Author:
Hufton, Joshua

Title:
Exploring phenotypic plasticity and gene co-expression networks in the jaw apparatus of speciating cichlid fishes

Exploring phenotypic plasticity and gene co-expression networks in the jaw apparatus of speciating cichlid fishes

Joshua Hufton

Submission date: 20/10/2022
Word count: 15,176

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master's by Research in the Faculty of Life Sciences.

Abstract

A major goal in evolutionary biology is to understand the molecular basis of phenotypic divergence during speciation. Given their high species richness, short evolutionary timescale coupled with exceptional ecological diversity, East African cichlid fishes represent model candidates for such investigations. Research for this thesis focuses on populations of the haplochromine cichlid *Astatotilapia calliptera*, in Lake Masoko, Tanzania, which are currently undergoing sympatric speciation into a shallow-water littoral ecomorph and deep-water benthic ecomorph. In Chapter 1 the study system is introduced, and genetic and epigenetic influences on phenotypic divergence are broadly summarised. In Chapter 2 I compare jaw morphology of wild-caught individuals of *A. calliptera*. I also quantify the jaw morphology of both ecomorphs when reared on divergent diets in common garden conditions. The results help to further our understanding of the extent that these phenotypes are determined by phenotypic plasticity. In Chapter 3 I link transcriptomic (RNA sequencing) and epigenetic (DNA methylation; whole genome bisulphite sequencing) data, to explore the potential for epigenetic mechanisms to promote divergent expression of functional genes during the early stages of ecological diversification. Specifically, I use a network analysis to identify co-expressed genes associated with ecologically relevant traits such as oral and lower pharyngeal jaw apparatus, and determine the overlap with those regions of the genome that are differentially methylated between the ecomorphs. In Chapter 4, I summarise the key insights provided by the study, and discuss research opportunities by this remarkable study system.

Acknowledgments

I would like to thank my supervisor's, Prof. Martin Genner and Dr Madeleine Carruthers for all their help and guidance throughout my degree, and for collecting the genetic data used in my third chapter. I would also like to thank Dr Andy Saxon for all his assistance in the laboratory, and for always being patient and helpful. I would also like to thank Prof George Turner, at Bangor University, for conducting the common garden experiments and supplying the samples used in chapter two. I would also like to thank my lab group and the School of Biological Sciences for all their support during my degree.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:  ……. Date: 20/10/2022

Table of Contents

List of Figures

List of Tables

Chapter 1: General introduction

1.1 Ecological speciation, adaptive radiation and phenotypic plasticity

Ecological speciation is the process by which barriers to gene flow develop between populations as a result of ecologically-based divergent selection between environments (Schluter, 2001; Rundle & Nosil, 2005; Nosil, 2012). The results of this process are observable in all life on earth, and it is a key concept in the field of evolutionary biology. Understanding the selection pressures driving ecological speciation has been a major goal for evolutionary biologists. Collectively, observational, experimental and genetic studies are steadily helping us to understand more about the mechanisms and pressures involved in ecological speciation (Rundle & Nosil, 2005; Pfennig et al., 2010; Salzburger, 2018). For example, in rough periwinkle (*Littorina saxatilis*), a marine gastropod of North Atlantic rocky shores, ecological speciation is taking place as two ecotypes diverge in phenotypes due to crab predation (Conde-Padín et al., 2007). The ecotypes differ in their average vertical distributions on the shore, and the phenotypic differences appear to have a strong genetic basis, suggesting that the reproductive isolation is an indirect consequence of divergent selection. In this particular case, the two ecotypes are partially overlapping, suggesting that strong divergent selection is required to drive divergence between ecotypes.

Adaptive radiation is defined as “the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage”. It occurs when one or more ancestors diverge into multiple reproductively isolated species using a range of ecological resources, and differ in traits used to exploit the resource (Schluter, 2000). This can be viewed as numerous successive ecological speciation events, resulting in an accumulation of phenotypic diversity. Ecological opportunity is thought to be one of the main features that enables adaptive radiation to occur. This means that there are available ecological resources that can be “evolutionarily exploited” (Stroud & Losos, 2016), in other words ecological niches that are available to be filled. Ecological opportunity can be provided by extinction events, that remove competition from the environment, making resources readily available and exploitable. One example is the Cretaceous–Paleogene mass extinction event which allowed birds and placental mammals to occupy new ecological niches (Hull, 2015). Ecological opportunity can also be provided by the availability of new habitat, such as when species colonise a novel

environment. An example of this is from Caribbean *Anolis* lizards, which have radiated into a multitude of different species occupying different ecological niches after colonising their island habitats. By contrast the mainland ancestral species are limited in diversity (Pinto et al., 2008). Many examples of adaptive radiation involve a trait that is a "key innovation" enabling a population to acquire a novel resource. In the case of *Anolis lizards*, the key innovation was the evolution of adhesive toepads, which enhanced their climbing ability and enabled occupation of arboreal niches.

Phenotypic plasticity is defined as "the ability of a given genotype to express different phenotypes in different environmental circumstances" (West-Eberhard, 1989, Pigliucci, 2001, Pfennig & West-Eberhard, 2021). In principle, plasticity may enable a species to develop the key innovations required to exploit an ecological opportunity, thereby helping to promote ecological speciation and adaptive radiation (Schneider & Meyer, 2017). Research addressing this issue now incorporates numerous biological areas, from evolutionary biology and genetics to developmental and behavioural ecology (Pfennig & West-Eberhard, 2021). Plasticity could impact evolution and promote adaptive diversification via "buying time" allowing a population to persist in a variable or novel environment, before new adaptations arise via genetic mutation fuelling phenotypic diversification (Pavey et al., 2010; Diamond & Martin, 2021). Alternatively, evolution could be promoted via "plasticity first" (Levis & Pfennig, 2016) or "flexible stem" (West-Eberhard, 2003; Schneider and Meyer, 2017) mechanisms, where genetic assimilation leads to the genetic fixation of a previously environmentally induced phenotype. These concepts are not mutually exclusive to one another, and the precise ways that phenotypic plasticity interacts are likely complex and not straightforward to resolve (Futuyma 2021).

The phenotypic traits generated from phenotypic plasticity often have an adaptive function, and in many cases have been shown to vary among heterogenous environments that are capable of contributing to diversification through natural selection. Evidence of adaptive phenotypic plasticity, whereby the plasticity of the trait has a direct fitness benefit, could have substantial implications for our understanding of evolution (Ledón-Rettig & Pfennig, 2011; Chevin et al., 2021; Pfennig & West-Eberhard, 2021). Spadefoot toads (*Scaphiopodidae*) from North America are a good model for studying ecological speciation and phenotypic plasticity. Species can be found in both sympatric and allopatric populations

and their development time is relatively short (Rice & Pfennig, 2010; Ledón-Rettig & Pfennig, 2011). Pfennig et al. (2006) conducted an observational study on the tadpoles of two species of spadefoot toads (*Spea multiplicata* and *Spea bombifrons*). Both species showed phenotypic plasticity based on their diet exposure, and could produce either omnivorous or carnivorous tadpole morphs. Their observations showed that the species composition and resource availability of a given pond would determine the morph of the tadpoles. Specifically, if only one of the species was found in a pond, then tadpoles would produce a similar proportion of both morphs. However, if both species were present, then *S. multiplicata* would produce more of the omnivorous morph, and *S. bombifrons* would produce more of the carnivorous morph. Both species therefore exhibit adaptive phenotypic plasticity, which allows them to maximise fitness in their immediate environment.

Natural selection in the form of resource acquisition is, in most cases, the strongest selection pressure acting on a population and can lead to reproductive isolation and speciation (Schluter, 2001). It is clear that phenotypic plasticity is almost ubiquitous in the natural world, and that it is in itself adaptive – subject to natural selection. However, the conditions that enable environmentally induced plastic traits to translate to genotypic and phenotypic fixation is still unclear (i.e. genetic assimilation). Whether or not a lineage is able to adapt to a selection pressure and subsequently speciate may be heavily dependent on the environmental and genetic mechanisms promoting phenotypic plasticity of an individual. Improving our understanding of the mechanisms likely holds the key to understanding the role of plasticity in evolution (Futuyma, 2021).

1.2 Genetics in East African cichlid fishes

Cichlids are an excellent model for the study of ecological speciation and phenotypic plasticity due to their extensive adaptive radiation (Kocher, 2004; Genner & Turner, 2005; Seehausen, 2006). East African cichlids are unparalleled when it comes to rapid phenotypic and taxonomic vertebrate diversification, and they represent one of the largest vertebrate species flocks on earth. They have diversified into 1000-2000 species during the last 5 million years, and Lake Malawi alone is estimated to contain 800-1000 species (Seehausen, 2006; Ivory et al., 2016; Salzburger, 2018). Extensive interspecific phenotypic variation can be found in a variety of traits (Genner & Turner, 2005), including jaw morphology (Conith &

Albertson, 2021), visual systems (Hofmann et al., 2009; Hahn et al., 2017) and lateral line systems (Edgley & Genner, 2019). Their substantial diversity and ongoing speciation can make it difficult to clearly delimit populations into species. Many cichlids are capable of interbreeding and generating hybrid crosses both in the wild and in aquaria, and therefore not all species may exhibit complete reproductive isolation. There has been significant progress in genetic studies on the East African cichlid models in recent years, summarised in recent literature reviews (Salzburger 2018; Svardal et al., 2020). Both reviews emphasise the relatively low genetic diversity and mutation rates of East African cichlids. In particular, the pattern of low interspecific differentiation seen within cichlid flocks is suggestive that that substantial genomic variation is not necessarily required for the evolution of large phenotypic diversity (Svardal et al., 2020).

Brawand et al. (2014) sequenced the genomes and transcriptomes of five African cichlid species, including two riverine species and one species from each of three major East African rift lakes: Tanganyika, Victoria and Malawi. They found more gene duplication events compared to most other teleosts, divergence in coding and non-coding genomic regions, and divergent gene expression. This study laid the foundations for subsequent studies that have explored molecular mechanisms generating phenotypic traits, including jaw morphology (Conith & Albertson, 2021) and visual systems (Hahn et al., 2017), in part by identifying protein coding genes that differ between species. The annotated genome of the Lake Malawi species *Metriaclima zebra* has become particularly important as a reference assembly in genomic studies on haplochromine cichlid diversity. By comparing three major lakes Brawand et al. (2014) enabled a greater understanding of cichlid diversity at the genomic level to be established, underpinning many subsequent studies on specific lakes, species, genes and phenotypes.

Malinsky et al. (2018) conducted large scale whole-genome sequencing of the major haplochromine cichlid lineages that comprise the Lake Malawi species flock. The lake is thought to have formed 5 million years ago (Mya), but palaeoecological data suggests that the lake was dry between 1.6 and 1 Mya. Stable deep-lake conditions only formed about 800 thousand years ago (Ivory et al., 2016), and it is thought that most of the species' richness has arisen since then, which makes Lake Malawi a relatively recent example of large-scale adaptive radiation. Malinsky et al. (2018) quantified genetic divergence of 134

individuals and 73 species and found interspecific diversity to be only slightly higher than intraspecific diversity, which is likely a result of a relatively low mutation rate, extensive gene flow between lineages and the relatively young age of the radiation. Within species diversity being relatively low suggests that low nucleotide diversity does not limit adaptive radiation or speciation. The Malinsky et al., (2018) study also suggests that *Astatotilapia calliptera* may be the possible common ancestor of the whole radiation, forming three successive haplochromine radiations interconnected by subsequent gene flow events. This makes *A. calliptera* key to understanding how East African cichlids have diversified. Notably this species is still found in multiple lakes and rivers in the Lake Malawi region (Malinsky et al., 2018) making it an excellent species for further study.

Recent genetic studies on East African cichlids have led to numerous insights into the genomic basis of diversification (Svardal et al., 2020). Since interspecific nucleotide diversity is relatively low given their vast phenotypic diversity, there must be specific molecular mechanisms at work that will enable us to explain the large phenotypic variation. Gene expression, and its regulation, is thought to be a key element explaining the large phenotypic variation and phenotypic plasticity found in East African cichlids. With the increased availability of genetic resources, including sequenced reference genomes for key species (Brawand et al., 2014) we can now investigate the molecular mechanisms underpinning adaptive phenotypes and ecological diversification with greater precision than has previously been possible.

1.3 Gene expression translating to phenotypic change

Genes can be turned on or off dependent on various molecular mechanisms. Whether a gene is being expressed, and how much it is expressed, determines to what extent a phenotype is observable in an individual. Gene expression is regulated in the cell via numerous molecular mechanisms and at different stages in the transition from DNA to protein. Gene expression can be inhibited or promoted by epigenetic mechanisms, including histone modifications and DNA methylation (Ng et al., 2019). Later, transcription occurs by RNA polymerase binding at promoter regions of DNA located upstream from the gene to be transcribed, and this initial stage can be inhibited or promoted by transcription factors. Once RNA is transcribed it needs to be converted to mRNA to leave the nucleus. At this point introns are spliced out so only exons remain and this can vary by a process called

alternative splicing (Bush et al., 2017). The amount of mRNA and its biochemical stability is also impacted by molecular processes. Once the mRNA is translated to a polypeptide chain various modifications can be made to it that impact its functionality and quantity in the cell (Ng et al., 2019). Next generation sequencing can be used to sequence RNA and study gene expression in an individual, or between cell types, and the availability of large databases of annotated genomic information from cichlids and other vertebrates (Kratochwil & Meyer, 2015) make this an increasingly accessible method. Expression quantitative trait analysis (eQTL) can be used to associate specific gene expression to a quantitative trait (Kratochwil & Meyer, 2015). Numerous genotype-phenotype associations in cichlid traits have been discovered using eQTL methodologies (Carruthers et al., 2022).

Cichlid jaw morphologies have evolved in parallel numerous times during cichlid adaptive radiation, and provide an example of how traits can be studied using eQTLs (Kocher, 2004). Cichlid jaw morphology is influenced by environmental factors, especially diet, and allows an individual or a species to occupy a specific niche and exploit the available resources (Liem, 1973). Gunter et al. (2013) conducted diet manipulation experiments on the East African molluscivore *Astatoreochromis alluaudi*, aiming to identify molecular markers underpinning plasticity in jaw shape driven by differences in mechanical strain imposed by soft and hard diets. The study identified candidate genes and transcriptional changes associated with jaw morphology, including mechano-responsive genes functional in bones and teeth of mammals. Schneider et al. (2014) conducted further diet-controlled experiments on the species, exploring gene expression during different stages of lower pharyngeal jaw (LPJ) development. They conducted further analyses on the specific genes identified by Gunter et al. (2013), which included analyses of co-expression and transcription factor binding sites to infer a regulatory pathway underpinning LPJ plasticity. More recently Singh et al. (2021) conducted a similar investigation focusing on both oral and lower pharyngeal jaws in three species of cichlid from Lake Tanganyika and one from Lake Malawi. They examined gene expression using a weighted gene co-expression network analysis (WGCNA), alongside a gene ontology analysis, and identified upstream transcriptional regulators/factors that form part of regulatory pathways for OJ and LPJ apparatus. These methods of analysis will be key to the future genotype-phenotype mapping, and will help to improve our understanding of specific gene expression networks (e.g. Kratochwil and Meyer, 2015; Chevin et al., 2021).

The study of differential gene expression is quickly becoming a key area of research in evolutionary studies, particularly in the context of ecological speciation, adaptive radiation and phenotypic plasticity. There are large amounts of phenotypic information on a variety of cichlid species (Kocher, 2004; Genner & Turner, 2005; Seehausen, 2006) that can now be expanded upon using next generation sequencing and advanced analytical techniques to identify the molecular basis of a given phenotype. This is important because the phenotype is under natural selection, and gene expression is key to closing that genotype-phenotype gap and creating a clearer picture of evolution as a whole (Kratochwil & Meyer, 2015; Chevin et al., 2021).

1.4 Epigenetic mechanisms in East African cichlid fishes

Epigenetic markers can be defined as non-genetic modifications to DNA or chromatin that cause changes to gene expression without changing the base DNA sequence (Duncan et al., 2014; Chevin et al., 2021). These mechanisms include DNA methylation, histone modifications and non-coding RNA, all of which can switch genes on or off thus controlling their expression and consequently influencing phenotypes (Duncan et al., 2014). These mechanisms can be environmentally induced, and are thought to be key to phenotypic plasticity (Morandin et al., 2019; Chevin et al., 2021).

DNA methylation is the transfer of a methyl group onto the C5 position of the cytosine to form 5-methylcytosine (Moore et al., 2013). DNA methylation can regulate gene expression by preventing the binding of transcription factors or by recruiting other proteins involved in gene silencing (Moore et al., 2013). These changes in gene expression can result in shifts in patterns of cell differentiation, and thus lead to phenotypic variation. DNA methylation can be transferred during cell replication via DNA methyltransferase's and is key in all multicellular organisms for cell specialization (Moore et al., 2013). As outlined in various review papers (Schneider & Meyer, 2017; Lafuente & Beldade, 2019; Chevin et al., 2021) using high throughput sequencing of bisulphite-treated DNA, we can now access information on the presence or absence of DNA methylation at unprecedented detail.

Driscoll et al. (2020) investigated DNA methylation patterns of brain and gonad tissues by sequencing bisulphite-treated DNA in the cichlid fish *Pelvacachromis pulcher*. They found strong evidence of epigenetic regulation of gene expression and development in ovaries or

testes, and epigenetic differences between brain and gonadal tissues. Vernaz et al. (2021) also performed methylome and transcriptome analysis on cichlids, specifically five closely related cichlid species from Lake Malawi. Their results revealed significant methylome divergence across species and tissues, including differentially methylated regions associated with transcriptional variation in ecologically-relevant genes, suggesting that epigenetic variation contributes to phenotypic variation and could facilitate rapid diversification and adaptive radiation.

Using our knowledge of the mechanisms and consequences of DNA methylation on gene expression (Moore et al., 2013; Ortega-Recalde & Hore, 2019) we can now begin to understand influences of methylation on specific genes, gene regulatory networks and phenotypes, and begin to incorporate these concepts into models of evolution and adaptive diversification (Smith & Ritchie, 2013; Kratochwil & Meyer, 2015; Schneider & Meyer, 2017; Chevin et al., 2021). The extent that DNA methylation is transgenerationally heritable is directly relevant to these models (Jablonka & Raz, 2009; Pfennig & Servedio, 2013; Miska & Ferguson-Smith, 2016). Although there is evidence of transgenerational heritability of methylation in sticklebacks (Hu et al., 2021), the mechanisms behind this are still unclear. In the context of cichlids it is worth noting that the extent that DNA methylation is retained or removed across generations is still largely unknown beyond preliminary indications of partial heritability (in Vernaz, et al. (2021)).

Epigenetic mechanisms provide us with further insight into how genotypes produce phenotypes. In conjunction with RNA datasets, we can create a much clearer picture of the phenotypic diversification of cichlids and find links between methylation patterns and gene expression. This builds upon the existing genomic data which has already yielded insights into evolution (Salzburger, 2018; Svardal et al., 2020). Furthermore, with the use of WGCNA (Langfelder & Horvath, 2008) and RNA datasets we can identify key gene sets involved with regulation of gene expression, while incorporating phenotypic data into analyses, thereby closing the gap between genotype and phenotype. This approach will be key to furthering our understanding of evolution, specifically concepts which are continually debated including phenotypic plasticity, ecological speciation, and adaptive radiation.

1.5 Study system

Lake Masoko is a small, isolated East-African crater lake in Tanzania, located in the northern region of the Lake Malawi catchment (Malinsky et al., 2015; Turner et al., 2019). For most of the year there is a strong thermocline and oxycline present at ~15 m, dividing the lake into a warmer (~28°C), well-oxygenated shallow-water habitat and a cooler (~24°C), poorly-oxygenated, deep-water habitat (Delalande, 2008). The lake contains two genetically segregating ecomorphs of the cichlid fish *Astatotilapia calliptera* that appear to be at an early stage of speciation (Malinsky et al., 2015). The shallow ecomorph inhabits the shallow waters (≤5m) and primarily feeds on "hard" diet of nearshore macroinvertebrates. The deep ecomorph inhabits the deeper waters (≥20m) and primarily feeds a "soft" diet of zooplankton (Malinsky et al., 2015). These differences in habitat and diet are associated with differences in phenotypes, with shallow morphs exhibiting larger lower pharyngeal jaws (LPJs) and more molariform LPJ teeth than deep morphs (Malinsky et al., 2015). The two ecomorphs also show differences in body shape (Malinsky et al., 2015). The extent of phenotypic differences present between ecomorphs, coupled with the low nucleotide diversity in the population, gives us the opportunity to investigate the role of gene expression, DNA methylation and phenotypic plasticity in contributing to adaptive phenotypic change during early-stage cichlid speciation.

1.6 Overall aims and thesis structure

The overall aim of research presented in this thesis is to contribute understanding to how phenotypic plasticity, differential gene expression and DNA methylation can contribute to adaptive divergence in functionally relevant traits.

Chapter 2 describes a study quantifying craniofacial phenotypic differences between the Lake Masoko *A. calliptera* ecomorphs. The study first aims to clarify the differences in oral jaw phenotypes among wild populations of the ecomorphs. Next, the study aims to determine the extent that oral jaw and lower pharyngeal phenotypes are heritable and/or influenced by plasticity, by studying fish from both ecomorphs reared on different diets in common garden laboratory conditions.

Chapter 3 describes a study of the transcriptomes and DNA methylation patterns from wild *A. calliptera* from Lake Masoko (Carruthers et al., 2022). The study first employs a weighted

gene expression network analysis (WGCNA) to group genes into modules based on co-expression and associations with phenotypes. Next, the extent of associations between genes in these modules and strongly differential methylation regions of the genome are quantified.

Chapter 4 includes a discussion of the key results from the thesis, and outlines directions for future research.

Chapter 2: Diet-induced phenotypic change in the jaw apparatus of *Astatotilapia calliptera* from Lake Masoko

2.1 Abstract

Phenotypic plasticity is hypothesised to have a key role in evolutionary processes such as ecological speciation, by allowing organisms the opportunity to persist and adapt to novel environments. To further understand the potential for phenotypic plasticity to contribute to evolutionary change in cichlid fishes, we investigated the oral and pharyngeal jaw apparatus of distinct deep and shallow ecomorphs of the East African haplochromine *Astatotilapia calliptera* from Lake Masoko. First, using wild samples, we show that shallow fish possess shorter oral jaws, accompanying the wider pharyngeal jaws with larger teeth already reported for this ecomorph. Second, using common garden experiments we show these morphological traits are partially heritable. We also found, however, that lower pharyngeal jaw morphology can be modified by diet. These findings suggest that jaw plasticity may have significance for the divergence of these ecomorphs in the natural environment. The results also suggest that jaw plasticity may have also taken a broader role in the evolution of East African cichlids, given the potential role of *A. calliptera* as an ancestral species of the large-scale Lake Malawi haplochromine radiation estimated to comprise over 800 species.

2.2 Introduction

Phenotypic plasticity, the capacity for organisms to shift aspects of their phenotypes depending on environmental cues, is thought to allow organisms the opportunity to adapt to novel environments and selection pressures (Pfennig & West-Eberhard, 2021). Freshwater fishes exhibit a variety of plastic traits. For example, Trinidadian guppies (*Poecilia reticulata*) reduce their conspicuousness in response to chemical predator cues by altering the colour pigmentations of their tails (Ruell et al., 2013). Clearfin livebearers (*Poeciliopsis lucida*) exhibit temperature dependent sex determination, leading to skewed sex ratios in fish reared in differing temperatures (Schultz, 1993). European eels (*Anguilla anguilla*) have head shapes influenced by diet, with eels fed on harder diets developing broader heads with a larger biting force (de Meyer et al., 2016). Such plastic traits may impact evolutionary processes by generating the phenotypic variation in populations, for natural selection to act upon. Consequently, phenotypic plasticity is thought to contribute to the diversification of life on earth (Pfennig & West-Eberhard, 2021; Pigliucci, 2001). Understanding which ecological factors impact a given plastic trait, and whether there is a heritable aspect to these traits, is at the forefront of evolutionary biology.

Labroid fishes, which include the wrasses, parrotfishes, damselfishes, surfperches and cichlids, possess pharyngeal jaws (Liem, 1973; Burress, 2015, Burress, 2016). These are a second set of jaws located in the pharynx, evolutionarily derived from gillrakers (Liem, 1973; Liem & Greenwood, 1981; Wainwright et al., 2012). Functionally, the pharyngeal jaw apparatus is decoupled from the oral jaw apparatus, with the pharyngeal jaws in these species primarily used for prey processing (for example grinding) and the oral jaws primary used for prey acquisition (for example suction feeding, ram feeding or biting). The evolution of the pharyngeal jaws has been considered to be a key innovation, allowing cichlids to adapt to capture and process a wide range of resources from fish to plankton and epilithic algae (Burress et al., 2016). It is thought this adaptive trait has been highly influential in allowing cichlids to radiate into the wide variety of ecological niches they currently occupy (Conith & Albertson, 2021; Gunter et al., 2013). Importantly, pharyngeal jaws can be a highly plastic trait in labroid fishes (Wainwright et al., 2012) and diet-controlled laboratory experiments have clearly shown the influence of the food types on their structure in cichlids (Gunter et al., 2013; Muschick et al., 2011).

Notably, in addition to pharyngeal jaws, oral jaws and craniofacial morphology also show strong phenotypic divergence in cichlid fishes (Conith & Albertson, 2021). Liem, (1973) noted not only the functional separation between the pharyngeal jaws and oral jaws, but also proposed the evolutionary decoupling of the jaws, and until recently it was thought that the pharyngeal jaws and oral jaws occupied different evolutionary trajectories. However, Conith & Albertson, (2021) have suggested that the two jaws are genetically coupled, sharing common (pleiotropic) genes, and exhibiting similar patterns of gene expression.

A central question in the evolution of cichlids is the extent that observed jaw divergence between closely related and recently diverged species in the wild is driven by purely genetic factors, or environmental factors such as the locally available food resources. Species of the genus *Astatotilapia* are potentially useful for investigating this question, because populations of the same species often occupy different types of habitats (ranging from rivers to large lakes) and can differ in jaw phenotypes between habitats (Malinsky et al., 2015). Moreover, representatives of this genus have significance because they are plausible contributors to larger scale radiations, including the Lake Malawi radiation that comprises over 800 species (Joyce et al., 2011; Malinsky et al., 2018).

Here, we explore variation in the oral and pharyngeal jaw apparatus of two ecomorphs of *Astatotilapia calliptera* from Lake Masoko, an isolated crater lake inside the Lake Malawi catchment. Here, the deep/benthic ecomorph dominates the deeper habitat (> 20m) and primarily feeds on a diet of offshore food resources (zooplankton). By contrast the shallow/littoral ecomorph dominates the shallow habitat (< 5m) and primarily feeds on a diet of inshore littoral prey (benthic invertebrates). Previous work has demonstrated genomic divergence between the ecomorphs, including identification of candidate genes involved in functional adaptations such as morphogenesis (Malinsky et al., 2015). In addition, divergence in lower pharyngeal jaw morphology (Malinsky et al., 2015; Carruthers et al., 2022) and the lower pharyngeal jaw transcriptome (Carruthers et al., 2022) have been clearly shown.

We build on this earlier work by testing for divergence in oral jaw apparatus between the two distinct ecomorphs in the natural populations. We also investigate plasticity and

heritability of pharyngeal and oral jaw apparatus, by rearing representatives of the ecomorphs on two distinctive diets in a laboratory common garden experiment.

2.3 Methods

Wild caught fish

Astatotilapia calliptera were collected from Lake Masoko in October 2019 using monofilament block nets and SCUBA (Carruthers et al. 2022). Samples were collected from shallow (< 5m) and deep (> 20m) environments, targeting deep and shallow ecomorphs, respectively (n=38; 18 deep, 20 shallow). Fish were then euthanized, laid flat and pinned to a polystyrene surface to be photographed and have standard length measured. Long term preservation of samples was in ethanol.

Laboratory set-up

Parental fish used to generate experimental individuals were collected from Lake Masoko in 2016 from < 5m (shallow) and > 20m (deep) habitats and housed at Bangor University. Deep and shallow groups were bred independently, using one father and two mothers from their respective environment and ecomorph. Offspring were reared in single clutch broods in standard aquarium conditions of soft tap water, a temperature of ~25°C, under fluorescent lighting (8am-8pm), and a diet of crushed aquarium flake food until they reached a standard length of 4cm. Fish were then transferred to single-fish tanks and separated into two diet treatments. Fish in the soft diet treatment were fed on crushed aquarium flakes, while those in the hard diet treatment were fed a mixed diet of hard pelleted food, and *Melanoides* snails. When the fish could reliably process *Melanoides* their diet was restricted to *Melanoides*. The respective diet treatments were continued for four months prior to sacrifice via anaesthetic overdose. At this point specimens were pinned to a flat surface, keeping samples as straight/flat as possible for ease of measuring later (n=22; 12 deep, 10 shallow, 11 hard-diet, 11 soft-diet, visual representation of the set up found in Fig. S1). Long term preservation of samples was in ethanol.

Craniofacial measurements

Samples were placed on a flat, white surface with a ruler for scale. Photographs of each sample were taken using a Nikon COOLPIX AW110 attached to a clamp positioned directly

above each sample. Initially a TPS file was generated using tpsUtil64 v1.81 (Rohlf, 2015) and then tpsDig2 v2.3.1 (Rohlf, 2015) was used for calibration and to take linear measurements of four craniofacial traits (Fig. 2.1).

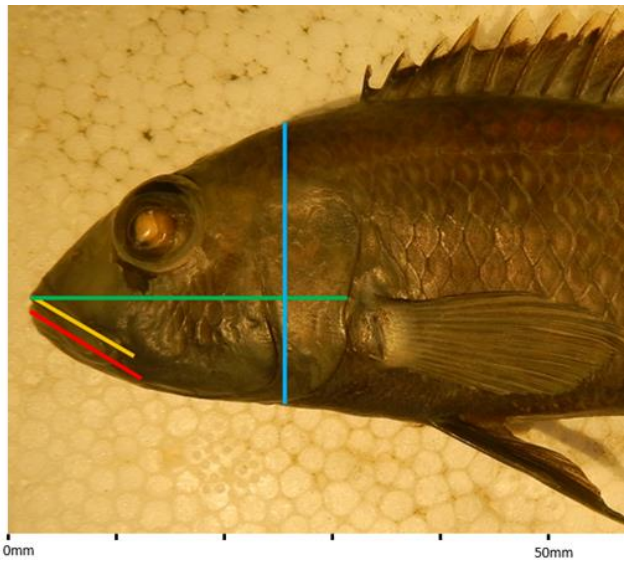


Figure 2.1. Craniofacial measurements. Coloured lines represent measurement taken in millimetres using "tpsDig232". Orange: upper oral jaw, red: lower oral jaw, green: head length, blue: head depth. Scale along the bottom of the image.

Computerized tomography (CT) scans

CT scans were performed on the laboratory-reared fish only. Wild samples had their lower pharyngeal jaws removed for molecular analysis immediately after collection. Scans were performed using the Nikon XTH225ST micro-CT scanning system at the XTM facility, Palaeobiology Research Group, University of Bristol. Fish were scanned in sets of 10, grouped according to similarities in body size. Scan groups were stacked in a rectangular plastic jar, using a layer of soft sponge between each sample to ensure there was a sufficient gap between each sample. The heads of each sample were orientated in the same position within the jar. Once all ten samples were in the jar, we placed layers of sponge on top until the jar was full, sufficiently compressed to ensure no movement during the scans.

Jars were placed in the scanner and orientated with the lower pharyngeal jaw (LPJ) as central as possible, then a 360-degree rotation of the jar ensured it would be captured throughout the scan. Parameters for the scans were as follows: beam energy of 80kV, average effective pixel size of 23µm, power fixed at 23.0W, frames to average at 86, projections at 2750. The scanner was programmed to scan two samples at a time and took roughly three hours to scan a set of ten samples.

Processing CT scans

Scans were converted into TPS files using CTpro v4.4.3. (Nikon Metrology) to find the centre of rotation for each sample and to crop the images. VGstudios v3.0 (Volume Graphics GmbH, 2016) was used to select and separate 3D segments of the scans, each containing one sample. Avizo lite v9.7.0 (ThermoFisher Scientific, 2018) was used to make a 3D reconstruction of each sample and then the lasso tool was used to isolate the LPJ from the rest of the sample. Once the LPJ for each sample had been isolated, 2D screen grabs of the LPJ's from a dorsal view were taken in orthographic mode with x and y axis scales (mm) added. TpsDig2 version 2.3.1 (Rohlf, 2015) was used to add landmarks and take three tooth width measurements (mm) from the 2D screen grabs using the configuration outlined in Fig. 2.2, after adjusting for scale.

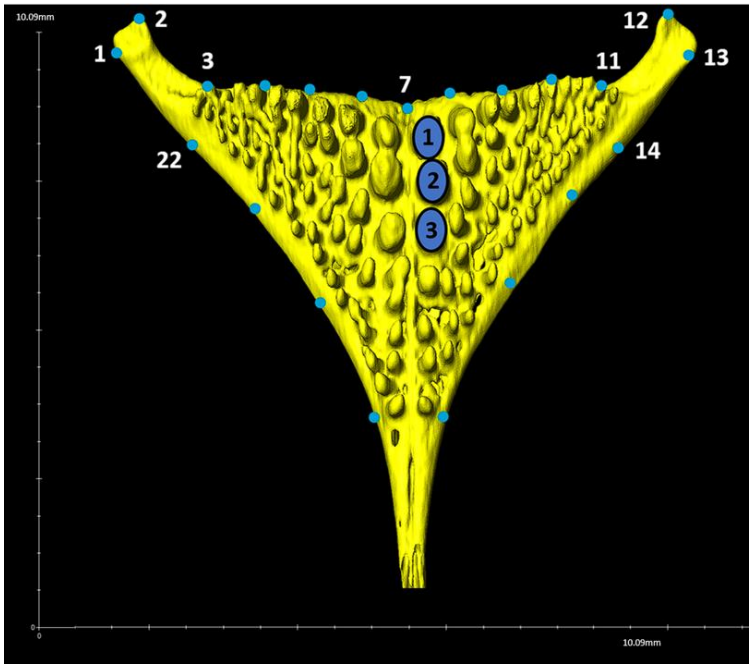


Figure 2.2: 2D screen grab of the dorsal view of an LPJ. Blue dots and white numbers outline the landmark scheme used to capture the LPJ shape. The three numbered blue circles with black outlines are the teeth that were measured for tooth width(mm), this was done from left to right. Scale along the right and the bottom of image.

Analysis

LPJ shape was analysed using a covariance principal component analysis on the landmark coordinates from each 2D image. Tooth width was analysed by taking an average of three tooth width measurements (mm). Craniofacial measurements were corrected for standard length by generating linear models from log-transformed trait measurements and log-transformed standard lengths, improving the normality of both measurements and the fit of each linear model. This was done separately for wild and lab datasets. In order to test the effects of ecomorph/parental ecomorph, treatment (lab only) and sex on each craniofacial

trait, I preformed analysis of variance (ANOVA) tests with the R package car (v3.1; Fox & Weisberg, 2019). All data analysis was performed in R v2021.09.1.

2.4 Results

Craniofacial measurements of wild fish

In wild samples there were significant positive associations between all craniofacial measurements and standard length; upper oral jaw (adjusted-R^2 = 0.518, $F_{1,36}$ = 40.72, P < 0.001), lower oral jaw (adjusted-R^2 = 0.584, $F_{1,36}$ = 52.99, P < 0.001), head length (adjusted-R^2 = 0.788, $F_{1,36}$ = 138.3, P < 0.001) and head depth (adjusted-R^2 = 0.636, $F_{1,36}$ = 65.69, P < 0.001).

Table 2.1: Anova tests for each corrected craniofacial measurement in wild samples, testing the effects of ecomorph and sex. Significant (* (P < 0.05), ** (P < 0.01), *** (P < 0.001)) results are in bold.

Measured variable	Predictor variables	Df	Sum Squared	F-value	*P*-value
Corrected upper oral jaw	Ecomorph	1	0.084712	18.0052	**0.0001604 *** **
	Sex	1	0.030339	6.4485	**0.0158468 ***
	Ecomorph:Sex	1	0.000083	0.0176	0.8953027
	Residuals	34	0.159965	-	-
Corrected lower oral jaw	Ecomorph	1	0.059201	12.5448	**0.001177 ** **
	Sex	1	0.029274	6.2033	**0.017797 ***
	Ecomorph:Sex	1	0.005266	1.1160	0.298236
	Residuals	34	0.160451	-	-
Corrected head length	Ecomorph	1	0.017820	7.4493	**0.009976 ** **
	Sex	1	0.002852	1.1923	0.282551
	Ecomorph:Sex	1	0.016572	1.1923	**0.012671 ***
	Residuals	34	0.081335	-	-
Corrected head depth	Ecomorph	1	0.00557	0.4333	0.51481
	Sex	1	0.03188	2.4819	0.12442
	Ecomorph:Sex	1	0.03871	3.0132	0.09165
	Residuals	34	0.43676	-	-

We found significant effects of ecomorph on upper/lower oral jaw and head length, deep benthic ecomorph has a longer upper/lower oral jaw and head than the shallow littoral ecomorph (Fig. 2.3). There was a significant effect of sex on upper/lower oral jaw, with males having larger jaws than females. We only found one significant interaction between sex and ecomorph, in head length (Table 2.1), with deep males having longer heads than

shallow males, and a larger difference in head length between sexes in the deep ecomorph compared to the shallow ecomorph.

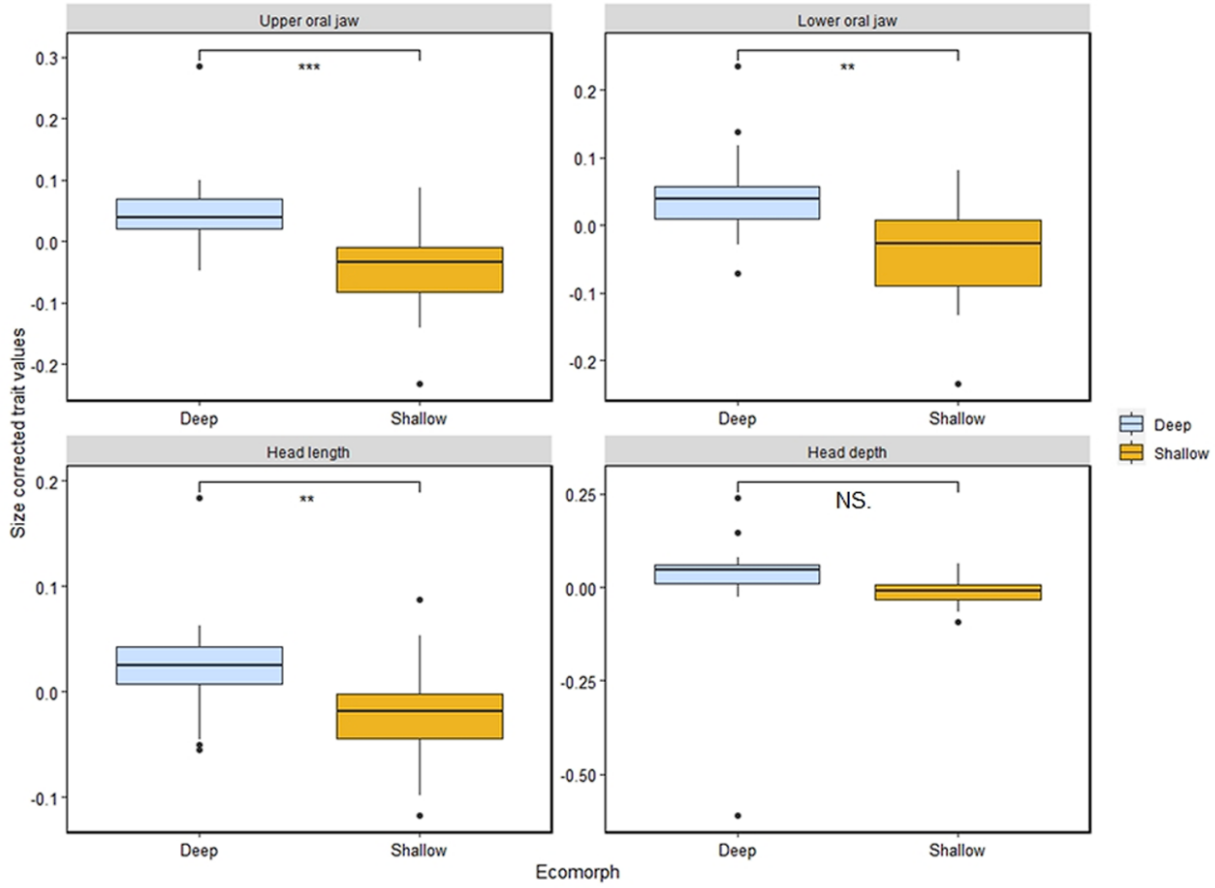


Figure 2.3: Size corrected trait values for wild Lake Masoko samples (n=38; 18 deep, 20 shallow). Boxes represent upper and lower quartiles and bar within the boxes represent the median, lines represent 1.5 times the interquartile ranges respectively and dots represent outliers. Significance labels NS. (non-significant), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) show the effect of ecomorph on a given corrected trait value.

Craniofacial measurements of laboratory-reared fish

In wild samples there were significant positive associations between all craniofacial measurements and standard length in the laboratory reared samples; upper oral jaw (adjusted-$R^2 = 0.829$, $F_{1,20} = 102.5$, $P < 0.001$), lower oral jaw (adjusted-$R^2 = 0.554$, $F_{1,20} = 27.09$, $P < 0.001$), head length (adjusted-$R^2 = 0.828$, $F_{1,20} = 102.3$, $P < 0.001$) and head depth (adjusted-$R^2 = 0.908$, $F_{1,20} = 208.8$, $P < 0.001$).

Table 2.2: Anova tests for each corrected craniofacial measurement within laboratory-reared fish, testing for the effects of treatment, parental ecomorph and sex. Significant (* ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$)) results are in bold.

Measured variable	Predictor variables	Df	Sum Squared	F-value	*P*-value
Corrected upper oral jaw	Treatment	1	0.000504	0.1111	0.7438
	Parental ecomorph	1	0.010063	2.2182	0.1586
	Sex	1	0.002287	0.5042	0.4893
	Treatment:parental ecomorph	1	0.003683	0.8119	0.3828
	Treatment:sex	1	0.003168	0.6983	0.4174
	Parental ecomorph:sex	1	0.000573	0.1263	0.7276
	Treatment:parental ecomorph:sex	1	0.000074	0.0163	0.9002
	residuals	14	0.063510	-	-
Corrected lower oral jaw	Treatment	1	0.011480	2.1274	0.166757
	Parental ecomorph	1	0.087635	16.2405	**0.001241 ****
	Sex	1	0.001277	0.2366	0.634201
	Treatment:parental ecomorph	1	0.004702	0.8714	0.366385
	Treatment:sex	1	0.006678	1.2375	0.284682
	Parental ecomorph:sex	1	0.000111	0.0207	0.887761
	Treatment:parental ecomorph:sex	1	0.000022	0.0041	0.949726
	residuals	14	0.075545	-	-
Corrected head length	Treatment	1	0.0015189	0.9762	0.339916
	Parental ecomorph	1	0.0203919	13.1057	**0.002785 ****
	Sex	1	0.0000008	0.0005	0.982117
	Treatment:parental ecomorph	1	0.0006780	0.4358	0.519888
	Treatment:sex	1	0.0002415	0.1552	0.699514
	Parental ecomorph:sex	1	0.0000969	0.0623	0.806520
	Treatment:parental ecomorph:sex	1	0.0003663	0.2354	0.635035
	residuals	14	0.0217834	-	-
Corrected head depth	Treatment	1	0.0003963	0.4053	0.534620
	Parental ecomorph	1	0.0130094	13.3048	**0.002638 ****
	Sex	1	0.0003014	0.3083	0.587512
	Treatment:parental ecomorph	1	0.0002969	0.3036	0.590295
	Treatment:sex	1	0.0009348	0.9561	0.344775
	Parental ecomorph:sex	1	0.0005640	0.5768	0.460150
	Treatment:parental ecomorph:sex	1	0.0008167	0.8353	0.376219
	residuals	14	0.0136891	-	-

We found that in laboratory-reared fish parental ecomorph had a significant effect on lower oral jaw, head length and head depth (Table 2.2). As observed in the wild fish, we found the deep benthic ecomorph has a longer lower oral jaw and head length than the shallow littoral ecomorph (Fig. 2.4). Additionally, we found that the deep benthic ecomorph has a shorter head depth than the shallow littoral ecomorph (Fig. 2.4).

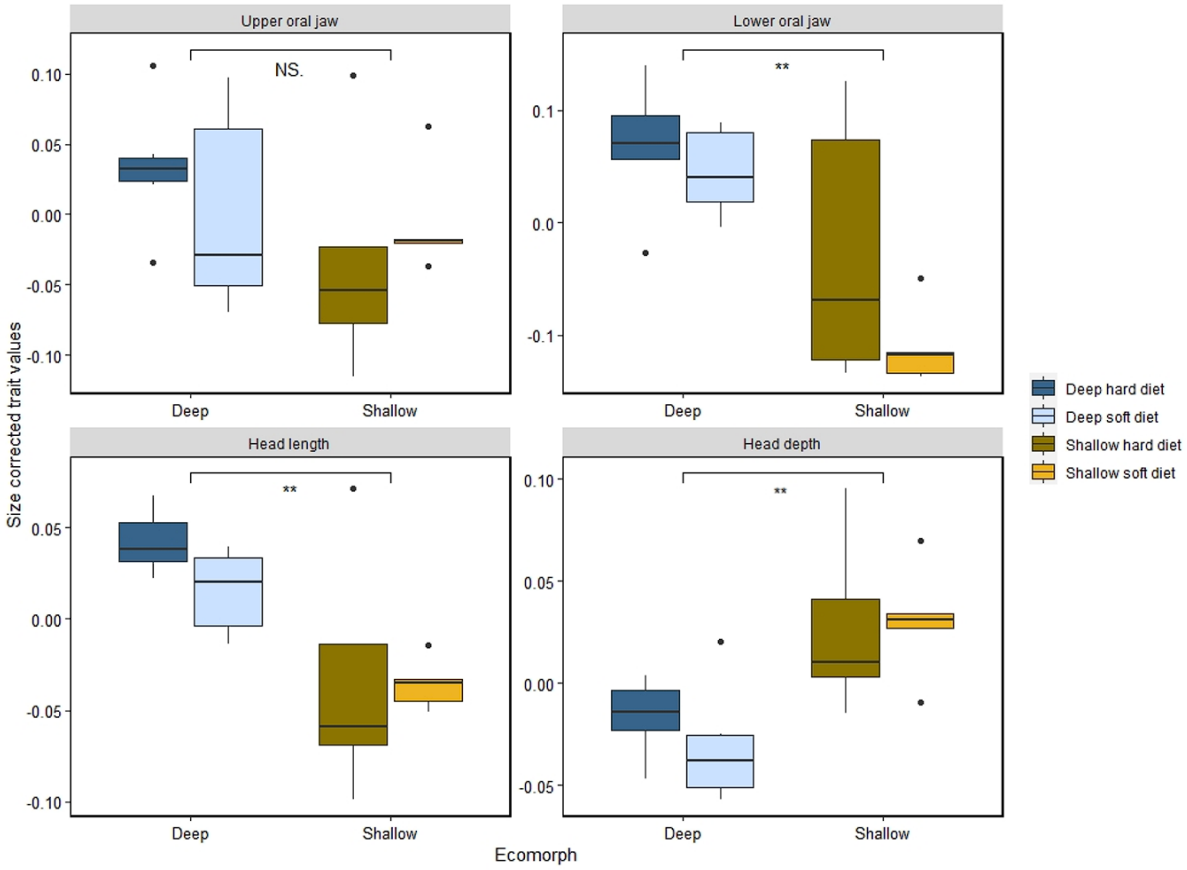


Figure 2.4: Size corrected trait values for diet-controlled lab samples (n=22; 12 deep, 10 shallow, 11 hard diet, 11, soft diet). Boxes represent upper and lower quartiles and bar within the boxes represent the median, lines represent 1.5 times the interquartile ranges respectively and dots represent outliers. Significance labels NS. (non-significant), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) show the effect of ecomorph on a given corrected trait value.

Lower pharyngeal jaw (LPJ)

Table 2.3: Anova tests for LPJ shape PC1, LPJ shape PC2 and average tooth width, testing for the effects of treatment, parental ecomorph and sex. Significant (* ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$)) results are in bold.

Measured variable	Predictor variables	Df	Sum Squared	F-value	*P*-value
LPJ shape PC1	Treatment	1	0.0003459	0.9258	0.3523
	Parental ecomorph	1	0.0235705	63.0860	**1.489e-06 *** **
	Sex	1	0.0003025	0.8096	0.3835
	Treatment:parental ecomorph	1	0.0000171	0.0458	0.8337
	Treatment:sex	1	0.0003034	0.8121	0.3827
	Parental ecomorph:sex	1	0.0003616	0.9679	0.3419
	Treatment:parental ecomorph:sex	1	0.0001779	0.4762	0.5014
	residuals	14	0.0052308	-	-
LPJ shape PC2	Treatment	1	0.0065445	23.7202	**0.0002477 *** **
	Parental ecomorph	1	0.0013983	5.0682	**0.0409552 ***
	Sex	1	0.0002518	0.9126	0.3556281
	Treatment:parental ecomorph	1	0.0002257	0.8182	0.3810148
	Treatment:sex	1	0.0005324	1.9295	0.1865099
	Parental ecomorph:sex	1	0.0005364	1.9442	0.1849512
	Treatment:parental ecomorph:sex	1	0.0000471	0.1708	0.6856716
	residuals	14	0.0038626	-	-
Average tooth width (mm)	Treatment	1	0.037797	44.4052	**1.072e-05 *** **
	Parental ecomorph	1	0.000019	0.0226	0.8826649
	Sex	1	0.020152	23.6755	**0.0002498 *** **
	Treatment:parental ecomorph	1	0.002213	2.5999	0.1291754
	Treatment:sex	1	0.002419	2.8419	0.1139847
	Parental ecomorph:sex	1	0.000594	0.6976	0.4176242
	Treatment:parental ecomorph:sex	1	0.000131	0.1539	0.7007331
	residuals	14	0.011917	-	-

We found that LPJ shape PC1 was significantly affected by parental ecomorph. Variation on PC1 showed that fish with deep ecomorph parents had LPJs that were smaller, narrower and with more inward angled wings relative to fish with shallow ecomorph parents. LPJ shape PC2 was significantly affected by diet treatment and parental ecomorph. Variation on PC2 showed that fish reared on the soft diet treatment had slightly smaller and narrower LPJs

compared to those reared on the hard diet treatment (figure 2.5). Average tooth width was significantly affected by diet treatment and sex (Table 2.3). Results showed that hard diet treatment led to a wider average tooth width compared to the soft diet treatment (figure 2.5), while males had a wider average tooth width relative to females.

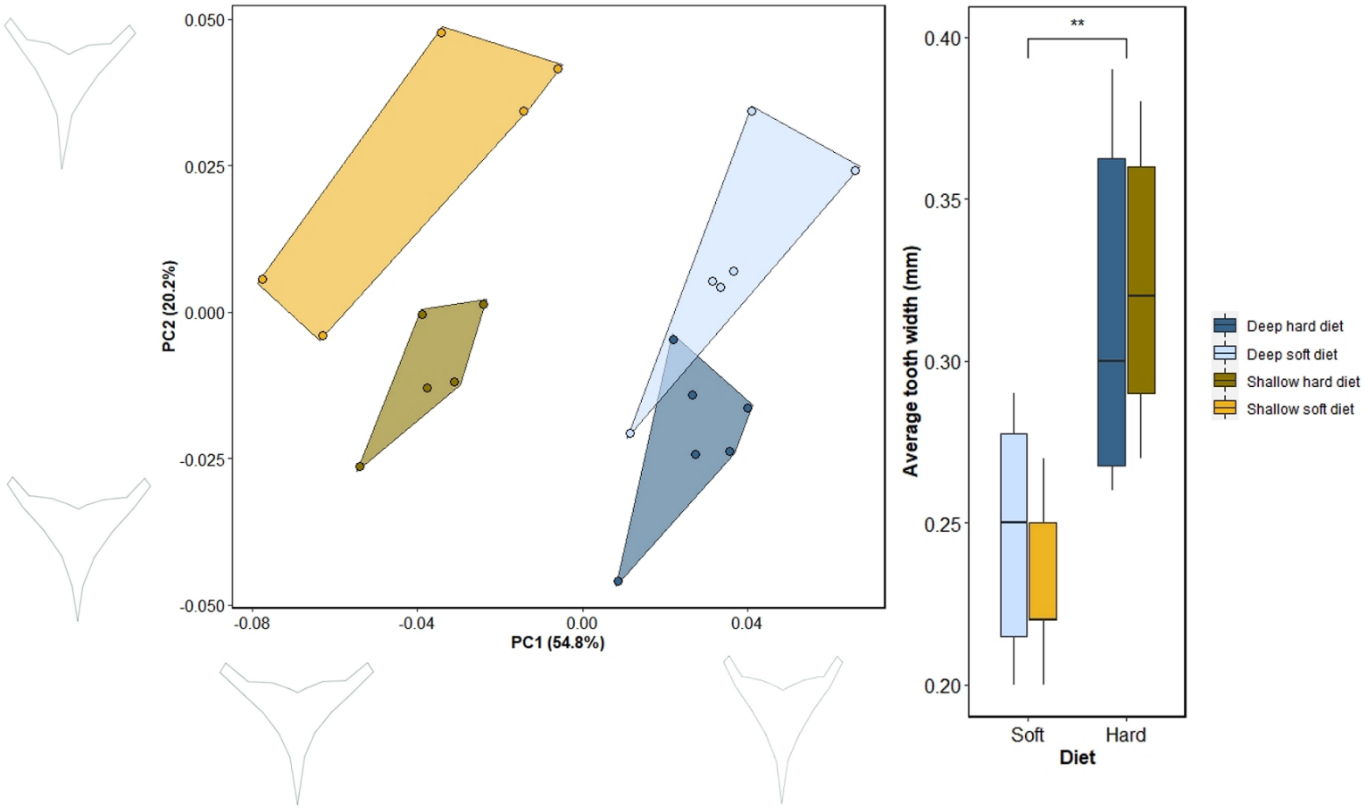


Figure 2.5: Lower pharyngeal jaw and tooth width results from diet-controlled laboratory-reared samples (n=22; 12 deep, 10 shallow, 11 hard diet, 11, soft diet). *Left*: Principal Component Analysis of lower pharyngeal jaw shape. Colours represent ecomorph and treatment. Numbers in brackets next to axis labels indicate the percentage of total variance captured by the axis. Wireframes provide a visual representation of shape at PC1: -0.08 and 0.08, and PC2: -0.05 and 0.05. *Right*: Differences in average tooth width (mm). Boxes represent upper and lower quartiles and bar within the boxes represent the median, lines represent 1.5 times the interquartile ranges respectively and dots represent outliers. Significance labels NS. (non-significant), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$) show the effect of treatment on average tooth width (mm).

2.5 Discussion

Oral jaw morphology

This study investigated the plasticity and heritability of oral apparatus of *A. calliptera* from Lake Masoko. We found that oral jaws showed phenotypic differences between ecomorphs in the wild, with the benthic ecomorph generally having larger oral jaws and longer heads

relative to the littoral ecomorph. These differences are consistent with a function in acquisition of different prey types, with larger and longer jaws potentially improving the ability of the benthic ecomorph to capture plankton in open water. Moreover, differences in oral jaw structures between the ecomorphs were maintained in laboratory conditions, and were not significantly affected by the diet on which they were reared. This suggests that oral jaw structures are not intrinsically phenotypically plastic, and instead are regulated by genotypic differences between the ecomorphs. However, the craniofacial measurements investigated here only capture the size of individual features of the head. This may have missed some of the more subtle differences in shape and teeth arising from the differing diet treatments. Analyses of the upper and lower oral jaw shape in common garden experiment samples, based on information from CT scans, may expand upon our findings and potentially find aspects of plasticity based on diet.

It is plausible that the differences between the ecomorphs in oral jaw structures are related to non-dietary factors (Malinsky et al., 2015). Maternal care via mouth brooding occurs in *A. calliptera* (Ribbink, 1990) and egg size and clutch size has been shown to differ among populations occupying differing environments (Parsons et al., 2017). Therefore, it is possible that differences in oral jaw size and head length could be a result of differing life history strategies between the two distinct ecomorphs. Further investigation into mouth brooding, life history traits and buccal cavity size of the *A. calliptera* from Lake Masoko would be necessary to confirm whether there are any trade-offs between feeding and maternal care (Tkint et al., 2012). Alternatively, the oxygen content of the water may impact oral jaw size. Bouton et al. (2002) investigated head shape of haplochromine cichlids populations from Lake Victoria and found oxygen levels were drivers of gill space and opercular volume. Since the deeper water habitat occupied by the benthic ecomorphs is low in oxygen when the lake is stratified (Delalande 2008), this may select for larger oral jaws and longer heads in the deep ecomorph.

Lower pharyngeal jaw morphology

Analyses of the lower pharyngeal jaws from fish reared in common garden experiments revealed a strong influence of parental ecomorph on jaw shape irrespective of diet. Specifically, the fish from the littoral ecomorph had broader jaws relative to those observed in the benthic ecomorph, consistent with differences observed among wild fish and with a

major role for heritable genetic determination of jaw shape. Additionally, the lower pharyngeal jaws showed differences between the diet treatments, with fish from the hard diet treatment developing wider and larger jaws, and wider teeth. These changes reflect the expected influence of phenotypic plasticity on jaw shape based on differing diet treatments. These findings are consistent with previous studies of pharyngeal jaw morphology following dietary manipulation in cichlids, including in the East African Alluaud's haplo *Astatoreochromis alluaudi* (Gunter et al., 2013) and the Central American Midas cichlid *Amphilophus citrinellus* (Muschick et al., 2011).

Evolutionary implications

Our results demonstrate a clear role of diet induced plasticity in pharyngeal jaw morphology but evidence for diet induced plasticity in oral jaw morphology is less clear. This partially supports the hypothesis that the two jaw structures have become evolutionarily decoupled and are following different evolutionary trajectories (Liem, 1973; Hulsey et al., 2016). It is plausible that this reflects differences in the genetic architecture of the traits, including the presence or absence of standing genetic variation that underpins phenotypic variation, and the selection pressures faced by the populations.

Plastic traits allow organisms to adapt and thrive in novel and heterogenous environments. It is possible that the plasticity of pharyngeal jaw apparatus has allowed *A. calliptera* to process a variety of diets and successfully occupy and adapt to both benthic and littoral environments in Lake Masoko (Joyce et al., 2011). An ability to survive and adapt to differing environments may facilitate and/or drive evolutionary processes such as ecological speciation (Lafuente & Beldade, 2019). It has been proposed that environmentally-induced phenotypic change (Levis & Pfennig, 2016; West-Eberhard, 2003) can be a precursor to genetic fixation of those traits, effectively buying time for genetic mutations to occur. Previous research has shown a limited number of fixed genetic variants between the Lake Masoko ecomorphs (Malinsky et al., 2015), despite substantial transcriptional divergence between ecomorphs (Carruthers et al., 2021). This suggests that an understanding of their phenotypic divergence will require a thorough understanding of the mechanisms underpinning differential gene expression, and the heritability of that gene expression (Pavey et al., 2010; Skelly et al., 2009).

While genetic mechanisms are likely to be important components underpinning both the observed morphological differences between ecomorphs, and the plasticity that has been induced in our experiment, it is also possible that epigenetic mechanisms such as DNA methylation are responsible for regulating gene expression may play a role (Pfennig & West-Eberhard, 2021). Recent research has discovered transcriptionally-functional epigenetic divergence among ecologically-divergent haplochromine cichlids in the Lake Malawi radiation (Vernaz, et al., 2021) and between the ecomorphs of *A. calliptera* in Lake Masoko (Vernaz, et al., 2022). Moreover, some evidence of transgenerational heritability of methylation marks has been identified (Vernaz et al., 2021). This provides a preliminary indication that methylation may have a role in enabling plastic morphological traits of *A. calliptera*, and be subject to selection. Our results lay the groundwork for further investigations into the underlying molecular mechanisms regulating lower pharyngeal jaw plasticity. Next steps would be associating environmentally induced phenotypic changes with changes in DNA methylation and gene expression, and looking at subsequent generations to assess if these changes are heritable.

Conclusions

This study has revealed that oral jaw morphology and head length show heritable differences in size in wild populations, but do not differ based on diet. Further investigations into other selection pressures present in Lake Masoko, and more detailed analysis into the shape and teeth of oral jaws will elaborate on our findings. By contrast, lower pharyngeal jaws show aspects of phenotypic plasticity and heritability which is likely to have contributed to the diversification of *A. calliptera*. Further work into the mechanisms driving divergence in gene expression of the populations, and those leading to environmentally-induced contrasting phenotypes, would help to clarify the role of environmental plasticity and transcriptome heritability in the adaptive radiation of cichlids.

Chapter 3: Gene co-expression networks uncover regulatory shifts in ecologically functional phenotypes

3.1 Abstract

Gene expression is thought to play an essential role in ecological speciation, by generating phenotypic variation in response to different local environments within diversifying populations. In the East African crater Lake Masoko, *Astatotilapia calliptera* is undergoing the early stages of ecological speciation into two distinct ecomorphs, a shallow-water littoral ecomorph and a deep-water benthic ecomorph. This system presents an excellent opportunity to investigate the molecular mechanisms underpinning adaptive phenotypic shifts during the initial stages of ecological diversification. Using whole transcriptome data from lower pharyngeal jaw tissue, we performed a gene network analysis, which identified candidate modules of co-expressed genes that were significantly associated with phenotypic shifts in craniofacial traits. Candidate modules were enriched for several ecologically relevant GO terms. Furthermore, we found that a high proportion of key regulator (hub) genes within candidate jaw apparatus modules exhibited divergent methylation patterns between ecomorphs, suggesting DNA methylation is an important regulator of jaw and craniofacial phenotypes in our focal species. Overall, these results show that gene expression is a crucial molecular mechanism that regulates the jaw apparatus of speciating *A. calliptera*, and that network co-expression analyses are a valuable tool for uncovering functional diversification of transcriptomes.

3.2 Introduction

Understanding the molecular basis of phenotypic traits that drive intraspecific variation, leading to local adaptation and ultimately ecological speciation, is a major goal in evolutionary biology. However, uncovering the molecular mechanisms underpinning phenotypic trait evolution is challenging. Rapid advances in next-generation sequencing, such as RNA-seq, have dramatically improved our ability to tackle such complex evolutionary questions by enabling high-resolution transcriptome-wide analysis of gene expression. Gene expression is fundamental in bridging the genotype and phenotype gap, as well as for clarifying gene x environment interactions (Kratochwil & Meyer, 2015). Gene expression is suggested to have two important roles in ecological speciation (Pavey et al., 2010). First, it potentially contributes to adaptive genetic divergence and facilitates reproductive isolation (Pavey et al., 2010). Second, it may enable population persistence in variable and novel environments through phenotypic plasticity. By associating transcriptomes to phenotypes, we can further our understanding of the molecular mechanisms generating functionally adaptive phenotypes, that act as targets of natural selection.

Relative to their extensive phenotypic diversity, East African cichlids demonstrate low genetic diversity compared to other vertebrates (Salzburger, 2018; Svardal et al., 2020). Therefore, it has been suggested that gene expression plays a major role in generating the substantial phenotypic diversity observed across cichlid radiations. Brawand et al. (2014) conducted a pivotal study on the genomes and transcriptomes of East African cichlids and found divergence in gene expression between species. Advancements in next-generation sequencing and high-performance computing have enabled numerous subsequent studies exploring gene expression in East African cichlids in more detail, in Lake Tanganyika (Gunter et al., 2017; El Taher et al., 2021), Victoria (Gunter et al., 2017) and Malawi (Vernaz et al., 2021). These studies have shown that transcriptome divergence between closely related species is commonplace within these lake radiations, potentially allowing adaptation in a variety of environments, and contributing to numerous ecological speciation events.

Epigenetic mechanisms can regulate gene expression, via histone modifications, non-coding RNAs and DNA methylation. These mechanisms can be environmentally induced and

bisulphite sequencing now allows us to look at DNA methylation patterns of an individual's genome in unprecedented detail. In conjunction with RNA sequencing data, we can begin to bridge the genotype-phenotype gap (Kratochwil & Meyer, 2015; Chevin et al., 2021), to improve our understanding of phenotypic divergence and ecological speciation. In East African cichlids DNA methylation mechanisms are beginning to be incorporated into gene expression studies. Their exact role of methylation in ecological speciation is still unclear, but East African cichlids provide an excellent study model for investigating this issue due to their huge phenotypic and transcriptomic diversity. Vernaz, et al. (2021) mapped the transcriptomes and methylomes of liver and muscle tissue from several Lake Malawi species. They found evidence of methylome divergence in ecologically relevant genes between species in both tissues, and species-specific differentially methylated regions (DMRs) were shared between tissues. These results suggest that DNA methylation is an important transcriptional regulator and might be an important mechanism contributing to ecological speciation.

This chapter explores transcriptome-wide gene expression divergence in the diverging ecomorphs of *Astatotilapia calliptera* from Lake Masoko. Data analysed in this chapter are taken from Carruthers et al. (2022), where whole transcriptomes of lower pharyngeal jaws from the deep water benthic and shallow water littoral ecomorphs were investigated. Carruthers et al. (2022) identified substantial shifts in gene expression between ecomorphs, associated with divergence in ecologically functional traits, and genetic variants under selection.

In this chapter, I build on this work by performing a weighted gene co-expression network analysis (WGCNA) to identify regulatory shifts in co-expressed gene modules directly associated with ecomorph divergence and variation in adaptive craniofacial phenotypes. I perform functional enrichment analysis of trait-associated modules to understand major molecular pathways underpinning trait diversification. I also identify key regulator 'hub' genes within candidate trait-linked modules and examine the association between key regulator genes and genes under divergent methylation. Finally, I examine whether genes identified as being hub genes, or under divergent methylation, or both, within trait-associated modules include any genes that have known roles in lower pharyngeal jaw (LPJ) plasticity and adaptation networks. Overall, this chapter will explore the association

between gene co-expression networks, DNA methylation and phenotypic variation in the diverging *A. calliptera* Lake Masoko population, advancing our understanding of the role of gene expression and DNA methylation in craniofacial divergence linked to trophic adaptation and ecological speciation.

3.3 Methods

Gene expression data

Gene expression data analysed in this chapter were taken from Carruthers et al. (2022). Briefly, total RNA was extracted from the LPJs of 38 wild individuals of *Astatotilapia calliptera* from Lake Masoko (18 deep-benthic and 20 shallow-littoral ecomorphs; *see section 2.3 Methods, wild caught fish*) and sequenced at a depth of approximately 30 million reads per sample. Sequences were aligned to the *Maylandia zebra* reference genome (UMD2a NCBI assembly: GCF_000238955.4) (Conte et al., 2019) using STAR *v*. 2.7.1a (Dobin et al., 2013), and HTseq *v*.0.11.1 (Anders et al., 2015) was used to quantify gene expression and generate read counts. Read counts were filtered to remove low coverage genes (<10 reads across 90% of samples), normalised and $\log_2$ scaled using the *rlog* function in DESeq2 v.1.28.1 (Love et al., 2014). This resulted in a final set of 19,239 expressed genes.

Gene co-expression network analysis

A weighted gene co-expression network analysis (WGCNA) was used to identify modules of co-expressed genes associated with functional divergence between the benthic and littoral ecomorphs. The R package WGCNA (v.1.70; Langfelder & Horvath, 2008) was used to construct a single network for all 38 individuals, based on the normalised count data and following the author's recommended pipeline. Briefly, we performed hierarchical clustering of the 38 samples based on their overall gene expression profiles to identify sample relationships and outliers (no outliers were identified; Fig. S2). Next, a gene x gene similarity matrix was constructed based on Pearson correlation coefficients, which included all genes and all samples, and provided a measure of gene co-expression (Fig. S3). Co-expression coefficients were then raised to a chosen power of 16 to create the adjacency matrix; soft thresholding power was chosen based on biologically motivated criteria according to scale free topology fit indexes (Fig. S3), selecting the lowest integer above the model fit of R^2=0.9.

WGCNA was then used to calculate the topological overlap matrix (TOM), which is a measure of network interconnectedness; i.e. the strength of a co-expression relationship between any two genes with respect to all other genes in the network. Genes with highly similar co-expression patterns were clustered using hierarchical clustering of the topological overlap dissimilarity matrix (1-TOM) with the R package flashClust (v.1.10; Langfelder & Horvath, 2012); method=average. Finally, network modules were defined using the dynamicTreeCut (v1.63; Langfelder et al., 2008) algorithm, with a minimum module size of 30 genes, module eigengene (ME) dissimilarity threshold of 0.25 to merge highly similar modules.

To identify co-expressed genes associated with ecomorph specific divergence in functional traits of interest, Pearson correlations were calculated between trait measurements and module eigengenes (defined as the first principal component of the module expression profile). Module-trait relationships were tested for one binary trait (ecomorph) and four craniofacial traits (upper oral jaw, lower oral jaw, head length and head depth; see Chapter 2 for details of linear measurements used for craniofacial phenotypes). Module-trait relationships were considered significant at *P-values* <0.05 after Benjamini-Hochberg (Benjamini & Hochberg, 1995) multiple testing correction.

Functional enrichment of trait-associated modules

To identify molecular pathways enriched in trait-associated co-expression modules we performed functional overrepresentation tests using the PANTHER classification tool (Thomas et al., 2003), for each module separately. Gene ontology (GO) classifications (Ashburner et al., 2000) were assigned using the *Danio rerio* protein database within PANTHER. We used Fisher's exact tests to test for significant overrepresentation of biological process GO terms and PANTHER pathways within modules. We specified the complete set of 19,237 genes expressed in our dataset as the background for all module overrepresentation tests.

Module hub gene identification

To identify genes which may act as key regulators of ecomorph/phenotypic divergence, we identified hub genes (i.e. the most highly connected genes within a given module), for all modules significantly associated with one or more of the five traits tested. Hub genes were

identified using two criteria, module membership and gene significance scores. Module membership scores reflect the connectivity of a gene with other genes in the module, calculated from the correlation between the expression level of a gene and the module eigengene. Gene significance is a genes correlation with a given variable. Genes within the top 10% quantile of both module membership and gene significance scores were considered hub genes. Where modules were significantly associated with multiple traits, we used the gene significance scores from the strongest trait association.

All analyses were conducted in R v2021.09.1. RScript code for the WGCNA analysis (Appendix, Script 1) and hub gene detection (Appendix, Script 2) can be found in the appendix.

DNA methylation patterns associated with co-expressed modules

To assess whether co-expressed gene modules were associated with shifts in DNA methylation we looked at the overlap between genes in trait-associated modules and genes that are under divergent methylation regulation between the littoral and benthic ecomorphs. Genes identified as being differentially methylated between ecomorphs were taken from Carruthers et al. (*in prep.*). Briefly, high molecular weight genomic DNA was extracted from LPJs. Whole genome bisulphite sequence (WGBS) data were generated at the Centre of Genomic Research facility (University of Liverpool), yielding an average of 200 million 150 bp paired-end reads per individual (n=38; RNA-seq and WGBS datasets are derived from the same set of 38 individuals). Bisulphite sequences were mapped to the *M. zebra* reference genome and methylated CpGs were called following the Bismark pipeline (v.0.23; Krueger & Andrews, 2011). CpG sites were filtered to remove low coverage sites (supported by less than five reads) and high coverage sites (defined as greater than the 99th percent quantile) to remove noise from repetitive elements and paralogous genes. Differentially methylation regions (DMRs) were identified using the Bioconductor package DSS (Dispersion Shrinkage for Sequencing data, v. 2.44.0; Wu et al., 2015). DMRs were filtered based on the following parameters: minimum sequence length of 50 bp, at least five CpGs present, significance threshold < 0.05, and a minimum of 10% methylation divergence between benthic and littoral ecomorphs. This resulted in a final set 4658 DMRs of which one or more were associated with 4177 differentially methylated (DM) genes.

3.4 Results

Gene co-expression network analysis

WGCNA successfully clustered all 19,237 genes into a final network consisting of 31 modules (Fig. S4). Remarkably, we found that 80.1% of expressed genes (Table S2) (22 out of 31 modules) showed significant trait associations (Fig. 3.1). Significant trait associated module sizes ranged from 37 to 3281 genes (Fig. S5). Of these, 16 modules were associated with divergent expression between ecomorphs. We also found three modules were significantly associated with both upper and lower oral jaw, and eight modules were significantly associated with head length (Fig. 3.1). No modules were significantly associated with head depth (Fig. 3.1). Interestingly, several modules were significantly associated with multiple traits. For example, the darkseagreen3 module was associated with ecomorph, upper/lower oral jaw and head length, and the blue module was associated with upper/lower oral jaw and head length (Fig. 3.1).

Functional enrichment of trait-associated modules

Function enrichment tests were performed on all modules that were significantly associated with a trait (Fig. 3.1); full gene ontology analysis can be found in the appendix (Table S1). Briefly, modules associated with ecomorph were enriched for 82 biological processes and 2 Panther pathways. Examples included, B cell receptor signalling pathway (black module) mitochondrial respiratory chain complex III assembly (brown module), fin morphogenesis (brown2) and nervous system development (darkorange2 module). Modules associated with head length were enriched for 31 biological processes and 3 Panther pathways. Examples included, cholesterol biosynthetic process (midnightblue module) and mRNA splicing, via spliceosome (mediumpurple2 module). Modules associated with multiple traits were enriched for 46 biological processes. Examples included, oxygen transport (coral1 module), intracellular signal transduction (darkseagreen3) and pharyngeal system development (blue module).

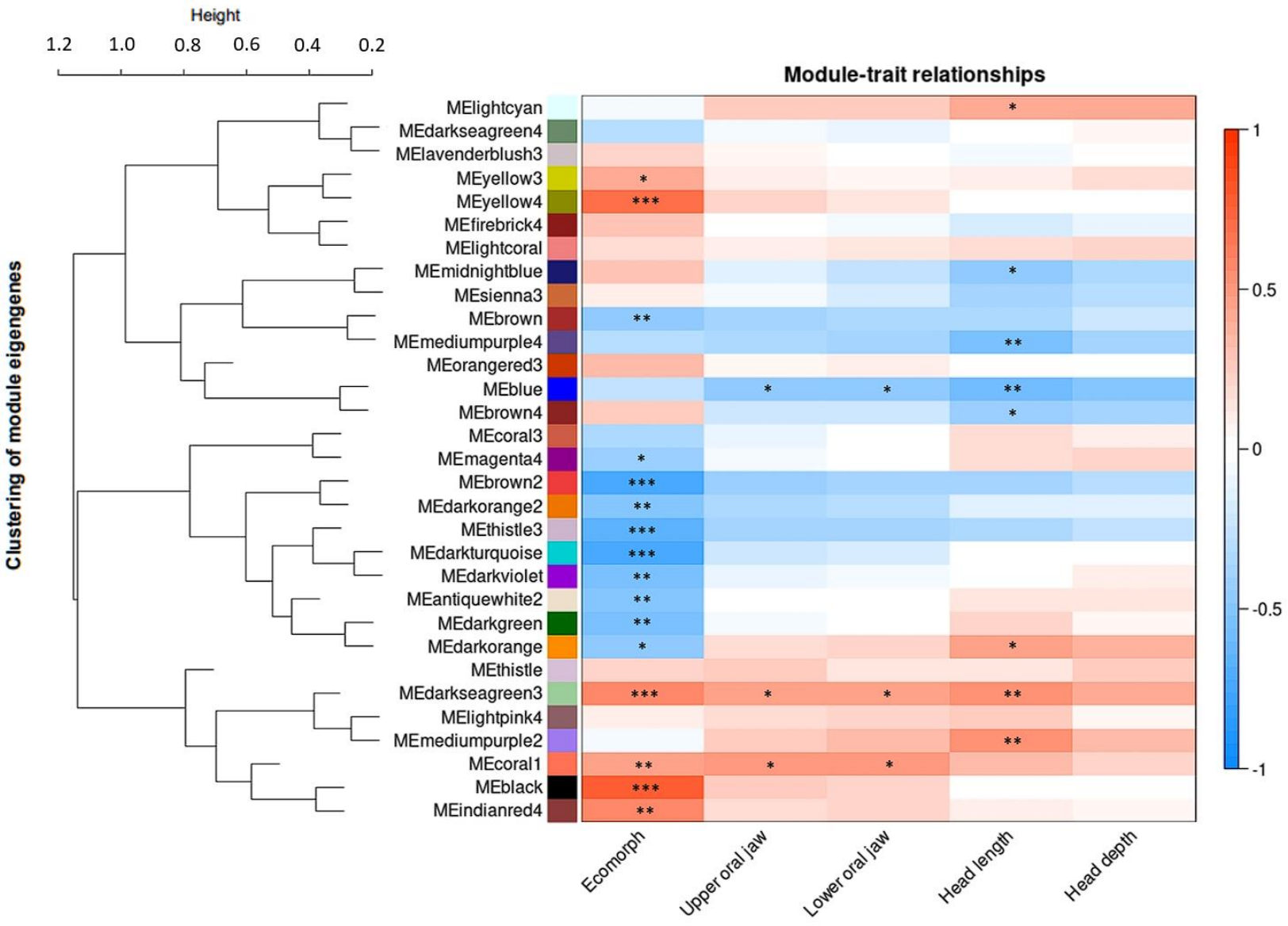


Figure 3.1: Heatmap showing the correlation WGCNA module eigengenes and measured traits (ecomorph, upper oral jaw, lower oral jaw, head length and head depth; n = 38). Each row corresponds to a module, identified on the left side by its colour (colours are assigned arbitrarily by WGCNA), the clustering of module eigengenes tree, corresponds to the average distance between all inter-cluster pairs, quantified by the scale above. Each column corresponds to a trait. Module-trait relationships are coloured based on their correlation: red represents a strong positive correlation and blue represents a strong negative correlation. Asterisks indicate significant module-trait associations (Benjamini-Hochberg corrected): * < 0.05, ** < 0.01, *** < 0.001 and blank cells indicate non-significant associations.

DNA methylation patterns associated with co-expressed modules

To investigate the role of DNA methylation in ecologically functional modules, we looked at the overlap between differentially methylated (DM) genes and genes within significant trait associated modules. DM genes were found in all significant trait-associated modules (Fig. 3.2) and 51.3% of DM genes were found within our significant modules (Table 3.1). We found that the distribution of DM genes across our dataset appeared evenly distributed, ranging from 11.5%-13.9% (Table 3.1). We found that 2144 out of 2586 DM genes within our expression dataset, were found within significant trait associated modules (hypergeometric test (Wang et al., 2015), 1 million simulations, *P* value = 0.011).

Table 3.1: Shows the number of genes and DM genes in significant trait associated modules: 22 modules significantly associated with one or more of our five traits, non-significant trait associated modules: 9 modules not significantly associated with one or more of our five traits, and not expressed: these are genes within the genome that were not found to be expressed in the RNA-seq dataset.

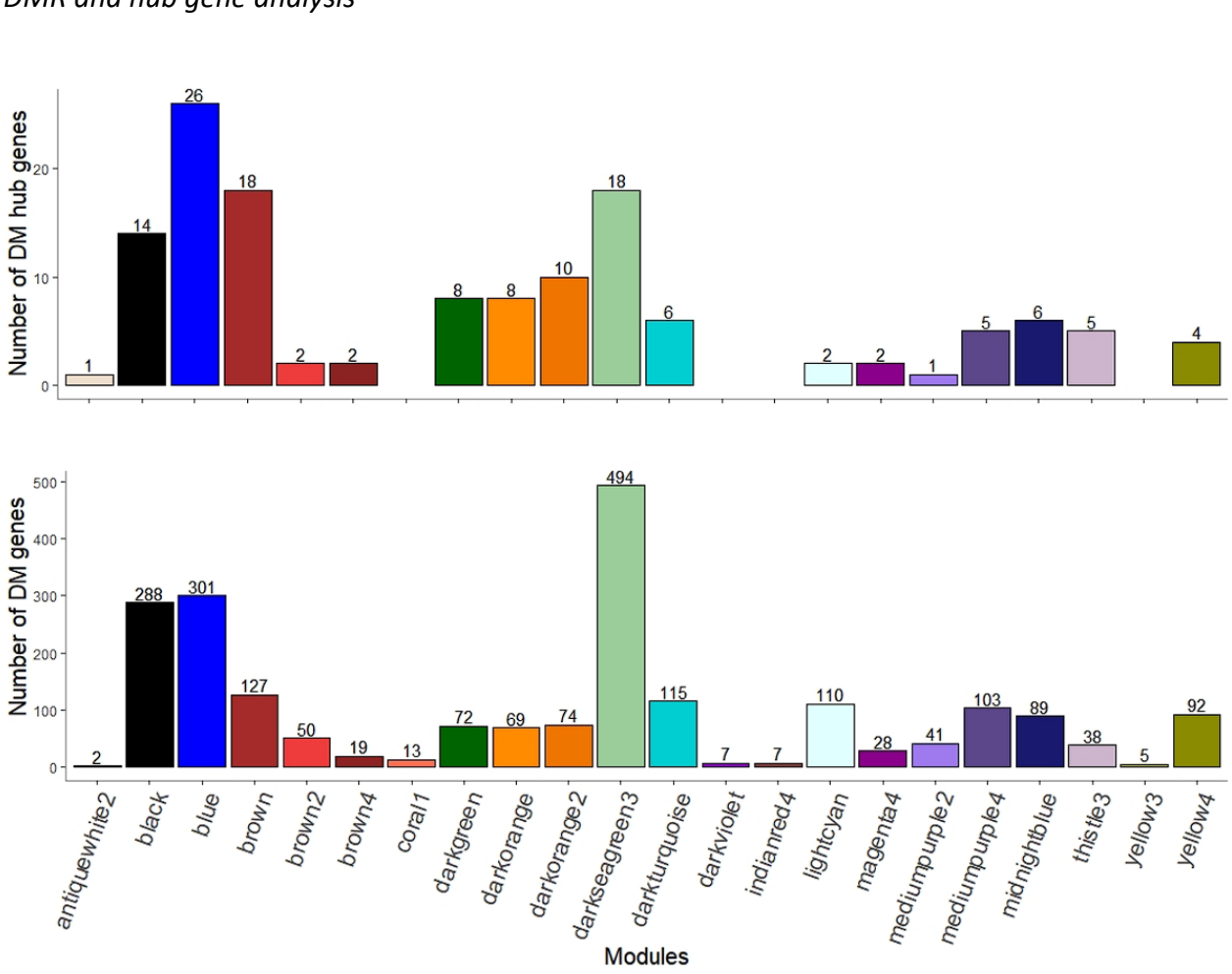
Dataset	**Number of genes (%)**	**Number of DM genes (%)**	**Percentage of DM genes/number of genes in dataset**
Significant trait associated modules	15408 (49.3%)	2144 (51.3%)	13.9%
Non-significant trait associated modules	3829 (12.3%)	442 (10.6%)	11.5%
Not expressed (excluded from the network)	11983 (38.4%)	1591 (38.1%)	13.3%
Total	31220 (100%)	4177 (100%)	13.4%

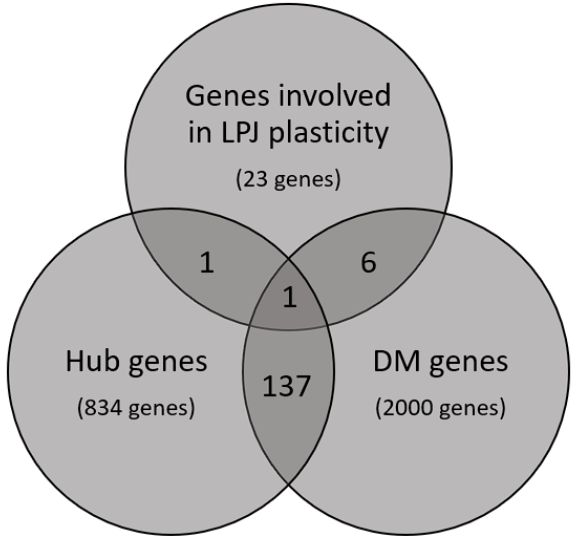
To investigate whether DNA methylation acts a potential regulator of functional co-expression divergence between ecomorphs, we first conducted a hub gene search for each of the 22 trait-associated modules. The number of hub genes within each module ranged from 1 to 228 (Fig. S6). Second, we determined whether key regulator 'hub' genes were associated with ecomorph-specific patterns of DNA methylation. We found that 138 out of 975 hub genes (14.2%) across 18 modules were associated with divergent DNA methylation (Fig. 3.2; table S2) (hypergeometric test (Wang et al., 2015), 1 million simulations, *P* value =

0.303). We found that the blue module contained the most DM hub genes (n = 26 genes; Fig. 3.2), and this overlap was more than expected by chance (hypergeometric test (Wang et al., 2015), 1 million simulations, *P* value = 0.031).

Additionally, we investigated whether genes identified as key regulators (hub genes) and/or genes under divergent methylation within co-expressed modules were genes known to be involved in cichlid fish LPJ adaptation and plasticity networks. We used a precompiled list of cichlid morphology genes described in Supplementary Table 6 of Carruthers et al. (2022). We identified eight of these previously identified cichlid morphology genes that were also identified as either hub genes, DM genes or both (Table S3), these were associated with four modules (blue, darkorange2, darkseagreen3 and lightcyan; Fig. 3.2).

DMR and hub gene analysis





Gene symbol	Module	DM	Hub
rxra	darkorange2	yes	yes
runx2	blue	no	yes
col6a1	blue	yes	no
fndc1	darkorange2	yes	no
lef1	blue	yes	no
ptch1	darkorange2	yes	no
rarg	lightcyan	yes	no
slmap	darkseagreen3	yes	no

Figure 3.2: Left panel: Bar plots number of DM genes and number of DM hub genes within each significant module, each bar is labelled with the number of genes in each module. Top right panel: Venn diagram showing overlap between hub genes, DM genes and known LPJ plasticity and adaptation genes (taken from Supplementary table 6 of Carruthers et al. (2022)). Bottom right panel: Table highlighting hub genes and DM genes that overlapped with established genes involved in LPJ plasticity.

Gene ontology analysis of candidate modules

Table 3.2: Gene ontology biological process terms that were statistically overrepresented within the darkorange2 (393 genes), lightcyan (755 genes), darkseagreen3 (3281 genes) and blue (1796 genes) modules. All *P*-values were Bonferroni corrected (FDR < 0.05).

Modules	GO term (GO ID)	Fold enrichment	*P*-value	FDR value
Darkorange2 (Significant for ecomorph)	nervous system development (GO:0007399)	2.17	7.03E-06	9.30E-03
	cellular macromolecule metabolic process (GO:0044260)	0.48	4.14E-05	3.66E-02
	nucleobase-containing compound metabolic process (GO:0006139)	0.29	4.46E-05	3.54E-02
	gene expression (GO:0010467)	0.21	4.59E-05	3.32E-02
Lightcyan (Significant for ecomorph)	cell-cell junction organization (GO:0045216)	5.13	1.56E-05	1.38E-02
	regulation of cellular process (GO:0050794)	1.25	4.75E-05	3.77E-02
	RNA processing (GO:0006396)	0.12	1.37E-05	1.36E-02
Darkseagreen3 (Significant for ecomorph, upper/lower oral jaw and head length)	intracellular signal transduction (GO:0035556)	1.48	2.81E-05	1.31E-02
	protein-containing complex assembly (GO:0065003)	0.47	6.03E-05	2.66E-02
	ncRNA processing (GO:0034470)	0.35	1.29E-04	4.88E-02
	ribonucleoprotein complex biogenesis (GO:0022613)	0.34	6.95E-05	2.91E-02
	translation (GO:0006412)	0.18	8.24E-08	2.18E-04
	tRNA metabolic process (GO:0006399)	0.17	1.30E-04	4.68E-02
Blue (Significant for upper/lower oral jaw and head length)	collagen fibril organization (GO:0030199)	6.78	9.06E-05	6.54E-03
	SMAD protein signal transduction (GO:0060395)	5.75	4.04E-06	5.53E-04
	pharyngeal system development (GO:0060037)	5.56	5.87E-04	2.83E-02
	positive regulation of pathway-restricted SMAD protein phosphorylation (GO:0010862)	5.52	5.80E-05	4.61E-03
	collagen metabolic process (GO:0032963)	5.32	1.60E-04	9.75E-03
	regulation of pathway-restricted SMAD protein phosphorylation (GO:0060393)	5.23	8.35E-05	6.19E-03
	collagen catabolic process (GO:0030574)	5.17	8.43E-04	3.81E-02
	negative regulation of BMP signalling pathway (GO:0030514)	4.52	2.23E-04	1.22E-02
	mucopolysaccharide metabolic process (GO:1903510)	4.32	3.01E-04	1.59E-02
	embryonic camera-type eye formation (GO:0060900)	4.3	5.79E-04	2.81E-02
	negative regulation of axon extension (GO:0030517)	3.93	1.01E-03	4.29E-02

	BMP signalling pathway (GO:0030509)	3.92	1.87E-04	1.09E-02
	notochord development (GO:0030903)	3.57	1.40E-04	9.00E-03
	vasculogenesis (GO:0001570)	3.55	1.11E-03	4.58E-02
	proteoglycan metabolic process (GO:0006029)	3.48	1.77E-04	1.06E-02
	cell-cell junction assembly (GO:0007043)	3.33	4.11E-04	2.07E-02
	negative regulation of Wnt signalling pathway (GO:0030178)	2.89	5.92E-04	2.82E-02
	chemical synaptic transmission (GO:0007268)	2.7	4.21E-05	3.64E-03
	axon guidance (GO:0007411)	2.56	7.80E-06	9.39E-04
	ameboidal-type cell migration (GO:0001667)	2.19	3.29E-04	1.73E-02
	cell-cell adhesion (GO:0098609)	2.09	8.37E-04	3.87E-02
	angiogenesis (GO:0001525)	2	8.53E-04	3.83E-02
	skeletal system development (GO:0001501)	1.97	1.95E-04	1.11E-02
	morphogenesis of an epithelium (GO:0002009)	1.81	3.81E-04	1.95E-02
	protein modification by small protein conjugation or removal (GO:0070647)	0.5	9.18E-04	4.05E-02
	intracellular protein transport (GO:0006886)	0.33	5.23E-05	4.20E-03
	chromosome organization (GO:0051276)	0.3	9.20E-05	6.58E-03
	ubiquitin-dependent protein catabolic process (GO:0006511)	0.27	1.33E-05	1.51E-03
	DNA repair (GO:0006281)	0.26	5.24E-04	2.59E-02
	mRNA processing (GO:0006397)	0.22	2.25E-04	1.22E-02
	negative regulation of gene expression (GO:0010629)	0.14	2.84E-04	1.51E-02
	mitochondrion organization (GO:0007005)	0.12	4.32E-05	3.65E-03
	translation (GO:0006412)	0.05	4.37E-08	1.09E-05
	RNA modification (GO:0009451)	< 0.01	1.04E-03	4.37E-02
	rRNA processing (GO:0006364)	< 0.01	8.71E-05	6.35E-03
	tRNA processing (GO:0008033)	< 0.01	1.03E-03	4.33E-02

Functional enrichment analyses identified several ecologically relevant GO terms associated with our four key modules (blue, darkorange2, darkseagreen3 and lightcyan; Table 3.2). The blue module, which was significantly correlated with upper and lower oral jaw length, and head length (Fig. 3.1), was enriched for 36 GO terms, including several terms that have known roles in shaping craniofacial phenotypes, such as skeletal development, pharyngeal system development, bone morphogenetic protein (BMP) and Wnt signalling pathways, and SMAD protein regulation (Table 3.2; Gou et al., 2015). We also found enrichment of terms related to neuron and blood vessel development. Given the blue module is significantly associated with three adaptive craniofacial traits, we suggest that this group of co-expressed genes are important modulators of craniofacial morphology and jaw phenotypes in our focal species.

The darkorange2 module, which was significantly correlated with ecomorph, was enriched for four GO terms, including macromolecule metabolism and nervous system development (Table 3.2). Darkseagreen3, which is significantly correlated with ecomorph, head length and upper and lower oral jaw length, was enriched for six terms, primarily associated with RNA metabolism and non-coding RNA processing (Table 3.2). Finally, the lightcyan module, which is significantly correlated with head length, showed statistical overrepresentation of three GO terms related to cellular process regulation and cell-cell junction organisation (Table 3.2).

Gene co-expression trait relationships in key modules

Finally, we examined the directional relationship between our four candidate modules (blue, orangered2, darkseagreen3 and lightcyan) and our measured traits (ecomorph, upper oral jaw length, lower oral jaw length and head length; head depth was excluded as it showed no significant association with any of the defined co-expressed modules). In the lightcyan module we can see some evidence of a positive relationship in our craniofacial measurements however only head length showed a significant positive association. This indicates that an up-regulation in lightcyan genes is associated with a longer head length. In the blue module we can see a significant negative relationship in all craniofacial measurements, which indicates that a down-regulation of blue module genes is associated with shorter upper/lower oral jaws, and head length. We can also see some evidence of more variation in eigengene scores in our shallow ecomorphs compared to deep ecomorphs

in our boxplot however the blue module was not significantly associated to ecomorph. In our darkorange2 module we can see a significant association to ecomorph, with an up-regulation of darkorange2 module genes in shallow ecomorph individuals. In the darkseagreen3 module we can see significant associations to ecomorph and all craniofacial measurements. This shows an up regulation in darkseagreen3 module genes in deep ecomorph and longer upper/lower oral jaws and head length. Similar to the blue module we see more variation in shallow ecomorph compared to deep, but we can see a clear significant positive relationship in all our craniofacial measurements.

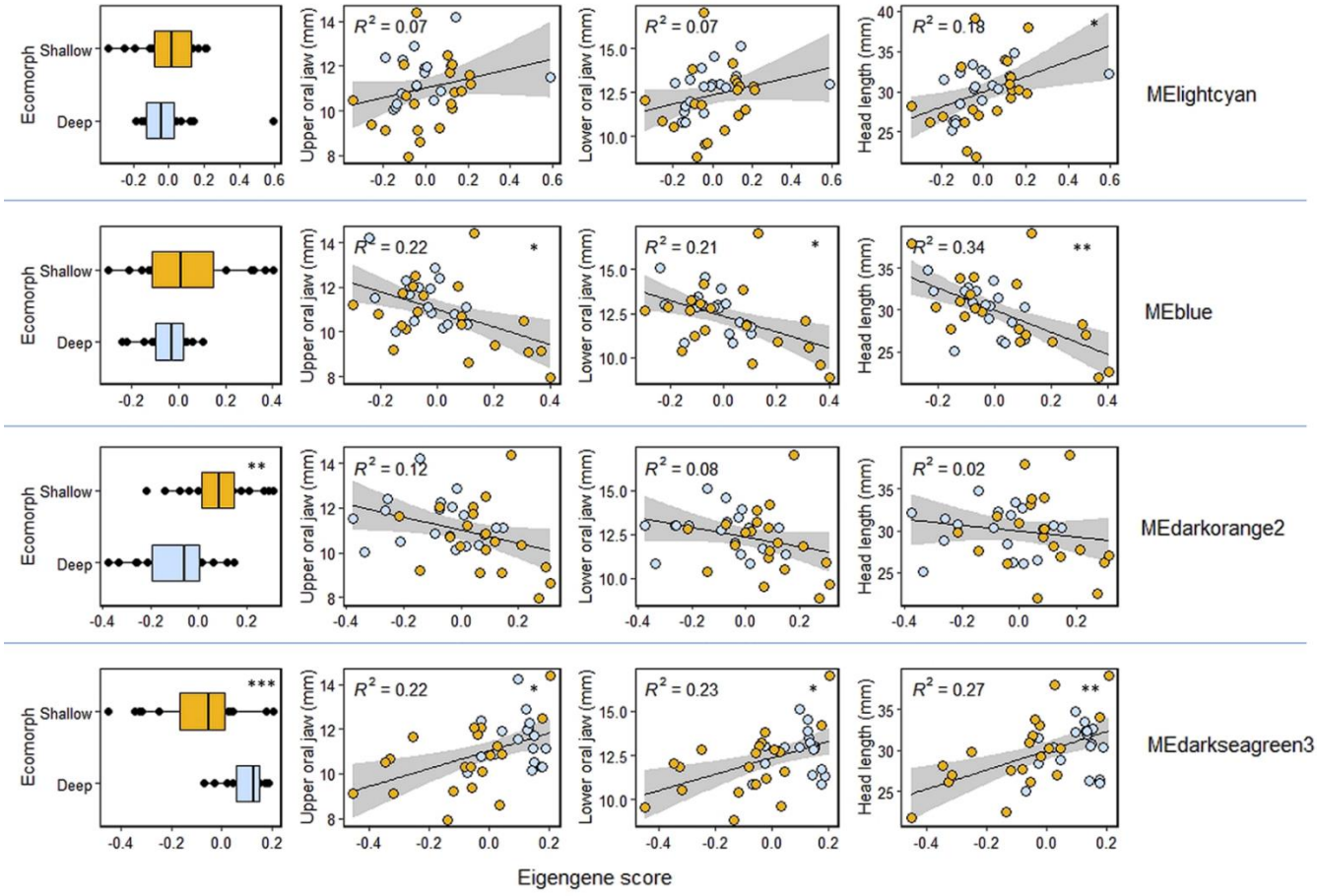


Figure 3.3: Module eigengene scores plotted against ecomorph, upper and lower oral jaw and head length traits, for our four candidate modules: lightcyan, blue, darkorange2 and darkseagreen3. Box plots for ecomorph traits represent upper and lower quartiles, the black bar within the boxes represent the median, horizontal lines represent 1.5 times the interquartile ranges respectively, and dots represent outliers. Scatter plots for continuous traits (upper and oral jaw length and head length), yellow points: shallow ecomorph, blue points: deep ecomorph. R^2 values in the top left corner of each plot indicate the proportion of variance accounted for by the dependent variable, represented by our linear regression lines. Asterisks denote the significance level of module-trait relationships in each plot, blank (non-significant), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$).

3.5 Discussion

In this chapter, I investigated gene co-expression in the diverging population of *A. calliptera* from Lake Masoko. Gene network analyses offer a powerful systematic approach to identify module transcriptional shifts underpinning functional phenotypes. Using a co-expression network approach (WGCNA), we identified numerous groups of co-expressed genes (modules) which were significantly associated to ecomorph and/or craniofacial measurements. By performing functional enrichment analysis on each module, we found overrepresentation of ecologically relevant gene ontology (GO) terms within these modules, which contributed to assigning functions to a given module. To assess the role of DNA methylation in the regulation of key genes, we identified the key regulators (hub genes) within each module, and compared these with genes that were differentially methylated (DM) between ecomorphs. We found these genes were no more likely to be differentially methylated than observed across the genome. However, we did identify a high number of DM hub genes within the blue module. To identify genes involved in jaw apparatus we compared our DM genes and hub genes with genes previously identified as involved in LPJ plasticity, which again highlighted the blue module as involved in the structure of jaw apparatus.

Co-expression network modules linked to traits

Using WGCNA we identified 22 modules significantly associated with trait data, indicating significant ecologically relevant co-expression within a large percentage of the transcriptome (80.1%). 16 modules were significantly associated with ecomorph, this confirms that there is substantial transcriptomic divergence between deep/benthic and shallow/littoral ecomorphs (Carruthers et al., 2022). We also found eight modules significantly associated with head length and three with upper/lower oral jaw. These results indicated that transcriptional differences are playing a role in diversification of craniofacial morphology in our focal species. Functional enrichment analysis of trait-significant modules showed that clustering based on co-expression tended to yield an overrepresentation of GO terms with similar functions. This allowed us to assign biological functions to modules (Table S1). Examples include immunity related terms in the black module, respiratory related terms

in the brown module, nervous system development in the darkorange2 module, and jaw apparatus related terms in the blue module.

DNA methylation in significant trait-associated modules

Differentially methylated genes appeared to be evenly distributed between expressed and non-expressed genes (Table 3.1). However, there was greater overlap between DM genes and significant trait-associated modules than expected. This may indicate that DNA methylation takes a role in driving diversification between ecomorphs and contributes to diversification in craniofacial morphology. The results support recent work that found evidence for epigenetic variants contributing to speciation and diversification in East African cichlids (Vernaz, et al., 2021 & 2022). However, the even distribution of differentially methylated genes (11.5%-13.9%) across the entire genome, regardless of whether the gene is expressed or not (Table 3.1), suggests that divergence in DNA methylation is ubiquitous.

DNA methylation in key regulatory 'hub' genes

We found that 14.2% of key regulatory 'hub' genes were differentially methylated between ecomorphs, but hypergeometric tests reveal this proportion to be no greater than expected given the extent of genome-wide methylation. This indicates that divergent DNA methylation patterns are not more likely to underpin key regulatory genes. However, the DM hub genes identified are still good candidates of key regulatory genes under regulation by epigenetic mechanisms within our focal species, and are a potentially useful targets for future study.

Jaw apparatus modules

Our study has identified gene clusters associated with ecologically-relevant morphological changes using transcriptomes from the ongoing speciation event in Lake Masoko (Table S1). Following on *Chapter 2* we chose to focus on the jaw apparatus diversification and its potential role in ecological speciation, utilizing the phenotypic data on craniofacial morphology (Fig. 2.3). The blue module was highlighted in multiple steps of our analysis; significant associations to craniofacial measurements (Fig. 3.2), strong overlap with DM genes (Fig. 3.2), numerous DM hub genes identified (Fig. 3.2), statistical overrepresentation of ecologically relevant GO terms (Table 3.2) and was found to contain genes known to be

involved in LPJ plasticity (Fig. 3.2). This strongly identifies it as a jaw apparatus module, and further confirms the involvement of this cluster of genes in jaw plasticity.

The blue module was one of just three (out of 31) modules significantly associated with three craniofacial traits (upper and lower jaw length and head length) (Fig. 3.1), highlighting the importance of this module in shaping divergent craniofacial morphology. These results support at least some transcriptional and functional coupling of pharyngeal and oral jaw apparatus evolution (Conith & Albertson, 2021) given that our RNA-seq data was generated from lower pharyngeal jaw tissue. Further investigation into the transcriptomes of oral and pharyngeal jaw apparatus of *A. calliptera* from Lake Masoko are likely to provide novel insights into the modularity of jaw evolution during the initial stages of ecological diversification and speciation.

Functional enrichment analysis of the blue module revealed 36 GO terms that were overrepresented. This included terms related to SMAD protein signal transduction. SMAD proteins are major transcription factors and have been suggested as good candidate genes in the regulation of variation in both LPJ and oral jaw apparatus in cichlids (Conith & Albertson, 2021). We also found terms related to regulation of BMP and Wnt signalling pathways, implicated in studies of cichlid craniofacial development (Parsons & Albertson, 2009; Parsons et al., 2014). Both pathways have been identified to be partially mediated by the transcriptional suppression/activation properties of SMAD proteins (Collery & Link, 2011; Gou et al., 2015; Ahi, 2016). Higher module eigengene scores (defined as the first principal component of the module expression profile) were associated with shorter lower/upper oral jaws and head lengths. This indicates that shorter craniofacial measurements (more common in the littoral/shallow ecomorph) are potentially linked to negative regulation of BMP and Wnt pathways. The exact role that BMP signalling pathway plays in craniofacial and pharyngeal jaw morphology is still unclear (Ahi, 2016). However there is some evidence of that downregulation of genes involved in the BMP signalling pathway results in shorter/more rounded skull profiles (Parsons & Albertson, 2009) and pharyngeal jaw phenotypes characteristic of soft diets (Gunter et al., 2013).

Overlap between hub genes and DM genes with LPJ master adaptation genes

The blue module was found to have a higher number of differentially methylated (DM) hub genes (18.8%, Fig. 3.2), despite not being the largest module (Fig. S5). This provides some evidence for increased regulation via DNA methylation of key regulatory genes, involved in craniofacial and jaw morphology. As these phenotypes are known to be under strong divergent selection and highly plastic in our focal species in Lake Masoko (Malinsky et al., 2015), this supports findings, by Vernaz, et al. (2022), of epigenetic contributions to speciation and environmentally induced changes in gene expression.

Of specific interest to LPJ plasticity, we identified two genes *runx2* and *lef1* within the blue module that are known master adaptation genes in cichlids (Fraser et al., 2013; Parsons et al., 2014; Schneider et al., 2014; Singh et al., 2017) that are involved in BMP and Wnt signalling pathways (Gou et al., 2015; Ahi, 2016) and are specifically associated with craniofacial development in teleosts (Ahi, 2016), including cichlids (Schneider et al., 2014). Additionally, the gene *col6a1,* also within the blue module (Fig. 3.2), was found to show increased expression in *Astatoreochromis alluaudi* LPJ tissues when reared on a hard diet treatment (Gunter et al., 2013; Schneider et al., 2014). All three of these genes were also differentially expressed between ecomorphs (Supplementary Table 6, Carruthers et al. (2022)). Taken together, our results highlight co-expressed genes in the blue module as potential major regulators of craniofacial morphology and oral jaw apparatus.

Our analysis also identified the darkorange2 module as a potential candidate in jaw and craniofacial morphology as it contained three DM genes known to be involved in LPJ plasticity networks (list genes; Fig. 3.2). This module was also significantly associated with ecomorph (Fig. 3.1 and 3.3). This potentially indicates that genes in the darkorange2 module are significant drivers in ecomorph divergence and speciation, with a basis in LPJ plasticity (Fig. 3.2) (Supplementary table 6, Carruthers et al. (2022)). Gene ontology analysis did not find any terms related to jaw apparatus but did find statistical overrepresentation of nervous system development. This, combined with apparent variation among ecomorphs, could indicate divergence based on habitat complexity (Huber et al., 1997).

Other ecological factors

The main focus of the present study was to identify co-expressed gene networks underpinning variation in jaw and craniofacial phenotypes within the diverging *A. calliptera* in Lake Masoko. However, our network analysis further identified several modules of co-expressed genes associated with ecomorph divergence, that were associated with a broad range of ecologically relevant functions (Table S1). For example, the black module, which was statistically overrepresented for GO terms related to immunity. This suggests there may be differences in immune response between deep and shallow ecomorphs, similar to patterns found in other studies (Rajkov et al., 2021; Carruthers et al., 2022; Vernaz, et al., 2022). A further example includes the brown module which had a number of statistically overrepresented GO terms related to mitochondria and ATP synthesis, this could be due to differences in water chemistry (dissolved oxygen and pH) between shallow and deep-water environments (Delalande 2008; Malinsky et al., 2015). This highlights the power of using WGCNA to uncover molecular mechanisms involved in divergent adaptive responses to a broad range of ecological factors, that would be otherwise difficult to measure. These co-expressed modules have been extracted from LPJ tissue which plays little role in these ecological functions, but WGCNA investigations into other tissues such as muscle, liver and gills may provide better insight into immunity and respiratory related phenotypes. Both the black and brown modules contained a higher number of DM hub genes (14 and 18 respectively, Fig. 3.2), and these represent candidates for further investigations into immune response and adaptions to hypoxic conditions.

Conclusions

Our results show that widespread regulatory shifts in gene expression are prevalent in speciating *A. calliptera* in Lake Masoko, and likely play an essential role in ecological speciation. Using a transcriptome-wide gene expression network analysis we identified key modules of co-expressed genes involved in shaping divergent jaw and craniofacial morphologies between the deep and shallow ecomorphs. This approach allowed us to specifically target gene modules underpinning adaptive shifts in jaw apparatus during the initial stages of ecological diversification, but we also captured gene modules underpinning other ecological functions which open up numerous opportunities for further research. Our results also investigated DNA methylation as a potential regulator of functionally adaptive

phenotypes. Altogether, our results show that differential DNA methylation was evenly distributed across expressed and non-expressed genes. We did find some evidence that differential methylation of key regulatory genes was more prevalent in our jaw apparatus (blue) module, indicating increased epigenetic regulation of phenotypes. Overall, this study reveals how gene co-expression network approaches can provide indications of the functional roles of ecologically relevant genes, and explore the epigenetic regulation of these genes.

Chapter 4: General discussion

Overview

In the research presented in this thesis, I investigated phenotypic plasticity in craniofacial traits associated with divergent trophic adaptation within a diversifying population of African cichlids, and investigate their molecular basis. I take advantage of the ongoing sympatric speciation of *Astatotilapia calliptera* in Lake Masoko to explore the role of phenotypic plasticity and gene expression divergence during the early stages of population diversification. The research has focussed on pharyngeal and oral jaw apparatus, which are thought to be one of the key innovations contributing to the vast diversity of East African cichlids. We used a gene co-expression network approach to link transcriptome shifts to phenotypic traits, to further understand the molecular basis of phenotypic diversity, and explore the evolutionary underpinnings of these phenotypes.

In Chapter 2, I report an investigation of phenotypic divergence and plasticity in jaw apparatus using wild and common garden reared individuals to explore to what extent phenotypes are heritable and/or induced by differing diets, i.e. phenotypically plastic. The results show that the lower pharyngeal jaw (LPJ) structure was phenotypically plastic and dependent on ecomorph lineages, with diet being a significant driver of this plastic response. These results suggest that pharyngeal jaw apparatus is undergoing specialisation based on differing diets between deep and shallow environments, but as expected in an ongoing speciation event, an aspect of phenotypic plasticity is still retained. Additionally, oral jaws did not show any significant plastic response to diet, but did show dependence on parental ecomorph lineages following phenotypic divergence patterns in the wild population. This may suggest that the phenotype has already undergone specialisation and is genetically fixed, or that diet is not the selection pressure driving oral jaw diversification. The resulting patterns of pharyngeal and oral jaw morphology suggests that they have a significant heritable genetic component (Malinsky et al., 2015). These findings demonstrate that oral jaw diversification may be driven by ecological factors in addition to diet, while LPJ diversification is more closely linked to diet (Muschick et al., 2011; Gunter et al., 2013). This partially supports the theory of evolutionary decoupling between the oral jaw and pharyngeal jaw apparatus (Liem, 1973), however further investigation into the phenotypic

diversification of oral jaws in *A. calliptera* would be required to elaborate on this. The prevalence of phenotypic plasticity and phenotypic variation in the jaw apparatus of *A. calliptera* appears to be a key feature which has allowed them to colonize niches in Lake Masoko so successfully, and may explain how this species has been able to more broadly colonise rivers and lakes across the Lake Malawi catchment, as well as neighbouring eastward-flowing river systems (Joyce et al., 2011).

In Chapter 3, I report research exploring the role of gene expression and DNA methylation in the phenotypic divergence of craniofacial morphology in our focal species. Using a co-expression network approach, we identified several groups of co-expressed genes (modules) linked to jaw apparatus, and other ecological functions. We found numerous modules significantly associated with ecomorphs and craniofacial measurements, showing that shifts in gene expression have a strong basis in functional adaptive phenotypes (jaw apparatus) and are therefore likely contributing to ongoing ecological speciation. We found significant overrepresentation of GO terms within significant trait-associated modules. For example, the blue module was highly overrepresented for numerous terms relating to jaw apparatus. In conjunction with our co-expression network, we explored differential methylation between ecomorphs in our population. Overall, we found that differential methylation was evenly distributed across the genome, but there was evidence of increased epigenetic regulation in significant trait-associated modules, and in key regulatory genes of the blue (jaw apparatus) module, but not of key regulator genes overall. This demonstrates the utility of co-expression networks in exploring the molecular basis of adaptive functional phenotypes which may be crucial in driving ecological speciation.

Overall, this work shows that phenotypic plasticity, gene expression and epigenetics likely all play a role in ecological speciation and consequently the vast adaptive radiation of East African cichlids.

Limitations

In the research presented in Chapter 2 we had very few males in these experiments, meaning our statistical tests for sex had relatively low power. In addition, we investigated oral jaw length, however a more detailed analysis of the dentition and shape of oral jaws

would yield greater insights into the differences in this apparatus and the influence of diet as a selection pressure, as seen in Conith & Albertson, (2021).

In chapter 3 all transcriptomes were taken from LPJ tissue which contained a mixture of different tissues, including bone, cartilage, muscle and blood. This may make it difficult to distinguish LPJ specific gene expression patterns, but can allow genes associated with other ecological functions to be examined. Singh et al. (2021) investigated the modularity of oral and pharyngeal jaw apparatus and raised this point in their methodology. They stated that "including bone and soft tissue in the dissection does not violate any biological assumptions about modularity as we know that there is cross-talk between the bone, cartilage, muscle pathways of the trophic apparatus". Therefore, any trends in co-expression detected in our network can be linked to LPJ specific patterns.

Future directions

Research presented in Chapter 2 identified two interesting avenues of future research. First, more detailed investigation into the divergence in oral jaw apparatus (dentition and shape) would shed more light on the relative contributions of diet and other ecological factors as drivers of diversity. As the *Astatotilapia* of Lake Masoko are part of an ongoing speciation event, and diet does not appear to significantly affect oral jaw length, research into functional drivers of oral jaw structures would help to inform us of the extent that oral and pharyngeal jaw apparatus are decoupled (Liem 1973). Secondly, we have shown considerable heritability in both oral and pharyngeal jaws. Investigating the genetic vs epigenetic heritability components of jaw apparatus using the next generations of our laboratory fish would be a significant step into understanding the molecular processes behind ecological speciation. Further research could allow us to explore theories on genetic assimilation of environmentally induced phenotypes (Gunter et al., 2017) and the molecular (epigenetic) mechanisms governing this process (Danchin et al., 2019).

Research presented in Chapter 3 presents numerous interesting directions for future research. Our initial investigation into the potential role of differential methylation in regulating functionally important gene expression modules revealed some interesting trends relating to the role of epigenetics in phenotypic divergence and ecological speciation. Further investigations into the methylomes of the *A. calliptera* ecomorphs using a WGCNA

approach to identify co-methylated modules underpinning adaptive traits may be incredibly insightful. In conjunction with gene expression networks such as ours, this could be used to look at module preservation between co-methylation and co-expression networks (as seen in, Morandin et al. (2019)), to elucidate key regulatory genes and modules that are potential targets of divergent natural selection. This would provide valuable and novel insights into the role of DNA methylation in driving phenotypic diversification and ecological speciation. Other avenues of future research involve investigations into highlighted candidate genes and their function in jaw apparatus development. Recent research has shown that craniofacial differences can be identified between divergent populations *A. calliptera* from the Lake Malawi catchment in the early stages of development (main lake (Salima) versus riverine (Mbaka river) populations; Marconi et al. (2022)). Recent work on cichlids using CRISPR-cas9 to target genes (Kratochwil et al., 2018; Clark et al., 2022) highlights emerging opportunities to investigate the molecular basis of craniofacial development in genes highlighted as potential candidates in shaping craniofacial phenotypes, such as *runx2, lef1* and *col6a1*.

Another interesting avenue for future research would be further investigation into head length, as a potential adaptation to low oxygen environments in the deep-water habitat in this study system. Research presented in Chapter 2 head length was found to be significantly longer in deep ecomorph individuals, and showed strong heritability. Functionally this may indicate an increase in gill space (Bouton et al. 2002) potentially due to hypoxia in the deep-water environment (Delalande 2008). In Chapter 3 we found 8 modules significantly associated with head length, some of which were enriched for several GO terms related to respiration. Gene co-expression networks generated from gill tissue from both ecomorphs, and collection of phenotypic data relating to gill rakers/size could provide valuable insights into adaptive response to low oxygen environments in the deep-benthic ecomorph.

Conclusions

To conclude, the research suggests that diet is a major factor associated with phenotypic plasticity and divergence in the lower pharyngeal jaw of *A. calliptera* in Lake Masoko, but in the oral jaw other ecological factors should be considered. Divergence in both jaw apparatus is heritable suggesting a substantial genetic basis in jaw diversity. Exploring the

epigenetic component of these phenotypes, and whether there is any heritability of epigenetic markers are interesting avenues for future research. Transcriptional shifts between the *A. calliptera* ecomorphs in Lake Masoko are widespread and underpin a diverse range of ecologically relevant functions. We identified co-expressed gene modules from LPJ tissues which had significant associations to craniofacial morphology. Our results further suggest that DNA methylation may play a key regulatory role within jaw gene modules and warrants further investigation. Taken together our results suggest that transcriptome diversity has a strong basis in adaptive functional phenotypes, allowing cichlids to thrive and adapt to novel environments. This is a key characteristic of East African cichlids and may have facilitated their ecological speciation and broad-scale adaptive radiation across the East African region.

References

Ahi, E. P. (2016). Signalling pathways in trophic skeletal development and morphogenesis: Insights from studies on *teleost* fish. Developmental Biology, 420(1), 11–31. https://doi.org/10.1016/j.ydbio.2016.10.003

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2), 162-169. https://doi.org/10.1093/bioinformatics/btu638

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bouton, N., de Visser, J., & Barel, C. D. N. (2002). Correlating head shape with ecological variables in rock-dwelling haplochromines (Teleostei: *Cichlidae*) from Lake Victoria. Biological Journal of the Linnean Society, 76(1), 39-48. https://doi.org/10.1046/j.1095-8312.2002.00046.x

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., … di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature, 513(7518), 375–381. https://doi.org/10.1038/nature13726

Burress, E. D., Duarte, A., Serra, W. S., & Loureiro, M. (2016). Rates of piscivory predict pharyngeal jaw morphology in a piscivorous lineage of cichlid fishes. Ecology of Freshwater Fish, 25(4), 590–598. https://doi.org/10.1111/eff.12236

Bush, S. J., Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2017). Alternative splicing and the evolution of phenotypic novelty. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1713), 20150474. Royal Society of London. https://doi.org/10.1098/rstb.2015.0474

Carruthers, M., Edgley, D. E., Saxon, A. D., Gabagambi, N. P., Shechonge, A., Miska, E. A., Durbin, R., Bridle, J. R., Turner, G. F., & Genner, M. J. (2022). Ecological speciation promoted by divergent regulation of functional genes within African cichlid fishes. BioRxiv. https://doi.org/10.1101/2022.01.07.475335

Chevin, L. M., Leung, C., le Rouzic, A., & Uller, T. (2021). Using phenotypic plasticity to understand the structure and evolution of the genotype–phenotype map. Genetica, 150, 209–221. https://doi.org/10.1007/s10709-021-00135-5

Clark, B., Elkin, J., Marconi, A., Turner, G. F., Smith, A. M., Miska, E. A., Juntti, S. A., & Emília Santos, M. (2022). oca2 targeting using CRISPR/Cas9 in the Malawi cichlid *Astatotilapia calliptera*. Royal Society Open Science, 9: 220077. https://doi.org/10.1098/rsos.220077

Collery, R. F., & Link, B. A. (2011). Dynamic Smad-Mediated BMP signaling revealed through transgenic zebrafish. Developmental Dynamics, 240, 712–722. https://doi.org/10.1002/dvdy.22567

Conith, A. J., & Albertson, R. C. (2021). The cichlid oral and pharyngeal jaws are evolutionarily and genetically coupled. Nature Communications, 12(1): 5477. https://doi.org/10.1038/s41467-021-25755-5

Conte, M. A., Joshi, R., Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Roberts, R. B., Carleton, K. L., Lien, S., & Kocher, T. D. (2019). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. GigaScience, 8(4): giz030. https://doi.org/10.1093/gigascience/giz030

Danchin, E., Pocheville, A., Rey, O., Pujol, B. and Blanchet, S. (2019), Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. Biological Reviews, 94, 259-282. https://doi.org/10.1111/brv.12453

Delalande, M. (2008). Hydrologie et géochimie isotopique du lac Masoko et de lacs volcaniques de la province active du Rungwe (Sud-Ouest Tanzanie). (Doctoral dissertation, Université Paris Sud-Paris XI).

de Meyer, J., Christiaens, J., & Adriaens, D. (2016). Diet-induced phenotypic plasticity in European eel (*Anguilla anguilla*). Journal of Experimental Biology, 219(3), 354–363. https://doi.org/10.1242/jeb.131714

Diamond, S. E., & Martin, R. A. (2021). Buying time: plasticity and population persistence. In Pfennig, D.W., (2021) Phenotypic plasticity & evolution, causes, consequences, controversies. 185-204. https://doi.org/10.1201/9780429343001-11

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Driscoll, R. M. H., Faber-Hammond, J. J., O'Rourke, C. F., Hurd, P. L., & Renn, S. C. P. (2020). Epigenetic regulation of gonadal and brain aromatase expression in a cichlid fish with environmental sex determination. General and Comparative Endocrinology, 296, 113538. https://doi.org/10.1016/j.ygcen.2020.113538

Duncan, E. J., Gluckman, P. D., & Dearden, P. K. (2014). Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? In Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 322(4), 208–220. https://doi.org/10.1002/jez.b.22571

Edgley, D. E., & Genner, M. J. (2019). Adaptive diversification of the lateral line system during cichlid fish radiation. IScience, 16, 1–11. https://doi.org/10.1016/j.isci.2019.05.016

El Taher, A., Böhne, A., Boileau, N., Ronco, F., Indermaur, A., Widmer, L., & Salzburger, W. (2021). Gene expression dynamics during rapid organismal diversification in African cichlid fishes. Nature Ecology and Evolution, 5(2), 243–250. https://doi.org/10.1038/s41559-020-01354-3

Fox, J., Weisberg, S., (2019). An R Companion to Applied Regression, Third edition. Sage.

Fraser, G. J., Bloomquist, R. F., & Streelman, J. T. (2013). Common developmental pathways link tooth shape to regeneration. Developmental Biology, 377(2), 399–414. https://doi.org/10.1016/j.ydbio.2013.02.007

Futuyma, D.J., (2021). How does phenotypic plasticity fit into evolutionary theory? In Pfennig, D.W., (2021) Phenotypic plasticity & evolution, causes, consequences, controversies. CRC press. 349-366. https://doi:10.1201/9780429343001-18

Genner, M. J., & Turner, G. F. (2005). The mbuna cichlids of Lake Malawi: A model for rapid speciation and adaptive radiation. Fish and Fisheries, 6(1), 482-488. https://doi.org/10.1111/j.1467-2679.2005.00173.x

Gou, Y., Zhang, T., & Xu, J. (2015). Transcription factors in craniofacial development. From receptor signaling to transcriptional and epigenetic regulation. Current Topics in Developmental Biology, 115, 377–410. https://doi.org/10.1016/bs.ctdb.2015.07.009

Gunter, H.M., Fan, S., Xiong, F., Franchini, P., Fruciano, C. & Meyer, A. (2013). Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. Molecular Ecology, 22(11), 4516–4531. https://doi.org/10.1111/mec.12417

Gunter, H. M., Schneider, R. F., Karner, I., Sturmbauer, C., & Meyer, A. (2017). Molecular investigation of genetic assimilation during the rapid adaptive radiations of East African cichlid fishes. Molecular Ecology, 26(23), 6634–6653. https://doi.org/10.1111/mec.14405

Hahn, C., Genner, M. J., Turner, G. F., & Joyce, D. A. (2017). The genomic basis of cichlid fish adaptation within the deepwater “twilight zone” of Lake Malawi. Evolution Letters, 1(4), 184–198. https://doi.org/10.1002/evl3.20

Hofmann, C. M., O’Quin, K. E., Justin Marshall, N., Cronin, T. W., Seehausen, O., & Carleton, K. L. (2009). The eyes have it: Regulatory and structural changes both underlie cichlid visual pigment diversity. PLoS Biology, 7(12): 1000266. https://doi.org/10.1371/journal.pbio.1000266

Hu, J., Wuitchik, S. J. S., Barry, T. N., Jamniczky, H. A., Rogers, S. M., & Barrett, R. D. H. (2021). Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). Genetics, 217(1): iyab001. https://doi.org/10.1093/GENETICS/IYAB001

Huber, R., van staaden, M. J., Kaufman, L. S., & Liem, K. F. (1997). Microhabitat use, trophic patterns, and the evolution of brain structure in African cichlids. Brain, Behavior and Evolution, 50(3), 167–182. https://doi.org/10.1159/000113330

Hull, P. (2015). Life in the aftermath of mass extinctions. Current Biology, 25(9), 941-952. Cell Press. https://doi.org/10.1016/j.cub.2015.08.053

Hulsey, C. D., García, F. J., León, D. E., León, L., Rodiles-Hernández, A. R., & Hernández, H. (2006). Micro and macroevolutionary decoupling of cichlid jaws: a test of Liem's key innovation hypothesis. Evolution, 60(10), 2096–2109. https://doi.org/10.1111/j.0014-3820.2006.tb01847.x

Ivory, S. J., Blome, M. W., King, J. W., McGlue, M. M., Cole, J. E., & Cohen, A. S. (2016). Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. Proceedings of the National Academy of Sciences of the United States of America, 113(42), 11895-11900. https://doi.org/10.1073/pnas.1611028113

Jablonka, E. V. A., & Raz, G. A. L. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology, 84(2), 131-176. https://doi.org/10.1086/598822

Joyce, D. A., Lunt, D. H., Genner, M. J., Turner, G. F., Bills, R., & Seehausen, O. (2011). Repeated colonization and hybridization in Lake Malawi cichlids. Current Biology, 21(3), 526. https://doi.org/10.1016/j.cub.2010.11.029

Kocher, T. D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. Nature Reviews Genetics, 5(4), 288–298. https://doi.org/10.1038/nrg1316

Kratochwil, C. F., Liang, Y., Gerwin, J., Woltering, J. M., Urban, S., Henning, F., Machado-Schiaffino, G., Hulsey, C. D., & Meyer, A. (2018). Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. Science, 362(6413), 457–460. https://doi.org/10.1126/science.aao6809

Kratochwil, C. F., & Meyer, A. (2015). Closing the genotype-phenotype gap: Emerging technologies for evolutionary genetics in ecological model vertebrate systems. BioEssays, 37(2), 213–226. https://doi.org/10.1002/bies.201400142

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics, 27(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Lafuente, E., & Beldade, P. (2019). Genomics of developmental plasticity in animals. Frontiers in Genetics, 10, 720. https://doi.org/10.3389/fgene.2019.00720

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics, 9: 559. https://doi.org/10.1186/1471-2105-9-559

Langfelder, P., & Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. Journal of Statistical Software, 46: 11. https://doi.org/10.18637/jss.v046.i11

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics, 24(5), 719–720. https://doi.org/10.1093/bioinformatics/btm563

Ledón-Rettig, C. C., & Pfennig, D. W. (2011). Emerging model systems in eco-evo-devo: The environmentally responsive spadefoot toad. Evolution and Development, 13(4), 391–400. https://doi.org/10.1111/j.1525-142X.2011.00494.x

Levis, N. A., & Pfennig, D. W. (2016). Evaluating “Plasticity-First” evolution in nature: Key criteria and empirical approaches. Trends in Ecology and Evolution, 31(7), 563-574. https://doi.org/10.1016/j.tree.2016.03.012

Liem, K. F. (1973). Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. Systematic Zoology, 22(4), 425–441. https://doi.org/10.2307/2412950

Liem, K. F., & Greenwood, P. H. (1981). A functional approach to the phylogeny of the pharyngognath teleosts. American Zoologist, 21(1), 83–101. https://doi.org/10.1093/icb/21.1.83

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12): 550. https://doi.org/10.1186/s13059-014-0550-8

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate

cichlid ecomorphs in an East African crater lake. Science, 350(6267), 1493-1498. http://doi.org/10.1126/science.aac9927

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nature Ecology and Evolution, 2(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Miska, E. A., & Ferguson-Smith, A. C. (2016). Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. Science, 354(6308), 59-63. https://doi.org/10.1126/science.aaf4945

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. In Neuropsychopharmacology, 38(1), 23–38. https://doi.org/10.1038/npp.2012.112

Morandin, C., Brendel, V. P., Sundström, L., Helanterä, H., & Mikheyev, A. S. (2019a). Changes in gene DNA methylation and expression networks accompany caste specialization and age-related physiological changes in a social insect. Molecular Ecology, 28(8), 1975–1993. https://doi.org/10.1111/mec.15062

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

Ng, I. C., Pawijit, P., Tan, J., & Yu, H. (2019). Anatomy and Physiology for Biomaterials Research and Development. Encyclopedia of Biomedical Engineering, 1–3, 225–236. https://doi.org/10.1016/B978-0-12-801238-3.99876-3

Nosil, P. (2012). Ecological Speciation (Oxford Series in Ecology and Evolution). Oxford University Press, USA. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001

Ortega-Recalde, O., & Hore, T. A. (2019). DNA methylation in the vertebrate germline: Balancing memory and erasure. Essays in Biochemistry, 63(6), 649–661. https://doi.org/10.1042/EBC20190038

Parsons, K. J., & Albertson, R. C. (2009). Roles for Bmp4 and CaM1 in shaping the jaw: Evo-devo and beyond. Annual Review of Genetics 43, 369–388. https://doi.org/10.1146/annurev-genet-102808-114917

Parsons, K. J., Trent Taylor, A., Powder, K. E., & Albertson, R. C. (2014). Wnt signalling underlies the evolution of new phenotypes and craniofacial variability in Lake Malawi cichlids. Nature Communications, 5:3629. https://doi.org/10.1038/ncomms4629

Parsons, P. J., Bridle, J. R., Rüber, L., & Genner, M. J. (2017). Evolutionary divergence in life history traits among populations of the Lake Malawi cichlid fish *Astatotilapia calliptera*. Ecology and Evolution, 7(20), 8488–8506. https://doi.org/10.1002/ece3.3311

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. In Annals of the New York Academy of Sciences, 1206, 110–129. https://doi.org/10.1111/j.1749-6632.2010.05765.x

Pfennig, D. W., & West-Eberhard, M.J. (2021). Phenotypic Plasticity & Evolution; Causes, Consequences, Controversies. Taylor & Francis. https://doi.org/10.1201/9780429343001

Pfennig, D. W., Rice, A. M., & Martin, R. A. (2006). Ecological opportunity and phenotypic plasticity interact to promote character displacement and species coexistence. Ecology, 87(3), 769-779. https://doi.org/10.1890/05-0787

Pfennig, D. W., & Servedio, M. R. (2013). The role of transgenerational epigenetic inheritance in diversification and speciation. Non-Genetic Inheritance, 1, 17-26. https://doi.org/10.2478/ngi-2013-0002

Pigliucci, M. (2001). Phenotypic plasticity: beyond nature and nurture. Johns Hopkins University Press.

Pinto, G., Mahler, D. L., Harmon, L. J., & Losos, J. B. (2008). Testing the island effect in adaptive radiation: Rates and patterns of morphological diversification in Caribbean and mainland Anolis lizards. Proceedings of the Royal Society B: Biological Sciences, 275(1652), 2749–2757. https://doi.org/10.1098/rspb.2008.0686

Rajkov, J., el Taher, A., Böhne, A., Salzburger, W., & Egger, B. (2021). Gene expression remodelling and immune response during adaptive divergence in an African cichlid fish. Molecular Ecology, 30(1), 274–296. https://doi.org/10.1111/mec.15709

Ribbink, A. J. (1990). Alternative life-history styles of some African cichlid fishes. Environmental Biology of Fishes, 28, 87–100. https://doi.org/10.1007/BF00751029

Rice, A. M., & Pfennig, D. W. (2010). Does character displacement initiate speciation? Evidence of reduced gene flow between populations experiencing divergent selection. Journal of Evolutionary Biology, 23(4), 854–865. https://doi.org/10.1111/j.1420-9101.2010.01955.x

Rohlf, F. J. (2015). The tps series of software. Hystrix, 26(1), 9-12. https://doi.org/10.4404/hystrix-26.1-11264

Ruell, E. W., Handelsman, C. A., Hawkins, C. L., Sofaer, H. R., Ghalambor, C. K., & Angeloni, L. (2013). Fear, food and sexual ornamentation: Plasticity of colour development in Trinidadian guppies. Proceedings of the Royal Society B: Biological Sciences, 280(1758): 201220193. https://doi.org/10.1098/rspb.2012.2019

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. Ecology Letters, 8(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. Nature Reviews Genetics, 19(11), 705–717. https://doi.org/10.1038/s41576-018-0043-9

Schluter, D. (2000). The Ecology of Adaptive Radiation. Oxford University Press.

Schluter, D. (2001). Ecology and the origin of species. Trends in Ecology & Evolution, 16(7), 372-380. https://doi.org/10.1016/S0169-5347(01)02198-X

Schneider, R. F., Li, Y., Meyer, A., & Gunter, H. M. (2014). Regulatory gene networks that shape the development of adaptive phenotypic plasticity in a cichlid fish. Molecular Ecology, 23(18), 4511–4526. https://doi.org/10.1111/mec.12851

Schneider, R. F., & Meyer, A. (2017). How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. Molecular Ecology, 26(1), 330–350. https://doi.org/10.1111/mec.13880

Schultz, R. J. (1993). Genetic regulation of temperature-mediated sex ratios in the livebearing fish *Poeciliopsis lucida*. Copeia, 4, 1148-1151. https://doi.org/10.2307/1447098

Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. Proceedings of the Royal Society B: Biological Sciences, 273(1597), 1987–1998. https://doi.org/10.1098/rspb.2006.3539

Singh, P., Ahi, E. P., & Sturmbauer, C. (2021). Gene coexpression networks reveal molecular interactions underlying cichlid jaw modularity. BMC Ecology and Evolution, 21(1):62. https://doi.org/10.1186/s12862-021-01787-9

Singh, P., Börger, C., More, H., & Sturmbauer, C. (2017). The role of alternative splicing and differential gene expression in cichlid adaptive radiation. Genome Biology and Evolution, 9(10), 2764–2781. https://doi.org/10.1093/gbe/evx204

Skelly, D. A., Ronald, J., & Akey, J. M. (2009). Inherited variation in gene expression. Annual Review of Genomics and Human Genetics, 10, 313-332. https://doi.org/10.1146/annurev-genom-082908-150121

Smith, G., & Ritchie, M. G. (2013). How might epigenetics contribute to ecological speciation? Current Zoology, 59(5), 686–696. https://doi.org/10.1093/czoolo/59.5.686

Stroud, J. T., & Losos, J. B. (2016). Ecological Opportunity and Adaptive Radiation. Annual Review of Ecology, Evolution, and Systematics, 47, 507–532. https://doi.org/10.1146/annurev-ecolsys-121415-032254

Svardal, H., Salzburger, W., & Malinsky, M. (2020). Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes Evolutionary radiations: periods of significant acceleration in the rate of species diversification in particular lineages. Annual Review of Animal Biosciences, 16(9), 55-79. https://doi.org/10.1146/annurev-animal-061220

ThermoFisher Scientific. (2018). Avizo 9 user´s guide. Thermo Scientific Avizo Software 9 User's Guide.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania. A., (2003). PANTHER: a library of protein families and

subfamilies indexed by function. Genome Research, 13, 2129-2141. https://doi:10.1101/gr.772403

Tkint, T., Verheyen, E., de Kegel, B., Helsen, P., & Adriaens, D. (2012). Dealing with food and eggs in mouthbrooding cichlids: Structural and functional trade-offs in fitness related traits. PLoS ONE, 7(2): e31117. https://doi.org/10.1371/journal.pone.0031117

Turner, G. F., Ngatunga, B. P., & Genner, M. J. (2019). The natural history of the satellite lakes of Lake Malawi. EcoEvoRxiv. https://doi.org/10.32942/osf.io/sehdq

Vernaz, G., Hudson, A. G., Santos, M. E., Fischer, B., Carruthers, M., Shechonge, A. H., Gabagambi, N. P., Tyers, A. M., Ngatunga, B. P., Malinsky, M., Durbin, R., Turner, G. F., Genner, M. J., & Miska, E. A. (2022). Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. http://doi.org/10.1038/s41559-022-01894-w

Vernaz, G., Malinsky, M., Svardal, H., Du, M., Tyers, A. M., Santos, M. E., Durbin, R., Genner, M. J., Turner, G. F., & Miska, E. A. (2021). Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes. Nature Communications, 12(1): 5870. https://doi.org/10.1038/s41467-021-26166-2

Wainwright, P. C., Smith, W. L., Price, S. A., Tang, K. L., Sparks, J. S., Ferry, L. A., Kuhn, K. L., Eytan, R. I., & Near, T. J. (2012). The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Systematic Biology, 61(6), 1001–1027. https://doi.org/10.1093/sysbio/sys060

Wang, M., Zhao, Y., & Zhang, B. (2015). Efficient test and visualization of multi-set intersections. Scientific Reports, 5: 16923. https://doi.org/10.1038/srep16923

West-Eberhard, M. J. (1989). Phenotypic plasticity and the origins of diversity. Annual Review of Ecology and Systematics. 20, 249-278. https://doi.org/10.1146/annurev.es.20.110189.001341

West-Eberhard, M. J. (2003). Developmental plasticity and evolution. In Developmental Plasticity and Evolution. https://doi.org/10.1093/oso/9780195122343.001.0001

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P., & Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data

without replicates. Nucleic Acids Research, 43(21): e141.
https://doi.org/10.1093/nar/gkv715

Appendix

Script 1. RScript for network construction using WGCNA

```
setwd("/home/em21733/Documents/Josh/WGCNA")

## Mostly follows code taken from
(https://wikis.utexas.edu/display/bioiteam/Clustering+using+WGCNA) unless stated otherwise

library(dynamicTreeCut)

library(fastcluster)

library(WGCNA)

library(tidyverse)

# Uploading data into R and formatting it for WGCNA

# This creates an object called "datExpr" that contains the normalized counts file output from
DESeq2

datExpr1 = read.table("Masoko_wild_log_expression_data.txt",header=T)

# removing SH02 # [c(1:20,22:39)]

datExpr<-datExpr1

# "head" the file to preview it

# Manipulate file so it matches the format WGCNA needs

row.names(datExpr) = datExpr$geneid

datExpr$geneid = NULL

datExpr = as.data.frame(t(datExpr)) # now samples are rows and genes are columns

dim(datExpr) # 38 samples and 19237 genes

# Run this to check if there are gene outliers

gsg = goodSamplesGenes(datExpr, verbose = 3)

gsg$allOK  ## Returns TRUE value

#####################################################################

#Create an object called "datTraits1" that contains your trait data

datTraits1 <- read.table("Masoko_wild_oral_jaw_phenotypes.txt",header=TRUE)

head(datTraits1)

#form a data frame analogous to expression data that will hold the clinical traits

rownames(datTraits1) = datTraits1$Sample_code

datTraits1$sample_code = NULL
```

```
table(rownames(datTraits1)==rownames(datExpr)) #should return TRUE if datasets align correctly, otherwise your names are out of order - TRUE is returned!

# removing ecomorph and sample code

datTraits<-datTraits1[c(2:6)]

datTraits<-datTraits %>%

  rename("Upper oral jaw" =Upper_Oral_Jaw) %>%

  rename("Lower oral jaw" =Lower_Oral_Jaw) %>%

  rename("Head depth" =Head_Depth) %>%

  rename("Head length" =Head_Length)

head(datTraits)

# You have finished uploading and formatting expression and trait data

# Expression data is in datExpr, corresponding traits are datTraits1

save(datExpr, datTraits, file="Masoko_wild_sample_WGCNA_input.RData")

load("Masoko_wild_sample_WGCNA_input_1.RData")

# Sample clustering to detect outliers

install.packages("flashClust")

library(flashClust)

WGCNATree = flashClust(dist(datExpr), method = "average")

sizeGrWindow(12,9)

par(cex = 0.6)

par(mar = c(5,5,5,5))

plot(WGCNATree, main = "Sample clustering to detect outliers", sub="", xlab="", cex.lab = 1.5,cex.axis = 1.5, cex.main = 2)

head(datExpr)

## At this point you will need to identify sample outliers and choose a soft threshold power.

## Code for this part of analysis was modified from (https://bioconductor.org/packages/devel/bioc/vignettes/CVE/inst/doc/WGCNA_from_TCGA_RNAseq.html)

## Power threshold for all popuations and ecotypes combined dataset is 3

powers = c(c(1:10), seq(from = 12, to=24, by=2))

options(stringsAsFactors = FALSE)

sft = pickSoftThreshold(datExpr, powerVector = powers, verbose = 3)
```

```
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
    xlab='Soft Threshold (power)',ylab='Scale Free Topology Model Fit,signed R^2',
    type='n', main = paste('Scale independence'));
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
    labels=powers,cex=1,col='red'); abline(h=0.90,col='red')
# Mean connectivity as a function of the soft-thresholding power
plot(sft$fitIndices[,1], sft$fitIndices[,5],
    xlab="Soft Threshold (power)",ylab="Mean Connectivity", type="n",
    main = paste("Mean connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5], labels=powers, cex=1,col="red")
#build a adjacency "correlation" matrix
# Pick soft power based on value just above the red thresholding line
enableWGCNAThreads()
softPower = 16
adjacency = adjacency(datExpr, power = softPower, type = "signed") #specify network type
# Construct Networks- USE A SUPERCOMPUTER IRL ----------------------------
#translate the adjacency into topological overlap matrix and calculate the corresponding
dissimilarity:
TOM = TOMsimilarity(adjacency, TOMType="signed") # specify network type
dissTOM = 1-TOM
# Generate Modules ------------------------------------------------------


# Generate a clustered gene tree
geneTree = flashClust(as.dist(dissTOM), method="average")
plot(geneTree, xlab="", sub="", main= "Gene Clustering on TOM-based dissimilarity", labels= FALSE,
hang=0.04)
#This sets the minimum number of genes to cluster into a module
minModuleSize = 30
dynamicMods = cutreeDynamic(dendro= geneTree, distM= dissTOM, deepSplit=2,
pamRespectsDendro= FALSE, minClusterSize = minModuleSize)
dynamicColors= labels2colors(dynamicMods)
MEList= moduleEigengenes(datExpr, colors= dynamicColors,softPower = 16)
```

```
MEs= MEList$eigengenes

MEDiss= 1-cor(MEs)

METree= flashClust(as.dist(MEDiss), method= "average")

save(dynamicMods, MEList, MEs, MEDiss, METree, file=
"Massokowild_WGCNA_signed_network.RData")

#plots tree showing how the eigengenes cluster together

plot(METree, main= "Clustering of module eigengenes", xlab= "", sub= "")

#set a threhold for merging modules.

MEDissThres = 0.25

merge = mergeCloseModules(datExpr, dynamicColors, cutHeight= MEDissThres, verbose =3)

mergedColors = merge$colors

mergedMEs = merge$newMEs

dev.off()

#plot dendrogram with module colors below it

plotDendroAndColors(geneTree, cbind(dynamicColors, mergedColors), c("Dynamic Tree Cut",
"Merged dynamic"), dendroLabels= FALSE, hang=0.03, addGuide= TRUE, guideHang=0.05)

moduleColors = mergedColors

colorOrder = c("grey", standardColors(50))

moduleLabels = match(moduleColors, colorOrder)-1

MEs = mergedMEs

dev.off()


save(MEs, moduleLabels, moduleColors, geneTree, file=
"Massokowild_WGCNA_merged_dynamic_tree.RData")

load("Massokowild_WGCNA_merged_dynamic_tree.RData")

# Relate gene expression modules to traits

# Correlate traits ----------------------------------------------------

#Define number of genes and samples

nGenes = ncol(datExpr)

nSamples = nrow(datExpr)

#Recalculate MEs with color labels

MEs0 = moduleEigengenes(datExpr, moduleColors)$eigengenes
```

```
MEs = orderMEs(MEs0)
head(MEs0)
write.csv(MEs, "Massokowild_WGCNA_module_eigengenes_May2022.csv")
moduleTraitCor = cor(MEs, datTraits, method = "pearson")
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples)
moduleTraitCor
MEs0
write.csv(MEs0, "Massokowild_module_eigengenes_May2022.csv")
#Print correlation heatmap between modules and traits
textMatrix= paste(signif(moduleTraitCor, 2), "\n(",
                  signif(moduleTraitPvalue, 1), ")", sep= "")
dim(textMatrix)= dim(moduleTraitCor)
par(mar= c(5, 10, 2, 2))
#display the corelation values with a heatmap plot
labeledHeatmap(Matrix= moduleTraitCor,
               xLabels= names(datTraits),
               yLabels= names(MEs),
               ySymbols= names(MEs),
               colorLabels= FALSE,
               colors= blueWhiteRed(50),
               textMatrix= textMatrix,
               setStdMargins= FALSE,
               cex.text= 0.55,
               zlim= c(-1,1),
               main= paste("Module-trait relationships"))
dev.off()
# write a function for extracting the p-values of an ANOVA
p.aov = function(x, y){
  return(summary(aov(as.matrix(x) ~ y))[[1]][[5]][[1]])
}
# put it all together
```

```
moduleTraitPvalue = cbind(apply(MEs, 2, function(x) p.aov(x,datTraits$Ecomorph)),
                          apply(MEs, 2, function(x) p.aov(x,datTraits$`Upper oral jaw`)),
                          apply(MEs, 2, function(x) p.aov(x,datTraits$`Lower oral jaw`)),
                          apply(MEs, 2, function(x) p.aov(x,datTraits$`Head length`)),
                          apply(MEs, 2, function(x) p.aov(x,datTraits$`Head depth`)))
moduleTraitPvalue
# correct for multiple testing - Benjamin Hochberg - aka FDR correction
moduleTraitPvalue.adj.ecomorph <- p.adjust(moduleTraitPvalue[,1], method = "BH")
moduleTraitPvalue.adj.UOJ <- p.adjust(moduleTraitPvalue[,2], method = "BH")
moduleTraitPvalue.adj.LOJ <- p.adjust(moduleTraitPvalue[,3], method = "BH")
moduleTraitPvalue.adj.HL <- p.adjust(moduleTraitPvalue[,4], method = "BH")
moduleTraitPvalue.adj.HD <- p.adjust(moduleTraitPvalue[,5], method = "BH")
Combined_data <- cbind(moduleTraitPvalue.adj.ecomorph,moduleTraitPvalue.adj.UOJ,
moduleTraitPvalue.adj.LOJ, moduleTraitPvalue.adj.HL, moduleTraitPvalue.adj.HD)
write.csv(Combined_data, ("Massokowild_WGCNA_BH_corrected_heatmap_values_May2022.csv"))
#Print correlation heatmap between modules and traits
textMatrix= paste(signif(moduleTraitCor, 2), "\n(",
                  signif(Combined_data, 1), ")", sep= "")
dim(textMatrix)= dim(moduleTraitCor)
par(mar= c(5, 10, 2, 2))

#display the corelation values with a heatmap plot
labeledHeatmap(Matrix= moduleTraitCor,
               xLabels= names(datTraits),
               yLabels= names(MEs),
               ySymbols= names(MEs),
               colorLabels= FALSE,
               colors= blueWhiteRed(50),
               textMatrix= textMatrix,
               setStdMargins= FALSE,
               cex.text= 0.55,
```

```
        zlim= c(-1,1),
        main= paste("Module-trait relationships"))
dev.off()
setwd("/home/em21733/Documents/Josh/WGCNA/Modules")
# Sig modules
MElightcyan <- names(datExpr[moduleColors=="lightcyan"])
write.csv(MElightcyan, file = "MElightcyan_module_HL_IDs.csv")
length(MElightcyan) # 755 genes SIG FOR HEAD LENGTH

MEyellow3 <- names(datExpr[moduleColors=="yellow3"])
write.csv(MEyellow3, file = "MEyellow3_module_eco_IDs.csv")
length(MEyellow3) # 37 genes SIG FOR ECOMORPH

MEyellow4 <- names(datExpr[moduleColors=="yellow4"])
write.csv(MEyellow4, file = "MEyellow4_module_eco_IDs.csv")
length(MEyellow4) # 849 genes SIG FOR ECOMORPH

MEmidnightblue <- names(datExpr[moduleColors=="midnightblue"])
write.csv(MEmidnightblue, file = "MEmidnightblue_module_HL_IDs.csv")
length(MEmidnightblue) # 893 genes SIG FOR HEAD LENGTH

MEbrown <- names(datExpr[moduleColors=="brown"])
write.csv(MEbrown, file = "MEbrown_module_eco_IDs.csv")
length(MEbrown) # 1452 genes SIG FOR ECOMORPH

MEmediumpurple4 <- names(datExpr[moduleColors=="mediumpurple4"])
write.csv(MEmediumpurple4, file = "MEmediumpurple4_module_HL_IDs.csv")
length(MEmediumpurple4) # 1019 genes SIG FOR HEAD LENGTH

MEbrown4 <- names(datExpr[moduleColors=="brown4"])
write.csv(MEbrown4, file = "MEbrown4_module_HL_IDs.csv")
```

```
length(MEbrown4) # 136 genes SIG FOR HEAD LENGTH

MEmagenta4 <- names(datExpr[moduleColors=="magenta4"])
write.csv(MEmagenta4, file = "MEmagenta4_module_eco_IDs.csv")
length(MEmagenta4) # 186 genes SIG FOR ECOMORPH

MEbrown2 <- names(datExpr[moduleColors=="brown2"])
write.csv(MEbrown2, file = "MEbrown2_module_eco_IDs.csv")
length(MEbrown2) # 261 genes SIG FOR ECOMORPH

MEdarkorange2 <- names(datExpr[moduleColors=="darkorange2"])
write.csv(MEdarkorange2, file = "MEdarkorange2_module_eco_IDs.csv")
length(MEdarkorange2) # 261 genes SIG FOR ECOMORPH

MEthistle3 <- names(datExpr[moduleColors=="thistle3"])
write.csv(MEthistle3, file = "MEthistle3_module_eco_IDs.csv")
length(MEthistle3) # 257 genes SIG FOR ECOMORPH

MEdarkturquoise <- names(datExpr[moduleColors=="darkturquoise"])
write.csv(MEdarkturquoise, file = "MEdarkturquoise_module_eco_IDs.csv")
length(MEdarkturquoise) # 954 genes SIG FOR ECOMORPH

MEdarkviolet <- names(datExpr[moduleColors=="darkviolet"])
write.csv(MEdarkviolet, file = "MEdarkviolet_module_eco_IDs.csv")
length(MEdarkviolet) # 59 genes SIG FOR ECOMORPH

MEantiquewhite2 <- names(datExpr[moduleColors=="antiquewhite2"])
write.csv(MEantiquewhite2, file = "MEantiquewhite2_module_eco_IDs.csv")
length(MEantiquewhite2) # 44 genes SIG FOR ECOMORPH

MEdarkgreen <- names(datExpr[moduleColors=="darkgreen"])
```

```
write.csv(MEdarkgreen, file = "MEdarkgreen_module_eco_IDs.csv")
length(MEdarkgreen) # 403 genes SIG FOR ECOMORPH

MEdarkorange <- names(datExpr[moduleColors=="darkorange"])
write.csv(MEdarkorange, file = "MEdarkorange_module_eco_hl_IDs.csv")
length(MEdarkorange) # 453 genes SIG FOR ECOMORPH HEAD LENGTH

MEdarkseagreen3 <- names(datExpr[moduleColors=="darkseagreen3"])
write.csv(MEdarkseagreen3, file = "MEdarkseagreen3_module_eco_uoj_loj_hl_IDs.csv")
length(MEdarkseagreen3) # 3281 genes SIG FOR ECOMORPH BOTH ORAL JAWS HEAD LENGTH

MEmediumpurple2 <- names(datExpr[moduleColors=="mediumpurple2"])
write.csv(MEmediumpurple2, file = "MEmediumpurple2_module_hl_IDs.csv")
length(MEmediumpurple2) # 250 genes SIG FOR HEAD LENGTH

MEcoral1 <- names(datExpr[moduleColors=="coral1"])
write.csv(MEcoral1, file = "MEcoral1_module_eco_uoj_loj_IDs.csv")
length(MEcoral1) # 90 genes SIG FOR ECOMORPH BOTH ORAL JAWS

MEblack <- names(datExpr[moduleColors=="black"])
write.csv(MEblack, file = "MEblack_module_eco_IDs.csv")
length(MEblack) # 1773 genes SIG FOR ECOMORPH

MEindianred4 <- names(datExpr[moduleColors=="indianred4"])
write.csv(MEindianred4, file = "MEindianred4_module_eco_IDs.csv")
length(MEindianred4) # 68 genes SIG FOR ECOMORPH
#################################################################
HUB_genes <- chooseTopHubInEachModule(
  datExpr,
  moduleColors,
  omitColors = "grey60",
```

```
  power = 8,
  type = "signed")
HUB_genes
write.table(as.data.frame(HUB_genes), "Massokowild_WGCNA_module_HUB_genes_April2022.txt",
sep = "\t")
# Gene relationship to trait in important modules: Gene Significance and Module Membership
# Define variable parity containing the parity column of datTrait
Upper_jaw = as.data.frame(datTraits$Corrected_Upper_Oral_Jaw);
names(Upper_jaw) = "Upper jaw"
# names (colors) of the modules
modNames = substring(names(MEs), 3)
geneModuleMembership = as.data.frame(cor(datExpr, MEs, use = "p"));
MMPvalue = as.data.frame(corPvalueStudent(as.matrix(geneModuleMembership), nSamples));
geneModuleMembership
names(geneModuleMembership) = paste("MM", modNames, sep="");
names(MMPvalue) = paste("p.MM", modNames, sep="");
write.csv(geneModuleMembership,
"WGCNA_Module_Membership_scores_Upper_oral_jaw_April2020.csv")
geneTraitSignificance = as.data.frame(cor(datExpr, Upper_jaw, use = "p"));
GSPvalue = as.data.frame(corPvalueStudent(as.matrix(geneTraitSignificance), nSamples));
names(geneTraitSignificance) = paste("GS.", names(parity), sep="");
names(GSPvalue) = paste("p.GS.", names(parity), sep="");
write.csv(geneTraitSignificance, "WGCNA_Gene_Significance_scores_April2022.csv")
```

Script 2. RScript for extracting hub genes from significant modules

```
setwd("C:/Users/joshu/OneDrive - University of Bristol/Documents/MRes/WGCNA/HUB")
#construct .csv file with gene significance and module membership for each significant module#####
data<-read.csv("MM and GS for sig modules edit.csv",header=T)
library(dplyr)
df1<-split(data,data$Sig)
ecomorph<-df1[[1]]
HL<-df1[[2]]
```

```
dfecomorph<-split(ecomorph,ecomorph$Module)
dfheadlength<-split(HL,HL$Module)
summary(dfecomorph)
aw2<-dfecomorph[[1]]
black<-dfecomorph[[2]]
brown<-dfecomorph[[3]]
brown2<-dfecomorph[[4]]
coral1<-dfecomorph[[5]]
darkgreen<-dfecomorph[[6]]
darkorange<-dfecomorph[[7]]
darkorange2<-dfecomorph[[8]]
darkseagreen3<-dfecomorph[[9]]
darkturquoise<-dfecomorph[[10]]
darkviolet<-dfecomorph[[11]]
indianred4<-dfecomorph[[12]]
magenta4<-dfecomorph[[13]]
thistle3<-dfecomorph[[14]]
yellow3<-dfecomorph[[15]]
yellow4<-dfecomorph[[16]]

blue<-dfheadlength[[1]]
brown4<-dfheadlength[[2]]
lightcyan<-dfheadlength[[3]]
mediumpurple2<-dfheadlength[[4]]
mediumpurple4<-dfheadlength[[5]]
midnightblue<-dfheadlength[[6]]

quantile(aw2$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE)#-0.625
quantile(aw2$MM, prob = 1 - 10/100, na.rm = TRUE) #0.924

filteredaw2<-filter(aw2,GS.ecomorph<"-0.625")
```

```
aw2HUB<-filter(filteredaw2,MM>0.924)

write.csv(aw2HUB, "aw2HUB.csv",row.names = F)

quantile(black$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.719
quantile(black$MM, prob = 1 - 10/100, na.rm = TRUE) #0.853

filteredblack<-filter(black,GS.ecomorph>0.719)
blackHUB<-filter(filteredblack,MM>0.853)

write.csv(blackHUB, "blackHUB.csv",row.names = F)

quantile(brown$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.519
quantile(brown$MM, prob = 1 - 10/100, na.rm = TRUE) #0.853

filteredbrown<-filter(brown,GS.ecomorph<"-0.519")
brownHUB<-filter(filteredbrown,MM>0.853)

write.csv(brownHUB, "brownHUB.csv",row.names = F)

quantile(brown2$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.695
quantile(brown2$MM, prob = 1 - 10/100, na.rm = TRUE) #0.862

filteredbrown2<-filter(brown2,GS.ecomorph<"-0.695")
brown2HUB<-filter(filteredbrown2,MM>0.862)

write.csv(brown2HUB, "brown2HUB.csv",row.names = F)

quantile(coral1$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.528
quantile(coral1$MM, prob = 1 - 10/100, na.rm = TRUE) #0.936
```

```
filteredcoral1<-filter(coral1,GS.ecomorph>"0.528")
coral1HUB<-filter(filteredcoral1,MM>0.936)

write.csv(coral1HUB, "coral1HUB.csv",row.names = F)

quantile(darkgreen$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.646
quantile(darkgreen$MM, prob = 1 - 10/100, na.rm = TRUE) #0.843

filtereddarkgreen<-filter(darkgreen,GS.ecomorph<"-0.646")
darkgreenHUB<-filter(filtereddarkgreen,MM>0.843)

write.csv(darkgreenHUB, "darkgreenHUB.csv",row.names = F)

quantile(darkorange$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.556
quantile(darkorange$MM, prob = 1 - 10/100, na.rm = TRUE) #0.861

filtereddarkorange<-filter(darkorange,GS.ecomorph<"-0.556")
darkorangeHUB<-filter(filtereddarkorange,MM>0.861)

write.csv(darkorangeHUB, "darkorangeHUB.csv",row.names = F)

quantile(darkorange2$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.554
quantile(darkorange2$MM, prob = 1 - 10/100, na.rm = TRUE) #0.823

filtereddarkorange2<-filter(darkorange2,GS.ecomorph<"-0.554")
darkorange2HUB<-filter(filtereddarkorange2,MM>0.823)

write.csv(darkorange2HUB, "darkorange2HUB.csv",row.names = F)

quantile(darkseagreen3$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.620
quantile(darkseagreen3$MM, prob = 1 - 10/100, na.rm = TRUE) #0.868
```

```
filtereddarkseagreen3<-filter(darkseagreen3,GS.ecomorph>0.620)
darkseagreen3HUB<-filter(filtereddarkseagreen3,MM>0.868)

write.csv(darkseagreen3HUB, "darkseagreen3HUB.csv",row.names = F)

quantile(darkturquoise$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.697
quantile(darkturquoise$MM, prob = 1 - 10/100, na.rm = TRUE) #0.853

filtereddarkturquoise<-filter(darkturquoise,GS.ecomorph<"-0.697")
darkturquoiseHUB<-filter(filtereddarkturquoise,MM>0.853)

write.csv(darkturquoiseHUB, "darkturquoiseHUB.csv",row.names = F)

quantile(darkviolet$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.646
quantile(darkviolet$MM, prob = 1 - 10/100, na.rm = TRUE) #0.923

filtereddarkviolet<-filter(darkviolet,GS.ecomorph<"-0.646")
darkvioletHUB<-filter(filtereddarkviolet,MM>0.923)

write.csv(darkvioletHUB, "darkvioletHUB.csv",row.names = F)

quantile(indianred4$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.686
quantile(indianred4$MM, prob = 1 - 10/100, na.rm = TRUE) #0.949

filteredindianred4<-filter(indianred4,GS.ecomorph>0.686)
indianred4HUB<-filter(filteredindianred4,MM>0.949)

write.csv(indianred4HUB, "indianred4HUB.csv",row.names = F)

quantile(magenta4$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.578
```

```
quantile(magenta4$MM, prob = 1 - 10/100, na.rm = TRUE) #0.860

filteredmagenta4<-filter(magenta4,GS.ecomorph<"-0.578")
magenta4HUB<-filter(filteredmagenta4,MM>0.860)

write.csv(magenta4HUB, "magenta4HUB.csv",row.names = F)

quantile(thistle3$GS.ecomorph, prob = 1 - 90/100, na.rm = TRUE) #-0.669
quantile(thistle3$MM, prob = 1 - 10/100, na.rm = TRUE) #0.856

filteredthistle3<-filter(thistle3,GS.ecomorph<"-0.669")
thistle3HUB<-filter(filteredthistle3,MM>0.856)

write.csv(thistle3HUB, "thistle3HUB.csv",row.names = F)

quantile(yellow3$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.504
quantile(yellow3$MM, prob = 1 - 10/100, na.rm = TRUE) #0.917

filteredyellow3<-filter(yellow3,GS.ecomorph>0.504)
yellow3HUB<-filter(filteredyellow3,MM>0.917)

write.csv(yellow3HUB, "yellow3HUB.csv",row.names = F)

quantile(yellow4$GS.ecomorph, prob = 1 - 10/100, na.rm = TRUE) #0.676
quantile(yellow4$MM, prob = 1 - 10/100, na.rm = TRUE) #0.867

filteredyellow4<-filter(yellow4,GS.ecomorph>0.676)
yellow4HUB<-filter(filteredyellow4,MM>0.867)

write.csv(yellow4HUB, "yellow4HUB.csv",row.names = F)
```

```
##########################################################################

quantile(blue$GS.headlength, prob = 1 - 90/100, na.rm = TRUE) #-0.572
quantile(blue$MM, prob = 1 - 10/100, na.rm = TRUE) #0.879

filteredblue<-filter(blue,GS.headlength<"-0.572")
blueHUB<-filter(filteredblue,MM>0.879)

write.csv(blueHUB, "blueHUB.csv",row.names = F)

quantile(brown4$GS.headlength, prob = 1 - 90/100, na.rm = TRUE) #-0.472
quantile(brown4$MM, prob = 1 - 10/100, na.rm = TRUE) #0.861

filteredbrown4<-filter(brown4,GS.headlength<"-0.472")
brown4HUB<-filter(filteredbrown4,MM>0.861)

write.csv(brown4HUB, "brown4HUB.csv",row.names = F)

quantile(lightcyan$GS.headlength, prob = 1 - 10/100, na.rm = TRUE) #0.451
quantile(lightcyan$MM, prob = 1 - 10/100, na.rm = TRUE) #0.834

filteredlightcyan<-filter(lightcyan,GS.headlength>0.451)
lightcyanHUB<-filter(filteredlightcyan,MM>0.834)

write.csv(lightcyanHUB, "lightcyanHUB.csv",row.names = F)

quantile(mediumpurple2$GS.headlength, prob = 1 - 10/100, na.rm = TRUE) #0.528
quantile(mediumpurple2$MM, prob = 1 - 10/100, na.rm = TRUE) #0.866

filteredmediumpurple2<-filter(mediumpurple2,GS.headlength>0.528)
mediumpurple2HUB<-filter(filteredmediumpurple2,MM>0.866)
```

```
write.csv(mediumpurple2HUB, "mediumpurple2HUB.csv",row.names = F)

quantile(mediumpurple4$GS.headlength, prob = 1 - 90/100, na.rm = TRUE) #-0.531
quantile(mediumpurple4$MM, prob = 1 - 10/100, na.rm = TRUE) #0.875

filteredmediumpurple4<-filter(mediumpurple4,GS.headlength<"-0.531")
mediumpurple4HUB<-filter(filteredmediumpurple4,MM>0.875)

write.csv(mediumpurple4HUB, "mediumpurple4HUB.csv",row.names = F)

quantile(midnightblue$GS.headlength, prob = 1 - 90/100, na.rm = TRUE) #-0.484
quantile(midnightblue$MM, prob = 1 - 10/100, na.rm = TRUE) #0.880

filteredmidnightblue<-filter(midnightblue,GS.headlength<"-0.484")
midnightblueHUB<-filter(filteredmidnightblue,MM>0.880)

write.csv(midnightblueHUB, "midnightblueHUB.csv",row.names = F)
```

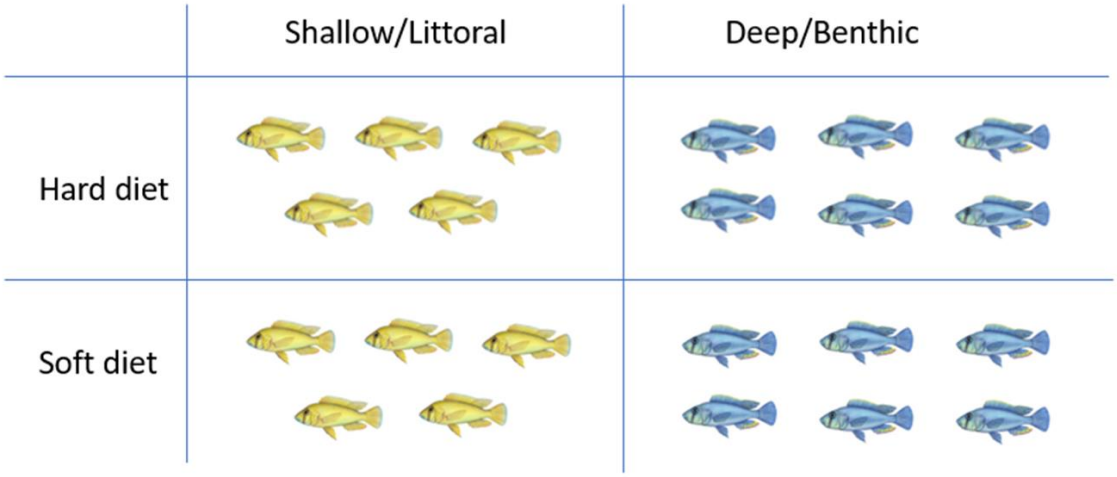


Figure S1: Visual representation of common garden experiments, yellow fish represent fish from shallow/littoral parents, blue fish represent fish from deep/benthic parents.

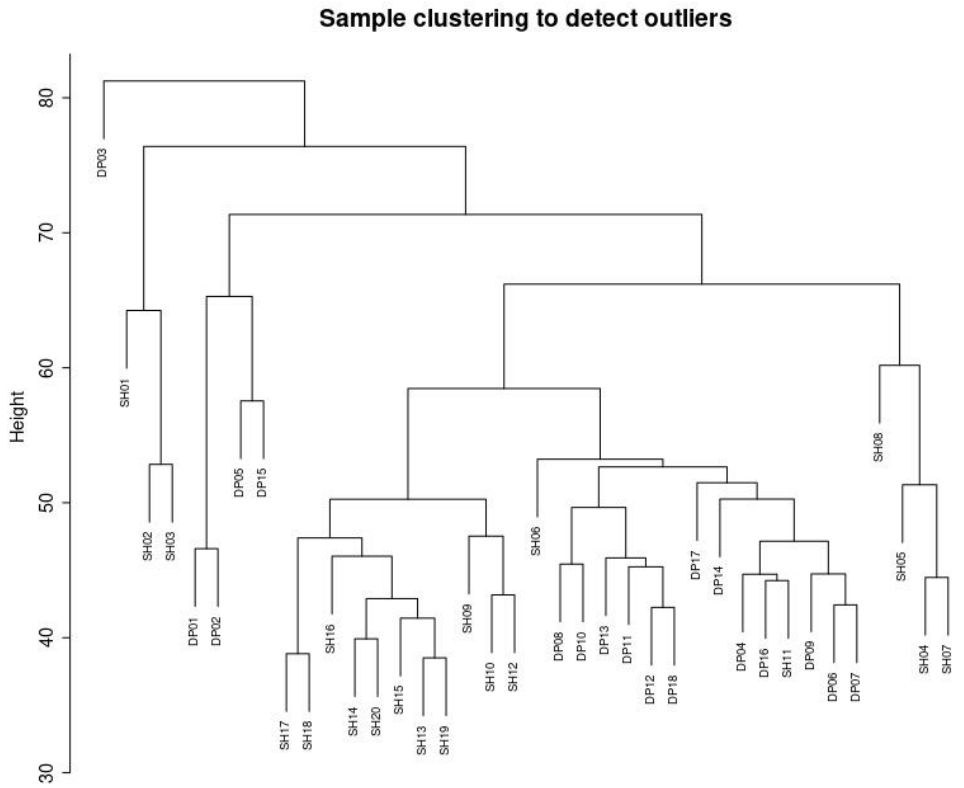


Figure S2: Hierarchical clustering dendrogram of normalised expression counts of 19,237 *M. zebra* annotated genes in 38 RNA-seq samples from LPJ tissue. DP = deep ecomorph, SH = shallow ecomorph.

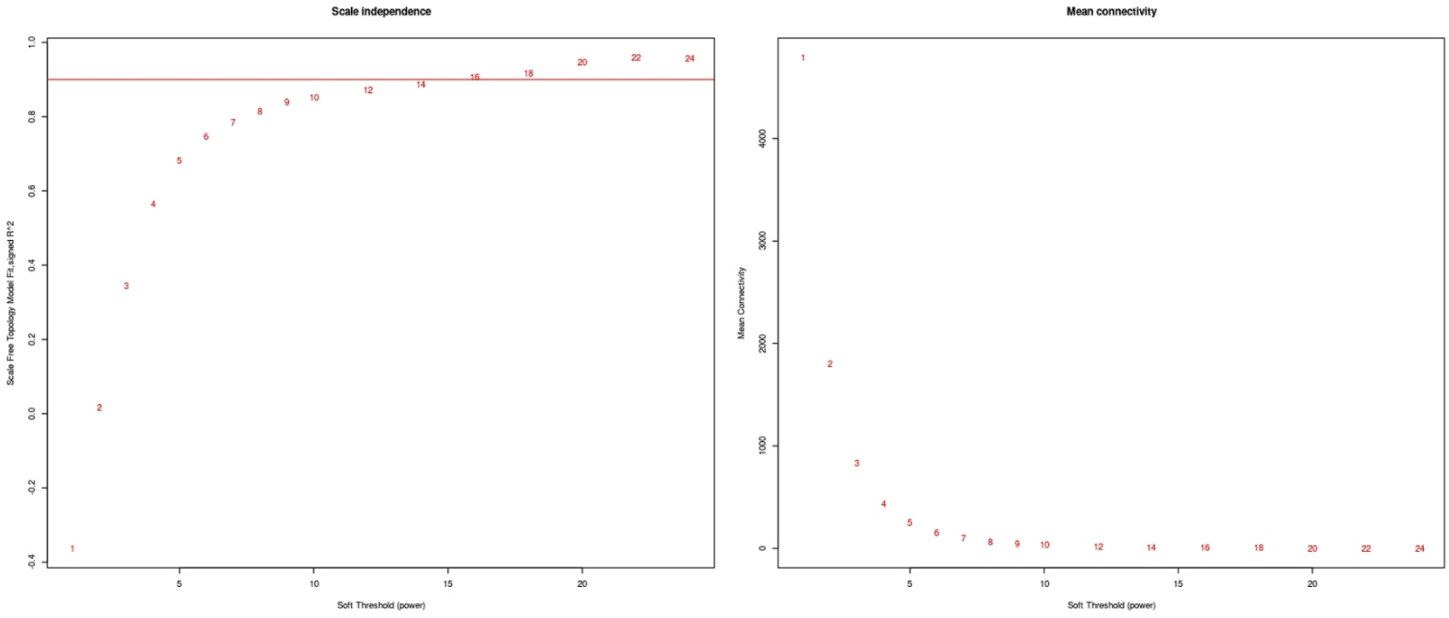


Figure S3: Choice of soft power for scale free network construction. The soft power was chosen according to the biologically motivated scale free topology criterion (Zhang and Horvath 2005): the lowest integer above model fit R^2=0.9 was 16.

Table S1: Gene ontology table for each significant module identified from (Fig. 3.1). GO terms in the table are all statistically overrepresented (*P* <0.05) within their given module against a reference *M. zebra* genome. All terms within the table are biological processes unless stated in the first column. All GO terms presented are child terms and represent the most specific subclass.

Modules significant for ecomorph	**GO term (GO ID)**	**Fold enrichment**	***P*-value**	**FDR value**
Black	interleukin-15-mediated signaling pathway (GO:0035723)	10.95	2.00E-04	1.37E-02
	B cell receptor signaling pathway (GO:0050853)	10.22	1.75E-08	3.32E-06
	complement activation, classical pathway (GO:0006958)	9.59	9.90E-05	7.71E-03
	phagocytosis, recognition (GO:0006910)	9.39	3.46E-04	2.15E-02
	positive regulation of B cell activation (GO:0050871)	8.96	1.39E-05	1.45E-03
	phagocytosis, engulfment (GO:0006911)	6.16	1.15E-04	8.47E-03
	regulation of immune effector process (GO:0002697)	5.16	6.93E-04	3.74E-02
	negative regulation of leukocyte activation (GO:0002695)	5.16	6.93E-04	3.72E-02
	regulation of lymphocyte proliferation (GO:0050670)	5.16	6.93E-04	3.69E-02
	positive regulation of cytosolic calcium ion concentration (GO:0007204)	4.56	3.91E-07	5.35E-05
	T cell activation (GO:0042110)	4.32	6.76E-05	5.54E-03
	negative regulation of cell-cell adhesion (GO:0022408)	4.16	3.18E-04	2.05E-02
	lymphocyte differentiation (GO:0030098)	3.89	5.04E-04	2.88E-02
	calcium-mediated signalling (GO:0019722)	3.83	1.01E-04	7.79E-03
	leukocyte chemotaxis (GO:0030595)	3.76	3.65E-04	2.25E-02
	innate immune response (GO:0045087)	3.65	3.31E-08	5.98E-06
	response to bacterium (GO:0009617)	3.3	3.45E-05	3.23E-03
	inflammatory response (GO:0006954)	2.78	2.42E-04	1.57E-02
	G protein-coupled receptor signalling pathway (GO:0007186)	2.66	1.08E-09	3.06E-07
	regulation of GTPase activity (GO:0043087)	2.11	9.61E-04	4.93E-02
	regulation of RNA metabolic process (GO:0051252)	0.64	6.26E-05	5.35E-03
	protein-containing complex organization (GO:0043933)	0.45	6.69E-04	3.72E-02
	chordate embryonic development (GO:0043009)	0.31	3.35E-04	2.11E-02
	cellular response to stress (GO:0033554)	0.25	4.26E-07	5.73E-05

	ubiquitin-dependent protein catabolic process (GO:0006511)	0.25	3.14E-05	2.97E-03
	ncRNA processing (GO:0034470)	0.24	7.29E-04	3.84E-02
	chromatin organization (GO:0006325)	0.15	4.53E-04	2.62E-02
	mRNA processing (GO:0006397)	0.13	1.03E-04	7.80E-03
	translation (GO:0006412)	0.11	1.12E-05	1.22E-03
	RNA splicing (GO:0008380)	0.08	1.40E-04	9.86E-03
	ribosome biogenesis (GO:0042254)	0.08	1.48E-04	1.03E-02
	mitochondrion organization (GO:0007005)	0.08	6.36E-05	5.32E-03
	regulation of translation (GO:0006417)	< 0.01	4.95E-04	2.85E-02
Brown	mitochondrial respiratory chain complex III assembly (GO:0034551)	11.16	6.60E-04	3.72E-02
	proton motive force-driven ATP synthesis (GO:0015986)	10.23	5.89E-07	7.55E-05
	mitochondrial electron transport, ubiquinol to cytochrome c (GO:0006122)	9.56	3.15E-04	2.12E-02
	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	8.37	2.33E-12	6.16E-10
	ATP synthesis coupled electron transport (GO:0042773)	8.37	2.33E-12	5.96E-10
	cytoplasmic translation (GO:0002181)	7.01	2.28E-10	4.65E-08
	tricarboxylic acid cycle (GO:0006099)	5.58	4.53E-04	2.77E-02
	mitochondrial translation (GO:0032543)	5.02	3.67E-04	2.33E-02
	ribosomal large subunit biogenesis (GO:0042273)	4.22	4.23E-05	3.65E-03
	translational initiation (GO:0006413)	3.72	3.86E-04	2.43E-02
	tRNA processing (GO:0008033)	2.87	2.43E-04	1.68E-02
	regulation of translation (GO:0006417)	2.76	5.88E-05	4.97E-03
	carboxylic acid metabolic process (GO:0019752)	1.87	2.83E-04	1.92E-02
	regulation of transcription by RNA polymerase II (GO:0006357)	0.62	5.89E-04	3.39E-02
	response to chemical (GO:0042221)	0.55	3.37E-04	2.18E-02
	regulation of signal transduction (GO:0009966)	0.54	2.10E-04	1.46E-02
	tissue development (GO:0009888)	0.54	7.58E-05	6.08E-03
	cell development (GO:0048468)	0.5	4.70E-05	4.02E-03
	circulatory system development (GO:0072359)	0.49	4.31E-04	2.67E-02
	neurogenesis (GO:0022008)	0.48	4.73E-04	2.83E-02
	anatomical structure formation involved in morphogenesis (GO:0048646)	0.43	3.98E-04	2.49E-02

	intracellular signal transduction (GO:0035556)	0.4	2.00E-06	2.18E-04
	cell morphogenesis (GO:0000902)	0.37	8.90E-04	4.88E-02
	cell migration (GO:0016477)	0.33	5.89E-05	4.93E-03
	system process (GO:0003008)	0.33	8.71E-04	4.80E-02
	G protein-coupled receptor signalling pathway (GO:0007186)	0.28	3.19E-04	2.11E-02
	regulation of anatomical structure morphogenesis (GO:0022603)	0.24	4.96E-04	2.90E-02
	enzyme-linked receptor protein signalling pathway (GO:0007167)	0.22	1.78E-04	1.31E-02
	protein phosphorylation (GO:0006468)	0.17	6.31E-08	1.04E-05
	actin filament organization (GO:0007015)	0.15	4.56E-04	2.77E-02
	regulation of cytoskeleton organization (GO:0051493)	0.15	4.65E-04	2.80E-02
	cell adhesion (GO:0007155)	0.09	3.56E-07	5.04E-05
	regulation of cell migration (GO:0030334)	0.09	6.76E-04	3.75E-02
Brown2	fin morphogenesis (GO:0033334)	13.54	6.95E-05	4.60E-02
	regulation of Wnt signalling pathway (GO:0030111)	5.53	6.14E-05	4.88E-02
	sensory system development (GO:0048880)	3.58	1.66E-05	2.20E-02
	regulation of transcription by RNA polymerase II (GO:0006357)	2.2	5.32E-05	5.28E-02
Darkgreen	regulation of transcription by RNA polymerase II (GO:0006357)	2.47	2.63E-09	1.39E-06
	organonitrogen compound biosynthetic process (GO:1901566)	0.13	6.00E-05	2.51E-02
Darkorange2	nervous system development (GO:0007399)	2.17	7.03E-06	9.30E-03
	cellular macromolecule metabolic process (GO:0044260)	0.48	4.14E-05	3.66E-02
	nucleobase-containing compound metabolic process (GO:0006139)	0.29	4.46E-05	3.54E-02
	gene expression (GO:0010467)	0.21	4.59E-05	3.32E-02
Darkturqoise	nitrogen compound metabolic process (GO:0006807)	0.72	1.66E-05	4.38E-02
	cellular macromolecule metabolic process (GO:0044260)	0.64	2.81E-05	4.47E-02
	gene expression (GO:0010467)	0.4	6.99E-06	2.78E-02
Thistle3 (pathway)	Gonadotropin-releasing hormone receptor pathway (P06664)	4.36	3.18E-04	4.62E-02
Yellow4	glycerophospholipid biosynthetic process (GO:0046474)	3.89	4.57E-06	1.21E-02
	nitrogen compound transport (GO:0071705)	1.94	1.00E-05	1.59E-02
	organic substance transport (GO:0071702)	1.77	2.61E-05	2.97E-02

Yellow4 (pathway)	Apoptosis signalling pathway (P00006)	3.65	1.69E-04	2.44E-02
Modules significant for head length	**GO term (GO ID)**	**Fold enrichment**	***P*-value**	**FDR value**
Lightcyan	cell-cell junction organization (GO:0045216)	5.13	1.56E-05	1.38E-02
	regulation of cellular process (GO:0050794)	1.25	4.75E-05	3.77E-02
	RNA processing (GO:0006396)	0.12	1.37E-05	1.36E-02
Midnightblue	serine family amino acid biosynthetic process (GO:0009070)	12.27	2.84E-04	3.31E-02
	cholesterol biosynthetic process (GO:0006695)	8.59	2.84E-04	3.41E-02
	snRNA processing (GO:0016180)	7.52	1.66E-04	2.10E-02
	ncRNA 3'-end processing (GO:0043628)	7.36	2.19E-05	4.04E-03
	protein N-linked glycosylation (GO:0006487)	5.52	1.35E-04	1.78E-02
	mRNA transport (GO:0051028)	4.91	1.30E-04	1.78E-02
	mRNA splicing, via spliceosome (GO:0000398)	4.42	4.31E-09	2.14E-06
	tRNA metabolic process (GO:0006399)	3.71	9.26E-07	2.37E-04
	DNA repair (GO:0006281)	2.7	1.21E-05	2.46E-03
	transcription, DNA-templated (GO:0006351)	2.55	4.63E-04	4.90E-02
	mitotic cell cycle (GO:0000278)	2.5	6.51E-05	1.01E-02
	chromosome organization (GO:0051276)	2.36	1.91E-05	3.69E-03
	ribonucleoprotein complex biogenesis (GO:0022613)	2.35	1.97E-04	2.40E-02
	establishment of protein localization (GO:0045184)	1.85	1.32E-04	1.78E-02
	cell communication (GO:0007154)	0.64	1.28E-04	1.82E-02
	signalling (GO:0023052)	0.64	1.48E-04	1.89E-02
	regulation of intracellular signal transduction (GO:1902531)	0.17	4.10E-05	6.64E-03
Midnightblue (Pathway)	Cholesterol biosynthesis (P00014)	9.54	6.54E-04	3.16E-02
	General transcription regulation (P00023)	6.61	1.62E-05	2.36E-03
	Transcription regulation by bZIP transcription factor (P00055)	4.19	3.87E-04	2.81E-02
Mediumpurple2	regulation of transcription by RNA polymerase II (GO:0006357)	2.79	6.23E-09	3.81E-06
Mediumpurple4	transcription by RNA polymerase II (GO:0006366)	3.89	3.34E-05	9.14E-03
	mRNA splicing, via spliceosome (GO:0000398)	3.24	1.14E-05	3.48E-03
	DNA repair (GO:0006281)	2.35	1.99E-04	4.06E-02

	system development (GO:0048731)	0.6	8.09E-06	2.57E-03
	anatomical structure morphogenesis (GO:0009653)	0.57	1.87E-04	3.90E-02
	regulation of response to stimulus (GO:0048583)	0.53	2.48E-04	4.58E-02
	regulation of cell communication (GO:0010646)	0.47	1.23E-04	2.71E-02
	regulation of signalling (GO:0023051)	0.47	9.39E-05	2.13E-02
	cell development (GO:0048468)	0.46	1.42E-04	3.05E-02
	cell surface receptor signalling pathway (GO:0007166)	0.39	8.37E-05	2.02E-02
	intracellular signal transduction (GO:0035556)	0.39	4.62E-05	1.18E-02
Modules significant for multiple variables	**GO term (GO ID)**	**Fold enrichment**	***P*-value**	**FDR value**
Coral1 (Significant for ecomorph and upper/lower oral jaw)	oxygen transport (GO:0015671)	> 100	9.95E-06	1.58E-02
	hydrogen peroxide catabolic process (GO:0042744)	64.75	2.90E-05	2.56E-02
	erythrocyte development (GO:0048821)	27.63	2.11E-05	2.10E-02
	embryonic hemopoiesis (GO:0035162)	16.6	1.77E-05	2.01E-02
Darkseagreen3 (Significant for ecomorph, upper/lower oral jaw and head length)	intracellular signal transduction (GO:0035556)	1.48	2.81E-05	1.31E-02
	protein-containing complex assembly (GO:0065003)	0.47	6.03E-05	2.66E-02
	ncRNA processing (GO:0034470)	0.35	1.29E-04	4.88E-02
	ribonucleoprotein complex biogenesis (GO:0022613)	0.34	6.95E-05	2.91E-02
	translation (GO:0006412)	0.18	8.24E-08	2.18E-04
	tRNA metabolic process (GO:0006399)	0.17	1.30E-04	4.68E-02
Blue (Significant for upper/lower oral jaw and head length)	collagen fibril organization (GO:0030199)	6.78	9.06E-05	6.54E-03
	SMAD protein signal transduction (GO:0060395)	5.75	4.04E-06	5.53E-04
	pharyngeal system development (GO:0060037)	5.56	5.87E-04	2.83E-02
	positive regulation of pathway-restricted SMAD protein phosphorylation (GO:0010862)	5.52	5.80E-05	4.61E-03
	collagen metabolic process (GO:0032963)	5.32	1.60E-04	9.75E-03
	regulation of pathway-restricted SMAD protein phosphorylation (GO:0060393)	5.23	8.35E-05	6.19E-03
	collagen catabolic process (GO:0030574)	5.17	8.43E-04	3.81E-02
	negative regulation of BMP signalling pathway (GO:0030514)	4.52	2.23E-04	1.22E-02

	mucopolysaccharide metabolic process (GO:1903510)	4.32	3.01E-04	1.59E-02
	embryonic camera-type eye formation (GO:0060900)	4.3	5.79E-04	2.81E-02
	negative regulation of axon extension (GO:0030517)	3.93	1.01E-03	4.29E-02
	BMP signalling pathway (GO:0030509)	3.92	1.87E-04	1.09E-02
	notochord development (GO:0030903)	3.57	1.40E-04	9.00E-03
	vasculogenesis (GO:0001570)	3.55	1.11E-03	4.58E-02
	proteoglycan metabolic process (GO:0006029)	3.48	1.77E-04	1.06E-02
	cell-cell junction assembly (GO:0007043)	3.33	4.11E-04	2.07E-02
	negative regulation of Wnt signalling pathway (GO:0030178)	2.89	5.92E-04	2.82E-02
	chemical synaptic transmission (GO:0007268)	2.7	4.21E-05	3.64E-03
	axon guidance (GO:0007411)	2.56	7.80E-06	9.39E-04
	ameboidal-type cell migration (GO:0001667)	2.19	3.29E-04	1.73E-02
	cell-cell adhesion (GO:0098609)	2.09	8.37E-04	3.87E-02
	angiogenesis (GO:0001525)	2	8.53E-04	3.83E-02
	skeletal system development (GO:0001501)	1.97	1.95E-04	1.11E-02
	morphogenesis of an epithelium (GO:0002009)	1.81	3.81E-04	1.95E-02
	protein modification by small protein conjugation or removal (GO:0070647)	0.5	9.18E-04	4.05E-02
	intracellular protein transport (GO:0006886)	0.33	5.23E-05	4.20E-03
	chromosome organization (GO:0051276)	0.3	9.20E-05	6.58E-03
	ubiquitin-dependent protein catabolic process (GO:0006511)	0.27	1.33E-05	1.51E-03
	DNA repair (GO:0006281)	0.26	5.24E-04	2.59E-02
	mRNA processing (GO:0006397)	0.22	2.25E-04	1.22E-02
	negative regulation of gene expression (GO:0010629)	0.14	2.84E-04	1.51E-02
	mitochondrion organization (GO:0007005)	0.12	4.32E-05	3.65E-03
	translation (GO:0006412)	0.05	4.37E-08	1.09E-05
	RNA modification (GO:0009451)	< 0.01	1.04E-03	4.37E-02
	rRNA processing (GO:0006364)	< 0.01	8.71E-05	6.35E-03
	tRNA processing (GO:0008033)	< 0.01	1.03E-03	4.33E-02

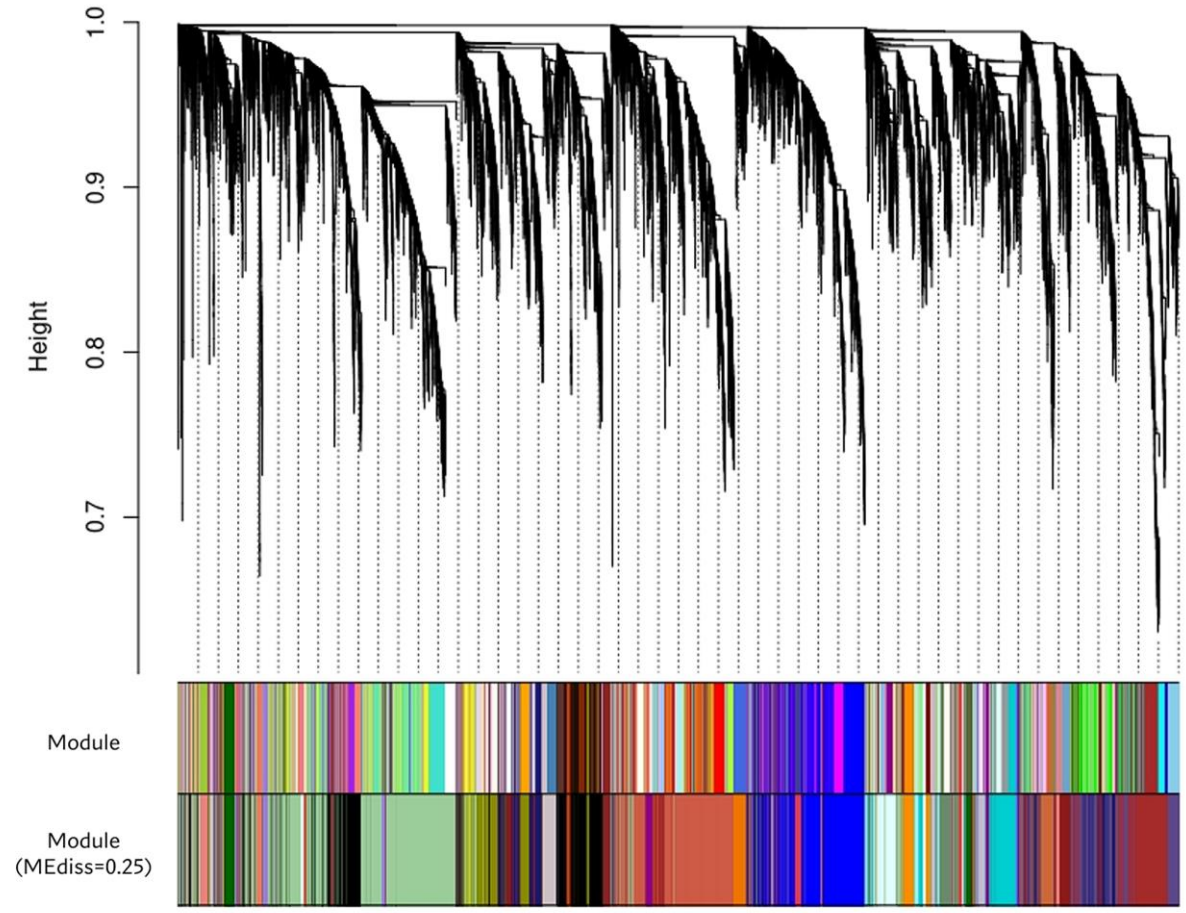


Figure S4: Visual representation of co-expression clustering. Average linkage clustering tree (dendrogram) based on topological overlap distance in gene expression profiles from wild samples (n=38; 18 deep, 20 shallow). Branches of the dendrogram correspond to co-expressed modules, which are each assigned to a colour in the module row. Modules were then further merged based on a module eigengene (ME) dissimilarity threshold of 0.25.

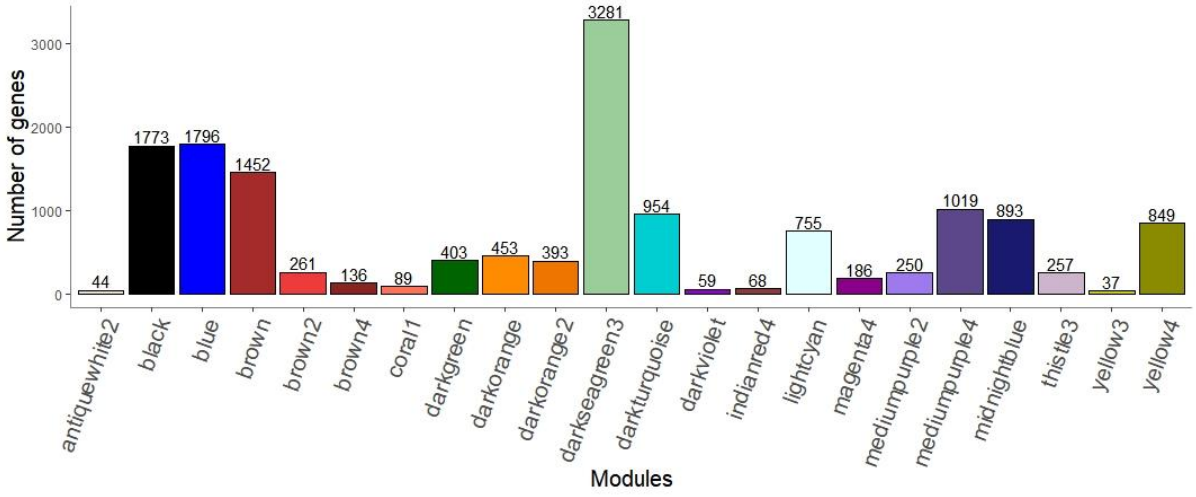


Figure S5: Bar plot number genes within each significant module, each bar is labelled with the number of genes in each module.

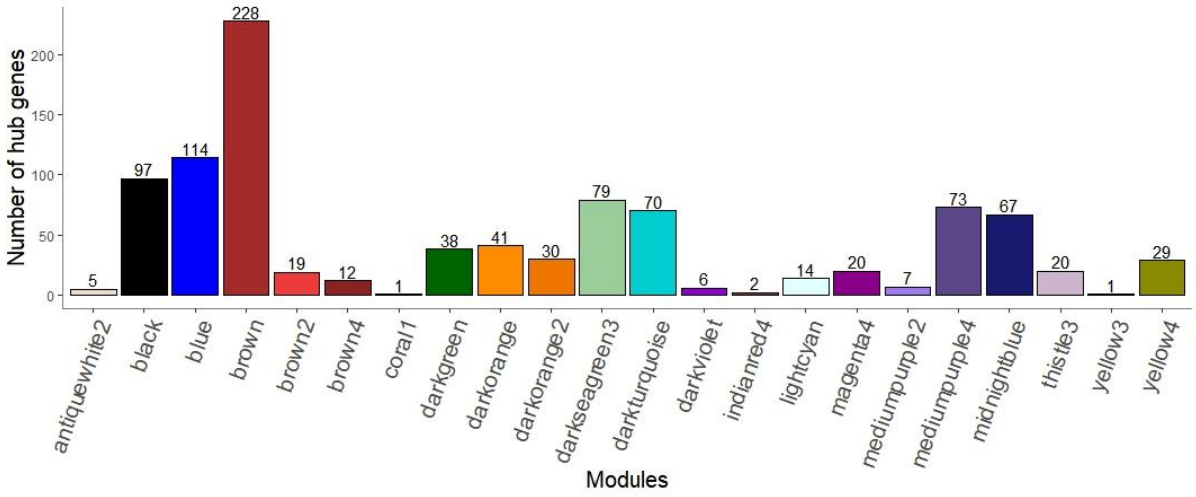


Figure S6: Bar plot number of hub genes within each significant module, each bar is labelled with the number of genes in each module.

Table S2: Number of genes within each dataset (gene list).

Dataset	Number of genes
Entire genome dataset	31220
Entire expression dataset	19237
Significant trait-associated modules	15408
Non-significant trait associated modules	3829
DMRs total	4658
DM genes total	4177
DM genes in entire expression dataset	2586
DM genes within significant modules	2144
DM genes within non-significant modules	442
Hub genes from significant modules	973
DM hub genes	138

Table S3: details on LPJ master genes located within significant trait-associated modules, that are either differentially methylated or key regulatory 'hub' genes. Taken from Supplementary table 6 in Carruthers et al., (2022).

Gene symbol	Gene name	Network	Study	Differentially expressed (DE)	Hub gene	DM gene	Methylation gain in …	Module
rxra	Retinoid X receptor alpha 2	retinoic acid signalling pathway	Hulsey et al. 2016	yes	yes	yes	Benthic	Darkorange2
runx2	runt-related transcription factor 2	multiple pathways/bone related	Schneider et al. 2014, Fraser et al. 2009, 2013	yes	yes	no	N/A	Blue
col6a1	collagen alpha-1(VI) chain	cichlid jaw plasticity network/matrix related	Gunter et al. 2013, Schneider et al. 2014	yes	no	yes	Benthic	Blue
fndc1	fibronectin type III domain-containing protein 1	cichlid jaw plasticity network	Gunter et al. 2013	no	no	yes	Benthic	Darkorange2
lef1	lymphoid enhancer-binding factor 1	wnt/beta-catenin signalling	Singh et al. 2017	yes	no	yes	Benthic	Blue
ptch1	patched 1	hedgehog pathway	Fraser et al. 2013, Roberts et al. 2011, Hu and Albertson 2014, Parsons et al. 2016	yes	no	yes	Benthic	Darkorange2
rarg	retinoic acid receptor gamma-A	retinoic acid signalling pathway	Hulsey et al. 2016	no	no	yes	benthic/littoral	Lightcyan
slmap	sarcolemma associated protein	cichlid jaw plasticity network	Gunter et al. 2013	yes	no	yes	Benthic	Darkseagreen3