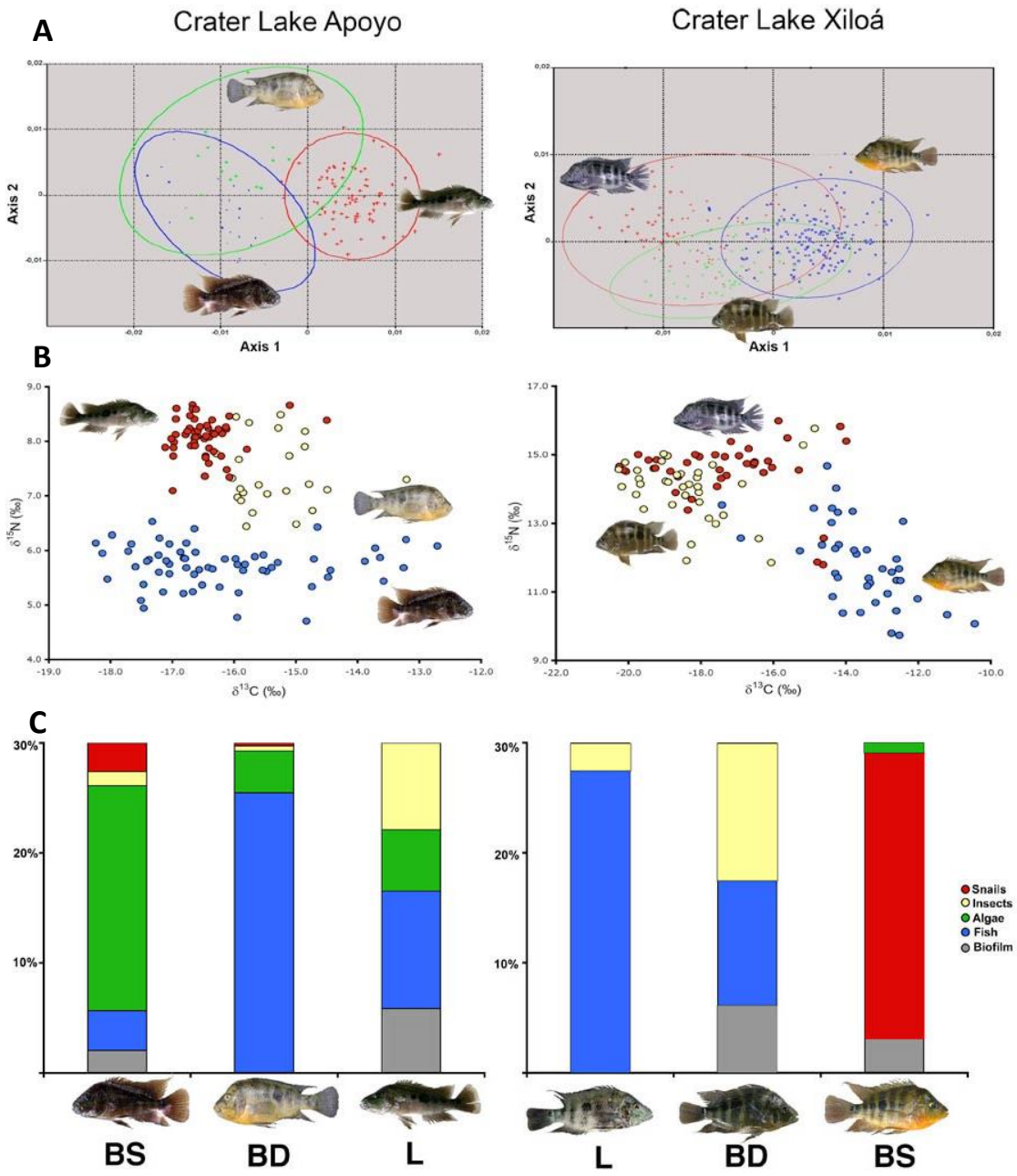


Figure III.3. A. Analysis of body shape in the two lakes discriminating the limnetic species from the two benthic species. B. Analysis of stable isotopes discriminating the benthic shallow species from the deeper benthic and limnetic species. C. Analysis of stomach content showing snails in the stomachs of the two shallow benthic species and a higher proportion of fish remnants in the two deeper benthic and limnetic species. D. Discriminant analysis with MHC genetic signature clearly discriminating the three species within each lake.

RAD-sequencing and SNP clustering analysis

We collected 62 individuals from the three species in crater L. Apoyo (20 *A. astorquii*, 21 *A. chancho*, and 21 *A. zaliosus*), and 64 samples from the three species in crater L. Xiloá (22 *A. Amarillo*, 13 *A. xiloaensis*, and 29 *A. sagittae*). The Illumina sequencing of the 126 individuals produced 1,902,616 clean sequences, with an average per individual of ~15 k sequences. The maximum sequence coverage was 45 k and the minimum 1.7 k (*A. astorquii*, 17.5 k*; A. chancho*, 18.5 k; *A. zaliosus*, 17 k; *A. Amarillo*, 16 k; *A. xiloaensis*, 12 k; *A. sagittae*, 9.5 k). Bayesian inference grouped fish in 6 different clusters corresponding to the 6 species. Each species within each lake was genetically distinct (Fig. III.4a). Supplementary Figure III.1 shows the plots of the 10 replicates of k =1-6 for both lakes. The analysis supports an optimal number of two populations per lake. This result indicates that the benthic deep species in both lakes still remain genetically undifferentiated from the benthic shallow, probably due to their very recent formation.

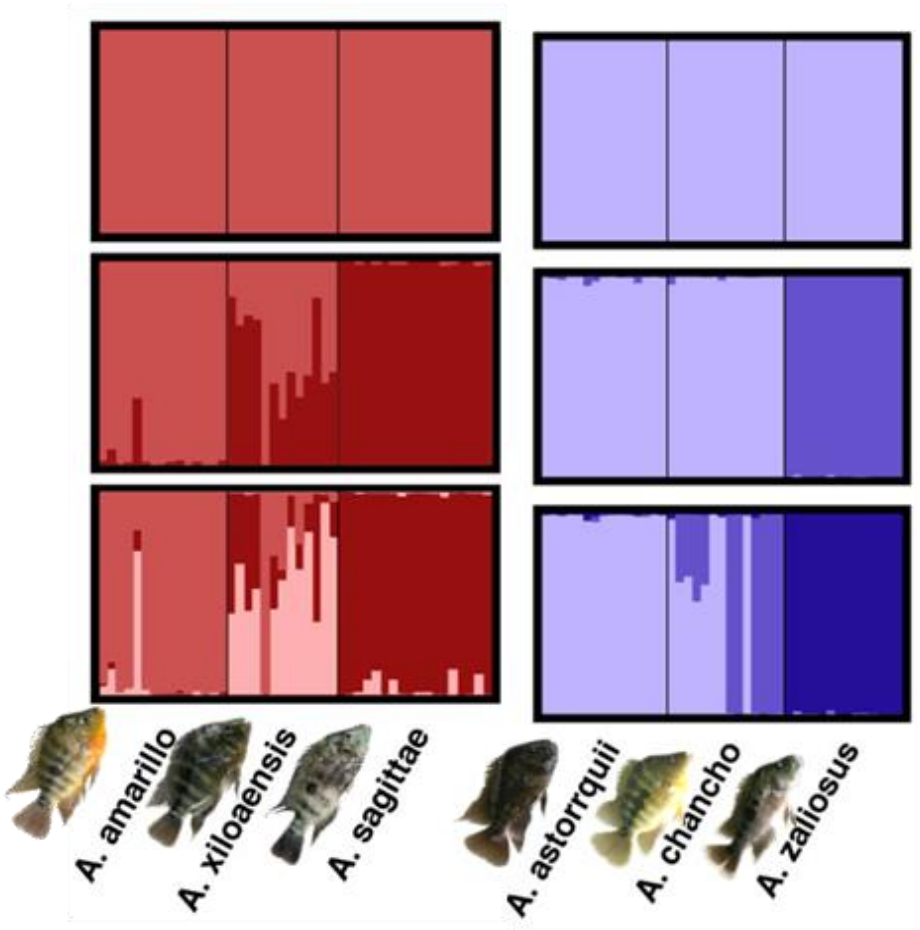


Figure III.4. Bayesian assignment analysis with Structure SNP data. In red the three species of crater L. Apoyo, and in blue the tree species of crater L. Xiloá

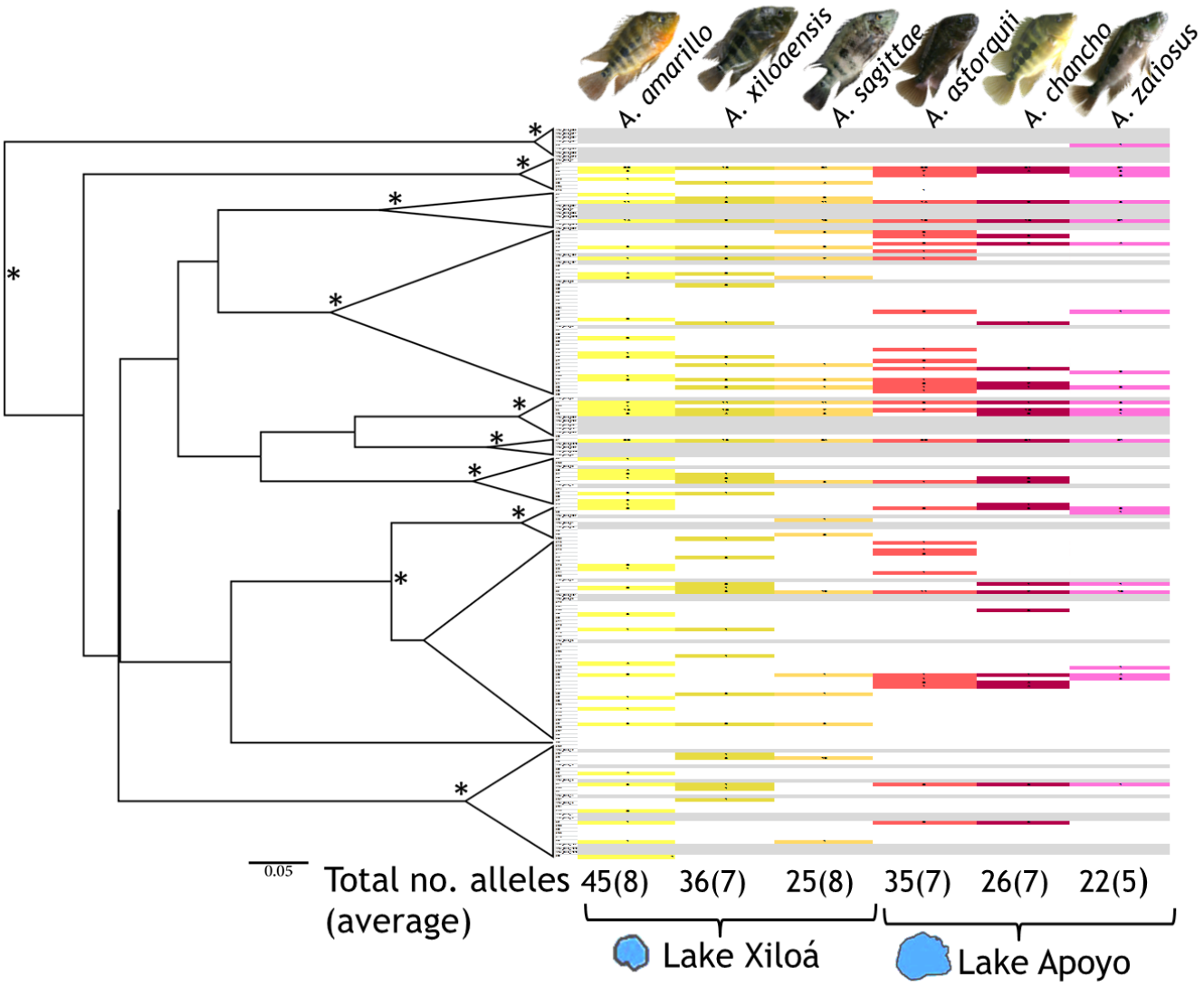


Figure III.5. Phylogenetic relationship between alleles of the six species studied (shallow benthic, *A. amarillo* and *A. astorquii*; deep benthic, *A. xilaoensis* and *A. chancho*; open water -limnetic species, *A. sagittae* and *A. zaliosus*). In relation to all alleles described, gray lines represent alleles from the characterization in chapter 1.

MHC diversity and genomic diversity

We obtained a total of 88 MHC IIB alleles for all 126 individuals analyzed. Species in crater L. Xiloá had 65 alleles while species in crater L. Apoyo had 45 alleles. Figure III.3 shows the number of alleles per species within lakes. We found an average of 7 alleles per individual (6.5 in crater L. Apoyo and 7.6 in crater L. Xiloá; min. 4, max 10). The six species varied in total number of alleles and average number of alleles per individual, although this difference was not statistically significant. Limnetic species

consistently had fewer alleles than the benthic species, and the limnetic species in crater L. Apoyo had lower number of alleles per individual than any other species.

Crater L. Apoyo contained alleles from 9 of the 11 described MHC supertypes (no allele from supertypes 10 and 12 were found; a single allele from supertypes 8 was found; alleles from supertype 4 were only abundant in *A. chancho*). Crater L. Xiloá had alleles from 9 of the 11 supertypes (no allele from 12 were found).

MHC alleles between the six species presented a high degree of TSP,. Despite large overlap in the segregating alleles, there was a unique genetic signature in each lake that distinguished species on the basis of MHC allele frequencies (Fig. III.4B). There are many shared alleles indicating TSP.

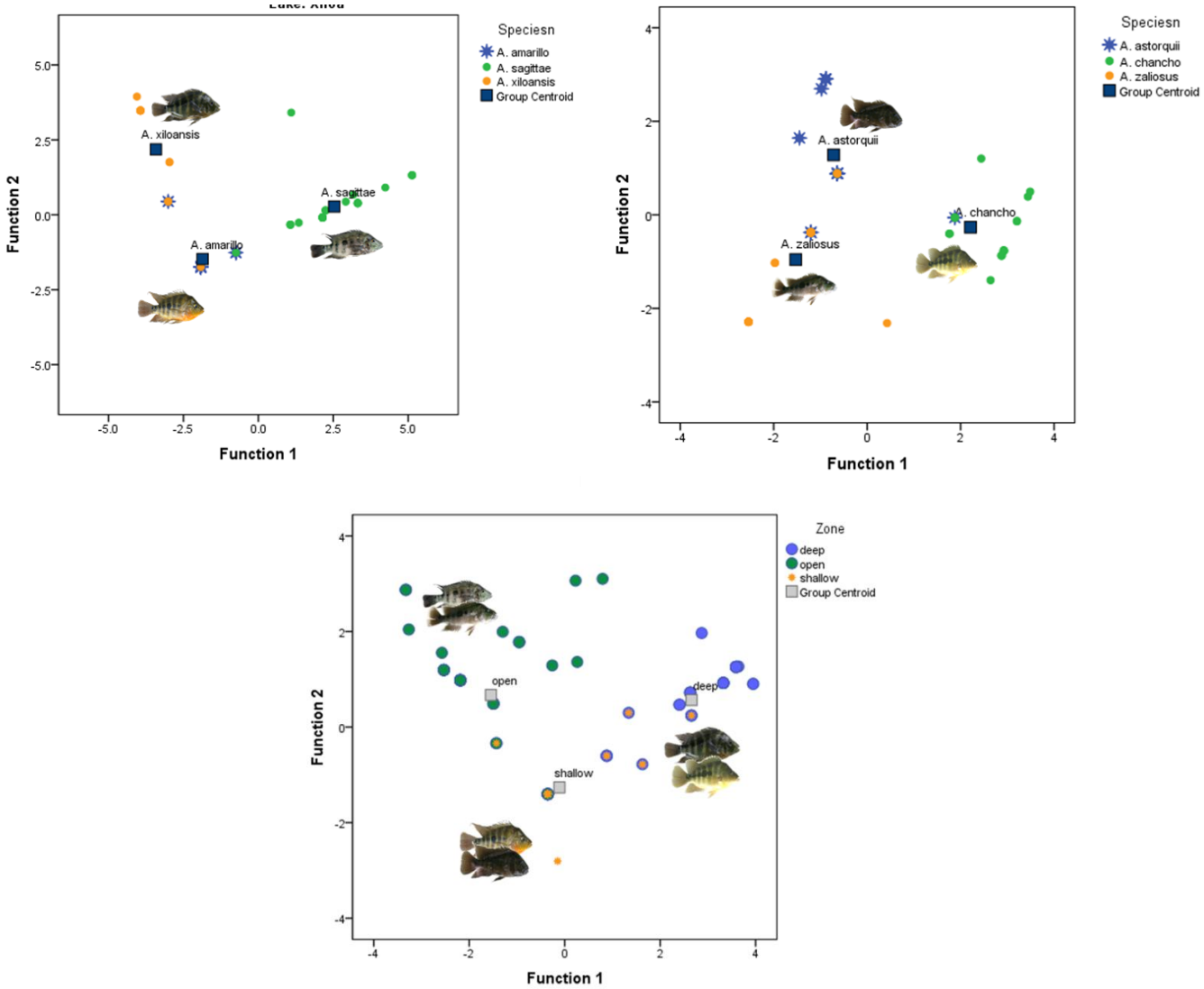


Figure III. 6. Canonical discriminant analysis of the most distinct MHC alleles combining the three pairs of species (shallow benthic, *A. amarillo* and *A. astorquii*; deep benthic, *A. xilaoensis* and *A. chancho*; open water -limnetic species, *A. sagittae* and *A. zaliosus*).

Discussion

We found that the three species pairs BS_{Apoyo}-$BS_{Xiloá}$, BD_{Apoyo}-$BD_{Xiloá}$, and L_{Apoyo}-$L_{Xiloá}$ are morphologically and ecologically equivalent. Morphometric analyses of overall body shape revealed that the benthic species have shorter and deeper bodies, while the limnetic species have lower and elongated bodies with elongated caudal peduncles (Fig. III.5A). Within crater lakes Apoyo and Xiloá the two benthic species are more similar to one another than they are to the limnetic ones. Analyses of the lower pharyngeal jaw showed that both BS species (*A. astorquii* in Apoyo and *A. amarillo* in Xiloá) have strong jaws with large rounded teeth ('molariform'). Stomach content analyses revealed that they both fed on snails plus plant material and *Aufwuchs*, a layer of green algae and diatoms admixed with detritus and sand carpeting rocks (Fig. III.5C).. The rare stable isotope analysis demonstrated that in both lakes the BS species have the lowest ^{15}N levels (Fig. III.5B), suggesting that BS is placed lowest in the trophic chain. The BD species (*A. chancho* in Apoyo and *A. xiloaensis* in Xiloá) have predator-like traits, more fragile pharyngeal jaws with smaller and pointed teeth ('papilliform'), and fish remnants filling their stomachs, again together with plant material and *Aufwuchs.* The limnetic species (*A. zaliosus* in Apoyo and *A. sagittae* in Xiloá), have a papilliform lower pharyngeal jaw apparatus, and show predator-like feeding habits. The BD and L species overlap considerably in their stable isotope signature and show high levels of ^{15}N (Fig. III.5B), indicating a higher position in the trophic chain.

The three species pairs BS_{Apoyo}-$BS_{Xiloá}$, BD_{Apoyo}-$BD_{Xiloá}$, and L_{Apoyo}-$L_{Xiloá}$ are the product of parallel speciation, since the lakes are isolated and the species assemblages of lakes Apoyo and Xiloá are reciprocally genetically independent. The species of Lake Apoyo form a monophyletic group suggesting that they have evolved *in situ* and sympatrically within the confines of the crater lake from an ancestor that most likely originated in nearby Lake Nicaragua (Barluenga *et al.* 2006; Barluenga & Meyer 2010). Lake Xiloá on the other hand, was most likely colonized by fish stocks from adjacent Lake Managua (Villa 1968; Barluenga & Meyer 2010), and although its Midas fauna does not seem to be of monophyletic origin (Barluenga & Meyer 2004, 2010), its genepool is distinct from that of all other lakes, and the Xiloá species are genetically more similar to one another than to populations and species from any other water-body in the area (Barluenga & Meyer 2010).

The most striking observation in the comparison of the radiations in crater lakes Apoyo and Xiloá is that the same sequence of speciation events occurred independently in these lakes. In both cases, the first step after colonization was the conquest of a novel habitat, the open water column, and the evolution of a limnetic species, with marked differences in body plan. The benthic niche partition along the depth axis only occurred at a later stage of the radiations and was accompanied by concomitant trophic specialization

leading to the split between the BS and the BD species. Genetic data corroborate this radiation in stages. Population assignment tests consistently revealed more pronounced differences between benthic and limnetic species than within the benthic types Fig. III.4A). Genetic diversity indices supports this scenario, too: The BS species are older, hence genetically more diverse, and expanded earlier; the L species are intermediate; and the BD species are the youngest, the least genetically diverse and either did not yet conform a population in demographic equilibrium, or underwent demographic expansion more recently than the other species. The benthic species pairs BS_{Apoyo}-$BS_{Xiloá}$ and BD_{Apoyo}-$BD_{Xiloá}$ also resemble each other more closely in overall morphology, but not in the ecologically important pharyngeal jaws, in diet (Fig. III.5C), where the BD species are more similar to the L species, in MHC composition (Fig. III.5D) where the BD species and L are quite distinct from each other in MHC profiles even though they both overlap considerably with the BS species. Thus, in both crater lakes the novel habitat, the deeper zone (which is absent from the older and larger lakes Managua and Nicaragua), has been occupied in two waves, apparently in the exact same sequence. First, a limnetic species evolved from a generalist ancestral benthic form as present in the large lakes, with the limnetic phenotype remaining restricted to the crater lakes. In a second wave of adaptation to deeper niches, the new BD species evolved. That the same phenotypic modifications occurred in parallel in crater lakes Apoyo and Xiloá, in spite of the lakes' different ages, the different source populations (L. Nicaragua *versus* L. Managua), and the different colonization histories (monophyletic *versus* paraphyletic origin), once more corroborates the causal role of natural selection in phenotypic evolution.

This study demonstrates that cichlid evolution can be somewhat predictable. That adaptive radiations may unfold in a predictable sequence is not a new idea (Schluter & McPhail 1993; Losos *et al.* 1998; Streelman & Danley 2003). Several vertebrate radiations appear to follow fixed evolutionary stages (Streelman & Danley 2003). Comparisons across taxa confirm that often the first axis of diversification involves habitat divergence. The early benthic-limnetic split in the crater lake Nicaraguan cichlid radiations provides one more example of this. The second axis of diversification involves trophic specializations. The secondary wave of trophic differentiation among the Nicaraguan crater lake benthic species fits again the proposed pattern. The third stage of evolutionary radiation relates to diversification in communication. This stage, missing in most radiations, is also not detected in the Nicaraguan crater lake cichlids. Future analyses on visual or chemical cues might confirm this observation.

How often evolution follows a fixed path is an intriguing question. Fish provide stunning examples of replicated evolution. The benthic-limnetic split is a frequent outcome of fish diversification, particularly after the colonization of lakes (Schluter & McPhail 1993; Robinson & Wilson 1994; Trapani 2003; Bernatchez 2004), this is also found in immune genes (Natsopoulou et al. 2012) that may have evolved in

parallel. In these cases virtually all taxa converge on equivalent body plans. This induces to think that lacustrine ecosystems offer similar and limited number of resources, and that most fish could have a limited number of designs able to deal with them. Although the similarities across pairs of benthic-limnetic species are enthralling, comprehensive ecological and morphological assays are still lacking in order to confirm the extensive parallelism.

Conclusions

Ecological speciation has taken place in two crater lakes in Nicaragua, where at least 3 species within the Midas cichlid have been formed in each lake. Speciation appears to have been mediated by natural selection through ecological specialization. Adaptation to the new environments was associated to variation in the immune response measured as MHC variation, especially in *A. sagittae* which presented a unique repertoire of alleles compared the other five species, however, the role of parasite mediated selection is still ambiguous and needs further study. We conclude that three equivalent eco-morphotypes have been formed in the two independent radiations in response to similar environments.

GENERAL DISCUSSION

The idea of speciation has been a contentious debate for centuries. Various theories have been proposed to explain how species form (Ridley 2003), and the prominent role of selection and ecological interactions has been of great interest over the last decades (Rundle et al. 2000; Egas et al. 2004; Rundle and Nosil 2005; Bolnick and Fitzpatrick 2007). In this work I hypothesize that parasite-mediated selection could initiate speciation or reinforce reproductive isolation in the Midas cichlid fish. To test this hypothesis I decided to study MHC genes because the variation in this immune gene has been linked to increased fitness, affecting traits such as parasite load (Wegner et al. 2003; Schad et al. 2005), mate-choice (Aeschlimann et al. 2003; Eizaguirre et al. 2010), parental effects (Turner et al. 2009; Kaufmann et al. 2014), fat reserves in males (Hablützel et al. 2014), survival (Sepil et al. 2013; Bateson et al. 2016), and lifetime reproductive success (Kalbe et al. 2009; Radwan et al. 2012; Sepil et al. 2013). For this reason MHC is an excellent gene complex to study evolution and it is sometimes considered a "magic trait" (Eizaguirre et al. 2009). In this work, I aimed to improve the understanding of the role that parasite-mediated selection plays in the speciation in Midas cichlid fish in several crater lakes in Nicaragua. I hypothesize that in contrasting environments parasite communities differ, and this will induce changes in the pools of MHC alleles. Mate choice related to MHC genotypes might trigger divergence between populations. The role of parasite communities in shaping host populations might be relevant both in sympatric and allopatric scenarios, and might alos paly a role in parallel situations. .

To do this I characterized MHC IIB genes *de novo* in Midas cichlids, I modeled their three dimensional structure and checked for key structural features, determined allele phylogeny and used NGS sequencing to MHC at a population level of several incipient and parallel speciation in sympatry and allopatry. I also determined the part that host-parasite interactions play in parallel ecological speciation. For this I take an in-depth look at a pair of equivalent species trios from two different lakes and determine if there is parallelism in their evolution based on genomic data, as well as other morphological and ecological data. This work sets a baseline reference for future studies involving mate choice experiments and parasite composition.

The first step was the challenging process of finding suitable primers to amplify Midas cichlid MHC IIB genes. By using tilapia and stickleback primers (Sato et al. 1998; Lenz et al. 2009) as references to start, I eventually were able to have primers that spanned the entire MHC genes. I tried the same primers on Spleen RNA to determine if all alleles were indeed expressed, however we discovered many more alleles in RNA then I were able to amplify in DNA using the same primer sets. Upon much investigation I was able to establish that extremely long (> 10 kb) introns were the cause of not being able to amplify all

alleles using intro-spanning primers on gDNA. Methodologically, to avoid this problem in future *de novo* MHC characterization studies, I recommend starting with primers set in the more conserved leader peptide (exon I) and the transmembrane region (in our case exon 5). By starting with these primers on RNA (cDNA) instead of gDNA you avoid the problems of intron length variants and initial primer placement in the hypervariable regions of Exon 2, as well as the amplification of unwanted pseudogenes. Once you have sequenced the resulting amplicons you can use them as reference to set primers spanning the region of exon 2 or 3 that is of interest. The only disadvantage of this strategy is not being able to set primers in the intro-exon boundary until you sequence said introns to account for their variability. MCH not being expressed in not of much concern if you use spleen tissue, as most studies find that all classical MHC is expressed (McConnell et al. 1998; Drozina et al. 2005; Li et al. 2014; Xu et al. 2016).

Once I completed the first main objective of this work, which was the characterizing the MHC molecules of the full range of alleles of the Midas cichlid species complex. I used phylogenetic analyses to establish the relationship between Midas cichlid MHC IIB alleles and estimate the selection pressures that each genetically-similar group is subjected to. However, the limitation to assign alleles to specific loci was problematic, as it cannot be distinguished between homologous alleles of the same locus and orthologous alleles from different loci. Genetic distance network analyses were used to group classical MHC alleles into 11 genetically similar groups, which were designated as MHC supertypes. Three supertypes were represented by a single allele. This finding may be due to different reasons.

We characterized 12 alleles belonging to Amci-DXB_040101- Amci-DXB_0406 that are nearly invariable in the fragment of exon 2 that was sequenced with NGS, meaning that sypertype 9 is in reality composed of 12 alleles that are indistinguishable in the fragment sequenced. Supertypes 10 and 12 have a rare allele present in only few individuals (18 and 4 respectively), and it is likely that with a larger sample size more alleles would be found. In order to establish the relationship between MHC of Midas cichlids and other fish, a phylogeny with the Midas MHC alleles was constructed, along with 20 alleles from 8 other cichlid and non-cichlid species, that showed clear signals of TSP. The results indicated that TSP along with balancing selection is the likely mechanism maintaining and promoting the diversity of MHC IIB alleles in this system.

The second main objective of this work was to describe the diversity of MHC class IIB alleles at a population level in several incipient Midas cichlid species in order to reveal the role that parasites may have had in divergence, and adaptation to new environments. Much like other species, the Midas cichlid show a pattern of a few dominant alleles present in a majority of the individuals sequenced, and some rare alleles present in only few individuals. The most abundant alleles (a1 - a6) comprise more than half of the

allele abundance, and 60 % of alleles (88 of the 146) are present in 5 or less individuals. The data show that despite this, most lake populations have a specific pattern of frequent alleles that are characteristic to each lake, even though most alleles are shared between several lakes or even all lakes.

The third objective of this thesis was to determine if MHC has evolved in parallel in two different lakes, along with parallel evolution of morphological traits. This has been found in other cichlids with respect to overall body morphology (Kocher et al. 1993; Clabaut et al. 2007), the morphology of the trophic apparatus (Rüber et al. 1999; Hulsey et al. 2008), coloration and pigmentation (Kocher et al. 1993; Seehausen 2006; Gavrilets et al. 2007), and brood-care behavior (Goodwin et al. 1998). The evidence of parallel evolution found in the three species pairs of crater lakes Apoyo and Xiloá, evident in their eco-morphology, is not so clearly explained with MHC data. Parallel evolution in the number of MHC alleles has been reported in stickleback *Gasterosteus aculeatus* (Natsopoulou et al. 2012). It might be needed a considerably larger sample to reach an equivalent conclusion in the Midas cichlid.

This work suggests that the pools of MHC alleles in the Midas cichlid do differ between environments within and among the Nicaraguan lakes. This finding shows that parasites are a selective pressure in both sympatric and allopatric conditions, However, without parasite community information it cannot be elucidated weather parasites are responsible for initiating, or for accelerating, divergence. The results show that the Midas cichlid have many MHC copies that are extremely diverse, as has been shown in other cichlid fish, and that MHC is subjected to positive but also balancing selection that maintains TSP. However, without mate choice experiments, or parasite composition there is a clear limitation to what can be concluded about MHC´s role in speciation.

Future Perspectives

To fully appreciate the MHC diversity in Midas cichlid species complex it is needed to sequence many more individuals of the same species and lake, as many alleles are rare at the population level and are detected in few individuals, this means that the total allele diversity will continue to increase with more sampling effort. To study the intricate process of host-parasite evolution, it would also be needed to have a comprehensive characterization of the parasite communities of each lake, this will enable to elucidate on the relationship of such parasites with each host species. The final key to really being able to study MHC in this species would be to be able to assign alleles to specific loci, however this is near impossible even with a well-annotated genome due to the unprecedented complexity and likely extensive copy number variation of cichlid MHC. One possible way to establish reference loci could be by using Bacterial

Artificial Chromosome (BAC) clones and long-read single molecule, real-time (SMRT) sequencing technology from Pacific Biosciences (PacBio) as implemented in (Viļuma et al. 2017). However it cannot be ruled out the likelihood of having very similar alleles from different loci that duplicated in the near past, alongside highly divergent alleles, and considerable recombination making it unclear if all future sequences could be assigned to specific loci. The findings of this work promise great potential in studying the processes of evolution and speciation in this model system and should be further studied despite its complexity as it holds the key to fully understanding the role that parasite-mediated selection has on speciation

CONCLUSIONS

I. The Midas MHCIIB gene structure is concordant with other cichlid species. The Midas cichlid has extremely variable intron lengths in some alleles, that can be up to 10 kb long making it very difficult to sequence using intron-spanning primers.

II: The Midas cichlid has a maximum reported of 26 alleles per individual (20 of which are classical MHC), implying that they have at least 13 MHC loci. They also present up to 12 non-classical MHC alleles per individual, meaning they have at least 6 non-classical loci. The 3D structure model of said alleles demonstrates that they have all the key components require to be functional.

III: The β_2 domain of all but the non-classical alleles were under selection, with up to 20 positively selected sites. This implies that they may contribute to the dynamic response of MHC to parasite-mediated selection.

IV: MHC allele composition of the Midas cichlid demonstrates a compelling case of trans-species polymorphism, not only with extensive sharing of alleles among species of *Amphilophus* but also sharing alleles with other distantly related fish species.

V: The immune genes in the Midas cichlid display enormous allelic richness and diversity. This diversity is larger than that described in most other fish, and only comparable to that found in African cichlids.

VI: Reference primers to study the Midas cichlid MHC at the population level have been established. However, it was not possible to assign alleles to specific locus due to their great polymorphism coupled with a lack of a good quality reference genome.

VII. At the population level, each lake has a unique signature of MHC allele composition. Crater lakes Asososca León and Xiloá are the most distinct, while lakes Managua, Nicaragua, Apoyo, and Masaya are more similar.

VIII. MHC alleles group into 12 genetically distinct groups which may reflect functional groups or allele supertypes. Allele groups vary greatly in size, with 55 alleles in one group, and a single allele in four groups.

IX: Some alleles are very abundant in all species and lakes, making a “core” set of alleles, potentially essential for survival in all lake environments. One allele was present in every single individual analyzed.

X. Individuals in all studied populations had an average of 7 alleles, belonging to 6 supertypes or functional groups. The optimal MHC supertype per individual differed among lake populations, may be due to different parasite pressures in each lake.

XI. Ecological speciation has taken place in two crater lakes in Nicaragua, where at least three species within the Midas cichlid have been formed in each lake.

REFERENCES

Aeschlimann PB, Häberli MA, Reusch TBH, et al (2003) Female sticklebacks Gasterosteus aculeatus use self-reference to optimize MHC allele number during mate selection. Behav Ecol Sociobiol 54:119–126. doi: 10.1007/s00265-003-0611-6

Albert JS, Reis RE (2011) Historical Biogeography of Neotropical Freshwater Fishes. University of California Press, Berkeley Los Angeles London

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

Anderson MJ (2001) A new method for non parametric multivariate analysis of variance. Austral Ecol 26:32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Arnegard ME, McGee MD, Matthews B, et al (2014) Genetics of ecological divergence during speciation. Nature. doi: 10.1038/nature13301

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. doi: 10.1093/bioinformatics/bti770

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Mol Ecol 10:237–251. doi: 10.1111/j.1755-0998.2009.02788.x

Bahr A, Wilson AB (2011) The impact of sex-role reversal on the diversity of the major histocompatibility complex: insights from the seahorse (*Hippocampus abdominalis*). BMC Evol Biol 11:121. doi: 10.1186/1471-2148-11-121

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376. doi: 10.1371/journal.pone.0003376

Barber I, Wright HA (2006) Effects of Parasites on Fish Behaviour: Interactions With Host Physiology. In: Fish Physiology. pp 109–149

Barlow GW, Baylis JR, Roberts D (1976) Chemical Analyses of some Crater Lakes in Relation to Adjacent Lake Nicaragua.

Barlow GW, Munsey JW (1976) The red devil-Midas-arrow cichlid species complex in Nicaragua. Pap Biol Sci 1:157–369.

Barluenga M, Meyer A (2004) The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? Mol Ecol 13:2061–2076. doi: 10.1111/j.1365-294X.2004.02211.x

Barluenga M, Meyer A (2010) Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus spp.*) in the Nicaraguan crater lakes. BMC Evol Biol 10:326. doi: 10.1186/1471-2148-10-326

Barluenga M, Stölting KN, Salzburger W, et al (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439:719–723. doi: 10.1038/nature04325

Bateson ZW, Hammerly SC, Johnson JA, et al (2016) Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. Mol Ecol 25:4730–4744. doi: 10.1111/mec.13793

Beaumont MA (2005) Adaptation and speciation: What can Fst tell us? Trends Ecol Evol 20:435–440. doi: 10.1016/j.tree.2005.05.017

Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res 27:573–580. doi: 10.1093/nar/27.2.573

Bernardi G (2013) Speciation in fishes. Mol Ecol 22:5487–5502. doi: 10.1111/mec.12494

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–77.

Bice DC (1985) Quaternary volcanic stratigraphy of Managua, Nicaragua: correlation and source assignment for multiple overlapping plinian deposits. Geol Soc Am Bull 96:553–566. doi: 10.1130/0016-7606(1985)96<553:QVSOMN>2.0.CO;2

Bijlmakers I, Sobalvarro A (1988) Ecological studies on the profundal benthic invertebrate fauna of Lake Managua, Nicaragua. Trop Freshwat Biol 1:30–41.

Blais J, Rico C, van Oosterhout C, et al (2007) MHC adaptive divergence between closely related and sympatric African cichlids. PLoS One 2:e734. doi: 10.1371/journal.pone.0000734

Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010) Extensive MHC Class II B Gene Duplication in a Passerine, the Common Yellowthroat (*Geothlypis trichas*). J Hered 101:448–460. doi: 10.1093/jhered/esq018

Bontrop RE (2006) Comparative Genetics of MHC Polymorphisms in Different Primate Species: Duplications and Deletions. Hum Immunol 67:388–397. doi: 10.1016/j.humimm.2006.03.007

Borghans J a M, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. Immunogenetics 55:732–739. doi: 10.1007/s00251-003-0630-5

Bouckaert R, Heled J, Kühnert D, et al (2014) BEAST 2: A Software Platform for Bayesian Evolutionary

Analysis. PLoS Comput Biol 10:1–6. doi: 10.1371/journal.pcbi.1003537

Bracamonte SE, Baltazar-Soares M, Eizaguirre C (2015) Characterization of MHC class II genes in the critically endangered European eel (*Anguilla Anguilla*). Conserv Genet Resour 7:859–870. doi: 10.1007/s12686-015-0501-z

Brawand D, Wagner CE, Li YI, et al (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381. doi: 10.1038/nature13726

Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58:191–202. doi: 10.1007/s00251-006-0085-6

Buckling A, Rainey PB (2002) The role of parasites in sympatric and allopatric host diversification. Nature 420:496–9. doi: 10.1038/nature01164

Bulut Z, McCormick CR, Bos DH, DeWoody AJ (2008) Polymorphism of alternative splicing of major histocompatibility complex transcripts in wild tiger salamanders. J Mol Evol 67:68–75. doi: 10.1007/s00239-008-9125-1

Bunn A, Korpela M (2016) Time Series Analysis in dplR. doi: 10.1017/CBO9781107415324.004

Burress ED (2015) Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. Hydrobiologia 748:7–27. doi: 10.1007/s10750-014-1960-z

Burri R, Promerová M, Goebel J, Fumagalli L (2014) PCR-based isolation of multigene families: lessons from the avian MHC class IIB. Mol Ecol 14:778–788. doi: 10.1111/1755-0998.12234

Bussing WA, Rica UDC (1976) Geographic Distribution of the San Juan Ichthyofauna of Central America with Remarks on its Origin and Ecology.

Butler EJ, Sun X-Z, Wertz N (2011) Immunoglobulin Polygeny: An Evolutionary Perspective. In: Gene Duplication. InTech, pp 1–28

Butlin RK, Debelle A, Kerth C, et al (2012) What do we need to know about speciation? Trends Ecol Evol 27:27–39. doi: 10.1016/j.tree.2011.09.002

Chen SL, Zhang YX, Xu MY, et al (2006) Molecular polymorphism and expression analysis of MHC class II B gene from red sea bream (*Chrysophrys major*). Dev Comp Immunol 30:407–418. doi: 10.1016/j.dci.2005.06.001

Chevin LM, Decorzent G, Lenormand T (2014) Niche dimensionality and the genetics of ecological speciation. Evolution (N Y) 68:1244–1256. doi: 10.1111/evo.12346

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 18:117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Cole GA (1976) Limnology of the Great Lakes of Nicaragua.

Consuegra S, Ellison A, Allainguillaume J, et al (2013) Balancing selection and the maintenance of MHC supertype variation in a selfing vertebrate. Proc R Soc London Ser B 280:20122854. doi: 10.1098/rspb.2012.2501.Despite

Coyne JA (2007) Quick guide Sympatric speciation. Curr Biol 17:787–788.

Csárdi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Syst 1695:1695.

Cummings SM, McMullan M, Joyce DA, van Oosterhout C (2010) Solutions for PCR, cloning and sequencing errors in population genetic analysis. Conserv Geneics 1095–1097. doi: 10.1007/s10592-009-9864-6

Darwin C (1859) The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 6th edn.

De Boer RJ, Borghans JAM, Van Boven M, et al (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55:725–731. doi: 10.1007/s00251-003-0629-y

Decaestecker E, De Gersem H, Michalakis Y, Raeymaekers JAM (2013) Damped long-term host-parasite Red Queen coevolutionary dynamics: a reflection of dilution effects? Ecol Lett 1455–1462. doi: 10.1111/ele.12186

Dijkstra JM, Katagiri T, Hosomichi K, et al (2007) A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics 59:305–321. doi: 10.1007/s00251-007-0198-6

Dionne M, Miller KM, Dodson JJ, et al (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution (N Y) 61:2154–64. doi: 10.1111/j.1558-5646.2007.00178.x

Dixon B, Van Erp SHM, Rodrigues PNS, et al (1995) Fish major histocompatibility complex genes: an expresion. Dev Comp Immunol 19:109–133.

Dobson A, Lafferty KD, Kuris AM, et al (2008) Colloquium paper: homage to Linnaeus: how many parasites? How many hosts? Proc Natl Acad Sci U S A 105:11482–11489. doi: 10.1073/pnas.0803232105

Doherty PC, Zinkernagel RM (1975) A biological role for the major histocompatibility antigens. Lancet (London, England) 1:1406–9. doi: 10.1016/S0140-6736(75)92610-0

Doytchinova IA, Flower DR (2005) In silico identification of supertypes for class II MHCs. J Immunol 174:7085–95.

Eimes JA, Townsend AK, Sepil I, et al (2015) Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. PeerJ 3:e853. doi: 10.7717/peerj.853

Eizaguirre C, Lenz TL (2010) Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. J Fish Biol 77:2023–2047. doi: 10.1111/j.1095-8649.2010.02819.x

Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. Ecol Lett 15:723–731. doi: 10.1111/j.1461-0248.2012.01791.x

Eizaguirre C, Lenz TL, Sommerfeld RD, et al (2010) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. Evol Ecol 25:605–622. doi: 10.1007/s10682-010-9424-z

Eizaguirre C, Lenz TL, Traulsen A, Milinski M (2009a) Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. Ecol Lett 12:5–12. doi: 10.1111/j.1461-0248.2008.01247.x

Eizaguirre C, Yeates SE, Lenz TL, et al (2009b) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. Mol Ecol 18:3316–3329. doi: 10.1111/j.1365-294X.2009.04243.x

Ejsmond MJ, Babik W, Radwan J (2010) MHC allele frequency distributions under parasite-driven selection: A simulation model. BMC Evol Biol 10:332. doi: 10.1186/1471-2148-10-332

Ekblom R, Saether SA, Fiske P, et al (2010) Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. Genetica 138:453–61. doi: 10.1007/s10709-008-9335-x

Elmer KR, Fan S, Gunter HM, et al (2010a) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Mol Ecol 19:197–211. doi: 10.1111/j.1365-294X.2009.04488.x

Elmer KR, Fan S, Kusche H, et al (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat Commun 5:5168. doi: 10.1038/ncomms6168

Elmer KR, Lehtonen TK, Kautt AF, et al (2010b) Rapid sympatric ecological differentiation of crater lake

cichlid fishes within historic times. BMC Biol 8:60. doi: 10.1186/1741-7007-8-60

Elmer KR, Lehtonen TK, Meyer A (2009) Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. Evolution (N Y) 63:2750–7. doi: 10.1111/j.1558-5646.2009.00736.x

Eschmeyer WN, Fong JD (2016) Catalog of Fishes. http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp. Accessed 31 Aug 2016

Evans ML, Dionne M, Miller KM, Bernatchez L (2012) Mate choice for major histocompatibility complex genetic divergence as a bet-hedging strategy in the Atlantic salmon (Salmo salar). Proc R Soc London Ser B 279:379–86. doi: 10.1098/rspb.2011.0909

Figueroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. Nature 335:265–267.

Figueroa F, Klein J (1986) The evolution of MHC class II genes. Immunol Today 7:78–81. doi: 10.1016/0167-5699(86)90089-7

Figueroa F, Mayer WE, Sültmann H, et al (2000) Mhc class II B gene evolution in East African cichlid fishes. Immunogenetics 51:556–575. doi: 10.1007/s002510000181

Figueroa F, Ono H, Tichy H, et al (1995) Evidence for Insertion of a New Intron into an Mhc Gene of Perch-Like Fish. Proc R Soc B Biol Sci 259:325–330. doi: 10.1098/rspb.1995.0048

Friedman M, Keck BP, Dornburg A, et al (2013) Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc R Soc London Ser B 280:20131733. doi: 10.1098/rspb.2013.1733

Galdino C a B, Ávila RW, Bezerra CH, et al (2014) Helminths Infection Patterns in a Lizard (Tropidurus hispidus) Population from a Semiarid Neotropical Area: Associations Between Female Reproductive Allocation and Parasite Loads. J Parasitol 100:864–7. doi: 10.1645/13-264.1

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution (N Y) 57:1707–22.

Gavrilets S (2004) Fitness Landscapes and the Origin of Species. Princeton University Press, NJ

Gavrilets S, Vose A, Barluenga M, et al (2007) Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. Mol Ecol 16:2893–909. doi: 10.1111/j.1365-294X.2007.03305.x

Geiger MF, McCrary JK, Schliewen UK (2010) Not a simple case - A first comprehensive phylogenetic hypothesis for the Midas cichlid complex in Nicaragua (Teleostei: Cichlidae:Amphilophus). Mol

Phylogenet Evol 56:1011–24. doi: 10.1016/j.ympev.2010.05.015

Getz WM, Salter R, Seidel DP, van Hooft P (2016) Sympatric speciation in structureless environments. BMC Evol Biol 16:50. doi: 10.1186/s12862-016-0617-0

Grimholt U (2016) MHC and Evolution in Teleosts. Bilology 5:6. doi: 10.3390/biology5010006

Guillén Bolaños TY (2006) Diagnóstico de los residuos sólidos y liquidos de la reserva natural laguna de apoyo.

Haase D, Roth O, Kalbe M, et al (2013) Absence of major histocompatibility complex class II mediated immunity in pipefish , *Syngnathus typhle* : evidence from deep transcriptome sequencing. Biol Lett 9:20130044–20130044. doi: doi.org/10.1098/rsbl.2013.0044

Hablützel PI, Vanhove MPM, Gregoir AF, et al (2014) Intermediate number of MHC class IIB variants relates to enlarged perivisceral fat deposits in blunt-head cichlids (*Tropheus spp.*). J Evol Biol 27:2177–2190. doi: 10.1111/jeb.12467

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.

Harf R, Sommer S (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, Gerbillurus paeba, in the southern Kalahari. Mol Ecol 14:85–91. doi: 10.1111/j.1365-294X.2004.02402.x

Harstad H, Lukacs MF, Bakke HG, Grimholt U (2008) Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (*Salmo salar*). BMC Genomics 9:193. doi: 10.1186/1471-2164-9-193

Havlicek J, Roberts SC (2009) MHC-correlated mate choice in humans: a review. Psychoneuroendocrinology 34:497–512. doi: 10.1016/j.psyneuen.2008.10.007

Hedrick PW (1992) Female choice and variation in the major histocompatibility complex. Genetics 132:575–581.

Hedrick PW (1999) Balancing selection and MHC. Genetica 104:207–214. doi: 10.1023/A:1026494212540

Hill AVS (1991) HLA Associations with Malaria in Africa: Some Implications for MHC Evolution. In: Molecular Evolution of the Major Histocompatibility Complex. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 403–420

Hofmann MJ, Bracamonte SE, Eizaguirre C, Barluenga M (2017) Molecular characterization of MHC

class IIB genes of sympatric Neotropical cichlids. BMC Genet 18:15. doi: 10.1186/s12863-017-0474-x

Huchard E, Raymond M, Benavides J, et al (2010) A female signal reflects MHC genotype in a social primate. BMC Evol Biol 10:96. doi: 10.1186/1471-2148-10-96

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170. doi: 10.1038/335167a0

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci 86:958–62.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32:415–35. doi: 10.1146/annurev.genet.32.1.415

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. doi: 10.1093/molbev/msj030

Hviid TV., Møller C, Sørensen S, Morling N (1998) Co-dominant Expression of the HLA-G Gene and Various Forms of Alternatively Spliced HLA-G mRNA in Human First Trimester Trophoblast. Hum Immunol 59:87–98. doi: 10.1016/S0198-8859(97)00259-0

Janeway CA, Travers P, Walport M (2005) Immunobiology: the immune system in health and disease.

Janeway CAJ, Travers P, Walport M (2001) Immunobiology: The Immune System in Health and Disease, 5th edn. Garland Science, New York

Jiang J, Li C, Zhang Q, Wang X (2015) Locus number estimation of MHC class II B in stone flounder and Japanese flounder. Int J Mol Sci 16:6000–6017. doi: 10.3390/ijms16036000

Kautt AF, Elmer KR, Meyer A (2012) Genomic signatures of divergent selection and speciation patterns in a “natural experiment”, the young parallel radiations of Nicaraguan crater lake cichlid fishes. Mol Ecol 21:4770–4786. doi: 10.1111/j.1365-294X.2012.05738.x

Kautt AF, Machado-Schiaffino G, Torres-Dowdall J, Meyer A (2016) Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? Ecol Evol 1–16. doi: 10.1002/ece3.2287

Kennedy CEJ, Endler JA, Poynton SL, McMinn H (1987) Parasite load predicts mate choice in guppies. Behav Ecol Sociobiol 21:291–295. doi: 10.1007/BF00299966

Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Zamudio KR (2010) Selection, trans-species polymorphism, and locus identification of major histocompatibility complex class IIβ alleles of New

World ranid frogs. Immunogenetics 62:741–751. doi: 10.1007/s00251-010-0476-6

Kikkawa EF, Tsuda TT, Sumiyama D, et al (2009) Trans-species polymorphism of the Mhc class II DRB -like gene in banded penguins (genus *Spheniscus*). Immunogenetics 61:341–352. doi: 10.1007/s00251-009-0363-1

Klein D, Ono H, O'hUigin C, et al (1993) Extensive MHC variability in cichlids of Lake Malawi. Nature 364:330–334.

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19:155–162. doi: 10.1016/0198-8859(87)90066-8

Klein J (1980) Generation of diversity at MHC loci: Implications for T-cell receptor repertoires. In: Fougereau M, Dausset J (eds) Immunology 80. London, Academic Press, pp 239–253

Klein J, Bontrop RE, Dawkins RL, et al (1990) Nomenclature for the major histocompatibility complex of different species: a proposal. Immunogenetics 31:217–219.

Klein J, Figueroa F (1986) Evolution of the major histocompatibility complex. Crit Rev Immunol 6:295–386.

Klein J, Sato A, Nagl S, et al (1998) Molecular trans-species polymorphism. Annu Rev Ecol Syst 29:1–21. doi: 10.1146/annurev.ecolsys.29.1.1

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet 41:281–304. doi: 10.1146/annurev.genet.41.110306.130137

Kobayashi KS, van den Elsen PJ (2012) NLRC5: a key regulator of MHC class I-dependent immune responses. Nat Rev Immunol 12:813–20. doi: 10.1038/nri3339

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288–98. doi: 10.1038/nrg1316

Kornfield I, Smith PF (2000) African Cichlid Fish: Model Systems for Evolutionary Biology. Annu Rev Ecol Syst 31:163–182.

Kropshofer H, Hämmerling GJ, Vogt AB (1999) The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. Immunol Rev 172:267–278. doi: 10.1111/j.1600-065X.1999.tb01371.x

Kullander SO (2003) Family Cichlidae. In: Check List of the Freshwater Fishes of South and Central America. pp 605–654

Kulski JK, Shiina T, Anzai T, et al (2002) Comparative genomic analysis of the MHC: the evolution of

class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95–122.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:msw054. doi: 10.1093/molbev/msw054

Kutterolf S, Freundt A, Pérez W, et al (2007) Late Pleistocene to Holocene temporal succession and magnitudes of highly-explosive volcanic eruptions in west-central Nicaragua. J Volcanol Geotherm Res 163:55–82. doi: 10.1016/j.jvolgeores.2007.02.006

Lanfear R, Calcott B, Kainer D, et al (2014) Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol 14:82. doi: 10.1186/1471-2148-14-82

Laurens V, Chapusot C, del Rosario Ordonez M, et al (2001) Axolotl MHC class II beta chain: predominance of one allele and alternative splicing of the beta1 domain. Eur J Immunol 31:506–515. doi: 10.1002/1521-4141(200102)31:2<506::AID-IMMU506>3.0.CO;2-P

Legendre P, Desdevises Y, Bazin E (2002) A statistical test for host-parasite coevolution. Syst Biol 51:217–34. doi: 10.1080/10635150252899734

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci-implications for evolutionary analysis. Gene 427:117–123. doi: 10.1016/j.gene.2008.09.013

Lenz TL, Eizaguirre C, Becker S, Reusch TBH (2009a) RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. BMC Evol Biol 9:57. doi: 10.1186/1471-2148-9-57

Lenz TL, Eizaguirre C, Kalbe M, Milinski M (2013) Evaluating patterns of convergent evolution and trans-species polymorphism at mhc immunogenes in two sympatric stickleback species. Evolution (N Y) 67:2400–2412. doi: 10.1111/evo.12124

Lenz TL, Eizaguirre C, Scharsack JP, et al (2009b) Disentangling the role of MHC-dependent “good genes” and “compatible genes” in mate-choice decisions of three-spined sticklebacks *Gasterosteus aculeatus* under semi-natural conditions. J Fish Biol 75:2122–2142. doi: 10.1111/j.1095-8649.2009.02410.x

Li H, Jiang L, Han J, et al (2011) Major histocompatibility complex class IIA and IIB genes of the spotted halibut *Verasper variegatus*: Genomic structure, molecular polymorphism, and expression analysis. Fish Physiol Biochem 37:767–780. doi: 10.1007/s10695-011-9476-1

Lighten J, Oosterhout C Van, Paterson IG, et al (2014a) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy

(Poecilia reticulata). Mol Ecol 14:753–67. doi: 10.1111/1755-0998.12225

Lighten J, Van Oosterhout C, Bentzen P (2014b) Critical review of NGS analyses for de novo genotyping multigene families. Mol Ecol 23:3957–3972. doi: 10.1111/mec.12843

Lynch S, Santos SG, Campbell EC, et al (2009) Novel MHC class I structures on exosomes. J Immunol 183:1884–91. doi: 10.4049/jimmunol.0900798

Maan ME, Seehausen O (2011) Ecology, sexual selection and speciation. Ecol Lett 14:591–602. doi: 10.1111/j.1461-0248.2011.01606.x

Madden DR (1995) The three-dimensional structure of peptide-mhc complexes. Annu Rev Immunol 13:587–622. doi: 10.1146/annurev.iy.13.040195.003103

Málaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, et al (1998) Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. Genetics 149:1527–1537.

Manousaki T, Hull PM, Kusche H, et al (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. Mol Ecol 22:650–69. doi: 10.1111/mec.12034

Matschiner M, Musilová Z, Barth JMI, et al (2016) Bayesian Phylogenetic Estimation of Clade Ages Supports Trans-Atlantic Dispersal of Cichlid Fishes. Syst Biol. doi: 10.1093/sysbio/syw076

McKaye KR (1986) mate choice and size assortative pairing be the cichlid fishes of Lake Jiloá, Nicaragua. J Fish Biol 29:135–150.

McKaye KR, Stauffer JRJ, Berghe EP Van Den, et al (2002) Behavioral , Morphological and Genetic Evidence of Divergence of the Midas Cichlid Species Complex in Two Nicaraguan Crater Lakes. Cuad Investig la UCL 12:19–47.

Meyer D, Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. Ann Hum Genet 65:1–26.

Milinski M (2014) Arms races, ornaments and fragrant genes: The dilemma of mate choice in fishes. Neurosci Biobehav Rev 46:567–572. doi: 10.1016/j.neubiorev.2014.08.005

Milinski M (2006) The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. Annu Rev Ecol Evol Syst 37:159–186. doi: 10.1146/annurev.ecolsys.37.091305.110242

Montenegro-Guillen S (1991) Limnological perspective of Lake Xolotlán (Managua): The PLALM. Hydrobiol Bull 25:105–109. doi: 10.1007/BF02291239

Mora C, Tittensor DP, Adl S, et al (2011) How many species are there on earth and in the ocean? PLoS

Biol 9:1–8. doi: 10.1371/journal.pbio.1001127

Murray BW, Shintani S, Sültmann H, Klein J (2000) Major histocompatibility complex class II A genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. Immunogenetics 51:576–586.

Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol Biol 11:116. doi: 10.1186/1471-2148-11-116

Natsopoulou ME, Pálsson S, Ólafsdóttir GÁ (2012) Parasites and parallel divergence of the number of individual MHC alleles between sympatric three-spined stickleback Gasterosteus aculeatus morphs in Iceland. J Fish Biol 81:1696–1714. doi: 10.1111/j.1095-8649.2012.03430.x

Neefjes J, Jongsma MLM, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11:823–36. doi: 10.1038/nri3084

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci 94:7799–7806. doi: 10.1073/pnas.94.15.7799

Niskanen AK, Kennedy LJ, Ruokonen M, et al (2014) Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population. Mol Ecol 23:875–89. doi: 10.1111/mec.12647

Nosil P (2012) Ecological speciation. Oxford University Press

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). Proc R Soc London Ser B 276:1119–1128. doi: 10.1098/rspb.2008.1525

Olsen LR, Kudahl UJ, Simon C, et al (2013) BlockLogo: Visualization of peptide and sequence motif conservation. J Immunol Methods 400–401:37–44. doi: 10.1016/j.jim.2013.08.014

Ono H, O'hUigin C, Tichy H, Klein J (1993a) Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Mol Biol Evol 10:1060–72.

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

Ottová E, Simková A, Martin J-F, et al (2005) Evolution and trans-species polymorphism of MHC class IIbeta genes in cyprinid fish. Fish Shellfish Immunol 18:199–222. doi: 10.1016/j.fsi.2004.07.004

Pang J, Gao F, Lu M, et al (2013) Major histocompatibility complex class IIA and IIB genes of Nile

tilapia *Oreochromis niloticus*: genomic structure, molecular polymorphism and expression patterns. Fish Shellfish Immunol 34:486–496. doi: 10.1016/j.fsi.2012.11.048

Pastoret P-P, Griebel P, Bazin H, Govaerts A (eds) (1998) Handbook of Veetebrate Immunology. Academic Press, San Diego, California 92101-4495, USA

Penn D, Potts W (1999) The evolution of mating preferences and major histocompatibility complex genes. Am Nat 153:145–164. doi: 10.1086/303166

Penn DJ, Ilmonen P (2005) Major Histocompatibility Complex (MHC). In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK, pp 1–7

Pérez W, Freundt A, Kutterolf S, Schmincke H-U (2009) The Masaya Triple Layer: A 2100 year old basaltic multi-episodic Plinian eruption from the Masaya Caldera Complex (Nicaragua). J Volcanol Geotherm Res 179:191–205. doi: 10.1016/j.jvolgeores.2008.10.015

Piertney SB, Oliver MK (2005) The evolutionary ecology of the major histocompatibility complex. Heredity (Edinb) 96:7–21. doi: 10.1038/sj.hdy.6800724

Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. Trends Ecol Evol 5:181–7. doi: 10.1016/0169-5347(90)90207-T

Pritchard JK, Stephens M, Donnelly P, Falush D (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959.

Rammensee H-G, Bachmann J, Stevanović S (1997) MHC Ligands and Peptide Motifs. Springer Berlin Heidelberg, Berlin, Heidelberg, Heidelberg

Relle M, Schwarting A (2012) Role of MHC-Linked Susceptibility Genes in the Pathogenesis of Human and Murine Lupus. Clin Dev Immunol 2012:1–15. doi: 10.1155/2012/584374

Reusch TBH, Langefors A (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback Gasterosteus aculeatus. J Mol Evol 61:531–541. doi: 10.1007/s00239-004-0340-0

Reusch TBH, Schaschl H, Wegner KM (2004) Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. Immunogenetics 56:427–437. doi: 10.1007/s00251-004-0704-z

Roesti M, Hendry A p., Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol Ecol 21:2852–2862. doi: 10.1111/j.1365-294X.2012.05509.x

Rogers W (1995) Female Choice Predicts the Best Father in a Biparental Fish, the Midas Cichlid (Cichlasoma citrinellum). Ethology 100:230–241. doi: 10.1111/j.1439-0310.1995.tb00327.x

Rogers W, Barlow GW (1991) Sex differences in mate choice in a monogamous biparental fish, the Midas cichlid (Cichlasoma citrinellum). Ethology 87:249–261. doi: 10.1111/j.1439-0310.1991.tb00250.x

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105:1118–1123. doi: 10.1073/pnas.0706851105

Rozenfeld AF, Arnaud-Haond S, Hernández-García E, et al (2008) Network analysis identifies weak and strong links in a metapopulation system. Proc Natl Acad Sci U S A 105:18824–9. doi: 10.1073/pnas.0805571105

Rundle HD, Nosil P (2005) Ecological speciation. Ecol Lett 8:336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Rymešová D, Králová T, Promerová M, et al (2017) Mate choice for major histocompatibility complex complementarity in a strictly monogamous bird, the grey partridge (Perdix perdix). Front Zool 14:9. doi: 10.1186/s12983-017-0194-0

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol Ecol 18:169–185. doi: 10.1111/j.1365-294X.2008.03981.x

Santos EM, Braasch I, Boileau N, et al (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat Commun 5:5149. doi: 10.1038/ncomms6149

Sato A, Dongak R, Hao L, Shintani S (2012) Organization of Mhc class II A and B genes in the tilapiine fish *Oreochromis*. Immunogenetics 64:679–690. doi: 10.1007/s00251-012-0618-0

Sato A, Figueroa F, O'hUigin C, et al (1998) Cloning of major histocompatibility complex (Mhc) genes from threespine stickleback, *Gasterosteus aculeatus*. Mol Mar Biol Biotechnol 7:221–231.

Sato A, Klein D, Sültmann H, et al (1997) Class I mhc genes of cichlid fishes: identification, expression, and polymorphism. Immunogenetics 46:63–72.

Scheibye K, Weisser J, Borggaard OK, et al (2014) Sediment baseline study of levels and sources of polycyclic aromatic hydrocarbons and heavy metals in Lake Nicaragua. Chemosphere 95:556–565. doi: 10.1016/j.chemosphere.2013.09.115

Schliewen UK, Klee B (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids. Front Zool 1:5. doi: 10.1186/1742-9994-1-5

Schluter D (2001) Ecology and the origin of species. Trends Ecol Evol 16:372–380.

Schwaiger FW, Weyers E, Epplen C, et al (1993) The paradox of MHC-DRB exon/intron evolution: alpha-helix and beta-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with beta-sheet codons. J Mol Evol 37:260–272. doi: 10.1007/BF00175503

Sebastian A, Herdegen M, Migalska M, Radwan J (2016) Amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Resour 16:498–510. doi: 10.1111/1755-0998.12453

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proc R Soc B 273:1987–1998. doi: 10.1098/rspb.2006.3539

Seehausen O, Butlin RK, Keller I, et al (2014) Genomics and the origin of species. Nat Rev Genet 15:176–92. doi: 10.1038/nrg3644

Selz OM, Pierotti MER, Maan ME, et al (2014) Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. Behav Ecol 0:1–15. doi: 10.1093/beheco/aru024

Sepil I, Lachish S, Hinks AE, Sheldon BC (2013a) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. Proc R Soc B Biol Sci 280:20130134–20130134. doi: 10.1098/rspb.2013.0134

Sepil I, Lachish S, Sheldon BC (2013b) Mhc-linked survival and lifetime reproductive success in a wild population of great tits. Mol Ecol 22:384–96. doi: 10.1111/mec.12123

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 18:1–18. doi: 10.1186/1742-9994-2-16

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B Biol Sci 277:979–988. doi: 10.1098/rspb.2009.2084

Star B, Jentoft S (2012) Why does the immune system of Atlantic cod lack MHC II? BioEssays 34:648–651. doi: 10.1002/bies.201200005

Star B, Nederbragt AJ, Jentoft S, et al (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210. doi: 10.1038/nature10342

Stauffer D, Aharony A (1994) Introduction to Percolation Theory. Taylor & Francis, London

Stauffer JR, McCrary JK, Black KE (2008) Three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Apoyo, Nicaragua. Proc Biol Soc Washingt 121:117–129. doi: 10.2988/06-37.1

Stiebens VVA, Merino SE, Chain FJJ, Eizaguirre C (2013) Evolution of immunogenes in the endangered

loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. BMC Evol Biol 13:95. doi: 10.1186/1471-2148-13-95

Sugden AM (2000) Ecology/Evolution: Sexual Selection and Speciation. Science (80-) 290:1055a–1055. doi: 10.1126/science.290.5494.1055a

Summers K, McKeon S, Sellars J, et al (2003) Parasitic exploitation as an engine of diversity. Biol Rev Camb Philos Soc 78:639–675. doi: 10.1017/S146479310300616X

Takahashi T, Koblmüller S (2011) The adaptive radiation of cichlid fish in lake tanganyika: a morphological perspective. Int J Evol Biol 2011:620754. doi: 10.4061/2011/620754

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967–78.

Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex. Genetics 130TSK92:925–938.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci 101:11030–5. doi: 10.1073/pnas.0404206101

Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi: 10.1093/molbev/msr121

Teifke JP, Klopfleisch R, Globig A, et al (2007) Pathology of natural infections by H5N1 highly pathogenic avian influenza virus in mute (Cygnus olor) and whooper (Cygnus cygnus) swans. Vet Pathol 44:137–43. doi: 10.1354/vp.44-2-137

Tellier A, Brown JKM (2007) Stability of genetic polymorphism in host-parasite interactions. Proc R Soc London Ser B 274:809–17. doi: 10.1098/rspb.2006.0281

Thibert-Plante X, Gavrilets S (2013) Evolution of mate choice and the so-called magic traits in ecological speciation. Ecol Lett 16:1004–13. doi: 10.1111/ele.12131

Trachtenberg E, Korber B, Sollars C, et al (2003) Advantage of rare HLA supertype in HIV disease progression. Nat Med 9:928–935. doi: 10.1038/nm893

Trowsdale J (1993) Genomic structure and function in the MHC. Trends Genet 9:117–122.

Trowsdale J (2011) The MHC, disease and selection. Immunol Lett 137:1–8. doi: 10.1016/j.imlet.2011.01.002

Turner GF (2007) Adaptive radiation of cichlid fish. Curr Biol 17:R827–R831. doi:

10.1016/j.cub.2007.07.026

van Staaden MJ, Smith AR (2011) Cutting the Gordian knot: Complex signaling in African cichlids is more than multimodal. Curr Zool 57:237–252.

Vanhove M, Hablützel PI, Pariselle A, et al (2016) Cichlids: a host of opportunities for evolutionary parasitology. Trends Parasitol 1536:1–13. doi: 10.1016/j.pt.2016.07.002

Via S (2009) Natural selection in action during speciation. Proc Natl Acad Sci U S A 106 Suppl:9939–9946. doi: 10.1073/pnas.0901397106

Vincek V, O'Huigin C, Satta Y, et al (1997) How large was the founding population of Darwin's finches? Proc R Soc London Ser B 264:111–118. doi: 10.1098/rspb.1997.0017

Waid RM, Reasly RL, Kenneth R, et al (1999) Zoogeografía ïctica de lagunas cratéricas de Nicaragua. Encuentro 51:65–81.

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. Mol Ecol 18:1818–1820. doi: 10.1111/j.1365-294X.2009.04173.x

Wielinga C, Thompson RCA, Monis P, Ryan U (2015) Identification of polymorphic genes for use in assemblage B genotyping assays through comparative genomics of multiple assemblage B Giardia duodenalis isolates. Mol Biochem Parasitol 201:1–4. doi: 10.1016/j.molbiopara.2015.05.002

Wilson AB, Noack-Kunnmann K, Meyer A (2000) Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. Proc R Soc London Ser B 267:2133–41. doi: 10.1098/rspb.2000.1260

Woelfing B, Traulsen A, Milinski M, Boehm T (2009) Does intra-individual major histocompatibility complex diversity keep a golden mean? Philos Trans R Soc B Biol Sci 364:117–28. doi: 10.1098/rstb.2008.0174

Wu C-I, Ting C-T (2004) Genes and speciation. Nat Rev Genet 5:114–22. doi: 10.1038/nrg1269

Wu CI (2001) The genic view of the process of speciation. J Evol Biol 14:851–865. doi: 10.1046/j.1420-9101.2001.00335.x

Xu S, Chen B, Zhou K, Yang G (2008) High similarity at three MHC loci between the baiji and finless porpoise: trans-species or convergent evolution? Mol Phylogenet Evol 47:36–44. doi: 10.1016/j.ympev.2007.05.026

Xu T, Chen S, Zhang YX (2010) MHC class II alpha gene polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). Dev

Comp Immunol 34:1042–1050. doi: 10.1016/j.dci.2010.05.008

Xu T, Liu J, Sun Y, et al (2016) Characterization of 40 full-length MHC class IIA functional alleles in miiuy croaker: Polymorphism and positive selection. Dev Comp Immunol 55:138–143. doi: 10.1016/j.dci.2015.10.019

Xu T, Sun Y, Shi G, et al (2011) Characterization of the Major Histocompatibility Complex Class II Genes in Miiuy Croaker. PLoS One 6:e23823. doi: 10.1371/journal.pone.0023823

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. doi: 10.1093/molbev/msm088

Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev 167:45–58.

Yewdell JW, Reits E, Neefjes J (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol 3:952–61. doi: 10.1038/nri1250

Zhang YX, Chen SL, Liu YG, et al (2006) Major histocompatibility complex class IIB allele polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). Mar Biotechnol 8:600–610. doi: 10.1007/s10126-005-6185-8

Zhou F, Dong Z, Fu Y, et al (2013) Molecular cloning, genomic structure, polymorphism and expression analysis of major histocompatibility complex class II B gene of Nile tilapia (Oreochromis niloticus). Aquaculture 372–375:149–157. doi: 10.1016/j.aquaculture.2012.10.032

APPENDIXES

Chapter I

Sup. Table I.1. List of samples used for this study, with their ID, species, lake of origin and whether we obtained gDNA, cDNA or both.

ID	Species	Lake	gDNA	cDNA
AM05	*Amphilophus citrinellus*	Lake Asososca León	✓	x
AM06	*Amphilophus citrinellus*	Lake Asososca León	✓	x
27D3	*Amphilophus citrinellus*	Lake Managua	✓	✓
AH65	*Amphilophus citrinellus*	Lake Masaya	✓	x
AQ72	*Amphilophus citrinellus*	Lake Masaya	✓	x
25D8	*Amphilophus citrinellus*	Lake Nicaragua	✓	✓
25D1	*Amphilophus citrinellus*	Lake Nicaragua	x	✓
25E7	*Amphilophus citrinellus*	Lake Nicaragua	✓	✓
25A2	*Amphilopus labiatus*	Lake Nicaragua	✓	✓
25B2	*Amphilopus labiatus*	Lake Nicaragua	✓	✓
27G4	*Amphilophus xiloaensis*	Lake Xiloá	x	✓
AL64	*Amphilophus xiloaensis*	Lake Xiloá	✓	x
AL11	*Amphilophus amarillo*	Lake Xiloá	✓	x

Sup. Table I-2. List of sequences used to design primers "MHC-Rev_3" all obtained from NCBI´s GeneBank.

Reference Number	Species
gi|378940681	*Oreochromis niloticus*
gi|378940681	*Oreochromis niloticus*
gi|379323047	*Oreochromis niloticus*
gi|332640059	*Oreochromis niloticus*
gi|379323049	*Oreochromis niloticus*
gi|378940681	*Oreochromis niloticus*
gi|EF540224.1	*Pseudotropheus flavus*
gi|2822471	*Poeciliopsis occide*
gi|380005973	*Epinephelus coioi*
gi|194338247	*xiphophorus multilineatus*
gi|387166486	*Sebastes caurinus*
gi|387167030	*Sebastes maliger*
gi|289223	*Aulonocara hansbaens*
gi|51449916	*Gasterosteus aculeatus*
gi|51449916	*Gasterosteus aculeatus*
gi|306966138	*Oncorhynchus mykiss*
gi|9937658	*Salmo trutta*
gi|205933548	*Salvelinus alpinus*
gi|115738470:29	*Dicentrarchus labrax*
gi|51449916	*Gasterosteus aculeatus*
gi|51449916	*Gasterosteus aculeatus*

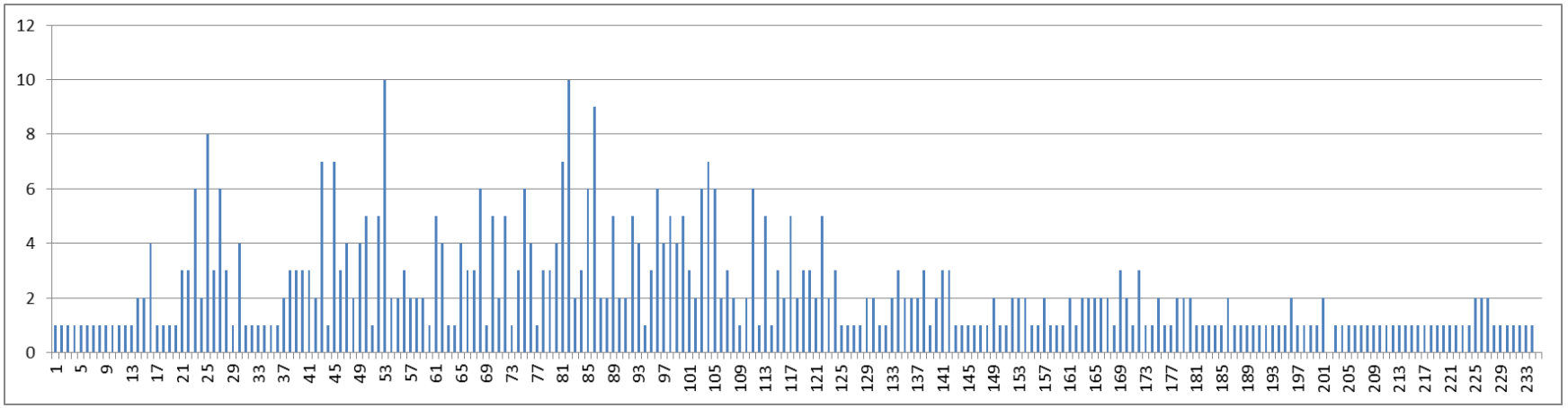
Sup. Table I-3. List of sequences of MHC IIB spanning exon 2 and 3 used from different fish species obtained from the literature to evaluate trans-species polymorphism.

Accesion #	Species/allele family	Source	Family
gi|387166334	*Sebastes caurinus*	Johansson et al unpublished	Sebastidae
gi|387166338	*Sebastes caurinus*	Johansson et al unpublished	Sebastidae
gi|387166402	*Sebastes maliger*	Johansson et al unpublished	Sebastidae
gi|478859903	*Miichthys miiuy*	Xu et al 2011	Sciaenidae
gi|7715334	*Pundamilia nyererei*	Figueroa et al 2000	Cichlidae
gi|7715336	*Haplochromis sp. 'rockkribensis'*	Figueroa et al 2000	Cichlidae
gi|357217114	*Oreochromis niloticus*	Zhou et al unpublished	Cichlidae
gi|379323037	*Oreochromis niloticus*	Gao et al unpublished	Cichlidae
gi|7715378	*Haplochromis xenognathus*	Figueroa et al 2000	Cichlidae
gi|394933462	*Oreochromis niloticus-DAB*	Sato et al 2012	Cichlidae
gi|394933466	*Oreochromis niloticus-DBB*	Sato et al 2012	Cichlidae
gi|394933470	*Oreochromis niloticus-DCB*	Sato et al 2012	Cichlidae
gi|394933474	*Oreochromis niloticus-DDB*	Sato et al 2012	Cichlidae
gi|394933476	*Oreochromis niloticus-DEB*	Sato et al 2012	Cichlidae
gi|394933479	*Oreochromis niloticus-DFB*	Sato et al 2012	Cichlidae
gi|394933483	*Oreochromis niloticus-DGB*	Sato et al 2012	Cichlidae
gi|394933486	*Oreochromis niloticus-DHB*	Sato et al 2012	Cichlidae
gi|394933490	*Oreochromis niloticus-DIB*	Sato et al 2012	Cichlidae
gi|394933494	*Oreochromis niloticus-DJB*	Sato et al 2012	Cichlidae
gi|394933497	*Oreochromis niloticus-DKB*	Sato et al 2012	Cichlidae

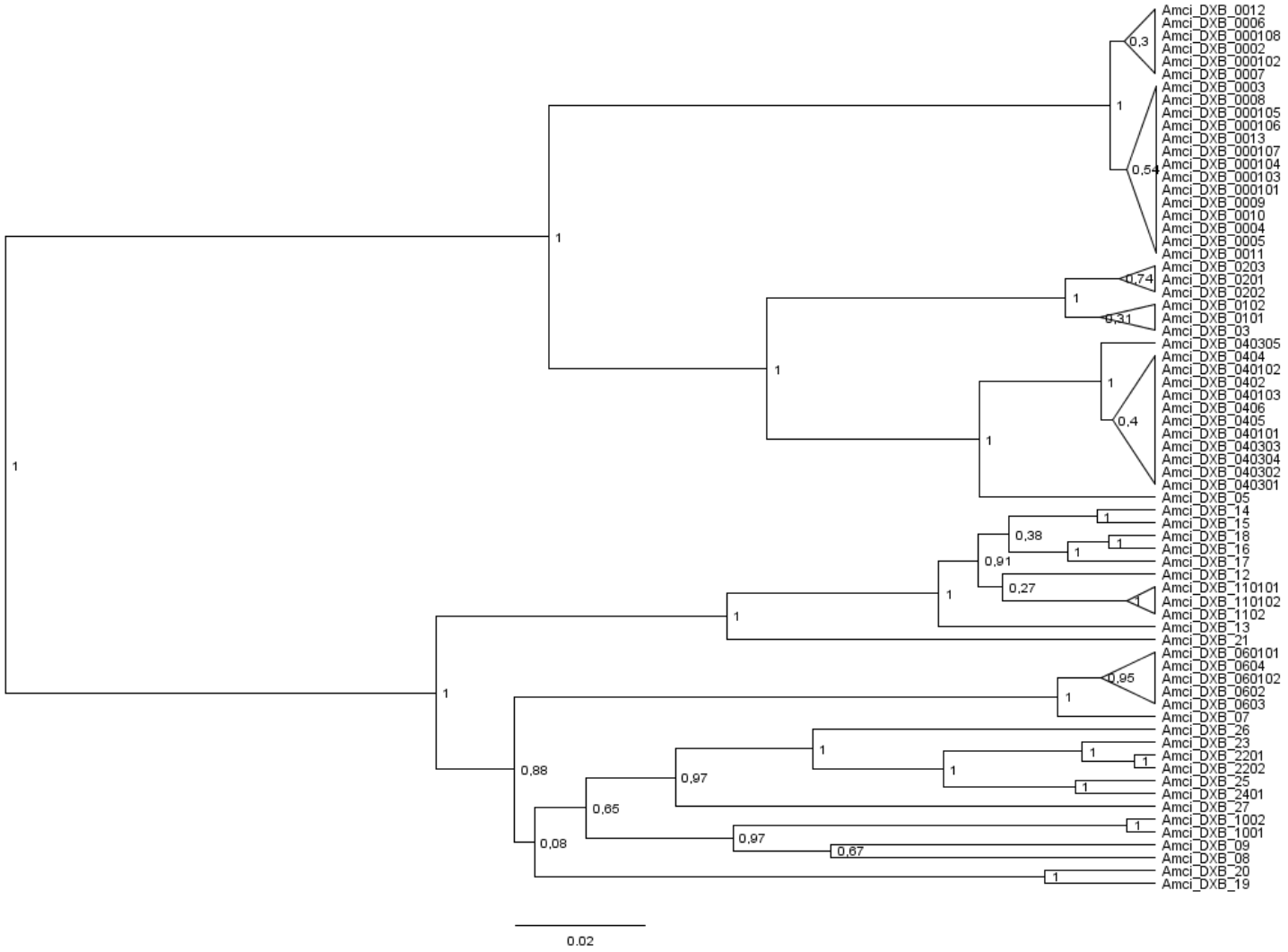
Sup. Table I-4. Information of the Protein homology models built for all alleles with Swiss-Model workspace. We report the templet used to build each model, the percent Identity with said templet, the global model quality estimation (GMQE), and the overall model quality (QMEAN4). The Summary of normalized QMean4 scores of all 3D Homology models compared with a non-reduntant set of Protein Data Bank (PDB) structures to infer Z-scores is also reported.

Allele	Templet	% Identity	GMQE	QMEAN4	Quality
Amci-DXB*000101	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*000102	4p23.1.D	39.18%	0.74	-2.7	✓
Amci-DXB*000103	1fv1.1.B	39.18%	0.71	-3.37	x
Amci-DXB*000104	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*000105	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*000106	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*000107	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*000108	4grl.1.B	40.74%	0.69	-3.5	x
Amci-DXB*0002	1fv1.1.B	37.61%	0.73	-3.43	x
Amci-DXB*0003	1fne.2.B	38.38%	0.73	-1.67	✓
Amci-DXB*0004	1fv1.1.B	39.09%	0.72	-3.27	x
Amci-DXB*0005	1fv1.1.B	38.37%	0.72	-2.41	✓
Amci-DXB*0006	4z7w.1.B	36.36%	0.7	-3.55	x
Amci-DXB*0007	1fv1.1.B	35.78%	0.71	-3.65	x
Amci-DXB*0008	1fv1.1.B	38.18%	0.7	-4.11	x
Amci-DXB*0009	1fv1.1.B	37.27%	0.71	-3.42	x
Amci-DXB*0010	1fv1.1.B	37.27%	0.71	-3.54	x
Amci-DXB*0011	1fv1.1.B	37.27%	0.71	-3.41	x
Amci-DXB*0012	4z7w.1.B	35.06%	0.7	-3.37	x
Amci-DXB*0013	1fv1.1.B	38.18%	0.71	-3.37	x
Amci-DXB*0101	4mcy.1.B	39.09%	0.73	-2.75	✓
Amci-DXB*0102	4mcy.1.B	39.09%	0.73	-2.75	✓
Amci-DXB*0201	4h26.1.B	39.81%	0.72	-3.03	x
Amci-DXB*0202	4h26.1.B	40.68%	0.7	-3.28	x
Amci-DXB*0203	4ov5.5.B	37.18%	0.71	-2.91	✓
Amci-DXB*03	4mcy.1.B	41.50%	0.73	-3.49	x
Amci-DXB*040101	1kt2.2.B	42.20%	0.73	-3.18	x
Amci-DXB*040102	1kt2.2.B	42.20%	0.73	-3.18	x
Amci-DXB*040103	1uvq.1.B	40.26%	0.71	-3.38	x
Amci-DXB*0402	4md5.1.B	41.67%	0.74	-2.91	✓
Amci-DXB*040301	1aqd.4.B	42.20%	0.73	-2.95	✓
Amci-DXB*040302	1aqd.4.B	42.20%	0.73	-2.95	✓
Amci-DXB*040303	1aqd.4.B	42.20%	0.73	-2.95	✓
Amci-DXB*040304	1aqd.4.B	42.20%	0.73	-2.95	✓
Amci-DXB*040305	4p23.1.D	41.96%	0.74	-2.8	✓
Amci-DXB*0404	1aqd.4.B	41.67%	0.74	-2.87	✓
Amci-DXB*0405	4p4r.1.B	39.80%	0.7	-4.21	x
Amci-DXB*0406	4y19.1.B	40.70%	0.74	-2.6	✓
Amci-DXB*05	3pdo.1.B	38.89%	0.72	-3.06	x
Amci-DXB*060101	4p46.1.B	35.98%	0.61	-3.01	x
Amci-DXB*060102	4i5b.1.B	40.74%	0.7	-3.75	x
Amci-DXB*0602	4i5b.1.B	39.81%	0.7	-3.7	x
Amci-DXB*0603	4i5b.1.B	40.74%	0.69	-3.84	x
Amci-DXB*0604	1fv1.1.B	33.87%	0.63	-3.59	x
Amci-DXB*07	3c5z.1.D	37.21%	0.68	-2.82	✓
Amci-DXB*08	1h15.1.B	36.36%	0.7	-4.26	x
Amci-DXB*09	1fv1.1.B	40.00%	0.72	-3.58	x
Amci-DXB*1001	4h25.1.B	41.67%	0.72	-3.37	x
Amci-DXB*1002	2g9h.1.B	39.81%	0.71	-3.79	x
Amci-DXB*110101	3c5z.1.D	35.64%	0.62	-3.66	x
Amci-DXB*110102	3c5z.1.D	35.64%	0.62	-3.66	x
Amci-DXB*1102	4p23.1.D	35.84%	0.68	-2.45	✓
Amci-DXB*12	1fv1.1.B	30.40%	0.61	-3.77	x
Amci-DXB*13	4mcy.1.B	34.55%	0.72	-2.31	✓
Amci-DXB*14	4i5b.1.B	31.20%	0.62	-2.52	✓
Amci-DXB*15	1fv1.1.B	34.68%	0.64	-3.18	x
Amci-DXB*16	1fne.2.B	33.06%	0.62	-3.41	x
Amci-DXB*17	4p23.1.D	35.45%	0.62	-3.44	x
Amci-DXB*18	4md5.1.B	32.26%	0.63	-3.5	x
Amci-DXB*19	1fv1.1.B	35.00%	0.69	-4.32	x
Amci-DXB*20	1fv1.1.B	34.68%	0.62	-4.32	x
Amci-DXB*21	4mcy.1.B	40.00%	0.74	-1.77	✓
Amci-DXB*2201	4p23.1.D	36.84%	0.63	-3.69	x
Amci-DXB*2202	4p46.1.B	37.04%	0.62	-3.41	x
Amci-DXB*23	3c5z.1.D	39.53%	0.68	-2.65	✓
Amci-DXB*2401	4is6.1.B	38.10%	0.62	-5.54	x
Amci-DXB*25	1bx2.1.B	47.06%	0.67	-2.52	✓
Amci-DXB*26	2g9h.1.B	35.48%	0.65	-2.81	✓
Amci-DXB*27	4grl.1.B	38.67%	0.65	-4.44	x

Sup. Fig. I.1. Amino acid variability ranging from 1-10 amino acids per site.

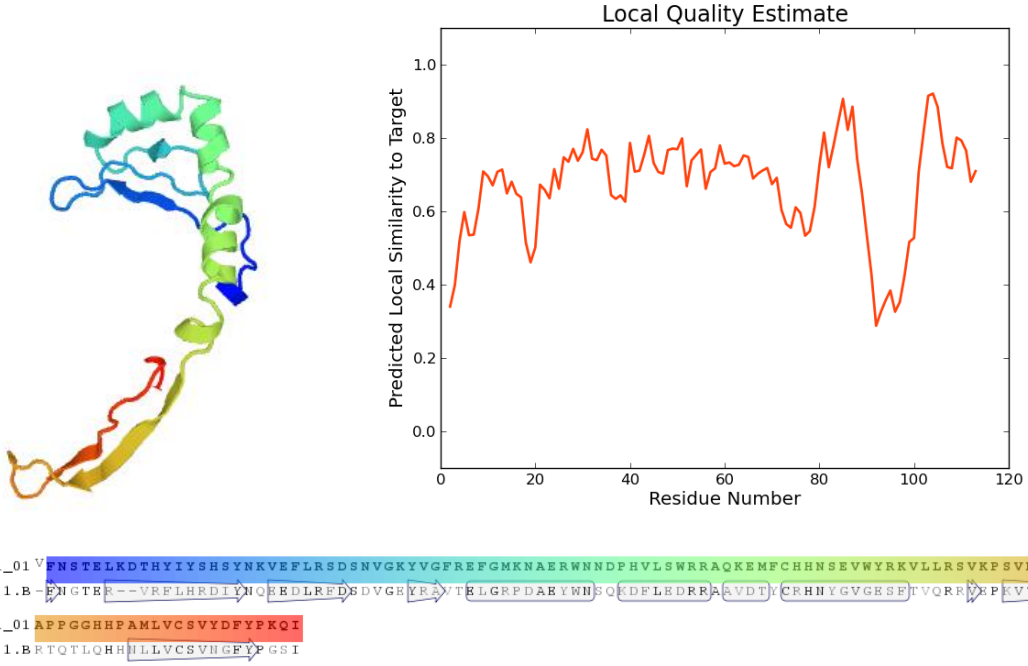


Sup. Fig. I.2. Phylogenetic inference relationship of all MHC IIB alleles with posterior probabilities shown for each node.

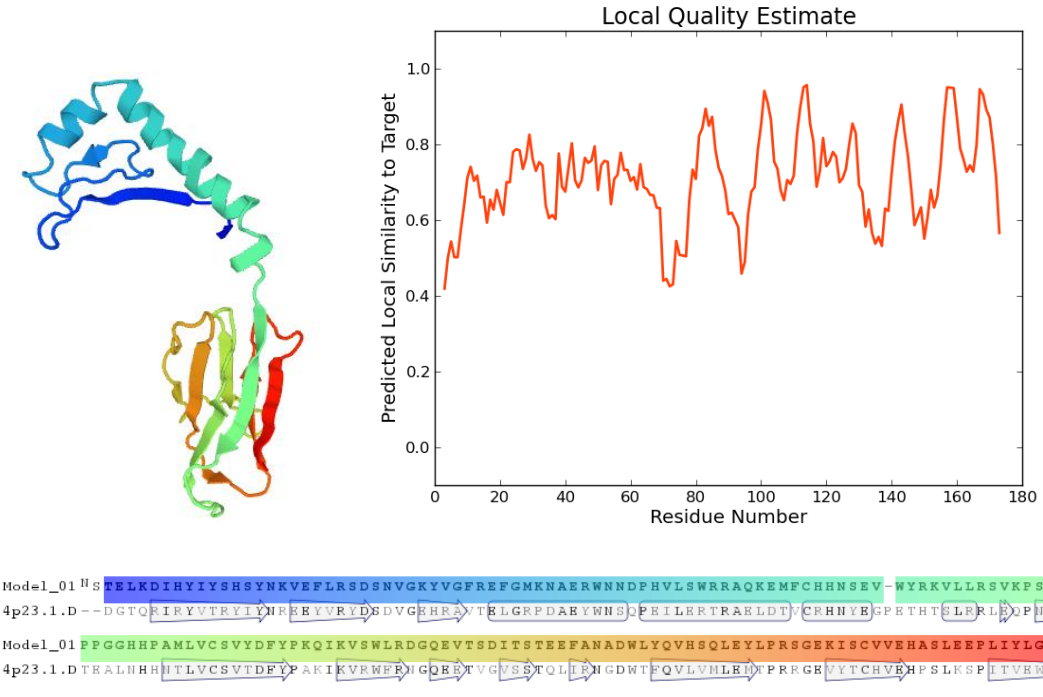


Sup. Figure I. 3. Protein homology #D models for all alleles built with Swiss-Model.

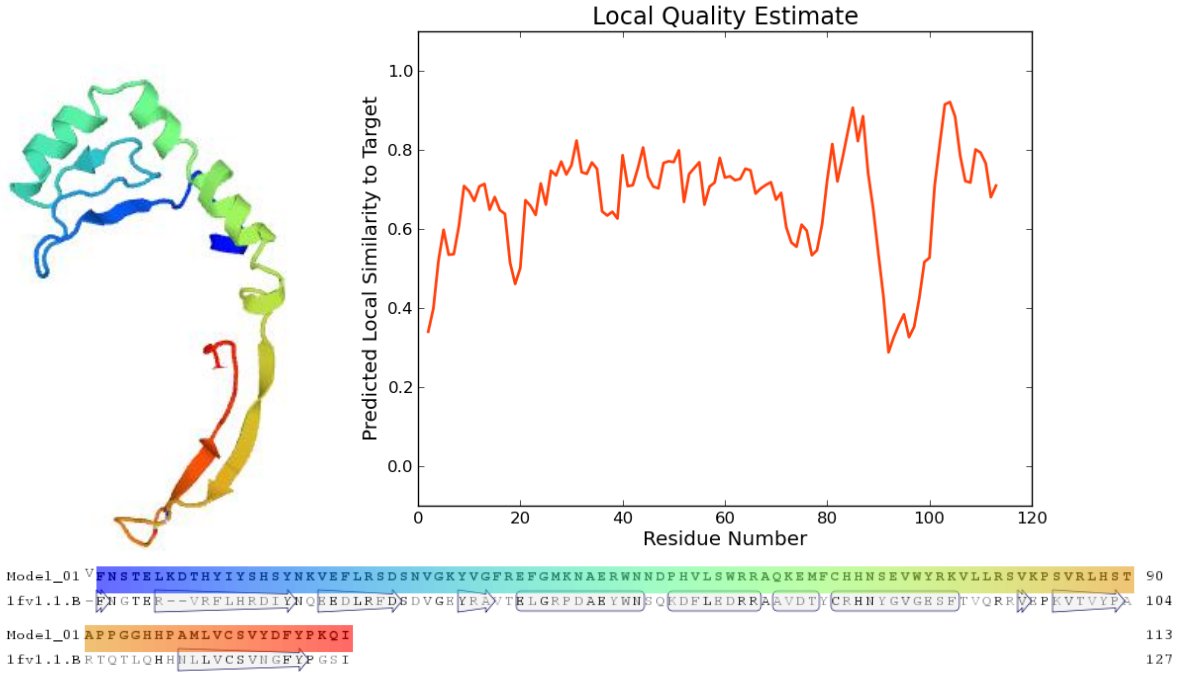
Amci-DXB*000101



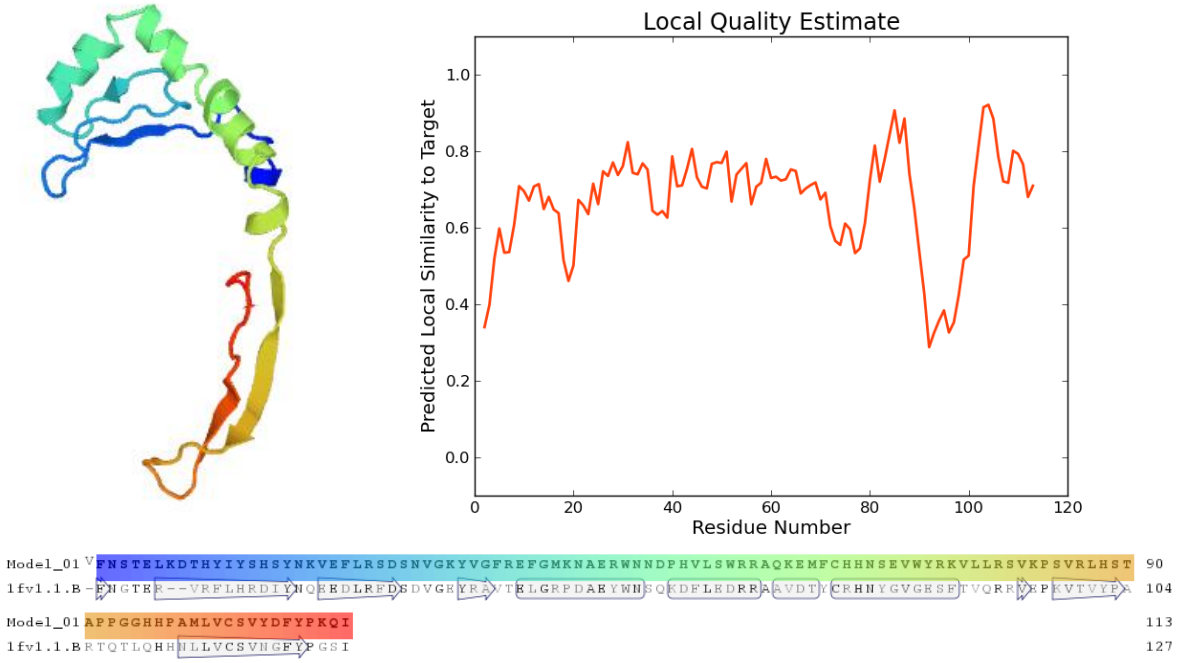
Amci-DXB*000102



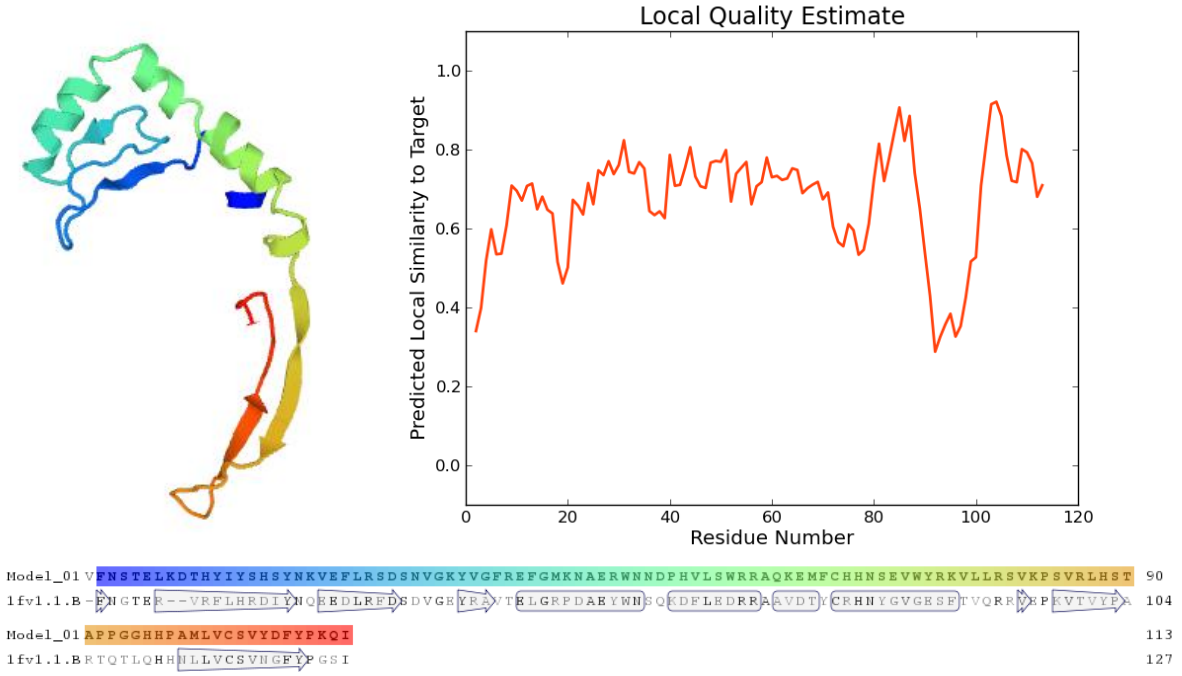
Amci-DXB*000103



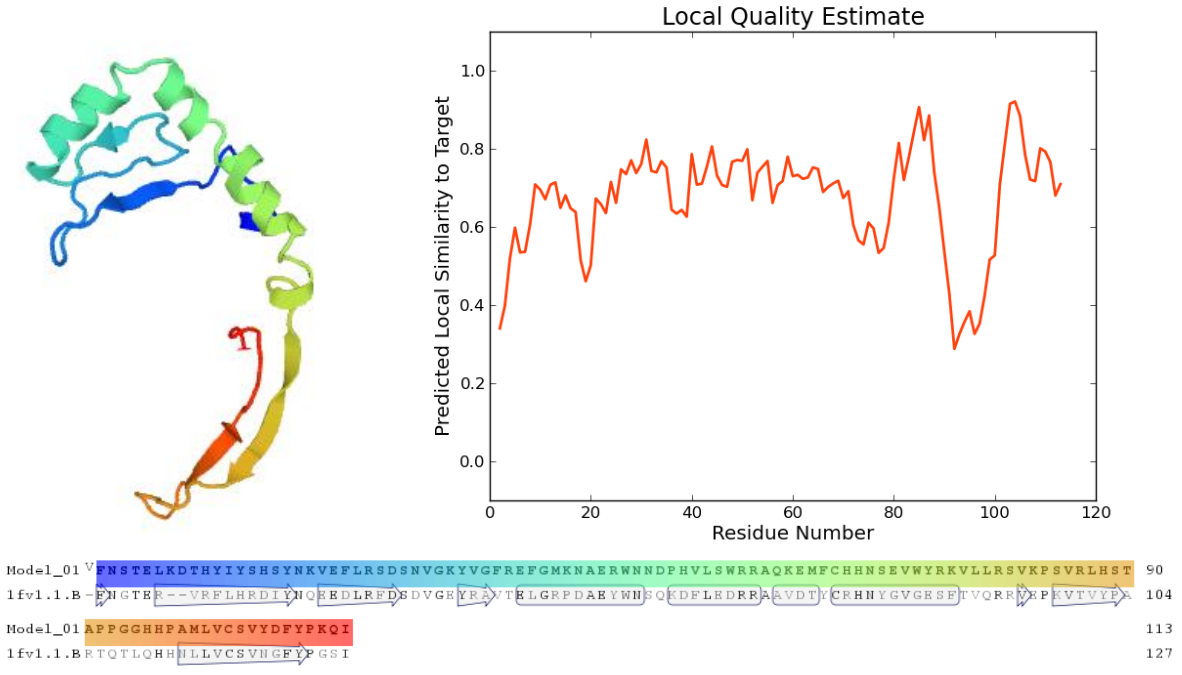
Amci-DXB*000104



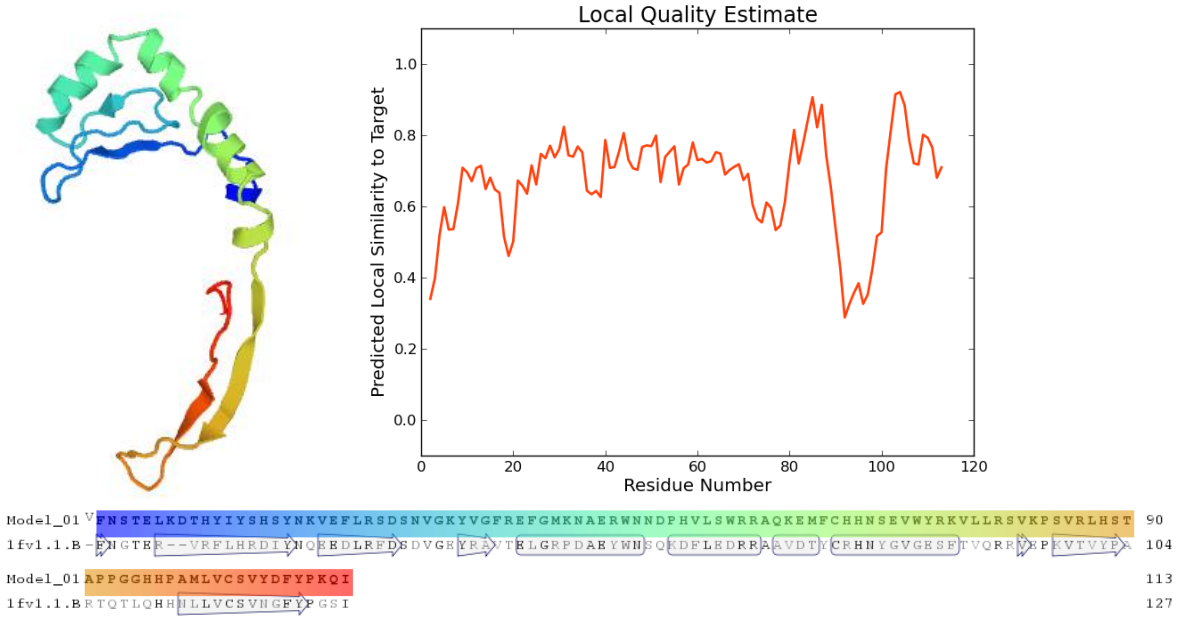
Amci-DXB*000105



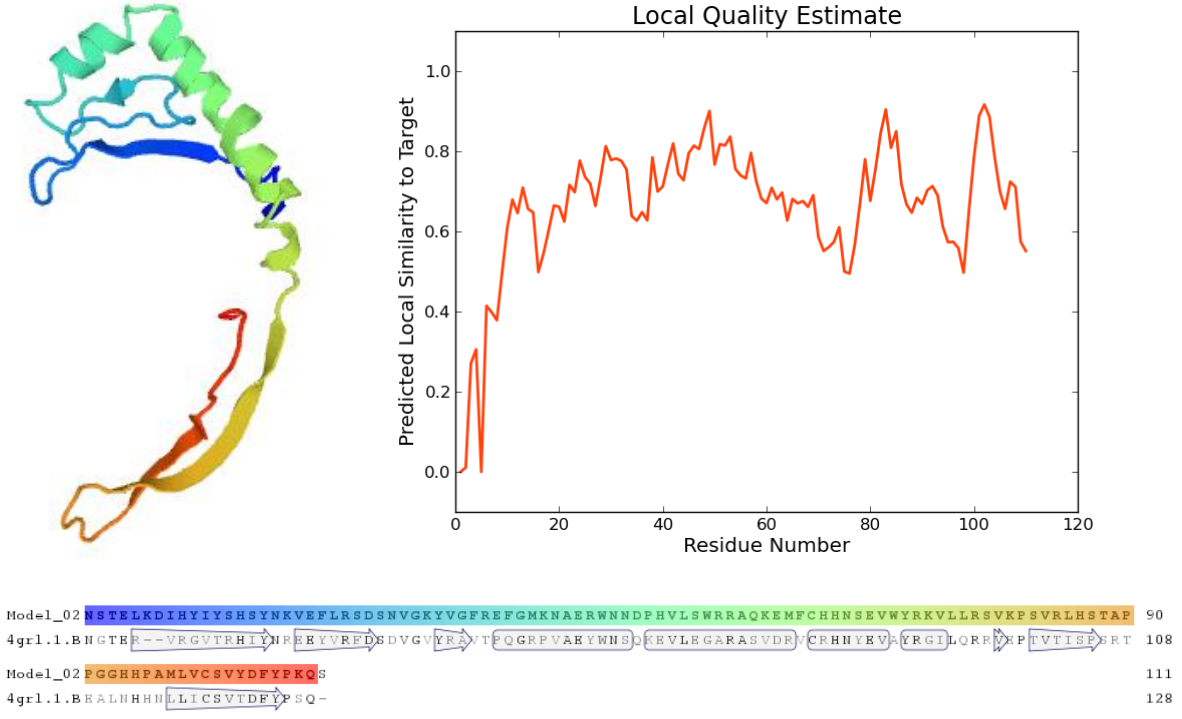
Amci-DXB*000106



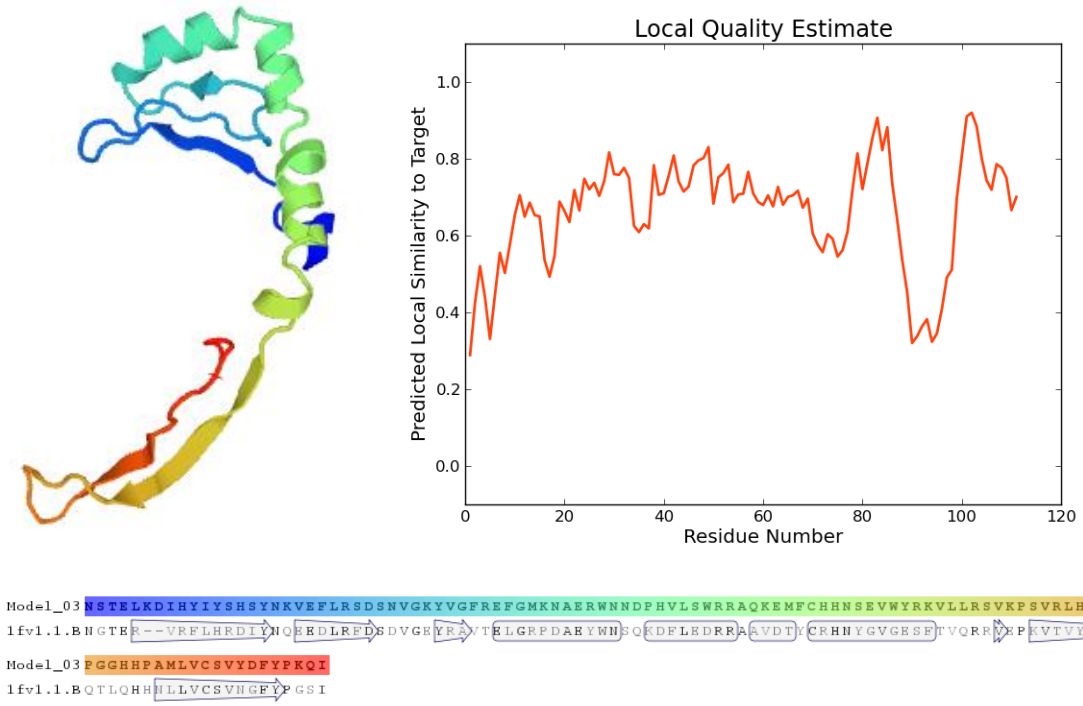
Amci-DXB*000107



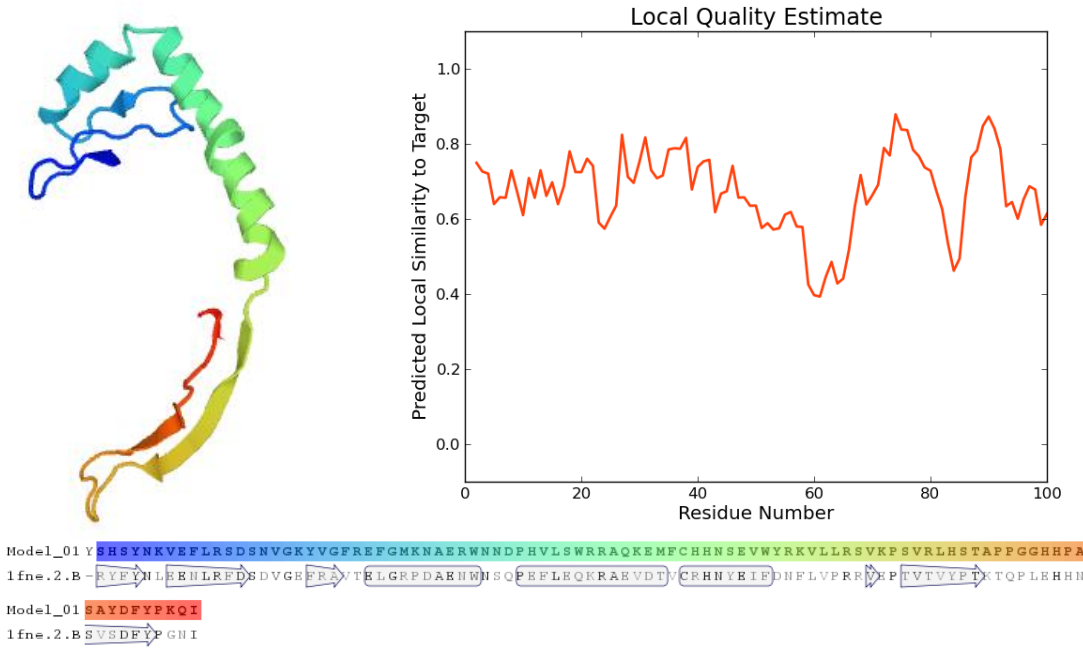
Amci-DXB*000108



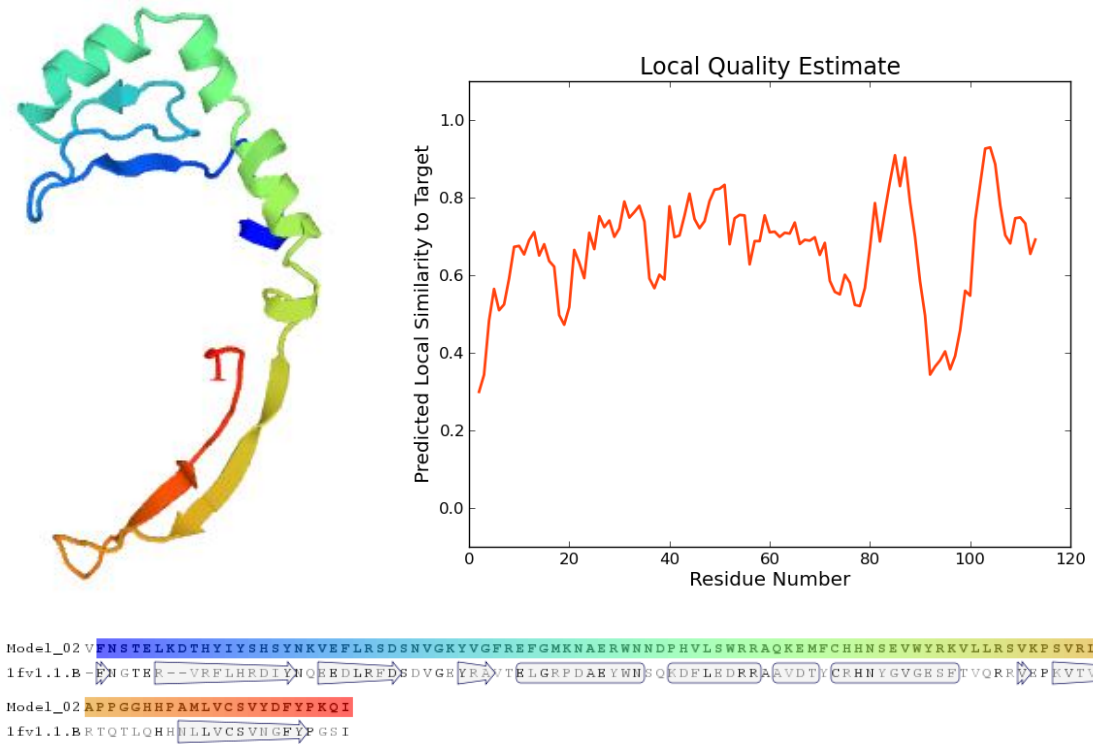
Amci-DXB*0002



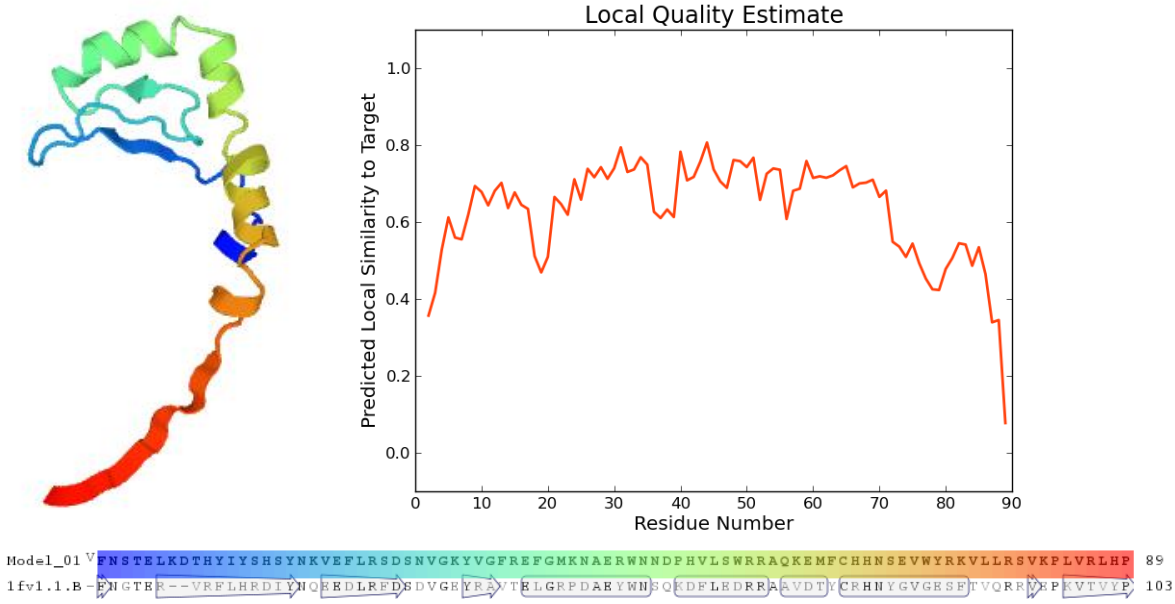
Amci-DXB*0003



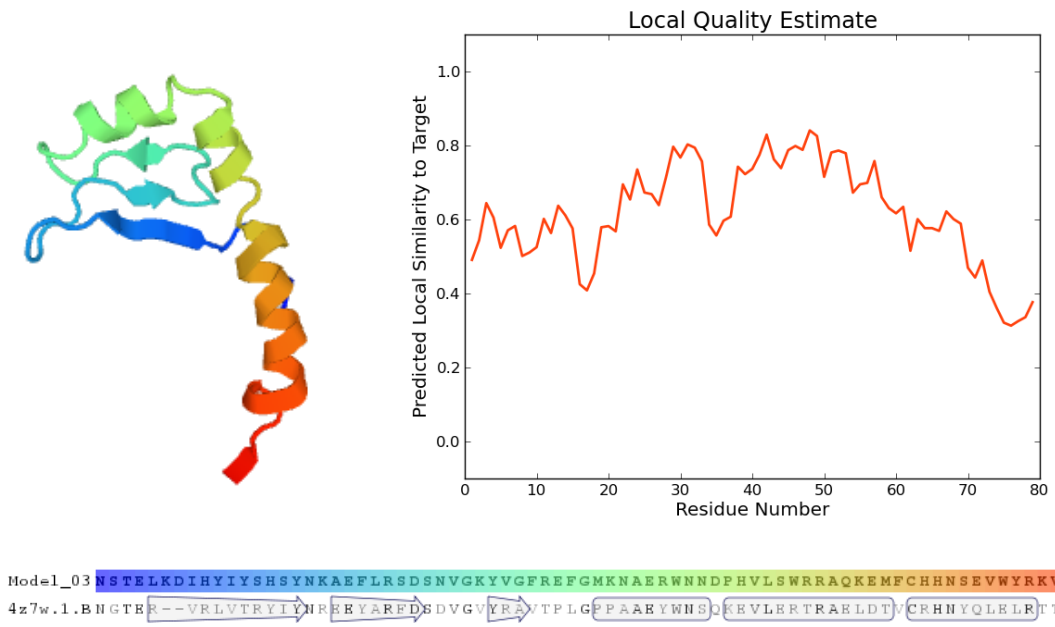
Amci-DXB*0004



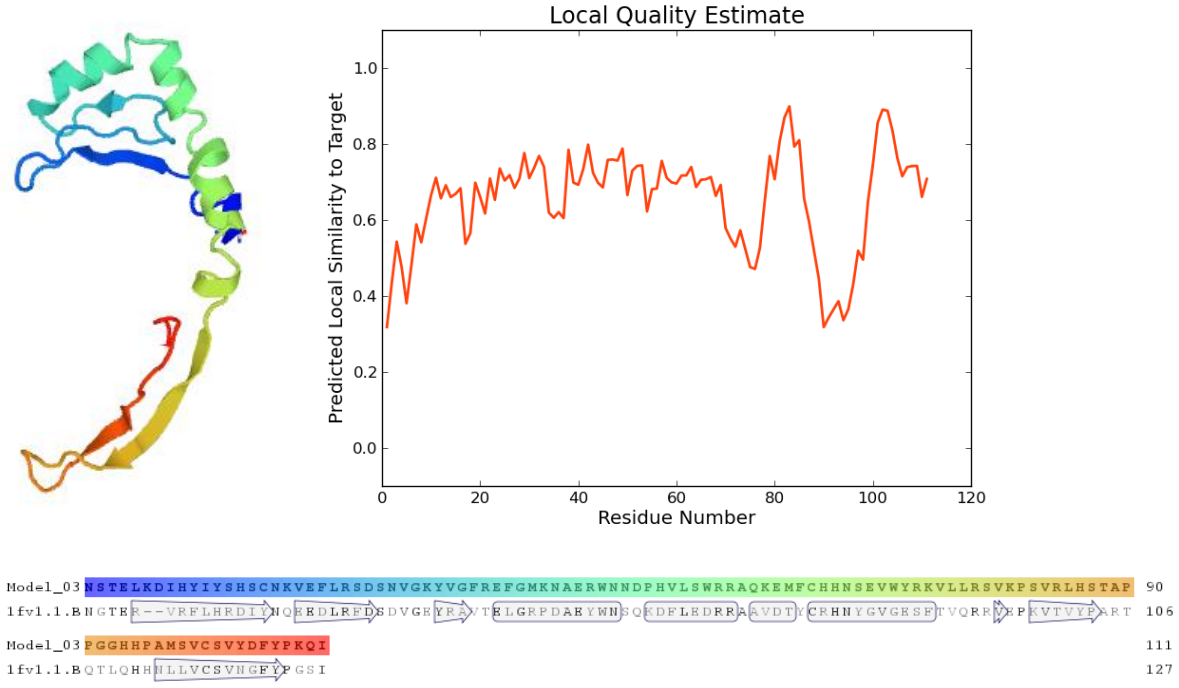
Amci-DXB*0005



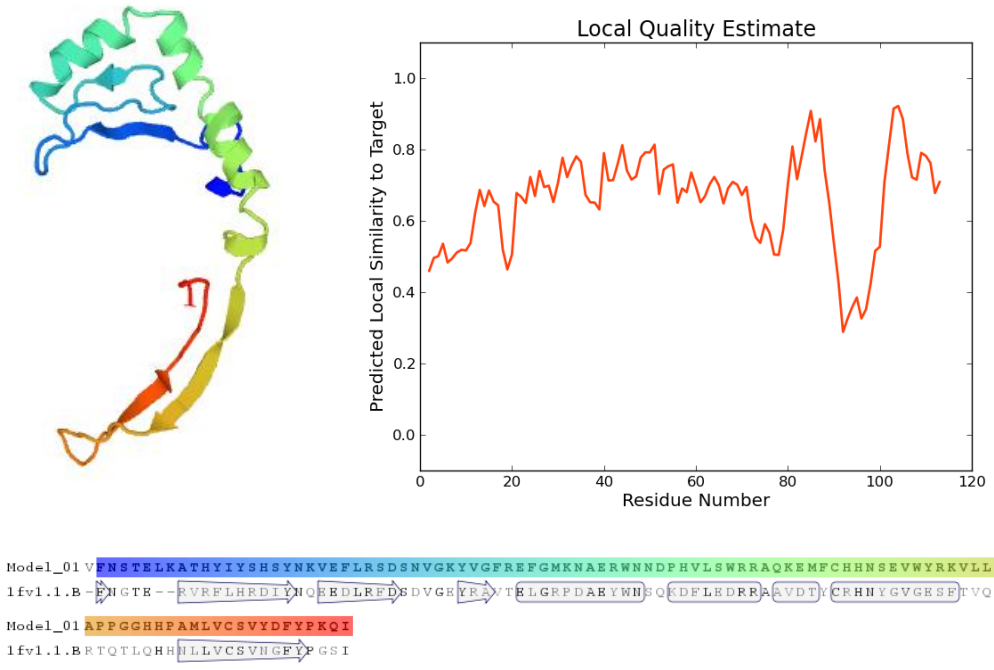
Amci-DXB*0006



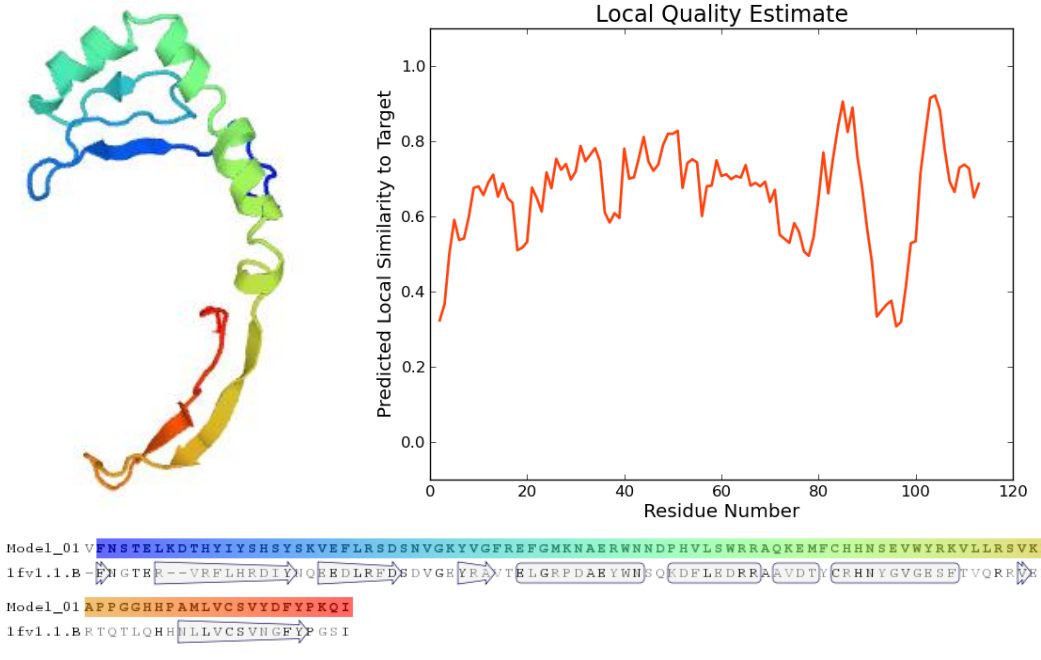
Amci-DXB*0007



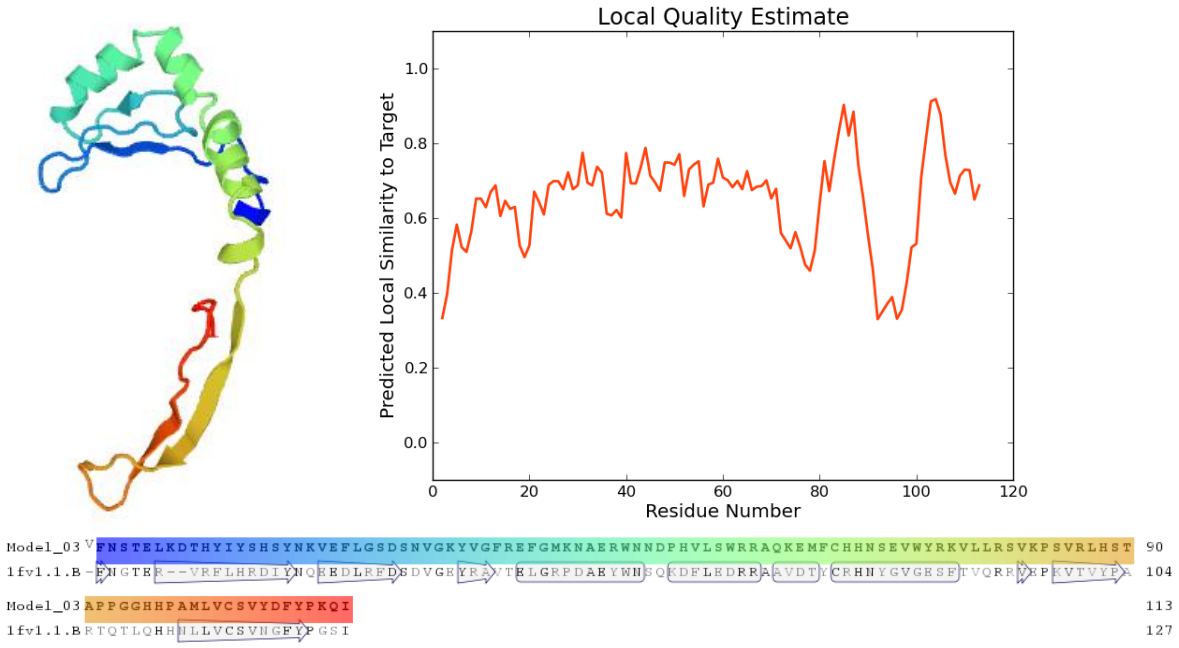
Amci-DXB*0008



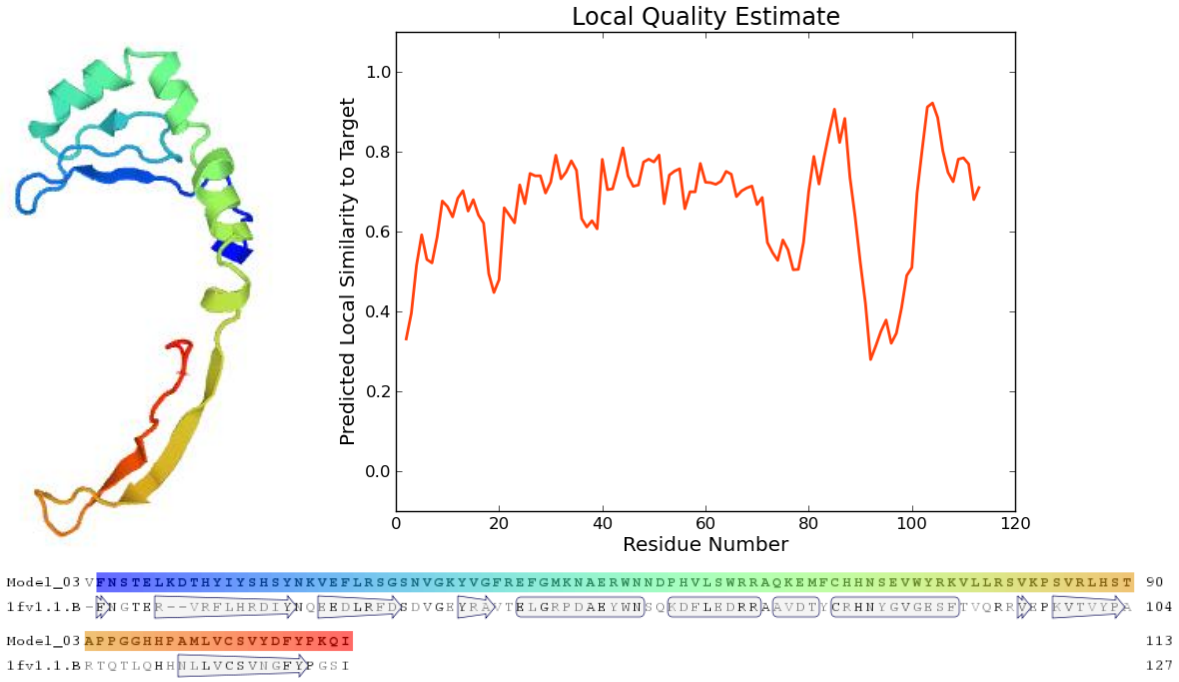
Amci-DXB*0009



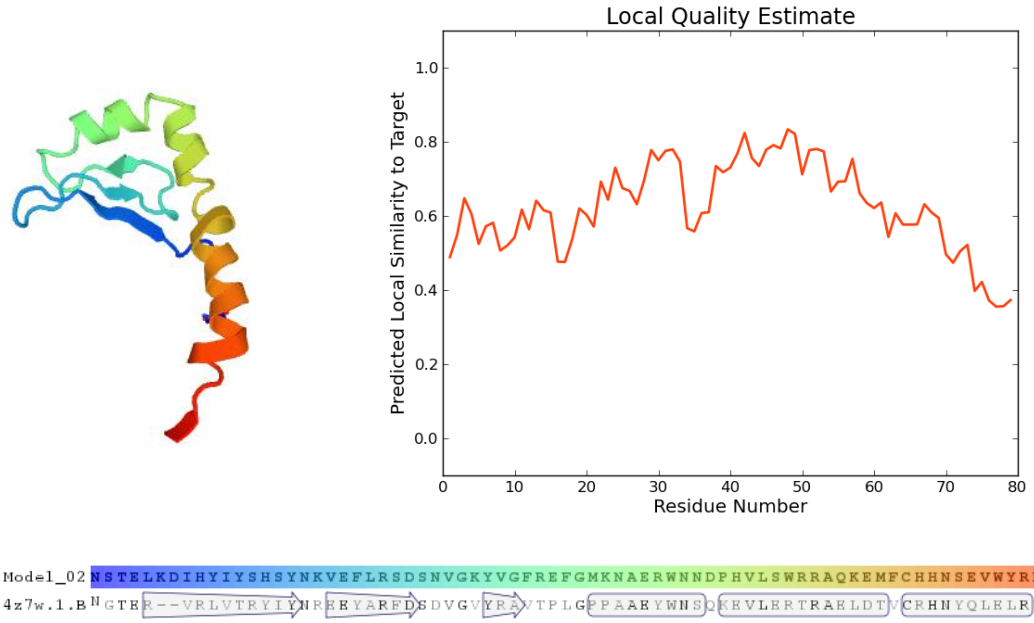
Amci-DXB*0010



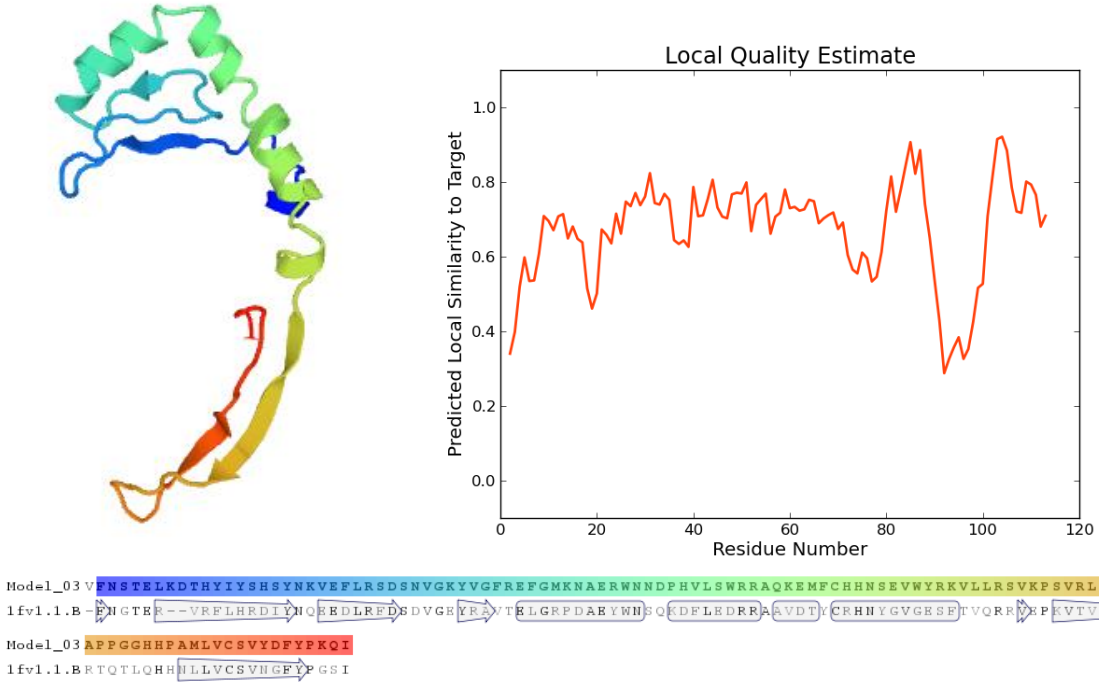
Amci-DXB*0011



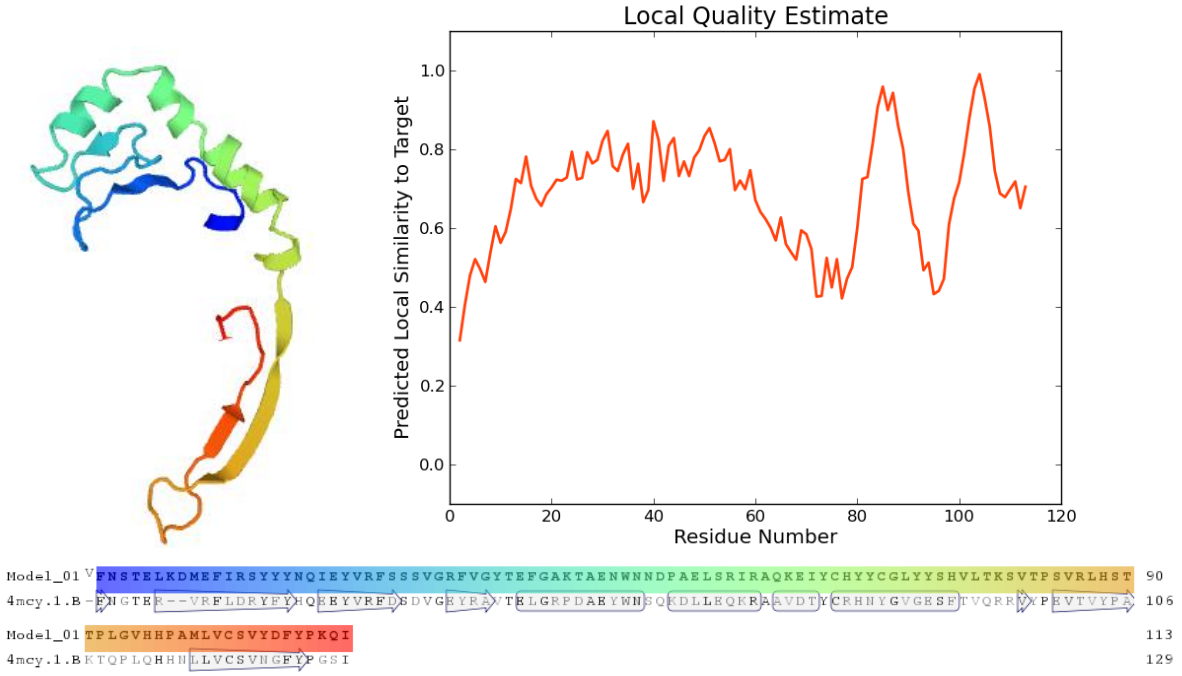
Amci-DXB*0012



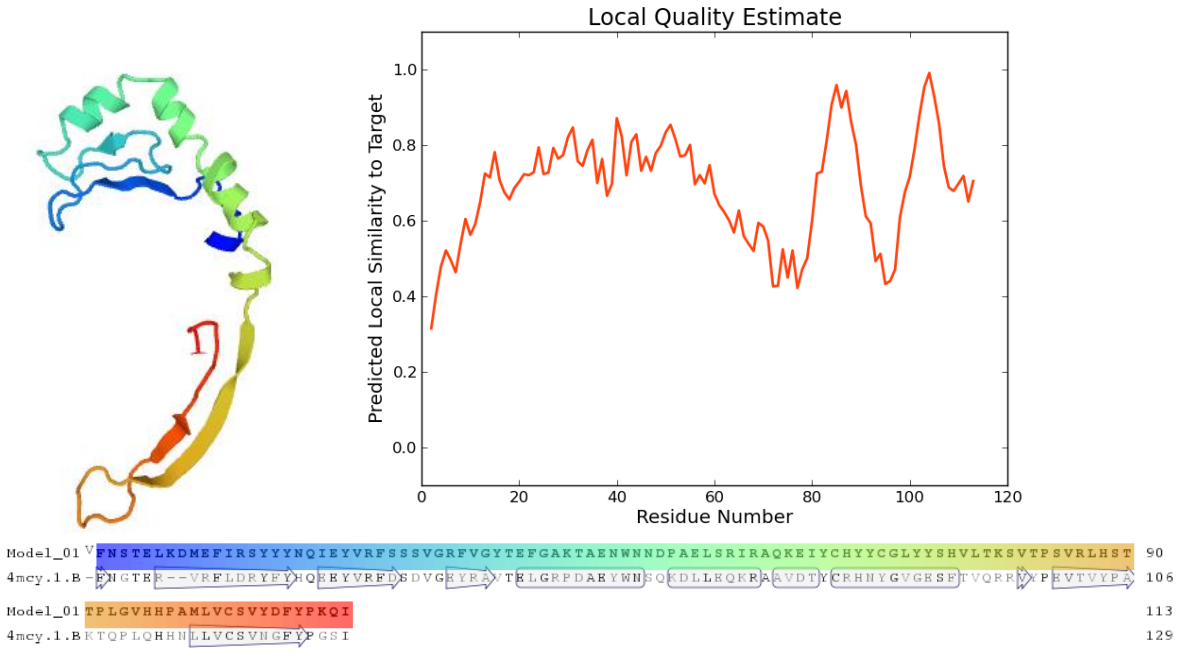
Amci-DXB*0013



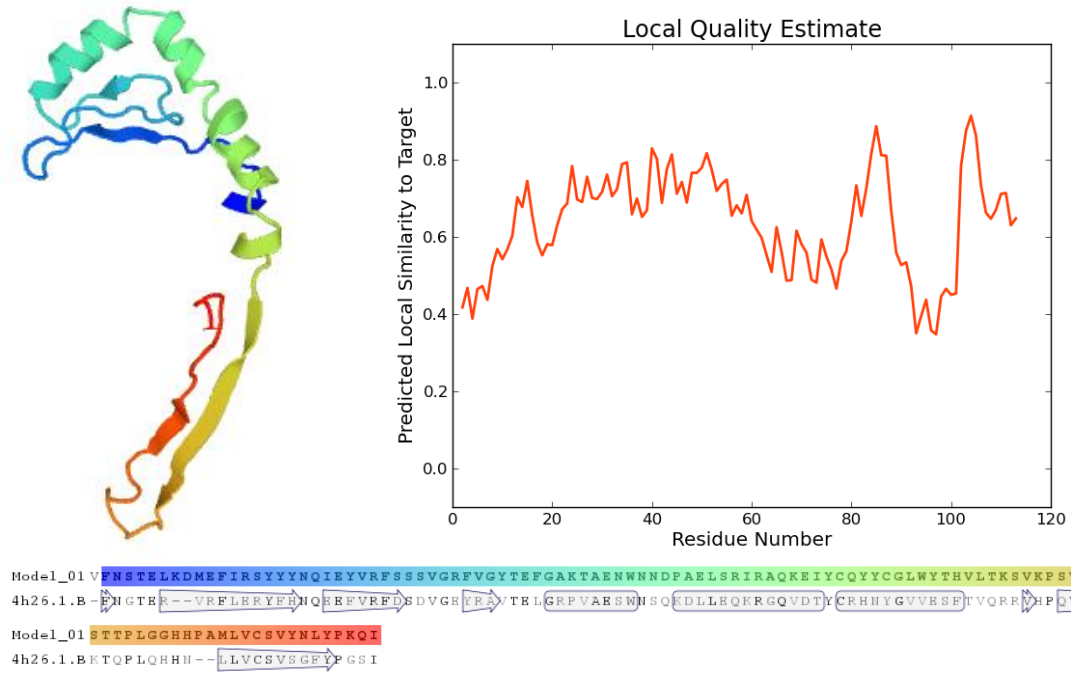
Amci-DXB*0101



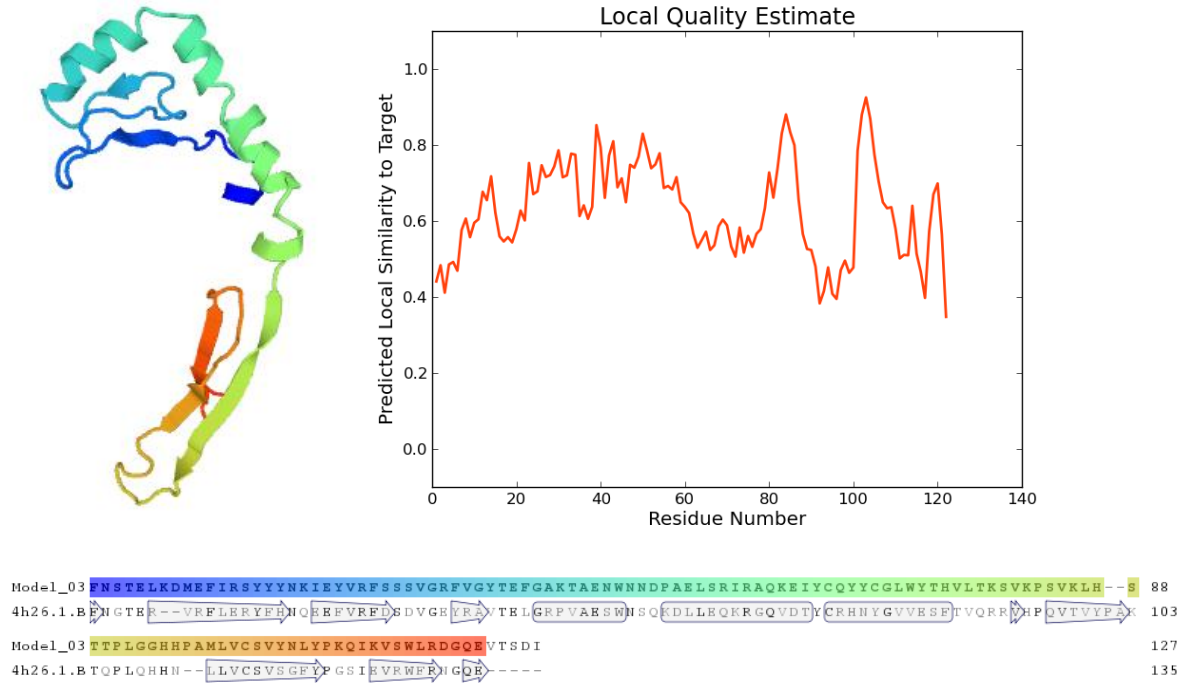
Amci-DXB*0102



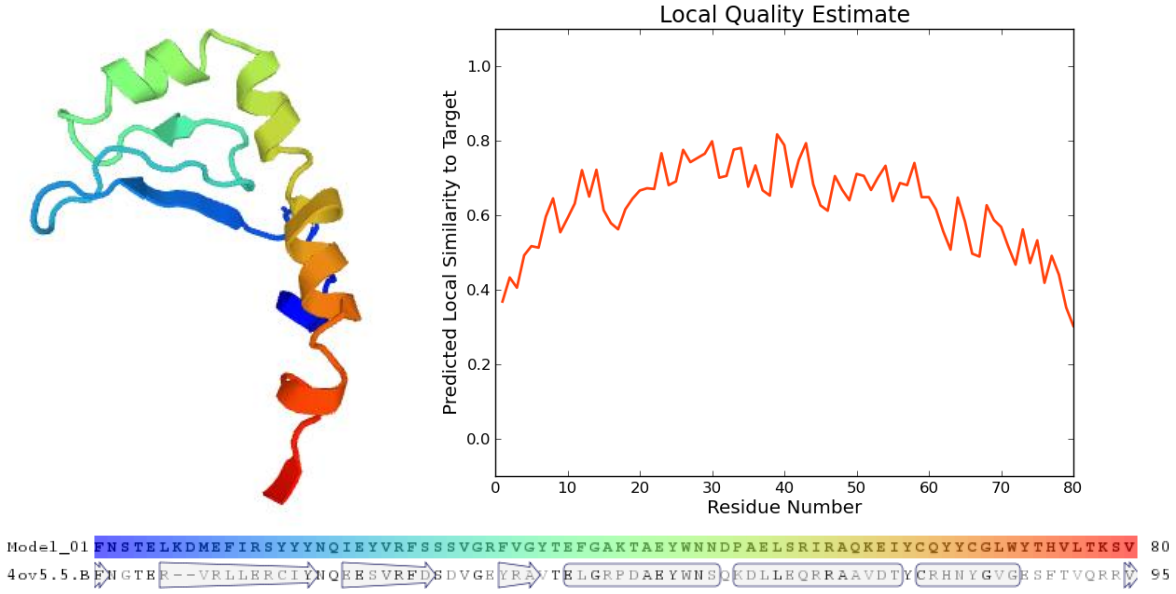
Amci-DXB*0201



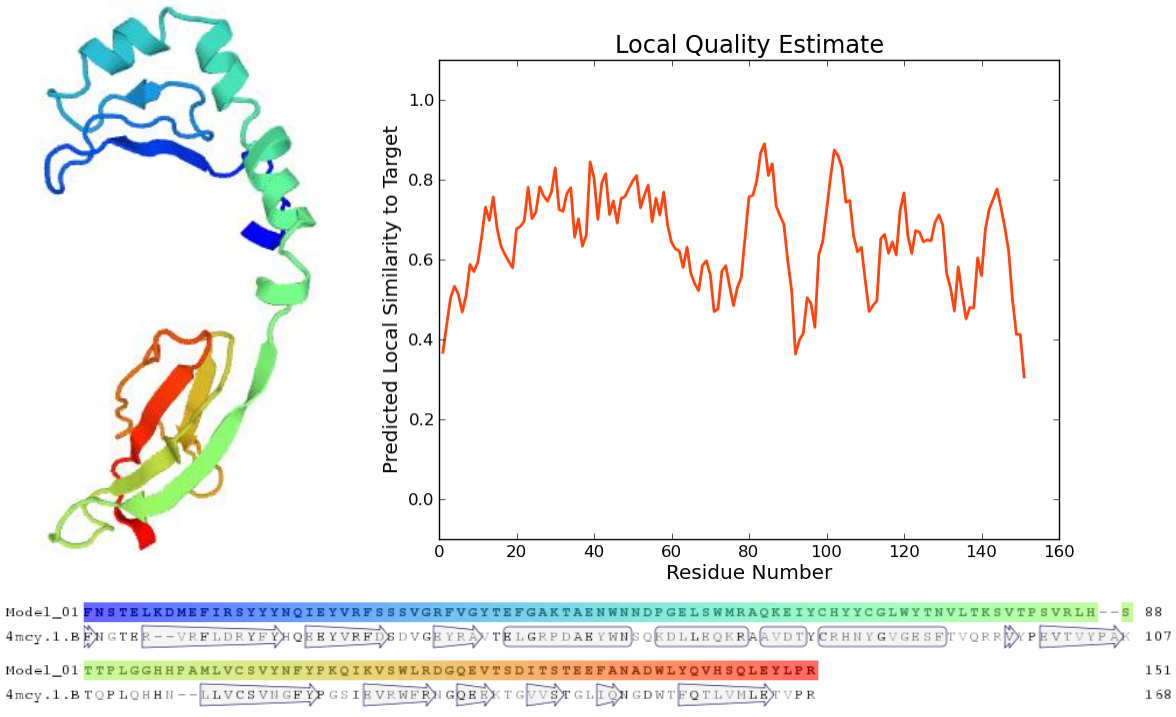
Amci-DXB*0202



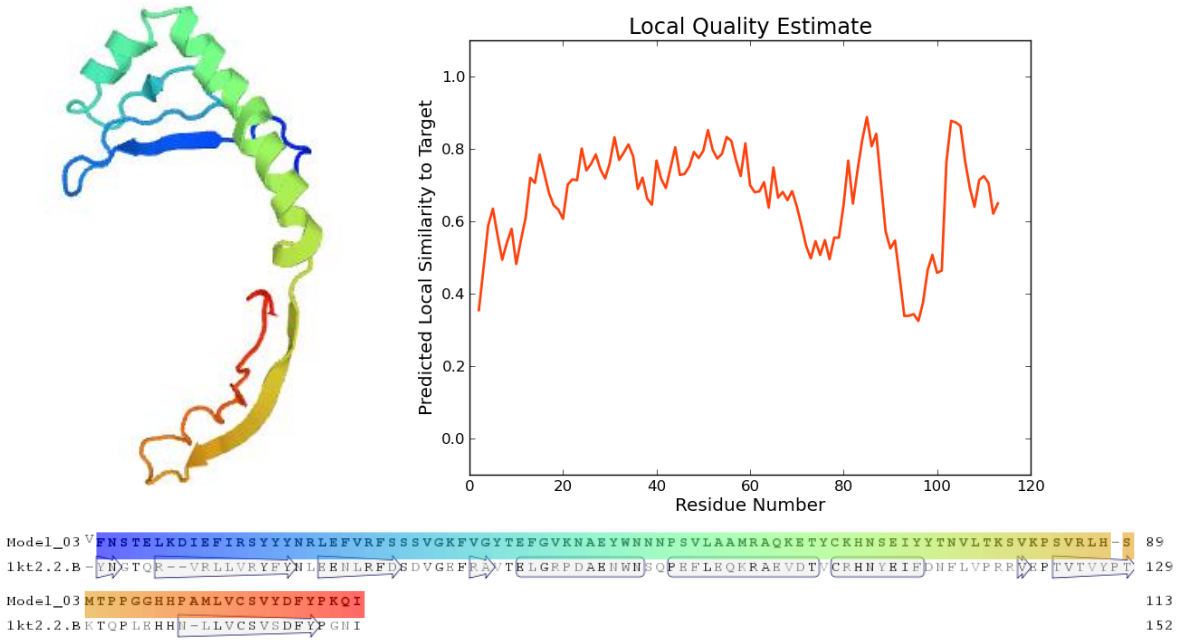
Amci-DXB*0203



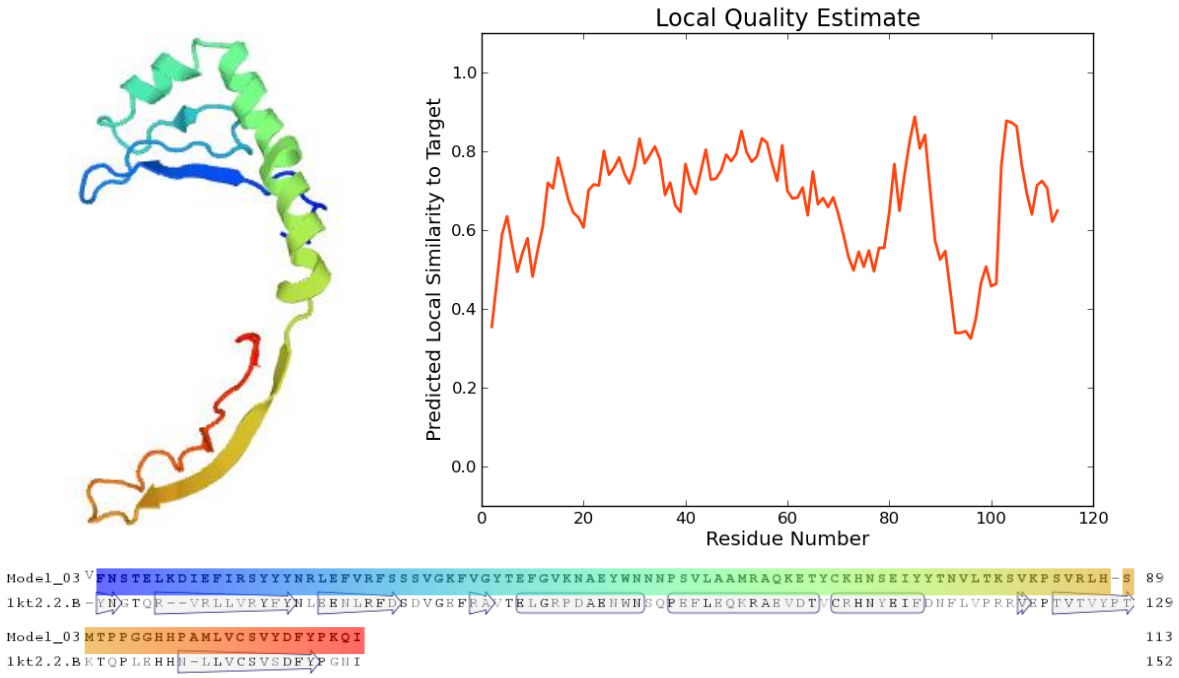
Amci-DXB*03



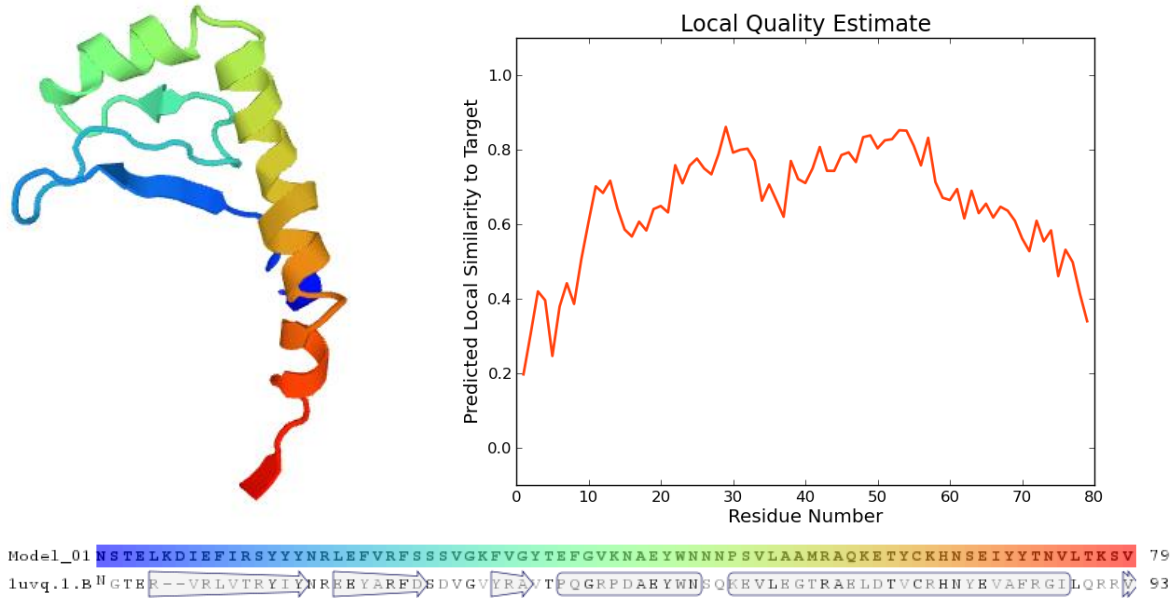
Amci-DXB*040101



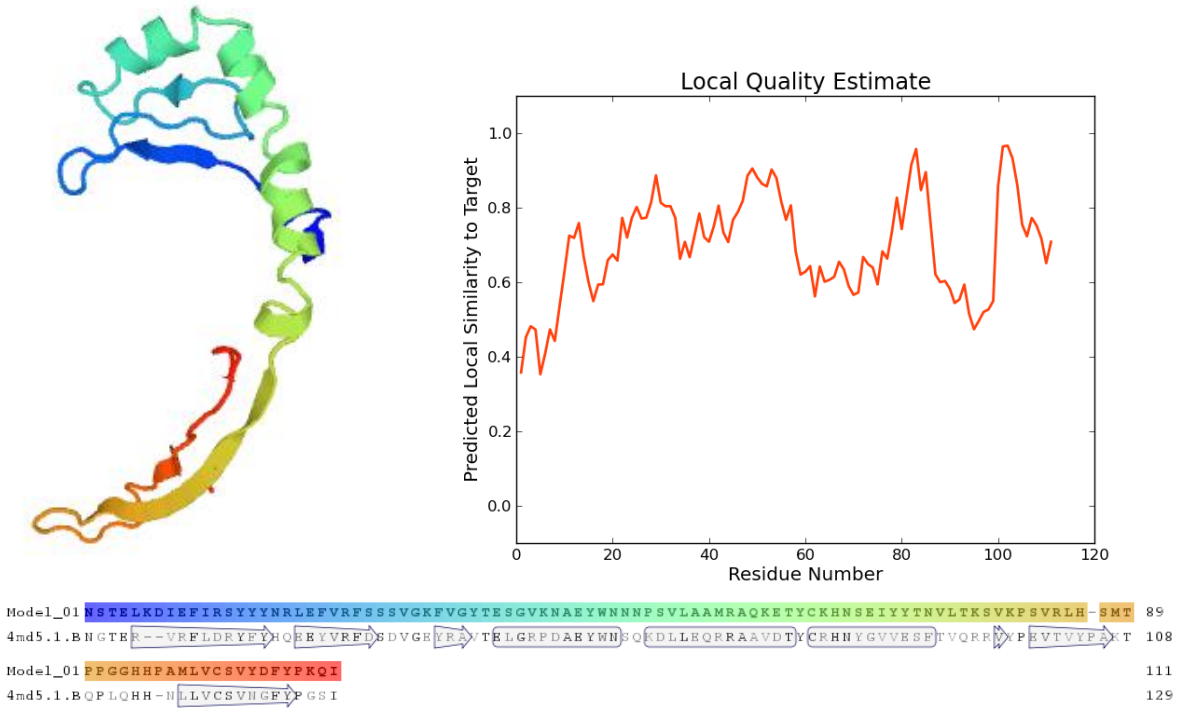
Amci-DXB*040102



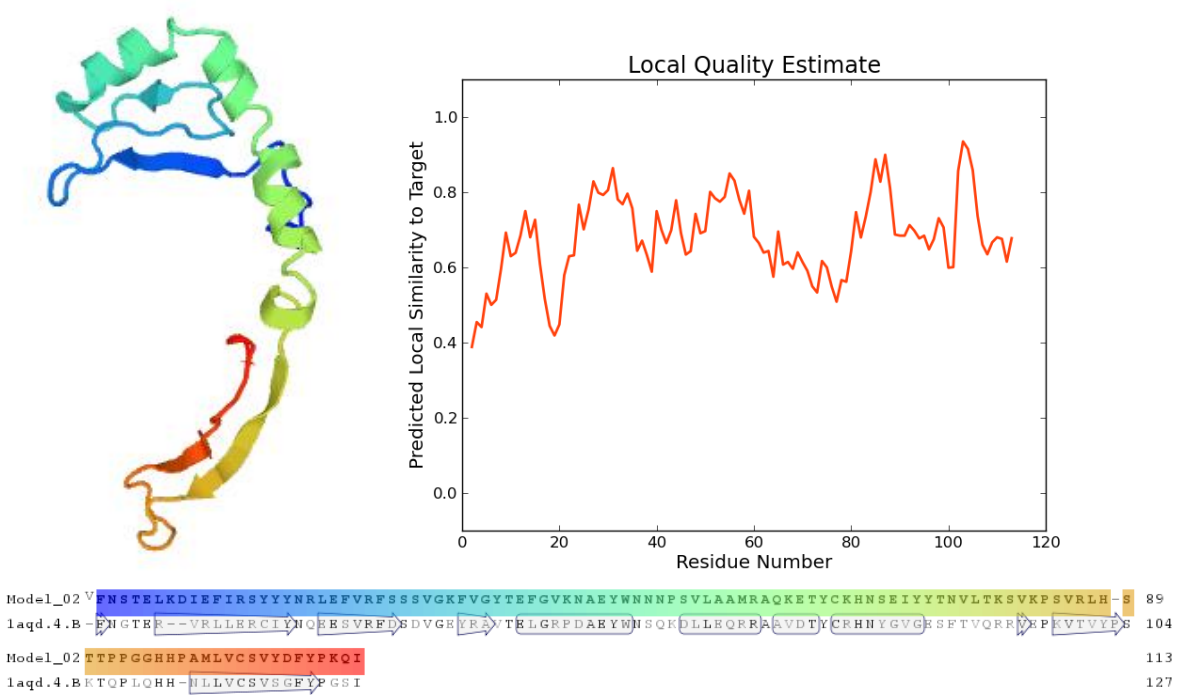
Amci-DXB*040103



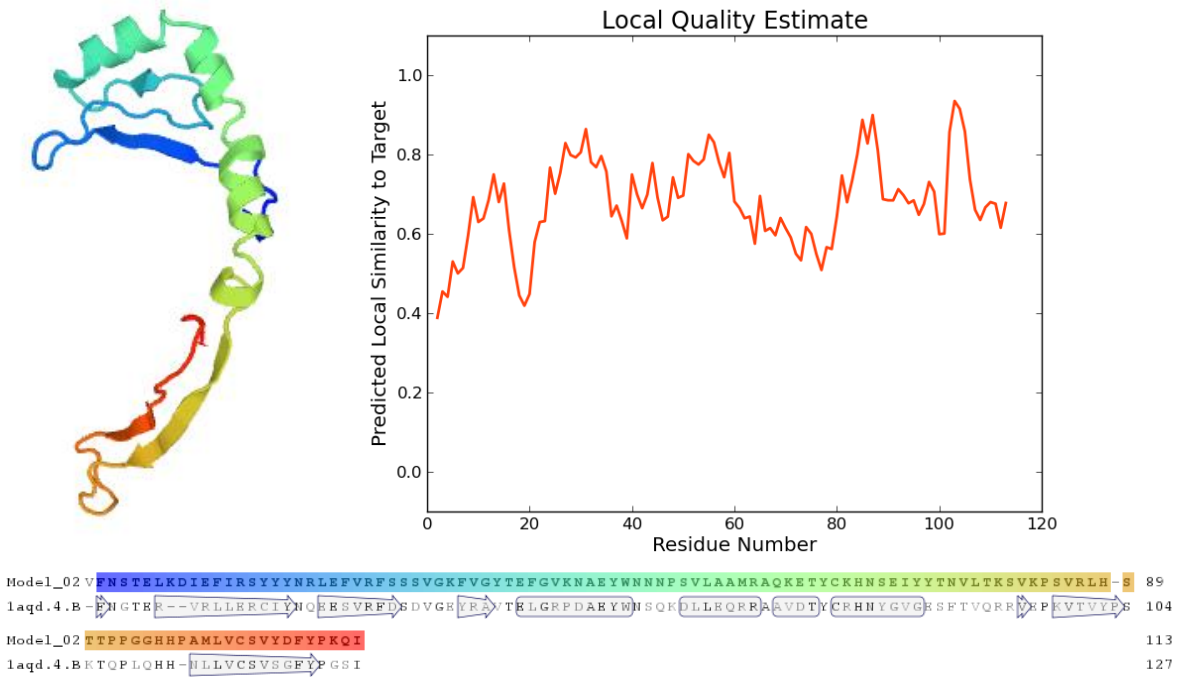
Amci-DXB*0402



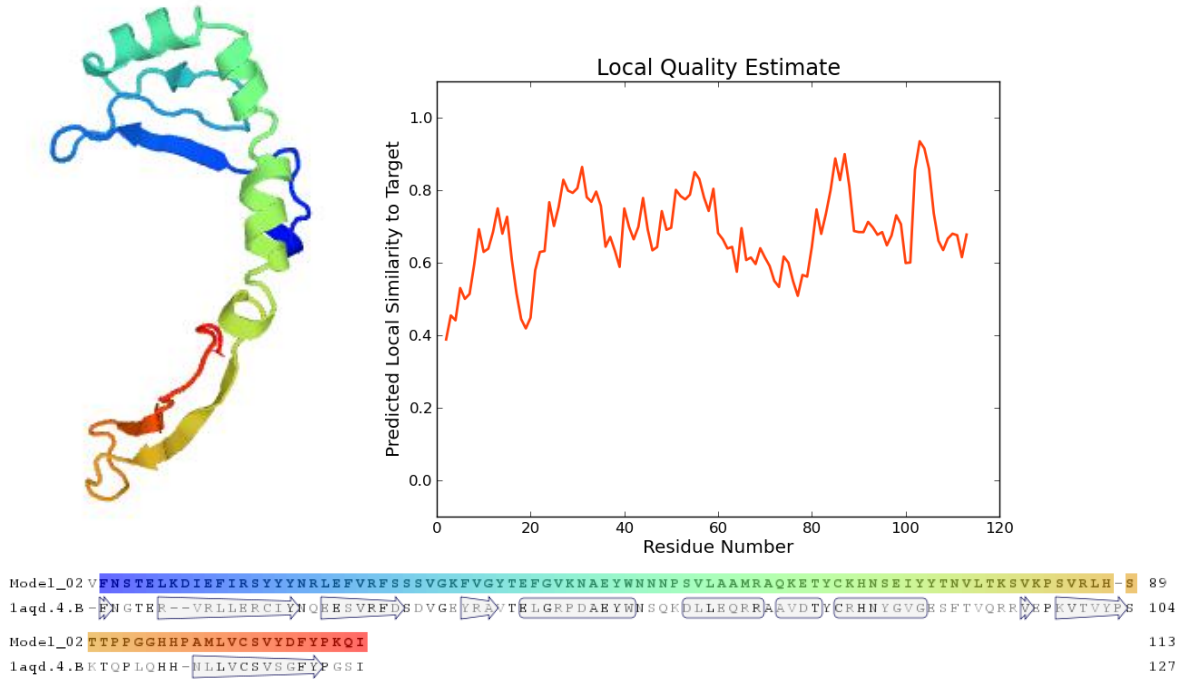
Amci-DXB*040301



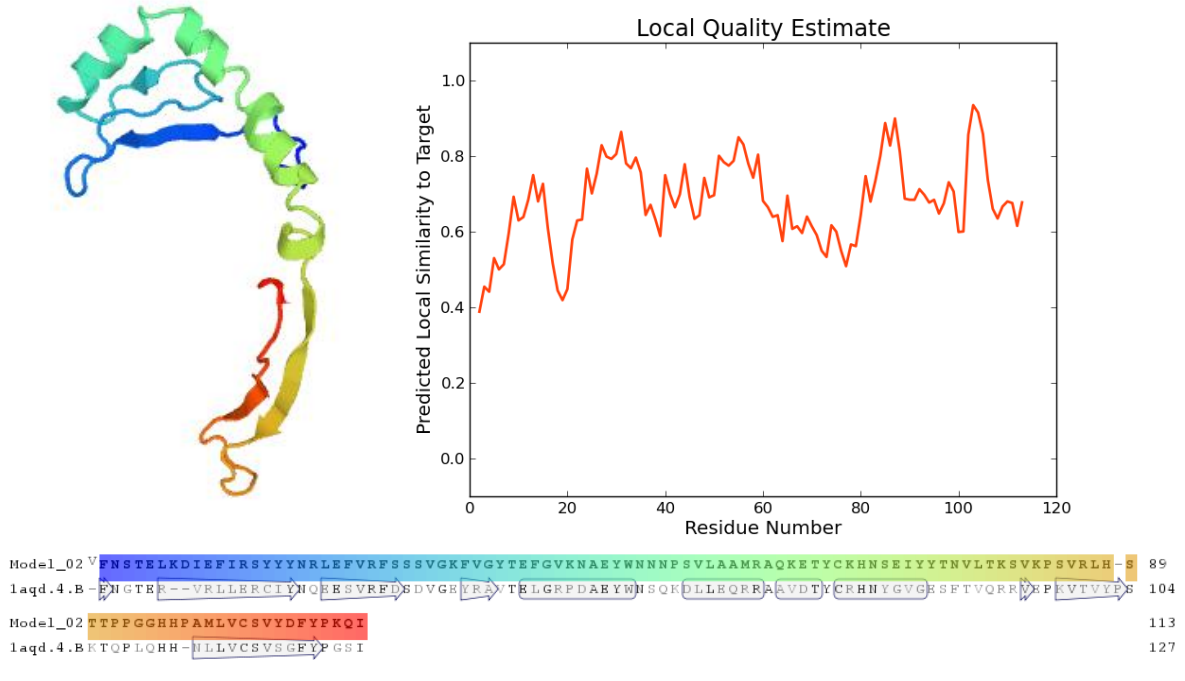
Amci-DXB*040302



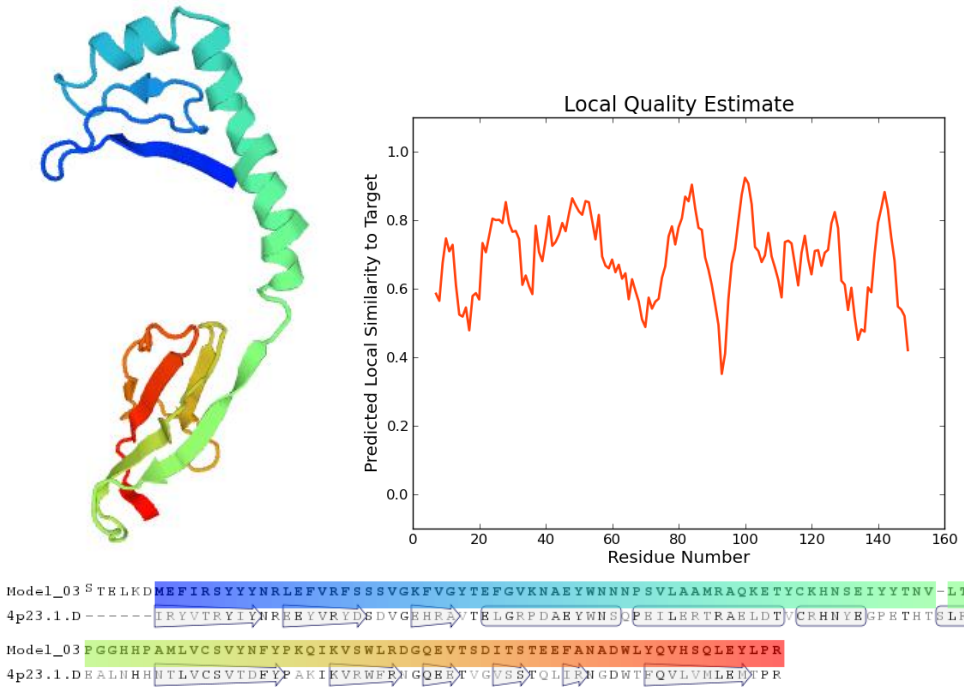
Amci-DXB*040303



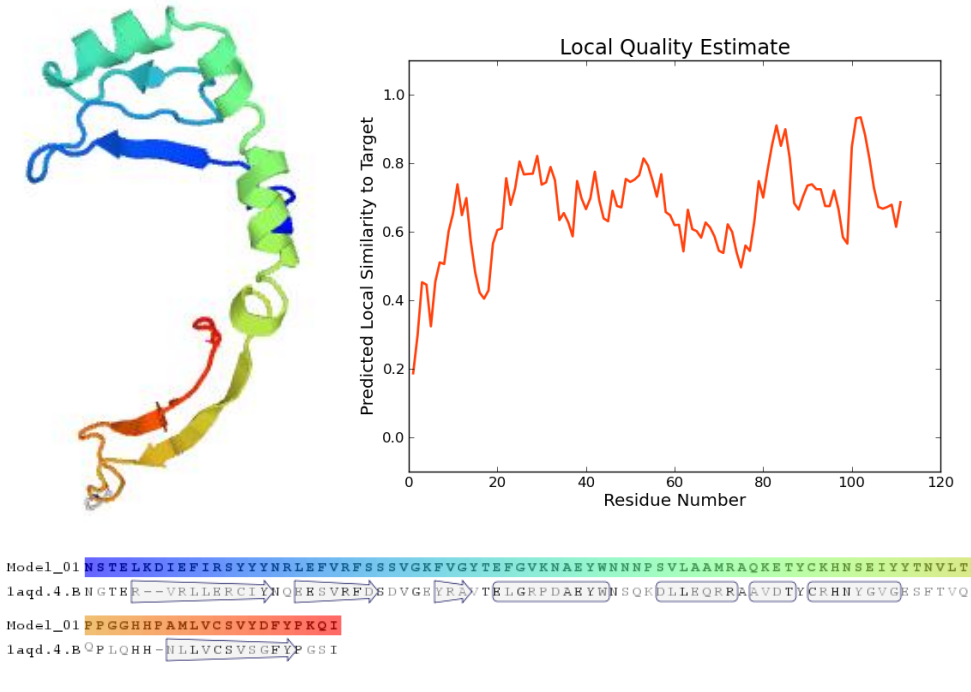
Amci-DXB*040304



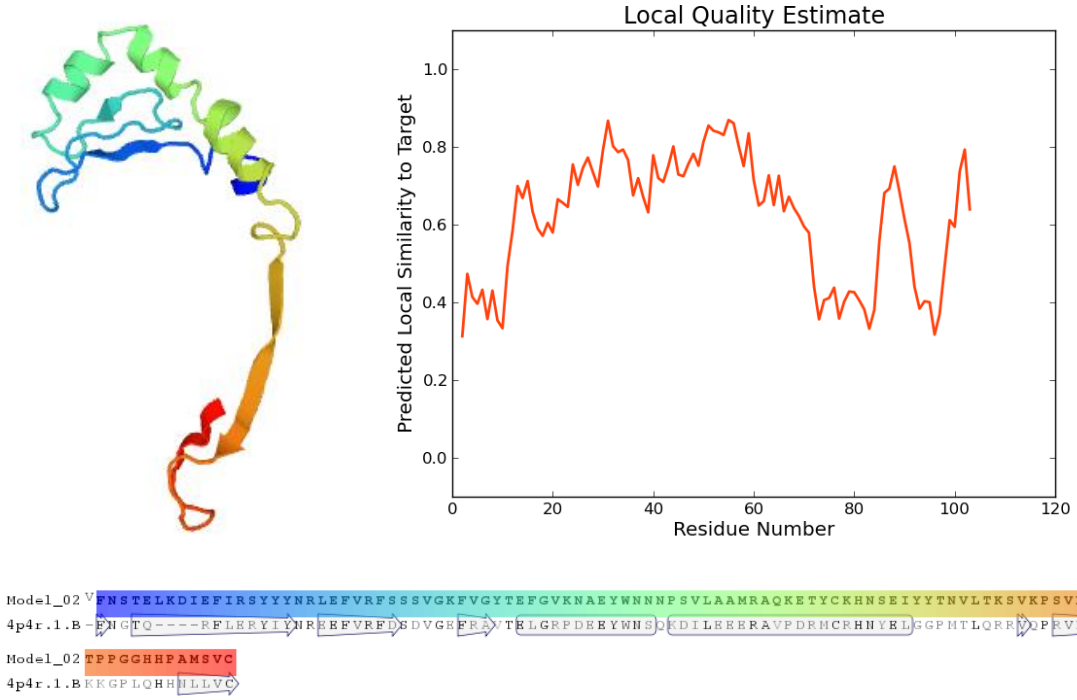
Amci-DXB*040305



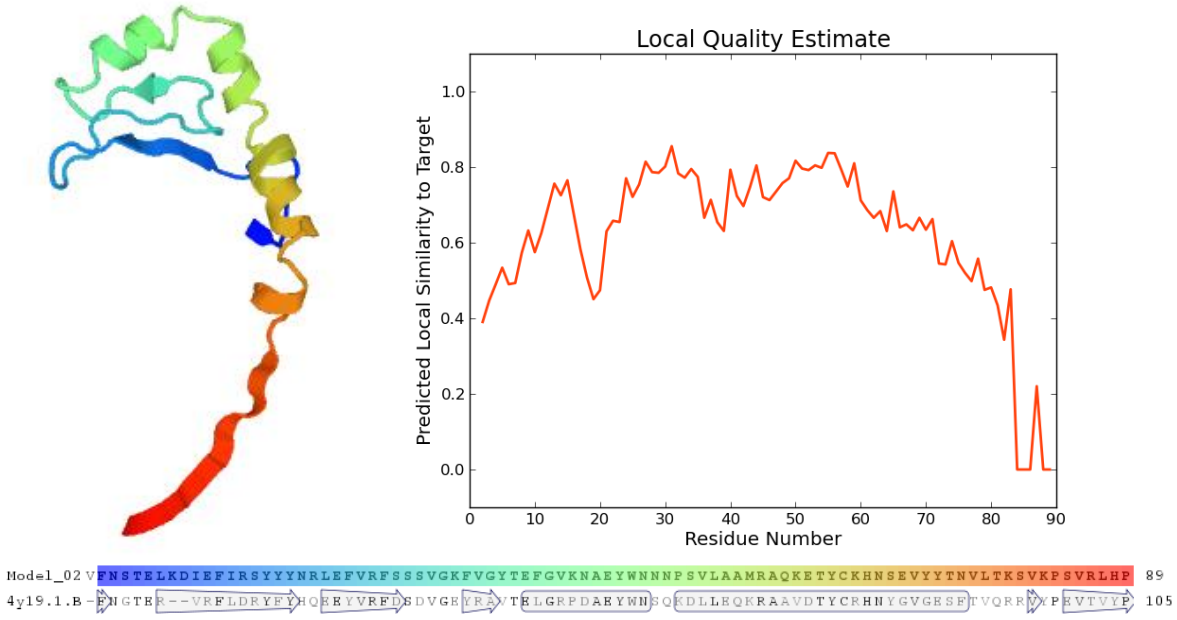
Amci-DXB*0404



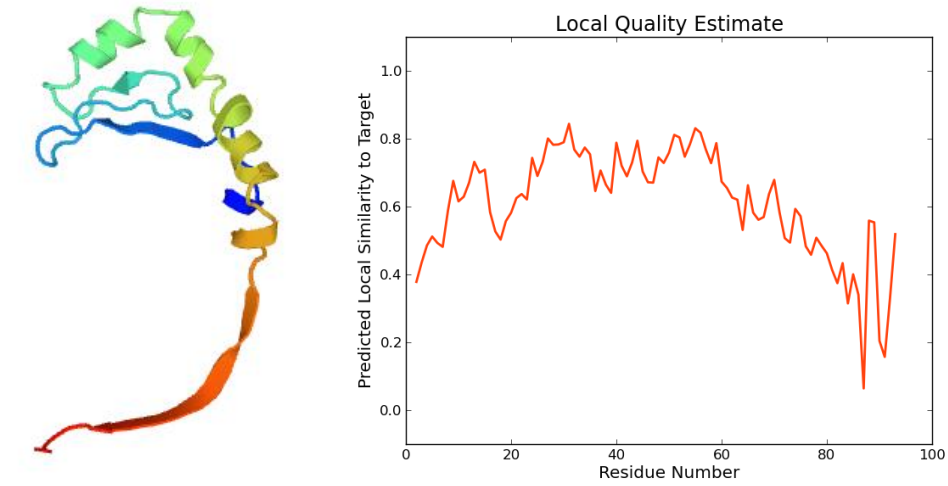
Amci-DXB*0405



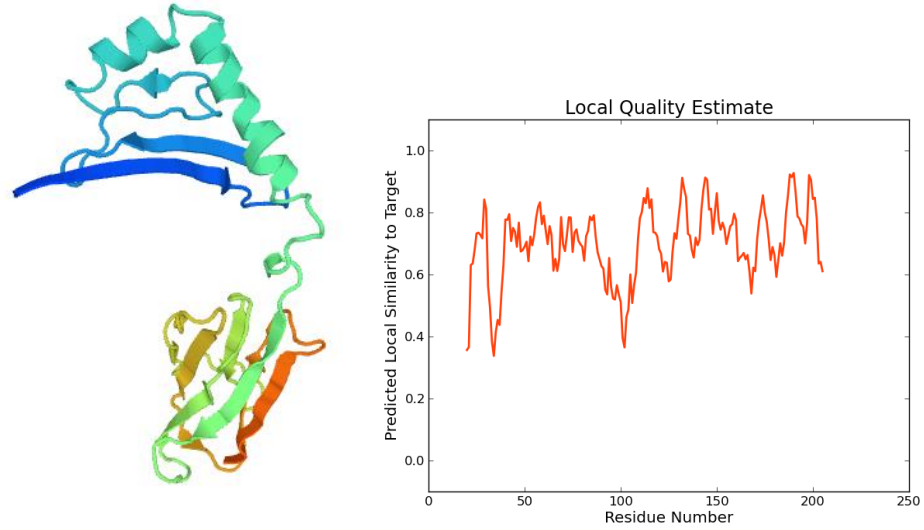
Amci-DXB*0406

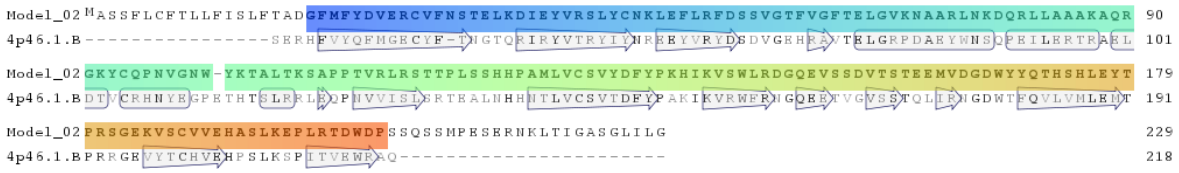


Amci-DXB*05

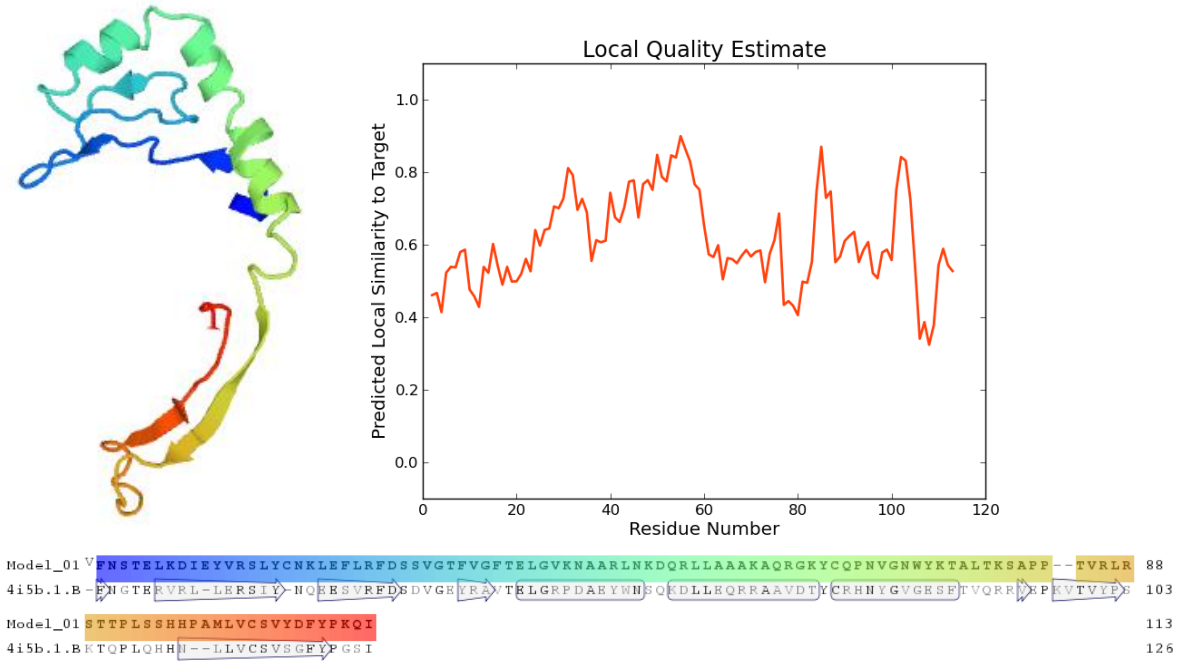


Amci-DXB*060101

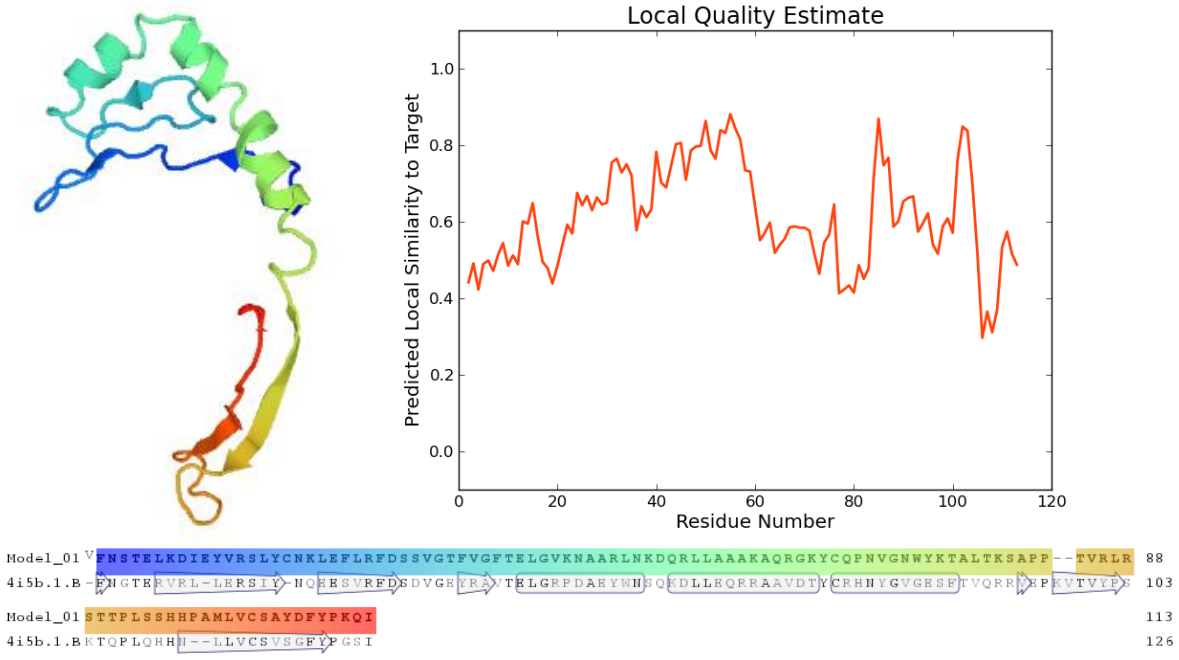




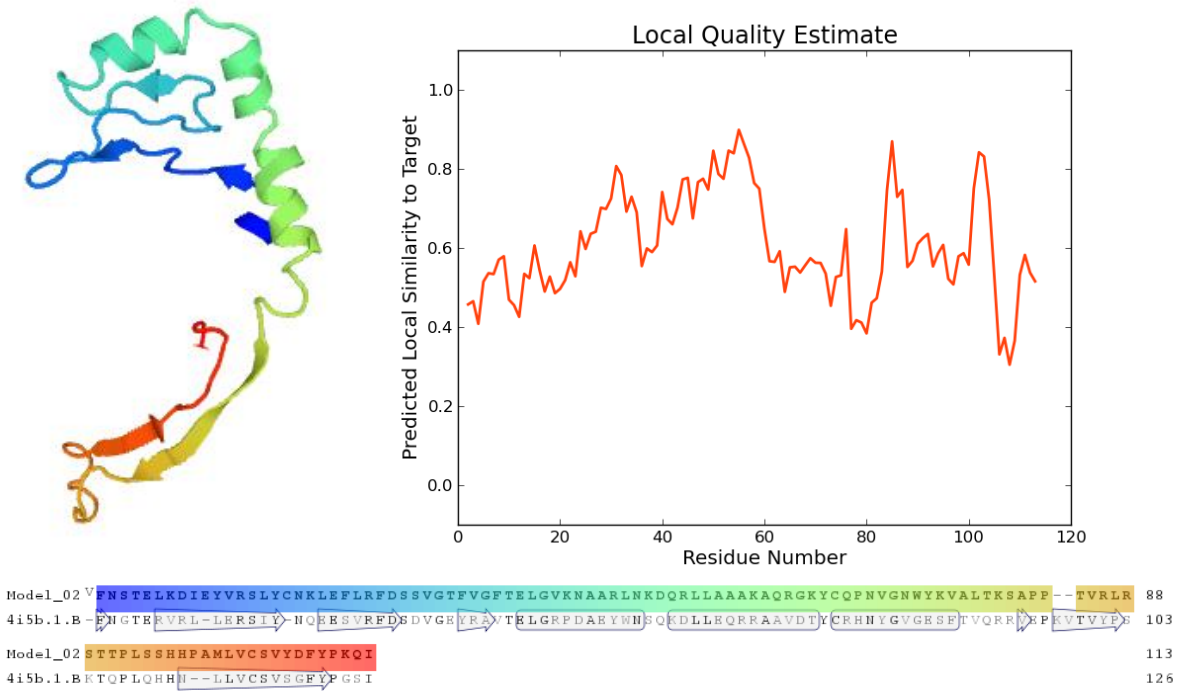
Amci-DXB*060102



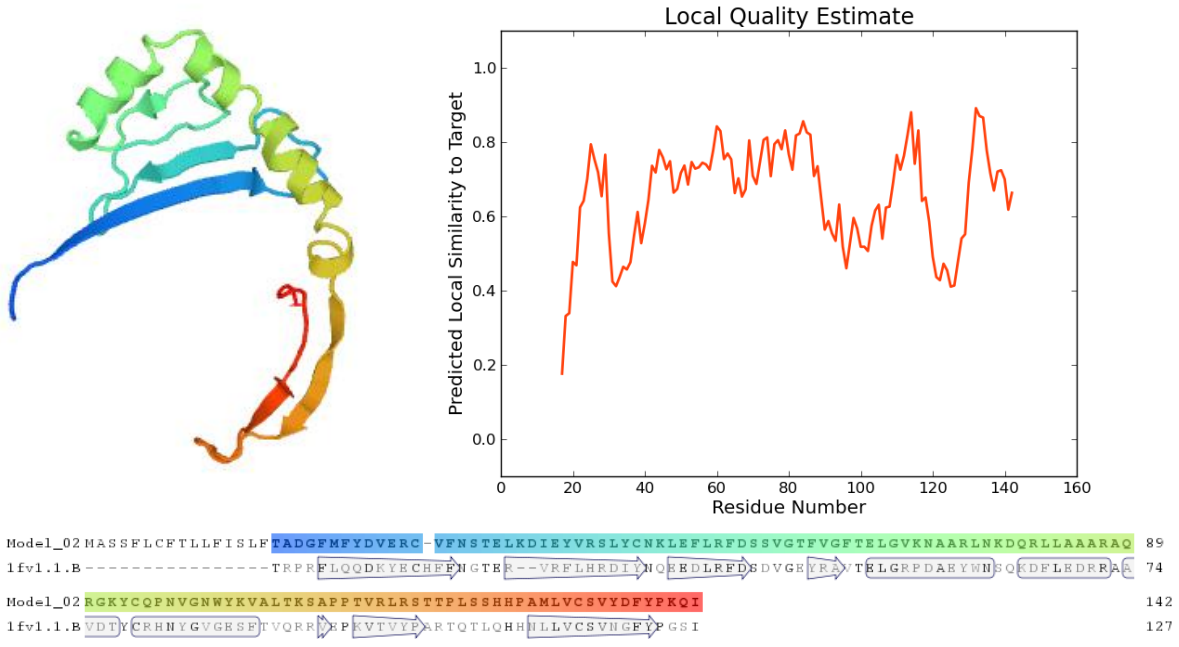
Amci-DXB*0602



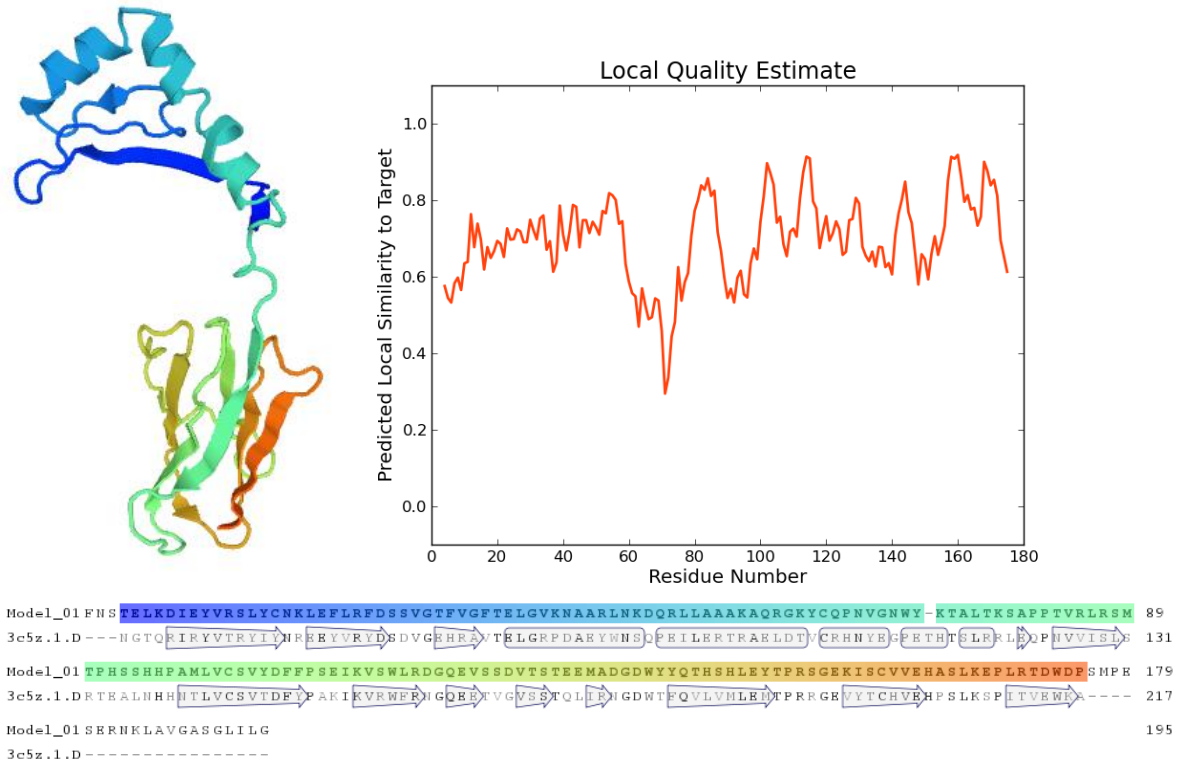
Amci-DXB*0603



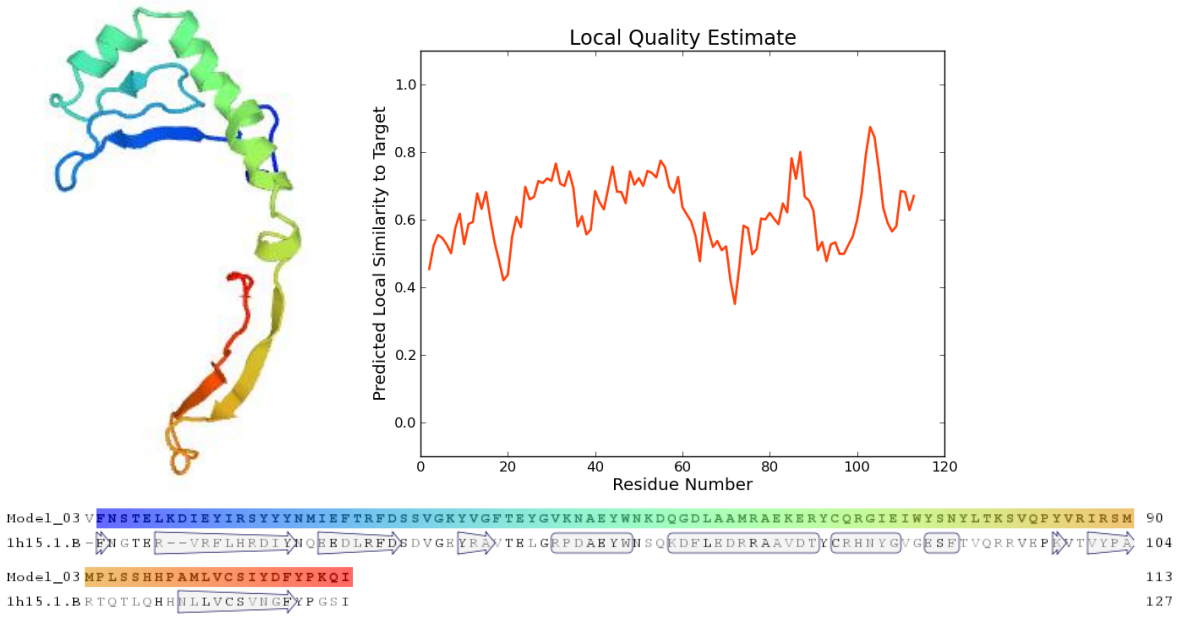
Amci-DXB*0604



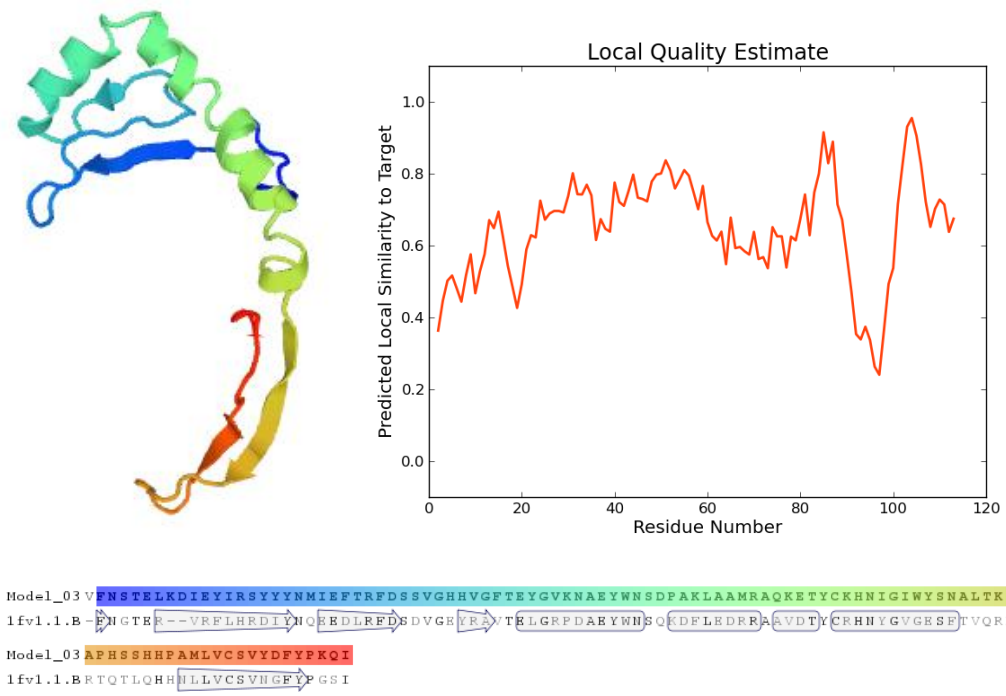
Amci-DXB*07



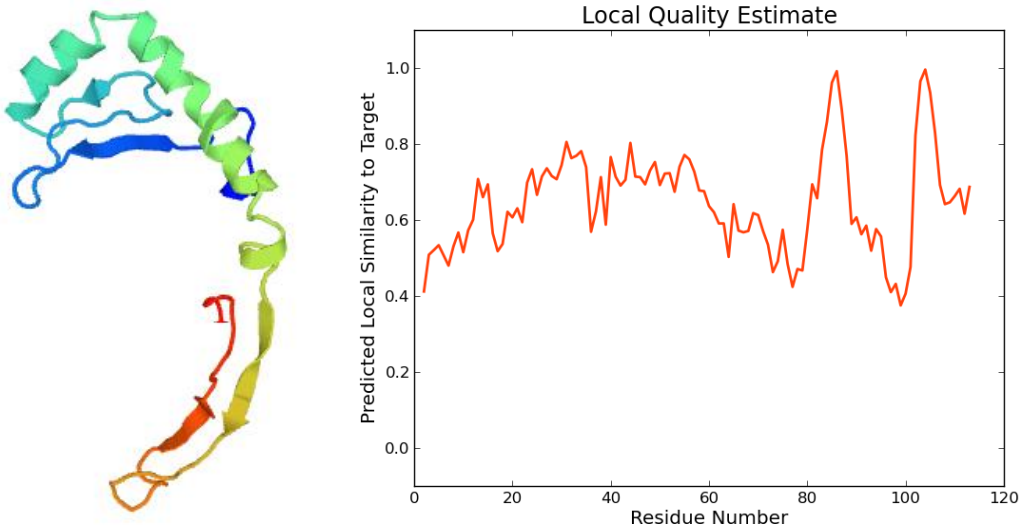
Amci-DXB*08

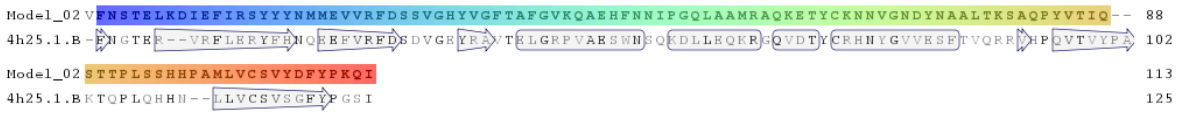


Amci-DXB*09

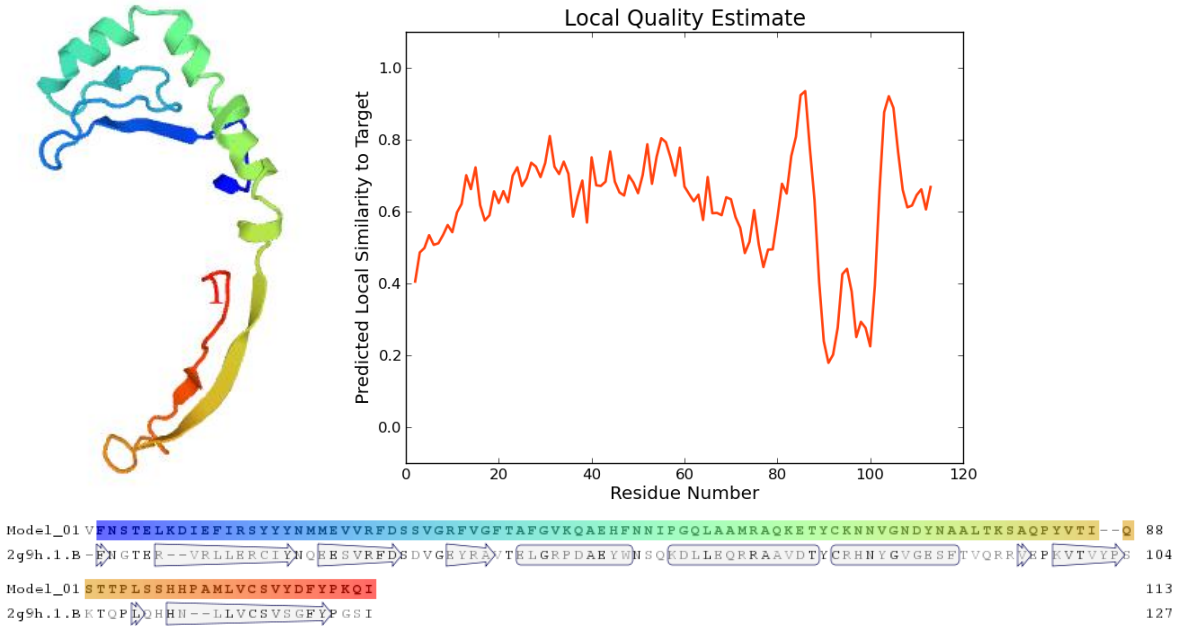


Amci-DXB*1001

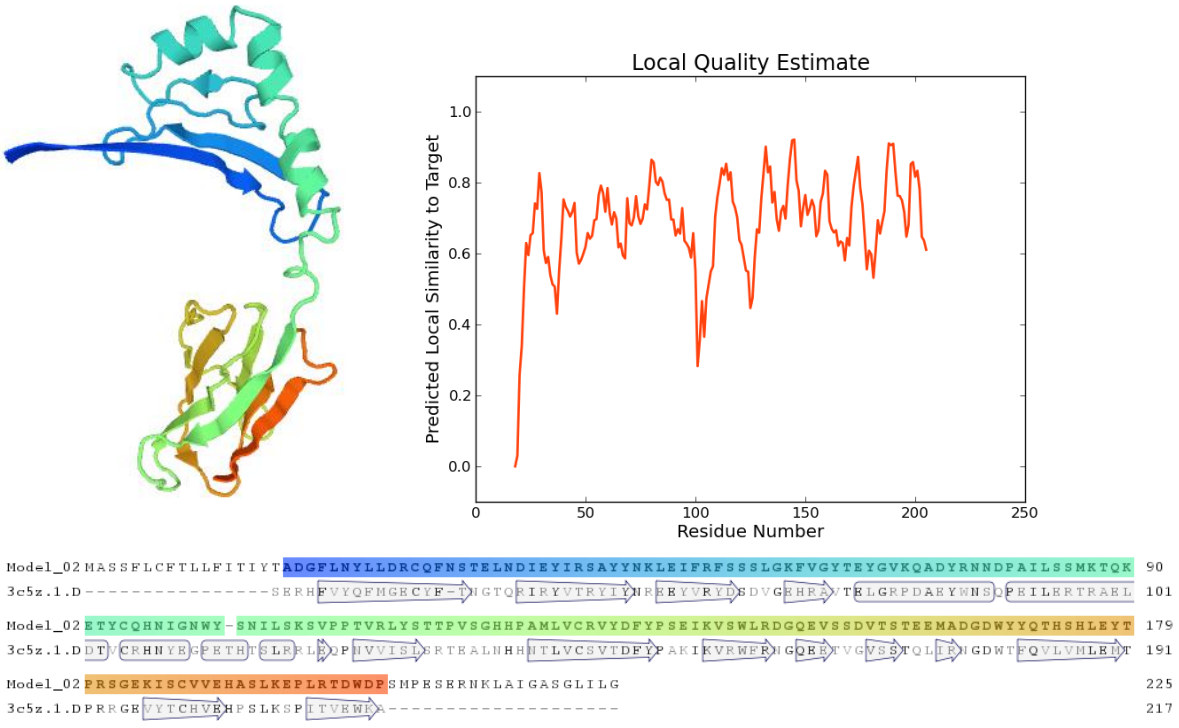




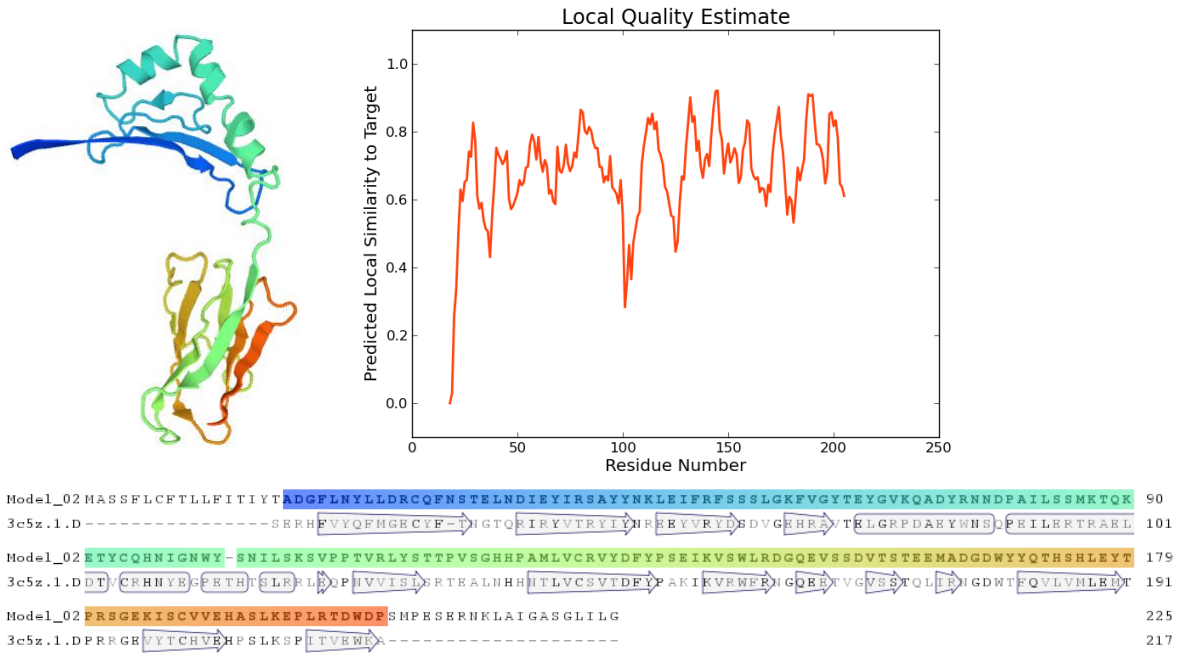
Amci-DXB*1002



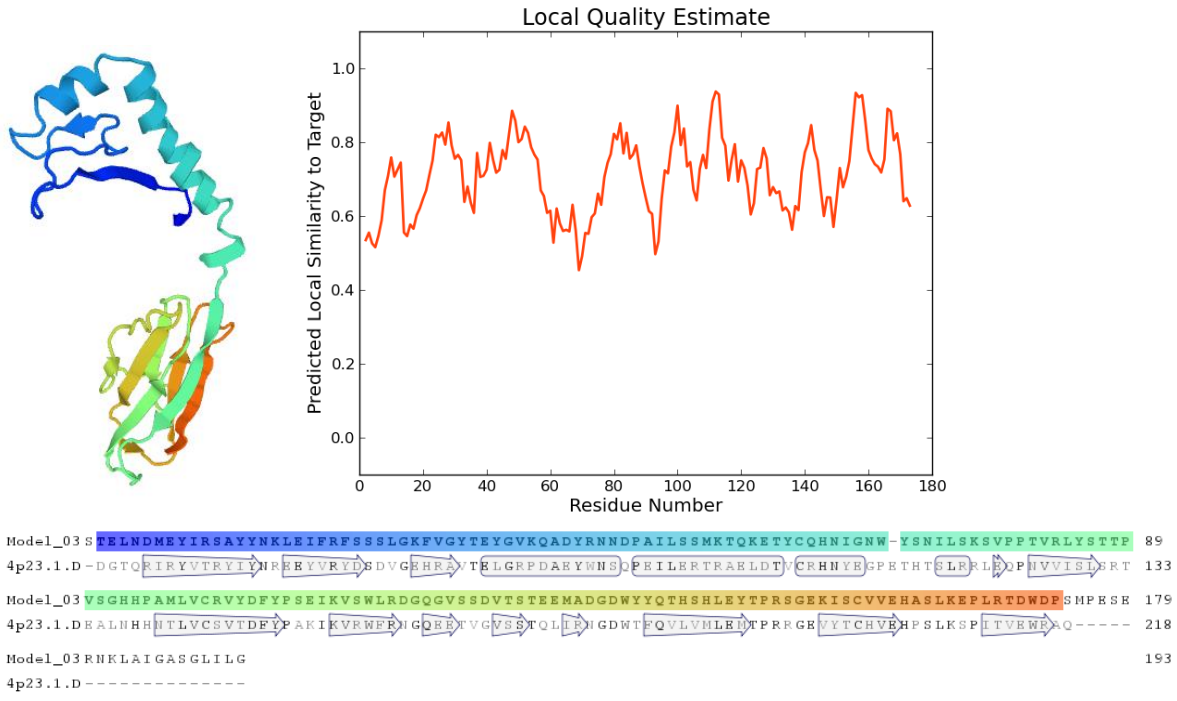
Amci-DXB*110101



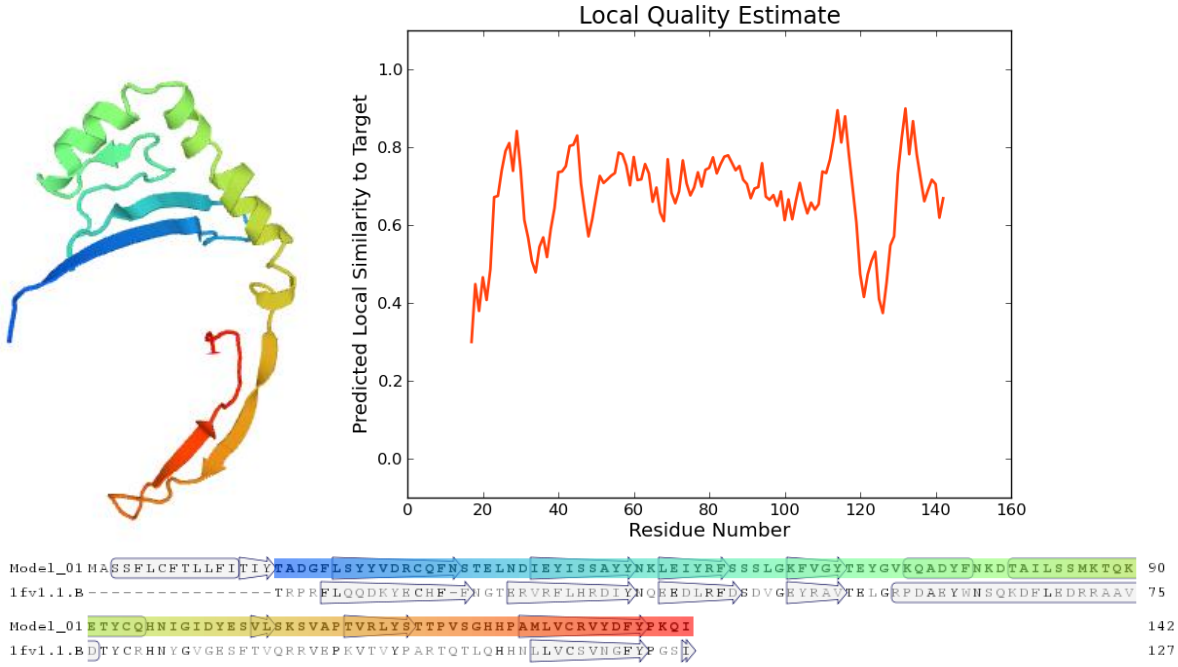
Amci-DXB*110102



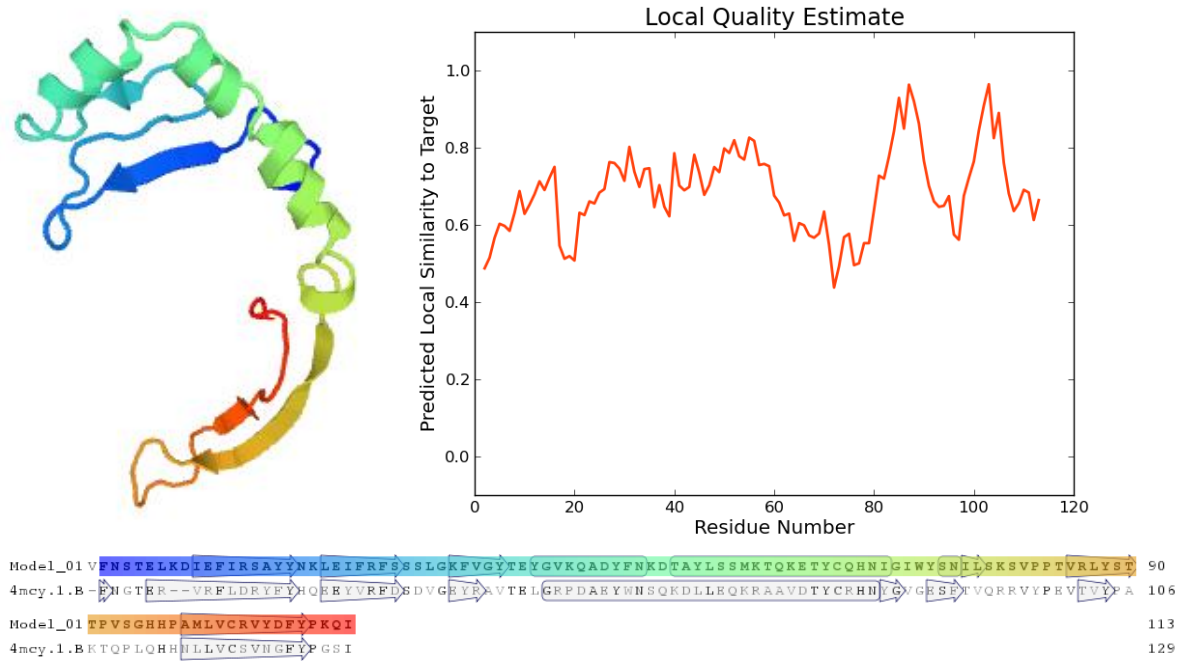
Amci-DXB*1102



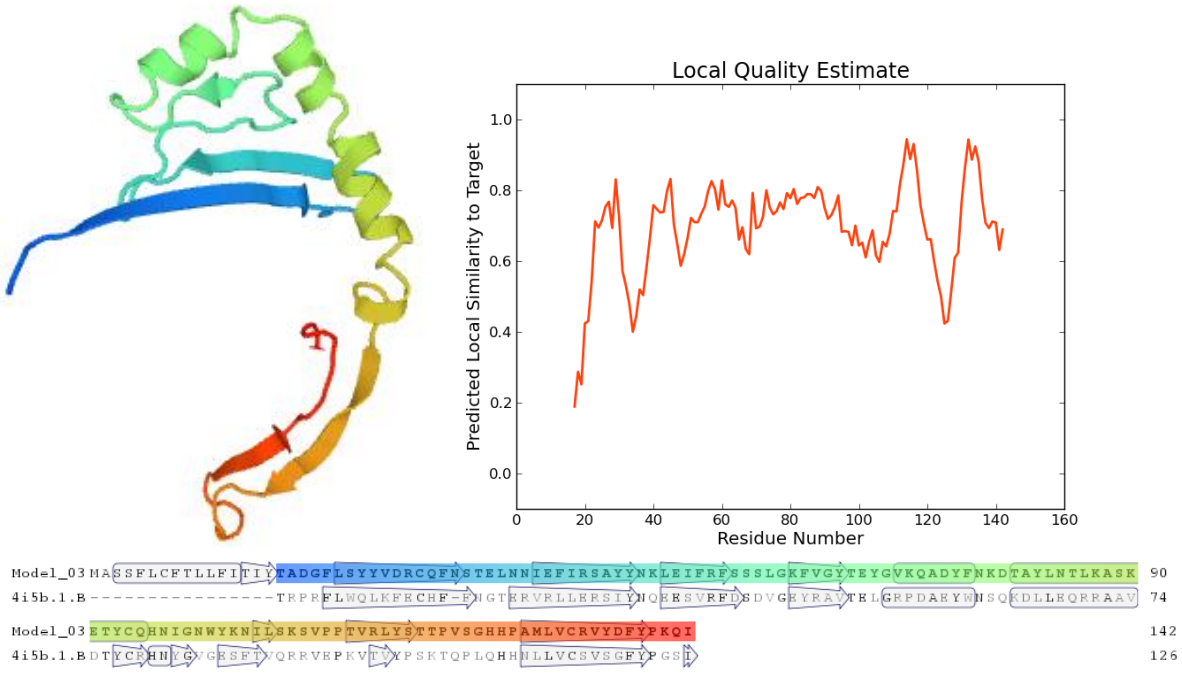
Amci-DXB*12



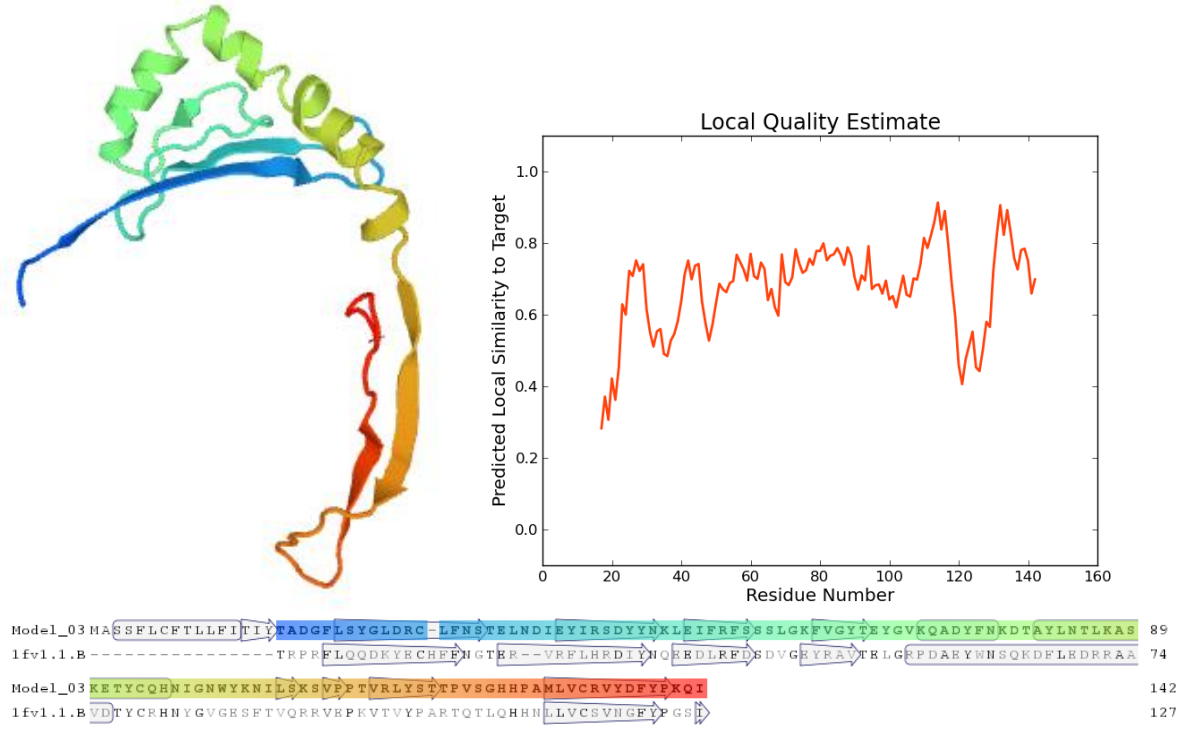
Amci-DXB*13



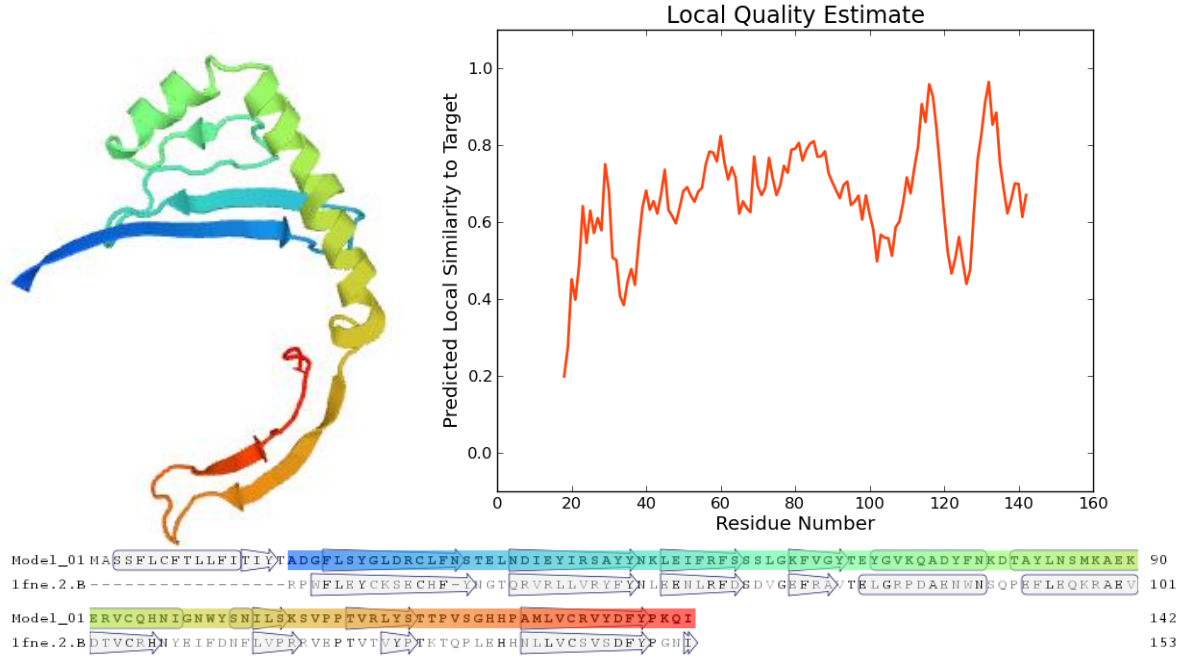
Amci-DXB*14



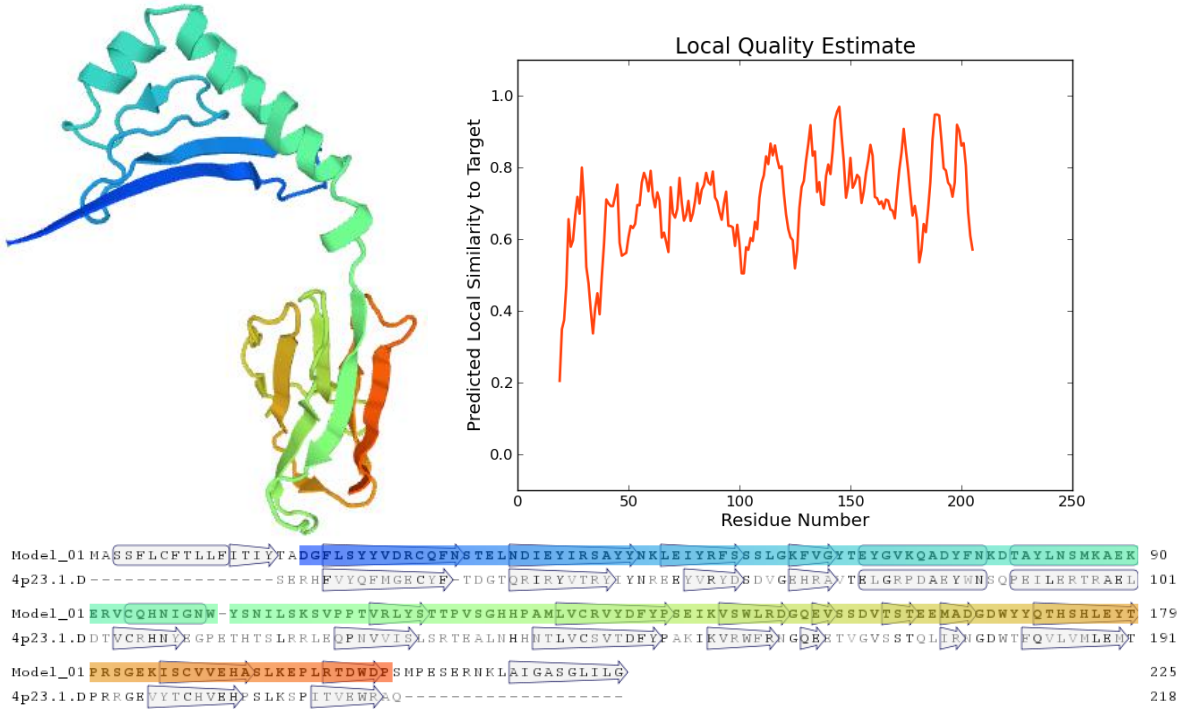
Amci-DXB*15



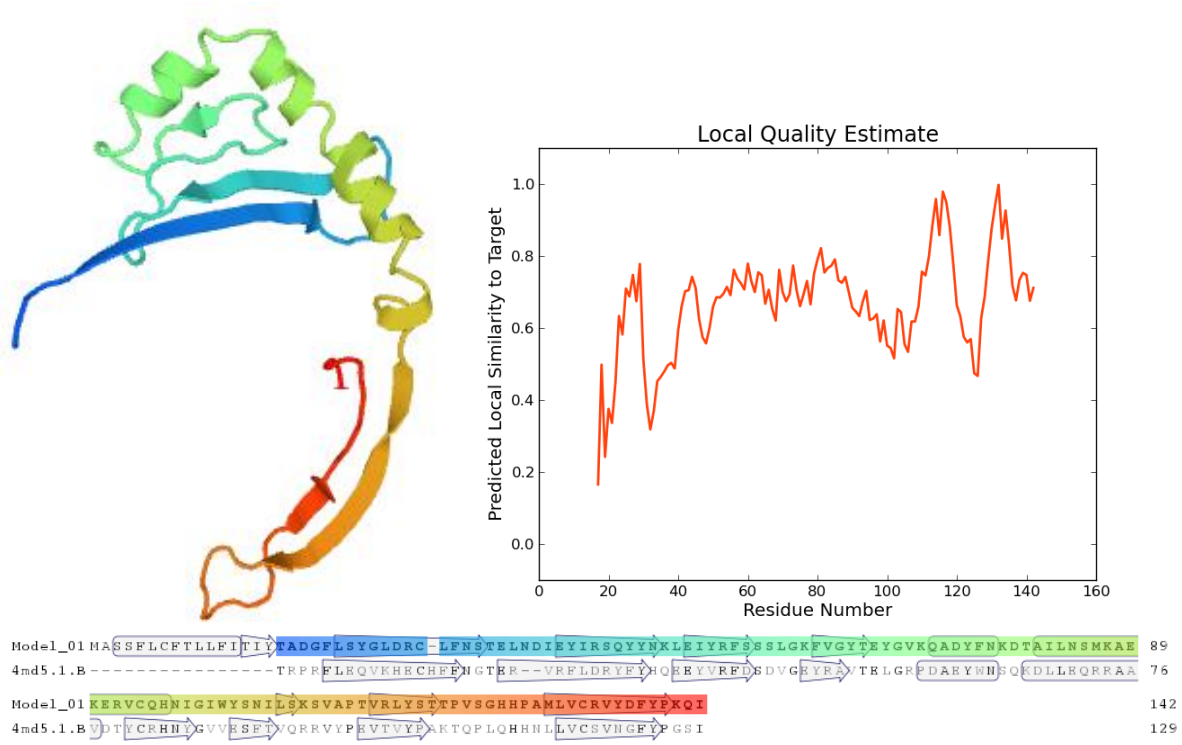
Amci-DXB*16



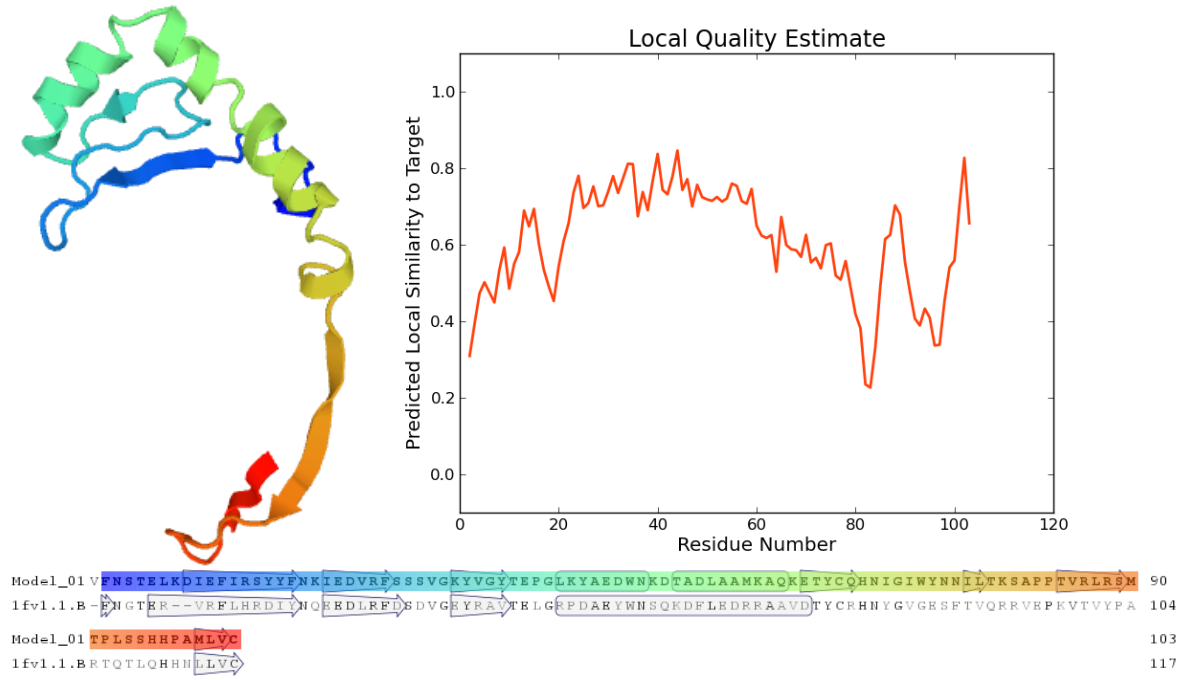
Amci-DXB*17



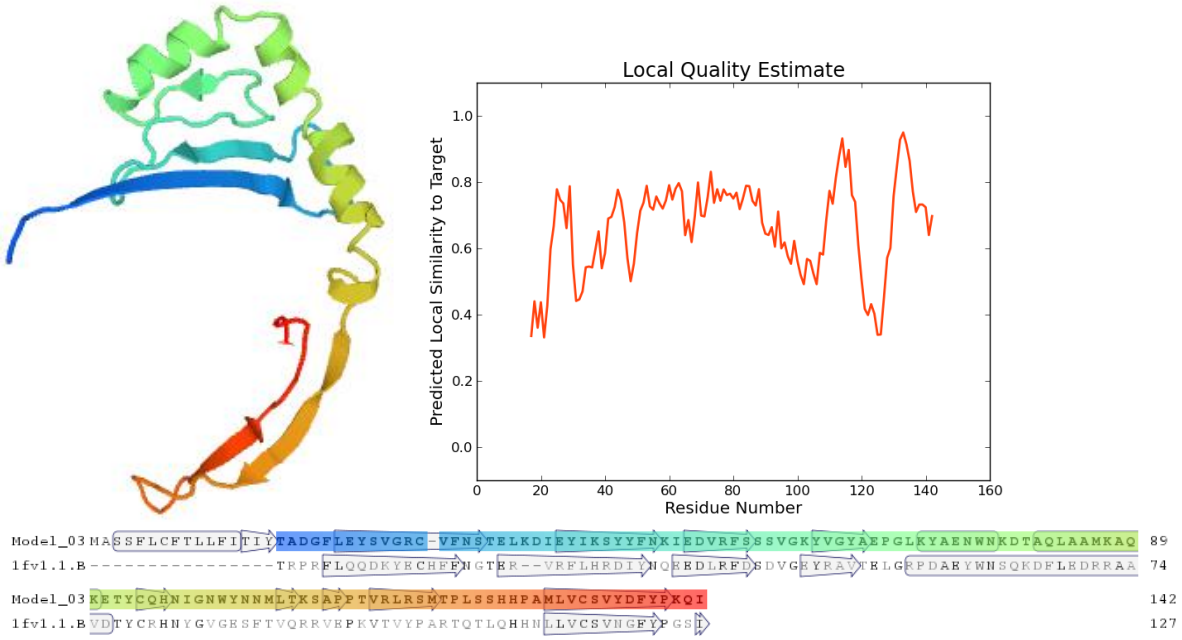
Amci-DXB*18



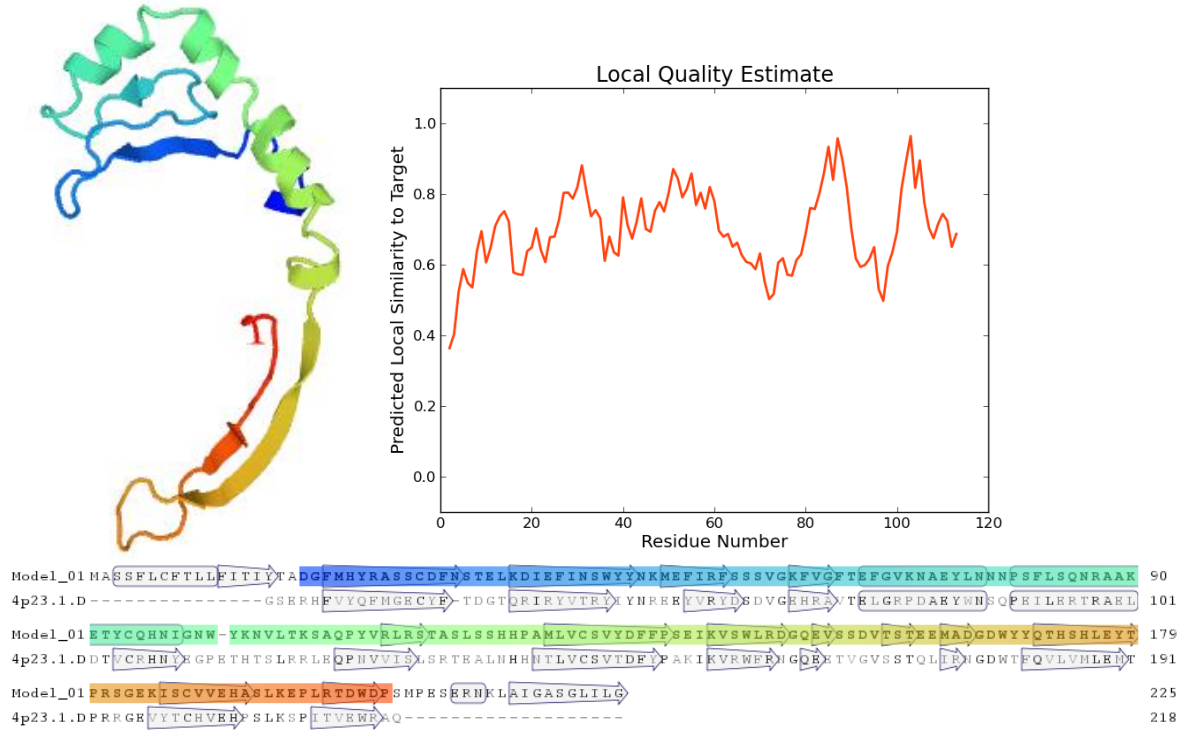
Amci-DXB*19



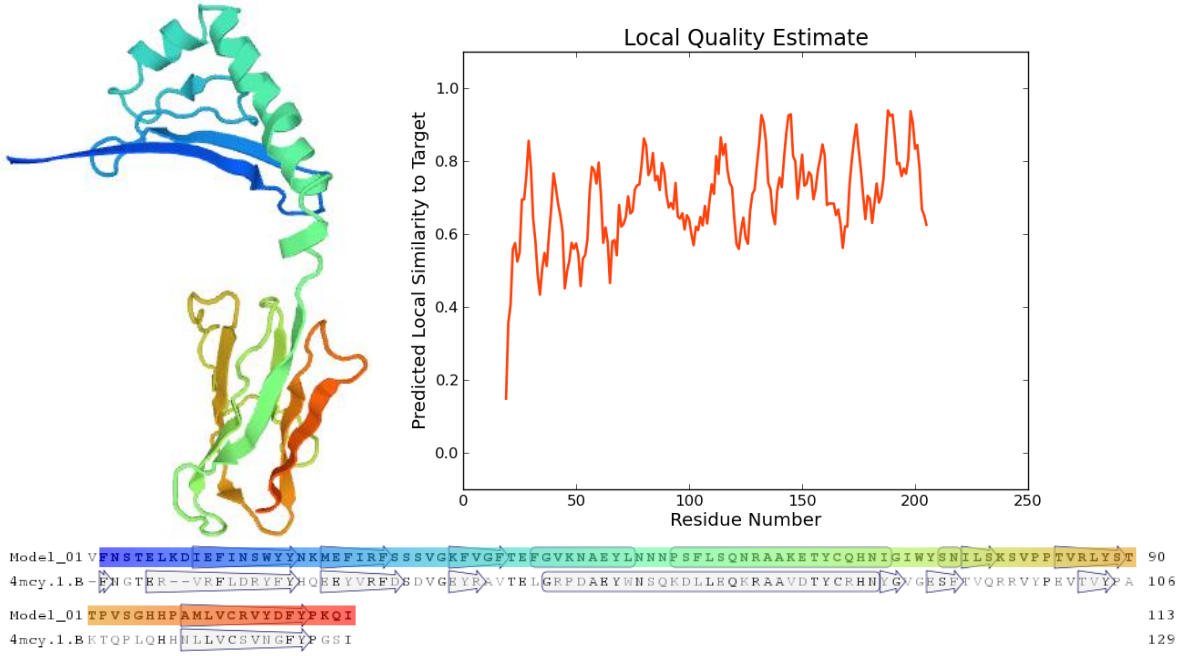
Amci-DXB*20



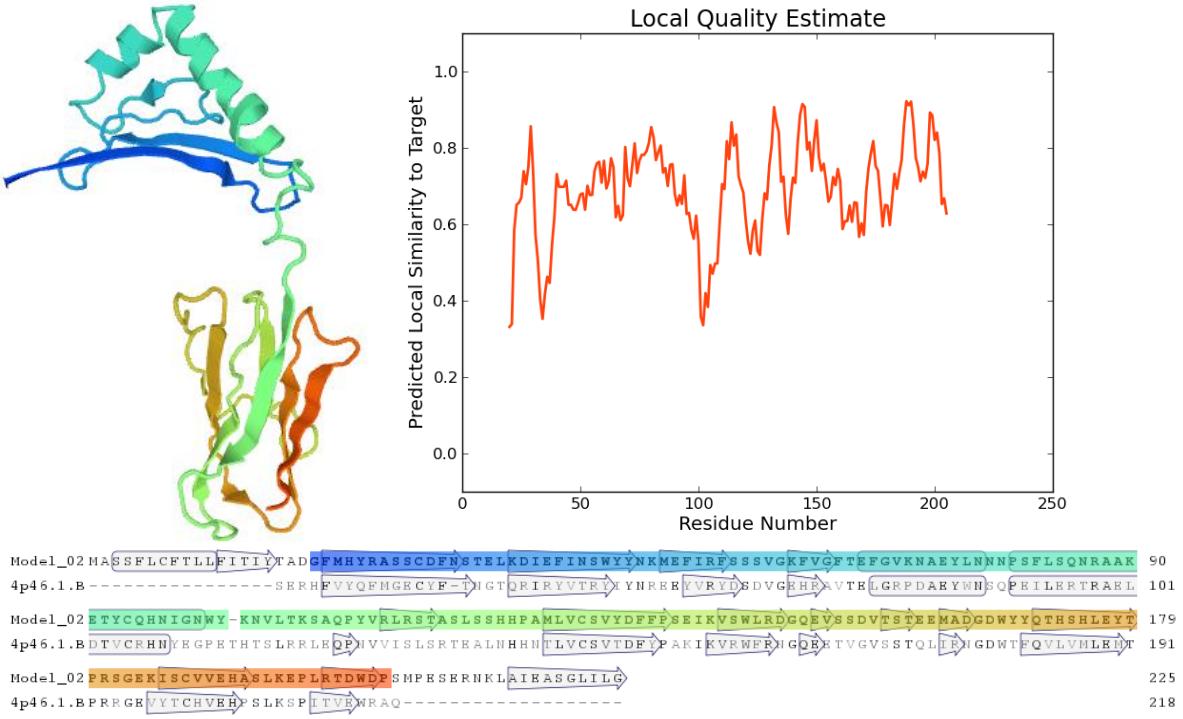
Amci-DXB*21



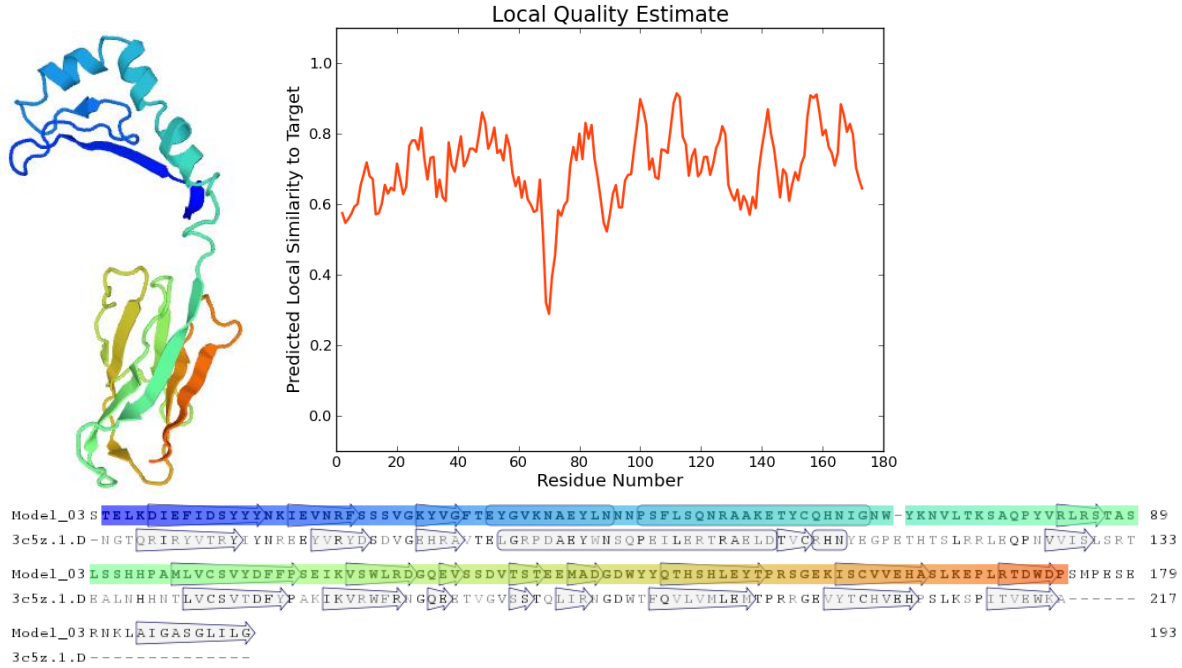
Amci-DXB*2201



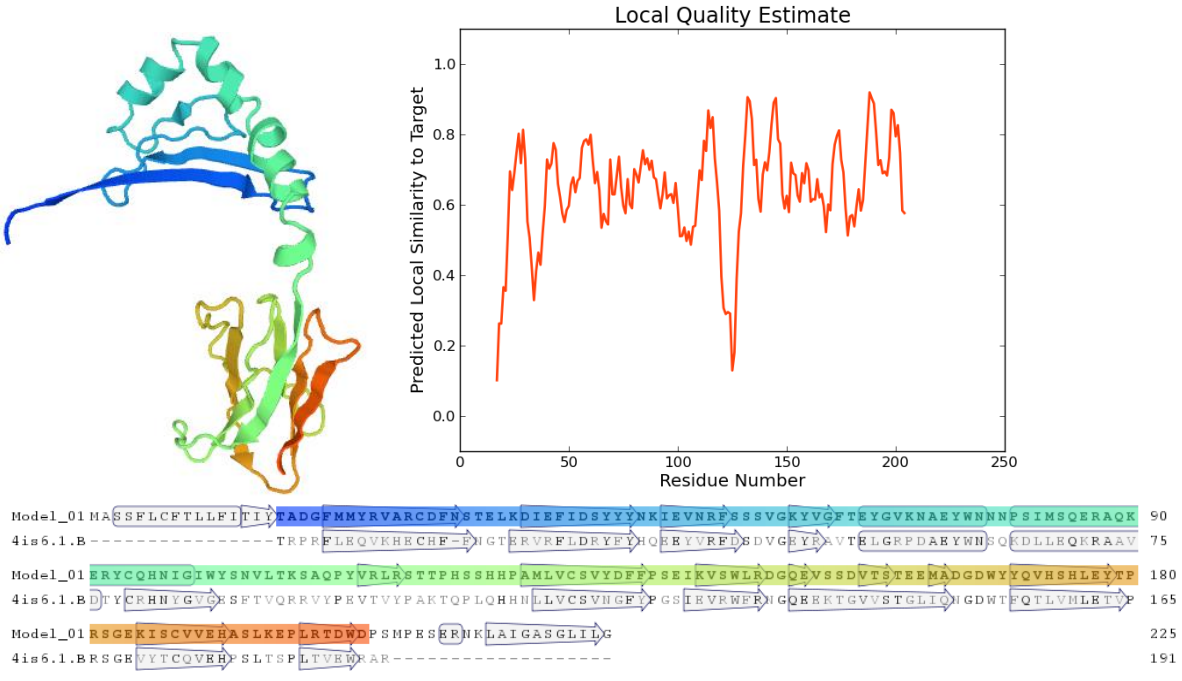
Amci-DXB*2202



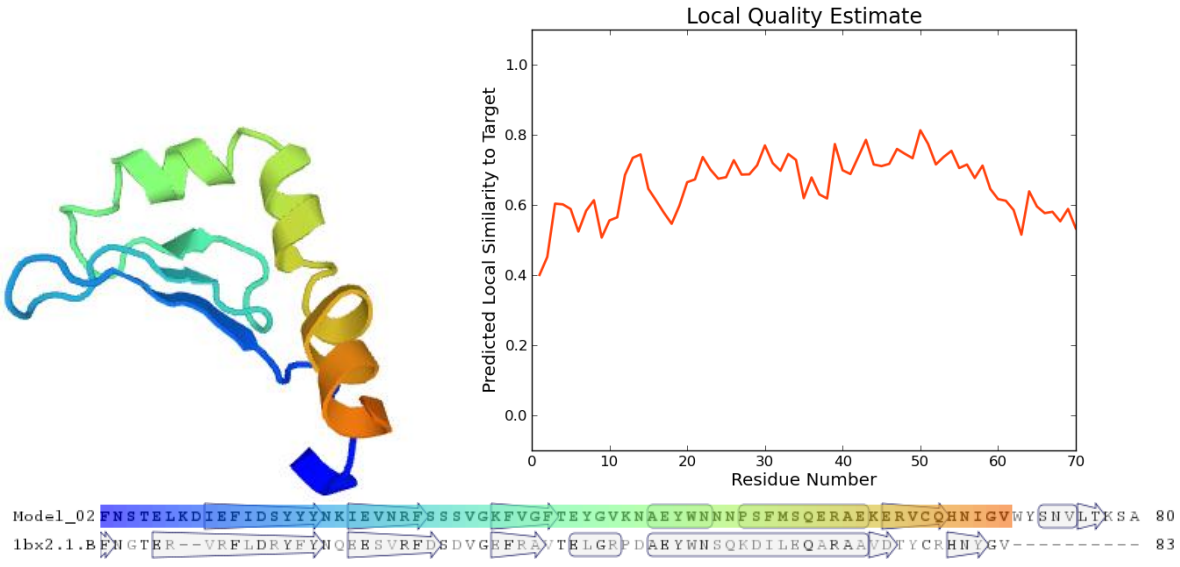
Amci-DXB*23



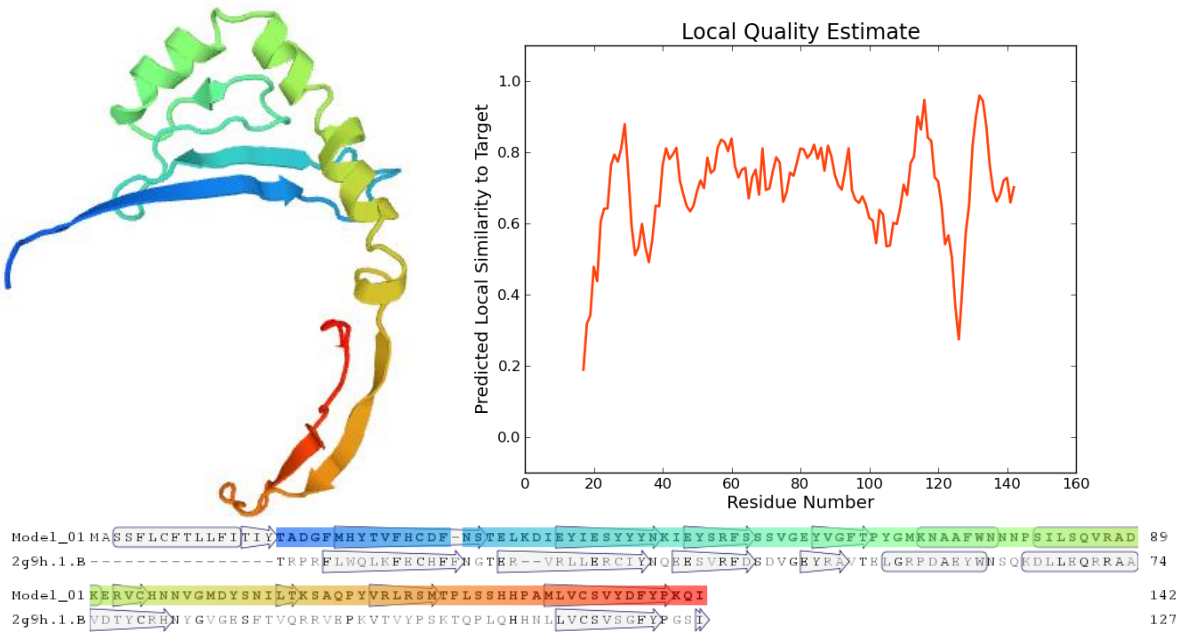
Amci-DXB*2401



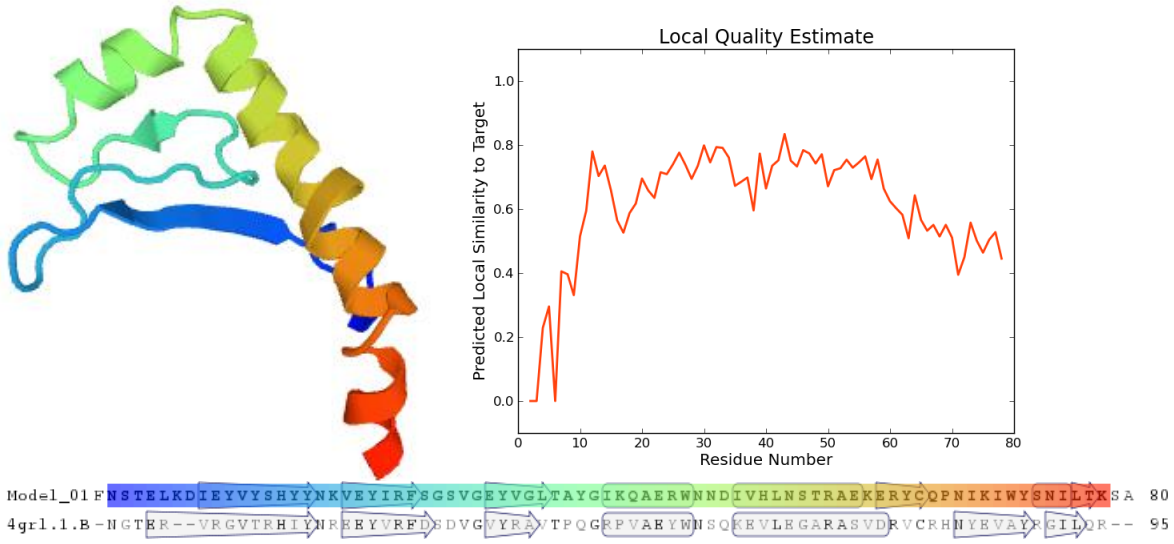
Amci-DXB*25



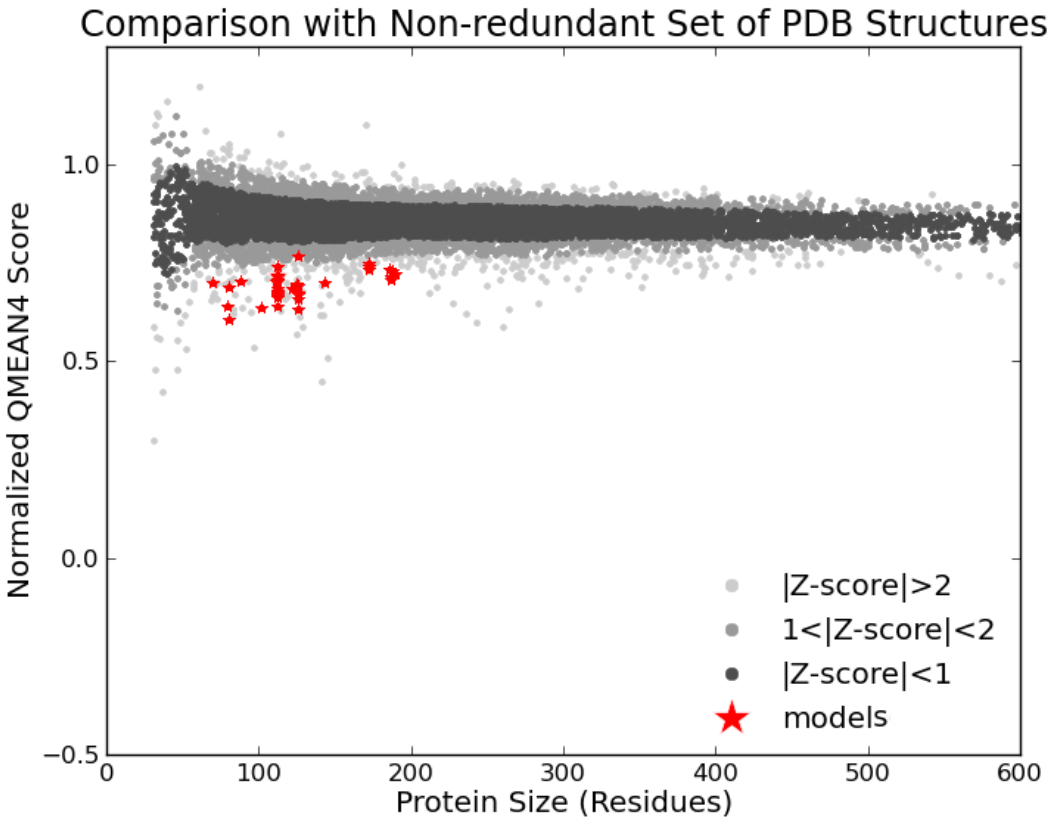
Amci-DXB*26



Amci-DXB*27


Local Quality Estimate
Predicted Local Similarity to Target
Residue Number
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
70
80
Model_01 FNSTELKDIEYVYSHYYNKVEYIRFSGSVGEYVGLTAYGIKQAERWNNDIVHLNSTRAEKERYCQPNIKIWYSNILTKSA 80
4grl.1.B -NGTER--VRGVTRHIYNREEYVRFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGARASVDRVCRHNYEVAYRGILQR-- 95

Sup. Figure I. 4. Summary of all estimated Z-scores for 3D models



Chapter II

Sup. Table II.1. List of samples with read count information for the Illumina Miseq runs after cleaning. QT stands for Quality trimmed

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Apoyo	*A. astorquii*	23A9	42,902	21,421	29,518	14,718
Apoyo	*A. astorquii*	23B1	71,403	35,666	60,786	30,337
Apoyo	*A. astorquii*	23C3	67,502	33,721	59,954	29,921
Apoyo	*A. astorquii*	23D2	16,343	8,162	8,466	4,204
Apoyo	*A. astorquii*	26E5	35,751	17,852	24,178	12,033
Apoyo	*A. astorquii*	AG12	38,735	19,346	29,092	14,510
Apoyo	*A. astorquii*	AG16	36,053	18,011	27,828	13,881
Apoyo	*A. astorquii*	AG18	49,490	24,722	41,458	20,689
Apoyo	*A. astorquii*	AG30	52,364	26,157	41,230	20,567
Apoyo	*A. astorquii*	AG34	56,883	28,409	46,528	23,216
Apoyo	*A. astorquii*	AG35	31,370	15,669	23,672	11,798
Apoyo	*A. astorquii*	AG36	51,621	25,786	42,438	21,177
Apoyo	*A. astorquii*	AG37	49,588	24,753	36,306	18,097
Apoyo	*A. astorquii*	AG38	32,028	15,995	20,512	10,230
Apoyo	*A. astorquii*	AG39	29,642	14,799	19,594	9,766
Apoyo	*A. astorquii*	AG40	36,423	18,193	27,944	13,939
Apoyo	*A. astorquii*	AG41	39,834	19,894	29,636	14,773
Apoyo	*A. astorquii*	AG42	95,757	47,824	82,828	41,341
Apoyo	*A. astorquii*	AG43	48,518	24,233	39,216	19,560
Apoyo	*A. astorquii*	AK04	31,573	15,767	19,510	9,726
Apoyo	*A. chancho*	23B3	54,968	27,447	40,646	20,265
Apoyo	*A. chancho*	23D3	60,365	30,150	49,500	24,696
Apoyo	*A. chancho*	23D6	42,565	21,258	29,324	14,604
Apoyo	*A. chancho*	24A6	60,745	30,339	48,106	23,976
Apoyo	*A. chancho*	24A7	39,594	19,772	29,018	14,470
Apoyo	*A. chancho*	24C2	80,988	40,453	69,818	34,844
Apoyo	*A. chancho*	25G5	28,042	14,001	20,982	10,440
Apoyo	*A. chancho*	25H4	50,900	25,414	33,894	16,894
Apoyo	*A. chancho*	25I8	32,561	16,263	23,670	11,799
Apoyo	*A. chancho*	AG05	48,873	24,416	37,702	18,804
Apoyo	*A. chancho*	AG06	66,792	33,366	57,636	28,766
Apoyo	*A. chancho*	AG10	45,495	22,722	34,592	17,244
Apoyo	*A. chancho*	AG14	36,202	18,087	25,162	12,553
Apoyo	*A. chancho*	AG15	30,985	15,479	20,882	10,411
Apoyo	*A. chancho*	AG45	35,208	17,575	23,596	11,754
Apoyo	*A. chancho*	AG52	104,610	52,255	89,976	44,914
Apoyo	*A. chancho*	AK06	47,838	23,898	37,526	18,712

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Apoyo	*A. chancho*	AK08	40,481	20,216	30,288	15,114
Apoyo	*A. chancho*	AK12	31,040	15,504	21,786	10,850
Apoyo	*A. chancho*	AK29	35,941	17,951	27,236	13,577
Apoyo	*A. chancho*	AK30	58,590	29,257	44,712	22,312
Apoyo	*A. chancho*	AK30	36,857	18,413	26,000	12,960
Apoyo	*A. zaliosus*	23B2	39,002	19,485	30,524	15,235
Apoyo	*A. zaliosus*	23D8	25,359	12,663	13,942	6,935
Apoyo	*A. zaliosus*	24B1	26,463	13,221	16,720	8,325
Apoyo	*A. zaliosus*	24B4	73,941	36,938	63,312	31,588
Apoyo	*A. zaliosus*	24B5	43,359	21,655	31,696	15,789
Apoyo	*A. zaliosus*	24B6	38,764	19,365	28,842	14,385
Apoyo	*A. zaliosus*	25G7	26,394	13,184	15,438	7,692
Apoyo	*A. zaliosus*	25G8	31,864	15,917	21,046	10,488
Apoyo	*A. zaliosus*	25G9	47,028	23,484	29,252	14,585
Apoyo	*A. zaliosus*	25H1	58,722	29,333	44,852	22,367
Apoyo	*A. zaliosus*	25H2	94,697	47,269	81,322	40,103
Apoyo	*A. zaliosus*	26F7	28,853	14,415	18,490	9,223
Apoyo	*A. zaliosus*	230	28,730	14,345	14,982	7,452
Apoyo	*A. zaliosus*	26E7	38,808	19,384	26,962	13,444
Apoyo	*A. zaliosus*	26E8	20,410	10,169	6,896	3,371
Apoyo	*A. zaliosus*	AG04	68,920	34,428	54,994	27,433
Apoyo	*A. zaliosus*	AG09	73,497	36,721	59,432	29,646
Apoyo	*A. zaliosus*	AG11	45,782	22,871	36,802	18,371
Apoyo	*A. zaliosus*	AG19	68,557	34,242	60,104	29,980
Apoyo	*A. zaliosus*	AG21	39,336	19,639	23,264	11,596
Apoyo	*A. zaliosus*	AJ80	56,147	28,042	39,752	19,822
A. León	*A.* sp2. *aff citrinellus*	AJ28	59,678	29,812	46,792	23,343
A. León	*A.* sp2. *aff citrinellus*	AJ29	60,042	29,997	45,742	22,821
A. León	*A.* sp2. *aff citrinellus*	AJ30	37,398	18,672	18,686	9,302
A. León	*A.* sp2. *aff citrinellus*	AJ32	93,551	46,726	80,882	40,372
A. León	*A.* sp2. *aff citrinellus*	AJ33	45,245	22,599	34,554	17,229
A. León	*A.* sp2. *aff citrinellus*	AJ35	47,918	23,939	37,316	18,618
A. León	*A.* sp2. *aff citrinellus*	AJ36	58,246	29,088	43,140	21,513
A. León	*A.* sp2. *aff citrinellus*	AJ38	58,203	29,073	43,852	21,848
A. León	*A.* sp2. *aff citrinellus*	AJ39	63,762	31,846	50,362	25,114
A. León	*A.* sp2. *aff citrinellus*	AJ41	36,084	18,017	22,726	11,318
A. León	*A.* sp2. *aff citrinellus*	AJ44	42,769	21,357	23,792	11,846
A. León	*A.* sp2. *aff citrinellus*	AJ45	37,356	18,656	25,484	12,700
A. León	*A.* sp2. *aff citrinellus*	AJ46	34,598	17,290	24,186	12,057
A. León	*A.* sp2. *aff citrinellus*	AJ48	39,604	19,763	21,726	10,652
A. León	*A.* sp2. *aff citrinellus*	AJ49	29,405	14,688	17,908	8,920
A. León	*A.* sp2. *aff citrinellus*	AJ51	34,736	17,354	18,466	9,196

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
A. León	*A.* sp2. *aff citrinellus*	AJ52	30,291	15,126	13,520	6,713
A. León	*A.* sp2. *aff citrinellus*	AJ54	23,854	11,916	10,856	5,392
A. León	*A.* sp2. *aff citrinellus*	AJ56	35,356	17,663	21,676	10,780
A. León	*A.* sp2. *aff citrinellus*	AJ57	37,214	18,580	18,184	9,036
A. León	*A.* sp2. *aff citrinellus*	AJ58	44,304	22,129	31,506	15,695
A. León	*A.* sp2. *aff citrinellus*	AJ59	47,303	23,633	35,764	17,835
A. León	*A.* sp2. *aff citrinellus*	AJ61	45,117	22,534	35,354	17,633
A. León	*A.* sp2. *aff citrinellus*	AM02	126,820	63,360	111,130	55,469
A. León	*A.* sp2. *aff citrinellus*	AM05	41,975	20,957	30,344	14,982
A. León	*A.* sp2. *aff citrinellus*	AM06	42,999	21,468	26,484	13,003
A. León	*A.* sp2. *aff citrinellus*	AM07	66,586	33,255	53,540	26,701
A. León	*A.* sp2. *aff citrinellus*	AM11	30,032	15,002	17,528	8,731
A. León	*A.* sp2. *aff citrinellus*	AM12	47,050	23,497	36,176	18,025
A. León	*A.* sp2. *aff citrinellus*	AM13	54,406	27,170	40,256	20,073
A. León	*A.* sp2. *aff citrinellus*	AM16	91,143	45,536	77,006	38,438
A. León	*A.* sp2. *aff citrinellus*	AM25	33,983	16,977	22,430	11,182
A. León	*A.* sp2. *aff citrinellus*	AM26	40,199	20,079	26,588	13,258
A. León	*A.* sp2. *aff citrinellus*	AM27	67,256	33,586	48,474	24,175
A. León	*A.* sp2. *aff citrinellus*	AM28	142,638	71,180	124,562	61,269
A. León	*A.* sp2. *aff citrinellus*	AM29	59,391	29,655	43,376	21,621
A. León	*A.* sp2. *aff citrinellus*	AM30	42,823	21,374	28,358	13,929
A. León	*A.* sp2. *aff citrinellus*	AM32	34,217	17,065	21,146	10,373
A. León	*A.* sp2. *aff citrinellus*	AM34	32,737	16,350	19,346	9,632
A. León	*A.* sp2. *aff citrinellus*	AM35	87,178	43,527	71,282	35,117
A. León	*A.* sp2. *aff citrinellus*	AM36	30,715	15,339	17,398	8,666
A. León	*A.* sp2. *aff citrinellus*	AM37	44,256	22,100	24,806	12,352
A. León	*A.* sp2. *aff citrinellus*	AM38	23,656	11,815	11,812	5,872
Managua	*A. citrinellus*	26H7	32,070	16,023	23,434	11,681
Managua	*A. citrinellus*	26I2	56,065	27,994	44,766	22,069
Managua	*A. citrinellus*	26I3	36,602	18,277	27,490	13,701
Managua	*A. citrinellus*	26I4	46,766	23,336	33,754	16,830
Managua	*A. citrinellus*	26I5	53,489	26,688	42,100	20,999
Managua	*A. citrinellus*	26I6	39,407	19,665	30,166	15,047
Managua	*A. citrinellus*	26I7	70,679	35,278	57,514	28,688
Managua	*A. citrinellus*	26I8	45,110	22,509	31,922	15,907
Managua	*A. citrinellus*	26I9	57,107	28,489	44,902	22,403
Managua	*A. citrinellus*	27A1	70,325	35,095	60,068	29,955
Managua	*A. citrinellus*	27A2	74,061	36,973	58,220	29,039
Managua	*A. citrinellus*	27D2	21,829	10,898	11,994	5,969
Managua	*A. citrinellus*	27D3	20,963	10,465	8,994	4,401
Managua	*A. citrinellus*	27D4	83,599	41,749	72,928	36,405
Managua	*A. citrinellus*	27D5	28,834	14,395	20,262	10,094

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Managua	*A. citrinellus*	27D6	28,196	14,072	16,790	8,360
Managua	*A. citrinellus*	28A9	28,539	14,245	14,534	7,249
Managua	*A. citrinellus*	28B1	33,585	16,765	22,138	11,042
Managua	*A. citrinellus*	28B5	23,150	11,555	11,426	5,681
Managua	*A. citrinellus*	28B7	55,296	27,580	39,588	19,738
Managua	*A. labiatus*	27D1	28,934	14,436	13,142	6,442
Managua	*A. labiatus*	28A2	43,339	21,639	32,500	16,029
Managua	*A. labiatus*	28A3	41,033	20,458	26,894	13,401
Managua	*A. labiatus*	28A4	35,114	17,533	23,054	11,340
Nicaragua	*A. citrinellus*	23H9	26,121	13,030	14,690	7,228
Nicaragua	*A. citrinellus*	23I1	31,816	15,884	20,114	9,905
Nicaragua	*A. citrinellus*	23I2	47,525	23,730	33,232	16,571
Nicaragua	*A. citrinellus*	23I3	64,971	32,441	52,354	26,129
Nicaragua	*A. citrinellus*	23I4	25,039	12,501	13,134	6,526
Nicaragua	*A. citrinellus*	23I5	32,795	16,359	15,208	7,567
Nicaragua	*A. citrinellus*	23I8	26,019	12,984	9,204	4,567
Nicaragua	*A. citrinellus*	23I9	22,897	11,421	9,946	4,940
Nicaragua	*A. citrinellus*	24A1	37,176	18,552	21,048	10,476
Nicaragua	*A. citrinellus*	25D1	35,669	17,782	13,454	6,663
Nicaragua	*A. citrinellus*	25D3	19,961	9,969	5,744	2,833
Nicaragua	*A. citrinellus*	25D4	19,234	9,585	6,484	3,169
Nicaragua	*A. citrinellus*	25D5	18,674	9,324	6,930	3,432
Nicaragua	*A. citrinellus*	25D7	17,923	8,944	4,990	2,458
Nicaragua	*A. citrinellus*	25D8	27,234	13,584	7,334	3,637
Nicaragua	*A. citrinellus*	25D9	29,605	14,778	12,864	6,392
Nicaragua	*A. citrinellus*	25F1	24,390	12,169	11,746	5,826
Nicaragua	*A. citrinellus*	26E3	20,721	10,341	7,790	3,863
Nicaragua	*A. citrinellus*	26E4	24,809	12,381	9,234	4,585
Nicaragua	*A. citrinellus*	26E5	25,696	12,821	6,834	3,367
Nicaragua	*A. citrinellus*	26E6	24,644	12,302	11,336	5,632
Nicaragua	*A. citrinellus*	26E7	27,608	13,787	15,676	7,801
Nicaragua	*A. citrinellus*	26E8	29,976	14,963	21,430	10,670
Nicaragua	*A. citrinellus*	26E9	28,932	14,421	13,266	6,588
Nicaragua	*A. labiatus*	23F9	30,064	14,999	16,614	8,267
Nicaragua	*A. labiatus*	23G1	44,742	22,322	32,322	16,092
Nicaragua	*A. labiatus*	23G4	80,778	40,309	65,664	32,758
Nicaragua	*A. labiatus*	23G6	39,174	19,568	27,406	13,657
Nicaragua	*A. labiatus*	23G9	22,177	11,071	11,884	5,906
Nicaragua	*A. labiatus*	23H2	27,112	13,526	14,412	7,162
Nicaragua	*A. labiatus*	23H3	33,664	16,793	16,330	8,118
Nicaragua	*A. labiatus*	23H4	47,865	23,879	32,610	16,241
Nicaragua	*A. labiatus*	23H7	46,701	23,297	32,972	16,430

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Nicaragua	*A. labiatus*	25A2	24,298	12,128	12,372	6,080
Nicaragua	*A. labiatus*	25B1	29,667	14,801	15,974	7,940
Nicaragua	*A. labiatus*	25B2	23,783	11,870	7,978	3,906
Nicaragua	*A. labiatus*	25B3	30,915	15,416	10,220	5,066
Nicaragua	*A. labiatus*	25B4	21,085	10,528	10,032	4,924
Nicaragua	*A. labiatus*	25B6	20,507	10,239	11,160	5,487
Nicaragua	*A. labiatus*	25B7	18,182	9,073	10,128	4,972
Nicaragua	*A. labiatus*	25B8	21,156	10,557	10,214	5,002
Nicaragua	*A. labiatus*	25B9	27,636	13,791	14,822	7,263
Nicaragua	*A. labiatus*	25C3	17,305	8,636	7,364	3,604
Nicaragua	*A. labiatus*	25C4	25,657	12,800	13,158	6,470
Nicaragua	*A. labiatus*	25C5	17,890	8,929	9,040	4,437
Nicaragua	*A. labiatus*	25C6	34,947	17,448	26,048	12,819
Nicaragua	*A. labiatus*	25C7	26,257	13,104	15,784	7,756
Nicaragua	*A. labiatus*	25C9	43,976	21,950	33,786	16,594
Nicaragua	*A. labiatus*	25F2	19,692	9,827	9,976	4,891
Masaya	*A.* sp1. *aff citrinellus*	AH50	59,580	29,731	48,400	24,140
Masaya	*A.* sp1. *aff citrinellus*	AH52	38,510	19,212	28,624	14,278
Masaya	*A.* sp1. *aff citrinellus*	AH53	64,358	32,119	53,840	26,855
Masaya	*A.* sp1. *aff citrinellus*	AH54	33,768	16,842	19,846	9,862
Masaya	*A.* sp1. *aff citrinellus*	AH55	32,831	16,394	22,860	11,401
Masaya	*A.* sp1. *aff citrinellus*	AH56	71,165	35,536	60,382	30,127
Masaya	*A.* sp1. *aff citrinellus*	AH57	26,313	13,134	17,662	8,802
Masaya	*A.* sp1. *aff citrinellus*	AH58	56,165	28,006	42,072	20,974
Masaya	*A.* sp1. *aff citrinellus*	AH59	22,625	11,276	10,662	5,298
Masaya	*A.* sp1. *aff citrinellus*	AH60	29,954	14,943	20,114	10,017
Masaya	*A.* sp1. *aff citrinellus*	AH61	30,839	15,390	21,334	10,630
Masaya	*A.* sp1. *aff citrinellus*	AH71	44,642	22,279	31,344	15,631
Masaya	*A.* sp1. *aff citrinellus*	AH72	69,546	34,731	57,410	28,650
Masaya	*A.* sp1. *aff citrinellus*	AJ01	32,576	16,265	21,188	10,555
Masaya	*A.* sp1. *aff citrinellus*	AQ38	27,663	13,811	16,182	7,951
Masaya	*A.* sp1. *aff citrinellus*	AQ66	30,800	15,380	18,260	9,094
Masaya	*A.* sp1. *aff citrinellus*	AQ68	36,372	18,145	19,072	9,489
Masaya	*A.* sp1. *aff citrinellus*	AQ70	39,903	19,916	25,162	12,529
Masaya	*A.* sp1. *aff citrinellus*	AQ78	27,494	13,715	13,208	6,558
Masaya	*A.* sp1. *aff citrinellus*	AQ79	42,289	21,107	30,280	15,088
Masaya	*A.* sp1. *aff citrinellus*	AQ81	29,346	14,640	14,078	7,002
Masaya	*A.* sp1. *aff citrinellus*	AH51	28,258	14,103	16,266	8,101
Masaya	*A.* sp1. *aff citrinellus*	AH62	59,260	29,554	41,066	20,466
Masaya	*A.* sp1. *aff citrinellus*	AH65	67,749	33,823	51,658	25,756
Masaya	*A.* sp1. *aff citrinellus*	AH68	79,699	39,761	67,266	33,576
Masaya	*A.* sp1. *aff citrinellus*	AQ67	102,354	51,093	86,564	43,191

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Masaya	*A.* sp1. *aff citrinellus*	AQ76	41,907	20,911	27,644	13,766
Masaya	*A.* sp1. *aff labiatus*	AH69	48,068	24,006	36,382	18,141
Masaya	*A.* sp1. *aff labiatus*	AJ02	27,812	13,885	15,944	7,939
Masaya	*A.* sp1. *aff labiatus*	AJ03	20,086	10,024	9,320	4,639
Masaya	*A.* sp1. *aff labiatus*	AJ05	27,183	13,546	12,564	6,250
Masaya	*A.* sp1. *aff labiatus*	AJ06	40,905	20,407	25,102	12,502
Masaya	*A.* sp1. *aff labiatus*	AJ11	33,458	16,682	20,346	10,138
Masaya	*A.* sp1. *aff labiatus*	AJ12	48,656	24,282	35,222	17,560
Masaya	*A.* sp1. *aff labiatus*	AQ31	53,300	26,600	40,206	20,042
Masaya	*A.* sp1. *aff labiatus*	AQ32	46,344	23,138	34,956	17,440
Masaya	*A.* sp1. *aff labiatus*	AQ33	34,471	17,209	23,454	11,702
Masaya	*A.* sp1. *aff labiatus*	AQ60	33,682	16,816	23,348	11,632
Masaya	*A.* sp1. *aff labiatus*	AQ62	53,605	26,738	34,458	17,184
Masaya	*A.* sp1. *aff labiatus*	AQ63	47,384	23,645	31,186	15,546
Masaya	*A.* sp1. *aff labiatus*	AQ65	35,149	17,539	21,200	10,553
Masaya	*A.* sp1. *aff labiatus*	AQ69	60,608	30,242	45,610	22,739
Masaya	*A.* sp1. *aff labiatus*	AQ73	41,792	20,855	26,670	13,279
Masaya	*A.* sp1. *aff labiatus*	AQ74	39,927	19,934	22,872	11,387
Masaya	*A.* sp1. *aff labiatus*	AQ80	51,068	25,506	39,914	19,903
Xiloá	*A. amarillo*	26A7	31,094	15,524	15,684	7,688
Xiloá	*A. amarillo*	27A7	29,376	14,657	19,164	9,408
Xiloá	*A. amarillo*	AL01	73,013	36,410	58,914	28,926
Xiloá	*A. amarillo*	AL02	73,081	36,469	59,690	29,334
Xiloá	*A. amarillo*	AL04	60,980	30,417	48,658	23,927
Xiloá	*A. amarillo*	AL09	50,278	25,101	40,554	19,963
Xiloá	*A. amarillo*	AL10	52,168	26,031	39,568	19,465
Xiloá	*A. amarillo*	AL11	56,127	28,014	42,778	21,075
Xiloá	*A. amarillo*	AL19	47,054	23,492	36,022	17,736
Xiloá	*A. amarillo*	AL21	27,078	13,518	17,236	8,481
Xiloá	*A. amarillo*	AL27	36,892	18,415	26,814	13,199
Xiloá	*A. amarillo*	AL28	39,674	19,780	27,218	13,375
Xiloá	*A. amarillo*	AL30	36,934	18,426	25,484	12,516
Xiloá	*A. amarillo*	AL34	44,293	22,094	33,010	16,222
Xiloá	*A. amarillo*	AL35	41,080	20,504	28,126	13,836
Xiloá	*A. amarillo*	AL69	28,768	14,357	16,912	8,302
Xiloá	*A. amarillo*	AN26	25,994	12,973	17,538	8,647
Xiloá	*A. amarillo*	AN61	37,608	18,780	27,854	13,699
Xiloá	*A. amarillo*	AP81	30,871	15,412	22,166	10,925
Xiloá	*A. amarillo*	AQ14	50,314	25,094	38,786	19,043
Xiloá	*A. amarillo*	AQ22	33,167	16,549	22,068	10,808
Xiloá	*A. amarillo*	AQ27	64,794	32,340	53,744	26,473
Xiloá	*A. sagittae*	AL17	37,231	18,583	27,208	13,403

Lake	Species	Sample ID	QT total reads	QT read pairs	Primer clipped total reads	Combined reads
Xiloá	*A. sagittae*	AL55	35,773	17,857	24,416	12,009
Xiloá	*A. sagittae*	AL57	16,183	8,076	6,148	3,008
Xiloá	*A. sagittae*	AL59	14,709	7,344	3,524	1,713
Xiloá	*A. sagittae*	AL70	35,913	17,927	24,398	12,000
Xiloá	*A. sagittae*	AL74	51,601	25,737	36,470	17,926
Xiloá	*A. sagittae*	AL75	26,026	12,991	12,808	6,266
Xiloá	*A. sagittae*	AL77	41,321	20,617	27,680	13,599
Xiloá	*A. sagittae*	AN01	27,525	13,743	15,162	7,438
Xiloá	*A. sagittae*	AN02	25,510	12,731	9,772	4,782
Xiloá	*A. sagittae*	AN04	27,934	13,939	17,170	8,438
Xiloá	*A. sagittae*	AN10	30,462	15,213	19,162	9,430
Xiloá	*A. sagittae*	AN11	34,262	17,103	24,436	12,038
Xiloá	*A. sagittae*	AN12	19,726	9,831	7,940	3,887
Xiloá	*A. sagittae*	AN13	23,380	11,668	10,576	5,189
Xiloá	*A. sagittae*	AN14	36,919	18,420	24,074	11,833
Xiloá	*A. sagittae*	AN16	43,227	21,578	30,236	14,913
Xiloá	*A. sagittae*	AN17	28,642	14,291	17,462	8,569
Xiloá	*A. sagittae*	AN20	61,381	30,638	50,672	24,970
Xiloá	*A. sagittae*	AN21	29,747	14,849	18,520	9,121
Xiloá	*A. sagittae*	AN24	19,015	9,487	8,150	3,992
Xiloá	*A. sagittae*	AN28	28,386	14,156	13,884	6,796
Xiloá	*A. sagittae*	AN31	23,616	11,787	10,588	5,183
Xiloá	*A. sagittae*	AN33	33,354	16,648	19,260	9,457
Xiloá	*A. sagittae*	AL54	26,469	13,208	14,054	6,900
Xiloá	*A. sagittae*	AL56	32,956	16,439	16,678	8,188
Xiloá	*A. sagittae*	AL63	24,587	12,272	8,066	3,938
Xiloá	*A. sagittae*	AL64	37,479	18,699	23,508	11,566
Xiloá	*A. sagittae*	AL73	42,590	21,266	27,272	13,446
Xiloá	*A. sagittae*	AL78	36,080	18,016	22,430	11,038
Xiloá	*A. sagittae*	AN32	40,611	20,262	26,996	13,240
Xiloá	*A. xiloaensis*	24D5	43,394	21,660	30,516	15,039
Xiloá	*A. xiloaensis*	26A6	35,151	17,545	16,168	7,944
Xiloá	*A. xiloaensis*	26A8	37,714	18,825	25,228	12,405
Xiloá	*A. xiloaensis*	26B2	34,713	17,331	19,982	9,819
Xiloá	*A. xiloaensis*	26B4	27,370	13,665	16,470	8,110
Xiloá	*A. xiloaensis*	26H1	29,586	14,750	14,774	7,243
Xiloá	*A. xiloaensis*	27G4	49,432	24,679	31,340	15,440
Xiloá	*A. xiloaensis*	27G6	39,906	19,912	23,174	11,349
Xiloá	*A. xiloaensis*	AQ15	46,481	23,197	34,694	17,087
Xiloá	*A. xiloaensis*	AQ20	49,221	24,573	36,178	17,786
Xiloá	*A. xiloaensis*	AQ24	33,361	16,650	21,956	10,798

Sup. Table II.2. Classification Function Coefficients for the most discriminating alleles. List of the 32 most discriminating alleles by a Fisher´s linear discrimination functions test. Valued in bold are indicating what to what lake the allele pertains.

Allele	Apoyo	Xiloá	Managua	Nicaragua	Masaya	A. León
a9	**9.417**	1.207	1.384	0.634	2.724	2.774
a11	3.558	1.517	0.401	0.590	2.579	**10.739**
a17	0.591	1.371	0.011	4.757	3.296	**6.111**
a38	0.635	1.435	**12.092**	1.229	1.782	1.863
a40	-0.713	2.060	**8.946**	**6.219**	2.034	2.523
a50	1.034	0.448	0.859	0.225	**14.173**	2.104
a55	-0.590	-1.113	0.599	0.312	**10.827**	0.539
a58	**10.263**	0.996	1.002	0.867	1.720	-0.365
a62	1.924	1.611	1.372	1.761	2.522	**19.036**
a111	-0.149	-0.972	-4.375	**10.882**	-2.391	-1.204
a112	-1.475	3.409	**19.293**	2.950	1.972	2.814
a115	4.238	**18.825**	1.684	2.848	1.789	1.696
a8	1.317	0.703	0.099	1.165	5.131	**33.071**
a15	1.068	**9.527**	0.512	0.773	2.958	2.169
a23	1.131	0.731	0.112	1.206	3.525	**22.397**
a34	3.154	-1.794	-1.156	0.402	**9.641**	2.522
a37	0.195	-0.113	3.030	1.168	**13.955**	3.184
a39	1.823	1.761	1.735	2.028	3.967	**24.631**
a42	1.466	**15.239**	1.383	1.478	1.341	1.841
a59	-0.164	**15.125**	0.197	0.525	1.343	1.354
a74	0.986	**15.641**	-5.354	0.685	0.601	0.537
a76	2.566	**9.914**	**9.149**	**-10.072**	1.423	1.372
a86	4.253	3.898	4.200	3.838	8.307	**47.112**
a12	**-9.181**	**12.987**	-1.793	-0.002	-1.891	-1.948
a68	-4.209	-0.658	0.260	**13.226**	-0.168	1.319
a31	-2.275	-1.664	-0.470	**11.121**	0.329	0.247
a64	0.147	**13.598**	0.464	0.959	0.107	-1.759
a5	2.041	2.812	4.657	1.843	2.050	4.404
a25	1.514	1.607	2.684	3.561	**9.757**	3.343
a3	**5.869**	4.932	1.983	3.096	0.594	0.456
a32	-0.817	-0.317	-0.647	**11.193**	-2.496	0.623
a48	0.791	0.458	0.538	0.058	**6.329**	**12.510**
(Constant)	-7.376	-9.712	-5.930	-7.126	-8.076	**-22.080**

Sup. Table II.2 List of alleles and the supertype they belong to.

Allele	Group	Allele	Group	Allele	Group	Allele	Group	Allele	Group	Allele	Group
a9	1	a89	1	a139	1	a69	2	a96	3	a72	5
a11	1	a94	1	a140	1	a71	2	a98	3	a80	5
a17	1	a100	1	a144	1	a74	2	a107	3	a123	5
a24	1	a101	1	a145	1	a76	2	a109	3	a141	5
a38	1	a102	1	a146	1	a77	2	a118	3	a142	5
a40	1	a105	1	a8	2	a81	2	a124	3	a143	5
a41	1	a106	1	a10	2	a84	2	a125	3	a5	6
a43	1	a108	1	a14	2	a86	2	a131	3	a6	6
a46	1	a110	1	a15	2	a87	2	a132	3	a16	6
a50	1	a111	1	a18	2	a88	2	a136	3	a25	6
a52	1	a112	1	a19	2	a92	2	a137	3	a133	6
a55	1	a113	1	a23	2	a93	2	a138	3	a3	7
a57	1	a115	1	a26	2	a95	2	a20	4	a4	7
a58	1	a116	1	a28	2	a97	2	a22	4	a7	7
a60	1	a117	1	a30	2	a99	2	a31	4	a13	8
a62	1	a119	1	a34	2	a103	2	a33	4	a32	8
a63	1	a120	1	a36	2	a104	2	a44	4	a67	8
a66	1	a121	1	a37	2	a12	3	a47	4	a1	9
a73	1	a122	1	a39	2	a21	3	a64	4	a27	10
a75	1	a126	1	a42	2	a29	3	a65	4	a48	11
a78	1	a127	1	a45	2	a51	3	a70	4	a90	12
a79	1	a128	1	a49	2	a53	3	a114	4		
a82	1	a129	1	a56	2	a54	3	a130	4		
a83	1	a134	1	a59	2	a68	3	a2	5		
a85	1	a135	1	a61	2	a91	3	a35	5		

Table II.4. Summary of allele composition per Group and lake.

			Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 10	Group 11*	Group 12
Lake	Apoyo	Frequency	50	46	13	8	62	33	57	1	0	1	0
		Proportion	80.6	74.2	21.0	12.9	100.0	53.2	91.9	1.6	0.0	1.6	0.0
	Asososca León	Frequency	27	37	5	7	43	36	30	0	0	5	0
		Proportion	62.8	86.0	11.6	16.3	100.0	83.7	69.8	0.0	0.0	11.6	0.0
	Managua	Frequency	22	13	12	7	24	22	18	2	1	0	0
		Proportion	91.7	54.2	50.0	29.2	100.0	91.7	75.0	8.3	4.2	0.0	0.0
	Masaya	Frequency	36	22	17	8	44	39	22	3	11	8	0
		Proportion	80.0	48.9	37.8	17.8	97.8	86.7	48.9	6.7	24.	17.8	0.0
	Nicaragua	Frequency	34	20	19	27	48	26	36	11	5	0	4
		Proportion	69.4	40.8	38.8	55.1	98.0	53.1	73.5	22.4	10.2	0.0	8.2
	Xiloa	Frequency	55	41	36	20	64	53	56	10	1	0	0
		Proportion	85.9	64.1	56.3	31.3	100.	82.8	87.5	15.6	1.6	0.0%	0.0
Total		**Frequency**	**224**	**179**	**102**	**77**	**285**	**209**	**219**	**27**	**18**	**14**	**4**
		Proportion	**78.0**	**62.4**	**35.5**	**26.8**	**99.3**	**72.8**	**76.3**	**9.4**	**6.3**	**4.9**	**1.4**

Table II.4. Chi-square Test of significance between allele frequencies.

Chi-Square Tests of allele groups per lake

Allele Group 1				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	13.255[a]	5	0.021	.022[b]	0.018	0.026
Likelihood Ratio	13.252	5	0.021	.026[b]	0.022	0.030
Fisher's Exact Test	12.617			.026[b]	0.022	0.030
N of Valid Cases	287					

Allele Group 2				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	27.913[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	29.235	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	28.579			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 3				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	30.974[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	32.831	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	32.023			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 4				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	31.097[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	29.910	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	29.176			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 5				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound

Pearson Chi-Square	4.146[a]	5	0.529	.484[b]	0.471	0.497
Likelihood Ratio	4.497	5	0.480	.484[b]	0.471	0.497
Fisher's Exact Test	4.261			.484[b]	0.471	0.497
N of Valid Cases	287					

Allele Group 6				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	36.170[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	36.316	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	34.750			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 7				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	32.780[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	32.535	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	31.852			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 8				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	21.996[a]	5	0.001	.000[b]	0.000	0.001
Likelihood Ratio	25.300	5	0.000	.000[b]	0.000	0.001
Fisher's Exact Test	21.537			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 10				Monte Carlo Sig. (2-sided)		
	Value	df	Asymptotic Significance (2-sided)	Significance	99% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	36.191[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	33.570	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	27.900			.000[b]	0.000	0.000

N of Valid Cases	287					

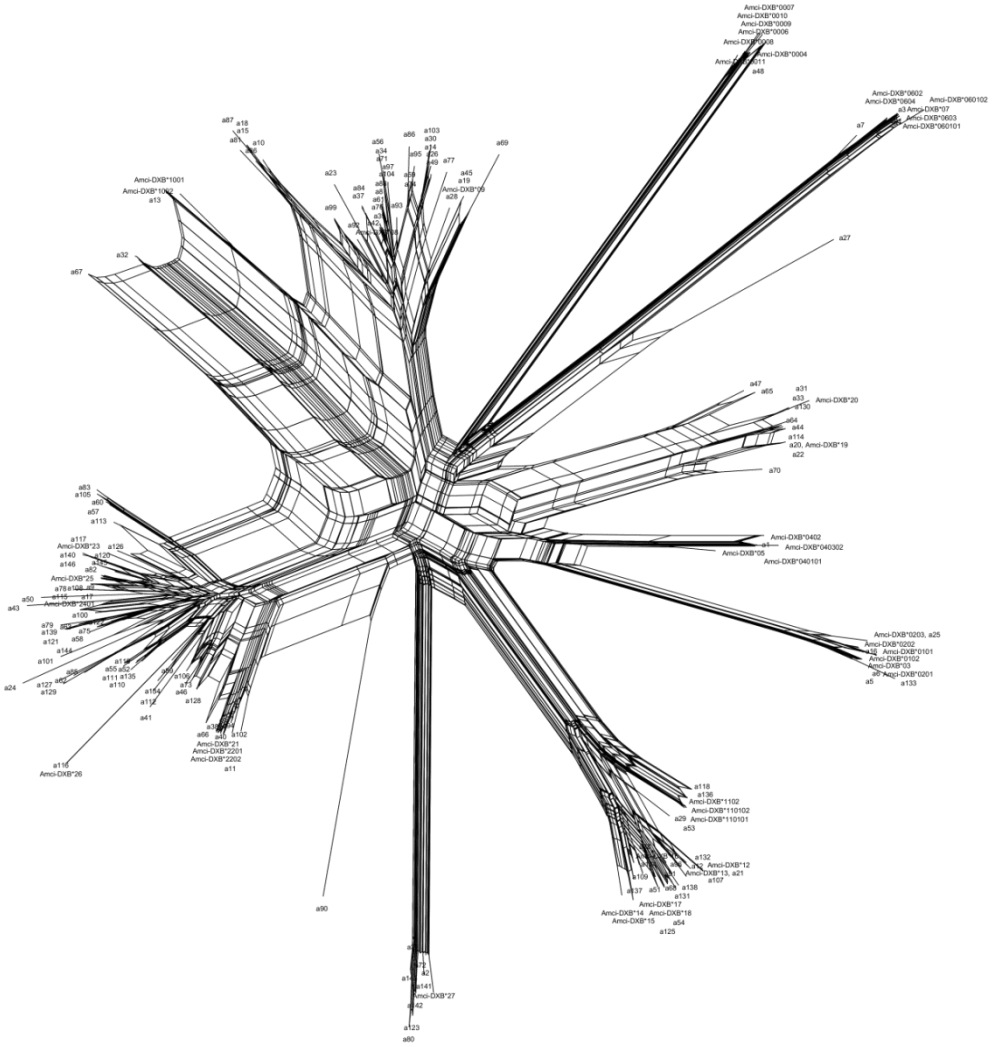
Allele Group 11				Monte Carlo Sig. (2-sided)		
					99% Confidence Interval	
	Value	df	Asymptotic Significance (2-sided)	Significance	Lower Bound	Upper Bound
Pearson Chi-Square	28.810[a]	5	0.000	.000[b]	0.000	0.000
Likelihood Ratio	28.607	5	0.000	.000[b]	0.000	0.000
Fisher's Exact Test	22.151			.000[b]	0.000	0.000
N of Valid Cases	287					

Allele Group 12				Monte Carlo Sig. (2-sided)		
					99% Confidence Interval	
	Value	df	Asymptotic Significance (2-sided)	Significance	Lower Bound	Upper Bound
Pearson Chi-Square	19.703[a]	5	0.001	.003[b]	0.002	0.005
Likelihood Ratio	14.421	5	0.013	.003[b]	0.001	0.004
Fisher's Exact Test	9.333			.003[b]	0.002	0.005
N of Valid Cases	287					

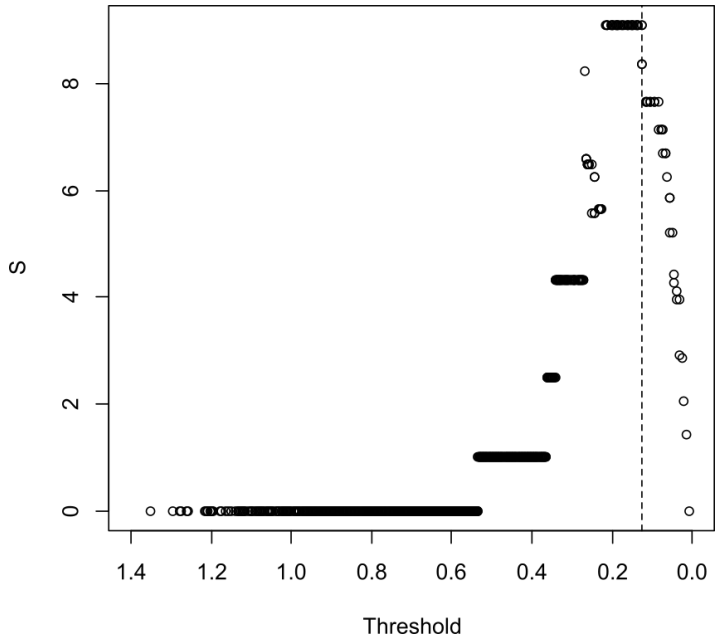
a. 6 cells (50.0%) have expected count less than 5. The minimum expected count is .33.

b. Based on 10000 sampled tables with starting seed 442399356.

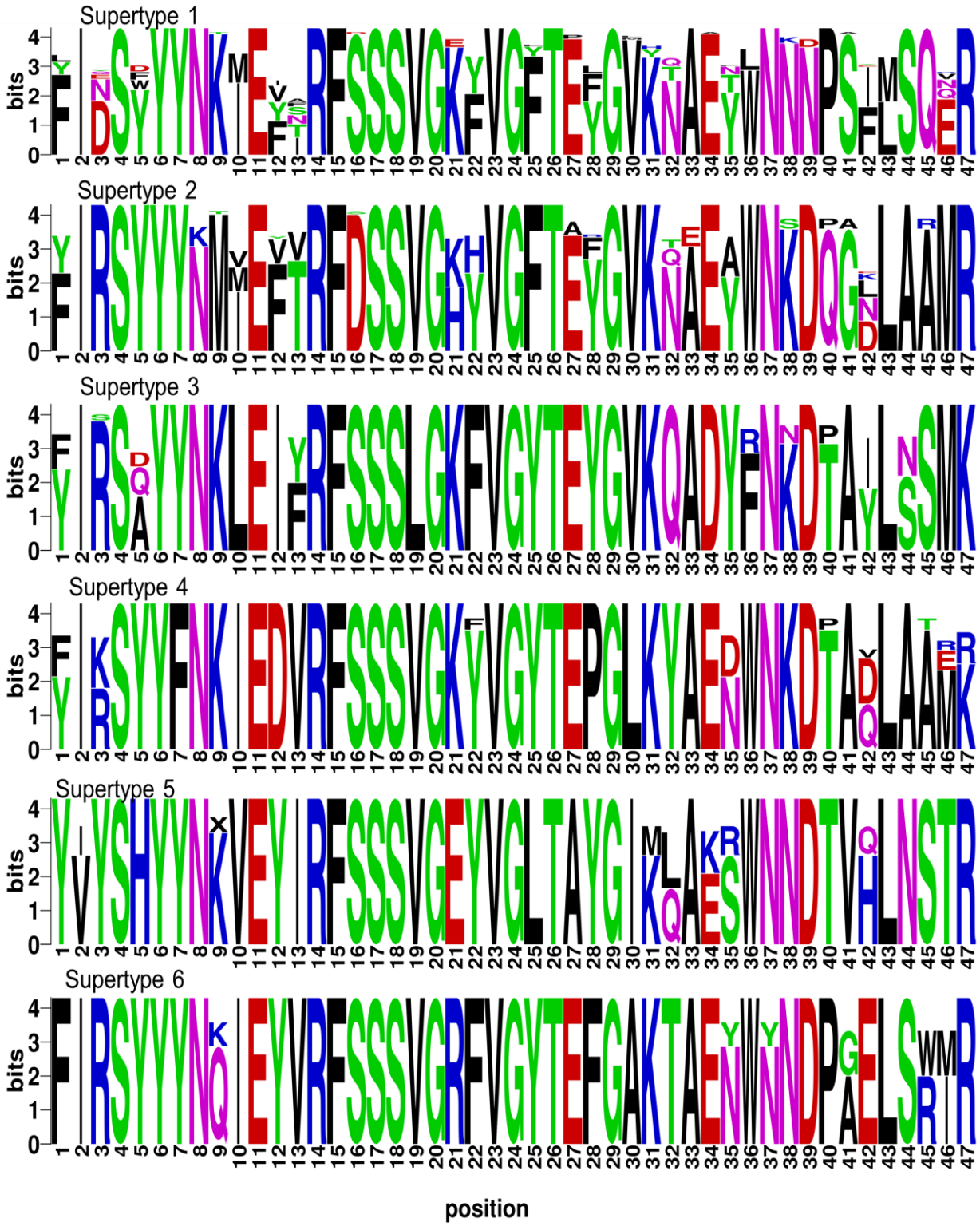
Sup. Figure II.1. Network analysis with all alleles from characterization and NGS

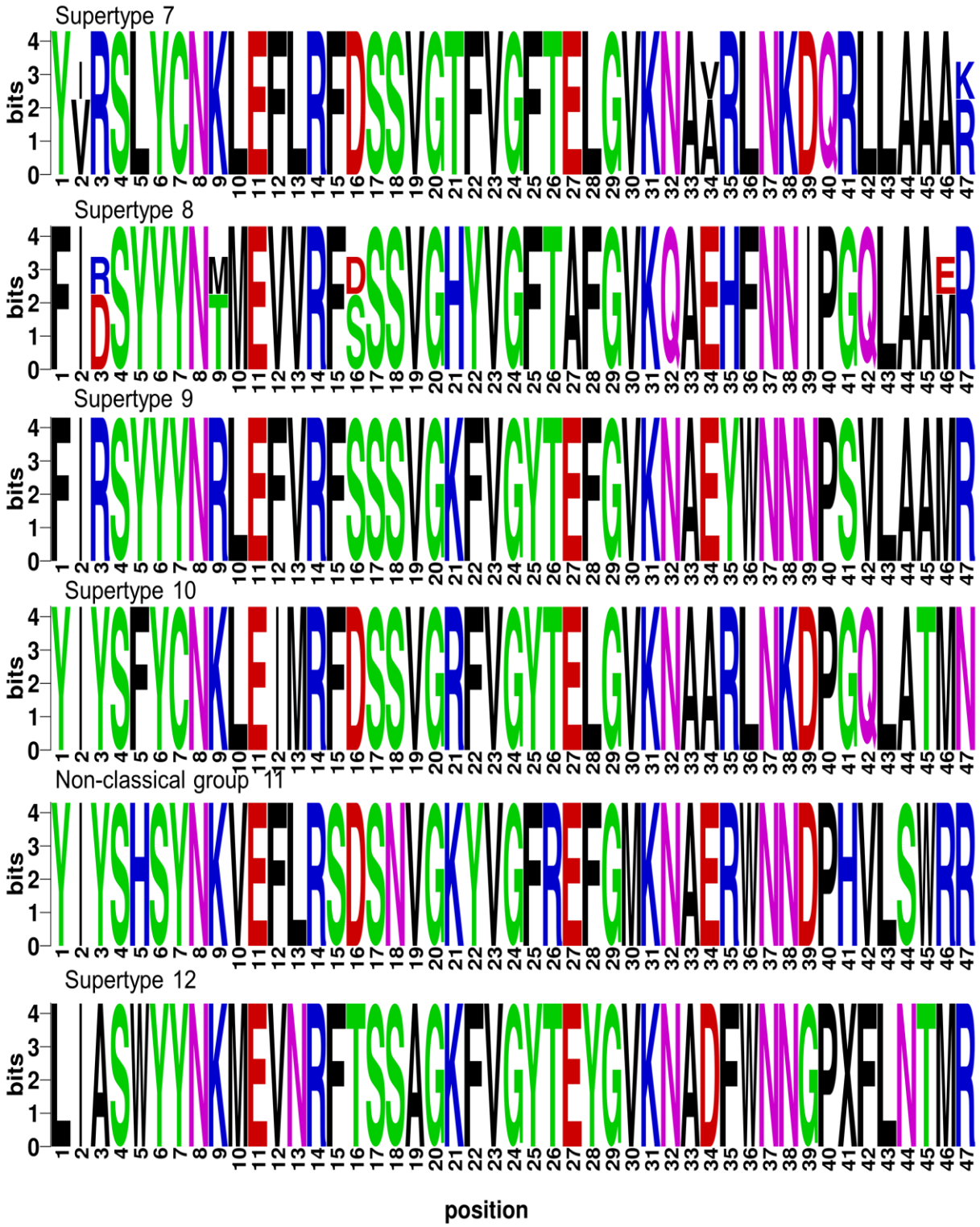


Sup. Figure II.2. Percolation analysis showing the threshold at which the alleles collapse into discrete clusters as a function of genetic distance threshold. This analysis identified Dp = 0.126 as the percolation threshold, and "S" is the average size of clusters excluding the largest one.



Suplementary Figure II.3. Block Logo visualization of peptide residue proportions. The log-transformed frequency are displayed in bits . residues at too low of a frequency for quality resolution are not displayed.




Supertype 7
Supertype 8
Supertype 9
Supertype 10
Non-classical group 11
Supertype 12
bits
position

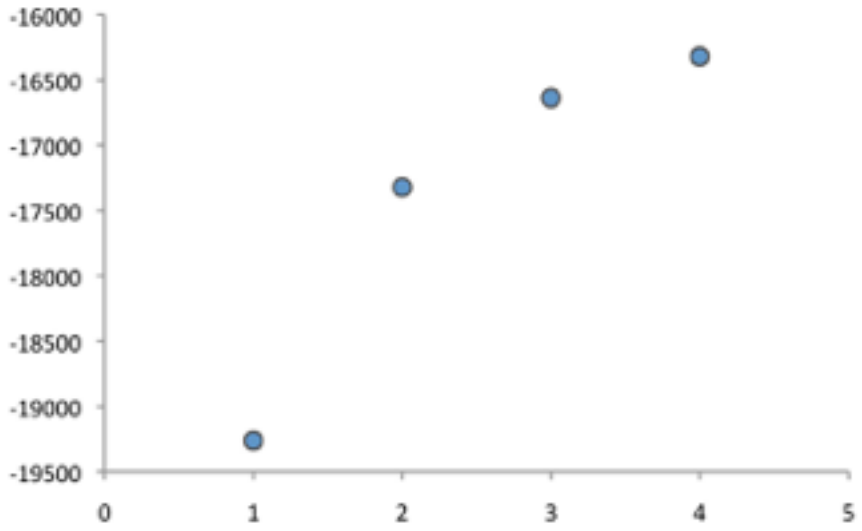
Chapter III

Table III.1. 17 MHC alleles that discriminate better the 6 species in the two crater lakes studied.

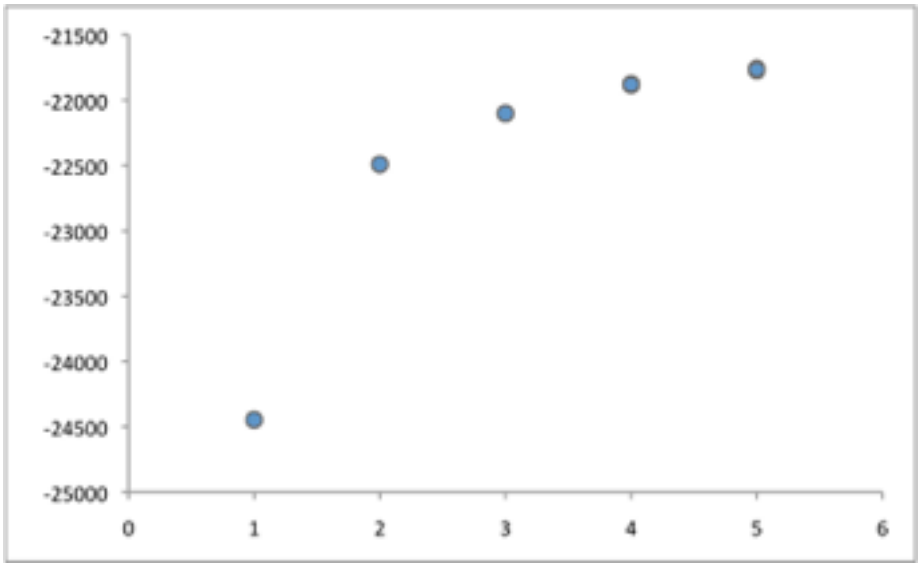
Classification Function Coefficients						
	A. amarillo	A. sagittae	A. xiloansis	A. astorquii	A. chancho	A. zaliosus
a7	0.971	27.721	0.692	-0.673	-0.084	-0.015
a9	0.024	0.077	0.693	9.888	14.792	9.852
a10	-0.660	-1.403	0.205	7.114	2.008	0.426
a12	1.033	34.664	-0.772	-10.474	-14.967	-9.889
a14	-0.070	-0.058	2.571	15.869	38.621	11.278
a17	0.020	0.403	10.026	3.962	11.503	2.760
a18	-0.075	-0.127	0.936	9.944	15.153	12.472
a24	2.175	12.709	0.517	-1.665	-0.386	-0.079
a28	-0.027	-0.085	-0.783	3.149	4.184	14.124
a36	-0.102	-0.135	2.324	15.491	34.739	10.068
a40	1.733	41.970	0.804	-0.448	-0.005	0.003
a58	0.066	0.211	1.923	13.882	30.716	11.591
a60	0.222	2.410	13.838	0.811	2.281	0.547
a75	-0.115	-0.182	1.795	9.208	24.799	4.531
a77	-0.237	-0.563	-1.557	-1.158	-24.334	-7.278
a100	-0.050	-0.033	2.032	9.789	30.050	5.659
a107	-0.900	-26.917	17.464	1.709	3.069	0.731
(Constant)	-1.940	-20.841	-5.416	-8.122	-21.604	-8.887

Sup. Figure III.1. Plot of log-likelihoods resulting from the clustering analysis with STRUCTURE. A. Analysis of samples in crater L. Xiloá (10 replicates). B. Analysis of samples in crater L. Apoyo (10 replicates).

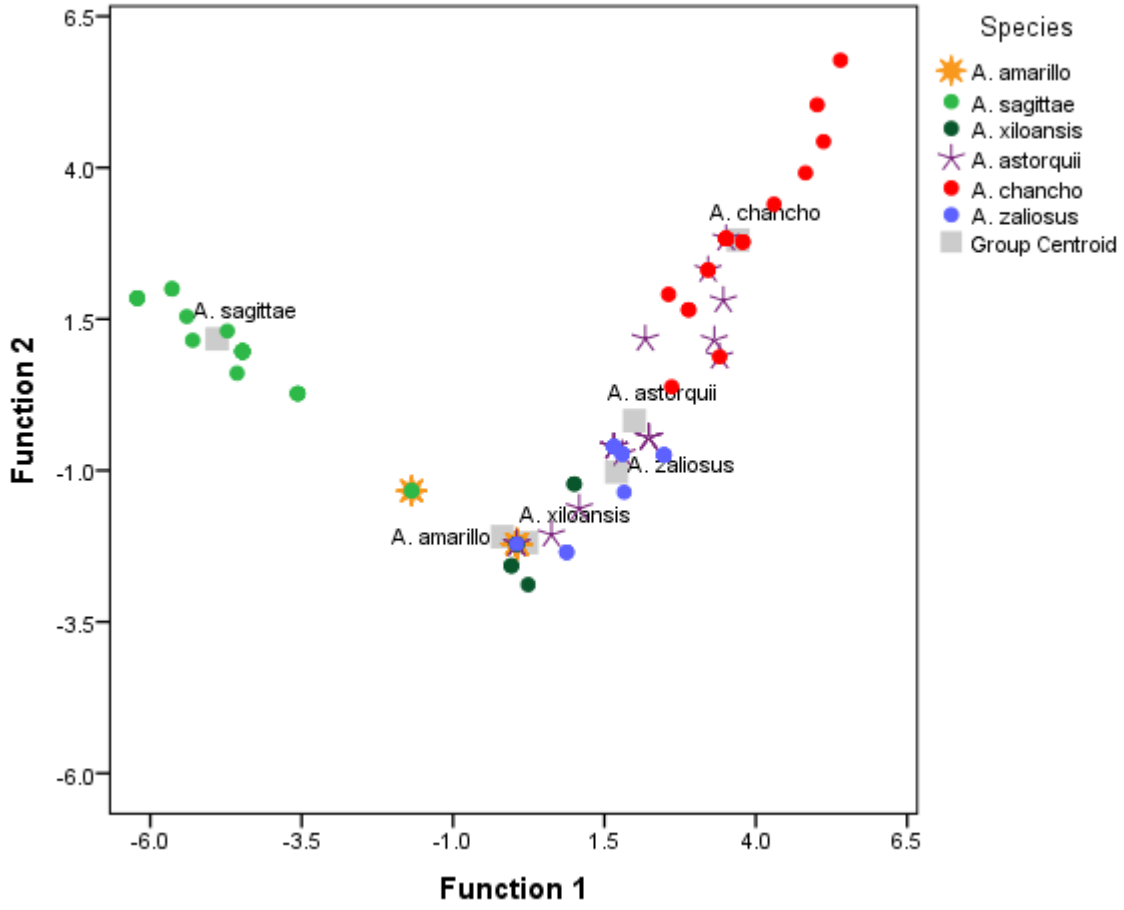
A.



B.



Sup. Figure III.2. Canonical discrimination analysis by species that shows the 17 most discriminating alleles.



Sup. Fig. III.3 Stomach and gut contents (coming soon)