THE GENERATION AND MAINTENANCE OF DIVERSITY IN A RAPID ADAPTIVE RADIATION

A Dissertation
Presented to
The Academic Faculty

By

Nicholas Francis Parnell

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Biology

Georgia Institute of Technology

December 2011

THE GENERATION AND MAINTENANCE OF DIVERSITY IN A RAPID ADAPTIVE RADIATION

Approved by:

Dr. J. Todd Streelman, Advisor
School of Biology
Georgia Institute of Technology

Dr. Joshua Weitz
School of Biology
Georgia Institute of Technology

Dr. Mark E. Hay
School of Biology
Georgia Institute of Technology

Dr. Daniel I. Goldman
School of Physics
Georgia Institute of Technology

Dr. Michael Goodisman
School of Biology
Georgia Institute of Technology

Date Approved: 2 August 2011

ACKNOWLEDGEMENTS

To my advisor, mentor and friend Todd I give my sincere thanks. His style of management, motivation, and teaching has made this a journey worth taking, not with an end in sight but rather a beginning. He asked for nothing but the best and taught me to want more, rather than settling for good enough. I thank Mark Hay for his candidness, honesty and scientific acumen – you truly are an eminent ecologist. I thank Mike Goodisman for advice, discussions, good times and producing incredible Ph.D.s and dear friends of mine. Thanks go to Joshua Weitz for his advice and for being a model of what we should all aspire to be as scientists. Thanks to Dan Goldman for his expertise, lab resources and providing a unique perspective. Finally, I give a very special thank you to Dr. James Haynes – my advisor, mentor, friend and colleague. "Doc" – you started this ball rolling and helped to keep it on line.

I thank my friends at Georgia Tech for entertainment, brilliant discussions, inspiration and making my time here richer. The Streelman "lab family" encompass some of my best friends and most respected colleagues – thank you for your input, help, and of course the hours of laughter. To my other friends, past and present, thank you. And of course Jed, thanks for everything and more.

Finally to my family I am thankful for the support and understanding that you have provided me on this journey. Thank you for always keeping a place for me at home no matter how far I have strayed. I guess the jokes about me being a professional student can stop now…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CIM	Composite interval mapping
cM	Centi-morgan
C-score	Checkerboard score
FT	Force transmission
fxn	Function
KT	Kinematic transmission
LF	*Labeotropheus fuelleborni*
LM	Lake Malawi
LOD	Logarithm of the odds
MM	*Melanochromis melanopterus*
m	meter
mm	Millimeter
MQM	Multiple QTL model
MTOM	Many-to-one-mapping
O	Homozygous orange-blotch color morph
OB	Orange-blotch color morph
PVE	Percent variance explained (estimated)
QTL	Quantitative trait locus
SD	Standard deviation
SE	Standard error
SES	Standardized effect size
SL	Standard length
SNP	Single nucleotide polymorphism
strx	Structure
TG (TS)	Transgression or transgressive segregation
V_E	Environmental variance

SUMMARY

The Lake Malawi cichlid fishes are a pre-eminent example of adaptive evolutionary radiation. The diversity of species (nearly 1000 extant) is mirrored by an array of variation in dozens of phenotypes (e.g. trophic morphology, tooth shape, color patterns, behavior, development). The unique characteristics of this system have produced unparalleled diversity with very little genetic differentiation between species. This dissertation is composed of four studies addressing different aspects of the variation in the LM cichlids and the mechanisms generating and maintaining this level of diversity at multiple biological levels.

Community-level diversity is investigated using null model analysis of species co-occurrence data. We detect signals of non-random community assembly at only the broadest and finest spatial scales. Based on the unique ecological and evolutionary characteristics of this assemblage we suggest that different mechanisms are responsible for these patterns. A 'core' group of species is posited to act as a foundation on which these diverse communities are created as a result of fine-scale species interactions. We identify both positive and negative depth-based correlations between species and suggest these interactions play an important role in species diversity in these fish.

The Lake Malawi cichlids exhibit an array of trophic morphologies which may play a role in the fine-scale species interactions described in chapter one. In the second chapter we build a genetic model to predict the evolution of jaw morphology and a complex functional jaw trait. We use a complex

biomechanical system, the 4-bar jaw linkage, to simulate trait evolution during interspecific hybridizations. We find rampant transgression (trait values beyond parental distributions) in jaw function in a large proportion of potential crosses. This result is characterized by a lack of novel morphological components but rather is the result of recombinations of existing component traits thus producing functional novelty. In the third chapter we create a laboratory cross of one of the parental combinations suggested from the genetic model. The results of this study serve as a proof of principle to the simulations as we observe a large proportion of transgressive 4-bar function in the F_2. As predicted this diversity is produced in the absence of transgressive morphology. We contrast these results between this complex system and data generated from several simple jaw lever traits and report differences in the patterns. Using quantitative trait locus (QTL) mapping approaches we examine the genetic basis for complex and simple jaw traits and discuss correlative patterns within and between systems.

Finally we examine the genetic architecture of sex-determination and color morphs in this hybrid cross. We find both ZW and XY sex systems segregating as well as linkage to sex-specific color patterns. Several loci and epistatic interactions are associated with sex-determination and color morphs in this cross. The orange-blotch (OB) color is found associated with ZW as predicted from previous work but a previously undescribed (in these species) male nuptial color (blue) is found associated with both ZW and XY genetic systems as well as other loci segregating for sex-determination. These results are discussed in the context of models of sex chromosome evolution as a result of sexual conflict and the

potential importance of sexual selection in the diversification of Lake Malawi cichlids.

Overall we observe various mechanisms generating and maintaining diversity at different levels of biological organization. We use community co-occurrence analyses, genetic simulation, and QTL analysis of an F2 hybrid population to examine these mechanisms in this rapidly radiating assemblage. These results bolster our understanding of the origins of diversity and the interplay between variation and aspects of evolution in all biological systems.

CHAPTER 1

INTRODUCTION

A central goal in biology is to understand the generation and maintenance of diversity across levels of biological organization. Elucidating mechanisms generating and maintaining genetic variation, phenotypic novelty and the resultant effects on species diversification has been the focus of evolutionary biology and ecology for decades. Adaptive radiations of organismal groups are historically and contemporarily of great interest to investigations of these topics, due in large part to the rapidity and extent of divergence in genetic, morphological, and ecological traits and configurations inherent to these systems. The integration of concepts, techniques and expertise from across ecological and evolutionary disciplines has broadened the scope of studies in diversity and helped to enhance our understanding of mechanisms important to evolutionary diversification.

Among the most prominent examples of adaptive radiations are the cichlid fishes of the East African rift lakes – Victoria, Tanganyika, and Malawi. The largest of these, Lake Malawi, is also the sight of the fastest and most species rich vertebrate radiation known with nearly 1000 extant species evolving in the last one million years. The Lake Malawi cichlid flock is characterized by not only an array of species but also phenotypic (e.g. tooth shape, jaw structure, coloration, behavior) and ecological (e.g. rich/abundant communities, strong dispersal limitations, high niche overlap) diversity. The rapidity of this radiation has

precluded large-scale genetic differentiation; however, advances in genomics and sequencing technology have revealed interesting patterns of local and regional population structuring, sequence divergence of specific molecular mechanisms, and signatures of selection. Unique to this system we have a natural mutagenic screen of dozens of phenotypes, species without genetic reproductive isolating mechanisms (hybridization is known), dense and species rich local communities, and a genetic record of the intricacies and patterns of rapid speciation and biological diversification. The characteristics of this system make it an excellent candidate for studies of how diversity is generated and maintained at the genetic, organismal, community and phylogenetic levels.

This dissertation engages questions about the generation and maintenance of diversity at multiple biological levels in the rapidly radiating Lake Malawi (LM) cichlid species flock. Chapter 2 provides an ecological window into the diversity of LM cichlids as we hypothesize as to the forces structuring cichlid distributions at varying spatial scales. Using null model analyses we examine species co-occurrence for patterns indicative of non-random structuring. We identify signals of non-random community assembly at the broadest and finest spatial resolutions, but not at intermediate scales. In the framework of the unique ecology and evolutionary history of this group we discuss these patterns as having different causative forces between scales. Focusing on the most commonly occurring species combinations we identify a core group that may form the foundation of cichlid rock-reef communities and play a role in producing and sustaining species richness and abundance.

Chapter 3 focuses on mechanisms affecting phenotypic diversity. This work is defined by a genetic model designed to explicate the dynamics of complex functional trait evolution during interspecific hybridization in LM cichlids. The model is constructed on the foundation of many-to-one-mapping (MTOM) of form to function whereby multiple morphological arrangements lead to the same biomechanical function. We focus on the 4-bar linkage of the anterior cichlid jaw, a complex system known to exhibit MTOM and be correlated with ecologically relevant functions (jaw protrusion, feeding modality, jaw speed). The phenotypes of interest are the four morphological components of the linkage and how they change with respect to changes in function, or kinematic transmission (KT) which is a proxy for jaw speed and protrusion in the 4-bar. Using simulations of interspecific hybridizations the simple genetic model explores the frequency of transgressive segregation, or hybrid traits outside of the parental distributions. Understanding how morphological and functional novelty can be produced in complex traits and how these changes arise with respect to the phenotypic system gives us the power to predict such outcomes, decipher the associations between form and function, and hypothesize expectations of this mechanism in natural populations.

The proof of principle for the genetic model is undertaken in Chapter 4 by generating one of the simulated hybrid crosses in the laboratory. We examine and contrast the resulting simple and complex anterior jaw traits and test hypotheses from the hybrid simulations in Chapter 3. We chose two species which are an archetypal example of many-to-one-mapping in the 4-bar jaw linkage – they have nearly identical function (KT) but different morphological

parts. Trait values beyond the parental distributions (transgressive) and the relationship between structural and functional transgression in this cross are quantified and compared between the simple and complex jaw systems. The genetic basis of the biomechanical function and underlying morphology is determined using quantitative trait locus mapping (QTL). We discuss evidence for genetic and phenotypic correlations and potential evolutionary constraints and trade-offs resulting therein.

In Chapter 5 we broaden the QTL analyses to investigate sex-determination systems and the genetic basis for color traits segregating in this cross. The interest here stems from a potential role for sex conflict-based sex chromosome evolution in speciation. The genetic architecture of multi-factorial sex systems is examined as well as associations with nuptial male colors and a largely female specific color morph. We discuss these results in the context of an antagonistic sex-color system and potential effects of polygenic sex-determination in the rapid and extensive evolutionary diversification of these fish.

This dissertation is the product of a multi-faceted investigation into the generation and maintenance of diversity at various biological levels. Focusing this study on the vastly diverse and rapidly radiating Lake Malawi cichlid fishes has allowed us to gain notable insights into these processes that would not be possible in other systems. We have used a variety of approaches including Monte Carlo ecological analyses, quantitative trait mapping, the construction and testing of genetic simulations, and morphometric and biomechanical measurements and modeling to create a dataset permitting us to test specific

hypotheses concerning these topics. Coalescing these approaches has facilitated new perspectives and helped develop our understanding of the extent of variation and diversity in this rapid adaptive radiation and how that diversity is generated, maintained, and transferred between biologically relevant scales.

CHAPTER 2

THE MACROECOLOGY OF RAPID EVOLUTIONARY RADIATION[1]

2.1 Abstract

A long-standing debate in ecology addresses whether community composition is the result of stochastic factors or assembly rules. Nonrandom, over-dispersed patterns of species co-occurrence have commonly been attributed to competition – a particularly important force in adaptive radiation. We thus examined the macroecology of the recently radiated cichlid rock-fish assemblage in Lake Malawi, Africa at a spectrum of increasingly fine spatial scales (entire lake to depth within rock-reef sites). Along this range of spatial scales, we observed a signal of community structure (decreased co-occurrence of species) at the largest and smallest scales, but not in between. Evidence suggests that the lakewide signature of structure is driven by extreme endemism and micro-allopatric speciation, while patterns of reduced co-occurrence with depth are indicative of species interactions. We identified a 'core' set of rock-reef species, found in combination throughout the lake, whose depth profiles exhibited replicated positive and negative correlation. Our results provide insight into how ecological communities may be structured differently at distinct spatial scales, re-emphasize the importance of local species interactions in community assembly, and further elucidate the processes shaping speciation in this model adaptive radiation.

2.2 Introduction

A central goal in ecology is to understand the patterns and processes that explain the organization of natural communities. A major focus investigates whether communities form as the result of stochastic processes or are constructed via assembly rules (i.e. competition; Connor and Simberloff 1979, Diamond 1975, Weiher and Keddy 1999). The former scenario is attributed to random species colonization, habitat gradients, and stochastic environmental effects, while the latter deterministically generates a community with a marked and predictable signature of co-occurrence due to species interactions (Gotelli 2000, Gotelli and Graves 1996, Gotelli and McCabe 2002). Competitive interactions form communities with species that co-occur less often than expected by chance (Diamond 1975), while species that do co-occur may differ significantly in key traits (e.g. body size or trophic morphology) that relax the degree of overlap in resource use (Dayan and Simberloff 2005). Although competition has been used synonymously with descriptions of structured communities, decreased species co-occurrence can be explained by other mechanisms. For instance, species may not co-occur because of diverged habitat choice; similarly, "historical checkerboards" can result from biogeography (e.g., dispersal barriers) and evolutionary history (e.g. allopatric speciation; Gotelli and McCabe 2002). Recent studies have linked expectations from reduced co-occurrence patterns to empirical verification of species interactions (competition; Adams 2007, Gotelli and Ellison 2002). Studies of community structure have assessed a variety of assemblages including salamanders (Adams 2007), ants (Gotelli and Ellison 2002, Lester *et al.* 2009), desert rodents (Brown *et al.* 2000), beetles (Ulrich and Zalewski 2006), marine reef

fishes (Connolly *et al.* 2005), and birds (Gotelli *et al.* 2010). Despite mixed individual results, a meta-analysis of 96 datasets revealed that non-random community structure (lower species co-occurrence) is more common in natural communities than expected (Gotelli and McCabe 2002). This finding is consistent with the idea that interspecific interactions play an important role in community organization.

While competition is only one possible mechanism producing structured communities, its importance in adaptive radiation is more straightforward (Schluter 2001, Schluter 2009, Schluter and Conte 2009). It is generally accepted that all adaptive radiations have involved a component of divergence due to competition for a limiting resource, usually producing disparity in functional characters, possibly resulting in ecological speciation (Gavrilets and Losos 2009, Losos and Ricklefs 2009, Streelman and Danley 2003). We reasoned that communities built recently from adaptively radiating species would show the expected signature of assembly rules. We thus examined the macroecology of a textbook rapid adaptive radiation, the cichlid fishes of Lake Malawi, Africa.

Lake Malawi (LM) houses some 600-1,000 species that have evolved in the last 1-2 million years, with little to no phylogenetic structure (Genner and Turner 2005, Hulsey *et al.* 2010, Hulsey *et al.* 2007, Loh *et al.* 2008, Seehausen 2006). The LM cichlids contain a lineage of rock-dwelling species (Genner and Turner 2005, Hulsey *et al.* 2007, Loh *et al.* 2008), known locally as the "*mbuna*," characterized by their strict habitat requirements. *Mbuna* dominate rocky habitats in densely packed communities consisting of dozens of species and hundreds of individuals

(Ribbink *et al.* 1983); these rock-fishes exploit all available niches, thus producing local communities built from a single lineage of closely related species. This condition is unique among assemblages examined for community structure (Brown *et al.* 2000, Connolly *et al.* 2005, Diamond 1975, Gotelli and Ellison 2002, Lester *et al.* 2009).

The rocky habitats of Lake Malawi are interspersed by sand flats and deep water, restricting the *mbuna* to near zero dispersal and extreme local population genetic structure (Arnegard *et al.* 1999, Danley *et al.* 2000, Markert *et al.* 1999, Rico and Turner 2002, Streelman *et al.* 2007, Streelman *et al.* 2004, van Oppen *et al.* 1997). The rock-reef *mbuna* thus live an island-like existence – a situation that has been associated with community structure in other systems (Diamond 1975, Lester *et al.* 2009, Simberloff and Connor 1981, Ulrich and Zalewski 2006). Under these conditions, a suite of traits has evolved in the *mbuna* including aggressive territoriality (Fryer 1959, Genner and Turner 2005, Genner *et al.* 1999, Merkert and Arnegard 2007), color-based assortative mating (van Oppen *et al.* 1998), high site fidelity (Hert 1992), and significant overlap in resource use (Bootsma *et al.* 1996, Martin and Genner 2009, Reinthal 1990, Ribbink *et al.* 1983). The distribution and abundance of the *mbuna* is coupled with minimal interspecific variation in body size but an extensive diversity in oral jaw morphology (McKaye and Marsh 1983, Parnell *et al.* 2008, Ribbink *et al.* 1983). *Mbuna* ecology and biology imply a propensity for competition between species (Albertson 2008, Genner and Turner 2005, Reinthal 1990, Ribbink *et al.* 1983), yet no conclusive manipulative data exist. Based on (i) the rapidity and extent of the LM cichlid radiation, (ii) the biology and ecology of the *mbuna* and (iii) the primacy of competition in

adaptive radiations, assembly rules seem obligatory to the formation and maintenance of these communities.

We used null model analyses (Gotelli and Graves 1996) to test for community structure in the LM *mbuna* cichlid assemblage. We employed this approach at multiple spatial scales to investigate how structure emerges as geographic resolution increases – a topic that remains unsettled in the literature (Brown *et al.* 2000, Connolly *et al.* 2005, Gotelli and Ellison 2002, Sanders *et al.* 2007). We made two central predictions: first, because this is a system in which local species interactions are thought to be pervasive and partly responsible for evolutionary diversification, we expected to detect the signature of community structure at the finest scale (depth within sites). Second, because this is a system characterized by extreme endemism and local population structure, we expected to detect the signature of community structure at the broadest scale (lakewide). Indeed, our results demonstrate that the signal of community structure changes with the scale of analysis -- as do our inferences of the forces producing this signal. It is only at the finest scale – that of depth within island sites – where we infer that species interactions shape LM cichlid communities. Therefore, assembly rules exist in communities of LM *mbuna* cichlids, but their influence is extremely localized and does not scale-up to broader spatial levels.

2.3 Materials and methods

2.3.1 The mbuna data

The data used in this study were originally published in a comprehensive survey of the rock-dwelling cichlids of LM (Ribbink *et al.* 1983). Ribbink and

colleagues (1983) made direct counts of species richness and abundance along belt transects at 18 sites throughout the lake and published them as 37 figures of species abundance by depth (kite plots). We estimated the abundance of each species at six discrete depths (1, 2.5, 5, 10, 15 and 20 m) by physically measuring the plots (e.g., Figure A.1). We selected a vertical segment across each depth and quantified the width of each species' plot at that point. In this way, we were able to extract (i) species abundance data at six depths that correspond to actual samples from Ribbink *et al.* (1983), as well as (ii) species presence-absence data at all sites. Species abundance data were used in calculations of diversity (see below) and correlations by depth while presence-absence data were used in null model analyses of species co-occurrence. Species richness was plotted against habitat 'island' size and fit to a power function to understand the cichlid assemblage in the context of island biogeography.

2.3.2 Analysis of community structure

The null model of co-occurrence predicts that species distributions are in accord with a random draw from the respective species pool, encompassing a given scale of analysis. To test this hypothesis, we organized the observed data into presence-absence matrices wherein species are listed in rows, sites are contained in columns and data consist of purely binary entries (1=presence, 0=absence). We analyzed the observed matrix with the EcoSim program, which uses a Monte Carlo algorithm to reorder community matrices based on row (species) and column (site) constraints (Gotelli and Entsminger 2009, Gotelli and Graves 1996). Co-occurrence patterns were examined at four spatial scales: 1)

across the entire lake (called 'lakewide' or 'regional,' 18 sites), 2) across two geographic groups of sites (called 'sub-regional' or 'north & south'), 3) across six transects within two sites each, one in the north and one in the south (termed 'local'), and 4) across the six discrete depths (termed 'local-vertical') at sites. A local-vertical simulation was run for each of the 18 sites individually to examine depth structuring within local communities. We used two null models for co-occurrence analyses: a fixed-fixed model and a fixed-proportional model (SIM9 and SIM4 in Gotelli 2000). SIM9 maintains column sums (i.e. # species at a site) and row sums (i.e. # sites where a species is found) from the observed matrix when producing each simulated matrix, thus preserving species occurrence frequencies and the number of species at sites. This model is suitable for examining 'island list' datasets and exhibits the most robust statistical properties against Type I and Type II errors (Gotelli 2000). The fixed-proportional model (SIM4) also maintains the total number of site occurrences for each species (row sums); however, the number of species at each site is proportional to the total number of site occurrences at that scale. This model is more sensitive to Type I errors than SIM9 but it has been shown to behave robustly in multiple tests. Although site occurrences are not identical to those in the observed matrix, they are proportional to one another, thus maintaining differences between sites (Gotelli 2000). Only SIM9 was used to analyze co-occurrence at the finest scale (local-vertical) because at this resolution the column proportionality of SIM4 approached an equi-probable model (SIM2; equal occurrence probability). Ten-thousand simulated matrices were produced for each model and scale, using the sequential swap algorithm (Gotelli and Entsminger 2001).

Several metrics can be used to test for patterns in the observed matrix, compared to the randomly simulated matrices produced by EcoSim. We chose the checkerboard score (C-score), which measures the average co-occurrence of all species pairs; the C-score has been shown to be robust to Type II errors (Gotelli 2000, Stone and Roberts 1990). This metric is calculated based on the number of shared sites and the number of unique sites between every species combination. The observed community matrix is scored and the probability of the observed score is calculated directly from the distribution of scores from the simulated (randomized) matrices. A significantly higher observed C-score (than the average from the randomized matrices) indicates less average pair-wise species co-occurrence and therefore a structured (non-random) community, while a C-score within the distribution for the simulated matrices indicates a community not different from the null model of random assembly (Gotelli 2000). The EcoSim program returns the observed score, the distribution of simulated matrix scores, the probability of a higher C-score in the observed data, and the standardized effect size (SES) for the observed C-score. The SES represents the distance (in standard deviations) that the observed score lies beyond the mean score of the simulated distribution whereby 95% of the simulated scores fall between -2 and 2. Calculating the SES for each separate analysis allows for meaningful comparison of results across spatial scales used in this study, as well as those of other studies where this metric is employed (Gotelli 2000).

2.3.3 A 'core' community

Lake Malawi houses a high number of species endemic to a single site (Figure A.2; Ribbink *et al.* 1983), a characteristic that may complicate investigations of species interactions due to (already) low levels of co-occurrence (Gotelli and Ellison 2002). Because we were interested not only in the presence of community structure, but also the forces responsible for shaping communities at multiple scales, we extracted a replicated group of co-occurring species for detailed investigation. Several methods have been employed to discover "core" and "satellite" communities in previous studies of other assemblages. Our data did not show certain indicators of a "core-satellite" assemblage: (i) the abundance-occupancy distribution is unimodal, (ii) there exists no positive correlation between the fraction of sites occupied and the average population size, and (iii) no shift is seen from lognormal to log-series between the "core" (>50% sites) and "satellite" (<3 sites) relative abundance distributions (Ulrich and Zalewski 2006). Therefore, to simplify the LM assemblage into a smaller, site-replicated community of species, we used two approaches. First we ranked each species by the total number of site occurrences and added each in order (most sites to single site) to a Shannon diversity index, which uses the relative abundance of each species to quantify a metric representing richness and evenness within an assemblage (Begon *et al.* 2006). At each step the proportional increase in diversity was calculated and the highest peak was used as a proxy for selection of a 'core' community. With this approach we generated a species list comprised of the most abundant and frequently occurring species across all sites in the lake, thereby including the majority of possible species interactions in this set. Next we used cluster analysis to examine

groupings based on a species-occurrence matrix. A bootstrapped hierarchical cluster (1000 iterations) was used to examine which species were grouped together based on this ecological measure of similarity (Sorensen or Bray distance) (R-package "pvclust"). Our clustering strategy to identify 'core' species is akin to other network approaches in ecology and evolution (Proulx *et al.* 2005). Both methods of choosing the 'core' group of *mbuna* species produced a condensed target assemblage for closer examination of species interactions and the forces acting in community structuring.

2.3.4 Species abundances by depth

We asked whether there were patterns of species abundances across the depth gradient within the list of 'core' species using iterative correlation analysis. Monte Carlo randomizations of species abundances by depth (summed across sites) allowed us to compare patterns in our observed data to those expected by random chance (1000 iterations). Product-moment correlation coefficients for observed data were compared to the distribution produced from randomized abundances and the tail probability of each observed correlation was calculated from its respective random family (Gotelli and Entsminger 2009). Significant correlations ($p<0.1$) of 'core' species abundances were used to construct a network of depth-based relationships between species. A less conservative significance level was chosen because of the small sample sizes available for this analysis and the fact that many strong, potentially relevant correlations ($r>0.75$) would be lost at the traditional threshold ($p<0.05$). However,

we realize that using a more liberal critical threshold may increase the chance of selecting spurious correlations.

To address repeated patterns of co-occurrence by depth, we analyzed each significant pairwise correlation (as above) on a site-by-site basis. Replication of correlation across sites was assessed qualitatively as the ratio of (i) repeated significant correlations ($p<0.1$) between two species to (ii) the total number of site co-occurrences for that pair. For instance, if two species show strong correlations ($p<0.1$) of the same direction at 5 sites and they co-occur at 7 sites, the weight of this interaction would be 5/7=0.71, while two species with strong correlations at 5 sites, but co-occurring at a total of 15 sites have a replication weight of 5/15=0.33. In this way the relative replicated correlation between two species can be examined across multiple sites with the weight of a single network edge. Significant correlations ($p<0.1$) by depth between species were used to construct a network, and site-by-site replication was superimposed on this network.

2.3.5 Species richness and depth distribution

To examine the effects of species richness on the depth profile of each 'core' species, we compared patterns in depth distributions at high and low richness sites. First, the abundance distributions of each species were examined across all 18 sites in order of increasing richness using a Jonckheere-Terpstra test. This more powerful analog to the Kruskal-Wallis test is used when samples have a natural ordering, such as richness among our sites (Terpstra and Magel 2003). Next, sites were binned into treatments groups of high and low richness for a

pooled test of species distributions (Kolmogorov-Smirnov). Among all 18 sites used in this study the average species richness was 20; therefore this value was used as the threshold to separate low and high richness sites (N=10 & 8 respectively). Each species' depth distribution was tested for differences in shape and shifts in scale (along the depth axis) occurring with variation in community richness (Kolmogorov-Smirnov 2-sample test). Changes along the depth axis were analyzed using raw depth data while distributional shape differences were examined by mean standardization, thus removing the effects of scale from the shape analysis.

2.4 Results

2.4.1 The Lake Malawi cichlid data

We analyzed previously published census data; Ribbink *et al.* (1983) compiled observations from 18 sites across the lake, including over 40,000 individuals from 138 cichlid species (14 genera). More than half (53%) of the species found were endemic to only one site, while all are endemic to Lake Malawi (Figure A.2; Genner and Turner 2005, Ribbink *et al.* 1983). We focused on the rock-dwelling *mbuna*, which included 134 species and 10 genera. Of the *mbuna*, only *Labeotropheus fuelleborni* was observed at all 18 sites, while *Metriaclima zebra* (found at 12 sites) had the highest overall abundance (6,851 individuals, 16.7% of total fish abundance). An average of 20 species (±1.8 SE) was found at each site and richness ranged from 9 to 36. Species richness was positively correlated with available rock-reef area (Figure 2.1; r^2=0.50), with a distribution well fit to the power function known from the theory of island

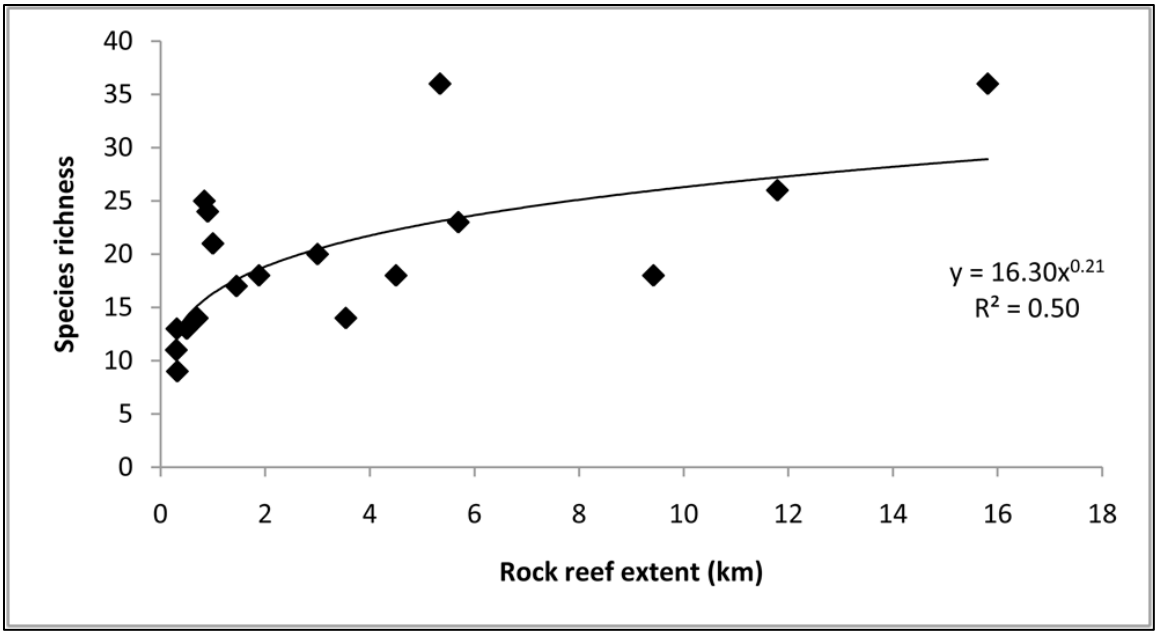

Figure 2.1. *Mbuna* distributions fit the power function familiar from the theory of island biogeography.

biogeography (MacArthur and Wilson 1967). The highest abundance of *mbuna* at any one site was ~7,300 individuals (36 species at Likoma Island).

2.4.2 Community structure in Lake Malawi cichlids

We randomized the species presence-absence matrix thousands of times to compare the observed pattern of co-occurrence to the distribution of matrices produced by chance. *Mbuna* communities showed the signature of community structure (higher C-scores, lower co-occurrence) at both the lakewide and sub-regional (north & south) scales (SES>3.65, $p \leq 0.01$) with SIM9. Notably, local communities showed compositions no different than random (SES<2.14, p>0.07; Table 2.1). The same pattern was observed with the fixed-proportional model (SIM4), excepting the southern sub-regional analysis -- which was no different than null (Table 2.1). A strong signal of community structure was

observed at the local-vertical scale (SES=33.26, p<0.00001; Table 2.1). Independent depth-occurrence results for each site further supported this finding, with non-random structure (SES>4.18, p<0.04) at 13 of 18 sites (72%), after correction for multiple tests.

Table 2.1. EcoSim analysis of Lake Malawi cichlid observed co-occurrence matrix vs. 10,000 randomized matrices. *Scale of analysis* is in descending order from broad to fine, *# sites [# spp]* denotes the number of sites or depths and species used at each scale, *SES* is standardized effect size (see methods) of observed C-score compared to mean of simulated matrices, *Probability of* $C_{obs}>C_{sim}$ is the probability of an observed C-score being greater than the average of random matrices, *Result* is determination of random or structured community based on analysis. Listed values correspond to fixed-fixed model (SIM9); values in parentheses correspond to fixed-proportional model (SIM4) where they differ. All p-values are corrected for multiple tests within a scale.

Scale of analysis	*# sites [# spp]*	*SES*	*Prob of* $C_{obs}>C_{sim}$	*Result*
Regional (lake-wide)	18 [134]	9.80 (3.61)	p<0.000001	SEGREGATED
Subregion – North	11 [113]	8.05 (2.59)	p<0.0004	SEGREGATED
Subregion – South	7 [47]	3.65 (-2.38)	p=0.01 (p>0.90)	SEGREGATED
Local (transect) – North	6 [36]	2.14 (0.33)	p>0.07	NULL
Local (transect) – South	6 [36]	0.37 (-1.37)	p>0.30	NULL
Local-vertical (depth) – across sites	6 [134]	33.26 (-)	P<0.00001 (-)	SEGREGATED
Local-vertical (depth) – 13/18 sites	6 [14-36]	4.18-22.79	p<0.04	SEGREGATED

2.4.3 The 'core' community

Because Lake Malawi *mbuna* contain a large proportion of narrowly endemic species (above), we sought to identify a 'core' assemblage found together throughout the lake. Two approaches converged on the same answer.

The highest proportional increases in the Shannon diversity index were observed at the addition of the 4th and 13th species by rank occurrence (7.6% and 6.8% increase respectively; Figure 2.2). This group of 13 species accounted for 33.5% of the total diversity and 47.2% of the total abundance of fish, while comprising 9.4% of species richness. The individual site richness explained by this focal group ranged from 15.4% to 77.8%.

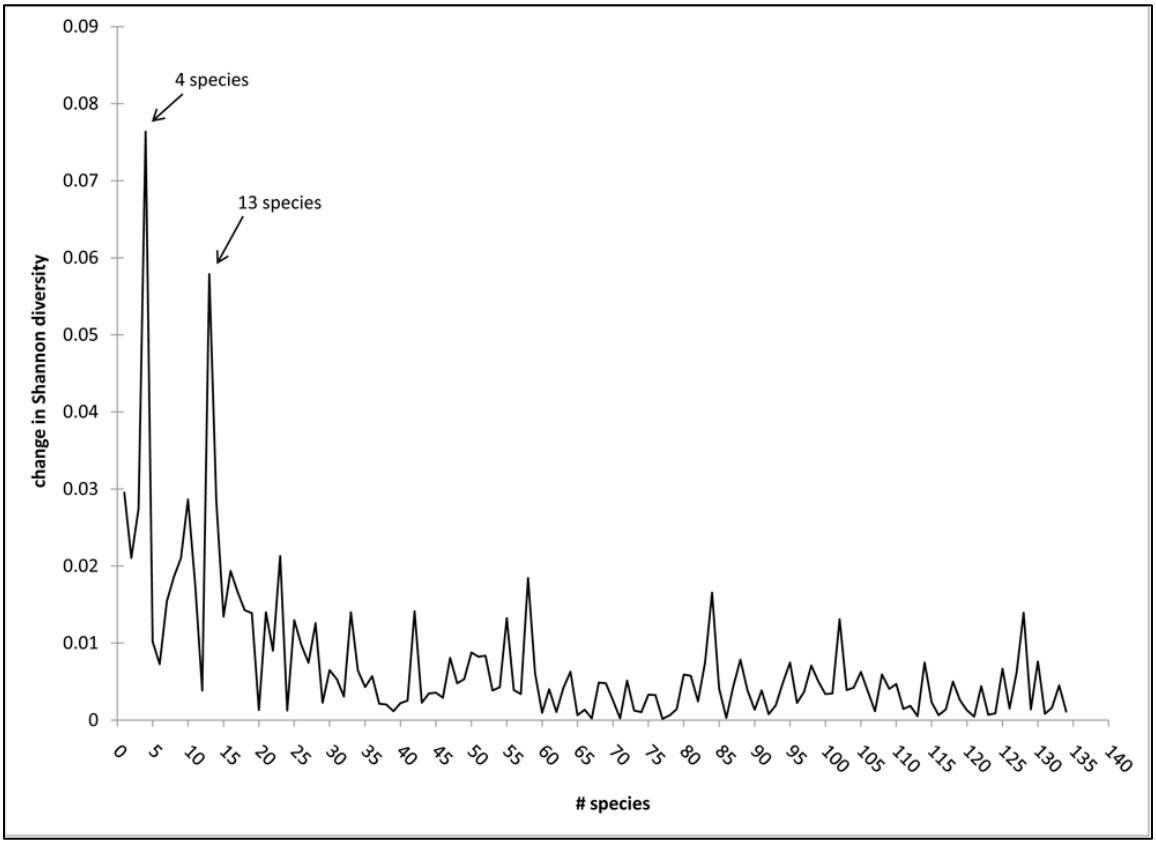

Figure 2.2. 'Core' community selection based on proportional increase in Shannon diversity calculation with addition of species in rank order of site occurrences (from most site occurrences to fewest). Arrows indicate the largest proportional increases at 4 and 13 species respectively.

Cluster analysis produced a 'core' grouping nearly identical to that of the ranked diversity index approach. Twelve species were grouped with a significant multi-scale bootstrapping probability ("approximately unbiased" p=0.97). The

13th species (*Cynotilapia afra*) found above was likely excluded from the cluster 'core' because it occurs predominantly at northern sites – thereby increasing the dissimilarity index upon which the clustering was based. The 12 species consensus 'core' was comprised of: *Genyochromis mento* (parasitic fin-biter), *Labeotropheus fuelleborni*, *Labeotropheus trewavasae* (specialized algal scrapers), *Labidochromis vellicans* (omnivore), *Melanochromis auratus* (omnivore), *Melanochromis melanopterus* (pursuit predator), *Melanochromis vermivorus* (omnivore), *Metriaclima zebra* (omnivore), *Petrotilapia genalutea, Petrotilapia tridentiger* (specialist diatom brushers), *Protomelas taeniolatus* (rock-crack algae sucker), *Tropheops tropheops* (herbivore picker). While one of the species (*Protomelas taeniolatus*) in this set is not a member of the *mbuna* evolutionary lineage (Hulsey *et al.* 2007, Loh *et al.* 2008), it does exhibit similar traits and occurs frequently enough in rocky habitats to account for a large proportion of fish observations (Genner and Turner 2005, Ribbink *et al.* 1983). Identification of 'core' species found together in various combinations at local reef sites permitted subsequent analysis of fine-scale interactions.

2.4.4 Depth distributions are correlated amongst species

Dispersal of cichlids from rock-reef to rock-reef is low (Arnegard *et al.* 1999, Rico and Turner 2002, Streelman *et al.* 2007, van Oppen *et al.* 1997); therefore each site in this study acts operationally as an independent community. To further investigate our observation of strongly non-random community structure with depth (above), we examined correlations of total abundance by depth between the 'core' species. We found both statistically significant positive and

negative correlations (Figure 2.3). Because 'core' species are represented in varied combinations at sites throughout the lake, we then asked if pair-wise abundance by depth correlations were replicated from site to site. Most of the relationships between 'core' species (85%) were consistent across multiple sites, indicative of repeated patterns of interaction with depth.

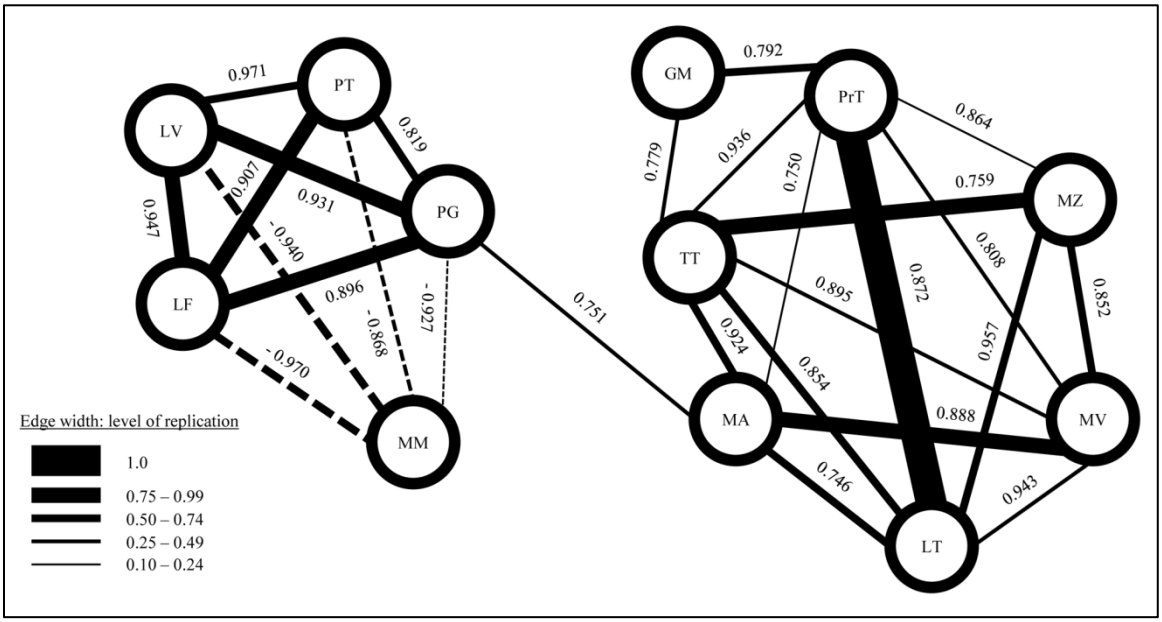

Figure 2.3. Network topology of depth-based product-moment correlations between 'core' Lake Malawi cichlid species. Nodes represent LM cichlid species; edges represent statistically significant correlations by depth (Monte Carlo regression analysis) with positive (grey) and negative (black) correlation coefficients. The level of replication of the relationships across sites is denoted by edge thickness (ratio of consistent correlations to total co-occurrences).

2.4.5 Richness alters species depth distributions

Lake Malawi cichlid communities differ in species richness. We wanted to know if and how species richness affected the depth distribution of 'core' species. Nine of 12 'core' species exhibited significantly different distributions across depth as sites increased in richness (Jonkheere-Terpstra, $p<0.05$). When

sites were binned by richness the depth profiles of 9 of the 12 'core' species differed significantly (Kolmogorov-Smirnov, $p<0.05$; Figure 2.4). Species shifted to both deeper (e.g. "LF") and shallower (e.g. "MM") distributions

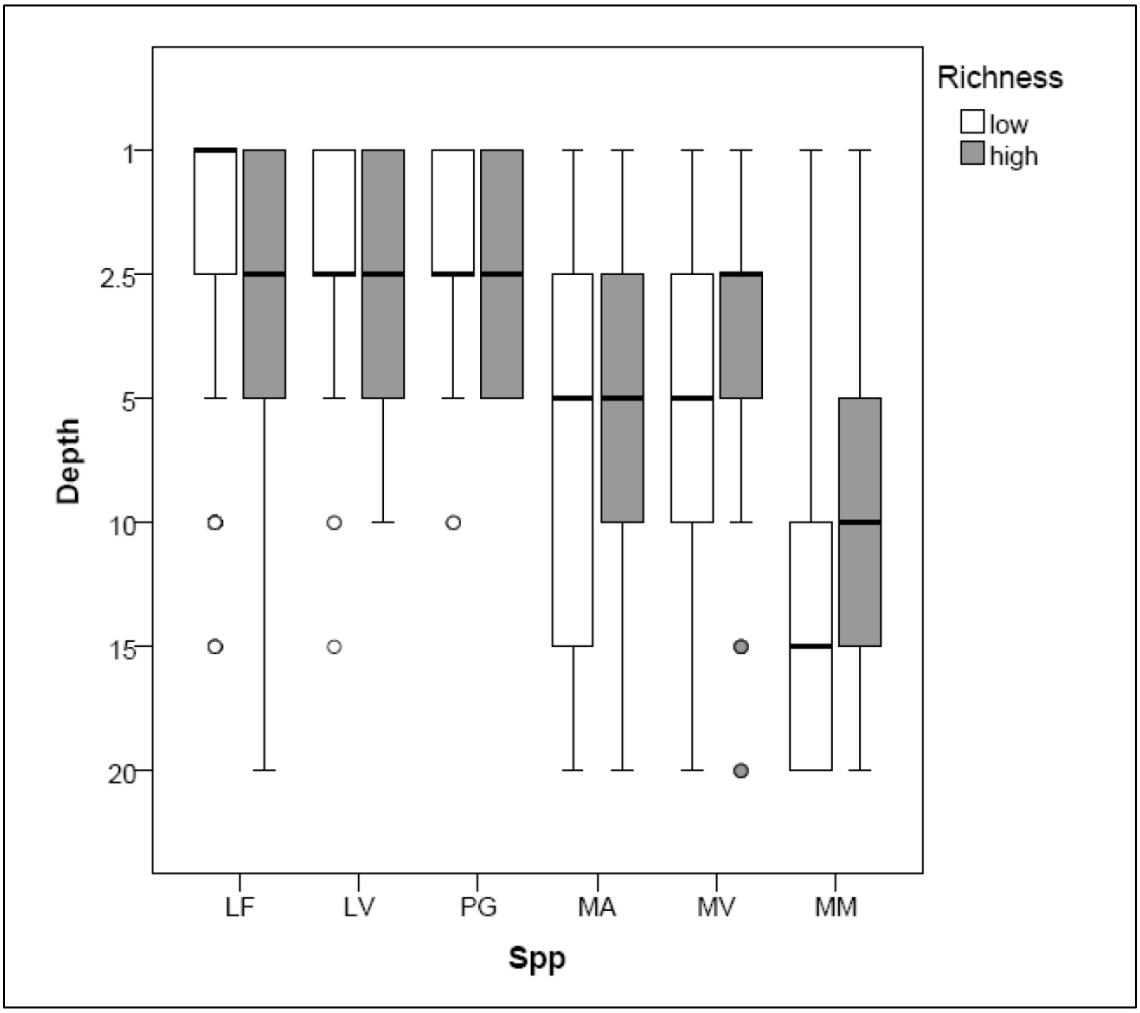

Figure 2.4. Depth distributions of 6 'core' species at low and high richness sites. Species are *L. fuelleborni* (LF), *L. vellicans* (LV), *P. genalutea* (PG), *M. auratus* (MA), *M. vermivorus* (MV), *M. melanopterus* (MM). Horizontal bar represents median depth, boxes are 25th – 75th quartiles, whiskers contain 95th percentiles, points are outliers.

between richness groups. There was no significant difference in fish density between high and low richness groups (Mann-Whitney, $p=0.66$), thus excluding the effect of total fish abundance on species' depth distributions.

2.5 Discussion

Biologists have long used the distributional patterns of species to infer the ecological and evolutionary forces that shape communities (Diamond 1975). The observation of over-dispersed species co-occurrence has been interpreted as evidence of competition (assembly rules), yet the same signature can be produced by other, quite distinct ecological and demographic factors (Gotelli and McCabe 2002). Despite the controversy surrounding competition's role in generating community structure, its importance in adaptive radiation is well accepted (Schluter 2001). We reasoned that if one could catch an adaptive radiation early in its history of diversification, one should be able to capture the signature of species interaction on community structure. To this end, we asked if we could detect the presence of community structure, at hierarchical geographic scales, in the young adaptive radiation of rock-dwelling *mbuna* cichlids from Lake Malawi. These species live and breed on rock-reef islands at high densities (e.g., 7 fish/m^2; Ribbink *et al.* 1983), consume effectively the same food items in different ways, and essentially do not disperse (Genner and Turner 2005).

We expected to detect the signature of community structure for LM *mbuna* at the broadest and finest scales of analysis, for different reasons (below). In fact, our results indicate that cichlid communities <u>are</u> under pressure from deterministic processes, and that these are extremely localized interactions that may shape depth distributions similarly across independent sites. Notably, local-vertical interactions do not scale-up, as *mbuna* communities at the local scale

(transects around the perimeter of sites) are assembled no differently than random.

2.5.1 Island biogeography and cichlid communities at broad scales

Mbuna species exhibit high rates of endemism (53% of species were present at only one rock-reef site, Ribbink *et al.* 1983), and extreme geographic structuring of population genetic variation (Danley *et al.* 2000, Markert *et al.* 1999, Mims *et al.* 2010, Streelman *et al.* 2007). At the broadest scales, we expected that (on average) species would co-occur less frequently than observed in random draws, not because they physically interact, but precisely because they do not. Although the simulations we utilized approximate the species site occupancy and the percentage of endemics in the empirical data, they do not account for the spatially auto-correlated nature of species co-occurrence due to geographic population structure. We reasoned that if the signature of reduced co-occurrence observed at regional and subregional (north & south) scales was influenced by species interactions, such interactions would necessarily involve the 'core' species, whose distributions span multiple sites. However, when we applied the null model co-occurrence analysis to the 'core' species only, at both the regional and subregional scales, we could not reject the hypothesis of random assembly (Table A.1). Thus, there is strong evidence that *mbuna* communities are indeed structured (i.e., different than random) at the broadest scales within Lake Malawi (Table 2.1), but that the cause of 'structure' is evolutionary and demographic history rather than ecological interaction. Our inference here runs counter to that suggested for

Danish avifauna by Gotelli *et al.* (2010), where the signature of species interactions on community structure was apparent even at the regional scale.

2.5.2 Local rock-reef cichlid communities are randomly distributed

Community analyses have traditionally focused on the regional scale (Brown *et al.* 2000, Diamond 1975, Weiher and Keddy 1999), neglecting intermediate scales within locales, perhaps because the data have not been available. Yet this scale represents an important one, linking (or not) processes that occur at the level of individual territories to patterns apparent across regions (Gotelli *et al.* 2010). In the *mbuna* system, this is the broadest scale at which species interactions would likely affect community structure --- individuals tend not to disperse from rock-reef island to island, but they will move around the perimeter of contiguous habitat (Trendall 1988). Notably, our analyses of community structure within two sites failed to detect the signal of decreased co-occurrence ($p>0.07$), and therefore we cannot reject the hypothesis of random assembly. Of course, there may be species interactions at this scale; however, they are apparently not strong nor consistent enough to produce exclusion along site perimeters.

2.5.3 Mbuna communities are structured by depth

Our analysis of cichlid co-occurrence within the depth regime at local sites revealed a striking pattern of non-random community structure (Table 2.1). Co-occurrence between *mbuna* species was lower than expected by chance ($p<0.01$) at the vast majority of sites (72% with Bonferoni correction, 94% before

correction). The few sites that did not exhibit structure were collectively characterized by small habitable areas and low species richness.

We wanted to more closely examine the potential causes of the pattern of community structure by depth revealed by presence-absence co-occurrence analysis. One of the limitations of the EcoSim approach is that it does not reveal *which species* contribute to the signal of community structure. We thus turned to a correlative analysis of species abundance by depth. Once again, because the vast majority of *mbuna* species are single-site endemics, we chose to focus our attention on widespread species, whose interactions (if observed) would likely affect multiple independent rock-reef communities. We used two methods to identify a set of 12 'core' species, found throughout the lake, with disproportionate effects on site diversity and richness. Abundances within this 'core' group exhibited both significant positive and negative correlations by depth, suggesting that structuring within sites may be the result of both positive and negative interactions (e.g. facilitation and competition) coupled with some degree of habitat partitioning (Figure 2.3). It is important to realize here that species with no overlap in their depth distributions (possibly due to competitive exclusion) cannot be analyzed for correlation, and thus our approach may underestimate the number and/or magnitude of negative interactions. Given this caveat, the observation of significant and replicated correlations with depth among 'core' *mbuna* species implies that interactions among these species may form the foundation of local rock-reef community structure, lake-wide – a modification of Fryer's (1959) "peaceful condominium" hypothesis that emphasizes both positive and negative interactions. Our correlational data are

supported by previous observation. For instance, at six of eight sites shared between *Petrotilapia tridentiger* and *Labeotropheus fuelleborni*, this pair is significantly positively correlated by depth. Both Reinthal (1990) and Albertson (2008) have proposed facilitatory feeding between *Petrotilapia spp.* and *Labeotropheus fuelleborni*, obligate diatom brusher and algal scraper respectively. In this case it has been explained in both ways: *Petrotilapia* individuals feed on diatoms after *L. fuelleborni* have trimmed the algae; or *L. fuelleborni* scrape algae after *Petrotilapia* have removed the diatoms.

2.5.4 Beyond pairwise interactions: the community as competitor

Mbuna rock-reef communities differ four-fold in species richness. This is taken to the extreme in the southeastern arm of the lake where two sites (Zimbawe Rock, Thumbi West island) separated by less than 3 km, hold 9 *vs*. 36 *mbuna* species, respectively. Our results suggest that *mbuna* communities are built as geographic endemics are added to some combination of 'core' species. We sought to understand how differences in richness among communities affected the depth distributions of 'core' species. Most (9 of 12) of the 'core' species exhibited significantly different depth distributions at low *vs*. high richness sites. These data suggest that both pairwise and community effects are important shapers of depth distributions in these rich and complex rock-reef environments.

2.5.5 What structures mbuna communities by depth?

The overwhelming signal in our data is one of replicated species interactions structuring communities at the finest scale of depth within local sites.

We speculate here about the root cause of structuring by depth. The first and most obvious interpretation is that *mbuna* species are structured across depth gradients because of variance in trophic resources. Numerous authors have observed that many ecological variables (e.g., light environment, wave action, rock size, algal abundance, abundance of silt and debris) vary with depth in Lake Malawi (Bootsma *et al.* 1996, Bouton *et al.* 1998, Genner *et al.* 1999, McKaye and Marsh 1983, Reinthal 1990, Stauffer and Posner 2006) and that species segregate by depth with associated, and assumed adaptive jaw morphologies (Albertson 2008). At any single depth, multiple species are thought to co-exist via habitat partitioning, as they exploit the same algal resources with divergent modes of feeding (Albertson 2008, Genner *et al.* 1999, Reinthal 1990). Our data, indicative of both positive and negative species interactions, are consistent with the conclusions of these studies.

Alternatively, *mbuna* cichlids may be segregated by depth because of variation in the light environment. *Mbuna* feed and breed in their territories, and male color patterns, filtered through the visual systems of females, are key to mating success (Pauers *et al.* 2008, Turner 1994). The greatest variations in species' abundances are found in the first 5 – 10 m at all sites, a result also reported in Ribbink *et al.* (1983). This depth range realizes the steepest gradient of light penetrance and absorption, thereby representing a bottleneck for the full visual spectrum (Dalton *et al.* 2010, Eccles 1974, Pauers *et al.* 2008). Positive correlations by depth could be attributed to a facilitation-type effect in which species of different colors co-exist more than expected due to decreased aggression (Pauers *et al.* 2008, Seehausen and Schluter 2004), while negative

correlations might indicate cases of competition for specific light environments (i.e., the shallowest depths). It is clear that these two hypotheses (trophic *vs*. light environment) are not mutually exclusive and that multiple factors might interact to produce the patterns we observe. We are now well placed to conduct manipulative experiments to evaluate the predictions of these hypotheses.

2.6 Conclusion

Taken together, our results indicate that (i) significant and replicated effects of non-random mechanisms shape vertical species distributions within sites and (ii) that the incredible diversity in the LM *mbuna* assemblage can be explained by a combination of extremely localized ecological interactions and the unique biogeographic and evolutionary attributes of these fishes. Notably, it is apparent that the forces structuring communities across depth gradients do not persist through broader local and regional scales. Based on our results and the speed and extent of the Malawi cichlid radiation, we suggest that similar patterns may exist in older, less species-rich but more extensively studied ecological systems (e.g., *Anolis* lizard and Darwin's finch ecotypes across islands). Indeed, investigations focused on these assemblages might provide additional insight as the implementation of phylogenetic community analyses should be possible.

2.7 Acknowledgements

We thank N. Gotelli for several helpful discussions of EcoSim and J. Weitz for discussions of choosing the 'core'. We are grateful for the insightful comments and criticism from L. Jiang, M. Hay and three reviewers. Thanks go to Z. Marion

and J. Sylvester for helpful discussions and to the Streelman Lab members for critiques of this work as it progressed. This research was funded by the NSF (IOS 0546423).

2.8 References

Adams, D.C. (2007). Organization of *Plethodon* salamander communities: guild-based community assembly. *Ecology*, 88, 1292-1299.

Albertson, R.C. (2008). Morphological divergence predicts habitat partitioning in a Lake Malawi cichlid species complex. *Copeia*, 689-698.

Arnegard, M.E., Markert, J.A., Danley, P.D., Stauffer, Jr., J.R., Ambali, A.J., & Kocher, T.D. (1999). Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* (Ahl) along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proc. R. Soc. B*, 266, 119-130.

Begon, M., Townsend, C.R. & Harper, J.L. (2006). *Ecology: From Individuals to Ecosystems* Blackwell, Oxford.

Bootsma, H.A., Hecky, R.E., Hesslein, R.H. & Turner, G.F. (1996). Food partitioning among Lake Malawi nearshore fishes as revealed by stable isotope analysis. *Ecology*, 77, 1286-1290.

Bouton, N., van Os, N. & Witte, F. (1998). Feeding performance of Lake Victoria rock cichlids: testing predictions from morphology. *J. Fish Biol.*, 53, 118–127.

Brown, J.H., Fox, B.J. & Kelt, D.A. (2000). Assembly rules: Desert rodent communities are structured at scales from local to continental. *Am. Nat.*, 156, 314-321.

Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005). Community structure of corals and reef fishes at multiple scales. *Science*, 309, 1363-1365.

Connor, E.F. & Simberloff, D. (1979). The assembly of species communities: chance or competition? *Ecology*, 60, 1132-1140.

Dalton, B.E., Cronin, T.W., Marshall, N.J. & Carleton, K.L. (2010). The fish eye view: are cichlids conspicuous? *J. Exper. Biol.*, 213, 2243-2255.

Danley, P.D., Markert, J.A., Arnegard, M.E. & Kocher, T.D. (2000). Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution*, 54, 1725-1737.

Dayan, T. & Simberloff, D. (2005). Ecological and community-wide character displacement: the next generation. *Ecol. Lett.*, 8, 875-894.

Diamond, J.M. (1975). Assembly of species communities. In: *Ecology and evolution of Communities* (eds Cody M.L. & Diamond J.M.). Harvard University Press, Cambridge, pp. 342-444.

Eccles, D.H. (1974). Outline of the physical limnology of Lake Malawi (Lake Nyasa). *Limn. & Ocean.*, 19, 730-742.

Fryer, G. (1959). Some aspects of evolution in Lake Nyasa. *Evolution*, 13, 440-451.

Gavrilets, S. & Losos, J.B. (2009). Adaptive radiation: contrasting theory with data. *Science*, 323, 732-737.

Genner, M.J. & Turner, G.F. (2005). The *mbuna* cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries*, 6, 1-34.

Genner, M.J., Turner, G.F. & Hawkins, S.J. (1999). Resource control by territorial male cichlid fish in Lake Malawi. *J. Anim. Ecol.*, 68, 522-529.

Gotelli, N.J. & Ellison, A.M. (2002). Assembly rules for New England ant assemblages. *Oikos*, 99, 591-599.

Gotelli, N.J. & Entsminger, G.L. (2001). Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia*, 129, 281-291.

Gotelli, N.J. & Entsminger, G.L. (2009). *EcoSim: Null models software for ecology*, v7.0. Acquired Intelligence Inc & Kesey-Bear; http://garyentsminger.com/ecosim.htm.

Gotelli, N.J. & Graves, G.R. (1996). *Null Models in Ecology*. Smithsonian Institution Press, Washington.

Gotelli, N.J. & McCabe, D.J. (2002). Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology*, 83, 2091-2096.

Gotelli, N.J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606-2621.

Gotelli, N.J., Graves, G.R. & Rahbek, C. (2010). Macroecological signals of species interactions in the Danish avifauna. *Proc. Natl. Acad. Sci.*,11, 5030-5035.

Hert, E. (1992). Homing and home-site fidelity in rock-dwelling cichlids (Pisces : Teleostei) of Lake Malawi, Africa. *Env. Bio. Fish*, 33, 229-237.

Hulsey, C.D., Mims, M.C. & Streelman, J.T. (2007). Do constructional constraints influence cichlid craniofacial diversification? *Proc. R. Soc. B*, 274, 1867-1875.

Hulsey, C.D., Mims, M.C., Parnell, N.F. & Streelman, J.T. (2010). Comparative rates of lower jaw diversification in cichlid adaptive radiations. *J. Evol. Bio.*, 23, 1456-1467.

Lester, P.J., Abbott, K.L., Sarty, M., Burns, K.C. (2009). Competitive assembly of South Pacific invasive ant communities. *BMC Ecol.*, 9, 3-12.

Loh, Y.H.E., Katz, L.S., Mims, M.C., Kocher, T.D., Yi, S.V. & Streelman, J.T. (2008) Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Gen. Biol.*, 9, R113.

Losos, J.B. & Ricklefs, R.E. (2009). Adaptation and diversification on islands. *Nature*, 457, 830-836.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton.

Markert, J.A. & Arnegard, M.E. (2007). Size-dependent use of territorial space by a rock-dwelling cichlid fish. *Oecologia*, 154, 611-621.

Markert, J.A., Arnegard, M.E., Danley, P.D. & Kocher, T.D. (1999). Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*: habitat transience, philopatry and speciation. *Mol. Ecol.*, 8, 1013-1026.

Martin, C.H. & Genner, M.J. (2009). High niche overlap between two successfully coexisting pairs of Lake Malawi cichlid fishes. *Can. J. Fish. Aq. Sci.*, 66, 579-588.

McKaye, K.R. & Marsh, A. (1983). Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia*, 56, 245-248.

Mims, M.C., Hulsey, C.D., Fitzpatrick, B.M. & Streelman, J.T. (2010). Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Mol. Ecol.*, 19, 940-951.

Parnell, N.F., Hulsey, C.D. & Streelman, J.T. (2008). Hybridization produces novelty when the mapping of form to function is many to one. *BMC Evol. Bio.*, 8, 122-132.

Pauers, M.J., Kapfer, J.M. & Fendos, C.E. (2008). Aggressive biases toward similarly coloured males in Lake Malawi fishes. *Biol. Lett.*, 4, 156-159.

Proulx, S.R., Promislow, D.E.L. & Phillips, D.C. (2005). Network thinking in ecology and evolution. *Trends Ecol. Evol.*, 20, 345-353.

Reinthal, P.N. (1990). The feeding habits of a group of herbivorous rock-dwelling cichlid fishes from Lake Malawi, Africa. *Env. Biol. Fish*, 27, 215-233.

Ribbink, A.J., Marsh, B.A., Marsh, A.C., Ribbink, A.C. & Sharp, B.J. (1983). A preliminary survey of the cichlid fishes of the rocky habitats in Lake Malawi. *S. Afr. J. Zoo.*, 18, 149-301.

Rico, C. & Turner, G.F. (2002). Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Mol. Ecol.*, 11, 1585-1590.

Sanders, N.J., Gotelli, N.J., Wittman, S.E., Ratchford, J.S., Ellison, A.M. & Jules, E.S. (2007). Assembly rules of ground-foraging ant assemblages are contingent on disturbance, habitat and spatial scale. *J. Biogeog.*, 34, 1632-1641.

Schluter, D. & Conte, G.L. (2009). Genetics and ecological speciation. *Proc. Natl. Acad. Sci.*, 106, 9955-9962.

Schluter, D. (2001). Ecology and the origin of species. *Trends Ecol. Evol.*, 16, 372-380.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323, 737-741.

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B*, 273, 1987-1998.

Seehausen, O. & Schluter, D. (2004). Male–male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. R. Soc. B*, 271, 1345-1353.

Simberloff, D. & Connor, E.F. (1981). Missing species combinations. *Am. Nat.*, 118, 215-239.

Stauffer, J.R. & Posner, I. (2006). An Investigation of the Utility of Feeding Angles Among Lake Malawi Rock-dwelling Cichlids (Teleostei: Cichlidae). *Copeia*, 2, 289-292.

Stone, L. & Roberts, A. (1990). The checkerboard score and species distributions. *Oecologia*, 85, 74-79.

Streelman, J.T. & Danley, P.D. (2003). The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.*, 18, 126-131.

Streelman, J.T., Albertson, R.C. & Kocher, T.D. (2007). Variation in body size and trophic morphology within and among genetically differentiated populations of the cichlid fish, Metriaclima zebra, from Lake Malawi. *Freshwater Biol.*, 52, 525-538.

Streelman, J.T., *et al.* (2004). Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Mol. Ecol.*, 13, 2471-2479.

Terpstra, J.T. & Magel, R.C. (2003). A new nonparametric test for the ordered alternative problem. *Nonpar. Stats.*, 15, 289-301.

Trendall, J. (1988). The distribution and dispersal of introduced fish at Thumbi west island in Lake Malawi, Africa. *J. Fish Biol.*, 33, 357-369.

Turner, G.F. (1994). Speciation in Lake Malawi cichlids: a critical review. *Adv. Limnol.*, 44, 139-160.

Ulrich, W. & Zalewski, M. (2006). Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos*, 114, 338-348.

van Oppen, M.J.H., *et al.* (1997). Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proc. R. Soc. B*, 264, 1803–1812.

van Oppen, M.J.H., *et al.* (1998). Assortative mating among rock-dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Mol. Ecol.*, *7*, 991-1001.

Weiher, E. & Keddy, P. (1999). *Ecological assembly rules: perspectives, advances, retreats.* Cambridge University Press, Cambridge.

CHAPTER 3

HYBRIDIZATION PRODUCES NOVELTY WHEN THE MAPPING OF FORM TO FUNCTION IS MANY TO ONE[2]

3.1 Abstract

Evolutionary biologists want to explain the origin of novel features and functions. Two recent but separate lines of research address this question. The first describes one possible outcome of hybridization, called transgressive segregation, where hybrid offspring exhibit trait distributions outside of the parental range. The second considers the explicit mapping of form to function and illustrates manifold paths to similar function (called many to one mapping, MTOM) when the relationship between the two is complex. Under this scenario, functional novelty may be a product of the number of ways to elicit a functional outcome (i.e., the degree of MTOM). We fuse these research themes by considering the influence of MTOM on the production of transgressive jaw biomechanics in simulated hybrids between Lake Malawi cichlid species. We characterized the component links and functional output (kinematic transmission, KT) of the 4-bar mechanism in the oral jaws of Lake Malawi cichlids. We demonstrated that the input and output links, the length of the lower jaw and the length of the maxilla respectively, have consistent but opposing relationships with KT. Based on these data, we predicted scenarios in which species with different morphologies but similar KT (MTOM species) would produce transgressive function in hybrids. We used a simple but realistic genetic model to show that

Parnell, N.F., C.D. Hulsey and J.T. Streelman. 2008. Hybridization produces novelty when the mapping of form to function is many to one. *BMC Evol. Biol.* 8:122-132.

transgressive function is a likely outcome of hybridization among Malawi species exhibiting MTOM. Notably, F_2 hybrids are transgressive for function (KT), but not the component links that contribute to function. In our model, transgression is a consequence of recombination and assortment among alleles specifying the lengths of the lower jaw and maxilla. We have described a general and likely pervasive mechanism that generates functional novelty. Simulations of hybrid offspring among Lake Malawi cichlids exhibiting MTOM produce transgressive function in the majority of cases, and at appreciable frequency. Functional transgression (i) is a product of recombination and assortment between alleles controlling the lengths of the lower jaw and the maxilla, (ii) occurs in the absence of transgressive morphology, and (iii) can be predicted from the morphology of parents. Our genetic model can be tested by breeding Malawi cichlid hybrids in the laboratory and examining the resulting range of forms and functions.

3.2 Introduction

Biologists are captivated by the evolution of new forms and functions (Gerhart and Kirschner 1997). Over large evolutionary scales, new traits like oxygen metabolism (Raymond and Segre 2006), flowers (Lawton-Rauh *et al.* 2000), limbs (Shubin *et al.* 1997) and the vertebrate neural crest (Gans and Northcutt 1983) have facilitated the ecological and numerical dominance of the lineages that possess them. On more recent timescales, organisms adapt to their environments by continued innovation and modification of these features. A

growing literature seeks to explain the genetic and developmental mechanisms of both macro- and micro-evolutionary novelty (e.g. Streelman *et al.* 2007).

Two recent but largely separate lines of research address the question. The first amounts to a renaissance in our understanding of the evolutionary role of hybridization (Barton 2001, Grant and Grant 1992). Hybrid organisms are generally unfit when compared to parent populations (Burke and Arnold 2001). However, hybrids may have greater fitness than their parents in extreme environments (Rieseberg *et al.* 2003), when environments fluctuate (Grant and Grant 1996), or after environmental and/or human disturbance (Ellstrand and Schierenbeck 2000). Hybridization contributes to 'creative' evolution in two major ways: by (i) providing hybridizing species with genetic variation for adaptive traits, or (ii) producing new species. Hybrid origin (or hybrid swarm) theories of evolutionary radiation combine these two ideas (e.g. Mallet 2007, Seehausen 2004). As an example of (i), Grant *et al.* (2004) used long-term study of Darwin's finches to show that directional introgression of alleles from *Geospiza fortis* to *G. scandens*, coupled with natural selection, could explain trends in beak shape and body size. The work of Rieseberg and colleagues on *Helianthus* sunflowers provides one of the best examples of (ii). Hybrid speciation in *Helianthus* is a product of chromosomal rearrangement (Rieseberg *et al.* 1999a) and ecological selection acting on pleiotropic genes in remote habitats (Rieseberg *et al.* 2003). *Helianthus* hybrid species specialize in extreme environments, where they out-compete their parents (Lexer *et al.* 2003). This phenomenon of transgressive segregation (TS), in which hybrids outperform parental forms, is more common than once believed. Rieseberg and others (1999b) summarized data from 1229

traits in 171 experiments and found that 91% of studies reported at least one transgressive trait and that 44% of traits were transgressive overall.

Importantly, Rieseberg *et al.* (1999b) also considered the genetic architecture of TS and proposed that complementary action of genes with additive effects was the major cause. Complementary gene action works as follows (see also Table 1 of Rieseberg *et al.* 1999b): species A (large) segregates 6 quantitative trait loci (QTL) for size and 4 of these contribute phenotypic effects of 'large,' while 2 contribute phenotypic effects of 'small;' species B (small) also segregates 6 QTL for size, 4 of which contribute phenotypic effects of 'small,' while 2 contribute phenotypic effects of 'large.' Species A and B produce hybrids that intercross to produce F_2; after recombination and independent assortment, some F_2 are larger than parent A (e.g., F_2 with 6 'large' QTL) and some are smaller than parent B (e.g., those with 6 'small' QTL). Thus, complementary gene action may be a general genetic mechanism of TS.

An upshot of this reasoning was that traits experiencing a history of strong directional selection were less likely to exhibit TS (Rieseberg *et al.* 2002). This was codified in Orr's QTL sign test (Orr 1998), which compares the distribution of direction of QTL effects to a random expectation, as a quantitative genetic metric of directional selection. Albertson and Kocher (2005) extended this logic to argue that directional selection and genetic architecture limit the evolutionary role of TS. Using Malawi cichlid hybrids, they showed that lower jaw shape, a trait under strong directional selection (Albertson *et al.* 2003), did not exhibit TS while the shape of the neurocranium, a trait not under directional

selection, did. They concluded that natural selection might constrain the degree of diversification that can be achieved via hybridization. This might hold, in particular, for aspects of the craniofacial skeleton involved in feeding, including the shape of skeletal components, the dentition, and simple levers involved in jaw function (Albertson *et al.* 2003, Albertson *et al.* 2005, Streelman and Albertson 2006).

A second line of research targeting the evolution of novelty departs from a concentration on trait morphology *per se* and instead focuses on function, or more specifically the relationship between form and function. Investigators have contrasted the relationship between shape and biomechanics for simple versus more complex lever systems in fish jaws to illustrate a general principle of organismal design (Alfaro *et al.* 2004). For instance, the 4-bar linkage system (Figure 3.1) models the rotation of the maxilla given a known input of lower jaw depression. The output of the model can be computed by geometry when component lengths and angles are measured. As with simpler, two component lever systems, one can estimate kinematic transmission (KT), in this case the amount of motion transmitted to the maxilla, through the linkage system, from lower jaw depression (Hulsey and Wainwright 2002). Jaw linkages with higher maxillary KT transmit more motion through the system and are predicted to evolve in fishes that feed on elusive prey (Hulsey and Garcia de Leon 2005). In contrast to simple levers, morphological variation is only a modest predictor of KT for the 4-bar system (Alfaro *et al.* 2004). Many different morphological designs can yield similar values of maxillary KT (Hulsey and Wainwright 2002); this is called many to one mapping (MTOM) of form to function. In essence, MTOM is an

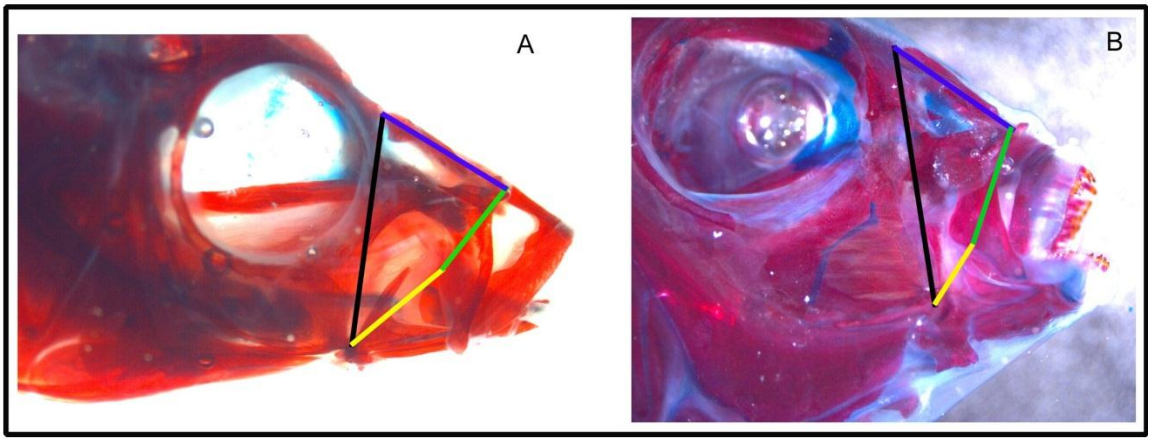

Figure 3.1. The components of the 4-bar linkage system are illustrated on cleared and stained cichlid heads of (A) Copadichromis eucinostomus (representative high KT) and (B) Chilotilapia rhoadesii (representative low KT). Yellow is the lower jaw (input) link; black is the fixed link, blue is the nasal link and green is the maxillary (output) link. Note differences in relative lengths of input and output links for these high and low KT exemplars.

example of functional redundancy. Diverse lineages that exhibit MTOM may be predisposed to evolutionary novelty, because manifold solutions exist among species for new ecological or functional demands (Alfaro *et al.* 2004). MTOM of form to function may be a general property of all but the simplest biomechanical systems and may influence system diversity, robustness, modularity and evolvability in complex ways (Hulsey *et al.* 2005).

We reasoned that MTOM of form to function might contribute to *functional transgression* in hybrid offspring. Just as complementary gene action describes hybrid trait transgression as the piling up of QTL of the same direction (e.g., 'large') from different parental genomes, we envisioned scenarios in which MTOM parental species might segregate 4-bar links to hybrids in combinations producing biomechanical output (KT) well beyond the parental range. To

evaluate this possibility, we characterized the 4-bar mechanism in Lake Malawi cichlids and simulated hybrids using a simple but realistic genetic model. Lake Malawi cichlids are apposite study organisms for this question. They exhibit tremendous morphological diversity on a background of genomic mosaicism (Hulsey *et al.* 2007, Loh *et al.* 2008, Won *et al.* 2005). Species hybridize in the wild (Smith *et al.* 2003, Smith and Kornfield 2002, Streelman *et al.* 2004) and speculation persists that hybridization has contributed to the adaptive radiation of this species flock (Seehausen 2004). Our simulations demonstrate that transgressive function is a likely outcome of hybridization among Malawi species exhibiting MTOM. Functional transgression is a product of recombination and assortment between alleles controlling the lengths of the lower jaw and the maxilla, occurs in the absence of transgressive morphology, and can be predicted from the morphology of parents.

3.3 Results and discussion

3.3.1 Diversity and correlation in the Malawi 4-bar linkage

We calculated maxillary KT from 86 Lake Malawi cichlid species (169 individual fishes) sampled in July 2005 or borrowed from the American Museum of Natural History. Malawi species exhibit a wide spread of KT values (range for individuals, 0.41 – 1.57; Figure 3.2; [see Table B.1 for species averages]) comparable to that observed in other fish lineages (Wainwright *et al.* 2004, Westneat *et al.* 2005). The bulk of the specimens had KT values between 0.61 – 0.80 (mean= 0.77; s.d.=0.16), corresponding to putatively intermediate jaw speeds and force potentials. Planktivores, on average, exhibited higher maxillary

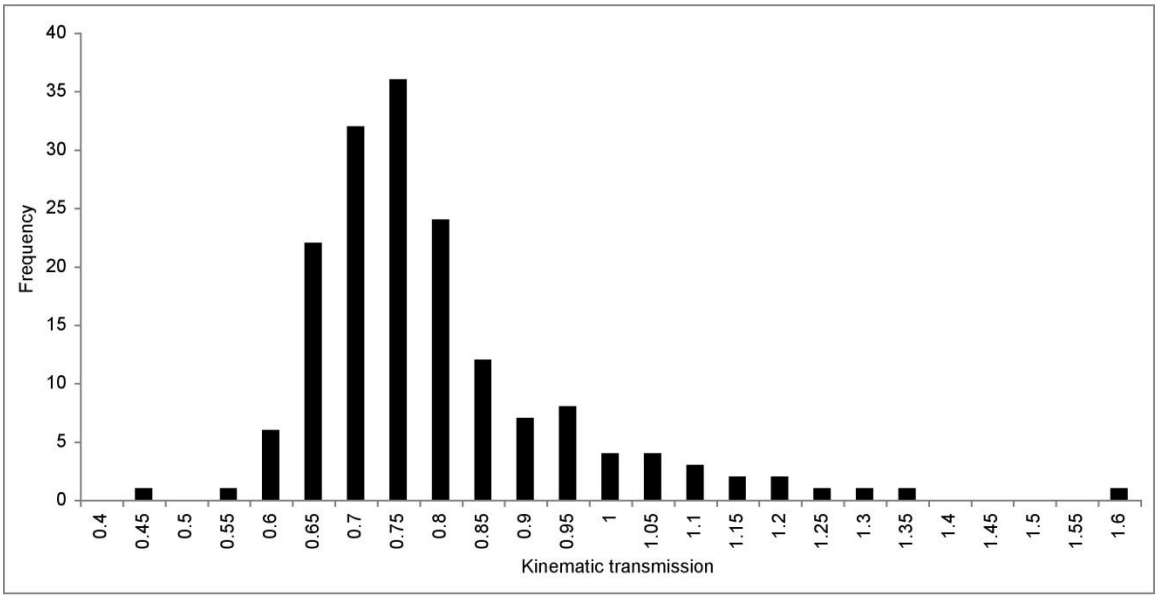

Figure 3.2. The distribution of kinematic transmission (KT) for 169 individuals from 86 Lake Malawi species shows that the majority of individuals have KT values between 0.65 – 0.80.

Table 3.1. Correlations (r^2) between links and KT among Lake Malawi cichlids are similar for uncorrected (below the diagonal) and phylogenetically independent contrasts (above the diagonal). Underlined values are negative correlations; * indicates statistical significance at p<0.0005.

	LJ	**Max**	**Nasal**	**Fixed**	**KT**
LJ		0.44*	0.00	0.47*	0.06
Max	0.33*		0.12	0.53*	<u>0.30*</u>
Nasal	0.06	0.08		0.12	0.11
Fixed	0.18*	0.54*	0.06		<u>0.07</u>
KT	0.23*	<u>0.15</u>*	0.00	<u>0.11</u>	

KT than other trophic groups [see Figure B.1]. The input (lower jaw, r=0.473; p<0.0005) and output (maxilla, r=-0.379; p<0.0005) links were significantly and antagonistically correlated with KT across the data set (Table 3.1).

We next asked whether correlations among links and KT were robust to the pattern of Malawi cichlid evolutionary history. It has been known for some time that trait correlations should be corrected for phylogeny because related individuals do not represent independent statistical samples (Felsenstein 1985). This is a particularly important and difficult problem in the Malawi cichlid assemblage where species are young (e.g., 1000 years, Won *et al.* 2005) and phylogenetic resolution is limited (Hulsey *et al.* 2007, Won *et al.* 2006). Table 3.1 (above the diagonal) shows evolutionarily independent correlations calculated using the phylogenetic topology shown in Figure B.2, employing an approach described in Hulsey *et al.* (2007). We use a single mitochondrial genealogy because the locus offers more resolution than nuclear genes in this rapidly evolving lineage (Hulsey *et al.* 2007, Won *et al.* 2006). The corrected correlations are similar to the uncorrected in all but one case. The lower jaw link, positively associated with KT using uncorrected data, is not associated with KT across the Malawi phylogeny. Further inspection of the data suggests that this is due to lack of power to detect a phylogenetic association (i.e., lack of divergent independent contrasts) as most rock-dwelling *mbuna* have relatively short lower jaws compared to most non-*mbuna* taxa.

Given this caveat, our empirical data support results from simulation in other studies (Alfaro *et al.* 2004, Hulsey and Wainwright 2002): longer lower jaws

on average yield higher values of KT while longer maxillae produce the opposite (see also Figure 3.1). Notably, lower jaw and maxillary links were positively correlated with one another in our data set (corrected r^2=0.44; p<0.0005; Table 3.1). Moreover, a simple ratio of LJ/Max lengths was a near perfect predictor of the fully parameterized calculation of KT (Figure 3.3; r^2=0.94; p<0.0005). Our data point to an appreciable linear component to variation in this 4-bar linkage system noted for its non-linear evolutionary dynamics (Alfaro *et al.* 2004, Hulsey and Wainwright 2002).

Comparison of species' trait values (corrected for body size) identified numerous cases of convergence in KT via divergence in link length, or MTOM of form to function (Alfaro *et al.* 2004). The correlations noted above help to explain this observation. For example, both *Cynotilapia afra* and *Pseudotropheus elongatus* have KT values of approximately 0.7, but they exhibit distinct morphologies. *Cynotilapia afra* has relatively long input and output links, while *P. elongatus* has relatively short elements of each. We reasoned that hybridization between species like these that are MTOM for function might produce F_2 hybrids with extreme KT values, if independent assortment recombines the long lower jaw of *C. afra* with the short maxilla of *P. elongatus* (high KT) or the long maxilla of *C. afra* with the short lower jaw of *P. elongatus* (low KT). We tested the generality of this prediction using a simple, but realistic genetic model.

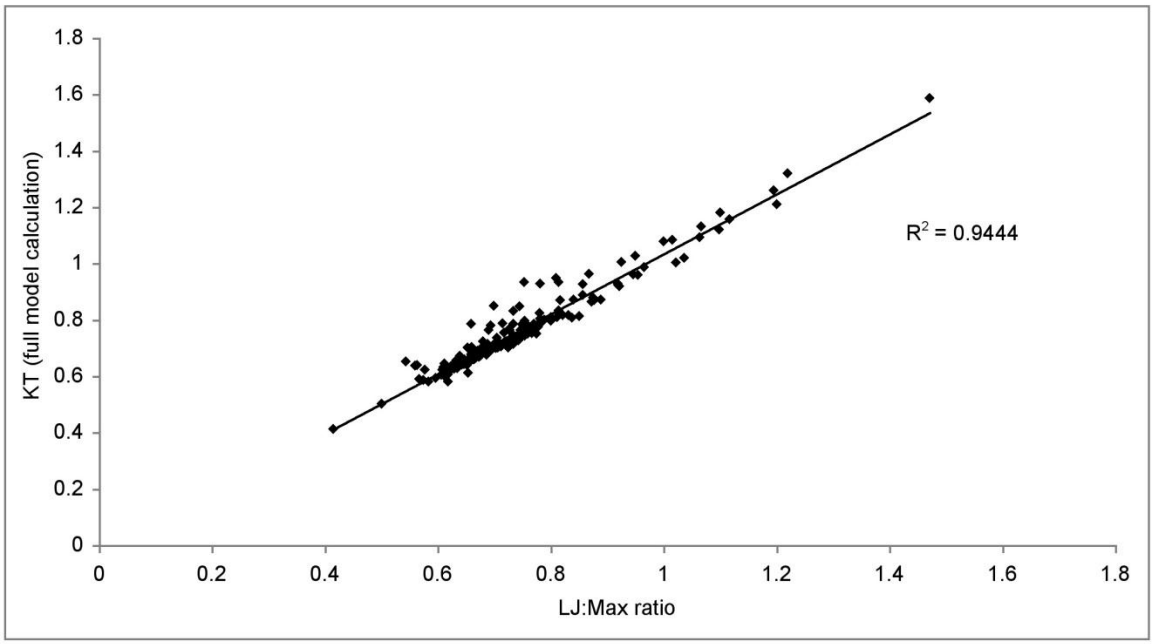

Figure 3.3. A simple ratio of input (lower jaw) to output (maxilla) links is strongly positively correlated with the fully parameterized calculation of KT.

3.3.2 Simulated hybrids of MTOM species are transgressive in function

Our Mendelian genetic model assumes no evolution and no environmental variance (V_E=0). We specified that size-standardized lengths of each of the four linkage components (Figure 3.1) were controlled by 4 independent loci (2 alleles at each locus). Links themselves were modeled to be genetically independent of one another. The effect of each allele was assumed to be equal and additive. Importantly, the assumptions of (i) number of loci per link, (ii) genetic independence of links and (iii) additivity of allelic affects are supported by empirical quantitative genetic data for the cichlid craniofacial skeleton (Albertson *et al.* 2003). For example, even though the lower jaw and maxilla links are positively (phenotypically) correlated in our data set, this is unlikely to be due to genetic linkage; QTL for the shape of these elements map to distinct chromosomal regions. We used this model to (a) determine if MTOM

species would produce hybrids transgressive for function, (b) estimate the percentage of F_2 progeny exhibiting transgression from specific Lake Malawi cichlid intercrosses, and (c) investigate the genetic contribution of each link to the phenomenon of functional transgression.

We assumed that each F_0 parent was homozygous at every locus; since every allele had additive and equal effects, no transgression was produced in the physical lengths of 4-bar links. By contrast, transgression in 4-bar function (KT) was found in 80% of simulated crosses (16/20, Table 3.2) in which we calculated KT for each and every of 6,561 composite F_2 genotypes. Note that we have employed a modified and conservative definition of transgression, specific to the distribution of KT produced by the model; an individual is 'transgressive' if its KT value lies ±1 standard deviation (or more) from the mean. The frequency of transgressive individuals, per cross, ranged from 0.1 – 34%, with up to 6% of the KT values lying two (or more) standard deviations beyond the mean for some (e.g., *C. afra* x *P. elongatus*; Figure 3.4).

Table 3.2. Simulated crosses of Lake Malawi cichlids produce transgression at appreciable frequencies. Transgressive (TG) individuals exhibit kinematic transmission (KT) at least one SD either side of the mean from the distribution of hybrid KT values. H/L indicates whether transgression occurs with values higher and lower than the parents.

Species	**TG F_2?**	**% TG**
Cynotilapia afra x Pseudotropheus elongatus	H/L	29
Protomelas fenestratus x Protomelas ornatus	H/L	32
Chilotilapia rhoadesii x Maravichromis subocularis	H/L	14.5
Copadichromis quadrimaculatus x Corematodus taeniatus	H	1
Cynotilapia afra x Metriaclima zebra	H/L	22
Protomelas ornatus x Protomelas taeniolatus	H/L	8
Protomelas spilopterus 'blue' x Protomelas fenestratus	H/L	21
Copadichromis virginalis x Chilotilapia rhoadesii	H	20
Petrotilapia nigra x Pseudotropheus elongatus	H/L	29
Metriaclima zebra x Labeotropheus trewavasae	H/L	9.2
Nimbochromis polystigma x Chilotilapia rhoadesii	H/L	31
Pseudotropheus elongatus x Labeotropheus fuelleborni	H	0.1
Metriaclima zebra x Pseudotropheus elongatus	H/L	34
Copadichromis mloto x Otopharynx picta	H	0.1
Labidochromis vellicans x Pseudotropheus elongatus	H/L	19
Ps. Tropheops 'orange chest' x Ps. Tropheus 'red cheek'	H	0.1
Copadichromis quadrimaculatus x Dimidochromis compressiceps	NO	
Copadichromis virginalis x Lethrinops altus	NO	
Copadichromis virginalis x Tyrannochromis macrostoma	NO	
Ps. Tropheops 'orange chest' x Labeotropheus fuelleborni	NO	

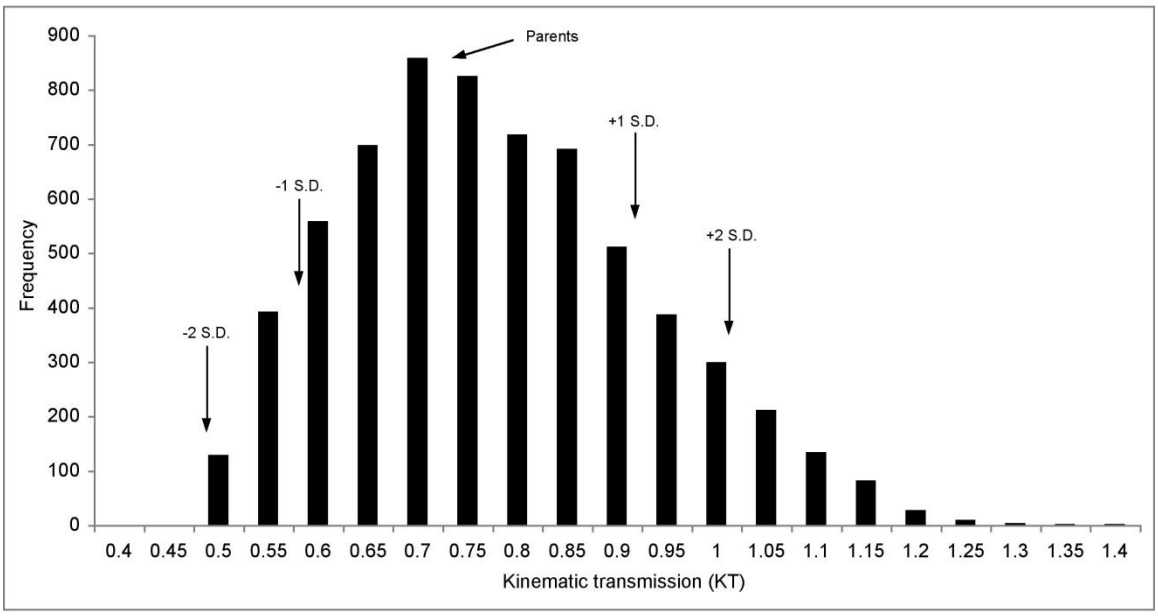

Figure 3.4. The distribution of KT values for simulated F2 hybrids from the intercross of Cynotilapia afra and Pseudotropheus elongatus highlights an appreciable frequency (29%) of individuals transgressive for function. Arrows indicate the KT of the parents and values ± 1 or 2 SD from the mean.

We used bubble plots to examine the genetic limits to transgression in each of the simulated crosses and the distribution of those limits among the four link lengths (Figure 3.5). These plots show the frequency of parental (F_0) alleles for each link in F_2 hybrids exhibiting transgressive KT, where hybrid KT is lower (A, C) or higher (B, D) than parental values. For example, in the cross between *Cynotilapia afra* and *Pseudotropheus elongatus*, transgressive KT lower than parental values (Figure 3.5A) is possible with nearly any combination of alleles for the nasal and fixed links. By constrast, transgression is observed in individuals with certain genotypic combinations of lower jaw (more *P. elongatus* alleles) and maxillary links (more *C. afra* alleles). Bubble plots for the cross between *Protomelas fenestratus* and *P. ornatus* are similar (i.e., negligible influence of nasal and fixed links, importance of lower jaw and maxillary links). These plots can be explained

by the starting morphology of the parents (F_0) and the noted relationship (from simulation and our empirical data above) between the input link (lower jaw), output link (maxilla) and KT. Both *Pseudotropheus elongatus* and *Protomelas fenestratus* have short maxillae and short lower jaws; *Cynotilapia afra* and *Protomelas ornatus* possess long elements of each. Recombination and independent assortment of alleles controlling lower jaw and maxillary links produce functional transgression in KT, according to this model.

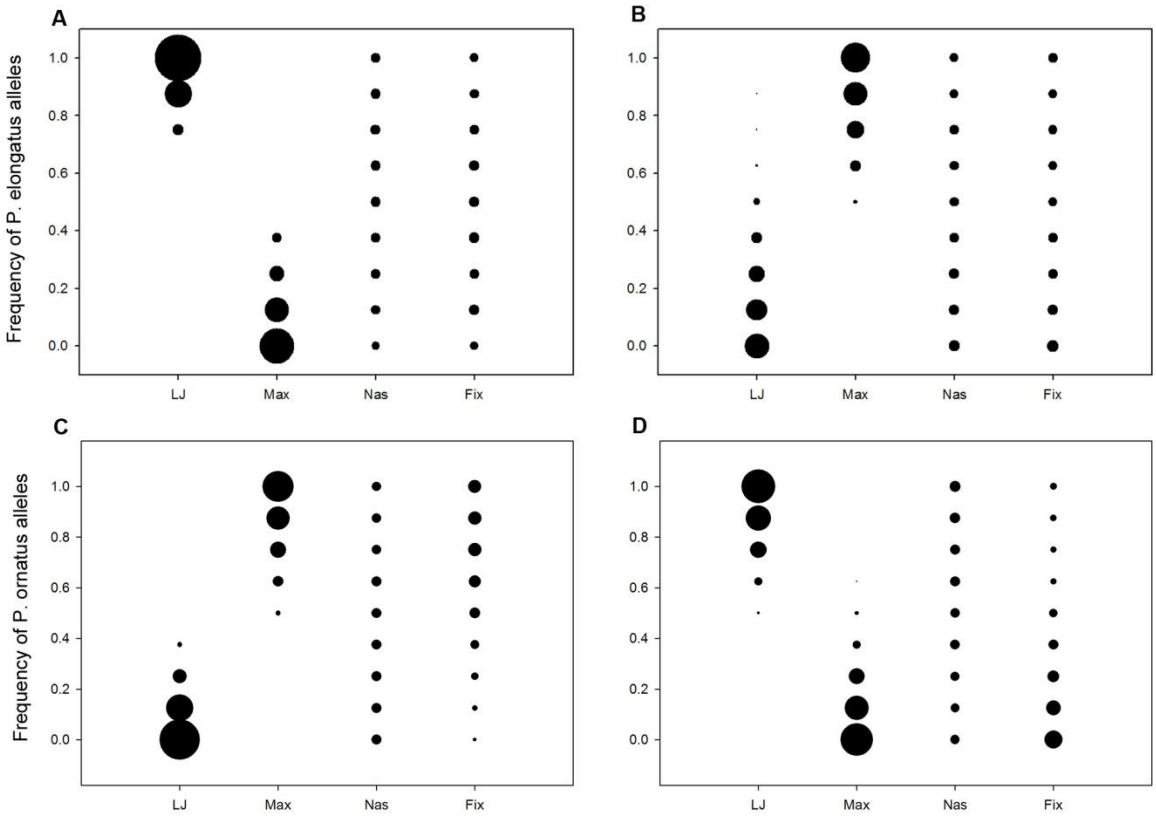

Figure 3.5. Bubble plots demonstrate the boundary conditions of transgression for Cynotilapia afra x Pseudotropheus elongatus (A, B) and Protomelas fenestratus x Protomelas ornatus (C, D). For both crosses, transgressive KT lower than (A, C) and higher than (B, D) the parents is dependent on allelic combinations of lower jaw and maxillary links.

Our choice of the 20 crosses to analyze presents a biased estimate of the proportion of crosses that would yield TS. However, because of the computation involved, a full analysis of all possible crosses in our data set is prohibitive. To overcome this, we realized that the simple ratio of lower jaw length to maxilla length is a strong predictor of KT from the fully parameterized calculation (Figure 3.3). We also realized, from the detailed analysis of the genetic model above, that transgression is usually produced by recombination between loci specifying the lengths of the lower jaw and maxilla. Combining these ideas, we devised a test to ask whether all pairwise crosses between any two of the 86 species in our data set would produce F_2 hybrids transgressive for KT. We generated a matrix encompassing all possible pair-wise combinations of size-adjusted lower jaw and maxilla lengths from all species in the data set. Roughly a third of these combinations (29%) produced transgression in KT value. We next used these data to ask whether the occurrence of transgression in F_2 could be predicted by the degree to which parents (F_0) are MTOM. As a proxy for MTOM, we used a simple metric of KT distance between the parents of any cross. Formally, if MTOM contributes to hybrid transgression in KT, we expect crosses that produce transgressive hybrids to be between parental species with smaller differences in starting KT. This is in fact what we observed (Figure 3.6; t-test, $p<0.001$).

3.4 Conclusion

We have described a general and likely pervasive mechanism that generates functional novelty. Simulated hybrid offspring among Lake Malawi cichlids exhibiting MTOM are transgressive for function at appreciable frequency.

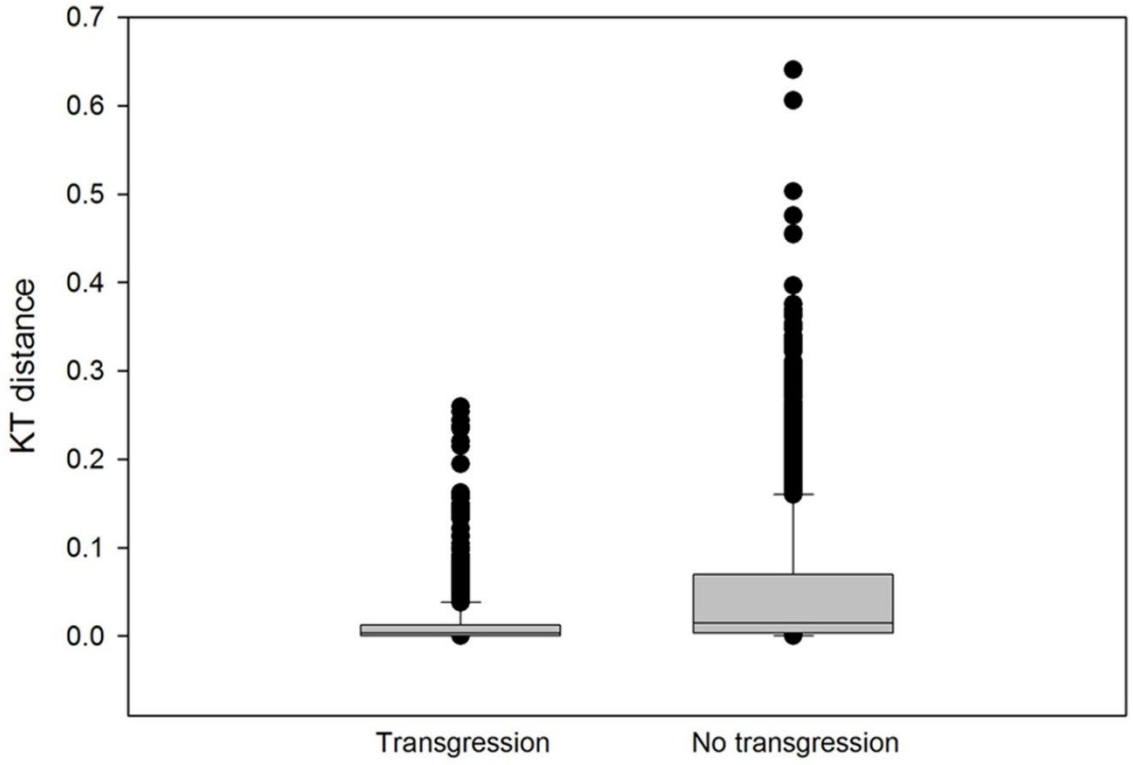

Figure 3.6. Box plots demonstrate a relationship between MTOM and transgression. Crosses that produce transgressive F_2 are between parental species with less difference in starting KT (a proxy for MTOM) than crosses that do not produce transgressive F_2 (t-test, $p<0.001$). The bar is the median value, the box is the 25th-75th percentile, whiskers are the 10th and 90th percentiles and the dots are outliers beyond the 5th and 95th percentiles.

Our modeling approach is noteworthy because it allows us to identify morphological components of the 4-bar mechanical system that contribute to TS. Functional transgression is a product of recombination and assortment between alleles controlling the lengths of the lower jaw and the maxilla and occurs in the absence of transgressive morphology. Transgression occurs when link lengths approaching parental values assort in new combinations in the F_2 generation. Novel function therefore is realized via new morphological combinations, but not transgressive lengths for any single link. Our model can be

tested by breeding Malawi cichlid hybrids in the lab and examining the morphological and mechanical diversity of their 4-bar linkages.

It is important to comment that the model we present shows how MTOM of form to function in the Malawi cichlid 4-bar linkage *could* generate TS in jaw biomechanics. One question raised by the modeling results is why *mbuna* species have not evolved higher maxillary KT (*mbuna* range=0.58 – 0.85). What maintains the positive phenotypic correlation between the lower jaw and maxilla links? One possibility is functional constraint imposed by other aspects of the feeding apparatus. For example, the lower jaw and maxilla both function to position the oral jaws for occlusion and contribute to the gape of the mouth. Mismatches in the length of upper and lower jaw elements may have negative functional ramifications that outweigh increased performance in jaw speed.

3.4.1 Recombinational evolution, MTOM and the limits to functional diversity

The genomic era has changed the way that evolutionary biologists think about hybridization. Gene flow among recently diverged species (or those evolved through pre-mating barriers) is common rather than the exception (Grant *et al.* 2004, Seehausen 2004, Won *et al.* 2005). Hybridization may provide an important source of genetic variation in the form of new trait combinations (Rieseberg *et al.* 2003) that sometimes help hybrid organisms maneuver across fitness landscapes (Mallet 2007). The phenomenon of TS, wherein hybrids outperform the parents, is a frequent observation in laboratory and natural outcrossings (Rieseberg *et al.* 1999b).

It has been suggested that strong directional selection and simple genetic architecture should limit TS because they point the phenotypic effects of QTL in the same direction; this would essentially constrain diversification by hybridization (Albertson and Kocher 2005). Our analysis shows how MTOM promotes TS in function, without transgressive morphology, via recombination and independent assortment. Because multiple morphological solutions exist for any function (i.e., value of KT), MTOM may buffer those constraints, imposed by directional selection, on the extent of mechanical diversity produced by hybridization. As the craniofacial skeleton evolves, functional demands of numerous mechanical systems shape the lengths and linkages of bony elements (Hulsey *et al.* 2005). MTOM of form to function in each of these systems might facilitate, or perhaps even require, the evolution of diversity.

3.5 Methods

3.5.1 Specimen information

Cichlid specimens were collected during July 2005 from various sites in southern Lake Malawi. Fishes were labeled with a unique identifier and fixed in 10% buffered formalin solution. Upon returning to the U.S., carcasses were transferred to 70% ethanol for storage. Additional specimens were borrowed from the American Museum of Natural History (AMNH) to supplement the collection. Up to three individuals of each species (depending on the number available) were cleared with trypsin and double-stained using alcian-blue (cartilage) and alizarin-red (bone). This method allows clear visualization of the

skeletal components while maintaining skull articulations (Hulsey and Garcia de Leon 2005). Cleared and stained fish were transferred to glycerin for storage.

3.5.2 Morphometrics and correlation

We measured the anterior jaw 4-bar linkage model as described (Hulsey and Wainwright 2002, Muller 1996, Westneat 1990). The four physical units of the linkage were quantified on 169 specimens of 86 Lake Malawi cichlid species. Up to three individuals were measured for each species using dial calipers (nearest 0.1 mm) to quantify each link (Figure 3.1). The four links (lower jaw or input, maxilla or output, nasal or coupler, and fixed or suspensorium) are described in detail, with their respective measurement landmarks, by Hulsey and Wainwright (2002). The linkage measurements were used to calculate the kinematic transmission (KT) of each jaw (Hulsey and Garcia de Leon 2005). Anterior jaw KT is used as an estimate of the speed of motion transmitted through the 4-bar linkage, and in addition, the measurement of KT simultaneously describes the force transmission (FT) through the 4-bar as the reciprocal of KT (Hulsey and Wainwright 2002). The method of KT calculation followed that of Hulsey and Garcia de León (2005) and was accomplished using an iterative function in Microsoft Excel® with a starting angle of 15^{o} and an input angle of 30^{o}.

Correlations were calculated between each of the 4-bar links and KT to examine the relationship therein. Because the structural components of the anterior jaw are highly correlated with body size (Hulsey and Wainwright 2002, NFP unpublished) each 4-bar link was corrected for fish standard length (SL) prior to correlation. We used the residuals of linear regression of link length to SL. Felsenstein (1985) has shown that correlations may not be statistically

independent when compared across taxa with common evolutionary history. Therefore, we calculated phylogenetically independent contrasts following an approach outlined in Hulsey *et al.* (2007). We constructed a phylogeny of 52 Lake Malawi species using sequences from the mitochondrial ND2 gene (see Figure B.2, Table B.1). All species were collected from the wild in Lake Malawi from various locations. DNA extractions, PCR parameters, and sequencing followed previously published protocols (Hulsey *et al.* 2007).

The ND2 gene was transformed into its three codon partitions. ModelTest 3.06 (Posada and Crandall 1998) was used to identify the most likely model of molecular evolution for each codon partition. Standard Bayesian analyses were executed to find the maximum log-likelihood topology using MrBayes 3.0 (Ronquist and Huelsenbeck 2003). Independent contrast analyses were performed using CAIC (Purvis and Rambaut 1995) in order to assess correlations among the SL-corrected residuals of the four link lengths and among those values and KT. We employed the highest log-likelihood tree from the above Bayesian search to correct for phylogeny (Figure B.2).

3.5.3 Genetic model of hybrid crosses

Several approaches can be taken to model the generation of novelty in jaw structures and for this project we chose a simple genetic model. Previous studies have examined how 4-bar function changes with different structural configurations, as well as how the 4-bar can evolve based on selection for different KT (Alfaro *et al.* 2004, Alfaro *et al.* 2005, Hulsey and Wainwright 2002, Wainwright *et al.* 2005). Also, Albertson and Kocher (2005) have examined the

phenotypic distribution of one interspecific Lake Malawi cross in terms of cranial and jaw morphology. For this paper we wanted a model that would not only describe the distribution of genotypes and phenotypes for specific intercrosses, but also allow the incorporation of empirical data from the Lake Malawi system.

We constructed a simple but realistic individual-based Mendelian genetic model, which incorporated assumptions of no mutation, no selection, and no error. We assumed that the size-standardized (below) length of each of the four anterior jaw links was controlled by 4 independent loci with 2 alleles at each locus, for a total of 8 alleles per link. Links themselves were modeled to be genetically independent of one another. The effect of each allele was assumed to be equal and additive, and therefore each of the four lengths (lower jaw, maxilla, nasal, fixed) for each parental (or F_0) specimen was divided by 8. Note that the assumptions of (i) number of loci per link, (ii) independence of links and (iii) additivity of allelic affects are supported by empirical quantitative genetic data for the cichlid craniofacial skeleton (Albertson *et al.* 2003). We controlled for differences in body size among species in the model by dividing each parental link length by SL. In this way, the heritable unit of link length can be thought of as the proportion of SL made up by each link. We also ran models using residuals from linear regression of link length on SL as input variables, and this did not change the outcome.

Using this rubric, each allele inherited from a parent would impart $1/8^{th}$ of the total length of that link to the offspring. Parents were assumed to be homozygous at all loci. The initial cross among F_0 produced F_1 progeny with

identical 4-bar lengths (and KT values) equal to the average of the two parents. A second cross within the F_1 progeny produced all potential parental and recombinant genotypes in the F_2 generation (6,561 possible). Hybrid link lengths were calculated by the summation of all allelic contributions totaled over the four loci per link. Thus, each composite F_2 genotype was assigned a 4-link phenotype, and KT was calculated for all. Some recombinant link lengths did not support a functional jaw linkage in some crosses. This result was taken to represent potentially unfit hybrid individuals within a cross and these individuals were left out of further analyses; this effect has been noted in prior simulation (Hulsey and Wainwright 2002). Frequency distributions of F_2 progeny were created (MS Excel® and SigmaPlot® 8.0) to compare hybrid KT values to parental KT. Bubble plots were generated (SigmaPlot® 8.0) to examine the contribution of parental alleles to link length in F_2 exhibiting transgressive KT. A total of 20 interspecific crosses were simulated with this model and F_2 KT values were examined for evidence of transgression. Species were chosen for these 20 iterations (i) if there is a report of hybridization between them in nature, (ii) if laboratory hybrids have been made, or (iii) if we were particularly interested in the cross from inspection of parental morphometrics.

These 20 simulated crosses present a biased look at the possibility of transgressive function in Lake Malawi cichlid hybrids. The calculations are iterative and time consuming. We wanted an unbiased means to ask if a hybrid cross between any two species would produce F_2 with transgressive KT. To accomplish this, we made use of the tight relationship between the fully parameterized model of KT and the simpler ratio of input to output links (Figure

3.3) and the relationship between input link, output link and KT observed in our data (see Results) and from previous simulation (Alfaro *et al.* 2004, Hulsey and Wainwright 2002). We generated a 86X86 species matrix containing all possible combinations of size-adjusted lower jaw and maxilla lengths from all species in the data set (7,310 combinations). We essentially asked if recombining the maxilla from one species with the lower jaw from the other (and vice versa) would produce hybrid linkages transgressive in KT. We carried out this operation for all pairs of species in the data set.

3.6 Acknowledgements

We thank members of the Streelman lab for comments on previous drafts of the manuscript. The research is supported by grants from the NSF (IOB 0546423) and the Alfred P. Sloan Foundation (BR 4499) to JTS.

3.7 References

Albertson RC, Kocher TD: Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution* 2005, 59:686-690.

Albertson RC, Streelman JT, Kocher TD, Yelick PC: Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proc. Natl. Acad. Sci. USA* 2005, 102:16287-16292.

Albertson RC, Streelman JT, Kocher TD: Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl. Acad. Sci. USA* 2003, 100:5252-5257.

Alfaro M E, Bolnick D I, Wainwright P C: Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in Labrid fishes. *Am. Nat.* 2005, 165:141-154.

Alfaro ME, Bolnick DI, Wainwright PC: Evolutionary dynamics of complex biomechanical systems: an example using the four-bar mechanism. *Evolution* 2004, 58:495-503.

Barton NH: The role of hybridization in evolution. *Mol. Ecol.* 2001, 10:551-568.

Burke JM, Arnold ML: Genetics and the fitness of hybrids. *Ann. Rev. Gen.* 2001, 35:31-52.

Ellstrand NC, Schierenbeck KA: Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc. Natl. Acad. Sci. USA* 2000, 97:7043-7050.

Felsenstein J: Phylogenies from gene-frequencies – a statistical problem. *Syst. Zool.* 1985, 34:300-311.

Gans C, Northcutt RG: Neural crest and the origin of vertebrates: a new head. *Science* 1983, 220:268-274.

Gerhart J, Kirschner M: *Cells, embryos, and evolution: Toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability.* Malden: Blackwell Science; 1997.

Grant P, Grant BR, Markert JA, Keller LF, Petren K: Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution* 2004, 58:1588-1599.

Grant PR, Grant BR: High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology* 1996, 77:500-509.

Grant PR, Grant BR: Hybridization of bird species. *Science* 1992, 256:193-197.

Hulsey CD, Fraser GJ, Streelman JT: Evolution and development of complex biomechanical systems: 300 million years of fish jaws. *Zebrafish* 2005, 2:243-257.

Hulsey CD, Garcia de Leon FJ: Cichlid jaw mechanics: linking morphology to feeding specialization. *Func. Ecol.* 2005, 19:487-494.

Hulsey CD, Mims MC, Streelman JT: Do constructional constraints influence cichlid craniofacial diversification? *Proc. Roy. Soc. Lond. B* 2007, 274:1867-1875.

Hulsey CD,Wainwright PC: Projecting mechanics into morphospace: disparity in the feeding system of labrid fishes. *Proc. Roy. Soc. Lond. B* 2002, 269:317-326.

Lawton-Rauh AL, Alvarez-Buylla ER, Purugganan MD: Molecular evolution of flower development. *Trends Ecol. Evol.* 2000, 15: 144-149.

Loh, Y.H.E., Katz, L.S., Mims, M.C., Kocher, T.D., Yi, S.V. & Streelman, J.T. (2008) Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Gen. Biol.*, 9, R113.

Lexer C, Welch ME, Raymond O, Rieseberg LH: The origin of ecological divergence in Helianthus paradoxus (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evolution* 2003, 57:1989-2000.

Mallet J: Hybrid speciation. *Nature* 2007, 446:279-283.

Muller M: A novel classification of planar four-bar linkages and its application to the mechanical analysis of animal systems. *Phil. Trans. Roy. Soc. Lond. B* 1996, 351:689-720.

Orr HA: Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 1998, 149:2099-2104.

Posada D, Crandall KA: MODELTEST: testing the model of DNA substitution. Bioinformatics 1998, 14:817-818.

Purvis A, Rambaut A: Comparative-analysis by independent contrasts (CAIC) – an Apple-Macintosh application for analyzing comparative data. Comp. *Appl. Biosci.* 1995, 11:247-251.

Raymond J, Segre D: The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 2006, 311: 1764-1767.

Rieseberg LH, Archer MA, Wayne RK: Transgressive segregation, adaptation and speciation. *Heredity* 1999b, 83:363-372.

Rieseberg LH, Raymond O, Rosenthal KM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzback AE, Donovan LA, Lexer C: Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 2003, 301:1211-1216.

Rieseberg LH, Whitton J, Gardner K: Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 1999a, 152:713-727.

Rieseberg LH, Widmer A, Arntz AM, Burke JM: Directional selection is the primary cause of phenotypic diversification. *Proc. Natl. Acad. Sci. USA* 2002, 99:12242-12245.

Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, 19:1572-1574.

Seehausen, O: Hybridization and adaptive radiation. *Trends Ecol. Evol.* 2004, 19:198-207.

Shubin N, Tabin C, Carroll S: Fossils, genes and the evolution of animal limbs. *Nature* 1997, 388:639-648.

Smith PF, Konings A, Kornfield I: Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. *Mol. Ecol.* 2003, 12:2497-2504.

Smith PF, Kornfield I: Phylogeography of Lake Malawi cichlids of the genus Pseudotropheus: significance of allopatric colour variation. *Proc. Roy. Soc. Lond. B* 2002, 269:2495-2502.

Streelman JT, Albertson RC: Evolution of novelty in cichlid dentition. *J. Exper. Zool. MDE* 2006, 306b:216-226.

Streelman JT, Gmyrek SL, Kidd MR, Kidd C, Robinson RL, Hert E, Ambali AJ, Kocher TD: Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Mol. Ecol.* 2004, 13:2471-2479.

Streelman JT, Peichel CL, Parichy DM: Developmental genetics of adaptation in fishes: the case for novelty. *Ann. Rev. Ecol. Evol. Syst.* 2007, 38:655-681.

Wainwright P C, Alfaro M E, Bolnick D I, Hulsey C D: Many-to-one mapping of form to function: a general principle in organismal design? *Int. Comp. Biol.* 2005, 45:256-262.

Wainwright PC, Bellwood DR, Westneat MW, Grubich JR, Hoey AS: A functional morphospace for the skull of labrid fishes: patterns of diversity in a complex biomechanical system. *Biol. J. Linn. Soc.* 2004, 82:1-25.

Westneat MW, Alfaro ME, Wainwright PC, Bellwood DR, Grubich JR, Fessler JL, Clements KD, Smith LL: Local phylogenetic divergence and global evolutionary convergence of skull function in reef fishes of the family Labridae. *Proc. Roy. Soc. Lond. B* 2005, 272:993-1000.

Westneat MW: Feeding mechanics of teleost fishes (Labridae, Perciformes) – a test of 4-bar linkages. J. morph. 1990, 205:269-295.

Won Y-J, Sivasundar A, Wang Y, Hey J: On the origin of Lake Malawi cichlid species. *Proc. Natl. Acad. Sci. USA* 2005, 102:6581-6586.

Won Y-J, Wang Y, Sivasundar A, Raincrow J, Hey J: Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Mol. Biol. Evol.* 2006, 23:828-837.

CHAPTER 4

THE GENETIC BASIS OF A COMPLEX FUNCTIONAL SYSTEM

4.1 Abstract

Contrasting simple and complex functional systems can provide insight into the processes of trait evolution. A prime example exists in the anterior jaws of many teleost fish whereby the lower jaw lever acts as a simple system and the anterior 4-bar linkage shows functional complexity in jaw kinematics. We have generated a large F_2 hybrid population by crossing two closely related Lake Malawi cichlid fish species with similar anterior jaw functions but different underlying morphology (many-to-one-mapping). A broad range of trait values is generated in these hybrids indicating transgressive segregation (extreme trait values) in morphological and functional traits but with stark differences between simple and complex biomechanical systems. QTL analysis identifies loci controlling jaw phenotypes in both systems with lower heritability for complex phenotypes. Positive and negative correlations exist between multiple traits as well as between simple and complex systems which may produce trade-offs in these functional and structural traits.

4.2 Introduction

The relationship between form and function has a profound effect on the dynamics of trait evolution. Many functional phenotypes are emergent properties of multiple interacting components and as such respond to evolutionary pressures in unique ways (Alfaro *et al.* 2004, Albertson and Kocher

2005). In complex functional systems manifold morphological combinations or configurations may produce identical functions. This mechanism, known as many-to-one-mapping (MTOM) is pervasive in biological systems and as such is expected to be a generality of organismal structure (Hulsey *et al.* 2005, Wainwright 2007).

There are several implications of MTOM for the generation of novelty and the evolution of phenotypes. Because a number of morphologies produce the same function there can be dramatic changes to structure without corresponding changes to the overall function. This can permit an expansion of an organism's overall morphospace without sacrificing the ancestral function. Also, a partial or complete decoupling of morphological and functional diversity through the redundancy of MTOM may lead to innovation in an adaptive functional trait whereby a novel function is produced from recombination of existing morphology (Parnell *et al.* 2008). Under the right conditions such innovations can lead to a shift into a new position of the adaptive landscape not previously realized (Alfaro *et al.* 2004, Hulsey *et al.* 2005, Wainwright 2007). In these ways the evolutionary pliability inherent to many-to-one-mapped traits is expected to play a major role in trait evolution and phenotypic diversity.

The anterior jaws of teleost fishes are excellent models of the contrast between simple and complex biomechanical systems and the differences between singular and plural mapping of form to function. A simple lever encompasses the lower jaw module in which movement transmitted through the

system (kinematic transmission or KT) is dictated by a ratio between the lengths of an output and an input lever (Figure 4.1). In this system one morphological

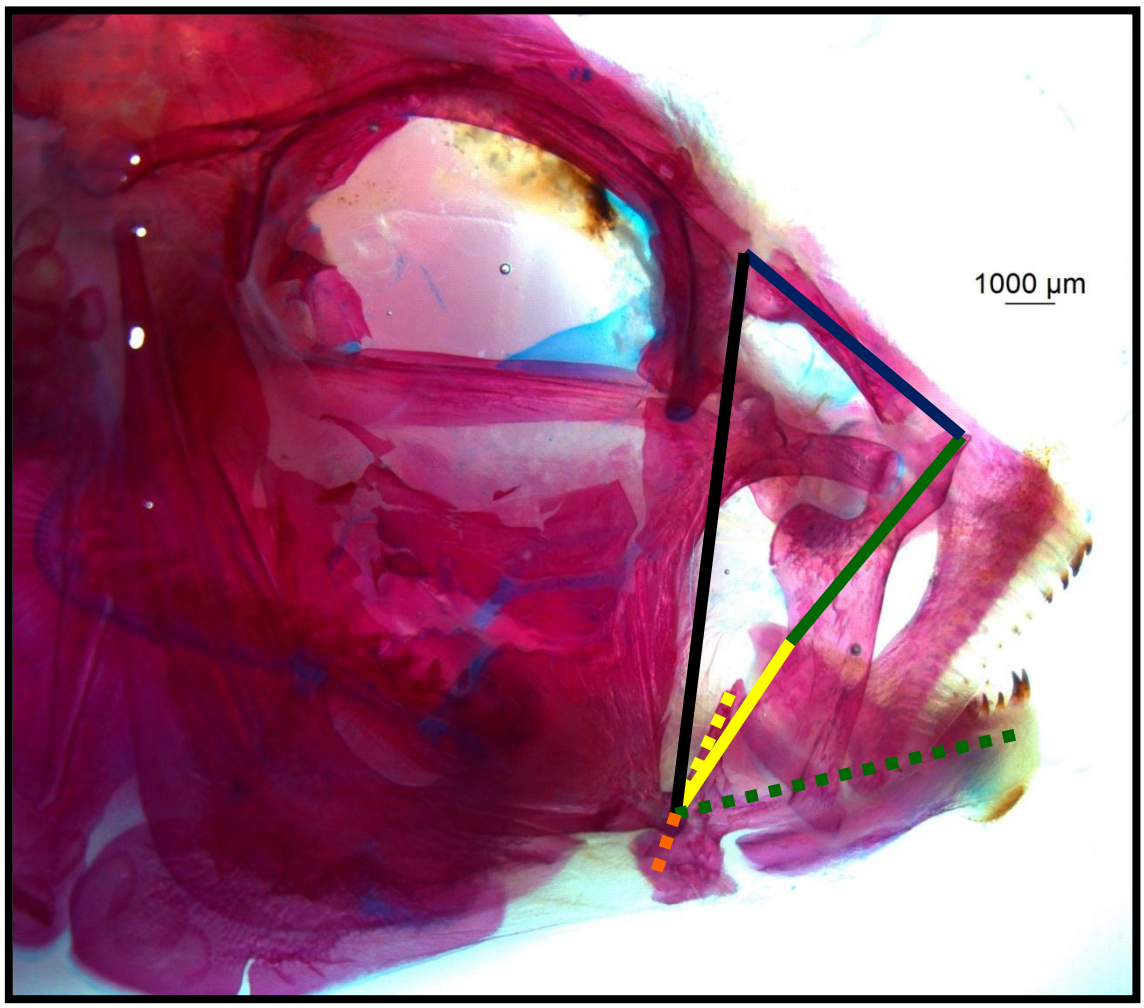

Figure 4.1. The structural components of the simple lever (dashed) and complex 4-bar linkage (solid) on a cleared and stained hybrid cichlid head. The simple lever is represented as: green dashed=out-lever (LJout); orange dashed=opening in-lever (LJin); yellow dashed=closing in-lever (LJincl). The 4-bar linkage is represented as: solid green=output (AJMax), solid yellow=input (AJLJ), solid blue=coupler (AJNas), solid black=suspensorium (AJFix).

combination (out-lever:in-lever) is mapped to a single functional value (one-to-one-mapping). The highly dynamic anterior jaw 4-bar linkage, also found in many teleost fishes, is distinct from the simple mechanics of the lever (Westneat 1990, Muller 1996). Function in the 4-bar is modeled as the rotation of the

maxillary link (output) given an input of lower jaw depression and constrained by the overall structural configuration (Figure 4.1). Four-bar jaw function can be quantified in terms of kinematic transmission (KT) by measuring the lengths of the four component links and computing the geometry of the system's mechanics (Hulsey and Wainwright 2002). Values of KT have been linked to ecologically relevant traits such as feeding modality and prey type. Higher values are associated with faster speeds, greater jaw protrusion and specialization on evasive prey while lower values denote slower speeds, greater force generation and specialization on durable, attached prey (Hulsey and Garcia de Leon 2005).

Simulations have revealed the behavior of the 4-bar linkage in response to selection for specific KT values, demonstrating the myriad solutions available with MTOM and the nonlinearity between morphology and function in this system (Alfaro *et al.* 2004). Recent theoretical work indicates that MTOM in the 4-bar coupled with a shuffling of morphological link lengths can produce transgressive segregation (extreme trait values; TS) in jaw function (KT) at appreciable frequencies without corresponding transgression in physical component links (Parnell *et al.* 2008). In the latter case the assumption of independent genetic control was made for each of the 4-bar structural traits based on empirical evidence (Albertson *et al.* 2003). With this genetic architecture, even in the presence of morphological and functional correlations genetic decoupling and independent assortment leads to transgression in function by recombination of the components of the complex system (Parnell *et al.* 2008). Contrarily, both simulation and experimental evidence indicate that these evolutionary

outcomes will be more constrained in the case of the simple jaw lever system in part due to a lack of MTOM (Alfaro *et al.* 2004, Albertson and Kocher 2005).

Traits that affect multiple aspects of organismal performance can be subject to functional trade-offs (Bennett and Lenski 2007, Wainwright 2007, Walker 2007). These trade-offs are the result of direct selection on one phenotype which produces correlated selection on related traits (Ghalambor *et al.* 2003, Walker 2007). For example, increasing the ratio of lower to upper hindlimb length in *Anolis* lizards affords greater jumping ability while concurrently decreasing sprinting speed (Bauwens *et al.* 1995, Toro *et al.* 2004). A similar effect is seen in the simple jaw lever as increasing the kinematics (speed) by lengthening the lower jaw will cause a decrease in the amount of force which is generated (Wainwright and Richard 1995). Albertson *et al.* (2005) have shown integration between lower jaw components as well as genetic correlations, both of which will facilitate such trade-offs. Complex traits, such as 4-bar KT, may experience fewer evolutionary trade-offs due to MTOM and the decoupling of morphology and mechanics. An increase in the number of interacting components can provide a greater number of morphological solutions for a given function and thus weaken the correlation between form and function (Alfaro *et al.* 2004). Recent work suggests that having more component traits for a given function may compensate for any trade-offs incurred during phenotypic changes as these complex functional systems realize mitigated performance trade-offs through compensatory changes in other related traits (Holzman *et al.* 2011).

In this paper we examine aspects of the evolution of structural and functional anterior jaw traits in F_2 hybrid Lake Malawi cichlids. We quantify the frequency of novel phenotypes in both simple and complex jaw systems and compare hybrid trait distributions with those found among the parental species and across this diverse and rapidly radiating group of fishes. We examine empirical evidence of MTOM operating in the complex 4-bar system as well as correlations between constituent morphological traits and trade-offs both within and between biomechanical systems. Using quantitative trait locus (QTL) mapping we discover and annotate the genomic regions controlling the morphological and functional phenotypes and examine the genetic basis of novel phenotypes resulting from interspecific hybridization. We test four specific hypotheses about the evolution of complex traits: (1) MTOM in the 4-bar can produce transgressive function without equivalent changes in component link lengths; (2) transgressive function will be the result of the recombination of existing link lengths in the 4-bar system (i.e. MTOM); (3) the complex 4-bar system will show higher functional transgression than the lever system; (4) the complexity of the 4-bar linkage and resulting variability in function will distill down to a linear component between the input and output links.

4.3 Methods

4.3.1 Production of F_2 mapping population

Two rock-dwelling (*mbuna*) cichlid species from Lake Malawi were chosen for interspecific hybridization based on previous work (Parnell *et al.* 2008). The dam (*Pseudotropheus elongatus*) was a member of an herbivorous species in

which female coloration approximates that found in the sire species (*Cynotilapia afra* Mara yellow), a primarily planktivorous feeder. These species are similar in body size, habitat utilization and geographic distribution, being found together at many locations in the lake (Ribbink *et al.* 1983). Although their primary prey differs in nature the complex functional kinematics of these species' jaw systems produce nearly identical measures of jaw speed, force and motion (Parnell *et al.* 2008). The principal of many-to-one-mapping (MTOM), whereby numerous physical arrangements produce the same functional characters (Wainwright *et al.* 2005), was integral to our choice of hybridizing species. The morphological links producing complex (4-bar) jaw function differ between species in that *P. elongatus* has short input and output lengths (AJLJ and AJMax, see below and Figure 4.1) while *C. afra* has longer lengths (Parnell *et al.* 2008). Incorporating these morphological differences (and functional similarities) in our experimental design we were able to test the hypothesis that MTOM can produce transgressive phenotypes during hybridization by recombination of existing morphology (short input/short output X long input/long output → short/long and long/short) without corresponding changes to the component structures. It should be noted that although our hybridization is lab-based these species co-occur at numerous sites throughout the lake in very close contact, with little spatial partitioning along shallow depth gradients (Parnell & Streelman 2011). These species are also some of several *mbuna* known to hybridize in the lake (Smith *et al.* 2003, Streelman *et al.* 2004, Genner and Turner 2005, Mims *et al.* 2010).

Several *C. afra* males were placed in a 189-L aquarium with *P. elongatus* females at a ratio of 1:2. Within a week the dominant male (and subsequent sire) had established an obvious hierarchy over his conspecifics and had fertilized the eggs of the dam which were retained in her mouth. Fin clips (25 mm^2 anal fin) were taken from both the dam and sire for RAD tag library preparation and the pair were kept in a breeding tank with dither fish to diffuse male aggression. Offspring were taken from the dam after hatching, grown up in net pens and moved to consecutively larger aquaria as they grew. This pair produced several F_1 broods which were maintained in 189-L aquaria for 2 years during which time they produced over 600 F_2 hybrid offspring. Each F_2 family was kept segregated from all other families and numbered by brood sequence (and individual). A total of 397 F_2 hybrids were euthanized with tricaine methane sulfonate (MS-222) and processed for DNA samples and a suite of morphological traits: intestine length, jaw morphometrics, jaw kinematics, tooth density and shape, taste bud density, sex and body color.

4.3.2 Morphometrics

All F_2 individuals were collected and sacrificed upon reaching approximately 75 mm total length (TL). Following euthanization with tricaine methane sulfonate (MS-222) fish were fixed in 10% buffered formalin for one week and then stored in 70% ethanol until processing (7-14 days). Individuals were cleared with trypsin and double-stained using alcian-blue (cartilage) and alizarin-red (bone). This method allows clear visualization of the skeletal components while maintaining bone articulations (Hulsey and Garcia de Leon

2005). Cleared and stained fish were transferred to glycerin (with thymol) for measurement and storage.

We measured and modeled the anterior jaw system as both a simple lever and a complex 4-bar linkage as described previously (Westneat 1990, Westneat 1994, Muller 1996, Hulsey and Wainwright 2002). The physical units of the levers and linkage were quantified on 349 hybrid offspring, both parental individuals, and 3 other individuals of each parental species. Each jaw unit was measured to the nearest 0.1 mm using dial calipers (see Figure 4.1). The simple lever components (opening input=in-lever (LJin), output=out-lever (LJout), closing input=closing in-lever (LJincl)), and the four links (input=lower jaw (AJLJ); output=maxilla (AJMax); coupler=nasal (AJNas); suspensorium=fixed (AJFix)) of the complex 4-bar system have been described in detail with their respective measurement landmarks (Westneat 1994, Hulsey and Wainwright 2002). Because the structural components of the anterior jaw are highly correlated with body size (Hulsey and Wainwright 2002) physical measurements were corrected as relative proportion of standard body length. Although the F_2 population was within a very narrow size range, and the parental species are very similar in adult size, some comparisons must be corrected for the relationship between body size and link length. For the sake of simplicity size-standardized measures were used for all analyses and calculations as the conversion does not change the quantification of function in any way. Size-standardized jaw measurements were used to calculate the kinematic transmission (KT) of each individual jaw in both the simple system (opening- and closing-KT) and the complex 4-bar as described (Westneat 1994, Hulsey and Garcia de Leon 2005). Anterior jaw KT is used as an

estimate of the speed of motion transmitted through the jaw system, while simultaneously describing the force transmission (FT) through the jaw as the reciprocal of KT (Hulsey and Wainwright 2002). High KT values are associated with fast moving (but weaker) jaws which are most effective capturing elusive prey, while low KT values are associated with slow (but forceful) jaws best suited to gripping and exerting force (Hulsey and Garcia de Leon 2005). Simple lever KT was calculated as the ratio of out-lever to the respective in-lever for opening and closing of the jaws (Westneat 1994). The method of 4-bar KT calculation followed that of Hulsey and Garcia de Leon (2005) whereby the four measured link lengths are used to estimate the length of a diagonal which bisects the 4-bar into two triangles. This diagonal is estimated along the distance between the AJLJ-AJFix articulation and the AJNas-AJMax articulation. Assuming a starting AJLJ-AJFix angle of 15° allowed the definition of all angles in the 4-bar by the law of cosines $\left[\cos\theta = \left(\frac{A^2+B^2-E^2}{2AB}\right)\right]$. An input angle of 30° was used to model the relevant rotation of the lower jaw allowing calculation of KT in the 4-bar as the output rotation of the maxilla.

4.3.3 DNA extraction and RAD library preparation

Genomic DNA was purified from sire and dam fin tissue using the DNeasy Blood & Tissue Kit (Qiagen) and a RAD library was prepared as described in Chapter 5.

4.3.4 Illumina Sequencing

The two RAD libraries were assembled and single nucleotide polymorphism (SNP) markers were detected and selected for genotyping in the intercross population as described in Chapter 5.

4.3.5 SNP genotyping in F2 individuals

A total of 382 F_2 individuals as well as sire and dam were genotyped at 384 SNP markers by the Emory Biomarker Service Center (Emory University). This method uses the Illumina BeadArray genotyping platform coupled with the GoldenGate assay in which oligonucleotide pool assays (OPAs) are designed specifically to discriminate between alleles at a SNP (Oliphant *et al.* 2002, Fan *et al.* 2003). Marker genotypes were examined to evaluate accuracy of initial sequencing during SNP discovery and 12 loci were removed as they did not segregate between parental lines. Another 12 markers that contained one heterozygous parental genotype were removed as well as 7 others that had low genotyping success in either parental or F_2 individuals. This left an informative set of 353 genotyped SNP markers in this cross.

4.3.6 Linkage map construction

A genetic linkage map was produced using genotype data from 382 individuals genotyped at 353 SNP markers with JoinMap® 3.0 software (van Ooijen & Voorrips 2001). The map was created using Kosambi's mapping function, a LOD threshold of 1.0, recombination threshold of 0.4, a jump threshold of 5.0, and a ripple function with no fixed order of loci. A LOD threshold of 4.0 was used to join 344 loci in 22 linkage groups with a total map size of 1082 cM and average marker distance 3.67±1.71 cM (see Figure C.1). The linkage map

generated from this cross was translated to the newly assembled tilapia genome as well as the *Metriaclima zebra* (a rock-dwelling "*mbuna*" species) genome by blasting SNP containing flanking sequences against assembled scaffolds (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html).

4.3.7 QTL mapping

The constructed linkage map was used to determine QTL locations for jaw morphology and kinematics in the F_2 population using the R/qtl package (Broman and Sen 2009). A total of 315 individual F2 were included in the analysis after removal of those with incomplete morphological and/or gender data. The size-standardized phenotypes included the simple lever (LJin, LJout, LJincl) and 4-bar linkage components (AJLJ, AJMax, AJNas, and AJFix) as well as function for simple lever opening and closing kinematic transmission (KT) and the complex 4-bar system (see Figure 4.1). We used an iterative approach by scanning for single QTL using standard and composite interval mapping (CIM), followed by two-dimensional scans to identify pairwise QTL interactions. Using the results of the previous steps we built multiple QTL models (MQM) incorporating QTL interactions and covariates. In the MQM process we used a forward-backward selection algorithm to add and remove QTL based on overall model effects and the effects of individual. Our dataset had nearly complete marker genotypes and thus we used the multiple imputation mapping method. Genotype-phenotype associations are scored using the logarithm of the odds (LOD) which represents the $\log_{10}$ likelihood ratio comparing the hypothesis of a QTL at a marker location to the null hypothesis of no QTL ($LOD=(n/2)\log_{10}(RSS_0/RSS_1)$; Broman and Sen

2009). The variance in a phenotype is assigned to each significant QTL and reported as percent variance explained (PVE) in the analysis output. The variance accounted for by a QTL is a proxy for the heritability of a trait and is calculated as $1\text{-}10^{-(2/n)LOD}$ (Broman and Sen 2009). Significance thresholds for LOD scores were estimated using 1000 permutations of phenotypes (randomized) relative to genotypes (constrained) to build a distribution of maximum genome-wide LOD scores. From this distribution the 95th percentile for LOD scores could be calculated to serve as a threshold for significant QTL associations (Broman and Sen 2009).

4.4 Results

4.4.1 Distribution of F_2 phenotypes

Anterior jaw phenotypes ranged widely in the F_2 population, outside of the expectation for intraspecific variation (and beyond parental distributions) and in some cases more similar to the distribution found across LM species (Table 4.1 and Figure 4.2; Parnell *et al.* 2008). Within this variation were unique patterns between structural and functional phenotypes, and between the simple and complex jaw systems.

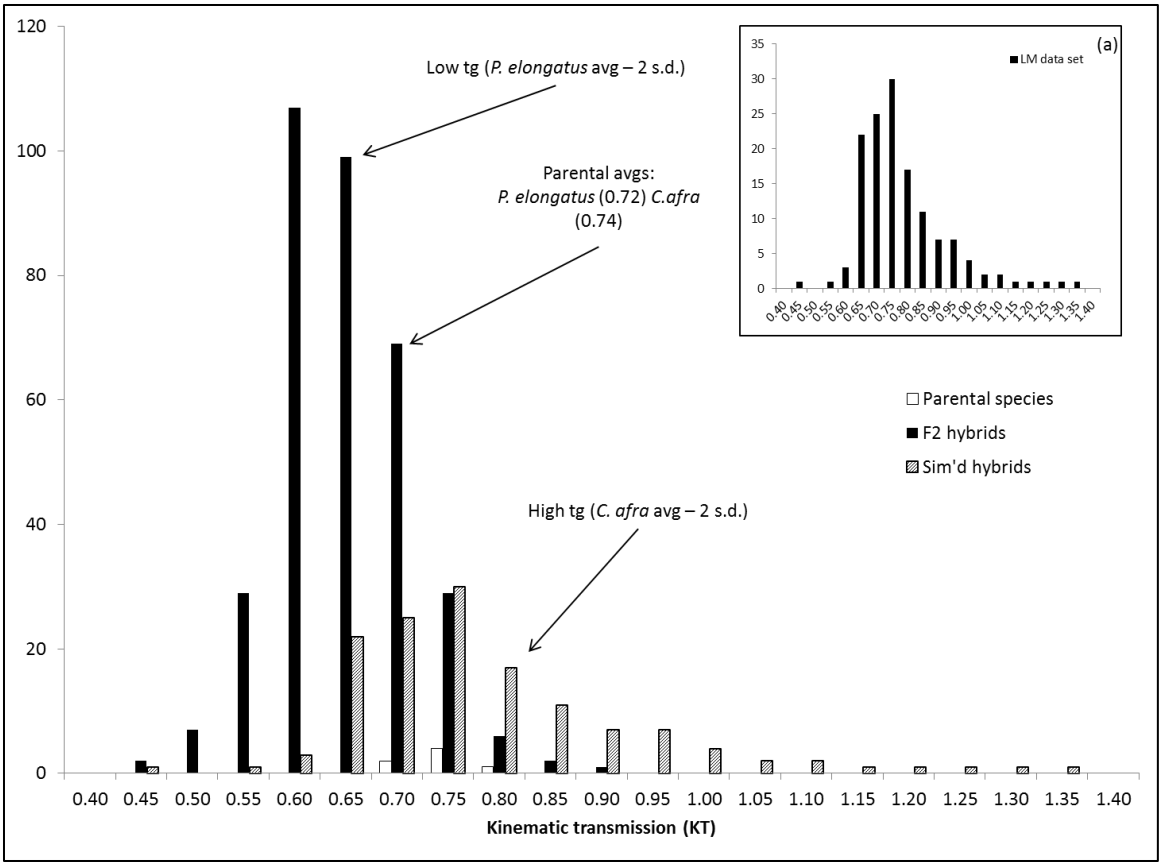

Figure 4.2. Distribution of 4-bar kinematic transmission (KT) in F_2 hybrids of *C. afra* and *P. elongatus*. Main plot represents F_2 (black bars) and parental species values (white bars) showing parental averages and ±1 and 2 standard deviations as thresholds for transgressive trait values (see Methods); right inset (a) represents expected distribution of 86 Lake Malawi cichlid species (Parnell *et al.* 2008).

Table 4.1. Anterior jaw trait values for F_2 hybrid cichlids and parental species (*C. afra* and *P. elongatus*). Mean, standard deviation, maximum and minimum values are included as well as calculated threshold for transgressive trait values (High/Low tg threshold; see Methods). Transgressive F_2 values are noted with bold font and underline if they range above (double underline) or below (single underline) the thresholds set from the parental data. Data includes measures of simple lever structure (opening in-lever=LJin; out-lever=LJout; closing in-lever=LJ incl) and function (closing kinematic transmission=Close KT; opening KT=Open KT) as well as complex 4-bar linkage structure (input=AJLJ; output=AJMax; coupler=AJNas; suspensorium=AJFix) and function (kinematic transmission=4-bar KT).

	Simple lever					4-bar linkage				
	LJ in	LJ out	LJ incl	Close KT	Open KT	AJLJ	AJMax	AJNas	AJFixed	4-bar KT
LM species set										
-mean	2.75	11.61	4.40	2.71	4.31	5.12	6.96	7.63	13.10	0.77
-standard deviation	0.54	2.62	0.64	0.39	1.08	0.94	1.27	1.15	1.44	0.16
-maximum	4.65	19.83	6.53	3.76	8.33	8.37	10.76	11.20	18.38	1.59
-minimum	1.67	6.30	3.00	2.03	2.28	3.24	4.45	5.03	10.16	0.41
C. afra										
-mean	2.64	10.88	4.76	2.30	4.12	5.33	7.23	5.81	13.36	0.74
-standard deviation	0.27	0.99	0.44	0.31	0.22	0.41	0.22	0.69	0.68	0.04
-maximum	2.98	11.79	5.31	2.57	4.50	5.87	7.55	6.70	13.99	0.78
-minimum	2.27	9.24	4.17	1.84	3.95	4.91	7.05	5.03	12.50	0.69
P. elongatus										
-mean	3.36	9.00	4.01	2.33	2.66	3.10	4.33	5.29	10.59	0.71
-standard deviation	0.31	0.68	0.40	0.28	0.35	0.26	0.25	0.26	0.68	0.02
-maximum	3.73	10.01	4.36	2.58	3.16	3.40	4.60	5.50	11.20	0.74
-minimum	2.97	8.56	3.44	2.07	2.36	2.90	4.10	5.00	9.86	0.70
High tg threshold	3.99	12.87	5.64	2.88	4.56	6.05	7.73	6.98	14.67	0.80
Low tg threshold	2.11	7.64	3.22	1.68	1.96	1.85	2.10	3.75	8.89	0.62
F_2										
-mean	2.61	9.17	3.55	2.62	3.61	3.96	6.46	5.13	11.88	0.62
-standard deviation	0.47	0.58	0.45	0.32	0.58	0.35	0.56	0.63	0.72	0.06
-maximum	**4.13**	11.05	5.50	**4.14**	**5.33**	5.14	**8.64**	**8.04**	**15.43**	**0.87**
-minimum	**1.64**	**7.50**	**1.97**	1.78	2.13	2.65	4.56	**3.38**	**8.09**	**0.45**
Proportion tg high	0.01	0.00	0.00	0.17	0.07	0.00	0.02	0.01	0.00	0.01
Proportion tg low	0.13	0.00	0.21	0.00	0.00	0.00	0.00	0.01	0.00	0.50
Fxn:strx tg ratio				0.80	0.48					13.50

To quantify novel phenotypes produced in this cross we employed a conservative threshold of transressive segregation (TS; extreme or novel trait values). It is expected that hybridization will produce phenotypes intermediate to the parents and therefore TS has been described as any trait value outside that of the parental species mean (Rieseberg *et al.* 1999). We have included parental

species distributions and estimates of intraspecific variation across the LM cichlid species flock to produce a conservative yet realistic definition of transgressive trait values. We use the highest and lowest parental species mean ±2 standard deviations relative to each trait as the threshold values for transgression (e.g. see Figure 4.2). This definition permits more phenotypic variation in the "normal" F_2 distribution than that found within any LM species and is a more stringent threshold than pooling parental species values.

4.4.2 Transgressive phenotypes in the anterior jaw systems

Transgressive phenotypes were observed in all but one of the structural and functional anterior jaw traits (Figure 4.1&4.2). The mean 4-bar KT of the F_2 (0.620±0.063) was significantly lower than that of the parents (*C. afra*=0.737±0.038; *P. elongatus*=0.715±0.022) with a much broader range (0.446-0.869; see Table 4.1). The distribution of 4-bar KT spanned large portions of a simulated data set as well as that measured across LM species (Figure 4.2; Parnell *et al.* 2008). Over 50% of the hybrid offspring were transgressive for 4-bar KT, the vast majority of which were lower than that of the parents. We observed limited (0-1.7%) transgressive component links of the 4-bar anterior jaw system as compared to the rampant transgression in function (Table 4.1). Of the four links in this complex jaw system the AJLJ showed no extreme values, the AJMax exhibited 1.7% above the parents and the AJNas and AJFix lengths varied outside the parental distributions in less than 1% of the offspring (Table 4.1). Between groups of functionally (KT) transgressive and non-transgressive

individuals only the input (AJLJ) and output (AJMax) 4-bar links differed significantly in length (t-test, $p<10^{-10}$).

Simple lever jaw kinematics (i.e. KT) are characterized by dividing the out-lever length (output) by the in-lever (input). Although the 4-bar linkage system is more complex and dynamic than a simple lever system there exists a clear pattern in the ratio of the AJLJ (input) and AJMax (output) links as a strong correlate with KT (r=0.9754; Figure 4.3). Although this proxy does not fully describe the kinematic properties of this complex system it is a very accurate predictor of KT as calculated with the fully parameterized 4-bar model. This ratio indicates a very strong opposing effect of the AJLJ (input) and AJMax (output) link lengths on 4-bar KT and identifies them as key components in this complex functional trait (Figure 4.4). Similar findings are reported among simulated hybrids of various LM species as well as across the LM cichlid species flock (Figure 4.3 and 4.4 insets; Parnell *et al.* 2008).

The size-standardized physical lengths of the simple lever system varied beyond the parental distributions (Table 4.1). A very low frequency of transgressive out-levers (LJout; 0.03%) were observed while each of the in-levers (LJin and LJincl) had more extreme values. For each in-lever the majority of lengths were shorter than the parentals (LJin=12.8% and LJincl=21.4%) although the opening

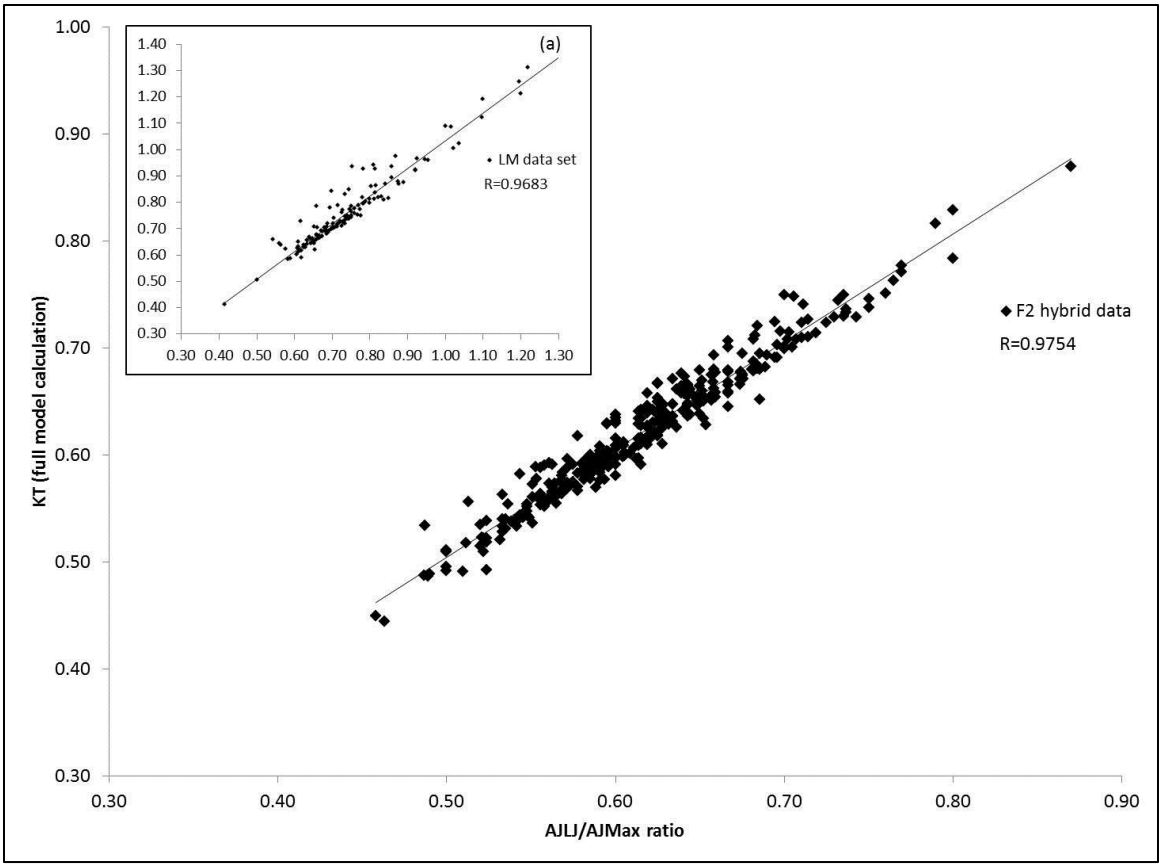

Figure 4.3. A simple ratio of 4-bar input (AJLJ) to output (AJMax) is strongly correlated with the calculated kinematic transmission (KT) from the fully parameterized model. Main plot represents F_2 hybrid data; inset (a) includes data from empirical set of 86 Lake Malawi cichlid species (Parnell *et al.* 2008).

in-lever had a small proportion (0.09%) longer than expected. Extreme simple lever functional values were seen in direct proportion to the magnitude of transgression in morphology. Jaw-closing KT was higher in 17.4% of individuals, opening-KT was higher in 6.6%, but neither function realized transgressive low values (Table 4.1). A ratio of functional (fxn) to morphological (strx) diversity clearly contrasts the evolutionary lability of the complex 4-bar (transgressive fxn:strx=13.50) with the constraints on the simple lever systems (closing=0.80 and opening=0.47; Table 4.1).

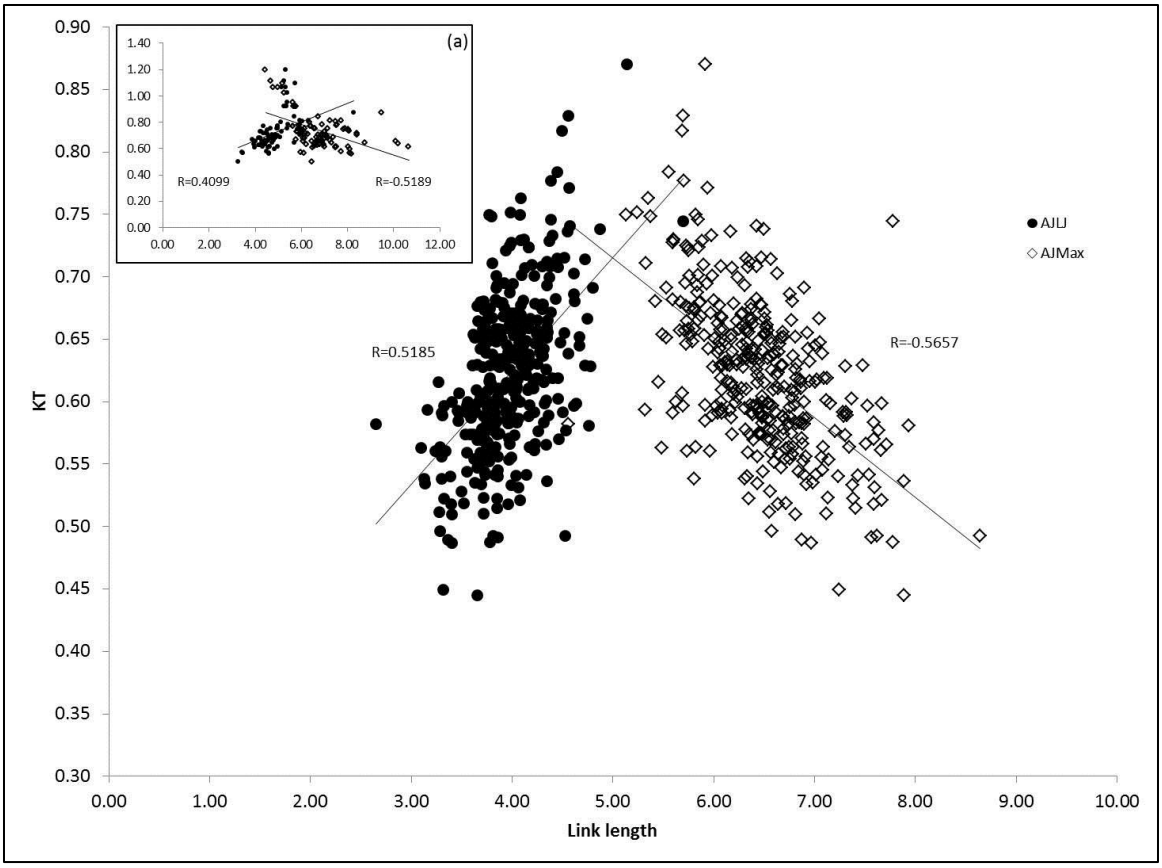

Figure 4.4. The input (AJLJ) and output (AJMax) links of the 4-bar are antagonistically correlated with kinematic transmission (KT). Main plot represents F_2 hybrid data; inset (a) includes data from 86 Lake Malawi species (Parnell *et al.* 2008).

4.4.3 Genetic basis of anterior jaw structure and function

4.4.3.1 Complex 4-bar linkage

Simultaneously scanning for two QTL revealed several genomic regions associated with the length of each of the component links in the 4-bar system (Table 4.2 and Figure C.2). Incorporating these loci into separate multiple QTL models (MQM) for each trait provided well supported estimates of model significance (LOD scores) and genetic variance explained (PVE). The loci associated with the input link (AJLJ) produced a two locus final model including

Table 4.2. Results of multiple QTL mapping (MQM) for loci controlling complex jaw structure and function in Lake Malawi hybrid cross. The 4-bar components and kinematic transmissions (KT) are included as phenotypes. The log-likelihood score (LOD), percent variance explained (PVE) and the p-value for the full model are indicated in the top row for each trait. The lower rows represent the results of dropping a putative QTL from the model one at a time. Putative QTL are indicated as chromosome@position in centimorgans (cM), "sex" represents inclusion of gender as a cofactor, and epistatic interactions are indicated by factor:factor notation. Significant results are indicated as p<0.05 (*), p<0.005 (**) and p<0.0005 (***).

AJLJ		df	LOD	PVE	P	Sig
	Full model	8	4.925	6.947	0.0045	**
	Drop one QTL:					
	7@35.8	6	4.521	6.358	0.0024	**
	18@14.0	6	3.723	5.206	0.0102	*
	7@35.8:18@14.0	4	3.334	4.648	0.0048	**
AJMax		df	LOD	PVE	P	Sig
	Full model	12	11.143	15.033	<0.0001	***
	Drop one QTL:					
	8@27.2	6	5.904	7.660	0.0002	***
	19@0.0	6	4.763	6.127	0.0017	***
	7@43.8	2	4.210	5.393	0.0001	***
	7@44.2	2	4.079	5.222	0.0001	***
	8@27.2:19@0.0	4	3.986	5.098	0.0014	***
AJNas		df	LOD	PVE	P	Sig
	Full model	8	6.234	8.711	0.0004	***
	Drop one QTL:					
	14@54.3	6	5.920	8.253	0.0002	***
	7@31.8	6	5.712	7.951	0.0003	***
	14@54.3:7@31.8	4	5.338	7.409	0.0001	***
AJFix		df	LOD	PVE	P	Sig
	Full model	5	12.059	16.164	<0.0001	***
	Drop one QTL:					
	7@43.8	2	7.012	9.050	<0.0001	***
	7@44.2	2	6.963	8.984	<0.0001	***
	sex	1	2.840	3.554	0.0003	***
4-bar KT		df	LOD	PVE	P	Sig
	Full model	8	6.031	8.440	0.0006	***
	Drop one QTL:					
	8@27.2	6	6.008	8.406	0.0001	***
	13@0.0	6	5.616	7.835	0.0003	***
	13@0.0:8@27.2	4	5.598	7.809	<0.0001	***

epistasis (LOD=4.93, PVE=6.96). The output length of the 4-bar linkage (AJMax) was associated with a total of four QTL including one epistatic interaction

(LOD=11.14, PVE=15.03). This model included two closely linked QTL sharing a chromosome (chr 7) with one of the input (AJLJ) loci but 8 centimorgans (cM) distal to it. The largest effect QTL for the AJMax was located on chromosome 8 at position 27.2 (8@27.2) and accounted for an estimated 7.7% of the variance in this trait. The coupler link (AJNas) was best explained by two QTL with epistasis (LOD=6.23) accounting for an estimated 8.71% of the phenotypic variance in this trait. This link shares a chromosome with the input (AJLJ) and output (AJMax) links (chr 7) but at a position 4 cM proximal to the former. The two QTL model best fit to the suspensorium (AJFix) link of the 4-bar included two of the loci associated with the AJMax (7@43.8 and 44.2). This model explained the largest proportion of phenotypic variance for all of the 4-bar link lengths (PVE=16.16, LOD=12.06) and was the only link found to have a significant variance effect of sex (3.55%) whereby female fixed lengths were longer.

Function of the 4-bar linkage (kinematic transmission, KT) was associated with two QTL exhibiting epistasis and accounting for 8.44% of the variance (LOD=6.03; Table 4.2). One of the QTL was found on a chromosome which was not detected in analyses for any of the constituent 4-bar link lengths (13@0.0). The largest effect QTL for this functional trait mapped to the exact location of the largest effect QTL for the maxilla link (AJMax; 8@27.2). As described above the AJMax is one of the key structural lengths correlated with function (KT) in the 4-bar system, and no other 4-bar links mapped to a locus shared with KT.

4.4.3.2 Simple jaw lever

Single and double QTL scans revealed marginal associations for simple lever traits and models with multiple QTLs and interactions revealed strong support for several loci and epistatic effects (Table 4.3). The full model for the opening in-lever (LJin) included eight loci on six chromosomes with several interactions. Together these QTL explained over 40% of the variance (PVE) in this trait with strong statistical support (LOD=36.59). Two QTL in the LJin model were closely associated with simple lever opening-KT (LM22@0.8 and 20@7.4). The model for opening-KT also included an interaction between these loci and explained a significant amount of variance in this functional trait (LOD=8.02, PVE=11.07; Table 4.3).

The genetic association for the out-lever (LJout) of the simple jaw system was best explained by a model including four QTL and three interactions (LOD=24.76, PVE=30.37; Table 4.3). The LJout was the only simple lever phenotype to exhibit a significant effect of sex (longer in females), which accounted for over 5% of the variance in this trait. The function of jaw-closing KT shared a chromosome with the LJout at a QTL located over 20 cM distal. A second QTL for closing-KT (17@7.9) was located on a chromosome also containing a QTL (~10 cM distal) for the closing in-lever (LJinc). Three other QTL and two interactions were significantly associated with LJinc accounting for over 18% of the variance in this trait (LOD=13.64; Table 4.3).

Table 4.3. Results of multiple QTL mapping (MQM) for loci controlling simple jaw structure and function in Lake Malawi hybrid cross. The simple lever components and kinematic transmissions (KT) are included as phenotypes. Column notations and abbreviations are as in Table 4.2.

LJ in		df	LOD	PVE	P	Sig
	Full model	40	36.59	41.43	<0.0001	***
	Drop one QTL:					
	15@3.7	10	13.90	13.20	<0.0001	***
	LM22@0.8	14	11.95	11.18	<0.0001	***
	5@26.6	10	11.61	10.84	<0.0001	***
	10@45.0	6	10.57	9.79	<0.0001	***
	8@27.2	6	8.79	8.03	<0.0001	***
	10@60.5	6	7.32	6.62	<0.0001	***
	15@14.0	6	6.25	5.60	0.0003	***
	20@7.4	6	5.19	4.61	0.0019	**
	5@26.6:10@45.0	4	9.44	8.67	<0.0001	***
	5@26.6:10@60.5	4	7.15	6.46	<0.0001	***
	15@3.7:LM22@0.8	4	4.55	4.03	0.0011	**
	15@3.7:8@27.2	4	4.11	3.63	0.0024	**
	LM22@0.8:15@14.0	4	4.41	3.90	0.0014	**
	LM22@0.8:20@7.4	4	4.34	3.84	0.0016	**
LJ out		df	LOD	PVE	P	Sig
	Full model	21	24.76	30.37	<0.0001	***
	Drop one QTL:					
	8@0.0	10	11.03	12.18	<0.0001	***
	13@43.0	6	9.17	9.99	<0.0001	***
	2@10.7	10	8.80	9.56	<0.0001	***
	12@49.9	6	5.73	6.08	0.0004	***
	sex	1	5.35	5.66	<0.0001	***
	8@0.0:13@43.0	4	7.67	8.26	<0.0001	***
	2@10.7:12@49.9	4	3.28	3.42	0.0070	**
	2@10.7:8@0.0	4	3.48	3.63	0.0048	**
LJ incl		df	LOD	PVE	P	Sig
	Full model	16	13.64	18.07	<0.0001	***
	Drop one QTL:					
	2@44.6	6	7.34	9.28	<0.0001	***
	21@17.7	6	6.84	8.61	<0.0001	***
	17@38.7	6	5.95	7.45	0.0002	***
	18@14.0	6	5.91	7.39	0.0002	***
	2@44.6:21@17.7	4	6.45	8.10	<0.0001	***
	17@38.7:18@14.0	4	5.27	6.56	0.0001	***
Close KT		df	LOD	PVE	P	Sig
	Full model	8	6.88	9.57	0.0001	***
	Drop one QTL:					
	2@32.7	6	6.78	9.43	<0.0001	***
	17@7.9	6	5.69	7.85	0.0003	***
	2@32.7:17@7.9	4	5.49	7.56	0.0001	***
Open KT		df	LOD	PVE	P	Sig
	Full model	8	8.02	11.07	<0.0001	***
	Drop one QTL:					
	20@7.4	6	7.45	10.23	<0.0001	***
	LM22@0.8	6	6.45	8.79	0.0001	***
	20@7.4:LM22@0.8	4	6.05	8.23	<0.0001	***

4.5 Discussion

In this study we have presented empirical evidence for the production of novel structural and functional phenotypes in biomechanical systems as a direct outcome of species hybridization. Within this diversity of phenotypes exist stark patterns that contrast important attributes of complex systems and lend support to hypotheses concerning the dynamics and functional trade-offs of trait evolution. We have identified transgressive function in the absence of transgressive morphology in the complex 4-bar but no such pattern in the simple lever system. Functional 4-bar transgression is the result of recombinations of the component links thereby changing function but not form. Although some extreme trait values exist in the simple lever system the complex system produced two orders of magnitude more functional diversity relative to morphological diversity. The complexity and MTOM of the 4-bar jaw system is contrasted by a strong linear component between two of the link lengths and KT. We find low heritability (proxy of PVE in these analyses) in the traits associated with the complex jaw system and relatively higher heritability for jaw levers and simple function. Finally, our analysis has detected genetic and morphological correlations between and within biomechanical systems which may act to both constrain and facilitate trait evolution.

4.5.1 MTOM produces diversity in structure and function

The complex functional phenotype (4-bar KT) produced in this cross ranged beyond the distributions seen in either of the hybridizing species. In fact the hybrid 4-bar KT distribution extends beyond that found throughout the rock-

dwelling clade that the hybridizing species are members of (*mbuna* KT range=0.58-0.85) and spans a significant portion of that found across all LM cichlid species (see Figure 4.2). Considering these trait distributions in the context of those within other species rich groups of fishes (e.g. Labridae) we see that the hybrid functional phenotypes encompass the variation found across millions of years of divergence, yet they reproduce naturally occurring KT values found in LM cichlids (Westneat 1995, Wainwright *et al.* 2004, Alfaro *et al.* 2005, Parnell *et al.* 2008). This outcome is in direct support of recent simulations of species hybridizations producing rampant functional transgressive segregation in 4-bar KT (Parnell *et al.* 2008).

Previous work has predicted that complex biomechanics may show extreme values while the constituent components do not (Alfaro *et al.* 2004, Parnell *et al.* 2008). The complexity and MTOM of the 4-bar system permit an increase in functional diversity in the absence of changes to morphology (Table 4.1). Less than 4% of all 4-bar links were outside the parental distributions yet over 51% of individuals had transgressive 4-bar function. Presented as a ratio of functional to structural transgression this represents a 13-fold difference (fxn:strx=13.50). Conversely, changes in simple lever structure produced corresponding magnitudes of change in function (fxn:strx<1.0) indicative of one-to-one-mapping (Alfaro *et al.* 2004). Although gross structural transgression was greater in the simple lever system the complex 4-bar showed a much higher magnitude of functional transgression in the absence of morphological component changes. These empirical results exemplify the effect of the redundancy and decoupling of morphology and mechanics which are inherent

to MTOM. The diversity in function without corresponding extreme morphology and difference of key link lengths between groups (TS vs. non-TS for KT) clearly indicates the recombination of 4-bar system components as the mechanism responsible for generating novel performance values.

4.5.2 Limited transgression occurs in the simple lever system

The simple lever system is known to exhibit morphological and genetic integration, correlations between structural components and linear (or one-to-one) form to function mapping (Alfaro *et al.* 2004, Albertson *et al.* 2005). Transgression in lower jaw KT is directly related to extreme lever lengths within the F_2 population (Table 4.1). Structural links of opening and closing in-levers were shorter than those found in the parents while only one individual showed transgression in out-lever. This is directly reflected in the functional phenotypes as all transgressive KT values were higher than expected (lever KT=out-lever/in-lever). However, the observation of extreme values in either morphology or KT of the simple jaw levers was unexpected. Albertson and Kocher (2005) reported that genetic architecture and past directional selection (Albertson *et al.* 2003) on the lower jaw system precluded it from producing transgression during hybridization. This argument is in line not only with their empirical results but also a meta-analysis of 96 published studies indicating that directional selection limits transgressive segregation due to a lack of antagonistic QTL (Rieseberg *et al.* 2003). While our results differ with theirs we do not suggest that this discounts the effect of directional selection on decreasing transgression. The species used in our cross were chosen specifically because they show MTOM in the 4-bar linkage

– nearly identical functions created by differing morphologies. The species used by Albertson and Kocher (2005) were chosen specifically because they had very different lower jaw lever systems. One of those species exhibits an extreme trophic specialization whereby the lower jaw is incredibly robust and powerful allowing it to scrape algae from the substrate much more efficiently than other species. In this case it is likely that directional selection has acted very strongly to produce such an adaptation, thereby limiting genetic variation in terms of antagonistic QTL. The species used in our study do not show extreme lower jaw lever phenotypes and therefore the assumption of strong historical directional selection here is not met and in fact we do see transgression in both form and function. Effect plots of the QTL detected for lower jaw traits indicate that our cross does segregate antagonistic alleles as evidenced by a lack of consistency in the direction of effects (e.g. Figure C.3).

4.5.3 Linearity within and between systems

The complex 4-bar system features some underlying linearity between components and function. The simple ratio of input (AJLJ) to output (AJMax) is found to be a near perfect predictor of the fully parameterized calculation of 4-bar KT (Figure 4.4; r=0.975). These links are positively correlated with each other (r=0.382) yet they have opposing effects on 4-bar function (AJLJ-KT r=0.518, AJMax-KT r=-0.566; Figure 4.3). These link lengths are critical to KT but are constrained by the overall geometry of the other two links (AJNas and AJFix). This linearity is at least partially due to genetic correlation between structure and function. Four-bar KT is associated with loci on two chromosomes but one of

those is precisely at a locus controlling the length of the output link (AJMax; 8@27.2). We found no QTL in common between the input link (AJLJ) and 4-bar KT and in fact no other loci controlling morphology associated with this function. The strong correlation between the input and output links is not due to genetic correlation (no shared loci) but is likely caused by the morphological association of correctly functioning biomechanical geometries – if these links are mismatched the jaws fail to work.

The simple and complex jaw systems do not share component structures; however, correlations exist between them. The length of the closing in-lever (LJinc) is strongly correlated with that of the 4-bar input (AJLJ) link (r=0.832; Figure C.4). Genetic correlation exists between these traits as loci controlling these structures are located within 1 cM on chromosome 18. Changes to the 4-bar input may impose correlated selection on the closing in-lever length and thus jaw closing KT as evidenced in the negative correlation between these functional traits (r=-0.320; Figure C.5). A reciprocal trade-off is not obligate in the 4-bar system due to the redundancy of MTOM and decoupling of form and function. The attributes of complexity serve to dissipate functional trade-offs, a mechanism not found in simple biomechanical systems (Wainwright 2007). In this case genetic linkage of structural jaw components between the simple and complex systems imposes an evolutionary constraint on function in only the simple lever (Holzman *et al.* 2011).

4.5.4 Complex systems and complex constraints

The distribution of traits in the F_2 suggests that a constraint may exist on the 4-bar, seen as the dearth of high transgressive KT values (Table 4.1). Nearly all of the extreme 4-bar values are low suggesting a limit to the potential link combinations. A similar limitation has been reported across the entire rock-dwelling *mbuna* clade based on empirical measurements and may be the result of yet undefined trade-offs impinging on the 4-bar linkage (Parnell *et al.* 2008). Unlike the lower jaw lever the 4-bar is composed of elements associated with both the upper and lower portions of the skull. The sheer number of morphological and functional traits dependent on the overall structure and geometry of the highly dynamic fish skull (Barel 1983, Carroll *et al.* 2004) facilitates the possibility that negative ramifications to another function may overwhelm any benefit gained by creating high KT 4-bar linkages. It is interesting to note that within the LM cichlids there do exist species with high 4-bar KT, but they are limited to the sister clade of the *mbuna*, which are associated with sand flats and open water feeding rather than rock reefs (Ribbink *et al.* 1983, Parnell *et al.* 2008). This delineation is at the base of the LM cichlid diversification and has been implicated as an early driver (divergent macrohabitats and trophic morphologies) of this adaptive radiation (Streelman and Danley 2003). The effect of MTOM to amplify evolutionary contingency (Alfaro *et al.* 2004, Wainwright 2007) may indicate this constraint is an ancestral limitation on 4-bar structure and function from the initial phylogenetic split of these sister clades thus precluding the generation of certain biomechanics.

The LM cichlid system has numerous unique attributes which make it an interesting case study in rapid adaptive evolution. An incredible diversity of

phenotypes, including trophic structure and function, have evolved in this group multiple times across this patchy and isolated island-like landscape. We have produced a realistic interspecific hybrid population between two rock-dwelling *mbuna*. These hybrids are marked with a wide diversity of jaw phenotypes that highlight contrasting patterns between simple and complex systems. Transgressive segregation produces functional phenotypes outside the expected hybrid distribution at high frequency in the complex 4-bar, and importantly these extreme functions are produced in the absence of morphological changes to the component links. This result is not found in the simple lever system where proportional changes in structure are necessary for changes in function. The attribute of many-to-one-mapping in the 4-bar linkage can explain such disjunction between morphological and functional diversity. We have shown from these results that 4-bar KT transgression is the result of recombinations of the existing variation in component links rather than the generation of novel morphology. Within the well documented complexity and MTOM of the 4-bar exists a strong linear component between the input and output lengths responsible for KT. In spite of this correlation the attributes of MTOM and decoupling of form to function serve to create very dynamic non-linear trait evolution in the complex system. The dynamics of the 4-bar exert trade-off constraints on the simple lever system through genetic and morphological correlations as well as an interesting limit to high KT in this lineage.

4.6 References

Albertson, R.C. and T.D. Kocher. 2005. Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution* 59:686-690.

Albertson, R.C., J.T. Streelman, T.D. Kocher. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl. Acad. Sci. USA* 100:5252-5257.

Albertson R.C., J.T. Streelman, T.D. Kocher, and P.C. Yelick. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternative feeding strategies. *Proc. Natl. Acad. Sci. USA* 102:16287–92

Alfaro, M.E., D.I. Bolnick, P.C. Wainwright. 2004. Evolutionary dynamics of complex biomechanical systems: an example using the four-bar mechanism. *Evolution* 58:495-503.

Alfaro, M.E., D.I. Bolnick, P.C. Wainwright. 2005. Evolutionary consequences of many-to-one-mapping of jaw morphology to mechanics in labrid fishes. *Am. Nat.* 165:E140-54.

Barel, C.D.N. 1983. Toward a constructional morphology of cichlid fishes (Teleostei, Perciformes). *Neth. J. Zool.* 33:357-424.

Bauwens, D., T. Garland Jr., A.M. Castilla, and R. Van Damme. 1995. Evolution of sprint speed in lacertid lizards: morphological, physiological, and behavioral covariation. *Evolution* 49:848–863.

Bennett, A.F. and R.E. Lenski. 2007. An experimental test of evolutionary trade-offs during temperature adaptation. *Proc. Natl. Acad. Sci. USA* 104:8649–8654.

Broman, K.W. and S. Sen. 2009. *A guide to QTL mapping with R/qtl.* Springer, New York, N.Y.

Carroll, A.M., P.C. Wainwright, S.H. Huskey, D.C. Collar, R.G. Turningan. 2004. Morphology predicts suction feeding performance in centrarchid fishes. *J. Exp. Biol.* 207:3873-81.

Fan, J.B. *et al.* 2003. Highly parallel SNP genotyping. *Cold Spring Harbor Symp. on Quant. Biol.* 68:69-78.

Genner M.J. and G.F. Turner. 2005. The *mbuna* cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries* 6:1–34.

Ghalambor, C.K., J.A. Walker, and D.N. Reznick. 2003. Multi-trait selection, adaptation, and constraints on the evolution of burst swimming performance. *Int. Comp. Biol.* 43:431–438.

Holzman, R., D.C. Collar, R.S. Mehta, and P.C. Wainwright. 2011. Functional complexity can mitigate performance trade-offs. *Am. Nat.* 177:E69-E83.

Hulsey, C.D. and F.J. Garcia de Leon. 2005. Cichlid jaw mechanics: linking morphology to feeding specialization. *Func. Ecol.* 19:487-494.

Hulsey, C.D. and P.C. Wainwright. 2002. Projecting mechanics into morphospace: disparity in the feeding system of labrid fishes. *Proc. R. Soc. Lond. B* 269:317-26.

Hulsey, C.D., G.J. Fraser, J.T. Streelman. 2005. Evolution and development of complex biomechanical systems: 300 million years of fish jaws. *Zebrafish* 2:243-257.

Mims, M.C., C.D. Hulsey, B.M. Fitzpatrick, J.T. Streelman. 2010. Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Mol. Ecol.* 19:940-951.

Muller, M. 1996. A novel classification of planar four-bar linkages and its application to the mechanical analysis of animal systems. *Phil. Trans. R. Soc. Lond. B* 351:689-720.

Oliphant A., D.L. Barker, J.R. Stuelpnagel, and M.S. Chee. 2002. BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 32:S56.

Parnell, N.F. and J.T. Streelman. 2011. The macroecology of rapid evolutionary radiation. *Proc. Roy. Soc. Lond. B* 278:2486-2494.

Parnell, N.F., C.D. Hulsey, and J.T. Streelman. 2008. Hybridization produces novelty when mapping of form to function is many to one. *BMC Evol. Biol.* 8:122-132.

Ribbink, A.J., A.C. Marsh, C.C. Ribbink, and B.J. Sharp. 1983. A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *S. Afr. J. Zool.* 18:149-310.

Rieseberg, L.H., M.A. Archer, R.K. Wayne. Transgressive segregation, adaptation and speciation. 1999. *Heredity* 83:363-372.

Rieseberg, L.H., O. Raymond, K.M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J.L. Durphy, A.E. Schwarzback, L.A. Donovan, C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211-1216.

Smith, P.F., A. Konings, I. Kornfield. 2003. Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. *Mol. Ecol.* 12:2497-2504.

Streelman, J.T. and P.D. Danley. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* 18:126–131.

Streelman, J.T., S.L. Gmyrek, M.R. Kidd, C. Kidd, R.L. Robinson, E. Hert, A.J. Ambali, T.D. Kocher. 2004. Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Mol. Ecol.* 13:2471-2479.

Toro, E., A.V. Herrel, D.L. Irschick. 2004. The evolution of jumping performance in Caribbean Anolis lizards: solutions to beiomechanical trade-offs. *Am. Nat.* 163:844-56.

van Ooijen, J.W. and R.E. Voorrips. 2001. Joinmap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, Netherlands.

Wainwright, P.C. 2007. Functional versus morphological diversity in macroevolution. *Ann. Rev. Ecol. Evol. Syst.* 38:381-401.

Wainwright, P.C. and B.A. Richard. 1995. Predicting patterns of prey use from morphology with fishes. *Env. Biol. Fishes* 44:97-113.

Wainwright, P.C., D.R. Bellwood, M.W. Westneat, J.R. Grubich, and A.S. Hoey. 2004. A functional morphospace for the skull of labrid fishes: patterns of diversity in a complex biomechanical system. *Biol. J. Linn. Soc.* 82:1-25.

Wainwright, P.C., M.E. Alfaro, D.I. Bolnick, C.D. Hulsey. 2005. Many-to-one-mapping of form to function: a general principle of organismal design? *Int. Comp. Biol.* 45:256-62.

Walker, J.A. 2007. A general model of functional constraints on phenotypic evolution. *Am. Nat.* 170:681–689.

Westneat, M.W. 1990. Feeding mechanics of teleost fishes (Labridae): a test of four-bar linkage models. *J. Morph.* 205:269-95.

Westneat, M.W. 1994. Transmission of force and velocity in the feeding mechanisms of labrid fishes. *Zoomorph.* 114:103-118.

Westneat, M.W. 1995. Feeding, function, and phylogeny: analysis of historical biomechanics in labrid fishes using comparative methods. *Syst. Biol.* 44:361-383.

CHAPTER 5

COMPLEX EPISTATIC INTERACTIONS DETERMINE SEX AND COLOR THROUGH MULTI-FACTORIAL ANTAGONISTIC SEX SYSTEMS IN LAKE MALAWI CICHLIDS

5.1 Abstract

The evolution of sex-determining systems and their effects on lineage diversification is a topic of much interest. Recent work has suggested an integral role of sexual conflict based sex chromosome evolution in models of speciation. In this study we use QTL analysis of an interspecific Lake Malawi cichlid cross to locate multiple loci responsible for sex determination and reproductively adaptive color phenotypes. We detect a complex system consisting of a major ZW locus on chromosome 5, a major XY locus on chromosome 7, and three other interactive loci both linked and unlinked to the major effect QTL. We present data supporting a previously undescribed association between a male heterogametic (XY) locus and an exclusively male nuptial color (blue), similar to the known association between ZW and the female dominant orange-blotch ("OB") color phenotype. These results provide empirical evidence for a complex antagonistic sex-color system in this species flock and suggest an important role and effect of polygenic sex-determining systems in rapid and extensive evolutionary diversification.

5.2 Introduction

One of the most fundamental and surprisingly diverse processes in the life history of an organism is the determination of sex. An array of environmental and

genetic controls has been discovered across divergent taxa yet the majority of sex-determining loci remain unidentified. Although sex-determining systems are frequently associated with a sex-specific heteromorphic chromosome pair (i.e. male XY or female ZW) more cryptic and complex systems are widespread across taxa (Ezaz *et al.* 2006). For example, among teleost fishes we find both simple and polygenic sex-determination, yet no general patterns exist and even the master sex-determining loci may vary within genera (Matsuda *et al.* 2002, Nanda *et al.* 2002, Mank *et al.* 2006). This diversity in such an integral genetic and developmental system sets sex-determination apart from other more highly conserved functional genetic pathways (Wilkins 1995, Carroll 2000).

The importance of reproductive isolation in speciation has led to interest in the role of sex-determining systems in this process (Kitano *et al.* 2009, Qvarnstrom and Bailey 2009, Meiklejohn and Tao 2010). Contemporary research has produced a bank of evidence in support of the role that natural selection plays in speciation (Schluter 2000, Streelman and Danley 2003, Losos and Ricklefs 2009, Schluter 2009), while sexual selection has received attention as an important contributor (Lande 1981, Ritchie 2007). Theoretical models of sympatric speciation have linked ecological and sexual selection as important drivers of diversification (Dieckmann and Doebelli 1999, Higashi *et al.* 1999, Kondrashav and Kondrashav 1999); enhanced by Lande *et al.* (2001) with the incorporation of color-based sexual selection and sexual conflict into two models of speciation. In these models an XY system is invaded by a dominant female allele (W) which functions as a male suppressor, as well as a W-linked novel color mutation and an unlinked locus for mate preference. Analytical results from

these models supported mechanisms of replicated sympatric speciation as a result of selection and interactions among sex-linked loci. The potential of sexually antagonistic forces producing intersexual conflicts is known (Rice 1992 & 1998, Chapman 2006). A possible result of such conflict is the evolution of new sex chromosomes through the action of an autosomal locus segregating differentially beneficial alleles based on sex. Selection in turn increases the frequency of sex-determining loci linked to these alleles, thereby serving as the impetus for a new sex system (Kocher 2004, Streelman *et al.* 2007). It is possible that rather than replacing an existing sex-determining system these new loci can remain polymorphic and thus create a complex interactive network of antagonistic sex-determining loci (Wilkins 1995, van Doorn and Kirkpatrick 2007).

The turnover of sex chromosomes as a result of sexually antagonistic selection has been suggested as important for the high speciation rates of some fishes (Seehausen *et al.*1999, Lande *et al.* 2001, Kocher 2004, Ross *et al.* 2009). In this paper we describe the existence of multiple sex-determining systems and the locations of several candidate loci in a hybrid cross between two Lake Malawi (LM) cichlid species.

Perhaps the fastest and most extensive vertebrate adaptive radiation known the LM cichlid flock consists of nearly 1000 species that have arisen in the past one million years (Won *et al.* 2005, Seehausen 2006). Throughout their history fluctuating lake levels coupled with island-like habitat fragments have driven a micro-allopatric diversification marked with extreme population structure, near zero dispersal and aggressive territoriality coupled with color-based assortative

mating (Hert 1992, van Oppen *et al.* 1997, Arnegard *et al.* 1999, Danley *et al.* 2000, Rico and Turner 2002, Genner and Turner 2005). Under these conditions natural selection has produced an array of feeding morphologies throughout the lake (Reinthal 1990, Albertson *et al.* 2005) while strong assortative mating is posited to have driven the striking variety of colors and patterns that exist across species (van Oppen *et al.* 1998).

The existence of multiple sex-determining systems (XY and ZW) in some LM species has been suggested using selective screening of sex-associated markers (Ser *et al.* 2010), but a conclusive genome-wide analysis does not exist. Past research suggests that a female heterogametic sex determiner (W) may be near to a female dominant color locus (orange-blotch or "OB"; Streelman *et al.* 2003) while an ancestral male dominant locus may exist on an adjacent chromosome (Roberts *et al.* 2009). Using QTL analysis of an F2 hybrid cross we identified multiple genomic regions responsible for polygenic sex-determination in these fish. We detail the association between the ZW locus and female-dominant OB color as well as a previously undescribed linkage between a male heterogametic XY and male-dominant blue nuptial color. This is the first genome-wide scan for sex-determining systems in LM cichlids and the first empirical data demonstrating the interplay between male and female sex and color postulated by theoretical models.

5.3 Methods

5.3.1 Production of F2 mapping population

Two rock-dwelling (*mbuna*) cichlid species from Lake Malawi were chosen for interspecific hybridization based on previous work (Parnell *et al.* 2008). The dam (*Pseudotropheus elongatus*) was a member of an herbivorous species in which female coloration approximates that found in the sire species (*Cynotilapia afra* Mara yellow type), a primarily planktivorous feeder (Figure 5.1). Several *C. afra* males were placed in a 189-L aquarium with *P. elongatus* females at a ratio of 1:2. Within a few weeks the dominant male (and subsequent sire) had established an obvious hierarchy over his conspecifics and had fertilized the eggs of the dam which were retained in her mouth. Fin clips (25 mm^2 anal fin) were taken from both the dam and sire for RAD tag library preparation and the pair was kept in a breeding tank with dither fish to diffuse male aggression. Offspring were taken from the dam after hatching, grown up in net pens and moved to consecutively larger aquaria as they grew. This pair produced several F_1 broods which were maintained in 189-L aquaria for 2 years during which time they produced over 40 broods consisting of more than 600 F_2 hybrid offspring. Each F_2 family was kept segregated from all other families and numbered by brood sequence (and individual). A total of 397 F_2 hybrids were euthanized with tricaine methane sulfonate (MS-222) and processed for DNA samples and a suite of morphological traits: intestine length, jaw morphometrics, jaw kinematics, tooth density and shape, taste bud density, sex, and color.

5.3.2 Sex determination and color assignment

Following euthanization with tricaine methane sulfonate (MS-222) hybrid individuals were photographed in a full body lateral view for color assignment.

We observed an unexpected emergent blue coloration in the F_2 which was not associated with the parental species but is found in other species in the lake, and therefore included it in our analysis. Color morphs were assigned binary coding as "OB" (1) or "NOT OB" (0) and "BLUE" (1) or "NOT BLUE" (0) depending on the pattern of coloration and the presence of blue/purple coloration on the body (see Figure 5.1). With this method it is possible for fish to be both OB and BLUE as a bluish OB pattern was produced in the F_2. The dam color, which lacks the blotches of OB, is the homozygous phenotype of OB (Roberts *et al.* 2009) and thus is included as such in scoring. Notes were made on variations in patterning and similarity to parental phenotypes. Fish were dissected and sex was determined from gross morphology (all fish were sexually mature) and assigned binary sex coding as male (1) or female (0).

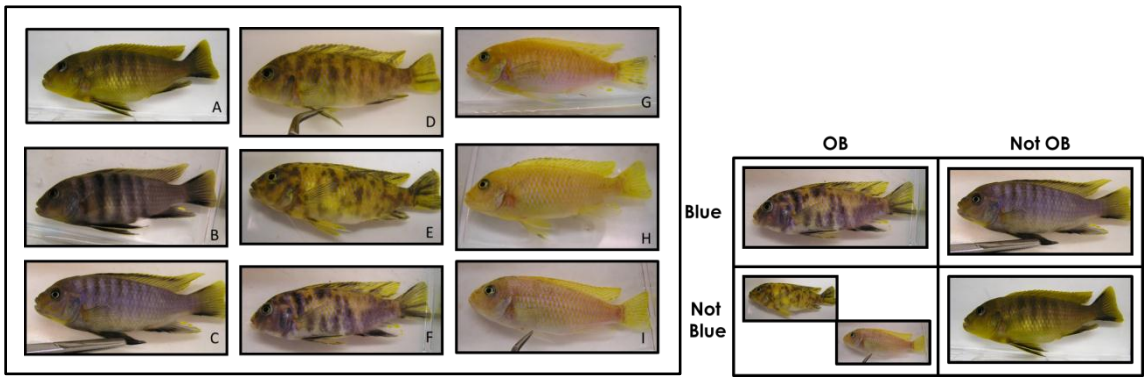

Figure 5.1. Color phenotypes of F_2 hybrid individuals. Left panel: (A) represents the sire, *Cynotilapia afra* and (I) displays the dam, *Pseudotropheus elongatus* ("O" phenotype=homozygous WW). (B), (C) and (F) show individuals scored as "blue" with (B) and (C) blue-barred and (F) blue-blotched ("blue-OB"). Individuals (D), (E) and (G)-(I) represent the "OB" phenotype with (G)-(I) displaying the "O" color (homozygous W) which cannot occur in males. (D) and (E) are representative of the OB phenotype (heterozygous ZW). Right panel: representative phenotypes for color scoring rubric.

5.3.3 DNA extraction and RAD library preparation

Genomic DNA was purified from sire and dam fin tissue using the DNeasy Blood & Tissue Kit (Qiagen). A sample of 10 µg DNA from each species was digested with the restriction endonuclease SbfI (recognition sequence CCTGCAGG) and RAD tag libraries were produced as in Baird *et al.* (2008). DNA was digested for 60 min at 37°C in a 50 µL reaction with 20 units (U) of SbfI (New England Biolabs [NEB]). Samples were heat-inactivated for 20 min at 65°C. 2.0 µL of 100 nM P1 adapters, a modified Solexa© adapter (2006 Illumina, Inc., all rights reserved) were added, each containing a unique multiplex sequence index (barcode) which is read during the first four nucleotides of the Illumina sequence read. In addition to the adapters 1 µL of 10 mM rATP (Promega), 1 µL 10× NEB Buffer 4, 1.0 µL (1000 U) T4 DNA Ligase (high concentration, Enzymatics, Inc.), and 5 µL H2O were added to the samples and each was incubated at room temperature (RT) for 20 min. Samples were again heat-inactivated for 20 min at 65°C, pooled and randomly sheared with a Bioruptor (Diagenode) to an average size of 500 bp. Samples were run out on a 1.5% agarose (Sigma), 0.5× TBE gel and DNA 300 bp to 800 bp was isolated using a MinElute Gel Extraction Kit (Qiagen). End blunting enzymes (Enzymatics, Inc.) were used to polish the ends of the DNA. Samples were purified using a MinElute column (Qiagen) and 15 U of Klenow exo− (Enzymatics) was used to add adenine (Fermentas) overhangs on the 3′ end of the DNA at 37°C. After subsequent purification, 1 µL of 10 µM P2 adapter, a divergent modified Solexa© adapter (2006 Illumina, Inc., all rights reserved), was ligated to the obtained DNA fragments at 18°C. Samples were again purified and eluted in 50 µL. The eluate was quantified using a Qubit fluorimeter and 20 ng of this product was used in a PCR amplification with 20 µL

Phusion Master Mix (NEB), 5 µL of 10 µM modified Solexa© Amplification primer mix (2006 Illumina, Inc., all rights reserved) and up to 100 µL H2O. Phusion PCR settings followed product guidelines (NEB) for a total of 18 cycles. Again, samples were gel purified and DNA bands from 300–700 bp size range were excised and diluted to 1 nM.

5.3.4 Illumina Sequencing

Two RAD libraries corresponding to cichlid dam and sire were run on an Illumina Genome Analyzer IIx at the University of Oregon High Throughput Sequencing Facility in Eugene, Oregon. Illumina /Solexa protocols were followed for paired end (2x60 bp) sequencing chemistry.

5.3.5 RAD LongRead Assembly

RAD LongRead uses mate-paired Illumina /Solexa data to assemble DNA sequence adjacent to restriction enzyme cleavage sites in a target genome. Unlike randomized short-insert paired end (SIPE) Illumina libraries, LongRead sequence data is characterized by a common or shared single end (SE) read that is anchored by the restriction enzyme digestion site and a variable paired end read. Similar strategies have been recently published as methodology for genome assembly in complex genomes using short-read Illumina /Solexa data (Hiatt *et al.* 2010).

Internal Floragenex sequence tools and perl scripts were used for processing of raw Illumina /Solexa data. First, raw fastq sequence data from each cichlid sample was segregated by the appropriate multiplex identifier

(MID) or DNA barcode assigned to each sample. To construct RAD LongRead contigs, data from *Cynotilapia afra* (the F_0 sire) was used to construct a reference assembly for SNP detection. The sequence from this individual was slightly higher quality than that of *Pseudotropheus elongatus* (the F_0 dam) and therefore it was chosen as the reference assembly. First, sequences with >5 poor Illumina quality scores (converted phred score of Q_{10} or lower) were discarded (typically <5% of all data). Remaining reads were then collapsed into RAD sequence clusters which share 100% sequence identity at the single end Illumina read. To maximize efficient assembly of sequences contributed from low-copy, single dose genome positions we imposed a minimum of 25x and maximum 500x sequence coverage at RAD single end reads. These thresholds have been empirically determined; single loci with coverage under 25x often display short and fragmented contig assemblies due to insufficient sequence coverage while loci with greater than 500 identical SE reads are often contributed from high-copy contaminant DNA or can be contributed by dosing from multiple genomic loci (e.g. repetitive class sequences). The variable paired end sequences for each common single end locus were extracted from these filtered sequences and passed to the Velvet® sequence assembler for contig assembly (Zerbino and Birney 2008). The contig data were further filtered to those with coverage higher than 4x (average 5.7x) and lengths greater than 200 bp (average 301.8 bp).

5.3.6 SNP Detection and Selection

Custom Floragenex short-read software using a Needleman-Wunsch algorithm was used to align paired-end Illumina / Solexa sequence reads from both cichlid samples to the reference cichlid sequence. Contigs were aligned between the parental species and compared for SNP selection and annotation. SNPs were assigned with a minimum of 2x coverage (two redundant reads) totaling 29,064 SNPs in 35,696 contigs. Relative to the estimated number of bases sequenced (~10.7 mb) this corresponds to a SNP rate of 0.27%, a value consistent with other published studies in LM cichlids (Loh 2008). The complete SNP panel was screened for those alleles free of polymorphic flanking regions (120 bp window) producing 9763 candidate SNPs for genotyping. From a subset of 1079 SNPs having at least 6x coverage or higher we selected 513 which were free of heterozygosity in either parent and contained the polymorphic SNP near the middle of the sequence to increase successful amplification. In addition, 10 SNPs from a previous cichlid sequencing project were included for a total of 523 markers. These SNPs were analyzed as potential candidates for genotyping with the Illumina® assay design tool, each generating a score from 0 to 1 (>0.6 have high probability of success). Markers with a designability score of 1.0 and a final score ≥0.84 were selected from our set, resulting in a final SNP genotyping dataset of 384 markers. In the final marker set 204 were 6-14x coverage and 180 were 15-30x coverage.

5.3.7 SNP genotyping in F2 individuals

A total of 382 F2 individuals (from 28 families) as well as sire and dam were genotyped at 384 SNP markers by the Emory Biomarker Service Center (Emory

University). We used the Illumina BeadArray genotyping platform coupled with the GoldenGate assay in which oligonucleotide pool assays (OPAs) are designed specifically to discriminate between alleles at a SNP (Oliphant *et al.* 2002, Fan *et al.* 2003). Overall genotyping success was 99.4% across all individuals, yielding a near complete dataset. Marker genotypes were examined to evaluate accuracy of initial sequencing during SNP discovery and 12 loci were removed as they did not segregate between parental lines. Another 12 markers that contained one heterozygous parental genotype were removed as well as 7 others that had low genotyping success in either parental or F_2 individuals. This left a fully informative set of 353 genotyped SNP markers in this cross.

5.3.8 Linkage map construction

A genetic linkage map was produced using genotype data from 382 individuals genotyped at 353 SNP markers with JoinMap® 3.0 software (van Ooijen and Voorrips 2001). The map was created using Kosambi's mapping function, a LOD threshold of 1.0, recombination threshold of 0.4, a jump threshold of 5.0, and a ripple function with no fixed order of loci. A LOD threshold of 4.0 was used to join 344 loci in 22 linkage groups with a total map size of 1082 cM and average marker distance 3.67 ± 1.71 cM (see Figure D.1). The linkage map generated from this cross was translated to the newly assembled tilapia and cichlid genomes (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html). This was accomplished by BLASTing SNP containing flanking sequences against both the

tilapia and cichlid scaffolds. A figure of scaffold hits and relative locations of QTL is included as a supplement (Figure D.1).

5.3.9 QTL mapping

The linkage map was used to determine genomic locations for sex, orange-blotch (OB) and blue color in the F_2 population using the R/qtl package (Broman and Sen 2009). We used an iterative approach by scanning for single QTL with standard and composite interval mapping (CIM), followed by two-dimensional scans to identify QTL x QTL interactions and detect additional QTL of marginal effects. Finally, using the results of the previous steps we built multiple QTL models (MQM) incorporating QTL interactions and covariates. In the MQM process we used a forward-backward selection algorithm to add and remove QTL based on overall model effects and the effects of single QTL as they were removed from the model. Genotype-phenotype associations are scored using the logarithm of the odds (LOD) which represents the $\log_{10}$ likelihood ratio comparing the hypothesis of a QTL at a marker location to the null hypothesis of no QTL ($LOD=(n/2)\log_{10}(RSS0/RSS1)$; Broman and Sen 2009). The variance in a phenotype is assigned to each significant QTL and reported as percent variance explained (PVE) in the analysis output. The variance accounted for by a QTL is a proxy for the heritability of a trait and is calculated as $1 - 10^{-(2/n)LOD}$ (Broman and Sen 2009). Significance thresholds for LOD scores were estimated using 1000 permutations of phenotypes relative to genotypes to build a distribution of maximum genome-wide LOD scores. From this distribution the 95th percentile LOD score could be calculated to serve as a threshold for significant QTL

associations (Broman and Sen 2009). Due to the nature of our dataset we used the Haley-Knott regression mapping method and binary phenotype modeling.

5.4 Results

5.4.1 Sex

Our dataset for the sex analysis included 353 individuals with complete phenotype data exhibiting a sex ratio of 42:58 male to female. Two QTL of major effect were detected on chromosomes 5 and 7 during the initial scan for sex loci (Figure 5.2). Scanning for two QTL simultaneously further supported these loci as strong effectors and epistatic interactors (full model LOD=37.5, interaction LOD=18.9). The scan for two QTL also indicated a second locus on chromosome 7 as an additive QTL with the locus on chromosome 5 (full model LOD=35.2, additive LOD=16.6) and revealed several other loci of interest (Figure 5.3 & Table 5.1). Using the results of the single and double scans, a multiple QTL model (MQM) was constructed and tested in multiple iterations, producing a significant final model (LOD=53.4). The full model incorporated five loci and three epistatic effects and accounted for over 50% of phenotypic variance (PVE=51.5; see Table 5.2).

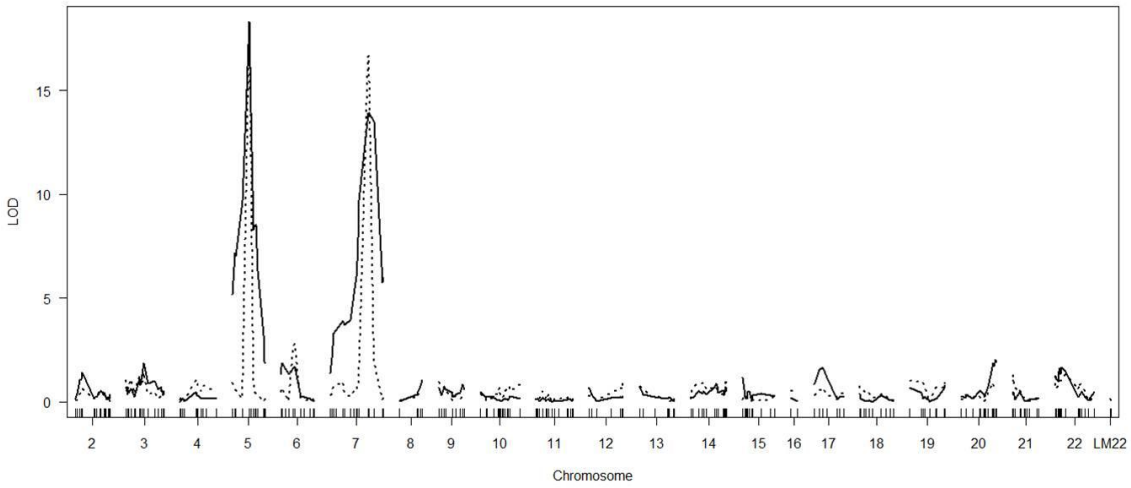

Figure 5.2. Single QTL scans for loci affecting sex. Solid line indicates standard interval mapping and dashed line indicates composite interval mapping. Horizontal axis denotes chromosome number with hash marks representing marker positions; vertical axis is log-likelihood score (LOD) of marker association with phenotype. Horizontal bar indicates significant LOD score (3.5; p<0.05).

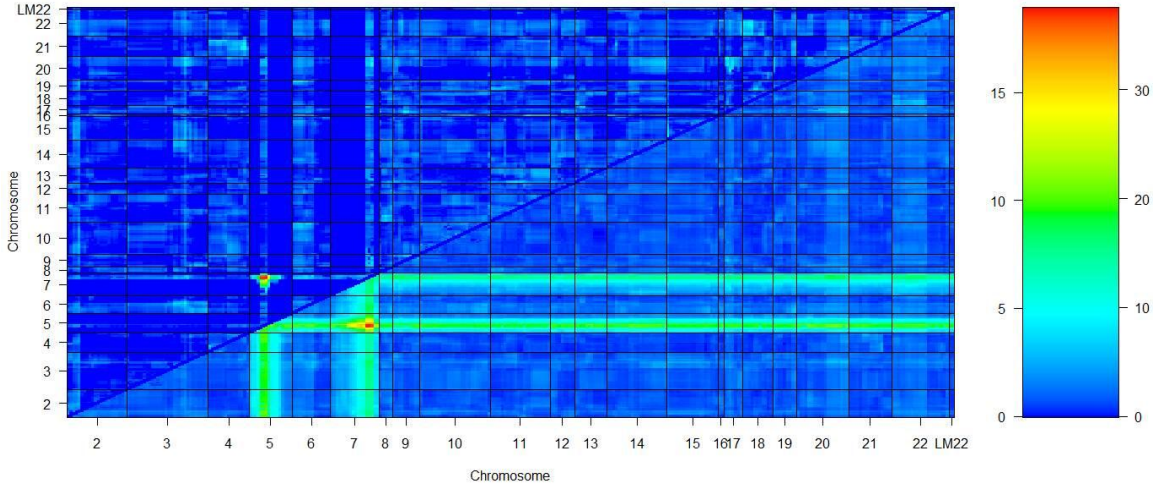

Figure 5.3. Simultaneous two QTL scan for loci affecting sex. Heat map plot represents log-likelihood score (LOD) for full 2-QTL model (below diagonal) and epistatic interaction between loci (above diagonal). LOD score increases with height on legend from 0 (blue) to highly significant (red). Horizontal and vertical axes represent chromosome number with location along chromosome increasing from left to right and bottom to top.

Table 5.1. Results of two QTL scan for loci affecting sex in Lake Malawi hybrid cross. Results are organized as full model including epistasis (FULL), the epistatic effect alone (INT), and the additive model (ADD) with corresponding locus positions (in cM) for both potential QTL. The log-likelihood score (LOD) is indicated for the full model with epistasis (LOD-F), the full model vs. the single QTL model (LOD-Fv1), the epistatic interaction alone (LOD-I), the additive model (LOD-A) and the additive model vs. the single QTL model (LOD-Av1).

	FULL	FULL	FULL	FULL v. 1	INT	ADD	ADD	ADD	ADD v. 1
Chr1:chr2	Pos 1	Pos 2	LOD-F	LOD-Fv1	LOD-I	Pos 1	Pos 2	LOD-A	LOD-Av1
c22 :c7	6.89	44.2	17	3.04	0.712	9.95	59.19	16.3	2.32
c5 :c20	26.62	49.4	21.7	3.16	0.843	26.62	49.42	20.9	2.31
c5 :c7	26.62	44.2	37.5	18.9	2.292	26.62	67.2	35.2	16.61
c5 :c6	26.62	20.7	20.8	2.23	0.106	26.62	20.7	20.7	2.12
c3 :c7	26.98	44.2	17.8	3.85	1.462	26.98	59.19	16.3	2.39
c7:c6	67.2	20.7	18	4.07	1.149	58.94	6.67	16.9	2.92

Table 5.2. Results of multiple QTL mapping (MQM) for loci controlling sex in Lake Malawi hybrid cross. The log-likelihood score (LOD), percent variance explained (PVE) and the X^2 p-value for the full model are indicated in the top portion. The lower table represents the results of dropping a putative QTL from the model one at a time. Putative QTL are indicated as chromosome@position in centimorgans (cM), "brood" represents inclusion of family as a cofactor, and epistatic interactions are indicated by factor:factor notation. Significant results are indicated as p<0.05 (*), p<0.005 (**) and p<0.0005 (***).

	df	LOD	PVE	Pvalue(Chi2)	
Full model	25	53.36	51.46	0	
Drop one QTL at a time ANOVA table:					
	df	LOD	PVE	Pvalue(Chi2)	Sig
5@26.6 (ZW)	6	29.15	23.50	<0.000001	***
7@67.2 (XY)	6	14.88	10.83	<0.000001	***
20@28.1 (X'Y')	12	11.44	8.14	<0.000001	***
3@27.0 (Z'W')	6	6.30	4.32	0.000006	***
7@44.2 (X"Y")	6	5.80	3.97	0.000164	***
brood	3	3.75	2.53	0.0006	***
20@28.1:7@67.2	4	6.10	4.18	0.000012	***
3@27.0:20@28.1	4	5.24	3.57	0.000075	***
5@26.6:7@44.2	4	4.09	2.77	0.000842	***
20@28.1:brood	2	1.86	1.24	0.013884	*

In the multiple QTL model the largest effect was seen at position 26.6 on chromosome 5 (5@26.6) with an estimated 23.5% phenotypic variance explained

(Table 5.2), coinciding with the position indicated in the single and double QTL scans. This region of chromosome 5 has previously been suggested as a ZW sex determining system in which the female allele (W) is dominant (Ser *et al.* 2010), and behaves in a similar fashion here (Figure 5.4). Individuals with WW genotype were unanimously female, regardless of genotypes at any other loci, while sex ratios (fraction of females) for ZW and ZZ genotypes were 0.54 and 0.43 respectively. Within the ZZ males we observed that all but four individuals had a homozygous "sire" or heterozygous genotype at the second major effect QTL (7@67.2) and full homozygous "sire" haplotypes were always male. This second QTL accounted for 10.83% variance (LOD=14.88) in sex (Table 5.2) and its action of effect suggests that it coincides with an ancestral XY sex determining system (Ser *et al.* 2010). We found that the ZW locus is dominant in sex determination to the XY and all other loci controlling sex (see below), however

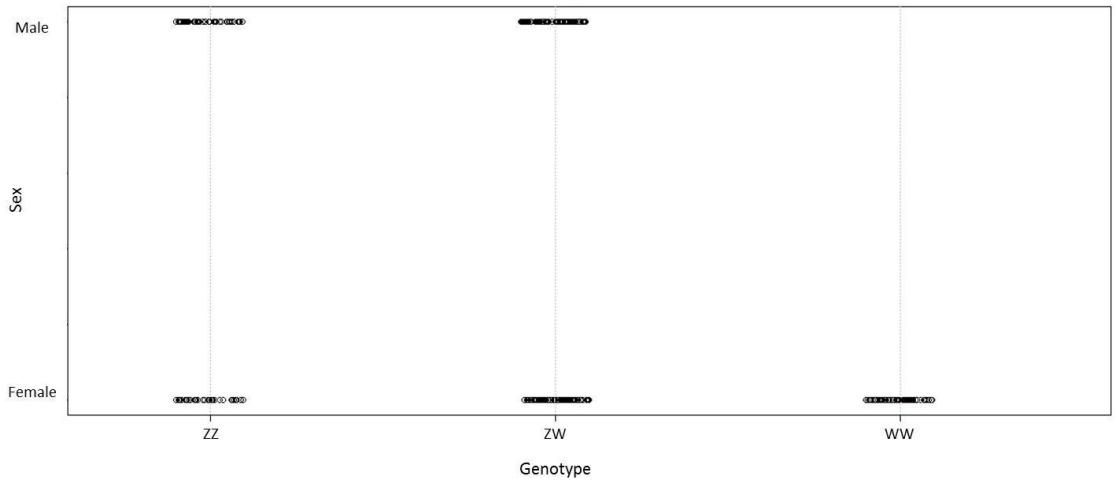

Figure 5.4. Association between gender and genotype at the putative ZW locus (chromosome 5@26.6 cM). All homozygous "W" individuals are female, while heterozygotes were slightly skewed towards females (54% female) and homozygous "Z" individuals were slightly skewed towards males (57% male) due to additional epistatic interactions among loci.

the effect of heterozygosity at the ZW is strongly affected by the genotype at the XY, as well as the effect of the other loci.

An unlinked QTL identified on chromosome 20 (20@28.1) exhibited significant effects (LOD=11.44, PVE=8.14) and epistasis with XY (Table 5.2). This locus behaves similarly to that of the putative XY (thus is labeled X'Y') and interacts with a fourth QTL (3@27.0), which has significant effects in the multiple QTL model (Table 5.2). This QTL at 27.0 cM on chromosome 3 behaves much like a second ZW (thus labeled Z'W'), except that it seems subordinate to the other loci in this cross. A fifth QTL was detected 23 centimorgans (cM) proximal to the putative XY (7@44.2). Genotypes here are similar to those at the distal XY site due to linkage; however, they differ in 37% of F_2 individuals. The behavior of this locus in sex determination appears to be as a complementary effector to the XY, functioning as another male dominant determiner (thus labeled X"Y"). Epistatic interactions had modest variance effects and relatively low LOD scores, yet they are integral to the strength of the overall model as well as explanation of haplotype effects on sex (Table 5.2). A significant effect of F_2 family ("brood") was detected in this analysis. This effect is most likely due to skewed sex ratios in a handful of families (4 skewed female, 1 skewed male) as well as a single family which was entirely male. Parentage of F_2 fish had no effect on sex when held as a covariate. The sex of some F_2 individuals (approximately 18%) were not as expected based on the QTL discovered in this analysis. These fish are most likely the result of other unmapped loci or non-genetic effectors in this multi-factorial system of sex determination (see Figure 5.4; environmental variables, sociality, diet, etc.).

5.4.2 Color

Complete color phenotype data were available for 340 F2 individuals in this cross for two color phenotypes: orange blotch (OB) and blue (see Figure 5.1). Almost 75% of F2 individuals were OB (or "O") while the remaining fish were either blue or colored similarly to the sire (see Figure 5.1). It is interesting to note that in the first filial generation 100% of individuals produced showed the OB phenotype. Previous work has indicated a single large effect QTL for the OB phenotype on cichlid chromosome 5 (Streelman *et al.* 2003). This locus was detected in each step of our analysis and found to explain over 90% of the phenotypic variance (Table 5.3, Figure D.2 & D.3). This QTL (5@26.6) is coincident with the largest effect QTL for sex which we infer to function as a ZW (female heteromorphic) sex determining locus. As suggested by previous work, our data indicate the female-linked color phenotype OB) is strongly associated with a female heteromorphic sex locus (Streelman *et al.* 2003). The presence of a single W allele at this locus expresses OB in all individuals, while homozygous WW individuals exhibit the "O" phenotype (Roberts, *et al.* 2009). Only homozygous sire (ZZ) genotypes did not exhibit OB (O).

Table 5.3. Results of multiple QTL mapping (MQM) for loci controlling orange blotch (OB) color in Lake Malawi hybrid cross. The log-likelihood score (LOD), percent variance explained (PVE) and the X^2 p-value for the full model are indicated in the top portion. The lower table represents the results of dropping a putative QTL from the model one at a time. Putative QTL are indicated as chromosome@position in centimorgans (cM), "sex" represents inclusion of sex as a cofactor, and epistatic interactions are indicated by factor:factor notation. Significant results are indicated as p<0.05 (*), p<0.005 (**) and p<0.0005 (***).

	df	LOD	PVE	Pvalue(Chi2)	
Full model	5	311.82	98.54	0	
Drop one QTL at a time ANOVA table:					
	df	LOD	PVE	Pvalue(Chi2)	Sig
5@26.6	4	308.58	94.24	<0.000001	***
sex	3	0.36	0.01	0.655	
5@26.6:sex	2	0.12	0.00	0.761	

The blue phenotype was an unexpected emergent trait in this cross between a homozygous OB ("O") dam and a black-barred yellow sire (see Figure 5.1). "Blue" individuals included those with blue blotches (approximating a blue OB pattern) as well as fish with black bars on blue background (similar to "BB" or "blue-barred" in Streelman *et al.* 2003; see Figure 5.1). Ten percent of F_2 exhibited the blue phenotype with the majority being "BB" type and our analysis revealed several candidate loci associated with this phenotype (full model LOD=40.69; Table 5.4, Figure D.4 & D.5). The largest variance in blue color was attributed to sex when held as a covariate (only males are blue; PVE=10.4) while the largest PVE for a QTL was at position 42.1 on chromosome 14 (Table 5.4). There were no effects of F_2 family ("brood") or parentage on this trait. A handful of loci of moderate effect were detected on several chromosomes throughout the genome including three epistatic interactions between six loci (Table 5.4). A significant QTL on chromosome 7 (position 43.8) was detected within 0.5 cM of

Table 5.4. Results of multiple QTL mapping (MQM) for loci controlling blue color in Lake Malawi hybrid cross. The log-likelihood score (LOD), percent variance explained (PVE) and the X^2 p-value for the full model are indicated in the top portion. The lower table represents the results of dropping a putative QTL from the model one at a time. Putative QTL are indicated as chromosome@position in centimorgans (cM), "sex" represents inclusion of gender as a cofactor, and epistatic interactions are indicated by factor:factor notation. Significant results are indicated as p<0.05 (*), p<0.005 (**) and p<0.0005 (***).

	df	LOD	PVE	Pvalue(Chi2)	
Full model	27	40.690	42.370	0	
Drop one QTL at a time ANOVA table:					
	df	LOD	PVE	Pvalue(Chi2)	Sig
sex	1	12.244	10.396	<0.000001	***
14@42.1	10	9.278	7.717	<0.000001	***
15@28.8	6	7.891	6.501	<0.000001	***
5@32.1	6	6.708	5.482	0.000027	***
6@35.3	2	6.589	5.379	<0.000001	***
LM22@0.8	2	6.384	5.175	<0.000001	***
18@39.8	2	4.826	3.893	0.0011	**
7@43.8	6	4.491	3.614	0.00209	**
14@42.1:18@39.8	4	2.329	1.847	0.02981	*
5@32.1:7@43.8	4	2.09	1.655	0.04726	*
14@42.1:15@28.8	4	2.086	1.651	0.04764	*

one of the male dominant sex determining loci (X"Y" above; 7@44.2). A locus of particular interest for the blue phenotype is that on chromosome 5 just distal (position 32.1) to the OB/ZW locus mapped in the analysis of sex (see above). As OB has been linked to females in nature, (Streelman *et al.* 2003, Roberts, *et al.* 2009, Ser *et al.* 2010) we find the blue phenotype linked to males.

5.5 Discussion

This study lends genome-wide empirical evidence to the genetic control of sex determination and color in cichlids and provides the first QTL evidence for a multi-factorial antagonistic sex-system in the LM species flock. Our results

support the description of cichlid chromosome 5 as hosting a female heterogametic (ZW) locus coincident with a female-associated OB color QTL (Streelman *et al.* 2003, Ser *et al.* 2010). We find strong evidence for additional sex loci including two male heterogametic (XY & X"Y") loci on chromosome 7, and complimentary X'Y' and Z'W' loci on unlinked chromosomes (20 and 3 respectively). Taken together in a comprehensive model these QTL account for over 50% of the gender variance in this interspecific cross with high statistical significance (LOD=53.36, $p<10^{-10}$). Two major QTL (PVE>10%) were detected (chr 5@26.6, ZW; 7@67.2, XY); however, all five loci have significant effects and are found to function in an obligate epistatic fashion in sex determination. The interactive effects between loci allows for the explanation of a larger proportion of individual genders when examined closely (see below).

5.5.1 A complex multi-factorial system

Using patterns of genotype-phenotype associations and percent variance explained in the final QTL model for sex we assigned a rank order to the loci therein as well as genotypic coding consistent with our locus assignments (ZW, XY, etc.; Table 5.2). All WW (dam genotype at locus 5@26.6) individuals were female (20.4% of individuals), while ZW and ZZ genotypes showed slightly skewed sex ratios (46:54% and 57:43% male to female respectively). The expectation in a purely ZW system is that the W allele confers the female sex, yet we do not see that here. Only homozygous W genotypes are unanimously female but the skew in sex ratios for ZW and ZZ genotypes is consistent with that expected in a pure ZW system (ZW more female, ZZ more male). The discrepancies between this ZW

system residing in a multi-locus sex determining environment and the expectation from a purely ZW system highlights the importance of the additional loci affecting sex.

A pattern consistent with XY effects was seen at the distal chromosome 7 QTL (7@67.2, XY) In this case the homozygous dam genotype (XX) was 90% female, the homozygous sire (YY) was 81% male, and XY individuals showed an equal sex ratio (50:50% male to female). Considering only the ZW (5@26.6) and XY (7@67.2) loci we were able to resolve the sex of over 53.8% of all F2 offspring (assuming W is dominant). It should be noted that both the proximal and distal QTL on chromosome 7 appear to function as male dominant heterogametic ("XY") loci in an additive manner (37% differing genotypes between both). It is likely that the cumulative effect of these genotypes in this complex system is dependent on the entire haplotype. The presence of these XY loci was suggested previously although that analysis did not have the power or analytical resources to further examine the interplay and effects therein (Albertson 2002). Two additional unlinked QTL showed significant effects on sex determination but with different mechanisms. First, chromosome 20 contains a locus (20@28.1, X'Y') that behaves similarly to both the XY and X"Y" loci found on chromosome 7. This locus was also found to interact with the fifth QTL in the sex model which seems to behave more like a female heterogametic determiner (3@27.0, Z'W').

5.5.2 Patterns of additive and epistatic effects

Patterns of interlocus effects are particularly evident when considering ZW individuals (Figure 5.4). As discussed, the expectation of inheriting the W allele is

that these offspring will be female. The importance of additivity among the QTL in this model is apparent when considering that all but 3 of the 82 ZW males are XY or YY at a minimum of two of the three XY loci (XY, X'Y', X"Y"), and over half are XY or YY at all three. This result implies an effect analogous to dose dependence based on the total number of alleles of each type and the genotypes at hierarchically superior loci (i.e. the ZW). Patterns of epistatic effects are also evident in ZWYY individuals. In this case the interaction between Z'W' and X'X' (3@27.0 x 20@28.1) is indicated by female sex in 8 of 13 individuals with this haplotype. A similar effect between Y'Y' and YY (20@28.1 x 7@67.2) can explain 7 of 7 individuals with these genotypes being male as well as 15 of 16 X'Y'Y"Y" male offspring. Similarly, 30 of 32 ZWXX hybrid individuals were female. The ZW individuals also exhibit a distinct pattern of dominance effects. Increasing heterozygosity at all loci (ZW→XY→X'Y'→X"Y"→Z'W') is directly related to an increasing number of male individuals in spite of the expected effects at the ZW loci (female heterogametic). Further support for QTL interactions were seen among ZZXY individuals. The epistatic effect between the ZW (5@26.6) and X"Y" (5@44.2) explains the observation of unanimous male phenotypes among all individuals with the ZZY"Y" haplotype, while 14 of 19 ZZX"X" offspring were female.

Examining patterns in offspring gender in the context of QTL haplotypes as modeled (ZW→XY→X'Y'→X"Y"→Z'W' plus epistatic effects) we were able to resolve the sex of 81.5% of all F2 individuals (Figure 5.5). We were unable to more clearly distill the QTL effects in the context of a complex haplotype but it is apparent that a significant portion of variance in gender is due to epistasis. We

cannot discount the effect of environmental influences on sex determination in fish (Baroiller *et al.* 2009) and we accept that these effects in cichlids in particular are not well described. The conditions in which we conducted our intercross experiment were well controlled in terms of temperature and other physical water parameters. However, social, dietary, and yet undescribed factors can still account for some undetermined variance in sex phenotypes, such as the significant effect of brood in our analysis (Table 5.2).

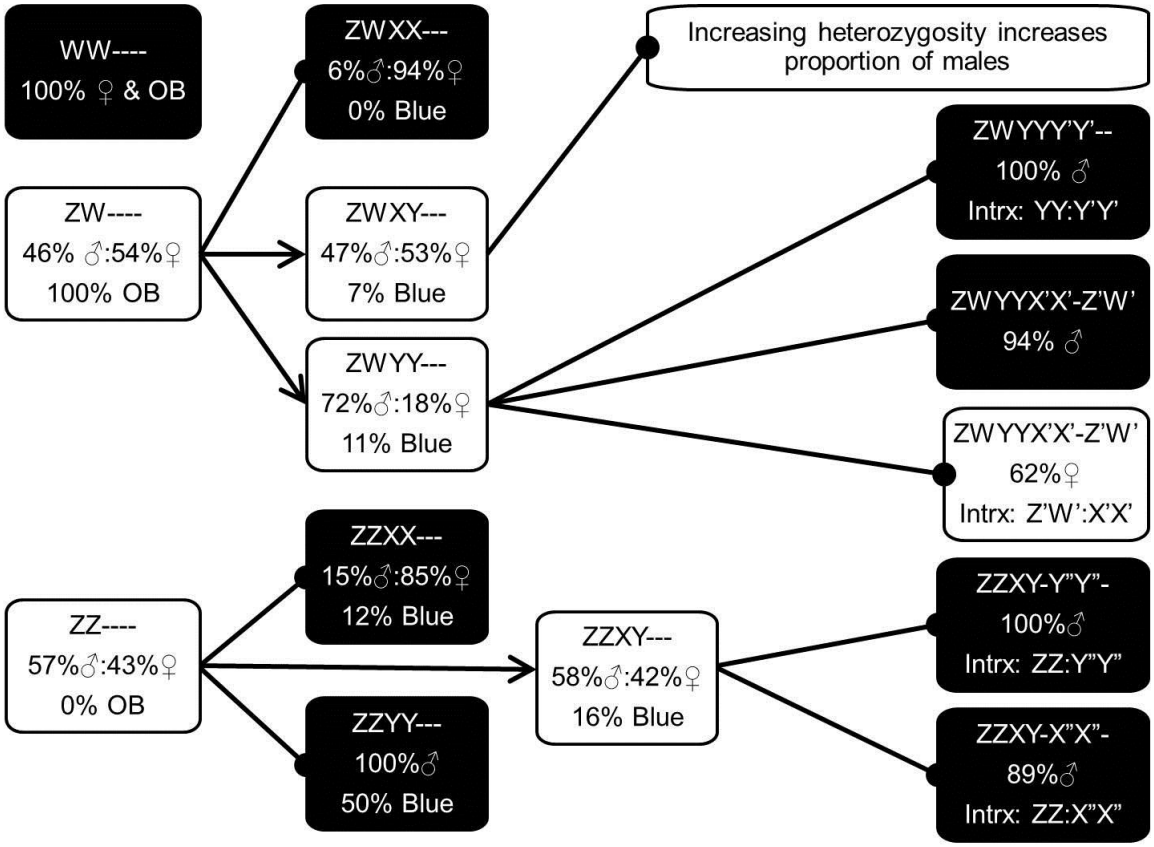

Figure 5.5. Patterns in gender and color phenotypes based on associated haplotypes at multiple loci as described in multiple QTL models. Filled polygons represent haplotypes that account for the majority of individuals or beyond which no pattern exists. Dashes within a haplotype represent locus genotypes which are indeterminate (and irrelevant) for the phenotype. Interactive locus effects are evident in several haplotypes as modeled and where applicable the epistatic interaction implicated in the phenotype is noted.

5.5.3 Sex, color and the evolution of sex determination

The coincidence of the QTL for the OB color phenotype with the female-dominant ZW has been suggested by previous work (Streelman *et al.* 2003, Roberts *et al.* 2009). We find that the W allele confers this trait to all individuals possessing it. Of 246 OB individuals 164 were female (66.7%) and all (100%) were genetically WW or ZW, while the 82 OB males were exclusively ZW. As expected no "O" colored individuals were male as the genotype at this locus is unanimously WW to produce this phenotype. In natural populations the rarity of OB males and lack of homozygous OB individuals suggests reduced fitness in males with this trait (Holzberg 1978, Seehausen *et al.* 1999b). This detriment is attributed to disruption of male nuptial coloration in an environment of strong color-based sexual selection (Roberts *et al.* 2009). Tight linkage between the OB locus and a female heterogametic sex determiner (ZW) has resolved the sexually antagonistic selection of this trait and thus may indicate an invasion of this system into the ancestral XY (van Doorn and Kirkpatrick 2007, Ser *et al.* 2010). This linkage may also represent a reduction in recombination around a sex determining locus, which is expected in the sequence of sex chromosome evolution (Rice 1987, Streelman *et al.* 2007). The presence of blue-OB individuals is interesting as it further supports the interplay between the sex determining loci on chromosome 5 (@26.6 cM; ZW) and chromosome 7 (@44.2; X"Y"). The blue-OB color and pattern is only possible if both the W and a Y" allele are inherited together.

Sexual conflict at the OB-ZW locus has been considered in the past; however we find an additional linkage between blue color phenotype and a putative XY locus. Blue color is completely linked to the male sex (YY or XY only) and the loci flanking two putative sex determiners (X"Y" and ZW) have significant effects on the blue phenotype. Invoking the model of chromosome evolution by sexual conflict (van Doorn and Kirkpatrick 2007) we can attribute linkage between the male dominant blue phenotype and male heterogametic sex loci as we have above for OB-ZW. Male coloration and other display traits have been found to be in strong linkage with sex chromosomes in several other fish species (Lindholm and Breden 2002, Basolo 2006). Once again sexually antagonistic fitness benefits push such linkages because trait expression in males is beneficial for reproduction but detrimental to females due to increased predation and/or intraspecific aggression (Bull 1983, Rice 1987). Much like the OB phenotype is beneficial to females, and thus in tight linkage with the ZW locus, it is likely that linkage between the male determining XY loci and the blue phenotype is the result of sexually antagonistic selection. Although not a single blue individual (blue OB or blue-barred) was female we did observe an appreciable number of females featuring the yellow with black bar pattern of the sire in this cross (see Figure 5.1). The blue variant of the barred pattern represents a novel putative male breeding plumage for each of these parental species, although it is common across other species in the lake. In this case a novel male nuptial color such as this could increase in frequency in a population due to its linkage with male dominant heterogametic loci.

Our results empirically identify several presumably autosomal sex-determining loci and their linkage to loci responsible for both female and male dominant colors in Lake Malawi cichlids. The presence of multiple sex-determining loci and coincidence of at least one of these with the OB color locus has been suggested previously (Streelman *et al.* 2003, Roberts *et al.* 2009, Ser *et al.* 2010). Albertson (2002) presented data indicating a sex locus near the position of the XY identified here, as well as a possible second locus of interest on the same chromosome but proximal to the first position – results directly supported by our findings here. The newly sequenced tilapia genome and collinearity with that of the LM cichlids allows us to compare and contrast our QTL to those loci previously described in this ancient cichlid lineage. Previous work on two tilapia species identified an XY system on chromosome 1 – the only linkage group we did not find correspondence with in our linkage map (Cnaani *et al.* 2008). In the same study the authors also report a ZW system in two other tilapia species on chromosome 3, a finding that corresponds with our Z'W' described in this study (3@27.0). Finally, in yet two other tilapia species the presence of both sex systems was reported as well as interactions between the unlinked loci (Lee *et al.* 2004, Cnaani *et al.* 2008). Multifactorial sex systems are predicted in the theoretical literature to be unstable and thus transient (Bull and Charnov 1977, Karlin and Lessard 1986). Based on the diversity and lability of sex systems in fish lineages and the rapidity and extent of the LM cichlid radiation the presence of such a complex mechanism of sex determination is not unexpected (Volff and Schartl 2001, Nanda *et al.* 2003, Kondo *et al.* 2004, Otake *et al.* 2006, Cnaani *et al.* 2008, Ross *et al.* 2009, Ser *et al.* 2010). Incorporating previous work

and our results we propose that the sex determining QTL are deployed in a hierarchical additive and epistatic framework. Assigning sex to the F2 population with this QTL rubric allowed us to explain over 80% of sex variation. Our results fit well into the model of sex chromosome evolution based on genetic conflicts and represent the first presentation of a multi-factorial sex system incorporating conflicts at multiple loci linked to sex-specific enhancer traits in both males and females. Our findings further highlight the role of sexual selection in speciation and the obligate interaction of this force along with ecological selection in shaping the radiation of this diverse lineage.

5.6 References

Albertson, R.C. 2002. *Genetic basis of adaptive radiation in East African cichlid fishes.* Ph.D. thesis, Univ. of new Hampshire, Durham, NH.

Albertson R.C., J.T. Streelman, T.D. Kocher, and P.C. Yelick. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternative feeding strategies. *Proc. Natl. Acad. Sci. USA* 102:16287–92

Arnegard, M.E., J.A. Markert, P.D. Danley, J.R. Stauffer, A.J. Ambali, and T.D. Kocher. 1999. Population structure and colour variation of the cichlid fishes Labeotropheus fuelleborni Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proc. Biol. Sci.* 266:119-130.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko, and E.A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376.

Baroiller, J.F., H. D'Cotta, and E. Saillant. 2009. Environmental effects on fish sex determination and differentiation. *Sex. Dev.* 3:118-135.

Basolo AL. 2006. Genetic linkage and color polymorphism in the southern platyfish (Xiphophorus maculatus): a model system for studies of color pattern evolution. *Zebrafish* 3:65–83.

Broman, K.W. and S. Sen. 2009. *A guide to QTL mapping with R/qtl.* Springer, New York, N.Y.

Bull J.J. 1983. Evolution of sex determining mechanisms. Benjamin Cummings, Menlo Park, CA. Bull, J.J., and E.L. Charnov. 1977. Changes in the heterogametic mechanism of sex determination. *Heredity* 39:1–14.

Carroll, S.B. 2000. Endless forms: The evolution of gene regulation and morphological diversity. *Cell* 101:577-580.

Chapman, T. 2006. Evolutionary conflicts of interest between males and females. *Curr. Biol.* 16:R744–R754.

Cnaani, A., B. Y. Lee, N. Zilberman, C. Ozouf-Costaz, G. Hulata, M. Ron, A. D'Hont, J. F. Baroiller, H. D'Cotta, D. J. Penman, *et al.* 2008. Genetics of sex determination in tilapiine species. *Sex. Dev.* 2:43–54.

Danley, P. D., J. A. Markert, M. E. Arnegard, and T. D. Kocher. 2000. Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* 54:1725–1737.

Dieckmann U. and M. Doebeli 1999. On the origin of species by sympatric speciation. *Nature* 400:354–57.

Ezaz, T., R. Stiglec, F. Veyrunes, and J.A.M. Graves. 2006. Relationships between vertebrate ZW and XY sex chromosome systems. *Curr. Biol.* 16:R736-R743.

Fan, J.B. *et al.* 2003. Highly parallel SNP genotyping. *Cold Spring Harbor Symp. on Quant. Biol.* 68:69-78.

Genner M.J. and G.F. Turner. 2005. The *mbuna* cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries* 6:1–34.

Hert, E. 1992. Homing and home-site fidelity in rock-dwelling cichlids (Pisces:Teleostei) of Lake Malawi, Africa. *Env. Biol. Fishes* 33:229-237.

Hiatt, J.B., R.P. Patwardhan, E.H. Turner, C. Lee, and J. Shendure. 2010. Parallel, tag-directed assembly of locally derived short sequence reads. *Nature Methods* 7:119-122.

Higashi, M., G. Takimoto, and N. Yamamura. 1999. Sympatric speciation by sexual selection. *Nature* 402:523-526.

Holzberg, S. 1978. A field and laboratory study of the behaviour and ecology of Pseudotropheus zebra (Boulenger), an endemic cichlid of Lake Malawi (Pisces: Cichlidae). *Zeit. Zool. Syst. Evol.* 16:171-187.

Karlin, S., and S. Lessard. 1986. *Sex ratio evolution.* Princeton Univ. Press, Princeton, NJ.

Kitano, J. *et al.* 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079-1083.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* 5:288–298.

Kondo, M., I. Nanda, U. Hornung, M. Schmid, and M. Schartl. 2004. Evolutionary origin of the medaka Y chromosome. *Curr. Biol.* 14:1664–1669.

Kondrashov A.S. and F.A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400:351–54.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* 78:3721–3725.

Lande, R., O. Seehausen, and J. J. van Alphen. 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* 112–113:435–443.

Lee, B.-Y., G. Hulata, and T. D. Kocher. 2004. Two unlinked loci controlling the sex of blue tilapia (Oreochromis aureus). *Heredity* 92:543–549.

Lindholm, A., and F. Breden. 2002. Sex chromosomes and sexual selection in poeciliid fishes. *Am. Nat.* 160:S214–S224.

Loh, Y.E., L.S. Katz, M.C. Mims, T.D. Kocher, S.V. Yi, and J.T. Streelman. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Gen. Biol.* 9:R113.

Losos J.B. and R.E. Ricklefs. 2009. Adaptation and diversification on islands. *Nature* 457:830-836.

Mank, J.E., D.E.L. Promislow, and J.C. Avise. 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* 87:83-93.

Matsuda M., Y. Nagahama, A. Shinomiya, T. Sato, C. Matsuda, *et al.* 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–63

Meiklejohn, C.D. and Y. Tao. 2010. Genetic conflict and sex chromosome evolution. *Trends Ecol. Evol.* 25:215-223.

Nanda, I., U. Hornung, M.Kondo, M. Schmid, and M. Schartl. 2003. Common spontaneous sex-reversed XX males of the medaka Oryzias latipes. *Genetics* 163:245–251.

Nanda I., M. Kondo, U. Hornung, S. Asakawa, C.Winkler, *et al.* 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y-chromosome of the medaka, Oryzias latipes. *Proc. Natl. Acad. Sci. USA* 99:11778–83

Oliphant A., D.L. Barker, J.R. Stuelpnagel, and M.S. Chee. 2002. BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 32:S56.

Otake, H., A. Shinomiya, M. Matsuda, S. Hamaguchi, and M. Sakaizumi. 2006. Wild-derived XY sex-reversal mutants in the medaka, Oryzias latipes. *Genetics* 173:2083–2090.

Parnell, N.F., C.D. Hulsey, and J.T. Streelman. 2008. Hybridization produces novelty when mapping of form to function is many to one. *BMC Evol. Biol.* 8:122-132.

Qvarnstrom, A. and R.I. Bailey. 2009. Speciation through evolution of sex-linked genes. *Heredity* 102:4-15.

Reinthal, P. N. 1990 The feeding habits of a group of herbivorous rock-dwelling cichlid fishes from Lake Malawi, Africa. *Env. Biol. Fish* 27, 215–233.

Rice, W.R. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41:911–14

Rice, W.R. 1992. Sexually antagonistic genes – experimental evidence. *Science* 256:1436–1439.

Rice, W.R. 1998. Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome. *Proc. Natl. Acad. Sci. USA* 95:6217–6221.

Rico, C., and G. F. Turner. 2002. Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Mol. Ecol.* 11:1585–1590.

Ritchie, M. G. 2007. Sexual selection and speciation. *Ann. Rev. Ecol. Evol. Syst.* 38:79–102.

Roberts, R. B., J. R. Ser, and T. D. Kocher. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* 326:998-1001.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro, and C. L. Peichel. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* 5: e1000391. doi:10.1371/journal.pgen.1000391.

Schluter, D. 2000. *The ecology of adaptive radiation.* Oxford Univ. Press, Oxford, U.K.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737-741.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc. Roy. Soc. B.* 273:1987-1998.

Seehausen, O., J.J.M. van Alphen, and R. Lande. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* 2:367-378.

Seehausen O., P.J. Mayhew, J.J.M. van Alphen. 1999b. Evolution of colour patterns in East African cichlid fish. *J. Evol. Biol.* 12:514–34.

Ser, J.R., R.B. Roberts, and T.D. Kocher. 2010. Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution* 64:486-501.

Streelman, J.T. and P.D. Danley. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* 18:126–131.

Streelman, J.T., R.C. Albertson, and T.D. Kocher. 2003. Genome mapping of the orange blotch colour pattern in cichlid fishes. *Mol. Ecol.* 12:2465-2471.

Streelman, J.T., C.L. Peichel, and D.M. Parichy. 2007. Developmental genetics of adaptation in fishes: the case for novelty. *Ann. Rev. Ecol. Evol. Syst.* 38:655-681.

van Doorn, G. S., and M. Kirkpatrick. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449:909–912.

van Ooijen, J.W. and R.E. Voorrips. 2001. Joinmap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, Netherlands.

van Oppen M. J. H., G. F. Turner, C. Rico, J. C. Deutsch, K. M. Ibrahim, R. L. Robinson, and G. M. Hewitt. 1997. Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proc. Biol. Soc.* 264:1803–1812.

van Oppen, M.J.H., G.F. Turner, C. Rico, R.L. Robinson, J.C. Deutsch, M.J. Genner and G.M. Hewitt. 1998 Assortative mating among rock-dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Mol. Ecol.* 7:991–1001.

Volff J-N. and M. Schartl. 2001. Variability of genetic sex determination in poeciliid fishes. *Genetica* 111:101–10

Wilkins, A.S. 1995. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* 17:71–77.

Won, Y-J., Y. Wang, A. Sivasundar, and J. Hey. 2005. On the origin of Lake Malawi cichlid species. *Proc. Natl. Acad. Sci. USA* 102:6581-6586.

Zerbino, D.R. and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Gen. Res.* 18:821-829.

CHAPTER 6

OVERALL CONCLUSIONS

Rapid evolutionary radiations have long drawn the attention of biologists. The unique attributes and features of these phenomena provide opportunities for experimental designs and tests of hypotheses untenable in other systems. The Lake Malawi cichlid fishes represent the fastest and most diverse vertebrate adaptive radiation on earth whereby hundreds of extant species, and potentially thousands of species in total, have evolved over the last one million years. This diversification has occurred in a unique geographic and ecological setting and under presumably strong natural and sexual selection pressures to produce an inimitable array of species and phenotypes in the face of little genetic differentiation. Despite decades of research our grasp of the exact mechanisms facilitating and sustaining these diversity generating processes remain cursory. This dissertation encompasses research aimed at addressing some of the many questions raised by this and other adaptive radiations. This work strives to enhance our understanding of the basic mechanisms generating and maintaining diversity at multiple biological scales and contribute to our knowledge of forces driving and opposing diversification.

In Chapter 2 we investigate species-level diversity using null model analyses of occurrence data. We find evidence of non-random community assembly at the broadest and finest spatial resolutions examined. Across 18 rock-reef sites dispersed throughout the lake we attribute decreased species co-occurrence to the evolutionary history of this species flock. The extreme

endemism, strong site fidelity and local population genetic structure of these fish lend support this alternative hypothesis to random species co-occurrence at the lake-wide level. Contrarily, at the finest (depth-within-site) resolution we attribute community structure to fine-scale ecological interactions shaping species distributions along depth gradients. We discuss data indicating competition for trophic, space, and reproductive resources as potential drivers of the negative correlations between species by water depth. Considering the group of species most commonly overlapping in distributions we report the identification of a "core" cichlid community, replicated at numerous sites throughout the lake. Developing these findings we conclude that the diverse fine-grain cichlid communities in the lake are a product of ecological interactions on the background of a replicated core group, thus producing diverse and abundant communities. These data and conclusions set up a myriad of experimental rubrics that can be used to test these hypotheses empirically and determine the strength of ecological interactions between various species in this assemblage.

Chapter 3 focuses more closely on a portion of the diversity of trophic traits in these species which may play a significant role in some of the interactions described above. We use a complex biomechanical system exhibiting many-to-one-mapping (MTOM) of form to function to investigate how novelty in function may arise during species hybridization. We model a mechanism which produces functional novelty in the majority of species combinations at evolutionarily relevant frequencies. By this mechanism transgression in function is possible in the absence of transgressive morphology, but rather is the result of recombinations of physical lengths of the maxilla and

lower jaw links, and is predictable from the morphology of the parent species. This study posits a likely pervasive source of novel functional phenotypes constructed from existing morphology and presents hypotheses which are readily testable in the LM system.

With the model of potential functional novelty described in Chapter 3 we were presented with a challenge to create hybrid individuals and assay whether extreme traits can be generated in this way. We add to this test by considering not only the complex 4-bar jaw system but also the simple jaw lever system which has been suggested to limit transgression due to its history of selection and lack of MTOM. As hypothesized we detect appreciable transgression in complex (4-bar) function without corresponding changes to morphological structures. Although extreme trait values are produced in the simple jaw lever as well it is characterized by obligate changes in morphology to produce transgressive function. The results of this cross indicate that the recombination of existing morphology is the mechanism creating novel functional phenotypes in the 4-bar and that the link lengths involved have a strong linear component to their respective morphologies. Genomic mapping of the structural and functional traits in both systems identify the genetic architecture responsible for novelty in these traits and provide a perspective on functional trade-offs and evolutionary constraints both within and between jaw systems. In following up the genetic model of novel complex phenotypes with a laboratory cross we have identified a probable mechanism to maintain high levels of variation within different morphological combinations among species as well as a basis for creative evolution of ecologically relevant phenotypes following hybridizations events.

Revisiting phenotypic variation and QTL analyses in Chapter 5 we engage questions concerning the determination of gender in these species as well as the highly variable, yet geographically replicated color morphs found in the LM cichlids. We confirm the genetic basis for a ubiquitous female-specific color and identify QTL associated with an emergent male nuptial pattern which has not been reported in these species previously. Linkage between these color phenotypes and several loci associated with sex-determination suggests the potential for sex-conflict based sex chromosome evolution. The two major sex systems, ZW and XY, are supplemented by three other loci and several epistatic interactions. We find the ZW linked to the female "OB" color while the male blue coloration is associated with this region and several other loci coinciding with XY systems. We discuss these findings in the context of an antagonistic sex-color system and the potential effects of polygenic sex-determination in the rapid and extensive evolutionary diversification of these fish. Along these lines we suggest that the maintenance of multiple sex loci linked to adaptive and nuptial colors in this rapidly evolving lineage highlights the importance of sexual selection in speciation. Considering the results of both QTL studies in this work we detect a genetic pattern whereby the jaw traits examined in Chapters 3 and 4 are clumped in the genome with many of those associated with sex and color. These "hotspots" represent genomic regions potentially exerting pleiotropic control of adaptive traits and thereby producing covariation in structural and functional phenotypes. Suggestion has been made that speciation is more rapid and likely when loci controlling ecologically relevant adaptive traits (i.e. jaw function) and those affecting mating preferences (i.e. coloration) become linked. The

association of reduced recombination around areas of sex determining loci could facilitate such linkage, making these areas prime genomic sites for relevant traits to evolve. Clumped loci such as these may explain parallel bursts of evolution seen in this and other adaptive radiations.

The results presented in this dissertation reveal various mechanisms responsible for the generation and maintenance of diversity at several biological levels of organization. The common thread to these studies is two-fold. First, we have chosen to investigate a system marked by exceptional evolutionary and ecological complexity and uniqueness. Detection of the patterns described here are largely possible because of the attributes of this biological assemblage yet they are likely pervasive in some capacity across all organisms. Second, the information collected and communicated in these chapters represents an amalgam of different methods, approaches and lines of research coalesced into a common theme. The data presented here support processes important for diversity in communities, novelty within simple and complex functional biomechanical phenotypes, differential and shared genetic controls for the determination of gender and adaptive color traits, as well the potential for combinations of these into genomic "hotspots". This work adds to our understanding of the forces and processes shaping evolutionary and ecological patterns while it also reveals more unknowns and serves as the impetus for other hypotheses about diversity in all species and at all levels.

6.1 Publications

1. Parnell, N.F., C.D. Hulsey, and J.T. Streelman. The genetic basis of a complex functional system. *In preparation.*

2. Parnell, N.F. and J.T. Streelman. Complex epistatic interactions determine sex and color through multi-factorial antagonistic sex systems in Lake Malawi cichlids. *In preparation.*

3. Parnell, N.F., C.D. Hulsey, and J.T. Streelman. 2008. Hybridization produces novelty when mapping of form to function is many to one. BMC Evol. Biol. 8:122-132.

4. Parnell, N.F. and J.T. Streelman. 2011. The macroecology of rapid evolutionary radiation. Proc. Roy. Soc. Lond. B. 278:2486-2494.

5. Hulsey, C.D., Mims, M.C., Parnell, N.F. andJ.T. Streelman. 2010. Comparative rates of lower jaw diversification in cichlid adaptive radiations. *J. Evol. Bio.*, 23, 1456-1467.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

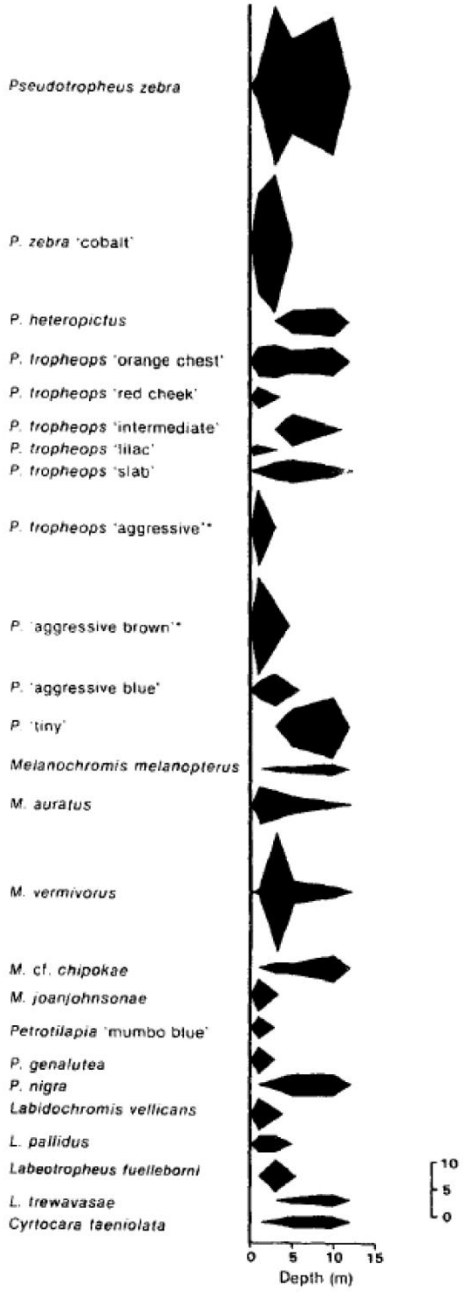

Figure A.1. Species abundance kite plots as presented in Ribbink *et al.* (1983). Horizontal axis represents water depth in meters, vertical axis denotes species present at site. Width of kites represents species abundance relative to scale bar. Measurements of species abundance were made at sampling depths in Ribbink *et al.* (1983) as described in methods.

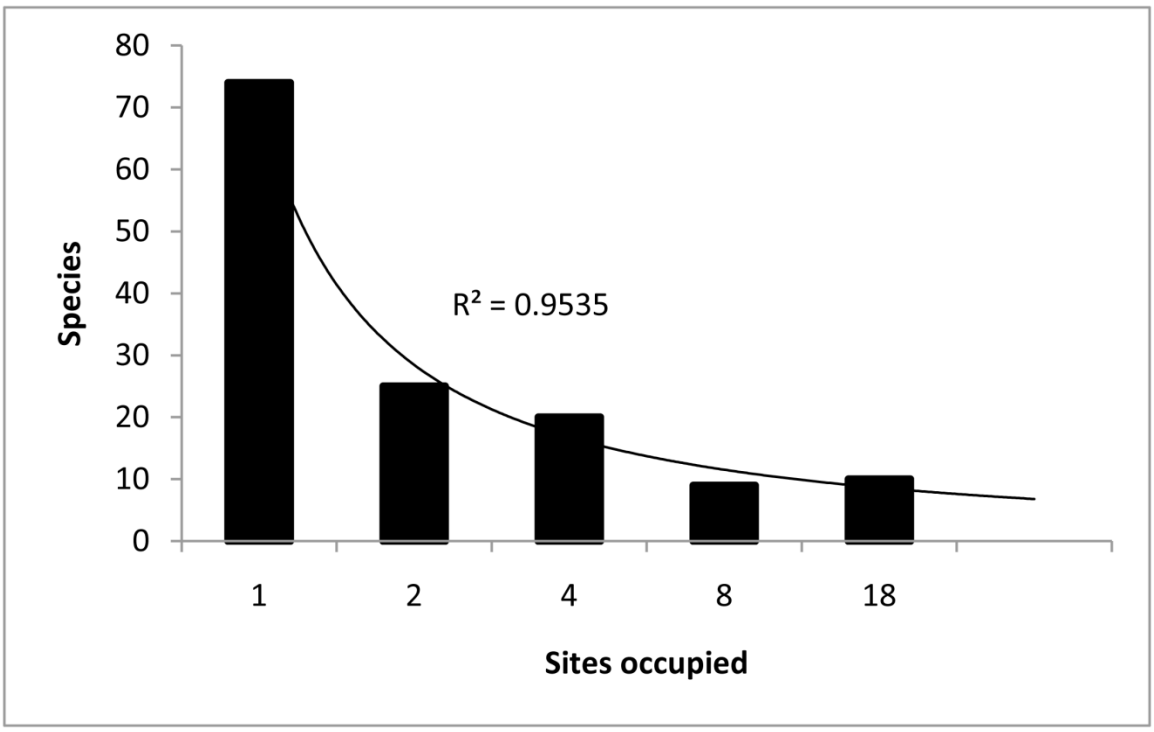

Figure A.2. Frequency of species site-occupancy. The majority (54%) of species in Lake Malawi are endemic to a single site ('singletons'). The distribution is well fit to a power function.

Table A.1. EcoSim analysis of Lake Malawi cichlid 'core' species' (n=12) observed co-occurrence matrix vs. 10,000 randomized matrices. Scale of analysis is in descending order, # sites [# spp] denotes the number of sites or depths and 'core' species used at each scale, SES is standardized effect size (see methods) of observed C-score compared to mean of simulated matrices, Probability of Cobs>Csim is the probability of an observed C-score being greater than the average of random matrices, Result is determination of random or structured community based on analysis. Listed values correspond to fixed-fixed model (SIM9; Gotelli 2000). All p-values are corrected for multiple tests within a scale.

Scale of analysis	# sites [# spp]	SES	Prob of $C_{obs}>C_{sim}$	Result
Regional (lake-wide)	18 [12]	1.58	p=0.077	NULL
Subregion – North	11 [8]	1.00	p<0.87	NULL
Subregion – South	7 [3]	0.00	p<0.99	NULL

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

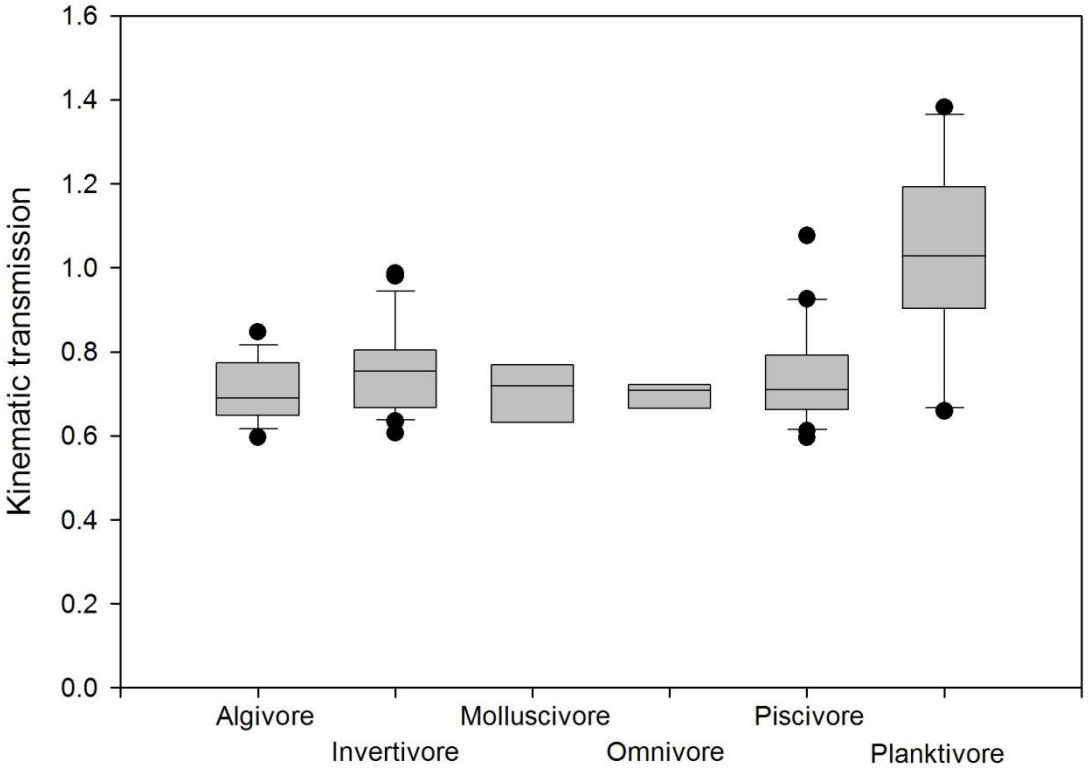

Figure B.1. Planktivores have higher KT, on average, than other trophic groups. Trophic groups are defined as in Hulsey *et al.* (2007; see Figure B.1). The bar is the median value, the box is the 25th-75th percentile, whiskers are the 10th and 90th percentiles and the dots are outliers beyond the 5th and 95th percentiles.

Figure B.2. Bayesian phylogenetic hypothesis for Lake Malawi cichlids, derived from the mitochondrial ND2 gene, and used to generate phylogenetically independent correlations among link lengths and links with KT. Posterior probability values are given for selected nodes.

Table B.1. Average KT, trophic group and GenBank accession numbers for all species used in this study. Trophic groups are defined as in Hulsey *et al.*2007.

Genus	species	KT	Diet	GenBank no.
Astatotilapia	calliptera	0.80	Omnivore	AF305280
Aulonocara	macrochir	1.00	Planktivore	
Aulonocara	stuartgranti	0.79	Invertivore	EU661720
Buccochromis	heterotaenia	0.66	Piscivore	EU661719
Chilotilapia	euchilus	0.63	Invertivore	EF585280
Chilotilapia	rhoadesii	0.67	Molluscivore	
Copadichromis	eucinostomus	0.99	Planktivore	EF585268
Copadichromis	mbenjii	0.96	Planktivore	EF585255
Copadichromis	mloto	1.08	Planktivore	
Copadichromis	quadrimaculatus	1.21	Planktivore	AF305314
Copadichromis	virginalis	1.38	Planktivore	
Corematodus	taeniatus	0.41	Fin/scales	
Cyathochromis	obliquidens	0.67	Algivore	
Cynotilapia	afra	0.74	Planktivore	EF585264
Cyrtocara	moorii	0.64	Invertivore	AY930089
Dimidiochromis	compressiceps	0.82	Piscivore	EF585267
Dimidiochromis	kiwinge	0.67	Piscivore	AF305322
Docimodus	evelynae	0.83	Cleaner	EF585252
Fossorochromis	rostratus	0.71	Piscivore	EF585281
Genyochromis	mento	0.60	Fin/scales	AF305297
Hemitaeniochromis	urotaenia	0.64	Piscivore	
Labeotropheus	fuelleborni	0.60	Algivore	EF585259
Labeotropheus	trewavasae	0.64	Algivore	EF585283
Labidochromis	gigas	0.74	Algivore	
Labidochromis	vellicans	0.85	Algivore	
Lethrinops	altus	0.78	Invertivore	
Lethrinops	longimanus	0.61	Invertivore	
Maravichromis	anaphyrmus	0.82	Molluscivore	AF305321
Maravichromis	epichorialis	0.60	Molluscivore	
Maravichromis	incola	0.69	Invertivore	
Maravichromis	lateristriga	0.76	Invertivore	
Maravichromis	mola	0.72	Molluscivore	EF585274
Maravichromis	spilostichus	0.70	Piscivore	
Melanochromis	auratus	0.68	Invertivore	AY930069
Metriaclima	aurora	0.67	Omnivore	EF585266
Metriaclima	callainos	0.71	Omnivore	EF585271
Metriaclima	heteropictus	0.71	Omnivore	
Metriaclima	zebra	0.72	Omnivore	DQ093114
Nimbochromis	fuscotaeniatus	0.70	Piscivore	
Nimbochromis	linni	0.75	Piscivore	EF585279
Nimbochromis	polystigma	0.70	Piscivore	EF585262
Nyassachromis	prostoma	1.19	Planktivore	EU661715
Otopharynx	pictus	1.05	Planktivore	EF585254
Otopharynx	heterodon	0.66	Planktivore	EF585278
Otopharynx	lithobates	0.80	Invertivore	

Table B.1. (continued)

Otopharynx	selenurus	0.74	Invertivore	
Otopharynx	tetrastigma	0.87	Invertivore	
Otopharynx	walteri	0.89	Invertivore	EU661716
Petrotilapia	"fuscous"	0.66	Algivore	
Petrotilapia	nigra	0.71	Algivore	EU661721
Placidochromis	johnstoni	0.99	Invertivore	
Placidochromis	longimanus	0.69	Omnivore	
Placidochromis	milomo	0.78	Invertivore	EF585251
Placidochromis	subocularis	0.75	Invertivore	
Protomelas	annectans	0.65	Invertivore	EU661718
Protomelas	fenestratus	0.67	Invertivore	AF305301
Protomelas	ornatus	0.66	Invertivore	EU661717
Protomelas	similis	0.70	Algivore	EU661714
Protomelas	spilonotus	0.74	Invertivore	
Protomelas	spilopterus "blue"	0.71	Invertivore	EF585253
Protomelas	taeniolatus	0.68	Algivore	AF305302
Protomelas	triaenodon	0.77	Algivore	
Pseudotropheus	crabro	0.72	Cleaner	EF585256
Pseudotropheus	elongatus	0.69	Algivore	EF585272
Pseudotropheus	livingstonii	0.58	Both	EF585273
Pseudotropheus	lucerna	0.78	Algivore	
Rhamphochromis	esox	1.08	Piscivore	AF305252
Rhamphochromis	lucius	0.93	Piscivore	
Rhamphochromis	macropthalmus	0.76	Piscivore	AF305249
Sciaenochromis	ahli	0.79	Piscivore	
Serranochromis	robustus	0.60	Piscivore	
Stigmatochromis	modestus	0.76	Piscivore	
Stigmatochromis	pleurospilus	0.91	Piscivore	
Stigmatochromis	woodi	0.71	Piscivore	AF305299
Taeniochromis	holotaenia	0.65	Piscivore	
Taeniolethrinops	praeorbitalis	0.98	Invertivore	AF305318
Tramitichromis	lituris	0.86	Invertivore	
Trematocranus	microstoma	0.76	Invertivore	
Trematocranus	placodon	0.72	Molluscivore	EF585261
Tropheops	"broad mouth"	0.69	Algivore	EF559101
Tropheops	"orange chest"	0.78	Algivore	EF585275
Tropheops	"red cheek"	0.81	Algivore	EF585265
Tropheops	gracilior	0.63	Algivore	EF585260
Tropheops	microstoma	0.62	Algivore	EF585258
Tyrannochromis	macrostoma	0.61	Piscivore	EF585257
Tyrannochromis	maculiceps	0.79	Piscivore	

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

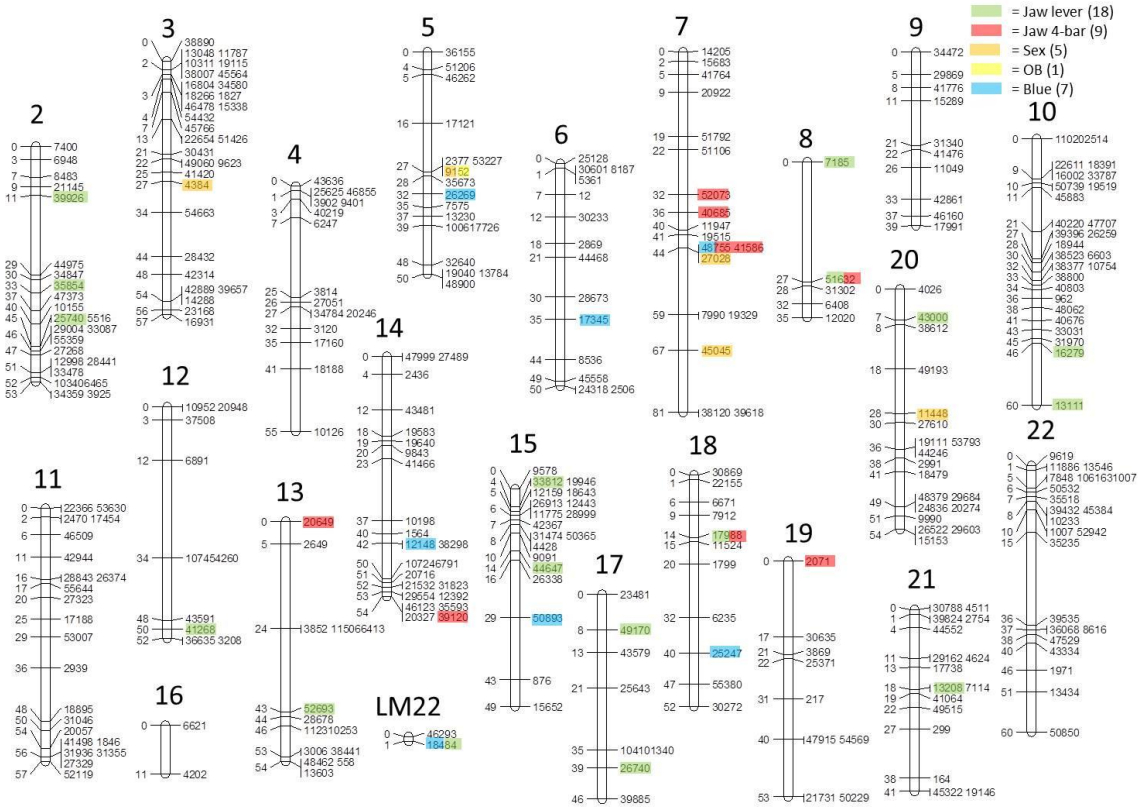

Figure C.1. Lake Malawi linkage map generated and used in this study. Numbers represent chromosome labels based on homology with the recently assembled tilapia genome (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html). Markers are single nucleotide polymorphisms (SNPs) with relative distance in centimorgans (cM) noted. Colored bars represent QTL detected for various traits measured in F2 hybrid population. Simple lever=green; complex 4-bar linkage=red; sex=orange; orange-blotch color=yellow; blue color=blue. For sex, orange-blotch and blue see Chapter 5.

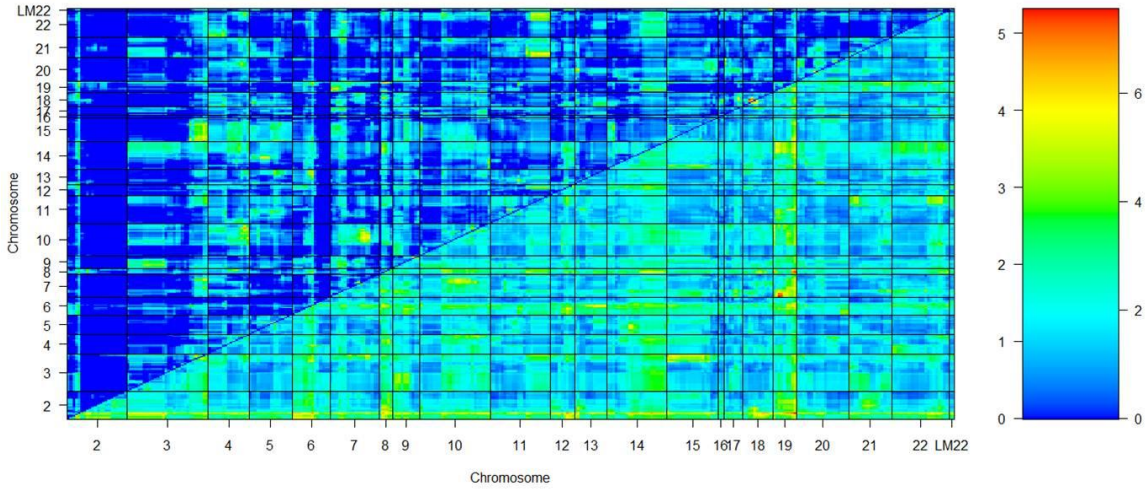

Figure C.2. Two QTL scan for loci affecting input link (AJMax) of 4-bar linkage. Heat map plot represents log-likelihood score (LOD) for full 2-QTL model (below diagonal) and epistatic interaction between loci (above diagonal). LOD score increases with height on legend from 0 (blue) to highly significant (red). Horizontal and vertical axes represent chromosome number with location along chromosome increasing from left to right and bottom to top.

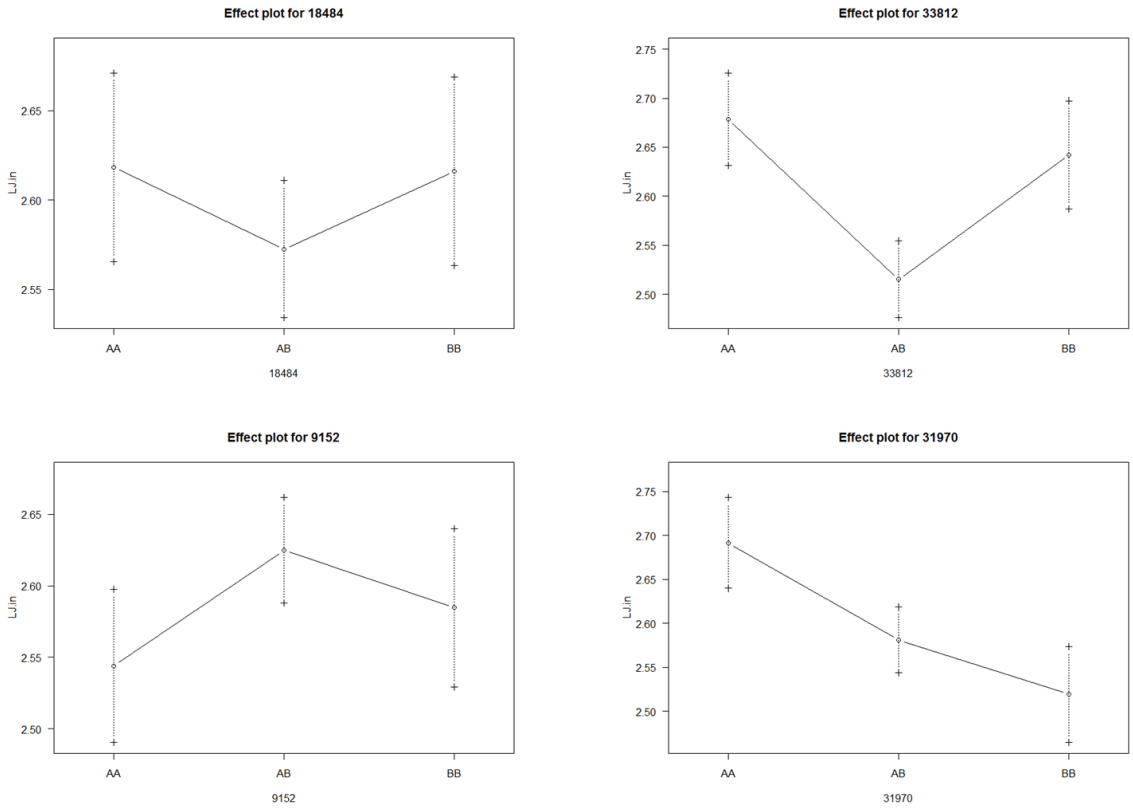

Figure C.1. QTL effect plots for the loci responsible for the highest PVE in the simple jaw in-lever (LJin). Marker 18484 corresponds to LM22@0.8 cM, 33812 is at 15@3.7, 9152 is at 5@26.63, and 31970 is at 10@45.0.

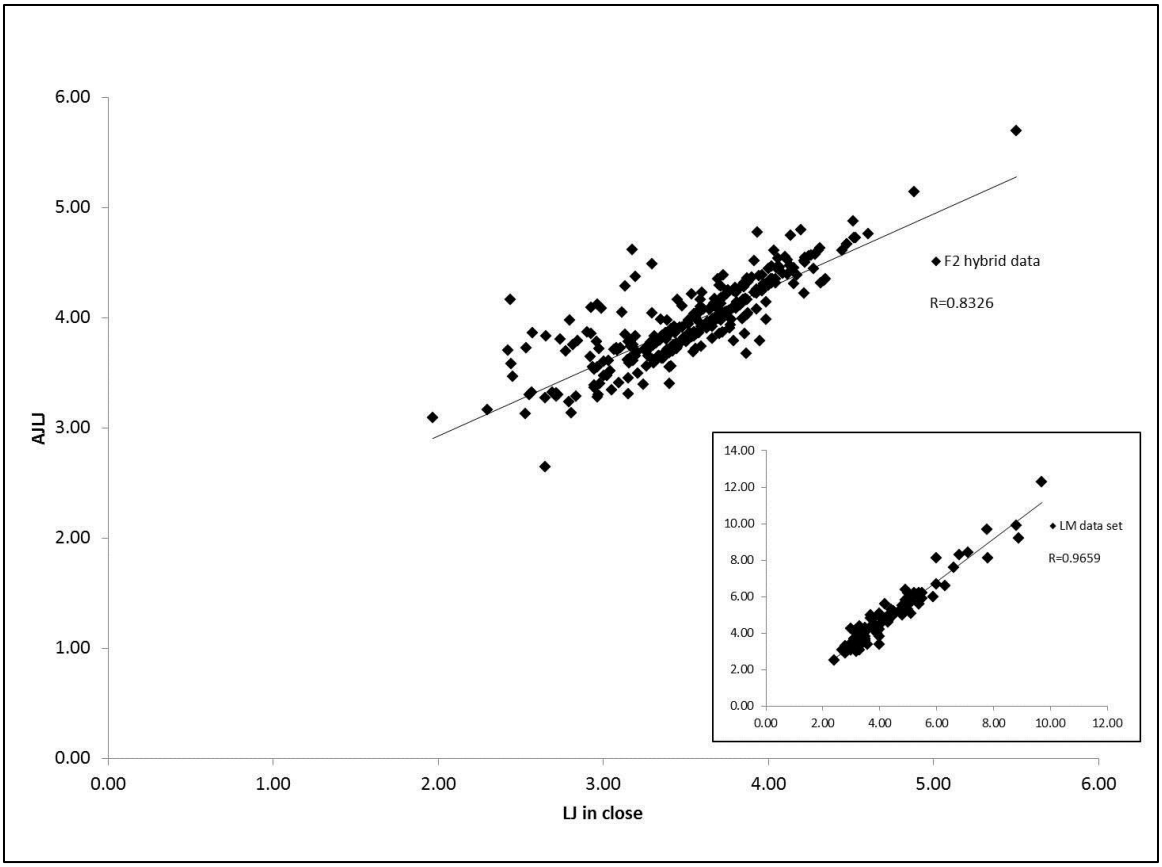

Figure C.4. Strong correlation between the simple lever closing in-lever (LJinc) and the 4-bar input link (AJLJ). Main plot represents F2 hybrid data; inset includes data from 86 species of Lake Malawi cichlids (Parnell *et al.* 2008).

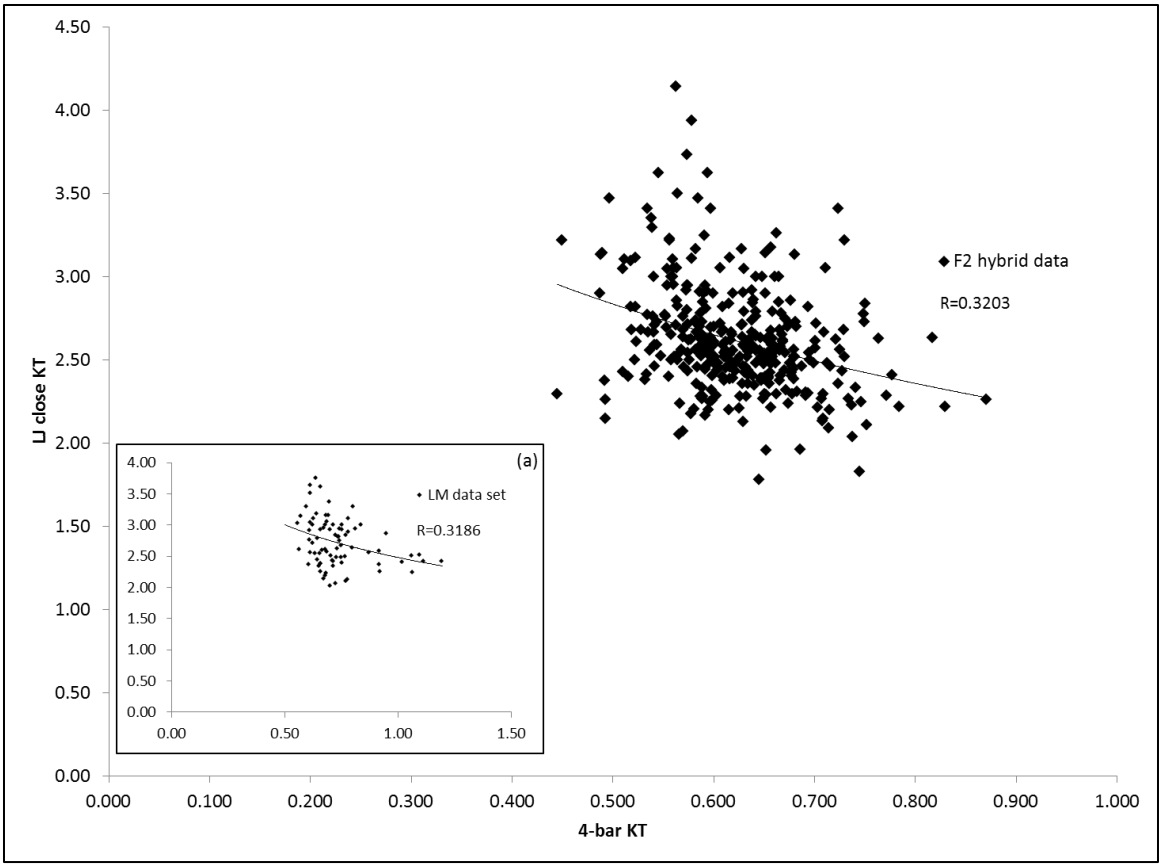

Figure C.5. Function of the complex 4-bar linkage (kinematic transmission; KT) is negatively correlated with function of the simple lever jaw closing system (closing KT). Main plot represents F2 hybrid data; inset includes data from 86 species of Lake Malawi cichlids (Parnell *et al.* 2008).

APPENDIX D

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

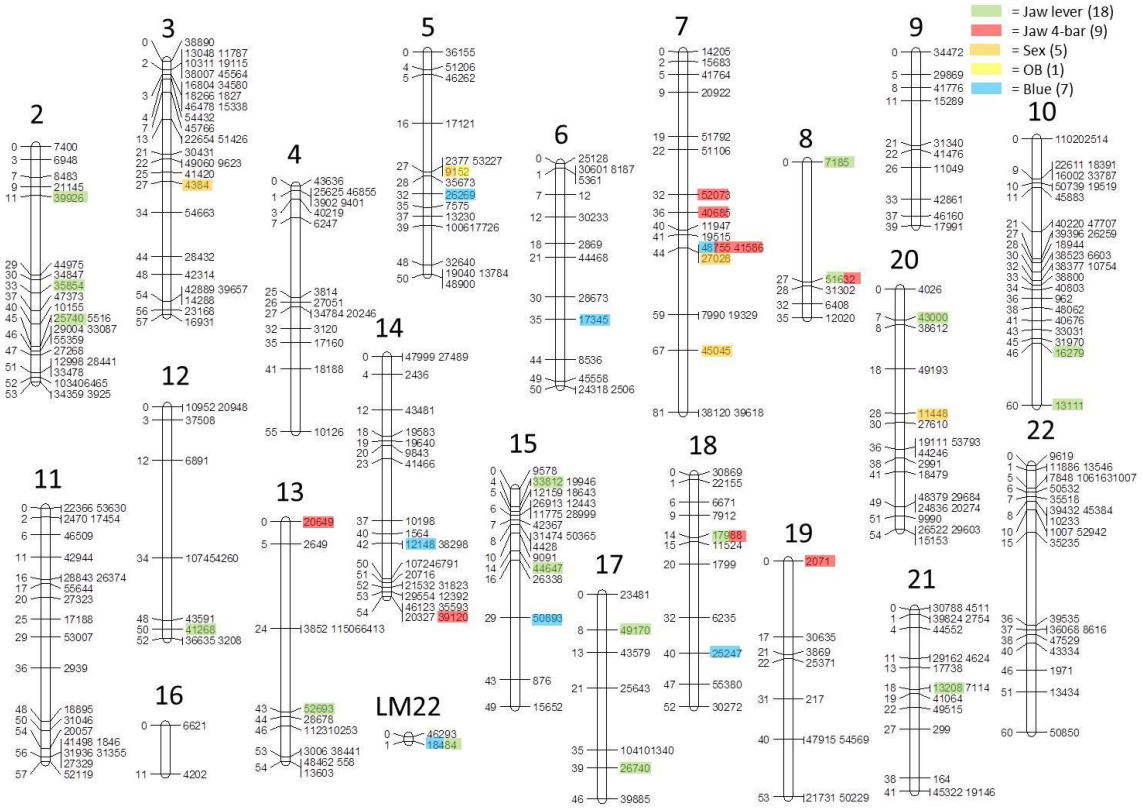

Figure D.1. Lake Malawi linkage map generated and used in this study. Numbers represent chromosome labels based on homology with the recently assembled tilapia genome (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html). Markers are single nucleotide polymorphisms (SNPs) with relative distance in centimorgans (cM) noted. Colored bars represent QTL detected for various traits measured in F2 hybrid population.; sex=orange; orange-blotch color=yellow; blue color=blue; simple jaw lever=green; complex 4-bar jaw linkage=red. For jaw traits please see Chapter 4.

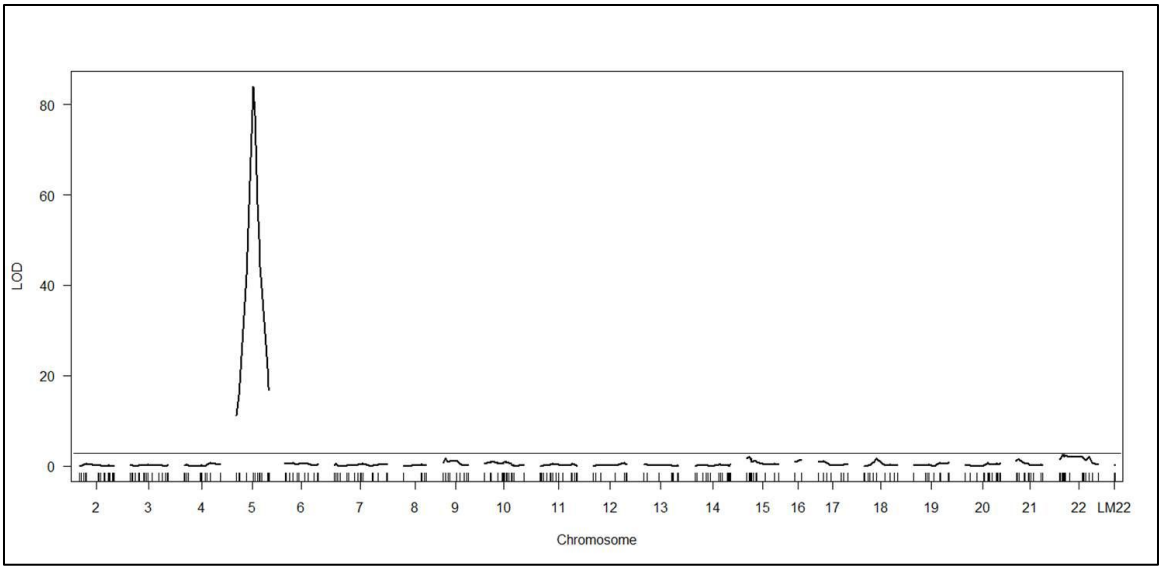

Figure D.2. Single QTL scan for loci controlling orange blotch (OB) color pattern. Horizontal axis denotes chromosome number with hash marks representing marker positions; vertical axis is log-likelihood score (LOD) of marker association with phenotype. Horizontal bar indicates significant LOD score (3.5; p<0.05).

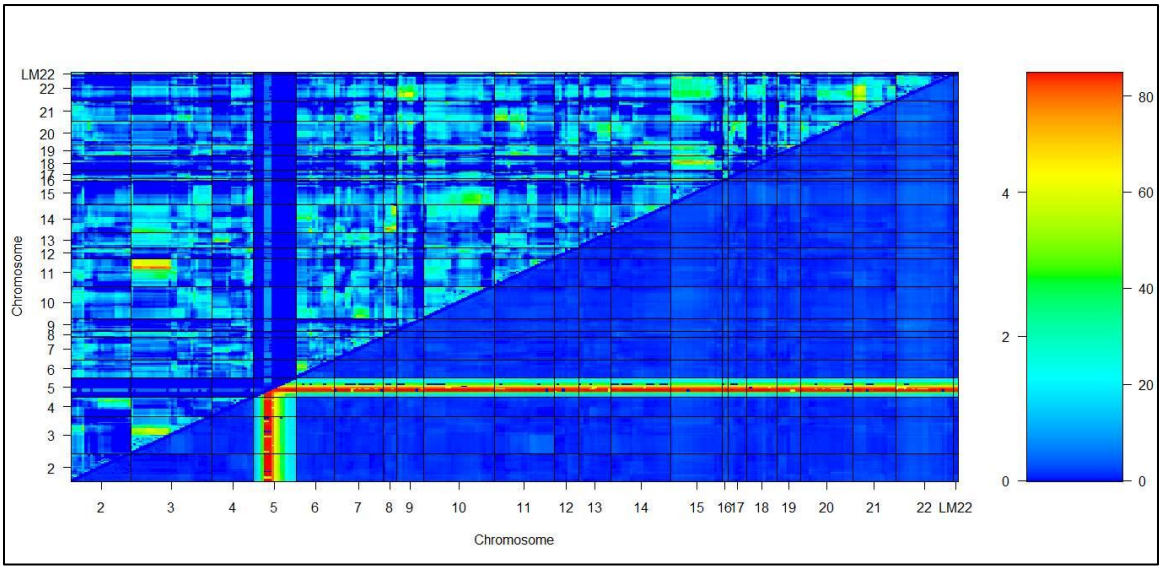

Figure D.3. Simultaneous two QTL scan for loci controlling orange blotch (OB) color pattern. Heat map plot represents log-likelihood score (LOD) for full 2-QTL model (below diagonal) and epistatic interaction between loci (above diagonal). LOD score increases with height on legend from 0 (blue) to highly significant (red). Horizontal and vertical axes represent chromosome number with location along chromosome increasing from left to right and bottom to top.

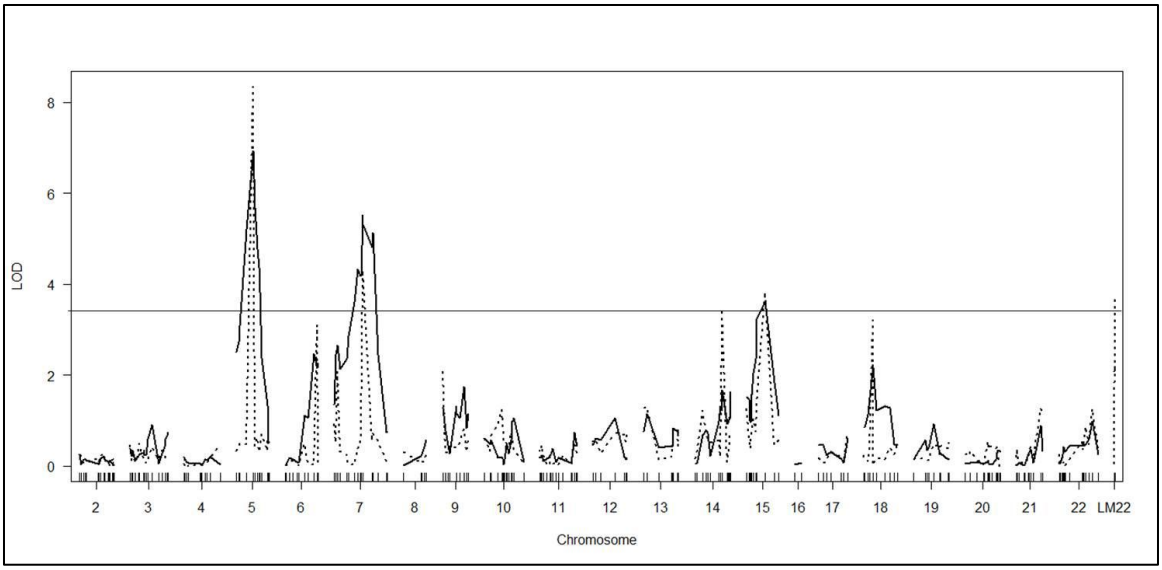

Figure D.4. Single QTL scans for loci controlling blue color. Solid line indicates standard interval mapping and dashed line indicates composite interval mapping. Horizontal axis denotes chromosome number with hash marks representing marker positions; vertical axis is log-likelihood score (LOD) of marker association with phenotype (see Methods). Horizontal bar indicates significant LOD score (3.5; p<0.05).

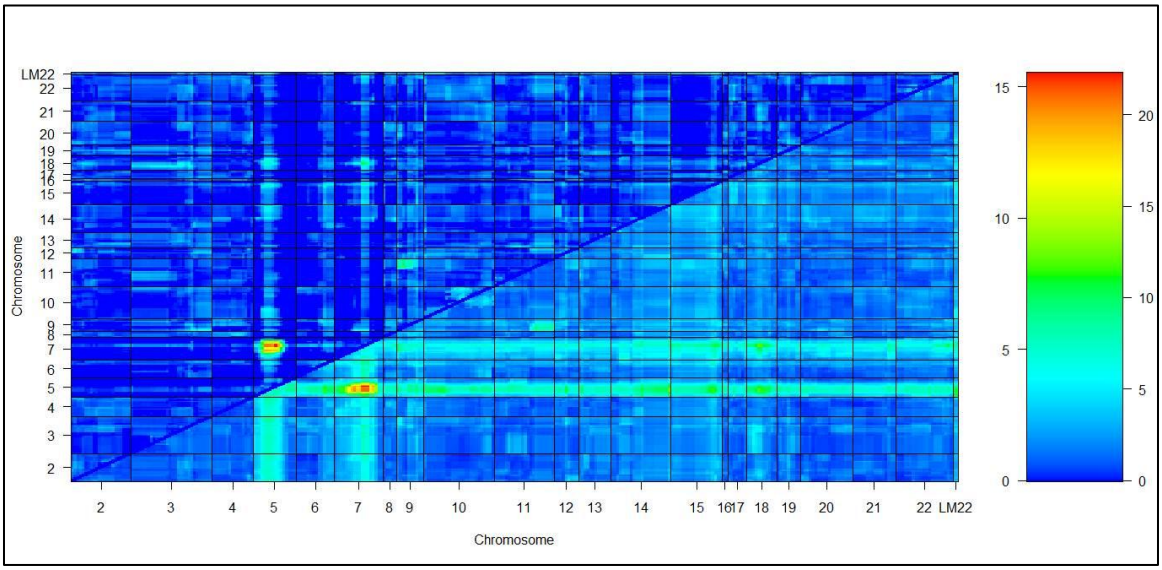

Figure D.5. Simultaneous two QTL scan for loci controlling blue color in Lake Malawi hybrid cross. Heat map plot represents log-likelihood score (LOD) for full 2-QTL model (below diagonal) and epistatic interaction between loci (above diagonal). LOD score increases with height on legend from 0 (blue) to highly significant (red). Horizontal and vertical axes represent chromosome number with location along chromosome increasing from left to right and bottom to top.