An Integrative Approach to Understanding Morphological Novelties: Anatomy, Development, Genetics, and Evolution of an Extreme Craniofacial Trait in East African Cichlids

Moira R. Conith
AN INTEGRATIVE APPROACH TO UNDERSTANDING MORPHOLOGICAL NOVELTIES: ANATOMY, DEVELOPMENT, GENETICS, AND EVOLUTION OF AN EXTREME CRANIOFACIAL TRAIT IN EAST AFRICAN CICHLIDS

A Dissertation Presented

by

MOIRA R. CONITH

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 1, 2018

Organismic & Evolutionary Biology
AN INTEGRATIVE APPROACH TO UNDERSTANDING MORPHOLOGICAL NOVELTIES: ANATOMY, DEVELOPMENT, GENETICS, AND EVOLUTION OF AN EXTREME CRANIOFACIAL TRAIT IN EAST AFRICAN CICHLIDS

A Dissertation Presented

by

MOIRA R. CONITH

Approved as to style and content by:

_________________________
R. Craig Albertson, Chair

_________________________
Courtney Babbitt, Member

_________________________
Jason M. Kamilar, Member

_________________________
Stephen D. McCormick, Member

_________________________
Paige S. Warren, Graduate Program Director
Organismic & Evolutionary Biology

_________________________
Patricia Wadsworth, Director
Interdepartmental Graduate Programs in the Life Sciences
DEDICATION

To my mother. Her intelligence and success as an engineer primed me to believe I could pursue a career in STEM.

To my husband. His love, support, and unwavering passion for science inspires me every day.

And to my daughter. Her curiosity and strength have reinvigorated these traits in me.
ACKNOWLEDGMENTS

I’d like to thank my co-authors for their hard work and commitment to this project: Yinan Hu, Andrew J. Conith, Maura A. Maginnis, Ilene Ea, Jacqueline F. Webb, and R. Craig Albertson.

I’d like to thank my committee for their guidance: Courtney Babbitt, Jason Kamilar, Stephen D. McCormick, and R. Craig Albertson.

I’d like to thank past and present members of the OEB community for creating such a collaborative and (most importantly) fun environment. I’d especially like to thank Penny Jaques and Maggie DeGregorio for their help navigating the ins and outs of graduate school.

I’d like to thank past and present members of the Albertson lab for doing cool science and making the lab a relaxed and supportive place to work. I’d especially like to thank Judy Bennett and Mary Packard for making sure the lab runs smoothly and providing exceptional fish care.

I’d like to thank the following people/organizations for their help with specific aspects of this project: Mary Catanese and Rolf Karls for help with immunohistochemistry; Madelaine Bartlett, Harry Klein, and Michael DeSalvo for their help in designing and analyzing qPCR data; Joy Richman for supplying beads for the TGFβ manipulation and critical advice for executing this experiment; Duncan Irschick and George Lauder for valuable discussions about skin biomechanics; Atsuo Iida for generously providing the adam12 mutant zebrafish; and Jacqueline F. Webb, J. Todd Streelman, Reade Roberts, The Royal Museum for Central Africa, Tervuren, Belgium, Cornell University Museum of Vertebrates, Ithaca, NY, The University of Michigan Musuem of Zoology, Ann Arbor, MI for providing cichlid specimens.

I’d like to thank my advisor R. Craig Albertson for being such a wonderful mentor over the years. This project wouldn’t exist without his ideas, enthusiasm, and guidance.

I’d like to thank my husband Andrew J. Conith for his love, support, and patience. And also for his stats consulting, available 24 hours a day.

Finally I’d like to thank these funding agencies for making this work possible: National Science Foundation; Organismic & Evolutionary Biology Graduate Program, UMass Amherst; The Graduate School, UMass Amherst.
ABSTRACT

AN INTEGRATIVE APPROACH TO UNDERSTANDING MORPHOLOGICAL NOVELTIES: ANATOMY, DEVELOPMENT, GENETICS, AND EVOLUTION OF AN EXTREME CRANIOFACIAL TRAIT IN EAST AFRICAN CICHLIDS

SEPTEMBER 1, 2018

MOIRA R. CONITH, B.S., SYRACUSE UNIVERSITY
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: R. Craig Albertson

Phenotypic novelties are an important but poorly understood category of morphological diversity that are often associated with elevated rates of diversification and/or ecological success. The aim of this dissertation is to explore a phenotypic novelty at many levels to contribute to our understanding of how these unique traits can arise (e.g., genetically, developmentally, and evolutionarily) as well as their ecological consequences (e.g., trait function). The extreme snout of the Lake Malawi cichlid fish Labeotropheus is used as a case study. The first chapter establishes the Labeotropheus snout as a model of phenotypic novelty by characterizing the gross morphology, genetic architecture, and growth patterns of the snout. Growth data show snout size diverges at early juvenile stages between Labeotropheus and a closely related ecological competitor. The second chapter reveals the tissue-level developmental origins and proximate molecular basis of the snout, providing evidence of the co-option of an existing genetic pathway in the evolution of this novelty. Specifically, histological staining shows that the exaggerated snout is composed of hypertrophied intermaxillary ligament which
forms a dynamic boundary with the surrounding loose connective tissue before anchoring to the epithelium. Further, protein manipulation, gene expression, and genetic mapping implicate the Tgfβ→Scx pathway in the development of this trait. We also identify and confirm adam12 as a novel candidate regulating ligament development. The third chapter expands the scope of this dissertation by using phylogenetic comparative methods to investigate snout shape in distantly related cichlid lineages. These data show that three cichlid lineages have converged on the exaggerated snout phenotype and models of evolution suggest that snout shape is under selection for feeding behavior. The fourth chapter explores intraspecific variation in snout morphology and environmental sensitivity. Comparative anatomy between two distinct populations of *Labeotropheus* coupled with lab experiments to manipulate feeding behavior show that both snout shape and the plastic response differs between populations. Overall, by taking an integrative approach and using the *Labeotropheus* snout as an example, the data presented in this dissertation provide a holistic understanding of a morphological novelty from proximate genetic mechanisms to ultimate ecological consequences.
# TABLE OF CONTENTS

| Chapter |
|------------------|------------------|
|                 | Page |
| **ACKNOWLEDGMENTS** | v |
| **ABSTRACT** | vi |
| **LIST OF TABLES** | xi |
| **LIST OF FIGURES** | xii |
| **CHAPTER** | |
| **1. GENETIC ARCHITECTURE OF AN EXAGGERATED CRANIOFACIAL TRAIT IN EAST AFRICAN CICHLIDS** | 1 |
| 1.1 Introduction | 1 |
| 1.2 Methods | 3 |
| 1.2.1 Anatomical Analyses | 3 |
| 1.2.2 Genetic Analyses | 5 |
| 1.3 Results | 6 |
| 1.3.1 Anatomical Analyses | 6 |
| 1.3.2 Genetic Analyses | 7 |
| 1.4 Discussion | 8 |
| 1.4.1 Developmental, genetic, and genomic bases of an exaggerated trait | 8 |
| 1.4.2 Candidate loci for flap development | 11 |
| 1.4.3 Roles for trait exaggeration in cichlid evolution | 13 |
| 1.4.4 Putative functions for an exaggerated flap | 14 |
| 1.5 Conclusions | 16 |
| **2. GENETIC AND DEVELOPMENTAL ORIGINS OF A NOVEL FORAGING ADAPTATION IN A LAKE MALAWI CICHLID GENUS** | 22 |
| 2.1 Introduction | 22 |
| 2.2 Methods | 24 |
| 2.2.1 Animals | 24 |
4. POPULATION-LEVEL DIFFERENCES IN THE ANATOMY AND PLASTICITY OF THE LABEOTROPHEUS SNOUT ................................................................. 68

4.1 Introduction .................................................................................................. 68
4.2 Methods ........................................................................................................ 71

  4.2.1 Animals ................................................................................................... 71
  4.2.2 Plasticity Experiments .............................................................................. 72
  4.2.3 Anatomy .................................................................................................. 73
  4.2.4 In situ Ligament Quantification ................................................................. 74

4.3 Results .......................................................................................................... 75

  4.3.1 Population-level differences in snout anatomy ...................................... 75
  4.3.2 Population-level differences in snout plasticity ...................................... 75

4.4 Discussion .................................................................................................... 76

  4.4.1 Snout Morphology is Population Specific ............................................. 76
  4.4.2 Phenotypic Plasticity Allows Convergence in Snout Morphology .......... 78

4.5 Conclusion ................................................................................................... 79

APPENDICES:

A: Supplementary information for chapter 1 ...................................................... 86
B: Supplementary information for chapter 2 ...................................................... 91
C: Supplementary information for chapter 3 .................................................... 105
D: Supplementary information for chapter 4 .................................................... 113

REFERENCES ..................................................................................................... 115
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Summary of significant QTL</td>
<td>21</td>
</tr>
<tr>
<td>3.1 Models of snout shape evolution</td>
<td>67</td>
</tr>
<tr>
<td>4.1 Results of t tests showing significant differences in snout anatomy between Likoma and Makanjila populations</td>
<td>85</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Sagittal sections of snout dissections</td>
<td>17</td>
</tr>
<tr>
<td>1.2 Relative snout size across different species of cichlid</td>
<td>18</td>
</tr>
<tr>
<td>1.3 Scaling relationship of snout size and body size</td>
<td>19</td>
</tr>
<tr>
<td>1.4 Kernel density plot of snout size in hybrids</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Morphology and development of the snout</td>
<td>39</td>
</tr>
<tr>
<td>2.2 Protein and gene expression in the snout</td>
<td>40</td>
</tr>
<tr>
<td>2.3 Morphological and genetic consequences of Tgfβ1 manipulation in the snout</td>
<td>41</td>
</tr>
<tr>
<td>2.4 Summary of QTL for snout size</td>
<td>42</td>
</tr>
<tr>
<td>2.5 QTL and gene expression data implicate adam12 as another candidate for snout size</td>
<td>43</td>
</tr>
<tr>
<td>2.6 µCT data demonstrate reduced ligament volume in adam12 zebrafish mutants</td>
<td>44</td>
</tr>
<tr>
<td>3.1 Snout morphospace reveals evolutionary, genetic, and ecological consequences of cichlid snout shape</td>
<td>64</td>
</tr>
<tr>
<td>3.2 Evolution of snout length and depth in Lakes Malawi and Tanganyika</td>
<td>65</td>
</tr>
<tr>
<td>3.3 Direct Red staining of dissected snouts shows species specific tissue-level anatomy</td>
<td>66</td>
</tr>
<tr>
<td>4.1 The novel snout of <em>Labeotropheus</em> endemic to Lake Malawi, Africa</td>
<td>81</td>
</tr>
<tr>
<td>4.2 Population level differences in the snout anatomy of <em>Labeotropheus</em></td>
<td>82</td>
</tr>
<tr>
<td>4.3 Differences in plasticity between Likoma and Makanjila</td>
<td>83</td>
</tr>
<tr>
<td>4.4 Reaction norms summarizing environmental sensitivity of snout phenotype in Likoma and Makanjila</td>
<td>84</td>
</tr>
</tbody>
</table>
CHAPTER 1
GENETIC ARCHITECTURE OF AN EXAGGERATED CRANIOFACIAL TRAIT IN EAST AFRICAN CICHLIDS

1.1 Introduction

The evolution of exaggerated traits has led to some of the most extreme and remarkable phenotypes among animals. Trait elaboration is commonly associated with processes of sexual selection wherein females typically prefer extreme male phenotypes as a way to assess mate quality (1). Notable examples include the giant horns of dung beetles, and the exaggerated tail feathers of birds of paradise. In these instances the trait is often sexually dimorphic, hypervariable between individuals, and has a steep allometric relationship with body size (2–4). Alternatively, trait exaggeration can also evolve under natural selection for increased locomotor performance (e.g., snake body, jerboa hindlimbs, bat wings), foraging efficiency (e.g., chameleon tongue, elephant trunk), or thermoregulation (e.g., fennec fox ears). In these instances, the trait is often sexually monomorphic, relatively invariable between individuals, and scales isometrically with body size (4).

The type of selection underlying an exaggerated trait may help to predict the underlying genetic and developmental mechanisms. For example, elaborated sexually selected traits are usually associated with male condition, and are thus highly sensitive to the environment (5). This environmental sensitivity may be mediated by hormones such that exaggerated traits develop with the onset of sexual maturity (e.g., in response to IGF signaling in beetle horns (6), or deer antlers (7)). Exaggerated traits under natural selection, on the other hand, are likely to be less
sensitive to the environment, and form earlier in development (e.g., bat wings). Regardless of the selective history, the fact that exaggerated traits are elaborations of existing traits suggests that the genetic architecture necessary to build the trait is already employed during development and only the level or timing of gene expression needs to be adjusted for the trait to grow. For example, digit elongation in bat wings is thought to be due to an increase in the level of Bmp2 expression during limb bud outgrowth (8). Additionally, increased appendage length in water striders is due to an expanded expression domain of Ubx (9). Given the modular nature of animal body plans and developmental processes, trait hypertrophy may be a mechanistically straightforward path to adaptation.

The cichlid fishes of East Africa are an ideal evolutionary model system to explore the genetic and developmental underpinnings of phenotypic innovation because of their high rates of speciation and diverse niche space. Within Lake Malawi, one genus in particular, Labeotropheus, includes two species with highly derived craniofacial morphologies (10), including an overdeveloped snout, which folds over on itself to form a flexible “flap” that rests along the rostral side of the upper jaw and superficially resembles a fleshy nose (see Figure 1.1). Both species in this genus possess this trait and have nearly identical ecologies – they both specialize on scraping attached filamentous algae from rocks. Observations in the field suggest, though it has never been experimentally tested, that the snout increases foraging efficiency and conserves energy by acting as a fulcrum to tear algae from rocks (11). While most other algae scraping species in the lake forage using a bite-and-jerk/twist mode (11, 12), Labeotropheus species employ a bite-and-
roll mode of feeding whereby they protrude their upper jaw ventrally toward the rocky substratum, take bites of tough filamentous algae, and then roll forward onto their snouts while retracting their upper jaw allowing algae to be leveraged off the rocks (11, Albertson personal observation). The very wide and square-shaped jaws of *Labeotropheus* likely enhances this behavior by providing a stable base upon which the flap/fulcrum can roll. This behavioral/morphological innovation, if true, may help to explain the ecological success of LF, resulting in both a widespread distribution around Lake Malawi and a larger body size than most other rock-dwelling species (11, 13). We are broadly interested in whether the flap has contributed to the ecological success of this species, and if so, why this exaggerated trait has only evolved once within the modern haplochromines. As a first step in this line of inquiry, revealing the underlying genetic and developmental mechanisms that lead to flap overgrowth will shed light on the evolutionary origins of this unique phenotype.

**1.2 Methods**

**1.2.1 Anatomical Analyses**

We measured flap size in developing (~1 month/1 cm standard length through adulthood/12 cm SL) *Labeotropheus fuelleborni* (LF; n=61) and *Tropheops* ‘red cheek’ (TRC; n=53) as well as in other adult Lake Malawi cichlids including *Labeotropheus trewavasae* (LT; n=3), *Cynotilapia afra* (CA; n=5), *Maylandia zebra* (MZ; n=4), *Tropheops* ‘red fin’ (TRF; n=10), and *Tropheops* ‘yellow chest’ (TYC; n=10). All fish were humanely sacrificed by prolonged exposure to ms-222 and/or
submersion in ice following IACUC protocol 2013-0101. Both LF and LT are obligate algal scrappers with conspicuous flaps. We do not report flap size for LT in the result section here. This is because LT has a dorsal-ventrally compressed and elongated body (14), which confounds the scaling relationship between flap size and body length (see below). However, when standardized by body depth, LT has a large flap that is similar in size to that of LF (data not shown). LT are also algal scrappers and nearly as widely distributed around the lake as LF, but far less abundant at any one locality and inhabit depths much greater than LF (13). The two *Labeotropheus* species also differ in their microhabitat; whereas LF feed mainly from the upper surfaces of rocks, LT tend to feed from the underside and vertical surfaces of rocks (11, 13). CA, on the other hand, is an obligate suction feeder. MZ and TYC are generalists that feed from both the rocky substrate and water column. TRC is a close ecological competitor of LF and LT, and feeds mainly on attached filamentous algae. However, unlike LF or LT it uses a bite-and-twist mode of feeding to accomplish this task. TRF is a benthic forager with a relatively wide jaw that will feed both by scraping rocks and sifting sediment. All fish were dissected to reveal a sagittal view of the flap, and imaged using a Leica DFC 450C digital camera mounted to a Leica M165 FC stereomicroscope. Flap size was measured using Image J as the distance between the dorsal most point of invagination to the ventral most tip of the flap (red bar in Figure 1.1).

We analyzed the scaling relationship between flap size and body size in developing LF and TRC using the SMATR package in R (15). After log transforming each size variable, we used standardized major axis to determine the line-of-best-fit.
We compared the slope of each species to a slope of 1 to determine if a hypothesis of isometry could be rejected. We also compared the slope and intercept across species to determine if the scaling relationships for this trait are consistent.

1.2.2 Genetic Analyses

We crossed a single LF female to a single TRC male to create a hybrid population in order to begin to elucidate the genetic basis of the flap. We measured flap size in F₁ (n=19) and F₂ (n=175) hybrids to investigate the genetic architecture of this trait. As a first step in this process, we followed Albertson et al. (2003a) to estimate the minimum number of effective genetic factors contributing to flap size by using the Castle-Wright Estimator (17–19). This equation utilizes means and variances of the trait in parental and hybrid lines to predict the minimum number of factors that underlie the difference in phenotype between the parental species.

Next, we used the F₂ hybrid population to perform a quantitative trait loci (QTL) analysis to detect the number and position of any loci associated with flap size. To accomplish this we used a previously constructed linkage map (20) to estimate marker position. The genetic map was built based on 268 F₂ individuals, but only 175 of these individuals were used for measuring the flap because this soft tissue trait requires a very precise dissection and only a subset were sufficiently intact to accurately measure flap size. Map construction and QTL analyses were performed in the R/QTL package of the R statistical language (21). Briefly, 268 F₂ hybrids and a panel of 40 wild-caught parental species were genotyped using
restriction-site-associated DNA sequencing (RAD-seq; (22)) which resulted in 42,724 high quality single nucleotide polymorphisms (SNPs). Markers that did not conform to Mendelian segregation in the F$_2$ were removed from the dataset, leaving 3,087 SNPs. During map construction, linkage groups were formed based on recombination frequency and checked based on their logarithm of odds (LOD) score and markers were ordered according to the shortest map length. SNPs with a high amount of missing genotype data and those not assigned to any linkage group were dropped from the dataset, resulting in a refined, dense genetic map with 946 markers across 24 linkage groups.

When more than two QTL underlie a quantitative trait, multiple QTL mapping is a robust and informative model fitting method (23). To this end, we identified significant QTL by building models of loci suspected to influence QTL under standard interval mapping and through automated backwards selection. The final model was fit using maximum likelihood in the Multiple QTL Mapping package of R/QTL (24). Epistatic interactions between loci were investigated using Haley-Knott regression in J-qtl (25). Permutation tests (n=1000) were used to estimate significance at the 0.05 level. Further details of map construction and QTL analyses in this cross can be found in Albertson et al. (2014).

### 1.3 Results

#### 1.3.1 Anatomical Analyses
LF individuals have a significantly larger flap size (p < 0.0001) than any other rock-dwelling cichlid species from Lake Malawi used in this study, including other filamentous algal scraping species (e.g., TRC, TRF), suction feeding limnetic species (e.g., CA), and generalist foragers (e.g., MZ, TYC; Figure 1.2). Additionally, we did not find evidence for sexual dimorphism in LF flap size (n=11 males; n=10 females; p=0.97). Among Tropheops species, we found that TRF has a marginally larger flap than either TRC or TYC (p=0.04). Similar to LF, TRF are benthic foragers characterized by relatively wide jaws. Whether a slightly larger flap is involved in foraging in this species remains to be investigated.

The flap develops isometrically in TRC (slope = 1.05, p = 0.33) whereas in LF the flap appears to be hyperallometric (slope = 1.31, p < 0.00001; Figure 1.3). Comparing these species, both the slope (p = 0.00007) and the intercept (p < 0.00001) are significantly different.

1.3.2 Genetic Analyses

Average flap size in both F₁ and F₂ hybrids was intermediate to that in LF and TRC. Compared to the F₁ hybrids, F₂ hybrids showed a broader range of flap size, which overlapped with parental distributions (Figure 1.4). These trends are consistent with a relatively simple genetic basis and additive mode of inheritance. Our Castle-Wright estimate of genetic factors that underlie differences in flap size based on patterns of inheritance was 4.83 (standard error 2.67). Our QTL analysis matched this estimate and identified 4 QTL significantly associated (at the genome-
wide 95% level) with flap size when standardized by body length. These QTL collectively accounted for 27.7% of the phenotypic variation in this trait (Table 1.1). For each QTL an additive mode of inheritance is supported over dominance. In addition, we detected no significant interactions among loci. Collectively, these data are consistent with a relatively simple genetic basis for flap size.

Evidence of selection acting on traits in a QTL analysis can be interpreted by looking at the direction of QTL effects (26), and, more formally, by performing a QTL sign test (27–29). While we did not detect enough QTL to perform a formal QTL sign test (a minimum of 6 are required; (27)), our QTL effects consistently show that the LF allele contributes to the development of a larger flap, whereas TRC alleles correspond to a smaller flap (Table 1.1). These data are consistent with the hypothesis that LF evolved large flaps due to strong and/or consistent directional selection.

1.4 Discussion

1.4.1 Developmental, genetic, and genomic bases of an exaggerated trait

We confirm that the fleshy, protruding snout (‘flap’) in LF is significantly larger than other cichlid species measured here, and contributes to its unique craniofacial shape. While other cichlids and fish in general have a homologous invagination of the skin and soft tissue, the exaggeration of this trait in *Labeotropheus* is novel within the modern haplochromine cichlids (i.e., Lake Malawi cichlids, Lake Victoria cichlids, and the Tropheini tribe from Lake Tanganyika (30). Analyses of the scaling relationship between flap size and body size show that the
flap develops isometrically (i.e., the slope does not differ from 1) in TRC but hyperallometrically in LF (i.e., the slope is greater than 1). Typically, hyperallometric traits are thought to have evolved under sexual selection and are characterized by steep scaling relationships with slopes 1.5-2.5 (31). In contrast, isometric traits are thought to be functionally relevant traits that must maintain the same relative size/shape of the trait over development in order to maximize performance. However, if foraging niche shifts over development and, for instance, the diet of a juvenile is different to the diet of an adult, the trait may not be constrained to isometry. We predict that this is the case with the flap in LF. For one, we did not detect evidence for sexual dimorphism in flap size in LF, which suggests that evolution under sexual selection is unlikely. In addition, the scaling relationship is hyperallometric (slope = 1.31), but not as steep as what is typically observed for exaggerated traits under sexual selection. Finally, like most rock-dwelling species, LF spend their early life history in the very upper reaches of the shoreline where larger/adult predators cannot reach them (12), and exist on a more generalized diet (32, 33). Thus, the shallow allometric relationship of the flap in LF may be due to animals taking progressively more filamentous algae as part of their diets as they get larger.

Both the Castle-Wright estimate and the QTL analysis are consistent in indicating a relatively small number of genetic factors underlie flap size (i.e., 4-5). This, along with the pattern of additive inheritance in hybrids suggests this trait has a relatively simple and tractable genetic basis. Furthermore, QTL analyses can reveal insights into the selective history of a trait by comparing the directionality of
the effects of each QTL (27). If alleles from each species contribute to both positive (i.e., increase trait value) and negative (i.e., decrease trait value) phenotypic effects, then genetic drift or stabilizing selection cannot be rejected as a possible mechanism of evolution. However, if the majority of alleles from one species contribute to increasing the trait value, while alleles from the other parent decrease trait value, then drift can be rejected and it is likely that directional selection is acting on the trait. This analysis has been used to show that directional selection led to speciation in Hawaiian crickets (34), adaptations of the oral jaws in Lake Malawi cichlids (29), and even that phenotypic diversification in plants and animals in general has largely been the result of directional selection (28). We were unable to perform a QTL sign test for the flap because at least six significant QTL (27) are required, but we can acknowledge the trend shown by our four QTL (similar to (35–37)), which consistently show that LF alleles contribute to a larger flap size. These data suggest that this trait has likely evolved in response to directional natural selection.

Our genetic map was anchored to the Lake Malawi cichlid genome, which allows QTL intervals to be directly compared to the underlying genomic sequence (e.g., (20)). Specifically, the 95% confidence intervals of the four QTL identified in this analysis collectively span over 60 Mb, and include 2,712 SNPs identified in natural populations of LF (n=20) and TRC (n=20), or roughly one SNP every 22kb. To search for candidate genes that might underlie flap development we focused on SNPs with an $F_{ST} \geq 0.57$, an empirical threshold of genetic divergence between cichlid genera estimated by Mims et al. (2010). This reduced the number of SNPs to 225. While combining QTL and outlier analyses from genome scans is becoming
more routine in the field, we are able to extend this approach given the nature of the
trait. Specifically, since LF possess the evolutionarily derived trait, we focused on
SNPs for which LF carried the derived allele when compared to six other cichlid
species (*Tropheops* ‘red cheek,’ *Maylandia* zebra, *Pundamilia nyerei*, *Astatotilapia
burtoni*, *Neolamprologus brichardi*, and *Oreochromis niloticus*). Evolutionary
relationships of these species are illustrated in Figure S1. Cichlid genomes may be
found at http://em-x1.gurdon.cam.ac.uk/. This reduced the number of SNPs to 74.
Finally, we examined genes adjacent to these 74 high SNPs, and identified their
functions. Genes that were close to the LOD peak within each QTL interval,
associated with high $F_{ST}$ value SNPs, and whose functions are consistent with tissue
hypertrophy are considered prime candidates. These data can be found in Table S1,
and are briefly discussed below.

**1.4.2 Candidate loci for flap development**

We cannot rule out the possibility that regions of the Malawi cichlid genome
are missing within our QTL intervals. It is also likely that our genome scans have
missed outlier SNPs within these regions. Nevertheless, this approach has proven
successful with respect to other traits segregating in this cross (20), and a few
notable candidates emerge for flap development here. The QTL on linkage group 7
(LG7) has a broad confidence interval, but its peak is centered over scaffold 0 at
position 8989175bp (i.e., marker c0.8989175, Table 1.1). The closest SNP with a
high $F_{ST}$ and for which LF carries the derived allele is at position 6854782bp, which
is 5’ to *bmp1*. This gene is notable as BMPs play key roles in craniofacial, connective
tissue, extracellular matrix and bone development. Bmp1 in particular has also been shown to regulate IGF1 signaling, which is significant given the broad roles for the Insulin/IGF pathway in exaggerated trait development (6, 7, 39, 40). The closest marker to the LOD peak for the QTL on LG8 corresponds to a SNP at 177665bp on scaffold 47. The SNP encodes a synonymous change within Protein kinase C-epsilon, and has an FST of 1.00. This gene is notable with respect to the flap phenotype as it plays important roles in fibroblast development, migration and adhesion to extracellular matrix (41). It has also been shown to mediate IGF1 signaling (42). The QTL on LG14 peaks at 7.656Mb on scaffold 8. At 8.842Mb there is a high FST intronic SNP (0.5796) within the gene beta-1 3-glucuronyltransferase 1, which encodes a protein that catalyzes proteoglycan synthesis (key components of extracellular matrix) (43), and is implicated in both facial clefting (44) and skeletal dysplasia (45) in humans. The LOD peak for the QTL on LG22 is centered on scaffold 40 (at position 3.023Mb). A notable outlier SNP on this scaffold (FST=0.7342, at 3987865bp) is located just 5' to the gene, GLIS family zinc finger 1 (glis1), which encodes a Kruppel-like zinc finger transcription factor (46). The expression of glis1 is upregulated by BMP signaling (47), and it can activate the Wnt/b-catenin pathway (48). These signaling pathways play key roles in myriad cellular (e.g., migration, proliferation, differentiation, apoptosis) and biochemical activities relevant to tissue hypertrophy, including links to IGF1 signaling (46, 49).
1.4.3 Roles for trait exaggeration in cichlid evolution

The evolutionary success of East African cichlids has been attributed to many factors, but roles for several “key innovations” have figured prominently in the literature, including the evolution of a modified pharyngeal jaw apparatus (50), as well as egg dummies and maternal mouth-brooding (30). Aside from such key innovations thought to have precipitated cichlid adaptive radiations as a whole, there are a number of novel exaggerated traits that are linked to the evolutionary success of species/lineages within each radiation. For instance an enlarged “forehead”, or nuchal hump, has arisen in both old and new world cichlids, develops from the overlying soft tissue of the skull as animals become reproductively mature, and is associated with reproductive behavior (51, 52). While hormonal control of nuchal hump development has been examined in new world cichlids (51), a genetic basis for this trait remains unknown. Another interesting soft-tissue novelty is the thick-lipped phenotype. Hypertrophied lips have evolved numerous times in both new and old world cichlid lineages, likely as an adaptation to facilitate foraging on benthic invertebrates (53). Notably, comparative transcriptomic analyses implicated different sets of genes in the development of hypertrophied lips in Central American and East African cichlids (53, 54). This observation combined with the recurrent nature of the phenotype suggests that there might be different genetic paths to evolving thick lips (although it cannot be ruled out that differences in gene expression might also be due to differences in the stage when tissues were harvested for RNA extraction).
Here we examine the developmental and genetic basis of another exaggerated soft-tissue trait. While the flap is limited to *Labeotropheus* among the modern haplochromine cichlids, there are at least two species within the older Ectodini tribe of cichlids endemic to Lake Tanganyika that possess a conspicuous flap that bears superficial resemblance to that of LF- *Asprotilapia leptura* and *Ophthalmotilapia nasuta*. Similar to LF, *A. leptura* lives and forages along the rocky, wave-swept shores of the lake. It also forages on attached filamentous algae and possesses many craniofacial adaptations that are similar to LF to facilitate this task, including a short, underslung lower jaw and closely spaced tricuspid teeth ((55); Concannon personal observation). *O. nasuta*, on the other hand, is a planktivore whose flap is sexually dimorphic. We postulate that in *Labeotropheus* and *A. leptura* the flap is evolving under natural selection for foraging efficiency and in *O. nasuta* the flap is evolving under sexual selection for mate choice. The evolution of conspicuous flaps in different cichlid lineages, possibly under different selection regimes, suggests a degree of evolutionary lability in this trait. Given the deep segregation of polymorphisms among East African cichlids (56), it would be interesting to know whether the same loci are implicated in the development of the same trait in different lineages and/or under different selective regimes.

1.4.4 Putative functions for an exaggerated flap

While the exact function of the flap remains elusive, we hypothesize that within the genus *Labeotropheus* it plays a mechanical role during feeding. This is based on several observations. First, the flap covers much of the anterior portion of
both the premaxilla and maxilla, and almost certainly influences the direction and/or extent of jaw protrusion in this species. Dissections of the flap reveal that it is comprised of rigid filamentous tissue that is firmly attached to the underlying bones. The upper jaw apparatus in cichlids (and indeed all acanthomorphs) is typically a highly mobile and protrusile structure (57–59) and we hypothesize that the flap in LF serves to effectively “lock” or “fuse” the upper jaw skeleton into place to facilitate the generation and/or absorption of force transmission through multiple skeletal elements as the animal forages on tough filamentous algae. The flap is also the first point of contact with the substrate during feeding ((11); Albertson personal observation), but doesn’t possess an abundance of sensory structures (i.e., taste buds, Concannon personal observation), which is consistent with it playing a mechanical rather than sensory role during feeding. Additionally, the flap may serve to increase foraging efficiency by decreasing the energetic demands of cropping algae from rocks. While most algivorous cichlids forage by taking a bite of algae and twisting their bodies to the side, LF may conserve energy by taking a bite and rolling forward onto the flap to leverage the algae off the rock (11). Finally, LF begins to develop a conspicuous flap soon after the yolk sac is completely absorbed and exogenous feeding becomes obligatory. Complete yolk absorption depends on several parameters such as temperature, but in *Labeotropheus* it occurs between 22-30 days post fertilization (dpf) in the wild (32) and 24-40 dpf in the lab (33). In the lab, flap size begins to diverge between LF and TRC around 27 dpf (data not shown). Thus, the onset of flap development coincides with the stage when LF fry begin foraging on their own. These coordinated
developmental events further implicate feeding mechanics as a primary role for the flap. Exactly how the flap functions during feeding, and the extent to which it confers a fitness advantage to individuals that possess it will be an interesting topic for future investigation.

1.5 Conclusions

Vertebrates display a broad array of phenotypic adaptations. Exaggerated morphologies, such as the flap in cichlids, represent one avenue for selection to drive adaptive evolution. By tinkering with existing phenotypes, novel traits can evolve which may lead to an increase in niche breadth. While some exaggerated traits evolve over and over again (e.g., feather elaboration in birds), others appear to be more restricted to certain lineages (e.g., the flap in cichlids). Uncovering the proximate genetic and developmental bases of these types of traits will allow us to better understand trait evolution and the forces (both extrinsic and intrinsic) that can limit evolution. In this way, the cichlid snout may serve as an informative model to better understand the evolution of exaggerated morphologies in general.
Figure 1.1 Sagittal sections of snout dissections

Sagittal sections of flap dissections of TRC (A) and LF (B). Flap size is indicated by a red bar. dnt = dentary; mx = maxilla; pmx = premaxilla.
Figure 1.2 Relative snout size across different species of cichlid
Relative flap size across different species of cichlid. LF have significantly larger flaps than any other cichlid (a, p < 0.0001). TRF have significantly larger flaps than other trophic species (b, TRC p = 0.02; TYC p = 0.04), but not MZ or CA (b,c).
Figure 1.3 Scaling relationship of snout size and body size
Scaling relationships of flap size to body size in developing LF (black open squares) and TRC (gray filled circles). Solid lines show lines-of-best-fit from standardized major axis. Dashed lines show isometric relationship with slope = 1.
Figure 1.4 Kernel density plot of snout size in hybrids
Kernel density plot showing the distribution of relative flap size of adult LF, TRC, and their F$_1$ and F$_2$ hybrids.
### Table 1.1 Summary of significant QTL

<table>
<thead>
<tr>
<th>LG</th>
<th>Position (cM)</th>
<th>95% CI (cM)</th>
<th>Nearest marker</th>
<th>LOD</th>
<th>PVE (%)</th>
<th>Mean phenotypic value</th>
<th>Add&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Dom&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LF/LF</td>
<td>LF/TRC</td>
<td>TRC/TRC</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>25-60</td>
<td>c0.8989175</td>
<td>3.24</td>
<td>8.17</td>
<td><strong>0.07971</strong></td>
<td>0.07846</td>
<td>0.00301</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>16-39</td>
<td>c47.177665</td>
<td>4.14</td>
<td>10.32</td>
<td><strong>0.07605</strong></td>
<td>-0.00664</td>
<td>0.09037</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>28-51</td>
<td>c8.7656076</td>
<td>3.64</td>
<td>9.13</td>
<td><strong>0.04180</strong></td>
<td>0.06590</td>
<td>0.05334</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>0-22</td>
<td>c40.3022819</td>
<td>4.48</td>
<td>11.12</td>
<td><strong>0.06591</strong></td>
<td>-0.00190</td>
<td>0.06818</td>
</tr>
</tbody>
</table>

<sup>a</sup> coefficient of additive effect  
<sup>b</sup> coefficient of dominant effect  
* indicates an increase in trait value.

LG = linkage group; CI = confidence interval; LOD = logarithm of odds; PVE = percent variance explained.  
All QTL are significant above 95% level (permuted 95% significance threshold LOD = 2.56).
2.1 Introduction

Species with novel phenotypes, which define the extremes of a collective morphospace, are a valuable resource for research across disciplines. In evolutionary biology, this phenotypic class can provide insights into the origins and constraints on morphological evolution (60–63). Engineers have long used novel phenotypes as inspiration for biology-based design and technology (64, 65). Such phenotypes can also serve as “evolutionary mutant models” of disease states, whereby the adaptive trait mimics a maladaptive condition in humans, and potentially provides insights into the genetic factors relevant for disease prediction and management (66). Finally, bizarre phenotypes serve to captivate the public and engage them in the context of education (67).

Novel phenotypes can arise as a dramatic reconfiguration of the body plan (e.g., seahorses), from the elaboration of existing structures (e.g., (8)), or as \textit{de novo} structures with no obvious precursor (e.g., (68)). Likewise, regulation at the molecular level can occur as the recruitment of existing genes and signaling pathways (e.g., through changes in timing, location, or amount of expression; (8, 69, 70)) or by the evolution of new genes (e.g., following gene/genome duplication events; (71, 72)). Unlike continuous phenotypic variation, novel morphologies highlight the opportunistic and flexible nature of evolution to act as both a tinkerer and an innovator (73). A major pursuit in evolutionary developmental biology is to
characterize the proximate genetic and developmental basis that can explain the origin of novel phenotypes. However, our understanding remains limited because most of our knowledge of development comes from model organisms, which are selected in part because of their conserved phenotypes (74, 75). Additionally, novel phenotypes often have ancient origins and/or occur in lineages that are not amenable to laboratory investigations. Here, we address the origin of a novel phenotype by identifying the genetic mechanisms underlying tissue-level changes of a unique craniofacial morphology in a genus of cichlid from Lake Malawi, Africa.

Cichlids are an iconic model system for the study of rapid and extensive craniofacial diversification (10, 76, 77). Within Lake Malawi, the genus *Labeotropheus* exhibits extreme craniofacial anatomy that defines the boundary of cichlid morphospace (10). Among other extreme phenotypes, the two species in this genus possess an exaggerated fleshy snout (e.g., Figure 2.1 A vs 2.1 B). We have previously shown that this elaborated snout folds in on itself, forming a flexible flap that extends over the upper lip, sometimes reaching the distal tip of the upper jaw dentition (78). It has been postulated that this protuberance enhances foraging efficiency by acting as a fulcrum to crop algae off rocks by leverage as an alternative to the energetically costly bite-and-twist mode of benthic feeding common in other cichlid species (11). In addition, prior genetic work suggests that this trait is under directional selection (78). *Labeotropheus* is also one of the most ecologically successful genera in the lake, as it has a cosmopolitan distribution and dominates the near shore rocky habitat (12). Indeed, although this trait is rare relative to the number of species that possess it (i.e., 2 out of over 800 cichlid species in Lake
Malawi), it is well represented in the lake given the ubiquity of *Labeotropheus*. Thus, as with novelties in general, the evolution of this putatively adaptive structure is coincident with a unique foraging mode and expanded ecological success.

We hypothesized that this unique structure develops as an expansion of the intermaxillary ligament, which runs laterally across the upper jaw and connects the left and right maxillary heads (59). Given the dynamic movement in fish skulls (79, 80), this class of tissue (e.g., ligaments) is likely to play important roles in foraging adaptations. However, relative to the bony skeleton, the evolution of ligaments is grossly understudied. Here we employ a combination of genetic and developmental approaches to test the hypothesis that elaboration of the intermaxillary ligament in *Labeotropheus* occurs via expansion of the Tgfβ→Scx candidate pathway (81–85).

Through these analyses, we also identified a novel candidate for ligament hypertrophy, *adam12*, and verify its roles in ligament development via functional analyses using the zebrafish model. Together these data support our main hypothesis, and contribute to an understanding of the genetic regulation of complex soft tissue morphologies and, more generally, of how phenotypic novelties evolve.

2.2 Methods

2.2.1 Animals

Cichlids used for experiments were derived from wild-caught Lake Malawi fish. They were reared and euthanized following standard protocols approved by the University of Massachusetts IACUC. Since both of the species in the *Labeotropheus* genus (i.e., *L. fuellborni*, LF, and *L. trewavasae*, LT) have exaggerated
snouts, we used both species depending on availability. Likewise, we used both
*Tropheops* ‘red cheek’ (TRC) and *Tropheops tropheops* (TT) in different experiments.
Fish are thus referred to as simply *Labeotropheus* or *Tropheops* throughout.

### 2.2.2 Histology

Histology was used to visualize and quantify tissue-level anatomy of the
snout. Animals for the developmental time series (LF, n=3; TRC, n=3) were collected
at 1.5 cm, 2.5 cm, and 4.0 cm standard length (SL). We have shown previously that
the snout scales with body size throughout the life of both LF and TRC, however
1.5 cm represents the time point at which the snout in LF is first apparent at a gross
morphological level (78). By 4.0 cm the snout is well-formed and clearly different in
the two species (78). While both species can reach 8–9 cm in lab, ~4.0 cm is also the
size when animals first begin to reach sexual maturity. In addition, animals for the
Tgfβ1 manipulation experiment (LT, n=3 Tgfβ1-treated, n=3 control; TT, n=3 Tgfβ1-
treated, n=3 control) were collected 7 days after bead implantation (see below) and
were 2.9–4.3 cm SL in size. In both experiments, serial sagittal sections were stained
with the Hall-Brunt Quadruple connective tissue stain (86). The amount of
intermaxillary ligament and loose connective tissue in a representative section were
quantified by measuring the area occupied by each of these tissues relative to a
standardized portion of the snout.
2.2.3 Immunohistochemistry

Immunohistochemistry was used to visualize presence and location of proteins in the snout of juvenile (~3.0 cm SL) LF. After blocking for non-specific staining, serial sagittal sections were incubated in primary antibody at 4°C overnight (Rabbit Anti-Scleraxis Polyclonal Antibody, Rabbit Anti-TGFβ1 Polyclonal Antibody, Rabbit Anti-ADAM12 Polyclonal Antibody; Bioss Antibodies) followed by incubation in a fluorescent secondary antibody at 4°C overnight (Goat anti-Rabbit IgG Alexa Fluor 488, Life Technologies). Finally, cell nuclei were counterstained with DAPI (Sigma).

2.2.4 Quantitative real-time PCR

qPCR was used to measure $tgf\beta1$, $scx$, 2 $smad4$ paralogs on linkage groups 7 and 11 (see below), and $adam12$ expression in juvenile (3.8–5.0 cm SL) LT (n=6), TRC (n=3), and TT (n=2) as well as to compare $scx$ expression following the Tgfβ manipulation experiment (LT, n=6 Tgfβ1-treated, n=6 control; TRC, n=5 Tgfβ1-treated, n=3 control; TT, n=3 Tgfβ1-treated, n=2 control). qPCR was also used to measure $scxa$ and $scxb$ in Adam12(-/-) mutant zebrafish (n=3) and several wild-type lines (caser n=1, AB n=2, EW n=2). RNA was isolated from homogenized snout tissue by the phenol chloroform extraction technique and standardized prior to reverse transcription. Levels of gene expression were measured using SYBR Green chemistry (Power SYBR Green Master Mix) and relative quantification was analyzed using the comparative $C_T$ method (87).
2.2.5 Tgfβ manipulation

Beads (Affi-Gel Blue Gel, Bio-Rad) soaked in either recombinant human Tgfβ1 protein (R&D Systems) or control buffer were implanted into the snouts of juvenile (2.9-5.3 cm SL) LT (n=9 Tgfβ1-treated, n=9 control), TRC (n=5 Tgfβ1-treated, n=3 control), and TT (n=6 Tgfβ1-treated, n=5 control). Because of the nature of this experiment, we used larger juvenile fish, but they were still in the range of those used for the histological analysis. A small incision was made in anesthetized fish parallel to the snout and a path was bored using a sewing needle to the tip of the snout. Four incubated beads were then pushed through the incision and guided to the tip of the snout, taking care to leave tissue as undisturbed as possible. All fish recovered from the surgery and were collected either 12 hours later for gene expression analysis or 7 days later for morphological analysis.

2.2.6 QTL mapping

Generation of the hybrid pedigrees and mapping strategies are described in Albertson et al. (20) and Parsons et al. (88). Further details may be found there or in Appendix B.

2.2.7 µCT scanning

We quantified the volume of ligamentous tissue in adam12 (-/-) zebrafish mutants using X-ray micro-computed tomography (µCT). Zebrafish do not possess an intermaxillary ligament. We instead focused on the cerato-mandibular ligament.
(CML), which runs from the medial surface of the mandible anteriorly to the medial surface of the ceratohyal posteriorly. We compared the volume of the CML between *adam12* mutant zebrafish (n=4) and wildtype casper zebrafish (n=4), the line in which the mutation was made. All fish were submerged for 5 hours in 2.5% Lugol iodine solution, which is absorbed into the soft tissues and acts as a contrast agent (89). We scanned all specimens at 5-micron resolution using an X-Tek HMXST 225 µCT scanner (Nikon Corporation) at 90kV and 75µA. We segmented out the CML using Mimics v19 (Materialise NV), and exported the 3D models before calculating volumes. We corrected the volumes by head depth and statistically compared the *adam12* to WT residual ligament volumes using a t-test in R (v3.4.0).

2.3 Results and Discussion

2.3.1 Anatomy and development of the snout

The snout in *Labeotropheus* is a complex and dynamic soft tissue structure compared to *Tropheops*, a closely-related near ecological competitor that lacks an exaggerated snout. Histological data showed two distinct, but interdigitating, tissue types within the snout – a Direct Red-positive organized connective tissue high in collagen (pink, Figure 2.1 C,D) and an Alcian Blue-positive loose connective tissue (blue, Figure 2.1 C,D). In both species, this basic configuration is maintained across the medio-lateral expanse of the snout, but for consistency we focused our comparisons between species and treatments on midline sections of this structure. The anatomy and organization of the Direct Red-positive tissue is consistent with
ligamentous tissue. To confirm this, we performed immunohistochemistry with an anti-Scleraxis antibody, which exhibited strong and consistent signal specific to this tissue (Figure 2.1 E,F). Further, this ligament appears to be the intermaxillary ligament, as it inserts on the left and right maxillary heads on either side of the ascending arm of the premaxilla (Figure 2.1 H,I), stretching medio-laterally across the upper jaw. Notably, this tissue differs from ordinary ligaments by invading the surrounding loose connective tissue (Figure 2.1 D) and anchoring to the overlying epithelium (Figure 2.1 D). Qualitatively, the degree of invasion and anchoring to the epithelium increases during ontogeny in *Labeotropheus* (Appendix B, Figure S1 A-C) whereas the interface between ligament and loose connective tissue is much less complex in *Tropheops* (Appendix B, Figure S1 D-F).

Ligaments and tendons are defined by their stereotypical connections: bone to bone in the case of ligaments, and muscle to bone for tendons. The few examples in the literature that describe departures from this anatomy highlight the functional significance of such novel arrangements. For instance, tendons in gecko feet insert directly onto toe pad integument, which increases stiffness and is vital for adhesion to surfaces during climbing (90). Likewise, stiffness of shark skin, which enhances hydrodynamics during swimming, is due to direct muscle to skin attachments (91). We speculate that the connection between the intermaxillary ligament and overlying skin may help to stiffen the exaggerated snout of *Labeotropheus*. A stiffer snout would provide a more robust fulcrum allowing *Labeotropheus* to pry algae from rocks in a manner that requires less energy compared to the bite-and-twist mode of feeding employed by most *Tropheops* species. The role of the integument
during feeding is largely unexplored, and more work is needed to determine the biomechanical implications of the interaction between the intermaxillary ligament and the epithelium in *Labeotropheus*. However, given the unique configuration of this tissue, we suggest that this would be a fruitful line of future research.

We also found that the snout exhibits dynamic growth. Histological data at three time points during juvenile development revealed marked differences between *Tropheops* and *Labeotropheus* in the pace and pattern of growth of both the intermaxillary ligament and loose connective tissue (Figure 2.1 J, Appendix B, Figure S1). In general, growth of both tissues is relatively modest in *Tropheops* with consistently more ligament than loose connective tissue at every stage, whereas in *Labeotropheus* growth is more vigorous, and at every stage there is more loose connective tissue than ligament. In addition, the pattern of snout growth is notable in *Labeotropheus* in that accelerated growth of the overlying connective tissue occurs early (i.e., 1.5-2.5 cm SL), and precedes accelerated growth of the ligament in larger fishes (i.e., 2.5-4.0 cm SL). The decoupling of growth rates between these tissues in *Labeotropheus* reveals a complex pattern of development. In addition, the accelerated growth of the loose connective tissue relative to the ligamentous tissue suggests that the inductive cues for ligament hypertrophy in *Labeotropheus* may arise from the loose connective tissue. Conversely, modest ligament growth in *Tropheops* may be due to more limited inductive signal from the growth-restricted loose connective tissue.
2.3.2 Molecular basis of exaggerated snout development

Tgfβ signaling is critical to ligament development via the transcriptional regulation of scleraxis (scx) (81–85), and we hypothesized that this pathway would be associated with snout development. Consistent with this, immunohistochemistry confirmed that Tgfβ1 is localized to the loose connective tissue of the snout (Figure 2.2 A), the putative source of the inductive signal for ligament overgrowth. We confirmed and quantified mRNA expression of tgfβ1 and scx in the snout using quantitative real-time PCR in Labeotropheus and Tropheops. We found that (1) both transcripts were expressed in the snout in both genera, (2) tgfβ1 was expressed at higher levels than scx in both genera, and (3) Labeotropheus had higher levels of tgfβ1 (T test, n = 6 Labeotropheus, n = 5 Tropheops, df = 8.7, t = 2.2, p = 0.05) and scx (T test, n = 6 Labeotropheus, n = 5 Tropheops, df = 6.9, t = 3.7, p = 0.008) compared to Tropheops (Figure 2.2 B).

The differential expression of these genes led us to experimentally manipulate this pathway to confirm its causal role in snout development. Specifically, we implanted either Tgfβ1-soaked beads or negative control beads into the snouts of juvenile Labeotropheus and Tropheops. We measured gene expression of the downstream target, scx, 12 hours after implantation (based on a time series analysis, Appendix B, Figure S3), and found that exogenous Tgfβ1 is indeed sufficient to increase levels of scx expression in both species (Tropheops, T test, n = 5 control, n = 8 Tgfβ1-treated, df = 7.3, t = -4.4, p = 0.003; Labeotropheus, T test, n = 6 control, n = 6 Tgfβ1-treated, df = 6.5, t = -4.4, p = 0.004; Figure 2.3 E). Additionally, both species responded with the same amount of increased scx expression
suggesting a conserved capacity to respond to Tgfβ1 signaling. Notably, this increase in scx expression also resulted in a tissue level response. Specifically, we assessed ligament morphology 7 days after bead implantation using histology and found that ligament size increased in Tgfβ1-treated animals in both species (*Tropheops*, T test, n = 3 control, n = 3 Tgfβ1-treated, df = 16.0, t = -2.5, p = 0.02; *Labeotropheus*, T test, n = 3 control, n = 3 Tgfβ1-treated, df = 14.1, t = -2.1, p = 0.05) whereas the amount of loose connective tissue did not change (Figure 2.3 A-D,F,G). In addition, the response of the ligament was dependent on the placement of the bead; there was more ligament growth and epithelial connections when beads were localized to the loose connective tissue overlaying the ligament compared to bead placement below the ligament (Appendix B, Figure S2). These data are consistent with our developmental time series and immunohistochemistry data and collectively demonstrate that (1) morphogenesis of the two tissues are molecularly decoupled, and (2) the inductive cue for ligament growth likely includes Tgfβ emanating from the overlying loose connective tissue.

2.3.3 Genetic basis for variation in snout size

Given the complexity of the snout phenotype and the wide range of interacting partners of Tgfβ, we wanted to know whether the causative genetic variants for snout size are associated with canonical members or known interactors of this pathway. To this end, we reanalyzed our published QTL data for snout morphology (78, 88) and performed additional analyses that capture different aspects of snout shape. We identified a total of 11 loci that contribute to either snout
length or depth (Figure 2.4, Appendix B, Table S1). Each of these QTL map to distinct regions of the genome, suggesting that snout length and depth are under independent genetic control. Likewise, separate loci were identified when fish were reared under alternate benthic and limnetic foraging environments, which supports the assertion that the environment can dictate regulation at the molecular level (88).

By anchoring our genetic linkage map to physical genomic sequence, we could determine whether QTL co-localize with members of the Tgfβ→Scx pathway or with other candidate genes for snout development. To this end, we mapped members of the canonical Tgfβ pathway (i.e., ligands, receptors, and intracellular Smads) and the transcriptional output for ligament development (i.e., scx). Notably, our QTL overlap with the intracellular transcriptional partner smad2 on linkage group 7, two paralogs of smad4 on linkage groups 7 and 11, as well as a scx paralog on linkage group 21 (Figure 2.4, Appendix B, Table S1). All four of these loci are associated with SNPs with outlier F_{ST} values when Labeotropheus are compared to either Tropheops specifically, or a more general subset of rock-dwelling cichlids (i.e., mbuna; Appendix B, Table S2). Consistent with the Tgfβ pathway being upregulated in Labeotropheus, expression data also show that the smad4 paralog on linkage group 7 (but not linkage group 11) is expressed at higher levels in Labeotropheus when compared to Tropheops (T test, n=6 Labeotropheus, n=5 Tropheops, df=9.0, t=2.4, p=0.04; Appendix B, Figure S4). These data provide additional genetic evidence of a role for Tgfβ→Scx in snout development, and identifies specific members of this pathway where causative mutations may be found.
Despite these cases of co-localization, many QTL for snout size do not overlap with Tgfβ pathway members, which provides an opportunity to identity other regulators of this pathway. To this end, we focused on a robust QTL on linkage group 13.1 with allele effects that maximize the difference in snout phenotype between parental genotypes (Appendix B, Figure S5, Table S1). We used a fine-mapping approach to narrow this QTL interval, and F\textsubscript{ST} data to identify fixed SNPs between *Labeotropheus* and *Tropheops* (Figure 2.5 A). Fine-mapping implicated two regions of peak association between genotype and phenotype (Figure 2.5 A, shaded regions), but only one of these regions contained variants with F\textsubscript{ST} -values of 1 (Figure 2.5 A, black circle). Notably, both of these SNPs fell within an intron of the gene *ADAM metalloproteinase domain 12 (adam12)*, a known regulator of Tgfβ signaling (92). Indeed, when *Labeotropheus* was compared to other rock-dwelling species, 21 additional intronic high F\textsubscript{ST} SNPs and one downstream SNP (Appendix B, Table S2) were identified, suggesting high potential for differential regulation in *Labeotropheus*.

Besides its role in Tgfβ signaling, Adam12 is known to mediate a host of complex cell/tissue behaviors, including proliferation, migration, hypertrophy, and invasion (93–95). These are consistent with the distribution and development of tissues in the snout of *Labeotropheus*. We used immunohistochemistry and an anti-Adam12 antibody to show that Adam12 is in fact expressed in the exaggerated snout of *Labeotropheus*, and is localized to the loose connective tissue, similar to the pattern of Tgfβ1 expression (Figure 2.5 B). We also quantified *adam12* expression in the snout of juvenile *Labeotropheus* and *Tropheops* using qPCR, and found that it is
expressed at relatively low levels compared to \(tgb\) or \(scx\), and at equal levels in \(Labeotropheus\) (i.e., \(2.28e^{-4} \pm 4.68e^{-5}\)) and \(Tropheops\) (i.e., \(2.31e^{-4} \pm 1.05e^{-4}\)) at this stage. Given the dynamic nature of this tissue as well as the complex regulatory capacity of Adam12 (92–95), we speculate that differences in expression may still occur, but at an earlier stage and/or in a transient manner.

To functionally explore the role of \(adam\) in ligament development, we took advantage of the recently described zebrafish \(adam\) mutant (96). Homozygous recessive mutants are viable and were reported as having no obvious morphological defects aside from reduced body size (96). Given our genetic mapping data, we hypothesized that animals lacking functional Adam12 would exhibit subtle differences in ligament development, and that defects would be apparent at both the transcript and phenotypic level. To this end, we measured mRNA expression of the two \(scleraxis\) paralogs in zebrafish, \(scxa\) and \(scxb\), in the pharyngeal skeleton of mutant and WT adult (~1yr) zebrafish. We found that \(scxa\) mRNA levels are lower in mutants compared to wildtype fish (T test, \(n=5\) wildtype, \(n=3\) mutants, \(df=5.9\), \(t=-2.3\), \(p=0.058\)). \(Scxb\) expression is lower than \(scxa\) in both mutant and wildtype animals, and shows similar (i.e., lower in mutants) but non-significant trends (T test, \(n=5\) wildtype, \(n=3\) mutants, \(df=4.0\), \(t=-1.5\), \(p=0.22\)) (Appendix B, Figure S6). We also measured the volume of a relatively large, functionally salient ligament that connects the mandible to the ceratohyal (cerato-mandibular ligament, Figure 2.6, Appendix B Figure S7), and found that fish lacking \(adam\) have smaller ligaments (T test, \(n=4\) wildtype, \(n=4\) mutants, \(df=4.4\), \(t=-4.2\), \(p=0.012\); Figure 2.6 B, Appendix B Figure S8). In addition, \(adam\) mutants exhibit
smaller tendons, and the bony processes on which tendons, ligament and other soft-tissues insert are more gracile and slender in mutants relative to those in wildtypes (i.e., arrow heads Figure 2.6, Appendix B Figure S9). Given that normal bone development and growth relies on mechanical input from the local soft tissue environment, these data suggest that more slender ligaments and tendons in \textit{adam12} mutants result in weaker mechanical input and the development of more gracile bone. Taken together, these results support our hypothesis that \textit{adam12} is necessary for normal ligament growth, and suggest new roles for this gene in craniofacial development.

Based on these data as well as the concordance between tissue anatomy in the snout and the known functions of Adam12, we hypothesize that Adam12 interacts with the Tgf\textbeta\textrightarrow Scx pathway to help mediate snout development and exaggeration in \textit{Labeotropheus}. Specifically, we predict that Adam12 increases Scx activity in the intermaxillary ligament, likely through Tgf\textbeta1 signaling (92) (Figure 2.5 C).

\textbf{2.4 Conclusions}

Evolution acts as both a tinkerer and innovator. This metaphor underscores the continuous and discontinuous nature of phenotypic variation among organisms. While continuous variation may characterize the majority of existing biodiversity, many examples of innovation (or saltation) also exist, especially in the fossil record (97). But what exactly is a phenotypic novelty? At what point does continuous variation cross over to become discontinuous variation? There is no clear consensus
in the literature on these points. Rather, much like the concepts of species or homology, the definition of novelty depends on context and level. Over 40 years ago, François Jacob suggested that evolution cannot produce novelties from scratch, but must work with what already exists (73). In other words, innovations at the organ level must arise by tinkering at the molecular level (73), and as we delve deeper into the molecular origins of different phenotypic novelties the exact nature of this tinkering is revealing itself. For instance, evolutionary loss of certain structures can be traced to the loss of genetic elements (98, 99). Alternatively, phenotypic gain has also been linked to the loss of genetic enhancers (100). Another recurrent theme in the study of novelties is the redeployment (co-option) of genetic/developmental networks in novel tissues/locations (70, 101). Finally, gene/genome duplication can facilitate the evolution of novelty by providing greater opportunities to tinker at the molecular level (71, 72). These examples suggest that there are many molecular paths to morphological novelty, and that saltatory evolution can arise from both large (e.g., genome duplication) and small (e.g., local deletion) mutational events. In this way developmental genetics is bringing new insights to an old debate between gradualists and saltationists by showing that a continuum of genetic changes can all lead to major shifts in morphology; however, more examples are needed.

While we are gaining unprecedented insights into the development and evolution of morphological variation, novelties are still under-represented in the literature. Here we describe a soft tissue novelty whose origins are associated with the recruitment of an existing signaling pathway (e.g., Tgfβ→Scx). However, it is worth noting that the evolution of the exaggerated snout may also have been
facilitated by larger-scale mutational events. When searching in public genomic databases (e.g., NCBI) teleost fishes appear to have undergone an expansion of *smad4*, with up to 4 paralogs in most teleosts (e.g., (102)). In theory, this ancestral gene duplication could lead to divergence in Tgfβ functioning in different tissues, which is consistent with recent evidence of selection on ancient gene duplicates in Malawi cichlids (103). It is therefore notable that our data show that three *smad4* paralogs co-localize to snout QTL, two of which define QTL peaks, and one that is differentially expressed in species with different snout sizes. Thus, it is possible that the evolution of the exaggerated snout in *Labeotropheus* is due, at least in part, to the molecular tinkering of *smad4* duplicates. While many questions regarding the development and evolution of this structure remain (e.g., there are several snout QTL with no obvious candidate genes), this work contributes to a growing body of literature by providing another example of the evolution of novelty by tinkering at the molecular level.
Figure 2.1 Morphology and development of the snout

Morphology and development of the snout. (A) An unremarkable, flat snout of most cichlids (and most fishes; represented here by *Tropheops* ‘red cheek’) compared to (B) the unique, exaggerated snout of *Labeotropheus* (represented here by *L. fueelleborni*). Images courtesy of Ad Konings. (C) Sagittal section of the snout in *Labeotropheus*. (D) Close-up of the black box in C shows the intermaxillary ligament (iml, pink) invading the surrounding loose connective tissue (lct, blue) and anchoring to the overlying epithelium (epi). (E-F, I) Immunohistochemical staining in *Labeotropheus* using anti-Scleraxis antibody (green) and cell nuclei counterstained with DAPI (blue). (E) Close-up of the upper black box in B shows the iml invading the surrounding lct. (F) Close-up of the lower black box in B shows the iml anchoring to the epi. (G) Schematic and corresponding (H) histology and (I) immunohistochemistry show the iml inserting onto the maxilla (mx). (J) Amount of lct and iml at three developmental time points in *Labeotropheus* and *Tropheops*. Each time point is represented by one individual with tissue measured in 2-3 sections close to the midline where the snout reaches its maximum size. Representative histological sections of both species at the 2.5cm SL time point are overlaid on the graph. Sections of the other time points are shown in SI Appendix, Fig. S1. Other abbreviations: pmx = premaxilla; SL = standard length.
Figure 2.2 Protein and gene expression in the snout

Protein and gene expression in the snout. (A) Tgfβ1 protein expression (green) in the lct of the snout of *Labeotropheus* via immunolabelling using anti-Tgfβ1 antibody and cell nuclei counterstained with DAPI (blue). (B) Quantitative PCR of *tgfb1* and *scx* expressed in *Labeotropheus* and *Tropheops* snouts. Asterisks indicate significant differences between species, p ≤ 0.05. Abbreviations as in Fig. 1.
Figure 2.3 Morphological and genetic consequences of Tgfβ1 manipulation in the snout

Morphological and genetic consequences of Tgfβ1 manipulation in the snout. Representative sagittal sections of the snout in (A) *Tropheops* negative control and (B) Tgfβ1-treated animals, and (C) *Labeotropheus* negative control and (D) Tgfβ1-treated animals. (E) Relative *scx* expression 12 hours after bead implantation in *Tropheops* and *Labeotropheus*. (F) Intermaxillary ligament area (iml, pink in A-D), and (G) loose connective tissue area (lct, blue in A-D) in *Tropheops* and *Labeotropheus* 7 days after bead implantation. Significant differences between negative control and Tgfβ1 treatments indicated by asterisks p ≤ 0.05 or double asterisks p < 0.005.
Summary of QTL results for snout size in hybrids between *Labeotropheus fueleborni* (LF) and *Tropheops* ‘red cheek’ (TRC). A genetic map is shown summarizing locations of QTL from the F₂ analyses (pink), F₃ benthic analysis (green) and F₃ limnetic analyses (blue). Vertical bars to the right of linkage groups represent the 95% confidence interval for each QTL and arrowheads represent QTL LOD peaks. In addition, the positions of canonical members of the Tgfβ pathway (i.e., ligands, receptors, and intracellular smads), interactors (i.e., adam12), and scleraxis paralogs are indicated on the linkage map. Raw data can be found in SI Appendix, Table S1.
Figure 2.5 QTL and gene expression data implicate adam12 as another candidate for snout size

QTL and gene expression data implicate *adam12* as another candidate for snout size. (A) Fine-mapping analysis of the QTL peak that maximizes allele effects (see SI Appendix, Fig. S5) showing average difference in snout size (i.e., average phenotypic effect, black line) between hybrids with homozygous genotypes at markers across scaffold 26 on linkage group 13.1 (i.e., across the QTL interval). Peaks (highlighted in gray) are regions where hybrids with the *Labeotropheus* genotype have the largest snouts and those with the *Tropheops* genotype have the smallest snouts. $F_{ST}$ data are also mapped onto the scaffold (blue dots). SNPs that fall above the blue dashed line exceed an empirical threshold for divergence between cichlid genera. There are two SNPs with $F_{ST}$-values of 1.0 that fall within a peak (black circle). Both of these markers are intronic in the gene *adam12*. (B) Adam12 protein expression is similar to Tgfβ in that it localizes to the lct of the snout in *Labeotropheus*. (C) Model of a proposed molecular pathway of exaggerated snout development. In the lct (blue) Adam12 may activate Tgfβ1, which in turn upregulates Scx in the iml (pink), leading to the expansion of this tissue. Abbreviations as in Fig. 1.
Figure 2.6 μCT data demonstrate reduced ligament volume in adam12 zebrafish mutants

μCT data demonstrate reduced ligament volume in adam12 zebrafish mutants. (A) Reconstructed 3D model in a wild-type zebrafish illustrating position of the cerato-mandibular ligament (cml) relative to mandible (den) and ceratohyal (ch). Maxilla (mx), premaxilla (pmx), and interopercle (iop) are also shown for perspective. (B) Reconstructed 3D model of the same structures in an adam12 mutant zebrafish. Note the marked reduction in the size of the cml (blue), as well as the coronoid process of the mandible (arrowheads).
CHAPTER 3

EVOLUTION OF A SOFT-TISSUE FORAGING ADAPTATION IN AFRICAN CICHLIDS: ROLES FOR NOVELTY, CONVERGENCE, AND CONSTRAINT

3.1 Introduction

New phenotypes arise when ecological opportunity is coupled with the release of constraints on a system. For instance, newly available niche space may lead to increased diversity through the release of cryptic genetic variation. Likewise, new genes or shifts in the developmental program (e.g., through heterotopy or heterochrony) can facilitate the evolution of novel structures. The ability of a species to capitalize on these dynamic processes is called evolvability (104).

The amount of ecological opportunities experienced by a population as well as the degree of constraints imposed during development can impact evolvability. Covariation between traits is one example of a type of constraint that can determine the direction and rate of evolutionary change (105). For instance, a suite of traits that are tightly integrated are expected to be evolvable in the direction of standing phenotypic variation. However, selection will be slow to move variation in other directions. Conversely, traits that are less integrated should be evolvable in many dimensions but at slower rates. In this way, evolvability is thought to shape patterns of variation at the microevolutionary level, but also contribute to macroevolutionary processes such as adaptive radiations and the evolution of novelty (reviewed in (105)).

Cichlid fishes are a model of adaptive morphological radiation and exhibit more biodiversity than any other vertebrate family. For example, in Lake Malawi
alone over 800 species have evolved in the last million years (106). Indeed, Lake Malawi and the other East African Great Lakes (i.e., Lakes Tanganyika and Victoria) are home to the largest species flocks, though cichlids can be found in freshwater rivers and lakes across Africa, South and Central America, India, the Middle East, and Madagascar (107). The ecological success of African cichlids, to name a few hypotheses, has been attributed to the versatility of their pharyngeal jaw apparatus (50), reproductive isolation due to species-specific color vision (108), dynamic niche landscapes produced by historical variability in lake climate (109), and the generation of novel genetic variation through extensive and recurrent hybridization (110). It is likely a combination of these and other factors (both intrinsic and extrinsic) that have shaped cichlid evolution.

Another contributing factor to cichlid diversity has been the evolution of extreme or novel phenotypes (53, 111, 112). For instance, a handful of cichlid species have evolved a structurally complex and exaggerated snout. The snout of most fish species is an unremarkable layer of soft tissue with a slight invagination in the skin just dorsal to the upper jaw. The folding of the skin presumably allows the skin to stretch during jaw protrusion. A unique elaboration of this structure is found within the genus *Labeotropheus* in Lake Malawi, and previous research has shown that in its exaggerated form, the snout increases in both the degree of invagination (i.e., snout length) and the amount of tissue contributing to its thickness (i.e., snout depth) (113). Additionally, the intermaxillary ligament, which runs mediolaterally across the upper jaw connecting the left and right maxillary heads and thus stabilizing the upper jaw, is enlarged in *Labeotropheus* and invades the surrounding
loose connective tissue before anchoring to the skin forming a novel ligament-epithelial boundary (113). It is thought that the protruding snout in *Labeotropheus* enhances foraging efficiency in these obligate algivores (11).

While genetic analyses of snout variation suggest that it has evolved due to directional selection (78, 113), it has not been examined in a broader phylogenetic framework. In fact, there is a notable paucity of soft-tissue traits (aside from muscles) in research on the evolution of foraging related phenotypes in general. Most work examining the origins of phenotypic variation in cichlids has focused on skeletal traits using either a phylogenetic (114–117), or genetic approach (76, 118–120). Here, we use morphological, genetic, and phylogenetic comparative methods to explore a soft tissue novelty – the exaggerated snout - across a sampling of Lake Malawi and Lake Tanganyika cichlids. We consider both intrinsic and extrinsic factors as we explore the idea that novel morphologies such as the exaggerated snout have contributed to the evolvability of cichlid lineages.

### 3.2 Methods

#### 3.2.1 Specimens

Lake Malawi cichlid specimens (n=1-3 individuals/species) were derived from wild-caught animals and either reared in the Albertson lab in a manner consistent with the University of Massachusetts Amherst institutional animal care and use committee, or were provided by colleagues J.T. Streelman (*Labeotropheus*...
trewavasae) and J.F. Webb (Tramitichromis sp., Aulonocara sturatgranti). Wild-caught Chilotilapia rhoadesii, Nimbochromis linni, Dimidiochromis compressiceps, and Cynotilapia afra were obtained from the aquarium trade. Hybrids between Labeotropheus fuelleborni (LF) and Tropheops red cheek (TRC) were derived from a single natural mating between a wild-caught female LF and wild-caught male TRC. F₁ progeny were intercrossed to produce F₂ hybrids (n=134). Lake Tanganyika specimens (n=3-5 individuals/species) were provided by The Royal Museum for Central Africa, Tervuren, Belgium; Cornell University Museum of Vertebrates, Ithaca, NY; and the University of Michigan Museum of Zoology, Ann Arbor, MI. Most of the Lake Tanganyika species we measured belong to the Ectodini tribe of cichlids. We targeted this group because, to our knowledge, it is the only other group in which an exaggerated snout has evolved. In particular the snout anatomy of two of its members - Asprotilapia leptura and Ophthalmotilapia nasuta – superficially resemble that of Labeotropheus in Lake Malawi. The Ectodini tribe is ecologically diverse, especially with respect to foraging and diet. Lake Malawi species were chosen based on ecological similarities to the Ectodini. More extensive sampling across Lake Malawi was not performed because of a lack of genetic data for some species and due to the poor resolution of the Malawi cichlid phylogeny (121).

3.2.2 Snout Anatomy

In all specimens, 4% paraformaldehyde fixed fish were dissected by making a midline incision down the front of the face to reveal a sagittal cross-section of the
snout. The right half of the snout was carefully dissected off the face and stored in 70% ethanol for further analyses. The left half of the snout, which remained intact, was imaged on a Leica DFC 450C digital camera mounted to a Leica M165 FC stereomicroscope. Snout length and depth were measured using ImageJ software (NIH). Snout length was measured as the distance from the ventral-most tip of the snout to the dorsal-most point of invagination. Snout depth was measured as the maximum distance from the skin to the proximal edge of the snout in the sagittal plane.

Tissue-level anatomy of the snout was revealed by whole-mount Direct Red staining of the dissected and isolated right half of the snouts of *Labeotropheus fuelleborni*, *Asprotilapia leptura*, and *Ophthalmotilapia nasuta*. The procedure was adapted from the Hall-Brunt Quadruple connective tissue stain (86) for its ability to differentially stain the intermaxillary ligament within the snout (113). The dissected snouts were rehydrated in a graded ethanol series ending with reverse osmosis water. Tissue was then incubated in 0.5% Direct Red for 3 minutes. Tissue was then quickly rinsed and imaged in 100% ethanol. Direct Red is readily water soluble, allowing tissue to be destained and restained as necessary.

### 3.2.3 Phylogenetic Tree Construction

We built a Bayesian, time-calibrated tree combining 33 Lake Tanganyika and 10 Lake Malawi cichlids to perform all of our phylogenetic comparative methods. We added 3 Lake Tanganyika cichlids and 10 Lake Malawi cichlids to a sample of 30
Lake Tanganyika cichlids originally compiled by Day et al. (122). We used two mitochondrial sequences to build our tree: \textit{NADH dehydrogenase 2 (nd2)} and the non-coding control region (cr). We generated time-calibrated trees using BEAST2 v2.4.5 (123), with two uniformly distributed calibrations on our tree based on the ages of Lake Tanganyika (9-12 million years (my)) and Lake Malawi (0.57-1 my). Further information on phylogenetic tree construction can be found in Appendix C. Specimen list and GenBank accession numbers for all taxa and sequences can be found in Appendix C, Table S1, XML file used for tree construction may be downloaded from Dryad (doi:10.5061/dryad.fh588ct).

\subsection*{3.2.4 Phylogenetic Comparative Methods}

Following tree construction we pruned all taxa from the tree with no associated snout data (10 taxa) and discarded snout data from taxa not present in the tree (3 taxa). Phylogenetic comparative methods were conducted on 33 total taxa, 10 from Lake Malawi and 23 from Lake Tanganyika. We first performed a phylogenetic size correction between snout morphology (length and depth) and body depth, and extracted snout length and depth residuals following the phylogenetic regression. We used the phyl.resid function in the Phytools R package to perform our size correction (124).

We found extreme sexual dimorphism in snout morphology in the Tanganyikan cichlid \textit{Ophthalmotilapia nasuta} (Appendix C, Figure S1). Indeed, the snout of \textit{O. nasuta} is known to increase with age in males and is largest in breeding,
territory-holding males (52, 125). Since the different sexes occupied such different regions of snout morphospace, we did not combine snout data to make a species average like we did for all other taxa. Instead, we used only male snout data to build the snout morphospace across the two lakes, in convergence tests, to reconstruct snout length and depth across the phylogeny, and to compare tissue-level snout anatomy. Conversely, we used only female snout data in analyses that relied on species ecology since presence of the snout isn’t associated with a shift in foraging, and the *O. nasuta* male is an extreme outlier when placed with other suction feeders.

To test whether cichlid taxa from different lakes occupy convergent regions of novel snout morphospace, we used the convnumsig function from the convevol R package (126). We tested whether four taxa residing in novel regions of snout morphospace (*Labeotropheus fuelleborni, L. trewavasae, Asprotilapia leptura*, and *Ophthalmotilapia nasuta*) could be considered convergent by calculating the number of transitions into each of those regions and comparing this to the simulated distribution of transitions.

We performed two different analyses to characterize rates of snout length and depth evolution. We first asked whether the rate of snout depth and snout length evolution was coupled or de-coupled in our cichlid sample. We conducted a likelihood procedure to compare evolutionary rates between the snout depth and snout length traits using the R function CompareRates.multTrait outlined in (127). Evidence for a difference in rates would suggest that distinct evolutionary and developmental processes control snout length and depth morphology. We then evaluated the rate of snout length and depth evolution between those taxa in Lake
Tanganyika and Lake Malawi. We used the R package mvMORPH to assess the Brownian rate of evolution ($\sigma^2$) from our snout length and depth data in each of the lakes using the functions mvBM and mvOU (128). Evidence for differences in the rate of snout length or depth evolution between lakes would suggest ample ecological opportunity and/or increased evolvability in the taxa inhabiting the lake with faster rates of snout evolution.

We compared the fit of five different evolutionary models to our snout length and depth data in a multivariate framework using the R package mvMORPH (128). We first examined the fit of snout morphology to two models: BM and a single-peak Ornstein–Uhlenbeck (OU) model. Support for the BM model would suggest that variation in snout morphology is uniformly increasing over time. Support for the OU model would suggest snout morphology is experiencing a constant pull toward an optimum value, indicative of stabilizing selection. These two models were compared to three multi-peak OU models whereby all taxa were assigned to habitat, diet, or feeding behavior categories based on published data (11, 52, 129). Support for a multi-peak model would suggest that selection is driving snout morphologies toward multiple optima based on a taxon’s habitat, diet, or feeding behavior. To account for uncertainty in character and phylogenetic history in our three multi-peak models, we stochastically mapped character transitions across a random sample of 100 trees from the Bayesian posterior distribution of time-calibrated trees (BPDT), and used the Stochastic Mutational Mapping on Phylogenies (SIMMAP) tool (130) from phytools (124) to generate 5 SIMMAP trees for each BPDT. We used the second-order Akaike’s information criterion (AICc), which
Corrects for small sample sizes, to select the best evolutionary model, which is favored over a competing model if the difference in AICc score is >2 units (131). We calculated a mean AICc score for each of the three multi-peak models. R code available on Dryad (doi:10.5061/dryad.fh588ct).

3.3 Results

3.3.1 Cichlid phylogeny

The relationships and divergence times of our Lake Malawi and Lake Tanganyika cichlids (Appendix C, Figure S2) are broadly congruent with previously published trees (e.g., (122, 132–135)). In general, nodes in Lake Tanganyika were well supported with high posterior probability (%PP). Lake Malawi cichlids divided into two clades, the rock-dwelling mbuna and the sand-dwelling utaka, both with high support (PP>0.95%). Within the mbuna, node support was lower (PP<75%). This uncertainty within Lake Malawi is consistent with many previous studies documenting incomplete lineage sorting and extensive gene flow within this group (121, 136–138).

3.3.2 Snout shape is convergent across lakes

The overall pattern of snout variation is strikingly similar between lakes (Figure 3.1 A). While variation in snout length is more extensive in the Ectodini from
Lake Tanganyika, Lake Malawi species have evolved similarly deep snouts. Notably, in both lakes, length and depth are phenotypically decoupled such that animals with similar snout lengths can differ dramatically in snout depth.

The results of the evolutionary analysis identified four taxa that have converged on exaggerated snout shape - *Labeotropheus fuelleborni* and *Labeotropheus trewavasae* from Lake Malawi, and *Asprotilapia leptura* and *Ophthalmotilapia nasuta* from Lake Tanganyika (Figure 3.1 B). By explicitly testing for convergent shifts, we found a significant level of convergence on those snout morphologies that lie within this region (convnumsig; P<0.01). This demonstrates that taxa in both Lake Malawi and Lake Tanganyika have converged upon similar exaggerated snout morphologies. We could not formally test for convergence between *A. leptura* and *O. nasuta* but their distant positions in the phylogeny and ancestral state reconstruction of their most recent common ancestor (Figure 3.2) suggests snout shape evolved independently in these two lineages as well.

While convergence is evident at the gross anatomical level, the tissue level anatomy of the snouts differed in conspicuous ways. We used whole-mount Direct Red staining in *L. fuelleborni, A. leptura*, and *O. nasuta* to visualize and quantify the intermaxillary ligament (IML) within the snout (Figure 3.3). While the IML is present in all three taxa, the proportion of tissues that make up the snout differ. In *L. fuelleborni* and *A. leptura*, at least 80% of the snout is composed of the IML and supporting loose connective tissue (LCT), with a relatively small amount of snout volume dedicated to skin and other tissue. Conversely, in *O. nasuta* the IML and LCT combine for only 50% of total snout volume, with the other half consisting of
hypertrophied skin and other underlying tissue(s). This difference is consistent with a kinematic role for the snout in *L. fuelleborni* and *A. leptura* (e.g., to facilitate algae scrapping), whereas the function of the snout in *O. nasuta* is likely not dependent on tissue expansion via the IML (e.g., it is likely used to attract females; see below for expanded discussion).

### 3.3.3 Decoupling of snout length and depth

In addition to snout length and depth being phenotypically decoupled, the evolutionary histories of these traits are also independent (Figure 3.2). That is, snout length and depth appear to coevolve in some taxa (e.g., *L. fuelleborni*, *L. trewavasae*, *A. leptura*, and *O. nasuta*) whereas they are decoupled in other taxa (e.g., *Nimbochromis linni*, *Xenotilapia caudafasciata*). The ancestral state of snout shape is intermediate in length, but relatively shallow in terms of depth (Figure 3.2).

We also find evidence for significant decoupling between the rate of snout length and depth evolution (AICc Constrained = 182; AICc Observed = 173; p < 0.01), with length evolving much faster (σ² = 0.76) relative to depth (σ² = 0.30) across all cichlids. When we assess the rate of snout evolution between lakes we find snout morphology in Lake Malawi taxa is evolving much faster for both length and depth (length σ² = 1.41; depth σ² = 0.64) relative to Ectodini taxa from Lake Tanganyika (length σ² = 0.47; depth σ² = 0.15).

Genetic data also provide evidence that snout length and depth are under independent control. Previous research has shown that snout length and depth map
to distinct regions of the cichlid genome (113). In addition, patterns of inheritance are distinct between snout length and depth. Specifically, in an F2 population of hybrids between *L. fuelleborni* and *Tropheops* red cheek (an ecologically similar species that lacks an exaggerated snout), the full range of snout lengths are recapitulated (x-axis, Figure 3.1 C). Alternatively, this recombinant population hasn’t recovered the full depth of *L. fuelleborni* (y-axis, Figure 3.1 C). Since we do not find evidence for dominance of *Tropheops* alleles at known QTL (78, 113), this pattern suggests a more complex genetic basis for snout depth variation.

3.3.4 Models of evolution suggest the snout is under selection for foraging behavior

Feeding behavior was the best-supported evolutionary model (Table 3.1). Feeding behavior garners very high model support (AICc weight >0.99), and is more than 10 AICc units away from the next model, diet. When excluding the sexually dimorphic *O. nasuta* male, all cichlids with highly exaggerated snout depths are specialized algae scrapers, while all those with exaggerated lengths are sand sifters (Figure 3.1 D). Support for the multi-peak OU model of feeding behavior indicates selection is likely pulling snout shape toward separate optima depending on how fish feed.
3.4 Discussion

3.4.1 Does convergence at the organ level suggest convergence at the molecular level?

Three lineages of cichlids from two of East Africa’s Great Lakes have independently converged on the same exaggerated snout. Specifically, long, deep snouts have evolved in the *Labeotropheus* genus from Lake Malawi as well as *A. leptura* and *O. nasuta* from Lake Tanganyika (Figure 3.1 B). Similarly, many other ecologically relevant traits have independently evolved in cichlids endemic to these two lakes, and in fact convergent evolution is considered a hallmark of East African cichlids (117, 139). Why is there so much convergence? Part of the answer may lie in the evolutionary origins of the cichlid radiation. Lake Tanganyika, the oldest of the three Great Lakes, is thought to have acted as an evolutionary reservoir and precipitated subsequent radiations in the other lakes and surrounding rivers (30). One consequence of this is deeply segregating polymorphisms in African cichlids, which raises the possibility for convergence at the molecular and/or genetic levels (38, 140). Few studies have investigated the molecular basis of convergent phenotypes in distantly related cichlid taxa, but there is some evidence in cichlids that convergence can occur both phenotypically and genetically (53). This phenomenon has also been observed in other taxa (e.g., (141–143)), though it is not always the case (e.g., (144–146)).

We have shown previously that the evolution of an extreme snout in *Labeotropheus* involved hypertrophy of the intermaxillary ligament via the
Transforming growth factor β (Tgfβ) signaling pathway (113). Given that expansion of the intermaxillary ligament is evident in the snouts of both *Labeotropheus* and *A. leptura* (Figure 3.3), this system seems poised to address the question of whether phenotypic convergence in a novel morphology is coupled with molecular convergence.

### 3.4.2 Snout length and depth are morphologically, genetically, and evolutionarily decoupled

In both Lake Malawi and the Ectodini tribe from Lake Tanganyika, snout length and depth are morphologically decoupled such that long snouts may be associated with those that are either deep or shallow (Figure 3.1 A). Modularity within a structure is common in functionally relevant traits, as is seen in the cichlid mandible (147), mammal dentary (148), and moth wings (149), and has important implications for trait evolvability (104).

In addition to morphological decoupling between these two dimensions of snout shape, we also find evidence of genetic modularity. The snout morphospace is skewed in F2 hybrids of two Lake Malawi species that differ in snout shape - *Labeotropheus fuelleborni* (LF) with a long and deep snout, and *Tropheops* red cheek (TRC) with an intermediate snout (Figure 3.1 C). Specifically, hybrids are intermediate in snout length between the two parental species (though the mean is shifted toward TRC), but the distribution is strongly biased toward TRC for snout depth. This pattern of inheritance suggests the genetic basis of snout length and
depth are distinct, which is consistent with previous data showing that QTL for snout length and depth do not overlap (113). In addition, given that hybrids have not recapitulated the full snout depth of LF, and snout QTL do not exhibit dominance for TRC alleles (78, 113), the development of a deep snout appears to require rare or a specific combination of alleles.

The pattern of genetic data showing a relatively more difficult path to developing a deep snout is consistent with our morphological data showing more continuous variation in snout length, compared to variation in snout depth, which appears more discontinuous (Figure 3.1 A). We do not believe this pattern is due to limited sampling, as deep snouts are rare among African cichlids (11, 52). Rather these patterns suggest that genetic or developmental constraints are influencing the evolution of this character complex, with more limited evolvability of snout depth compared to snout length. This assertion is supported by our phylogenetic analyses, which show that within both lakes, snout length is evolving faster than depth.

That Lake Malawi cichlids have achieved similar snout depth diversity to that in the Ectodini tribe from Lake Tanganyika over a much shorter timespan confirms that snout depth in Malawi is under strong natural selection. It is also interesting to note that even though snout shape is evolving faster in Malawi than Tanganyika, the four taxa with extreme snout depths have converged on similar phenotypes (Figure 3.1 B). That is, there may be an upper limit on snout depth. This may be because snout size has reached its optimum and a deeper snout would inhibit performance. For instance, since the snout is mechanically integrated with the upper jaw via the
intermaxillary ligament, there may be a functional constraint preventing growth beyond what is observed in *Labeotropheus* and *A. leptura*.

### 3.4.3 The snout is under selection for foraging behavior

Models of evolution implicate the snout in feeding behavior, which is intuitive given its proximity to, and mechanical integration with, the upper jaw (Table 3.1). Indeed, foraging eco-types occupy distinct regions of snout morphospace (Figure 3.1 D), though more extensive sampling is required to determine the extent to which snout shape is reliably associated with foraging behaviors. Nonetheless, this analysis provides experimental evidence of the functional significance of snout shape, and offers further opportunities for study. For instance, snout length may be related to the ability to protrude the upper jaw such that it limits the amount of protrusion, and the degree of protrusion is likely to influence ability to capture prey (58). Given this, future studies could investigate the relationship between feeding performance and snout length, as well as the association between snout length and the degree of jaw protrusion. Data like these would contribute to a more holistic understanding of the complicated dynamics at play during fish feeding as well as enhance the sparse literature on soft tissue traits in feeding biomechanics.

Snout depth may also impact feeding performance. With the exception of *O. nasuta* (discussed below), the taxa with exaggerated snouts (i.e., the two *Labeotropheus* species and *A. leptura*) are all algae scrapers in rocky habitats (Figure
3.1 D, Appendix C Figure S3). This is consistent with an untested hypothesis suggested to explain the large snout of *Labeotropheus* – that this protuberance functions as a fulcrum to tear algae off rocks by leverage rather than the energetically costly bite-and-twist mode of feeding common to other aligvores (11).

Tissue-specific staining of the snout supports this hypothesis. The intermaxillary ligament and the surrounding loose connective tissue in both *L. fueleborni* and *A. leptura* comprises at least 80% of the snout and includes ligamentous projections and attachment to the overlying skin (Figure 3.3). While the functional significance of this arrangement has not yet been specifically explored, given the important roles for ligaments in fish feeding mechanics (150), it seems likely that the hypertrophied ligament is a key contributor to feeding performance in these obligate algae scrapers.

3.4.4 The curious case of *Ophthalmotilapia nasuta*

*O. nasuta* is a planktivorous suction feeder that lives in an intermediate habitat at the interface between sand and rocks. There is extreme sexual dimorphism in the snout (Appendix C, Figure S1) such that it continues to grow with age in males and is especially large in breeding, territory-holding males (52, 125). Since there is no evidence of a difference in feeding behavior between males and females, we excluded the exaggerated snout of the *O. nasuta* male from our evolutionary models and instead relied on the females to calculate a species average. Given the natural history, feeding ecology, and sexual dimorphism of *O.
nasuta, we propose that the snout evolved by sexual selection for mate choice in this lineage.

The histological staining of this species also suggests the snout does not engage in the same functional roles as it does in Labeotropheus and A. leptura. While the ligament and loose connective tissue makes up 80-95% of the snout in these two algae scrapers, it only accounts for 50% of snout volume in male O. nasuta, while skin and other tissue make up the remainder of the snout (Figure 3.3). It’s possible that if the snout is an organ used for mate choice but is not under selection for feeding performance, O. nasuta may have overcome the genetic/developmental constraint imposed on expanding snout depth by using skin and other tissue to increase snout depth rather than ligament. Additionally, since O. nasuta is a suction feeder, there may have been selection against hypertrophy of the intermaxillary ligament if it could interfere with jaw protrusion, leaving fewer options for tissues that could expand to create the exaggerated snout.

3.5 Conclusions: Evolvability of a rare exaggerated phenotype

Trait evolvability is maximized when ecological opportunity is coupled with the relaxation of constraints, whether genetic, developmental, or functional. Morphological, genetic, and evolutionary data presented here suggest the snout is a modular trait with length and depth evolving independently and with distinct genetic bases. Decoupling between length and depth allows the snout to evolve in multiple dimensions, resulting in a morphospace that is not a simple linear
relationship. In particular, we propose the following scenario: The modular nature of the snout has allowed this phenotype to evolve in many dimensions, resulting in the scattered morphospace shown here. While snout length varies along a continuum, the discontinuous distribution of snout depth coupled with a complicated genetic architecture and slower rates of evolution suggests that evolvability is more limited along this dimension. However, *Labeotropheus* from Lake Malawi and *A. leptura* from Lake Tanganyika have overcome this putative constraint and have convergently evolved a novel snout morphology – i.e., extreme snout depth due to a hypertrophied intermaxillary ligament with complex anatomical architecture. The development of an enlarged snout is proposed to enhance foraging efficiency in shallow, wave-swept habitats, which likely resulted in increased fitness, and ultimately the fixation of alleles for this novel trait (as is seen in (78, 113)). Finally, we suggest that the ecological expansion into this foraging niche has facilitated the cosmopolitan, lake-wide distributions of both *A. leptura* and *Labeotropheus* (11, 52). While the evolution of other phenotypes is clearly also associated with the ecology of these lineages (e.g., ventrally oriented oral jaws, dorsally displaced eyes, elongated guts, etc.), an exaggerated snout appears to be a novel feature that is key to the success of both lineages. Similar to other instances whereby the evolution of novelty is associated with evolutionary success (151–156), we propose that the success of these species is due, at least in part, to the increased performance and fitness consequences associated with an exaggerated snout.
Figure 3.1 Snout morphospace reveals evolutionary, genetic, and ecological consequences of cichlid snout shape

Snout morphospace reveals evolutionary, genetic, and ecological consequences of cichlid snout shape. (A) Phylogenetically- and body size-corrected snout morphospace in species from Lake Malawi (black circles) and Lake Tanganyika (red triangles). Inset photo showing exaggerated snout of *Labeotropehus fuelleborni* (courtesy of Ad Konings at cichlid press). (B) Phylomorphospace of snout shape showing convergence between taxa surrounded by the purple ellipse. Taxa are represented by black circles and nodes on the phylogeny are represented by smaller grey circles. AL = *Asprotilapia leptura*, LF = *Labeotropehus fuelleborni*, LT = *Labeotropehus trewavasae*, ON = *Ophthalmotilapia nasuta*. (C) Body size-corrected snout shape in *Labeotropehus fuelleborni*, *Tropheops* red cheek, and their F2 hybrids. Note the hybrids range most of snout length but do not access the same depths as *Labeotropehus*. (D) Phylogenetically- and body size-corrected snout morphospace in Lakes Malawi (circles) and Tanganyika (triangles) colored by feeding behavior. The *Ophthalmotilapia nasuta* male with the large snout was excluded from this analysis because the trait is extremely sexually dimorphic in this species and confounds the relationship between feeding and snout shape.
Figure 3.2 Evolution of snout length and depth in Lakes Malawi and Tanganyika

Evolution of snout length and depth in Lakes Malawi and Tanganyika. Snout shape is predicted for ancestral taxa along the phylogeny, represented by color where cooler colors represent smaller snouts and warmer colors represent larger snouts. Bolded taxa are the four species that have converged on snout shape. Note snout length and depth can covary, and be either short and shallow as in (a) or long and deep as in (b and d). It can also be decoupled as in (c), which shows a species with a long and shallow snout.
Direct Red staining of dissected snouts shows tissue-level anatomy differs between *Labeotropheus fuelleborni* (LF), *Asprotilapia leptura* (AL), and *Ophthalmothilapia nasuta* (ON). In LF and AL, where the snout is under selection for algae scraping, the intermaxillary ligament (IML) and supporting loose connective tissue (LCT) make up 80-95% of the snout. Conversely, these tissues make up only 50% of the snout in ON.
Table 3.1 Models of snout shape evolution

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>AICc</th>
<th>dAICc</th>
<th>wtAICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>OU - Feeding Behavior</td>
<td>-31.560</td>
<td>99.355</td>
<td>0.000</td>
<td>0.996</td>
</tr>
<tr>
<td>OU - Diet</td>
<td>-37.017</td>
<td>110.269</td>
<td>10.914</td>
<td>0.004</td>
</tr>
<tr>
<td>OU</td>
<td>-51.307</td>
<td>121.141</td>
<td>21.786</td>
<td>0.000</td>
</tr>
<tr>
<td>OU - Habitat</td>
<td>-44.213</td>
<td>124.662</td>
<td>25.307</td>
<td>0.000</td>
</tr>
<tr>
<td>BM</td>
<td>-75.488</td>
<td>161.976</td>
<td>62.621</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Models of snout shape evolution ranked by Akaike’s information criterion (AICc). A multi-peak Ornstein-Uhlenbeck (OU) model of feeding behavior is best supported. This model is compared to other multi-peak OU models (Diet and Habitat), a single-peak OU model, and a Brownian motion (BM) model. Log likelihood (LL), change in AICc scores (dAICc) and AICc weights (wtAICc) are also reported. The *Ophthalmotilapia nasuta* male with the large snout was excluded from this analysis because the trait is extremely sexually dimorphic in this species and confounds the relationship between ecology and snout shape.
CHAPTER 4

POPULATION-LEVEL DIFFERENCES IN THE ANATOMY AND PLASTICITY OF THE LABEOTROPHEUS SNOUT

4.1 Introduction

The production of phenotypic variation is a vital first step for evolution by natural selection, and the types and magnitude of variation that are exposed to selection can influence the direction and rate of divergence. For instance, highly canalized traits may be slow to evolve, whereas traits that exhibit a high degree of variability have the potential to be more evolvable (157). Patterns of variation and their ecological consequences, have been extensively studied across a range of organisms (158–164). However, the mechanisms through which variation is generated remains an important but open line of research. While much recent attention has been given to genetic sources of phenotypic variation (e.g., (143, 165–167)) it is increasingly appreciated that the environment plays an important role. For example, cryptic genetic variation can be “released” when populations are exposed to novel environments (168). Phenotypic plasticity is also an important source of variation that can ultimately facilitate adaptive radiations (169), lead to speciation, and/or the evolution of novelty (170). Determining the relative contributions of genetic and environmental factors to phenotypic variation is a major goal of evolutionary biology. Here, we explore population-level differences in the anatomy and phenotypic plasticity of a novel craniofacial trait in a Lake Malawi cichlid lineage.
The *Labeotropheus* genus from Lake Malawi is easily distinguished from other rock-dwelling mbuna cichlids by its fleshy, protruding snout (Figure 4.1B). In general, the unremarkable snout of most fish has two basic features – a mild invagination of the skin which likely allows the skin to extend during jaw protrusion, and a small ligament (the intermaxillary ligament) stretching mediolaterally through the snout anchoring to the left and right maxillary heads, which helps to keep the premaxilla functionally integrated with the rest of the feeding apparatus. Species within the genus *Labeotropheus* have taken this phenotype to the extreme. Specifically, in the snout of *Labeotropheus*, the skin invagination can extend beyond the teeth of the upper jaw, the intermaxillary ligament along with the surrounding loose connective tissue is expanded, and the intermaxillary ligament interdigitates with the overlaying connective tissue and anchors to the epithelium. Our previous work has explored the genetic variation that regulates the development of this bizarre trait (78), and demonstrated that hypertrophication of this trait is associated with increased expression of the ligament transcription factor *scleraxis*, likely mediated by the Transforming Growth Factor β pathway (113). However, we have also shown that the genetic basis for snout size is dependent on the environment in which the population is reared (88).

While the precise function of the snout is unknown, there is evidence that this trait is used in feeding. *Labeotropheus* are obligate benthic algivores whose craniofacial anatomy is well adapted to scrape algae from rocks. Indeed, evidence from evolutionary models and genetic data indicate the snout is evolving under selection for foraging behavior (78, 171). This is consistent with a hypothesis
proposed by Konings (11) that the snout enhances foraging efficiency in *Labeotropheus* by acting as a fulcrum to tear algae off rocks by leverage rather than the energetically costly bite-and-twist feeding behavior common to other algivorous cichlids. The fact that the intermaxillary ligament in *Labeotropheus* is expanded in size and anchors to the skin lends support to the fulcrum hypothesis, because this type of interaction may increase skin stiffness (e.g., as it has been shown in sharks (91) and geckos (90)) and a stiffer snout should provide a more robust fulcrum.

*Labeotropheus* have a cosmopolitan distribution around Lake Malawi and are found at nearly every rocky shore. The size and underlying limnology of the lake mean that different populations of *Labeotropheus* should be exposed to different environmental conditions (12). This, combined with low levels of gene flow between populations of rock-dwelling cichlids in general, sets the stage for intraspecific phenotypic variation to evolve. Indeed, Konings (11) observed that *Labeotropheus* in more turbulent parts of the lake grow larger snouts, supposedly to provide a larger, more stable base during feeding to avoid getting swept away. However, this has not been formally tested.

Here, we characterize the snout anatomy of *Labeotropheus fueelleborni* from two distinct populations. Makanjila is a relatively sheltered bay along the shallow southeast arm of the lake. Likoma, on the other hand, is an exposed island along the deeper eastern side of the lake over 200km north of Makanjila (Figure 4.1 A). On the whole, environmental conditions in Makanjila are more unstable because of temperature fluctuations that cause more upwelling and nutrient cycling (172, 173). Given that the evolution of phenotypic plasticity is typically favored in
unpredictable environments, we also assessed the degree of plasticity in both populations. Overall, we find that the two *Labeotropheus fuelleborni* populations have distinct snout phenotypes in both a gross morphological- and tissue-level analysis. We also show that populations differ in their ability to respond to a changing environment. In all, these data contribute to our understanding of how variation at the macroevolutionary scale (i.e., the evolution of novel morphologies) can achieve local adaptation via differential sensitivity to the environment.

4.2 Methods

4.2.1 Animals

*Labeotropheus fuelleborni* (LF) were derived from Lake Malawi wild-caught fish from either Makanjila or Likoma populations and housed in the Albertson lab at the University of Massachusetts Amherst. They were fed algae and egg yolk flake food (unless otherwise noted) and were reared in 40-gallon glass aquaria maintained at 28.5°C ± 1°C on a 14 hr light/10-hr dark daily cycle. Salinity and pH were also automatically monitored and dosed. Embryos were collected from mouth-brooding females after 3 days and kept in aerated flasks dosed with methylene blue. Once the yolk was absorbed they were moved to the glass aquaria.
4.2.2 Plasticity Experiments

*Makanjila*

LF from Makanjila were reared in 40-gallon glass aquaria and subjected to a benthic diet treatment (n=26) or a control diet (n=12) for 5 months starting around 1 month of age. Both diets consisted of the same amount and type of food (a mixture of algae and egg yolk flake food) presented in different ways. Flake food for the benthic diet was ground up and mixed with food-grade agar to form a paste. This paste was spread on lava rocks and allowed to dry, resulting in a food that the fish had to bite and scrape off the rocks. The algae/yolk food mixture for the control fish was delivered as large flakes, which was intended to serve as a biomechanically undemanding diet (e.g., it minimized the amount of suction feeding that was needed, and animals did not have to leverage food from rocks). A limnetic diet treatment (n=12) was also carried out wherein juvenile LF were fed the algae and egg yolk flakes finely ground and sprinkled in the water column such that fish were suction feeding from the surface of the water. Since this additional diet treatment was statistically indistinguishable from control animals, these data are only presented as supplemental information (Appendix D Figure S1, Table S1).

*Likoma*

LF from Likoma were subjected to a lab environment that was designed to mimic their natural environment (i.e., a benthic diet of scraping food from rocks and turbulent water). Specifically, juvenile LF (n=13) were reared from 1 to 6 months of age in 40-gallon glass aquaria with a wave maker generating water flow 24
hours/day. Fish were fed a mixture of algae and yolk flake food hardened onto lava rocks (see above). Control LF from Likoma (n=6) were reared in identical conditions except without a wave maker and they were fed the same diet Makanjila control fish. The sample size of Likoma fish was smaller and so we did not perform a limnetic treatment. Given that snout morphology did not differ between limnetic and control fish in Makanjila, we do not believe that this is a major confounding issue.

4.2.3 Anatomy

When necessary, animals were euthanized in MS-222 following methods approved by the University of Massachusetts Institutional Animal Care and Use Committee, fixed in 4% paraformaldehyde, and stored in 70% ethanol. The snout was dissected by making a sagittal incision on the midline from the top of the head to the upper jaw. The skin from the right side of the face was cut away to reveal a sagittal view of the center of the snout (Figure 4.1 C). Specimens were imaged using a Leica DFC450 C digital microscope camera mounted to a Leica M165 FC stereomicroscope. Snout length was measured from digital images using imageJ (NIH) by taking the linear measurement from the dorsal-most point of invagination to the tip of the snout. Snout depth was measured as the maximum depth between the proximal edge of the intermaxillary ligament and distal epithelium. Snout morphology was standardized by body depth. We analyzed differences in snout size
between populations or plasticity treatments using $t$ tests in the R statistical language (version 3.4.1).

### 4.2.4 In situ Ligament Quantification

Tissue-level anatomy of the snout was revealed by whole-mount Direct Red staining in control LF from Likoma (n=6) and Makanjila (n=5) as well as LF from the plasticity experiments - Makanjila benthic (n=12) and limnetic treatments (n=6) and Likoma benthic/wave (n=13). The procedure was adapted from the Hall-Brunt Quadruple connective tissue stain (86) for its ability to differentially stain the intermaxillary ligament within the snout (Figure 4.1 D; (113)). The dissected snouts were rehydrated in a graded ethanol series ending with reverse osmosis water. Tissue was then incubated in 0.5% Direct Red for 3 minutes. Tissue was then quickly rinsed and imaged in 100% ethanol. Direct Red is readily water soluble, allowing tissue to be destained and restained as necessary.

Previous work has shown that the tissue immediately distal to the intermaxillary ligament is loose connective tissue, which stains blue in a full HBQ stain (113). Here, we quantified surface area of intermaxillary ligament and the surrounding loose connective tissue in a sagittal section at the midline of the snout. These tissues were quantified from digital images using imageJ (NIH) as the surface area from the tip of the snout to the dorsal edge of the ascending arm of the premaxilla. In all analyses, the ratio of intermaxillary ligament to loose connective tissue was used to assess the relative proportion of each tissue type. Amount of
tissue was standardized by body depth, by taking residuals from regression analyses.

4.3 Results

4.3.1 Population-level differences in snout anatomy

We find that snout shape varies between Likoma and Makanjila populations (Figure 4.2, Table 4.1). Specifically, Likoma individuals possess short, deep snouts with roughly equal amounts of intermaxillary ligament (IML) and loose connective tissue (LCT). Makanjila individuals, on the other hand, possess relatively long, shallow snouts with more IML compared to the amount of LCT. These differences between populations are all significant (Table 4.1).

4.3.2 Population-level differences in snout plasticity

In addition to morphological differences between the populations, we found that populations differed in their ability to respond plastically to different foraging conditions. Since Likoma is an island, and exposed to high levels of wave action, we sought to mimic these conditions in the lab by making animals feed from rocks while also being exposed to wave-action (i.e., a wave maker). Makanjila is in a more shallow and protected region of the lake, and we therefore simply made fish forage from rocks without introducing the extra variable of wave-action. There were no
obvious behavioral differences in how fish fed under each condition. Both Likoma and Makanjila fish foraged vigorously from the top sides of rocks when they were introduced to the tank.

The benthic and turbulent conditions in the Likoma populations did not lead to any significant changes in snout length, depth, or the amount of IML/LCT (Figure 4.3 A-E, Table 4.1). However, in Makanjila, where only one environmental variable was changed, we found that all aspects of snout morphology were altered in the benthic treatment. Specifically, snouts of animals raised on a benthic diet are shorter, deeper, and have close to equal amounts of IML and LCT (Figure 4.3 F-J, Table 4.1). Notably, the benthic treatment in Makanjila animals leads to a snout morphology that closely resembles the Likoma population.

4.4 Discussion

4.4.1 Snout Morphology is Population Specific

We find that the characteristic snout of *Labeotropheus fuelleborni* (LF) varies between populations. Specifically, LF from Makanjila have longer but shallower snouts that are composed of more intermaxillary ligament (IML) than the surrounding loose connective tissue (LCT). LF from Likoma have shorter but deeper snouts that are composed of roughly equal amounts of IML and LCT (Figure 4.2). Intraspecific divergence in trait morphology is expected if different environments are producing distinct selective pressures. We don’t know whether selective
pressures are different between Likoma and Makanjila, but there does seem to be environmental variation that may be shifting the niche of LF between these populations. For instance, Likoma is an exposed island and it likely experiences more wave action than the sheltered bay of Makanjila. Additionally, greater temperature fluctuations at Makanjila cause more unpredictable nutrient cycling compared to the more moderate and stable conditions at Likoma (172, 173). Given these different environments, it is perhaps not surprising that an important feeding adaptation such as the snout is diverging between these populations.

What then is the benefit of the different snout morphologies? Snout length is likely related to the degree of jaw protrusion such that a longer snout (i.e., longer invagination of the skin) enables more extreme jaw protrusion. Jaw protrusion is normally considered with respect to suction feeding kinematics, however LF also protrude their upper jaws, only it is directed down toward the substrate (174). In the case of algae scrapers like LF, a shorter snout may mean the animal needs to be closer to the substrate during feeding because jaw protrusion is limited. This may occur in more turbulent parts of the lake, such as Likoma, when animals need to be closer to the substrate while feeding due to increased wave action. A shorter snout might also allow for more rapid jaw protrusion, as, all other variables being equal, the time to maximum protrusion should be shorter.

The functionality of a deep snout is more difficult to predict. However, Konings’ (11) snout-as-a-fulcrum hypothesis may provide some insights. If true, then snout depth should act as a moment arm in feeding. That is, as an animal grasps filamentous algae in its jaws and begins to roll over its snout, a deeper snout,
and hence longer moment arm (i.e., increased distance between the tip of the teeth and rostral most edge of the snout), may make it even easier to tear algae off rocks.

The above discussion is, of course, a matter of speculation, however it lays out a set of testable hypotheses. For instance, Likoma LF have short, deep snouts relative to Makanjila LF. Therefore, we would expect that compared to LF from Makanjila, LF from Likoma should (1) protrude their jaws faster (e.g., take more bites per unit time), and (2) be able to generate greater “tearing” forces (e.g., dislodge thicker strands of algae). It also remains to be tested the degree to which the LF foraging habitat differs between Likoma and Makanjila. For example, are LF at Likoma truly exposed to more turbulent waters? Is the structure of algal beds different between locations? Unfortunately, the answer to these and other questions need to be in hand before we can assess whether population level differences in snout shape have arisen due to selection or neutral processes.

4.4.2 Phenotypic Plasticity Allows Convergence in Snout Morphology

We found that every aspect of snout morphology measured here is canalized in the Likoma population, in spite of manipulating multiple aspects of their environment (Figure 4.3 A-E, Table 4.1). In Makanjila, on the other hand, the benthic diet elicited a change in every aspect of snout morphology (Figure 4.3 F-J, Table 4.1). Specifically, Makanjila LF reared under a benthic diet developed shorter, deeper snouts with roughly equal amounts of IML and LCT. This phenotype is very similar to the Likoma LF snout (Figure 4.4). This suggests a few things. First, consistent with
theory (175), it makes sense that the Makanjila population would maintain flexibility in the snout phenotype since the environment is likely less stable due to more extreme temperature fluctuations than what’s seen in the center of the lake where Likoma is located. In fact, this flexibility in snout morphology in Makanjila may be actively selected for given that periodic upwellings should result in a greater abundance of limnetic prey items. Thus, during times of upwelling LF at Makanjila may develop longer snouts to accommodate greater jaw protrusion to capture limnetic prey, whereas when such prey is scarce LF may develop short/deep snouts for benthic foraging. In addition, that the snout of Likoma LF appears to be robust to environmental perturbation is also consistent with theory, whereby plasticity is predicted to be energetically expensive to maintain. In other words, when environments are predictable, animals are expected to lose plasticity (or never to evolve it in the first place). Lastly, since the benthic treatment in Makanjila resulted in the same snout phenotype as that seen in Likoma, this suggests there may be an optimum and stereotypical benthic snout phenotype.

4.5 Conclusion

The adaptive radiation of cichlid fishes is an extraordinary example of vertebrate diversity, both in terms of species richness and morphological disparity. The genus *Labeotropheus* defines one boundary of cichlid morphospace due to a suite of traits (e.g., wide, ventrally-directed jaws) that facilitate its benthic lifestyle. Our previous work has shown that a novel snout morphology is another adaptation
for its algivorous diet (171) and may elevate feeding performance (11). While there are many other benthic cichlids in Lake Malawi, *Labeotropheus* is known to be one of the more ecologically successful lineages because of its wide geographic distribution and large body size (11). It’s possible that the novel snout contributes to this success given the relationship between novelties and ecological success in general.

Here, we expand our understanding of the evolution of novelty in this system by uncovering intraspecific variation that has presumably allowed for local adaptation to different environments. Not only do we see differences at the gross morphological and tissue-level anatomy of the *Labeotropheus* snout in different populations, but we also find variation in the capacity to respond to environmental cues. This difference in a plastic response shows that this phenotype is evolutionarily labile.
Figure 4.1 The novel snout of *Labeotropheus* endemic to Lake Malawi, Africa

(A) *Labeotropheus* is endemic to Lake Malawi, Africa. The Likoma population surrounds an island exposed to wave action of the lake. The Makanjila population is in a sheltered bay. (B) *Labeotropheus* has a novel snout anatomy compared to other Lake Malawi cichlids (represented here by *Labeotropheus twevavasae*; Image courtesy of Ad Konings). (C) Sagittal cross-section of the snout reveals hypertrophied tissue as well as an invagination of the skin that forms a flexible “flap” of tissue. (D) Direct Red staining shows much of the tissue of the snout is the collagen-rich intermaxillary ligament (IML, staining red). Note the dynamic boundary between the IML and overlying tissue.
Figure 4.2 Population level differences in the snout anatomy of *Labeotropheus* Makanjila individuals have significantly (A) longer and (B) shallower snouts compared to those from Likoma. (C) Makanjila individuals have a higher ratio of intermaxillary ligament (IML) to loose connective tissue (LCT). Schematics summarizing the anatomy of (D) Likoma and (E) Makanjila individuals. The IML is shown in pink and the LCT is blue. Significance indicated by * p < 0.05
Figure 4.3 Differences in plasticity between Likoma and Makanjila

(A-E) Snout anatomy is canalized in the Likoma population. There is no significant difference in (A) snout length, (B) depth, or (C) ratio of intermaxillary ligament (IML) to loose connective tissue (LCT). Schematics summarizing the anatomy of (D) control and (E) benthic individuals. The IML is shown in pink and the LCT is blue.

(F-J) Snout anatomy is plastic in the Makanjila population. Individuals in the benthic treatment have significantly (F) shorter and (G) deeper snouts with a (H) lower ratio of IML to LCT compared to control individuals from the same population. Schematics summarizing the anatomy of (I) control and (J) benthic individuals. Significance indicated by * p < 0.05
Figure 4.4 Reaction norms summarizing environmental sensitivity of snout phenotype in Likoma and Makanjila

Reaction norms summarizing environmental sensitivity of snout phenotype in Likoma and Makanjila. Note, snout phenotype in Likoma (gray lines) is relatively stable in a benthic environment whereas the snout phenotype in Makanjila (black lines) shifts according to the environment. Additionally, the benthic environment remodells the Makanjila snout towards the Likoma snout phenotype (i.e., line types converge in the benthic environment).
Results of \( t \) tests showing significant differences in snout anatomy between Likoma and Makanjila populations. Significant differences in snout anatomy between treatment groups within the Makanjila population suggest phenotypic plasticity whereas a lack of significance between treatment groups in the Likoma population suggest trait canalization.

### Table 4.1 Results of \( t \) tests showing significant differences in snout anatomy between Likoma and Makanjila populations

<table>
<thead>
<tr>
<th>Trait</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Populations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Snout Length</td>
<td>Likoma</td>
<td>6</td>
<td>-0.129</td>
<td>-2.506</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>12</td>
<td>0.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Snout Depth</td>
<td>Likoma</td>
<td>6</td>
<td>0.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>12</td>
<td>-0.070</td>
<td>4.357</td>
<td>0.001</td>
</tr>
<tr>
<td>Relative IML/LCT</td>
<td>Likoma</td>
<td>6</td>
<td>-0.300</td>
<td></td>
<td>-4.149</td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>5</td>
<td>0.360</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Populations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Snout Length</td>
<td>Likoma</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Snout Depth</td>
<td>Likoma</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative IML/LCT</td>
<td>Likoma</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Makanjila</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Likoma**

- Relative Snout Length: Benthic 13 -0.015 -0.569 0.581
- Relative Snout Depth: Benthic 13 -0.016 0.997 0.335
- Relative IML/LCT: Benthic 13 0.001 -0.032 0.975

**Makanjila**

- Relative Snout Length: Benthic 26 -0.037 2.700 0.012
- Relative Snout Depth: Benthic 26 0.068 -3.660 0.001
- Relative IML/LCT: Benthic 12 -0.234 3.224 0.025
Evolutionary relationships of cichlid species used to determine whether LF carried the derived allele. Species abbreviations are identical to Table S1. Phylogeny after Salzburger et al. (2005).
List of candidate genes for flap development within QTL 95% confidence intervals on four different linkage groups (LG). SNPs were filtered based on outlier $F_{ST}$ values (i.e., $\geq 0.57$) and LF possessing the derived allele. Cichlid species abbreviations are as follows: LF, *Labeotropheus fuelleborni*; TRC, *Tropheops* ‘red cheek’; MZ, *Maylandia zebra*; AB, *Astatotilapia burtoni*; PN, *Pundamilia nyerei*; NB, *Neolamprologus brichardi*; ON, *Oreochromis niloticus*. Markers closest to QTL peak LOD scores are in boldface and followed by an asterisk. Candidate genes discussed in the text are boldfaced and italicized. Genotypes for each species are given as A, T, G, or C. Dash marks indicate missing data, “D” indicates a deletion that encompasses the SNP of interest. Note that the marker closest to the QTL peak on LG14 below (“7981298”) is different that the nearest marker listed in Table 1 (“c8.7656076”) because c8.7656076 has a low $F_{ST}$ and so was not included in this table. The marker at 7981298 is the next closest marker with a high $F_{ST}$.

<table>
<thead>
<tr>
<th>QTL #1</th>
<th>Scaffold number</th>
<th>Position (bp)</th>
<th>$F_{ST}$</th>
<th>LF</th>
<th>TRC</th>
<th>MZ</th>
<th>AB</th>
<th>PN</th>
<th>NB</th>
<th>ON</th>
<th>Adjacent gene</th>
<th>SNP relative to gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>973190</td>
<td>0.773</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>sorting nexin 19</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1325166</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>sortilin-related receptor L (DLR class) A repeats</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1325183</td>
<td>1.000</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>sortilin-related receptor L (DLR class) A repeats</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2639114</td>
<td>1.000</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>breast carcinoma amplified sequence 4</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4486020</td>
<td>0.706</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>golgi autoantigen golgin subfamily a 2</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5071706</td>
<td>0.950</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>mediator of RNA polymerase II transcription subunit 13</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>6854671</td>
<td>0.855</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>AP2 associated kinase 1/bone morphogenetic protein 1a</td>
<td>intronic/5'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>6854782</td>
<td>1.000</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>AP2 associated kinase 1/bone morphogenetic protein 1a</td>
<td>intronic/5'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>8989175*</td>
<td>0.950</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>glutamate receptor ionotropic, NMDA 3A</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>11488535</td>
<td>0.950</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>glutamate receptor ionotropic, NMDA 3A</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>13419971</td>
<td>0.814</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>leucine rich repeat transmembrane neuronal 4</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>13088722</td>
<td>0.945</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>von 2 guanine nucleotide exchange factor</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>14345619</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>V-set and immunoglobulin domain containing 8a</td>
<td>intronic</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>14967006</td>
<td>0.900</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>BarH 1b</td>
<td>5'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>14971946</td>
<td>0.769</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>BarH 1b</td>
<td>5'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>15218636</td>
<td>0.903</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>zinc finger 2275</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>16149180</td>
<td>1.000</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>euchromatic histone-lysine N-methyltransferase 1b</td>
<td>exonic</td>
</tr>
<tr>
<td>QTL #2</td>
<td>LG</td>
<td>Scaffold number</td>
<td>Position (bp)</td>
<td>$F_{ST}$</td>
<td>LF</td>
<td>TRC</td>
<td>MZ</td>
<td>AB</td>
<td>PN</td>
<td>NB</td>
<td>ON</td>
<td>Adjacent gene</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>----------------</td>
<td>--------------</td>
<td>--------</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----------------</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1622976</td>
<td>0.855</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>taperin</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>16285882</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>natriuretic peptide receptor B/guanlylate cyclase B [NPR2]</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>18725296</td>
<td>1.000</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>uncharacterized LOC102304234</td>
</tr>
<tr>
<td>7</td>
<td>77</td>
<td>1009058</td>
<td>0.812</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>protein phosphatase 1 regulatory (inhibitor) subunit 3B</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>77</td>
<td>1482098</td>
<td>0.751</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>WD repeat domain 36</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>77</td>
<td>1978929</td>
<td>0.900</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>folypolyglutamate synthase</td>
<td>5'</td>
</tr>
<tr>
<td>7</td>
<td>77</td>
<td>2013045</td>
<td>0.950</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>LIM domain only 4a</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>45</td>
<td>50973</td>
<td>0.855</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>ATPase H+ transporting lysosomal V1 subunit B2</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>592</td>
<td>37147</td>
<td>0.900</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>D</td>
<td>G</td>
<td>no gene on this scaffold</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>340749</td>
<td>0.654</td>
<td>A</td>
<td>-</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>eukaryotic translation initiation factor 4 gamma 2</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>731960</td>
<td>0.889</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>leucine-rich repeat and IQ domain-containing protein 1</td>
<td>exonic</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>98</td>
<td>2246548</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>myosin 9a</td>
<td>exonic</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>2593202</td>
<td>0.857</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>myosin 9b</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>159729</td>
<td>0.900</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>CCR4-NOT transcription complex subunit 4</td>
<td>3'</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>162660</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>CCR4-NOT transcription complex subunit 4</td>
<td>exonic</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>809525</td>
<td>0.950</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A disintegrin and metalloproteinase with thrombospondin motifs 20</td>
<td>5'</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>1617848</td>
<td>0.900</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>SRSF protein kinase 2</td>
<td>intronic</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>1991227</td>
<td>0.857</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>peroxisome proliferator-activated receptor alpha</td>
<td>intronic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QTL #3</th>
<th>LG</th>
<th>Scaffold number</th>
<th>Position (bp)</th>
<th>$F_{ST}$</th>
<th>LF</th>
<th>TRC</th>
<th>MZ</th>
<th>AB</th>
<th>PN</th>
<th>NB</th>
<th>ON</th>
<th>Adjacent gene</th>
<th>SNP relative to gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>8</td>
<td>4146160</td>
<td>0.661</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>uncharacterized LOC101484193</td>
<td>3'</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>4600666</td>
<td>1.000</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>non-LTR retrotransposon Rex6</td>
<td>5'</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>6918800</td>
<td>1.000</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Yip1 interacting factor homolog B</td>
<td>3'</td>
</tr>
<tr>
<td>QTL</td>
<td>LG</td>
<td>Scaffold number</td>
<td>Position (bp)</td>
<td>FST</td>
<td>LF</td>
<td>TRC</td>
<td>MZ</td>
<td>AB</td>
<td>PN</td>
<td>NB</td>
<td>ON</td>
<td>Adjacent gene</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
<td>----------------</td>
<td>--------------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>7981298*</td>
<td>0.769</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>Willebrand factor A domain-containing protein 5A</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>8245183</td>
<td>0.773</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td></td>
<td>beta-1 3-glucuronyltransferase 1a</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>8842204</td>
<td>0.580</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td>Down syndrome cell adhesion molecule b</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>10440285</td>
<td>0.900</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td>Down syndrome cell adhesion molecule b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>125</td>
<td>1272507</td>
<td>0.814</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td>RNA-directed DNA polymerase from mobile element jockey</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>125</td>
<td>1591783</td>
<td>1.000</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td>limbic system-associated membrane protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QTL #4</td>
<td>22</td>
<td>168</td>
<td>5735</td>
<td>1.000</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td>UPF0556 protein C19orf10 homolog</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>168</td>
<td>402786</td>
<td>1.000</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td></td>
<td>rho/rac guanine nucleotide exchange factor (GEF) 18b</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>168</td>
<td>402816</td>
<td>1.000</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>obscurin cytoskeletal calmodulin and titin-interacting RhoGEF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>168</td>
<td>730099</td>
<td>1.000</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td>reg</td>
<td>chromosome 19 open reading frame 26</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>246451</td>
<td>1.000</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td></td>
<td>regulatory factor X-associated ankyrin-containing protein</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>754066</td>
<td>0.659</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>1240012</td>
<td>1.000</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td></td>
<td></td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>3022819*</td>
<td>0.808</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td>HAUS augmin complex subunit 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>3266945</td>
<td>0.903</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>dynein heavy chain 6 axonemal or E-selectin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>3830869</td>
<td>1.000</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td></td>
<td>chol</td>
<td>sterol O-acyltransferase 1</td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>3987865</td>
<td>0.734</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td>GLIS family zinc finger 1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40</td>
<td>4283073</td>
<td>1.000</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>sterol O-acyltransferase 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>273340</td>
<td>1.000</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>?</td>
<td>T</td>
<td></td>
<td>TM2 domain containing 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>382448</td>
<td>1.000</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>NACHT and WD repeat domain containing 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>838506</td>
<td>0.889</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td>WWC family member 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>1116440</td>
<td>0.900</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>Nance-Horan syndrome a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>1488593</td>
<td>0.903</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>cytoplasmic polyadenylation element binding protein 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>1818498</td>
<td>0.734</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td>furry homolog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>2173492</td>
<td>0.734</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>ligand of numb-protein X1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>3403027</td>
<td>0.692</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>methyltransferase like 13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SNP relative to gene
<table>
<thead>
<tr>
<th>QTL</th>
<th>LG</th>
<th>Scaffold number</th>
<th>Position (bp)</th>
<th>F&lt;sub&gt;ST&lt;/sub&gt;</th>
<th>LF</th>
<th>TRC</th>
<th>MZ</th>
<th>AB</th>
<th>PN</th>
<th>NB</th>
<th>ON</th>
<th>Adjacent gene</th>
<th>SNP relative to gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>49</td>
<td>3492002</td>
<td>3492002</td>
<td>1.000</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>si:dkey-239i20.4/2</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>3526426</td>
<td>3526426</td>
<td>0.728</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>-</td>
<td>si:dkey-239i20.4</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>4082148</td>
<td>4082148</td>
<td>0.950</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>cytochrome P450 2J2</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>49</td>
<td>4082180</td>
<td>4082180</td>
<td>0.950</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>cytochrome P450 2J2</td>
<td>3'</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Methods

Animals

Cichlids were reared in 40-gallon glass aquaria on a 14hr light/10hr dark daily cycle in the Albertson lab at the University of Massachusetts Amherst. Water was automatically regulated to keep temperature, salinity, and pH consistent. Animals were fed a mixture of algae and egg yolk flake food. When necessary animals were euthanized in MS-222 following standard protocols approved by the University of Massachusetts IACUC and fixed in 4% paraformaldehyde.

Histology

All histology was performed in the Webb lab at the University of Rhode Island. In both the developmental time series and bead experiments, dissected whole heads were embedded in Paraplast Plus, serial sagittal sections were cut at a thickness of 8 microns, and stained (see main text methods). Sections were imaged on a Leica DFC450 C digital microscope camera mounted to a Leica DM1000 light microscope and measurements were made using ImageJ software (NIH). The amount of intermaxillary ligament and loose connective tissue in a representative section were quantified by measuring the area occupied by each of these tissues relative to a portion of the snout standardized to both body size and the distance to which the tissue folds in on itself, “snout length”.

91
**Immunohistochemistry**

Dissected whole heads were embedded in 1.5% agar/5% sucrose and serial sagittal sections were cut at a thickness of 20 microns at -25°C on a Leica CM1850 cryostat. A 3-day staining protocol was used. Briefly, 2N hydrochloric acid was used for antigen retrieval and 2% bovine serum albumin/5% normal goat serum were used to block non-specific staining. Tissue was then incubated in primary antibody at 4°C overnight at the following concentrations: Rabbit Anti-Scleraxis Polyclonal Antibody 1:500, Rabbit Anti-TGFβ-1 Polyclonal Antibody 1:50, Rabbit Anti-ADAM12 Polyclonal Antibody 1:75 (Bioss Antibodies). All antibodies were designed from human epitopes with percent similarity to the amino acid sequence in cichlids as follows: Scx 86.67%, Tgfβ1 81.25%, and Adam12 91.67%. Tissue was then incubated in a fluorescent secondary antibody (Goat anti-Rabbit IgG Alexa Fluor 488, Life Technologies) at 4°C overnight at a concentration of 1:800. Finally, cell nuclei were counterstained with 0.01% DAPI (Sigma). Stained slides were imaged on a Nikon Eclipse TE2000 inverted microscope.

**Quantitative real-time PCR**

Dissected snout tissue (cichlids) or right half of head with eye, gill, and brain removed (zebrafish) for RNA extraction was homogenized using the Bullet Blender Storm Tissue Homogenizer and stainless steel UFO beads (Next Advance, Averill Park, NY, USA). RNA was isolated from homogenized snout tissue by the phenol chloroform extraction technique using TRIzol Reagent (Ambion Life Technologies). Genomic DNA was degraded by incubating in DNase enzyme (Invitrogen). The
amount of RNA was quantified spectrophotometrically (NanoDrop 2000, Thermo Scientific) and standardized across all samples to 57.7ng/μL (cichlids) or 200ng/μL (zebrafish). RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit (Ambion Life Technologies).

For cichlids, Primer3 software (http://bioinfo.ut.ee/primer3/) was used to design primers against adam12, tgfβ1, scleraxis, smad4 LG7, smad4 LG11, and two reference genes – ubiquitin and β-actin (adam12: forward, 5’-
CTGCAGGATGACCGTTGA-3’, reverse, 5’-CGGCAAGGTCTCTTTACTG-3’; tgfβ1: forward, 5’-CCACAGAACAAAAGGAGGAGG-3’, reverse, 5’-
GATGACGGTGCTTGTAGT-3’; scx: forward, 5’-TCGCACCAATTCTGTCAACA-3’, reverse, 5’-TCGCATCTTTTGACAAGTTCTCTGT-3’; smad4 LG7: forward, 5’-
TGCATCCAGGAACAAGGCTA-3’, reverse, 5’-GGGGTGCTTGTGATTGACA-3’; smad4 LG11: forward, 5’-CTTAGCGACCACTCTGTGTTT-3’, reverse, 5’-
CAGGTCGAACACCTTGATGT-3’; ube: forward, 5’-GAAAAAGCTTCCCAGCAGT-3’, reverse, 5’-GATCAGGGACCACGAACATC-3’; act: forward, 5’-
GTATGTGCAAGGGCCGATT-3’, reverse, 5’-TTCTGACCATAACCCACCAT-3’). Primers were designed across exon-exon boundaries to exclude potential amplification of genomic DNA. Primer specificity was validated by comparing product size and sequence of amplicons following amplification from gDNA and cDNA templates. All zebrafish primers were either from the literature (scxb and β-actin (31); 18s (51)) or were purchased from MilliporeSigma (scxa, gene ID 100034489, pair 1). All primer efficiencies were ≥ 90% except for adam12 87%, tgfβ1 88%, smad4 LG7 88%, and smad4 LG11 87%.

93
Levels of gene expression were measured using SYBR Green chemistry (Power SYBR Green Master Mix) in the StepOnePlus Real-Time PCR System (Applied Biosystems). Three technical replicates for each sample were included on every run. Relative quantification was analyzed using the comparative \( C_T \) method (48). We reanalyzed data with the lowest primer efficiencies using the Pfaffl method, which takes primer efficiency into account and requires an \textit{a priori} designation of a control group (52). One effect of this analysis is that differences in expression between target genes are lost when data is standardized to a control group. Since the overall pattern of the data did not change between analyses, we present all data analyzed with the comparative \( C_T \) method in order to preserve gene-level differences in expression and also to be able to consistently present all data as relative expression. Raw \( C_T \) values are deposited in Dryad.

\textit{Tgßβ manipulation}

The beads (~200 μm diameter) were incubated in Tgß1 protein or the control buffer for 1 hour on ice prior to implantation. Fish were anesthetized in MS-222 (0.3 mg/mL) prior to bead implantation. All fish recovered from the surgery and were collected either 12 hours later for gene expression analysis (see Fig. S3 for justification of collection at 12 hours) or 7 days later for morphological analysis.

\textit{QTL mapping}

A mapping cross between \textit{Labeotropheus fuelleborni} (LF) and \textit{Tropheops} ‘red cheek’ (TRC) was previously generated, producing hybrid offspring with a
phenotypic and genetic mosaic of the parentals. A genetic map built from $F_2$ hybrids (49) and a subset of RAD-seq markers ($n=364$) was used to construct a map in $F_3$ hybrids (34). This design allowed us to take advantage of more recombination events and therefore smaller mapping intervals. See Albertson et al. (49) and Parsons et al. (34) for more details of the cross and mapping approach.

We reviewed previously published QTL data of snout (“flap”) length in $F_2$ (21) and $F_3$ hybrids (34). We also repeated these analyses to identify QTL for snout depth. We used fine-mapping to follow up on and refine the QTL interval on linkage group 13.1 identified in (21). By binning animals by their genotype (i.e., homozygous LF, homozygous TRC, or heterozygous), we were able to look, on a marker-by-marker basis within our interval on LG13.1 (~150kb intervals), at the average difference in snout size between parental genotypes. Loci that maximized this difference were prioritized. To further refine this interval, we then overlayed $F_{ST}$ data between a panel of wild-caught LF and TRC over the interval from the fine-mapping analysis. Given that this trait exhibits discontinuous variation in the wild (21) we focused on loci with $F_{ST}$ values of 1.0, indicating complete segregation between LF and TRC. Using this combination of approaches, we were able to narrow our search for candidate loci underlying snout development from genome-wide (1.06Gb) to QTL interval (6.45Mb) to fine-mapping interval (6kb) to $F_{ST}$-compliant (2 SNPs within 10bp).
Representative histological sagittal sections of the snout at three time points in development – 1.5 cm, 2.5 cm, and 4.0 cm standard length in (A-C) *Labeotropheus* and (D-F) *Tropheops*. 
Representative histological sagittal sections of snout morphology 7 days after bead implantation showing the effect of bead placement. (A-B) When the bead (blue circle) localized beneath the ligament there is an increase in ligamentous tissue (pink) in (B) Tgfβ1-treated animals compared to (A) negative controls, but not an increase in ligamentous connections to the epithelium. Note, the tissue surrounding the bead in B appears more organized and fibrous (i.e., more ligament-like) compared to A. (C-D) When the bead (not shown here for the sake of image clarity) localized to the loose connective tissue (blue) above the ligament there is an increase in both the amount of ligamentous tissue and the number of ligamentous connections to the epithelium in (D) Tgfβ1-treated animals compared to (C) negative controls.
Relative scleraxis expression standardized to (A) β-Actin and (B) Ubiquitin in a trial of the bead experiment. TGFβ1 soaked beads were implanted in the snout and scleraxis expression was measured at 12 hours, 24 hours, and 48 hours and compared to control animals with beads soaked only in the control buffer to determine the window of sensitivity for the full experiment.
Relative expression of smad4 paralogs on linkage groups 7 and 11 in Labeotropheus and Tropheops cichlids. Differences in expression analyzed by t test (smad4 LG11, $p = 0.28$; smad4 LG7, $p = 0.04$).
Differences in average allele effects between LF/LF and TRC/TRC genotypes at each QTL peak are arranged in descending order. Modes of inheritance can be interpreted from the relative size of the black and grey bars such that equal size indicates additivity, a small gray bar indicates dominance towards the TRC allele, and a smaller black bar indicates dominance towards the LF allele. The QTL on linkage group 15 exhibits overdominance.
Relative expression of *scxa* and *scxb* paralogs in adam12 (-/-) and WT zebrafish. Differences in expression analyzed by t test (*scxa*, $p = 0.058$, *scxb*, $p = 0.22$).
Direct red staining of an adult AB WT zebrafish jaw. Direct red stains both ligaments and tendons, making them easy to see under light microscopy. In this preparation, the certatohyal-mandibular ligament (CML) can be seen extending from the medial surface of the dentary bone (DNT) just below the coronoid process (CP) anteriorly (a), over the dorsal edge of the interopercle bone (IOP) (b & c), and onto the lateral surface of the ceratohyal bone (CH) posteriorly (c’). This preparation verifies the reconstruction of the CML from CT scans (Fig. 6 in the main text). In the main panel at top, the arrows indicate the path of the CML, the arrowhead shows the tendinous insertion of the second subdivision of the adductor mandibulae onto the dentary bone, the letters a-c correspond to the regions shown at higher magnification, and c’ shows the lateral face of the posterior region of the CH where the CML inserts. The abbreviation QU refers to the quadrate bone. The scale bar equals 1mm for the top panel. In panels a-c’ the scale bars all equal 200μm, the arrows point out the CML, and the arrowhead in c’ shows the insertion of the CML onto the CH. A note on the naming of the CML: Our use of this name is purely descriptive, as there is no mention of this ligament in the literature describing soft tissue anatomy in zebrafish. A CML has been described for damselfishes, and it serves an important role in feeding in this group, however whether these ligaments are homologous remains unknown at this time.
Volumetric data for the cerato-mandibular ligament are presented for WT control (i.e., casper) and adam12 mutant zebrafish. Residual values are plotted. Dark dots represent the mean for each group. Whiskers show standard errors of the mean, and grey dots represent values for individuals in each group. Statistical results based on these data are presented in the main text.
Direct red staining of adam12 mutant fish and casper WT controls verifies and extends data generated from CT scans (Fig. 6 in the main text). The medial surface of a WT lower jaw apparatus is shown to the left; adam12 is shown to the right. The two fish shown were size matched, with the mutant animal being slightly larger (WT = 2.3 cm SL; adam12 = 2.4 cm SL). In all panels, anterior is to the left, dorsal is up, and the scale bars equal 200 μm. The tendon of the second subdivision of the adductor mandibulae (A2T) is clearly thicker in WT versus mutant animals (we cannot get volumetric data for this structure from CT scans, because we cannot distinguish tendon from muscle). The coronoid process (CP) of the dentary is also much smaller, especially in the anterior-posterior dimension, in the mutant compared to WT. This is notable because this process is the area of insertion for multiple ligaments, tendons, and connective tissues (e.g., mandibular-maxillary ligament), and thus should remodel itself to accommodate mechanical input from these tissues. The observation that the CP is smaller in mutant animals is consistent with a defect in these soft tissues. In other words, smaller ligaments/tendons should lead to lower mechanical load over time and less buildup of bone around the CP. Along the dorsal surface of the interopercle bone (IOP), the CML can be seen in both WT (bottom left) and mutant (bottom right) animals. In mutants this structure is noticeably reduced, which is consistent with the volumetric data presented in the main text.
**APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 3**

**Phylogenetic Tree Construction**

We built a Bayesian, time-calibrated tree of a select suite of cichlids from Lake Tanganyika (n=33) and Lake Malawi (n=10) to perform all of our phylogenetic comparative methods. Day et al. (122) originally compiled the majority of the Lake Tanganyika cichlids present in our phylogeny to which we added 3 taxa, alongside the 10 Lake Malawi cichlids. A specimen list and GenBank accession numbers for all taxa and sequences can be found in Table S1. We used two mitochondrial sequences to build our tree: *NADH dehydrogenase 2* (*nd2*) and the non-coding control region (*cr*). Both *nd2* and *cr* were initially aligned using MUSCLE contained in the AliView v1.19 alignment viewer and editor (176), and any additional alignment was conducted manually following visual inspection.

Time-calibrated trees were generated using BEAST2 v2.4.5 following the construction of an input file in BEAUti v2.4.5 (123). All analyses were performed using the CIPRES Science Gateway v3.3 computing cluster (177). We selected the GTR+Γ substitution model based on Akaike Information Criteria (AIC) fit using jModelTest v2.1.10 (178). We partitioned our data into three subsets, to account for the different substitution rates of the genes as well as the third codon of *nd2*.

We used a log-normal distributed relaxed molecular clock and assigned a pure birth model (Yule) as a prior for the branching process. All other parameters were left to their default settings. We placed two geological calibrations on our tree based on the ages of Lake Tanganyika and Lake Malawi. We calibrated the age of Lake Tanganyika as 9-12 million years and Lake Malawi as 0.57-1 million years using a uniformly distributed prior and forced monophyly of all ingroup taxa (132, 133). Both calibrations assume that the initial cichlid radiations in Lake Tanganyika and Lake Malawi occur during lake formation.

We performed four independent runs for 2×10^7 generations sampling every 2000 generations. We checked convergence of the four runs using Tracer v1.6 and discarded 20% of the trees from each run for burn-in using Log Combiner v2.4.5. We created a maximum clade credibility tree with median node heights from the combined trees using TreeAnnotator v2.0.3.

**Phylogenetic Comparative Methods**

**Phylogenetic Size Correction**

We performed a phylogenetic size correction between snout morphology (length and depth) and body depth, and extracted snout length and depth residuals following the phylogenetic regression. We standardized by body depth instead of body length because some species with large snouts were deep-bodied (e.g., *Labeotroheus fuelleborni* and *Ophthalmotilapia nasuta*) while others had more shallow bodies (e.g. *Labeotroheus trewavasae* and *Asprotilapia leptura*). To remove the effects of size we first calculated the expected variance-covariance (VCV) matrix from the tree under Brownian motion (BM). Under BM, trait variation among lineages is expected to accumulate linearly over time. We then compute the least
squares regression coefficients from a regression of snout morphology on body depth using generalized least squares and the VCV matrix to control for phylogenetic correlation of our traits among species. Finally, we used the least squares estimates of the regression slope and intercept to calculate the size independent residuals following (179). We used the phyl.resid function in the Phytools R package to perform our size correction (124). Accounting for phylogenetic non-independence during size correction significantly reduces the variance and type I error rate to a nominal level relative to non-phylogenetically size corrected data (180).

Convergence Tests
To test whether cichlid taxa from different lakes occupy convergent regions of novel snout morphospace, we used the convnumsig function from the convevol R package (126). Convnumsig performs 1000 evolutionary simulations along our cichlid phylogeny using parameters derived from the observed snout length and depth data to build a simulated null distribution of convergent transitions. We tested whether four taxa residing in novel regions of snout morphospace (\textit{Labeotropheus fuelleborni}, \textit{L. trewavasae}, \textit{Asprotilapia leptura}, and \textit{Ophthalmotilapia nasuta}) could be considered convergent by calculating the number of transitions into each of those regions and comparing this to the simulated distribution of transitions.

Rates of Morphological Evolution
We performed two different analyses to characterize rates of snout length and depth evolution. We first asked whether the rate of snout depth and snout length evolution was coupled or de-coupled in our cichlid sample. We used likelihood procedure to compare evolutionary rates between the snout depth and snout length traits. This approach compares the likelihood of a model where snout length and snout depth evolve at different evolutionary rates to the likelihood of a model where length and depth are constrained to evolve at the same evolutionary rate. We conducted the likelihood analysis using the R function CompareRates.multTrait outlined in (127). Evidence for a difference in rates would suggest that distinct evolutionary and developmental processes control snout length and depth morphology. We then evaluated the rate of snout length and depth evolution between those taxa in Lake Tanganyika and Lake Malawi. We used the R package mvMORPH to assess the Brownian rate of evolution ($\sigma^2$) from our snout length and depth data in each of the lakes using the functions mvBM and mvOU (128). Evidence for differences in the rate of snout length or depth evolution between lakes would suggest ample ecological opportunity and/or increased evolvability in the taxa inhabiting the lake with faster rates of snout evolution.

Evolutionary Models
We compared the fit of five different evolutionary models to our snout length and depth data in a multivariate framework using the R package mvMORPH (128). We examined the fit of snout morphology to five total models: Brownian motion (BM), single-peak Ornstein–Uhlenbeck (OU), and three 4-peak OU models whereby
species were assigned to different discrete regimes for habitat (OU - Habitat), diet (OU - Diet), and feeding behavior (OU – Feeding Behavior) based on published data (11, 52, 129).

To account for uncertainty in character and phylogenetic history in our three multi-peak models, we stochastically mapped character transitions across a random sample of trees from the Bayesian posterior distribution of time-calibrated trees (BPDT). We randomly sampled 100 trees from the BPDT and generated 5 stochastic character maps on each tree using the Stochastic Mutational Mapping on Phylogenies (SIMMAP) tool (130) from phytools (124). This resulted in a total of 500 simulated character history trees for each of the three multi-peak models and ran a combined multivariate dataset containing the length and depth residuals over each tree. We used the second-order Akaike's information criterion (AICc), which corrects for small sample sizes, to select among the best evolutionary models. We calculated a mean AICc score for each of the three multi-peak models. The best-fitting model is determined by the lowest AICc score, and is favored over any other model if the difference in AICc score is greater than two units (131).
Body-size corrected snout morphospace in Lakes Malawi (black circles) and Tanganyika (red triangles) showing the distance between *Ophthalmotilapia nasuta* male and females in morphospace.
Phylogenetic tree of Lake Tanganyika and Lake Malawi cichlids projected onto a geological timescale. Purple bars represent 95% highest probability density age ranges. The posterior probability (PP) of each node is binned into three categories and denoted by three different colored circles: black PP>95%, gray 75%<PP<95%, white PP<75%. Time calibration points are illustrated by red circles at two nodes: the node at the root of the tree (#1), and the node containing all the Lake Malawi cichlids (#2). The Malawi radiation is hypothesized to be split into two major clades, the sand-dwelling utaka and the rock-dwelling mbuna, a relationship that we recover here and illustrate on the tree.
Phylogenetically- and body size-corrected snout morphospace in Lakes Malawi (circles) and Tanganyika (triangles) colored by (A) diet and (B) habitat. The Ophthalmotilapia nasuta male with the large snout was excluded from this analysis because the trait is extremely sexually dimorphic in this species and confounds the relationship between ecology and snout shape.
<table>
<thead>
<tr>
<th>Species</th>
<th>Control Region</th>
<th>ND2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lake Malawi</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aulonocara_stuartgranti</td>
<td>NC_029380</td>
<td>EU661720</td>
</tr>
<tr>
<td>Chilotilapia_rhoadesii</td>
<td>AY911729</td>
<td>JX122940</td>
</tr>
<tr>
<td>Cynotilapia_africa</td>
<td>AY911740</td>
<td>EF585264</td>
</tr>
<tr>
<td>Dimidiochromis_compressiceps</td>
<td>NC_018561</td>
<td>EF585267</td>
</tr>
<tr>
<td>Labeotropheus_fuelleborni</td>
<td>U90773</td>
<td>EF585259</td>
</tr>
<tr>
<td>Labeotropheus_trewavasae</td>
<td>AY911790</td>
<td>EF585283</td>
</tr>
<tr>
<td>Maylandia_zebra</td>
<td>KT166981</td>
<td>KT166981</td>
</tr>
<tr>
<td>Nimbochromis_linni</td>
<td>NC_018558</td>
<td>EF585279</td>
</tr>
<tr>
<td>Tramitichromis_brevis</td>
<td>AY911840</td>
<td>AF305320</td>
</tr>
<tr>
<td>Tropheops_red_cheek</td>
<td>JN628862</td>
<td>EF585265</td>
</tr>
<tr>
<td><strong>Lake Tanganyika</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asprotilapia_leptura</td>
<td>AF400701</td>
<td>Y337772</td>
</tr>
<tr>
<td>Aulonocranus_dewindti</td>
<td>AY339007</td>
<td>Y337782</td>
</tr>
<tr>
<td>Callochromis_macrops</td>
<td>AY339051</td>
<td>Y337795</td>
</tr>
<tr>
<td>Callochromis_melanostigma</td>
<td>AY339047</td>
<td>N/A</td>
</tr>
<tr>
<td>Callochromis_pleurospilus</td>
<td>Z21735</td>
<td>Y337771</td>
</tr>
<tr>
<td>Callochromis_stappersii</td>
<td>AY339048</td>
<td>Y337775</td>
</tr>
<tr>
<td>Cardiopharynx_schoutedeni</td>
<td>AY339000</td>
<td>Y337791</td>
</tr>
<tr>
<td>Cunningtonia_longiventralis</td>
<td>AY338999</td>
<td>Y337780</td>
</tr>
<tr>
<td>Cyathopharynx_furcifer</td>
<td>AY338981</td>
<td>Y337781</td>
</tr>
<tr>
<td>Cyprichromis_leptosoma</td>
<td>AY339053</td>
<td>Y337786</td>
</tr>
<tr>
<td>Ectodus_descampsii</td>
<td>AY339014</td>
<td>Y337790</td>
</tr>
<tr>
<td>Enantiopus_melanosoma</td>
<td>AY339022</td>
<td>Y337770</td>
</tr>
<tr>
<td>Eretmodus_cyanostictus</td>
<td>DQ054911</td>
<td>DQ055010</td>
</tr>
<tr>
<td>Grammatotria_lemairii</td>
<td>AY339018</td>
<td>Y337787</td>
</tr>
<tr>
<td>Lestradea_perspicax</td>
<td>Z21745</td>
<td>Y337765</td>
</tr>
<tr>
<td>Lestradea_stappersii</td>
<td>AY338997</td>
<td>Y337792</td>
</tr>
<tr>
<td>Microdontochromis_rotundiventralis</td>
<td>AY339021</td>
<td>Y337793</td>
</tr>
<tr>
<td>Microdontochromis_tenuidentatus</td>
<td>AY339019</td>
<td>Y337784</td>
</tr>
<tr>
<td>Ophthalmotilapia_boops</td>
<td>AY338987</td>
<td>Y337773</td>
</tr>
<tr>
<td>Ophthalmotilapia_heterodonta</td>
<td>Z95998</td>
<td>EF679254</td>
</tr>
<tr>
<td>Ophthalmotilapia_nasuta</td>
<td>AY338989</td>
<td>Y337783</td>
</tr>
<tr>
<td>Ophthalmotilapia_ventralis</td>
<td>AY338993</td>
<td>Y337774</td>
</tr>
<tr>
<td>Oreochromis_tanganicae</td>
<td>AY929940</td>
<td>AF317240</td>
</tr>
<tr>
<td>Xenotilapia_bathyphila</td>
<td>AY339028</td>
<td>Y337789</td>
</tr>
<tr>
<td>Xenotilapia_boulengeri</td>
<td>AY339029</td>
<td>HM135111</td>
</tr>
<tr>
<td>Xenotilapia_caudafasciata</td>
<td>AY339035</td>
<td>Y337777</td>
</tr>
<tr>
<td>Xenotilapia_cf_bathyphila</td>
<td>AY339026</td>
<td>Y337768</td>
</tr>
<tr>
<td>Xenotilapia_flavipinnis</td>
<td>AY339034</td>
<td>Y337794</td>
</tr>
<tr>
<td>Xenotilapia_longispinis</td>
<td>AY339037</td>
<td>Y337779</td>
</tr>
</tbody>
</table>
Xenotilapia_ochrogenys       Z21750     AY337767
Xenotilapia_papilio_sunflower   AY339045   AY337776
Xenotilapia_sima               AY339039   AY337785
Xenotilapia_spilooptera        AY339043   AY337788

GenBank accession numbers of species used to build the phylogeny.
Snout anatomy in Makanjila is sensitive to a benthic diet but is stable in a limnetic environment. Significance indicated by * p < 0.05
Results of separate ANOVA with Tukey correction for different aspects of snout anatomy in the Makanjila population. Significant (p < 0.05) values are in bold.

<table>
<thead>
<tr>
<th>Trait</th>
<th>F</th>
<th>Group</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative snout length</td>
<td>3.16</td>
<td>Control - Limnetic</td>
<td>0.4546</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control - Benthic</td>
<td><strong>0.0419</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limnetic - Benthic</td>
<td>0.5292</td>
</tr>
<tr>
<td>Relative snout depth</td>
<td>11.87</td>
<td>Control - Limnetic</td>
<td>0.9786</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control - Benthic</td>
<td><strong>0.0011</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limnetic - Benthic</td>
<td><strong>0.0005</strong></td>
</tr>
<tr>
<td>Relative IML/LCT</td>
<td>12.08</td>
<td>Control - Limnetic</td>
<td>0.3045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control - Benthic</td>
<td><strong>0.0005</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limnetic - Benthic</td>
<td><strong>0.0160</strong></td>
</tr>
</tbody>
</table>
REFERENCES


