Phylogenomic analyses show repeated evolution of hypertrophied lips among Lake Malawi cichlid fishes

Paul Masonick[1], Axel Meyer[1], C. Darrin Hulsey[1,2]

[1]Department of Biology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany

[2]Current Address: School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland

Corresponding author: Darrin Hulsey

E-mail: darrin.hulsey1@ucd.ie

Abstract

Cichlid fishes have repeatedly evolved an astounding diversity of trophic morphologies. For example, hypertrophied lips have evolved multiple times in both African and Neotropical cichlids and could have even evolved convergently within single species assemblages such as African Lake Malawi cichlids. However, the extremely high diversification rate in Lake Malawi cichlids and extensive potential for hybridization has cast doubt on whether even genome-level phylogenetic reconstructions could delineate if these types of adaptations have evolved once or multiple times. To examine the evolution of this iconic trait using protein-coding and non-coding single nucleotide polymorphisms (SNPs), we analyzed the genomes of 86 Lake Malawi cichlid species, including 33 *de novo* resequenced genomes. Surprisingly, genome-wide protein-coding SNPs exhibited enough phylogenetic informativeness to reconstruct inter- and intra-specific relationships of hypertrophied lip cichlids, although non-coding SNPs provided better support. However, thinning of non-coding SNPs indicated most discrepancies come from the relative smaller number of protein-coding sites and not from fundamental differences in their phylogenetic informativeness. Both coding and non-coding reconstructions showed that several “sand-dwelling” hypertrophied lip species, sampled intraspecifically, form a clade interspersed with a few other non-hypertrophied lip lineages. We also recovered *Abactochromis labrosus* within the rock-dwelling "mbuna" lineage, starkly contrasting with the affinities of other hypertrophied lip taxa found in the largely sand-dwelling “non-mbuna” component of this radiation. Comparative analyses coupled with tests for introgression indicate there is not widespread introgression between the hypertrophied lip lineages and taken together suggest this trophic phenotype has likely evolved at least twice independently within Lake Malawi.

Key words

whole genome resequencing, mbuna, parallelism, coding vs. non-coding

Significance

Convergent evolution is widespread in nature. While this phenomenon is known to occur across cichlid fishes found in different parts of the world, in this study we used genome-wide SNPs to resolve whether a specialized trophic morphology, hypertrophied lips, evolved convergently among cichlids endemic to Lake Malawi. Our analyses provided well-supported inferences of relationships even within closely related species and showed that hypertrophied lips likely evolved at least twice independently within the two major radiations of Lake Malawi cichlids.

Introduction

Evolution does repeat itself and convergently evolved adaptations speak to the non-randomness of phenotypic evolution. As species diversify, ecological selective pressures could commonly result in replicate evolution within divergent lineages and drive them to converge on similar adaptive phenotypes. Convergent evolution is an especially common feature of adaptive radiations. For instance, Darwin's finches (Grant and Grant 2008), Caribbean *Anolis* lizards (Losos et al. 1998), and three-spined stickleback fishes (Rundle et al. 2000; Marques et al. 2017) all have repeatedly evolved the same structural modifications. Most of this replicate evolution, whether it occurred over long timeframes independently or rapidly through convergence, has been documented in different geographic localities. The exceptionally diverse radiations of cichlid fishes offer classic examples of these allopatrically derived convergent phenotypes (Meyer 1993; Kocher et al. 1993; Stiassny and Meyer 1999; Salzburger et al. 2005; Salzburger 2009; Elmer and Meyer 2011; Muschick et al. 2012; Kratochwil et al. 2018; Ronco et al. 2021). However, parallelism can also occur within the same geographic region and even within the same closely related lineage (Elmer et al. 2011 PTRS; Torres-Dowdall and Meyer 2021). Yet, when similar phenotypes arise in the same geographic locations over short timeframes, especially in highly diverse radiations, it is difficult for even genome-wide data to evaluate whether traits have arisen only once, evolved independently, originated in parallel, been retained as ancient polymorphisms, or are shared among taxa due to hybridization (Hulsey et al. 2019; Kautt et al. 2020). To evaluate between these evolutionary alternatives, i.e., whether a particular adaptive phenotype arose once or multiple times in a classic example of adaptive radiation, we evaluated the ability of both protein-coding and non-coding data obtained from whole genome resequencing to delineate the evolution of hypertrophied lips in Lake Malawi cichlids.

Cichlids have acquired a huge diversity of trophic morphologies that are specialized for different feeding niches, and hypertrophied lips are one of the most easily recognized phenotypes that have evolved multiple times independently (Burress et al. 2013; Colombo et al. 2013; Manousaki et al. 2013; Henning and Meyer 2014; Baumgarten et al. 2015; Henning et al. 2017). This distinct trophic innovation found in both Neotropical and African cichlid lineages is typically associated with fish that forage in rocky substrates where the lips may act as a seal to help suck prey from in between narrow cracks and crevices (Ribbink et al. 1983; Baumgarten et al. 2015), absorb stress from repeated contact with hard and rough surfaces (Fryer and Iles 1972; Greenwood 1974), and/or enhance prey detection by providing an enlarged area for taste receptors (Oliver and Arnegard 2010). This morphology also exhibits strong feeding tradeoffs with hypertrophied lip fish being more efficient at extracting prey from crevices but less apt at capturing evasive prey in open water (Elmer et

al. 2010; Colombo et al. 2013; Machado-Schiaffino et al. 2017). Lip size may not only be important in terms of natural selection, but also play a role in sexual selection since assortative mating based on lip size has been found in polymorphic populations (Machado-Schiaffino et al. 2014; Machado-Schiaffino et al. 2017). Considerable plasticity in the trait has also been observed in the laboratory possibly hinting that this trait can be acquired and lost over short timeframes (Machado-Schiaffino et al. 2014; Machado-Schiaffino et al. 2017). The relative ease of diagnosing this qualitative phenotype that has testable ecological consequences makes it a model trait for studying adaptive evolution and convergence.

The hypertrophied lip phenotype is also one of the iconic examples of a trait that has arisen independently in all three major cichlid adaptive radiations that are endemic to East Africa's largest great lakes. This specialized morphology is found in *Haplochromis chilotes* from Lake Victoria, *Lobochilotes labiatus* from Lake Tanganyika, as well as eight species (including one undescribed species) native to Lake Malawi (Fryer and Iles 1972; Oliver and Arnegard 2010). Given the recurrent evolution of hypertrophied lips across these major East African cichlid lineages, it is plausible that the phenotype has also evolved independently multiple times in Lake Malawi (Hulsey et al. 2018). With roughly 850 species of haplochromine cichlids inhabiting this large African lake (Figure 2), the opportunity for adaptive traits to arise repeatedly in Malawi is considerable (Danley and Kocher 2001; Genner and Turner 2005; Konings 2007). The vast majority of Lake Malawi endemic cichlids belong to the tribe Haplochromini, but in this lake, have traditionally been placed in two main lineages: the primarily rock-dwelling mbuna (Genner and Turner 2005) and the largely sand-dwelling non-mbuna (Danley and Kocher 2001). These major ecomorphological groups have also consistently been recovered as distinct clades in molecular phylogenetic analyses (Meyer et al. 1990; Meyer 1993; Meyer et al. 1994, 1996; Hulsey et al. 2017; Hulsey et al. 2018; Malinsky et al. 2018). Additionally, all but one species of hypertrophied lip taxa are classified as non-mbuna haplochromines and have been assorted into different genera largely on the basis of body pigmentation patterns (Fryer and Iles 1972; Arnegard and Snoeks 2001; Snoeks 2004; Konings 2007). The one putative mbuna hypertrophied lip species, *Abactochromis labrosus,* is an evolutionarily enigmatic cichlid (Trewavas 1935; Eccles and Trewavas 1989; Oliver and Arnegard 2010). Only a single phylogenetic study has included *A. labrosus* along with a limited number of taxa (15 Lake Malawi species in total) and this was based solely on mitochondrial control region sequences that inferred this species split at the base of the mbuna radiation (Meyer et al. 1996). Hypertrophied lips provide a readily diagnosable and potentially phylogenetically labile phenotype that could provide a more general model to test alternative hypotheses about what we can infer regarding how novel traits tend to evolve in Lake Malawi.

Previous work with mitochondrial DNA, nuclear AFLP loci, and ultraconserved elements (UCEs) have all repeatedly highlighted the issues in obtaining a clear phylogenetic consensus for the Malawi radiation (Kocher et al. 1995; Salzburger and Meyer 2004; Hulsey et al. 2010; Joyce et al. 2011; Hulsey et al. 2013; Hulsey et al. 2018). In seeking to address whether the hypertrophied lip phenotype originated multiple times among Lake Malawi cichlids, Hulsey et al. (2018) evaluated the relationships of hypertrophied lip taxa from several genera by analyzing single nucleotide polymorphisms (SNPs) from UCEs. Their results suggested that *Cheilochromis euchilus*, *Eclectochromis ornatus*, *Placidochromis milomo*, and *Placidochromis* "Mbenji fatlip", all taxa that forage along rocky shores (Froese and Pauly 2021), along with several normal-sized lip species form a clade and that this remarkable phenotype and trophic guild originated just once among non-mbuna haplochromines (Hulsey et al. 2018). Conversely, the disparate placement of hypertrophied lip species in reconstructions based on whole genome sequences suggests that the enlarged lip condition either evolved more than once or possibly reverted to the ancestral condition in several closely related lineages (Malinsky et al. 2018). However, a lack of intra-specific sampling with respect to these taxa in that study limited what can be inferred regarding their species-level relationships. Further, these previous studies contrasted with findings from earlier analyses based on mitochondrial data that found these hypertrophied lip taxa to be highly polyphyletic with *P. milomo* inferred to be nested even within the mbuna (Hulsey et al. 2010; Joyce et al. 2011).

Phylogenomic studies have traditionally relied more heavily on coding sequences due to the relative ease of PCR-amplification and the ease of identifying orthologous sequences in conserved amino acids that facilitates straightforward alignment of homologous sequences (Townsend et al. 2008; Thomson et al. 2010). However, when compared with other faster-evolving non-coding regions, the more conserved nature of coding sequences may limit the ability to resolve the evolutionary history of recent and rapid adaptive radiations (Meyer 1994; El Taher et al. 2021). Further, coding sequences could more often be functionally constrained, potentially prone to convergent evolution, and carry signal incongruent with the true species tree (Steinke et al. 2006; Parker et al. 2013). In contrast, non-coding regions are generally less susceptible to convergence, exhibit greater variability because of their faster substitution rates, and are informative at shallow evolutionary timescales (Meyer et al. 1990; Chojnowski et al. 2008; Yu et al. 2011; Foley et al. 2015). Given these factors, non-coding data may provide more robust phylogenetic signal than coding sequence data in a rapid radiation like the Lake Malawi cichlids and be especially useful for parsing the number of times traits like hypertrophied lips have evolved (Chen et al. 2017). However, in the era of whole genome resequencing, a distinction

between using coding and non-coding DNA for phylogenetics could seem extraneous since both can be readily obtained from the same completely sequenced genomes. Nonetheless, there is an ever-increasing ability to combine genome-wide data with subsets of the genome such as RAD-tag markers or transcriptome sequences (Sharma et al. 2014; Rahman et al. 2021). Transcriptome sequences for instance primarily produce information about protein-coding sequences and a heavy reliance on this component of the genome could be problematic (Lemmon and Lemmon 2013; Yang and Smith 2014; Cheon et al. 2020; Smith and Hahn 2021). Distinguishing the relative ability of coding and non-coding sequences to resolve phylogenetic relationships at various stages of diversification among Lake Malawi cichlids could inform not only future sequencing strategies but influence inferences concerning phenotypic convergence.

Robust species tree reconstructions play a crucial role in testing for convergence and one of its most powerful uses is to reveal whether phenotypically similar traits have originated once or multiple times (Omland 1999; Revell 2012). Nevertheless, our knowledge of Malawi cichlid relationships and ability to draw meaningful conclusions regarding the evolution of the group has long been impeded by the limited ability of molecular markers to provide phylogenetic resolution (Salzburger and Meyer 2004, Hashem et al. 2020). A number of factors that have complicated the phylogenetic reconstruction of Malawi cichlids are also shared with the even faster evolving Lake Victoria cichlid radiation. These factors include the impressive phenotypic diversity of these fishes, the recent ages of the radiations (~2 million years old and perhaps as young as ~0.4 Mya) (Hulsey et al. 2010; Friedman et al. 2013; Genner et al. 2015; Meyer et al. 2017), exceptionally low interspecific genetic divergence (Meyer et al. 1990; Malinsky et al. 2018), rampant incomplete lineage sorting (Moran and Kornfield 1993), and the high potential for hybridization within the clades (Mims et al. 2010; Brawand et al. 2014). In both Lakes Victoria and Malawi, all of these factors could make resolving whether a trait like hypertrophied lips has originated multiple times intractable.

Additionally, even if phylogenetic inference typically supports the non-monophyly of a trait such as hypertrophied lips, it is difficult to discount that a trait might have only evolved a single time. This is because traits might also have arisen once but be lost multiple times thereby appearing to have evolved repeatedly. The evolutionary loss of eyes in cave adapted fishes (Coghill et al. 2014), flightlessness in island birds (Wright et al. 2016), and lack of terrestrial adult stages in neotenic salamanders (Riley et al. 2003; Bonett et al. 2021) all have likely occurred multiple times and should not lead to the erroneous inference that distinctive and complex traits such as eyes, wings, or adult forms evolved multiple times. Therefore, even the best phylogenetic reconstructions of trait evolution make it difficult to

ascertain how many times a trait was gained and lost among exceptionally closely related taxa. Furthermore, despite the fact that traits might seem to have arisen repeatedly in more phylogenetically disparate groups, introgression can play a role in the phylogenetic distribution of these "convergent" phenotypes (Stern 2013).

Inferring whether convergent phenotypes have evolved could be impacted substantially by the degree of introgression that has occurred within the Lake Malawi radiation (Figure 1). Given that many lineages of Malawi cichlids do hybridize (Stauffer et al. 1996; Streelman et al. 2004; Mims et al. 2010), interspecific gene flow even across evolutionarily disparate lineages could play a large role in the phylogenetic distribution and putative independent evolution of a trait like hypertrophied lips. In addition to the difficulties incomplete lineage sorting pose for phylogenetic inference, if genetic admixture was rampant during the diversification of Lake Malawi cichlids, tree-like phylogenetic signal may be too severely obscured to resolve relationships at any level and leave us unable to assess convergence in any traits (Figure 1A). Even with a better resolved phylogeny, if gene flow was particularly extensive among hypertrophied lip taxa, we might fail to recover conspecifics as monophyletic groups and could have difficulty inferring the number of origins of a trait (Figure 1B). Alternatively, if phylogenetic structure is clearly retraceable and conspecifics are found to be monophyletic even in the presence of low levels of hybridization (Figure 1C), we could obtain a clear indication of whether the hypertrophied lip trait was likely independently derived in multiple lineages.

Whole genome resequencing now offers the opportunity to resolve relationships among closely related Malawi cichlids and should provide the power to reveal cases of within-lake convergent evolution (Brawand et al. 2014; Conte et al. 2019). In this study, we employed whole genome resequencing to explore the evolution of the hypertrophied lip phenotype among Lake Malawi cichlids, evaluated the ability of SNPs from coding and non-coding regions to reconstruct these and other relationships at different tree depths, and tested whether substantial gene flow could explain the phylogenetic distribution of the hypertrophied lip phenotype. To assess the evolutionary history of hypertrophied lip taxa, we sampled *Abactochromis labrosus* along with four hypertrophied lip non-mbuna species (three of which were represented by multiple intraspecific samples), additional lineages of normal-lipped sand-dwelling non-mbuna, and a range of rock-dwelling mbuna species. This broad taxonomic sampling enabled us to test i) whether hypertrophied lip non-mbuna species evolved repeatedly, ii) are closely related to each other, and iii) finally narrow down the enigmatic phylogenetic position of *Abactochromis* as possibly the only hypertrophied lip mbuna. As it can be difficult to know whether traits evolved independently or via allele sharing and adaptive introgression, we also tested for interspecific gene flow across the

radiation with a focus on hypertrophied lip species within Malawi cichlids to see if large-scale introgression could readily explain any inferences of the phenotype's repeated origins.

Results and Discussion

Phylogenomics of hypertrophied lips in Lake Malawi

Our phylogenetic analyses are based on 1,352,537 SNPs derived from whole genome resequencing of 86 Lake Malawi cichlid species that included 33 newly generated genome sequences (Supplementary Table 1). We recovered robust phylogenetic hypotheses of Malawi cichlid relationships based on concatenation (IQ-TREE, Nguyen et al. 2015) and multi-species coalescent-based (SVDquartets - Chifman and Kubatko 2014) approaches that inferred largely similar species trees across both the non-coding and coding SNP datasets (Figures 2 & 3, Supplementary Figure 1). All analyses found distinct, well supported clades containing the rock-dwelling mbuna and primarily sand-dwelling non-mbuna (nodes B and C, respectively Figures 2 and 3, Supplementary Figure 1: 100 UFBS / 100 BS). Further, the hypertrophied lip taxa were unambiguously resolved as polyphyletic (node A of Figures 2 and 3 represents the inferred MRCA of all hypertrophied lip species). *Abactochromis labrosus* is clearly nested within the mbuna and shares a close affinity to the other rock-dwelling lineages *Labidochromis*, *Iodotropheus sprengerae*, and the two *Melanochromis* species examined. This result corroborates relatively recent taxonomic work that proposed that *A. labrosus* is distinct from species of *Melanochromis* but is still likely a member of the mbuna (Oliver and Arnegard 2010). However, the two *Labidochromis* species sampled are not monophyletic with *L. ianthinus* and *I. sprengerae* grouping together and *L. gigas* clustering with *Melanochromis*. The hypertrophied lip taxa *Cheilochromis euchilus*, *Eclectochromis ornatus*, *Placidochromis milomo*, and *Placidochromis* "Mbenji fatlip" form a clade within the non-mbuna that also contains a few non-hypertrophied lip species (*Chilotilapia rhoadesii*, *Hemitaenichromis spilopterus*, and *Placidochromis johnstoni*) (node D, Figures 1 and 2: 100 UFBS / 99 BS (non-coding) / 67 BS (coding)). *Cheilochromis euchilus* and the normal-lipped *C. rhoadesii* are recovered as sister taxa in both the ML and SVDquartet reconstructions. While *P. milomo* and *P. johnstoni* compose another group according to our IQ-TREE reconstruction, SVDquartets found *P. milomo* as the sister to the other hypertrophied lip taxa + *C. rhoadesii*, *H. spilopterus*, and *P. johnstoni*. Our phylogenetic results support the notion that there has likely been repeated evolution of the hypertrophied lip phenotype in Lake Malawi cichlids.

As we did not recover a group exclusively composed of hypertrophied lip non-mbuna, the question remains as to whether enlarged lips arose once or multiple times among these taxa. We investigated this further with maximum likelihood based ancestral state

reconstruction. From this analysis, the approximately 50% probability that the ancestor of any specific hypertrophied lip species and its close relatives possessed enlarged lips suggests that there could have been multiple transitions to the hypertrophied-lip phenotype within this relatively small sub-clade of the non-mbuna (see reconstructions for lineages nested under node D, Figure 2, Supplementary Figure 2). An alternative explanation for this pattern would be that the ancestor of these lineages evolved hypertrophied lips and that there were multiple reversals back to the much more common normal-lip morphology (e.g., with possible reversals occurring in *P. johnstoni* and/or *C. rhoadesii*). In contrast to this relative ambiguity, the probability that the common ancestor of the non-mbuna hypertrophied lip species and *Abactochromis* had hypertrophied lips is exceedingly low (0.1% - node A, Figure 2), and it is highly unlikely that the most recent common ancestor of the mbuna or non-mbuna exhibited this phenotype (0.5% - node B, 0.1% - node C, Figure 2). Based on our phylogenetic inferences coupled with ancestral state reconstruction, we can confidently conclude that there clearly are at least two independent origins of hypertrophied lips in the Lake Malawi cichlids (one in the mbuna and one in the non-mbuna).

However, as missing taxa may bias ancestral state reconstruction (Omland 1999; Salisbury and Kim 2001), any ancestral state reconstruction in this diverse radiation should be interpreted with caution. *Otopharynx pachycheilus*, *Lichnochromis acuticeps*, and *Trematocranus pachychilus* are three other Malawi haplochromine cichlids with enlarged lips that were not sampled in this study. However, according to previous analysis, *L. acuticeps* is closely related to the hypertrophied lip non-mbuna (Hulsey et al. 2018). *Otopharynx* represents another genus of non-mbuna, and in line with the notion that it represents an artificial grouping (Arnegard and Snoeks 2001), is rendered polyphyletic in all our analyses (Figures 2 and 3, Supplementary Figure 1). Therefore, the relation of *O. pachycheilus*, a rare deep-water species, to the other hypertrophied lip taxa remains untested and could represent an additional origin of the hypertrophied lip phenotype in Malawi cichlids. Likewise, if the recently described *Trematocranus pachychilus* (Dierickx et al. 2018) is indeed a congener of *Trematocranus placodon*, a species which was analyzed herein, this may represent yet another instance of convergent evolution of hypertrophied lips. Although logistically quite difficult to perform on this incredibly diverse radiation, a more quantitative assessment of the size and diversity of tissues contributing to the enlarged lips would further allow us to evaluate the degree of the repeated origins of these hypertrophied lip phenotypes.

Several other more generalizable phylogenetic patterns emerged from our analyses of the genomes of the 86 species of closely related Lake Malawi cichlid fishes. SNPs obtained from whole genome resequencing provided considerable power to test the

monophyly of diagnosed taxa, particularly when concatenated for analysis with maximum likelihood in IQ-TREE. Yet, we obtained insufficient resolution for many of the relationships within the mbuna based on bootstrapping of the SVDquartet reconstructions. This stood in contrast to the high support across many of the mbuna estimated through ultrafast bootstrapping in IQ-TREE. Additionally, depending on the reconstruction method, the pelagic genera *Rhamphochromis* and *Diplotaxodon*, along with the possibly previously misidentified *Mylochromis lateristriga and Pallidochromis tokolosh*, comprise the earliest diverging group of the Malawi radiation (IQ-TREE) as has been found previously using mtDNA data (Meyer et al. 1994, 1996). However, these taxa are inferred to be either the sister to the non-mbuna or mbuna + *A. calliptera* clade based on SVDquartet analysis of non-coding or coding SNPs, respectively. The position of *Astatotilapia calliptera*, a haplochromine also found commonly outside of Lake Malawi, varied depending on reconstruction method and the SNP dataset used. In most analyses, *A. calliptera* is sister to the mbuna. Yet, in our SVDquartets tree derived from non-coding SNPs, *A. calliptera* was recovered as sister to the rest of the Lake Malawi radiation (its traditional placement – Meyer, Montero, Spreinat 1994, 1996; Malinsky et al. 2018). While many genera were recovered as monophyletic with high support across analyses (i.e., *Buccochromis*, *Dimidiochromis*, *Labeotropheus*, *Nimbochromis*, and *Taeniolethrinops*), there are several taxonomic groupings that appear to be artificial taxonomic entities. *Placidochromis* is highly polyphyletic and several species exhibiting normal sized lips (*Placidochromis electra* and *Placidochromis longimanus*) were placed outside of the "sand-dwelling" non-mbuna hypertrophied lip clade, corroborating what has been reported in previous studies (Hulsey et al. 2018; Malinsky et al. 2018). *Lethrinops*, *Mylochromis*, and *Otopharynx* are also polyphyletic. Further, while IQ-TREE found a well-supported, distinct grouping of *Copadichromis* species with *Nyassachromis prostoma* nested among them, SVDquartets failed to recover any *Copadichromis* species together. A previous phylogenetic study by Hulsey et al. (2018) based on SNPs extracted from UCEs recovered a topology that does not strongly contradict that reported here for hypertrophied lip species. This study however covered a narrower taxonomic sampling compared to that of our present investigation and in general provided much weaker topological support. The linkage among SNPs and whether they were protein-coding or non-coding was also not assessed for these UCEs.

Comparisons across genomic regions

Given that coding regions are believed in general to carry less phylogenetic signal than faster-evolving non-coding regions among closely related lineages, we would expect coding SNPs to yield more poorly resolved trees (lower branch/node support values) for the Malawi radiation than those from non-coding regions. Additionally, we might expect non-

coding SNPs to provide better resolution among more recently diverged lineages, especially between populations or species of Malawi cichlids that have diverged so recently (Figure 4). Thinning the number of non-coding SNPs to the same size data set as the coding DNA allowed us to evaluate whether these two data types provided fundamentally different phylogenetic signals or simply represent different sized data partitions of the genome. Despite these caveats, the distributions of node support derived from our IQ-TREE and SVDquartet analyses were largely similar when considering coding and non-coding datasets of equal size (~50K SNPs) (Figure 4). Pairwise t-tests indicated no significant difference between the mean support values derived from these datasets. Furthermore, at both deep (≥10 nested terminals) and shallow (<10 nested terminals) nodes, the number and distribution of support values were comparable between these two data types. However, we did observe significant differences in resolution when comparing either of these smaller datasets to the entire non-coding SNP dataset (1,190,719 SNPs). In all, the much larger complete non-coding dataset provided better resolved trees with fewer ambiguities. As coding SNPs appear to be as informative as non-coding site per site, incorporating SNPs from transcriptomic data for instance could contribute to the resolution of many rapidly radiating lineages including these Lake Malawi cichlids. However, the much greater amount of phylogenetically informative sites obtained from non-coding SNPs suggests that whole genome resequencing might be a more powerful approach for fully resolving Malawi cichlid relationships as well of those of other large radiations. This approach is particularly useful for phylogenetic investigation in groups for which high-quality reference genomes exist and such resources are becoming more readily available, especially among vertebrates (Brawand et al. 2014; Conte et al. 2019; Kautt et al. 2020). Natural selection can influence the retention of standing genetic variation in both coding and non-coding sequences and hybridization can also drastically alter the genetic substrate exposed to selective forces (Seehausen 2004). Depending on which of these regions happen to be more greatly affected, discordance in tree topology between loci may arise and ultimately impact our ability to resolve clades. Given this caveat, it is likely advantageous to reconstruct phylogenies using both types of sequence data whenever possible.

Gene flow among hypertrophied lip Malawi cichlids

Because hybridization might readily explain the appearance of hypertrophied lips in both mbuna and non-mbuna clades, we conducted tests for hybridization across the Lake Malawi fauna with Dsuite (Malinsky et al. 2021). This program applies ABBA-BABA tests to biallelic SNPs using sets of four taxa and assumes a pectinate tree topology usually denoted as (((P1,P2),P3),O). The outgroup (O) is used to define the ancestral allele (A) from the derived allele (B) and site patterns (i.e., BBAA, ABBA, and BABA) are counted across SNPs.

Under the null model where only ILS is present (i.e., no gene flow, *D*=0), ABBA-BABA patterns are expected to occur in equal frequency, but a significant divergence from this indicates potential introgression between P3 and either P1 or P2 (Malinsky et al. 2021). Across the entire dataset many (x48,869) significantly different (*p*-value <0.001) ABBA-BABA patterns were found and the *D*-statistic values of these ranged from 0.2258–0.0071 (Supplementary Table 2, Supplementary Figure 3). Similarly, Malinsky et al. (2018) also inferred numerous instances of gene flow within the Malawi radiation through the analysis of SNP data.

Yet, while introgression appears to be rampant across Lake Malawi species, *Abactochromis* and the hypertrophied lip non-mbuna do not show elevated patterns of introgression with each other relative to other taxa. Nevertheless, of the 1,322 statistically significant trios involving potential hybridization with *A. labrosus*, 81 were suggestive of gene flow between *A. labrosus* and several hypertrophied lip non-mbuna species (Supplementary Table 2). However, of these, only three instances are represented among the highest 100 *D*-statistic scores calculated across all taxa. In general, the strongest evidence for gene flow was found to have occurred between the hypertrophied lip taxa and other Lake Malawi species such as *A. calliptera*. (Supplementary Figure. 3).

Comparisons with other lake radiations and future directions

Most instances of convergent evolution have occurred in allopatric settings, but our results here add to the growing number of cases of replicate evolution that have arisen among closely related lineages inhabiting the same body of water (Muschick et al. 2012; Hulsey et al. 2019). Our results also highlight that Lake Malawi harbors multiple hypertrophied lip species while the other East African Great Lakes, Victoria and Tanganyika, each have only one known species (Fryer and Iles 1972; Konings 2007; Oliver and Arnegard 2010). This disparity in species number may be due in part to Lake Malawi containing two distinct major radiations (the mbuna and non-mbuna) as well as the fact that the many isolated rocky reefs around the lake offer more extensive opportunities for adaptation and acquisition of novel phenotypes. Additionally, Lake Victoria, the youngest of the three lakes, has experienced the collapse of incipient species into hybrid swarms due to the loss of environmental heterogeneity and clear water habitats (Seehausen et al. 1997; Meier et al. 2017). This rampant hybridization could explain why hypertrophied lip taxa have not evolved, or at least been inferred to have evolved, repeatedly or speciated as seems to be the case for Lake Malawi. In contrast, given the relatively ancient age of Lake Tanganyika compared to Lakes Malawi and Victoria, additional hypertrophied lip species may have gone extinct resulting in the single widespread species found today (*Lobochilotes labiatus*). Because of

the much longer timeframe over which the Tanganyikan radiation has evolved (Friedman et al. 2013; Irisarri et al. 2018), reduced gene flow among the more ancient lineages in Lake Tanganyika may have also played a role in limiting new hypertrophied lip species from emerging across the phylogeny. Additionally, it is possible that the basic genetic architecture underlying the hypertrophied lip phenotype is significantly different among cichlid lineages and even among these three lakes (Henning et al. 2017; Kautt et al. 2020). Adaptive introgression influencing the presence of enlarged lips, as a result, could be less likely depending on this trait's (yet unknown) genetic complexity.

Future studies of the hypertrophied lip evolution in Lake Malawi cichlids should focus more extensively on the genomic basis of this trait. These enlarged lips could be highly polygenic or they could result from changes in only one or two loci (Colombo et al. 2013; Manousaki et al. 2013; Henning et al. 2017; Kautt et al. 2020). The exact genomic architecture of the hypertrophied lips in Malawi might influence how likely a single hybridization event could have contributed to repeated evolution of this trait across disparate phylogenetic lineages of Malawi cichlids (Riley et al. 2003; Mallet 2005; Taylor and Larson 2019). For instance, unlike traits that have been found to have a more polygenic basis such as body shape and pharyngeal jaw morphology, GWA mapping of lip size in Central American Midas cichlids has revealed high genomic associations at only two loci of major effect and the potential for introgression across multiple lineages (Kautt et al. 2020). A simpler genetic basis to lip size might generally allow it to readily introgress (Mallet 2005; Taylor and Larson 2019). While introgression can play an integral role in shaping the genomic architecture of hybridizing lineages in adaptive radiations (Edelman et al. 2019), our current analyses provide no evidence that extensive interspecific introgression in Lake Malawi cichlids has contributed to the within-lake convergent evolution of hypertrophied lips.

Conclusion

Within the Lake Malawi cichlid radiation, the evolution of hypertrophied lips has occurred multiple times. Even genome-wide protein-coding SNPs exhibited power to robustly reconstruct the relationships among Lake Malawi cichlids, but the much larger non-coding SNP dataset provided better resolved inferences of relationships even within closely related Lake Malawi species. Both coding and non-coding phylogenomic reconstructions supported the monophyly of intraspecific sampling of several non-mbuna species with hypertrophied lips. These species also form a relatively closely related clade interspersed with a few other sand-dwelling non-hypertrophied lip lineages. Additionally, our phylogenomic and comparative analyses coupled with tests for introgression are most consistent with hypertrophied lips having evolved independently in the sand-dwelling non-mbuna and rock-dwelling mbuna Lake Malawi cichlids. Future, whole genome-wide inference based studies

of Malawi cichlid relationships will continue to shed greater light on the patterns and processes of phenotypic and molecular evolution in this rapidly evolving adaptive radiation.

Materials and Methods

Whole genome resequencing, mapping, and variant discovery

We generated new whole genome sequences for 33 individuals representing 24 cichlid species from Lake Malawi (Supplementary Table 1). Our samples included five of the eight known Lake Malawi species with hypertrophied lips: *Abactochromis labrosus*, *Cheilochromis euchilus*, *Eclectochromis ornatus*, *Placidochromis milomo*, and *Placidochromis* "Mbenji fatlip". High-molecular-weight DNA was extracted from fin or muscle tissue using a QIAGEN Dneasy Blood & Tissue Kit while including an RNase A treatment step. DNA integrity was verified on agarose gels and concentrations were determined on a QuBit fluorometer. Genomic libraries were prepared with Illumina TruSeq DNA Nano kits targeting insert sizes of 350-bp and then paired-end sequenced (2 x 150 bp) on Illumina HiSeq platforms at the Beijing Genome Institute. Four individuals were pooled per lane with the aim of generating an approximate genome coverage of 20x per individual. Our genomic dataset was supplemented with additional short-read WGS data for 60 individuals sequenced by Malinsky et al. (2018) and seven individuals by Scherz et al. (2022) (Supplementary Table 1).

Following demultiplexing, unmapped BAM files were generated from the raw FASTQs with Picard Tools v2.7.1 (*FastqToSam*) while marking Illumina adapters in the process (*MarkIlluminaAdapters*). Reads were then converted back to FASTQ format (*SamToFastq*) and mapped against the 22 chromosome assemblies of the latest version of the *Maylandia zebra* reference genome (GCA_000238955.5: M_zebra_UMD2a of Conte et al. 2019) using bwa -mem v0.7.17 (Li 2013). Metadata stored in the original unmapped BAM files were then added to the aligned BAM files using Picard *MergeBamAlignment* and PCR duplicates were annotated with Picard *MarkDuplicates*.

Variant discovery and genotype calling were preformed while considering all samples together using freebayes v1.3.1 (Garrison and Marth 2012) and implementing standard quality filters (a minimum mapping quality 30 and a minimum base quality of 20). To speed up variant calling, we ran freebayes in parallel over separate 1Mb regions spanning all 22 chromosomes and then concatenated the resulting VCFs into a single file with bcftools v1.3.1 *concat*. Hard quality filters were applied using the vcffilter script from the vcflib package (https://github.com/vcflib/vcflib) (command: -s -f "QUAL > 1 & QUAL / AO > 10 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1"). The vt tools *normalize* and *uniq* (Tan et al.

2015) were then applied to normalize variants and remove duplicates. Further variant filtering was conducted with bcftools v1.11 to set individual genotypes with depth ("DP") less than 10x or greater than 50x (approximately twice the raw mean depth per sample) and genotype quality ("GQ") less than 30 as missing ("./."). We also included only SNPs with minor alleles present more than once ("MAC>=2"), excluded any sites at which no alternate alleles remained after the filtering above ("AC==0") or where only alternate alleles were called ("AC==AN"), and removed possible false positive singletons by excluding sites with a minor allele frequency of less than or equal to 5% ("MAF<=0.05"). Lastly, we excluded sites with more than 20% missing data ("F_MISSING>0.2") and/or that were not biallelic SNPs ("-m2 -M2 -v snps"). This filtering scheme yielded a master VCF file containing 1,352,537 SNPs which was subsequently divided into separate datasets containing SNPs from non-coding and coding regions based on CDS annotations in the *M. zebra* reference genome (Conte et al. 2019). SNPs in these parsed datasets were further filtered based on linkage disequilibrium (LD) with bcftools 1.11-88 (Li et al. 2009; Danecek et al. 2021) using the +prune plugin. For each dataset, the squared correlation (r^2) between alleles at each pair of loci within windows of 500 kb (+prune parameter --window 500,000) was calculated and highly linked SNPs ($r^2 > 0.9$; +prune parameter --max 0.9) were discarded. This reduced the size of the non-coding and coding datasets to 1,190,719 and 54,021 SNPs, respectively, which were lastly converted into NEXUS and PHYLIP formatted alignments with IUPAC ambiguity codes applied to heterozygous sites using the python script vcf2phylip.py (Ortiz 2019) (https://github.com/edgardomortiz/vcf2phylip).

Phylogenomic analysis

To assess which hypertrophied lip species of cichlids in Lake Malawi compose monophyletic groups, phylogenetic analyses were conducted on the non-coding and coding datasets with *Astatotilapia bloyeti* (a non-Lake Malawi species) designated as an outgroup. Maximum likelihood trees were inferred from the SNP datasets using IQ-TREE v1.6.12 (Nguyen et al. 2015). For this analysis, the ascertainment bias correction was applied to correct for the absence of invariant sites in the sequence alignment (command -m MFP+ASC) (Kalyaanamoorthy et al. 2017) and 1,000 ultrafast bootstrap replicates (command -bb 1000; UFBoot Minh et al. 2013) were preformed to assess branch support.

To augment our inferences of relationships, we also used SVDquartets (Chifman and Kubatko 2014) as implemented in PAUP* v4.0a166 (Swofford 2002) to reconstruct the species tree from the non-coding and coding datasets under its coalescent-based framework. Unlike summary methods that rely on *a priori* reconstructed gene trees to

estimate the species tree, this program uses sequence data directly to infer quartet trees and performs well even in the presence of gene flow (Long and Kubatko 2018). For our analysis, the multispecies coalescent tree model was selected and individual samples were each assigned to a taxon partition (i.e., their respective species). We exhaustively sampled all quartets (3,751,519 in total) and inferred the species tree using the Quartet FM algorithm (QFM; Reaz et al. 2014). Subsequently, one hundred bootstrap replicates for each data partition were carried out to assess branch support.

After obtaining reconstructions from these analyses, we compared the resolving power of coding versus non-coding SNPs at various nodal depths. Ultrafast bootstrap values from IQ-TREE and bootstrap support values from SVDquartet were compared across the analyses conducted with the coding and non-coding datasets. For a fair comparison between the contrasting datasets, we thinned the VCF file containing only the non-coding SNPs to a size comparable to that of the coding dataset (~x50k SNPs) using the bcftools +prune plugin. For these comparisons, nodes that have ten or more nested terminal branches were referred to as “deep” and those with fewer than ten tips as “shallow”. Significance was assessed using pairwise t-tests of mean support values (α=0.05) as implemented by the 'stat_compare_means' function of the R package *ggpubr* (Kassambara 2020).

Maximum likelihood reconstruction of ancestral states was performed using the 'fastAnc' function of RStudio package *phytools* (Revell 2012). The presence and absence of hypertrophied lips was categorized as a discreet variable and the transition probabilities between these two states were considered to be equal. The probabilities of lineages possessing either hypertrophied lips or not at each node for the non-coding SNP-based tree derived with IQ-TREE are displayed in Supplementary Figure 2 and are represented for relevant nodes as pie-diagrams in Figure 2.

Analysis of gene flow

Because we found hypertrophied lip cichlid species of Lake Malawi to fall within both the mbuna and non-mbuna radiations, we investigated their genomic histories further. To assess the degree of interspecific gene flow across all ingroup taxa including that between hypertrophied lip species, we calculated genome-wide Patterson’s *D* (ABBA/BABA) statistics as implemented in the program Dsuite (Malinsky et al. 2021). This test is applied to biallelic SNPs across four taxa and assumes a pectinate tree topology ordered as (((P1,P2),P3),O). The outgroup (O) helps to define the ancestral allele (A) from the derived allele (B) and site patterns (BBAA, ABBA, and BABA) for each SNP are counted. Under the null model where only ILS is present (i.e., no gene flow, *D*=0), ABBA-BABA patterns are expected to occur

with equal frequency, but a significant divergence from this indicates that introgression may have happened between P3 and either P1 or P2 (Malinsky et al. 2021). Using the 1,352,537 SNPs from the master VCF file and *Astatotilapia bloyeti* set as the outgroup, we assessed all possible three taxon combinations (102,340 in total) with the 'Dtrios' function. Each trio was ordered so that the BBAA pattern was maximized in the output. Standard jackknife blocks (x20) were used to determine if the resulting *D*-statistic values differed significantly from zero. To account for multiple tests, *p*-values were adjusted in RStudio 4.0.3 by applying the false discovery rate method of Benjamini and Hochberg (1995) with the *stats* package (command: p.adjust(p_values, method = "fdr")). An α of 0.001 was applied to identify statistically significant *D*-statistic values. To visualize species pairwise comparisons of *D*-statistic scores, a heatmap was generated using the Ruby script plot_d.rb (available at: https://github.com/mmatschiner).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG 447185000) to [CDH & AM]. Samples were collected in 2010 by CDH with a permit issued by the Malawi Parks Service. We thank members of the Meyer Lab for their helpful comments and suggestions pertaining to this research.

Author Contributions

C.D.H. conceived the project and collected the samples. P.M. conducted the data processing and analyses. P.M. and C.D.H. drafted the manuscript. All authors read, revised, and approved the final manuscript.

Data Availability

Raw whole genome resequencing reads for the 33 newly sequenced taxa have been deposited in NCBI's Sequence Read Archive (See Supplementary Table 1). Tree files obtained from our analyses are available as supplementary material.

References

Arnegard ME, Snoeks J. 2001. New three-spotted cichlid species with hypertrophied lips (Teleostei: Cichlidae) from the deep waters of Lake Malaŵi/Nyasa, Africa. *Copeia* 2001:705–717.

Baumgarten L, Machado-Schiaffino G, Henning F, Meyer A. 2015. What big lips are good for: On the adaptive function of repeatedly evolved hypertrophied lips of cichlid fishes. *Biological Journal of the Linnean Society* 115:448–455.

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)* 57:289–300.

Bonett RM, Ledbetter NM, Hess AJ, Herrboldt MA, Denoël M. 2021. Repeated ecological and life cycle transitions make salamanders an ideal model for evolution and development. *Developmental Dynamics*:1–16.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–381.

Burress ED, Duarte A, Serra WS, Loueiro M, Gangloff MM, Siefferman L. 2013. Functional diversification within a predatory species. *PLoS ONE* 8.

Chen MY, Liang D, Zhang P. 2017. Phylogenomic resolution of the phylogeny of laurasiatherian mammals: Exploring phylogenetic signals within coding and noncoding sequences. *Genome Biology and Evolution* 9:1998–2012.

Cheon S, Zhang J, Park C. 2020. Is phylotranscriptomics as reliable as phylogenomics? *Molecular Biology and Evolution* 37:3672–3683.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30:3317–3324.

Chojnowski JL, Kimball RT, Braun EL. 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene* 410:89–96.

Coghill LM, Hulsey CD, Chaves-Campos J, García de Leon FJ, Johnson SG. 2014. Next generation phylogeography of cave and surface *Astyanax mexicanus*. *Molecular Phylogenetics and Evolution* 79:368–374.

Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, Barluenga M, Salzburger W. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology* 22:670–684.

Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerdinger WJ, Roberts RB, Carleton KL, Lien S, Kocher TD. 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* 8.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10.

Danley PD, Kocher TD. 2001. Speciation in rapidly diverging systems: Lessons from Lake Malawi. *Molecular Ecology* 10:1075–1086.

Dierickx K, Hanssens M, Rusuwa B, Snoeks J. 2018. *Trematocranus pachychilus*, a new endemic cichlid from Lake Malawi (Teleostei, Cichlidae). *ZooKeys* 743:153–166.

Eccles D, Trewavas E. 1989. Malawian cichlid fishes: the classification of some haplochromine genera. Herten: Lake Fish Movies.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, Challis R *et al*. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594–599.

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* 5:243–250.

Elmer KR, Kusche H, Lehtonen TK, Meyer A. 2010. Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philosophical Transactions of the Royal Society B* 365:1763–1782.

Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology and Evolution* 26:298–306.

Foley NM, Thong VD, Soisook P, Goodman SM, Armstrong KN, Jacobs DS, Puechmaille SJ, Teeling EC. 2015. How and why overcome the impediments to resolution: Lessons from rhinolophid and hipposiderid bats. *Molecular Biology and Evolution* 32:313–333.

Friedman M, Keck BP, Dornburg A, Eytan RI, Martin CH, Hulsey CD, Wainwright PC, Near TJ. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proceedings of the Royal Society B: Biological Sciences* 280.

Froese R, Pauly D. 2021. FishBase. *World Wide Web electronic publication* [Internet]. Available from: www.fishbase.org

Fryer G, Iles TD. 1972. The cichlid fishes of the great lakes of Africa. Edinburgh: Oliver and Boyd.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Genner MJ, Ngatunga BP, Mzighani S, Smith A, Turner GF. 2015. Geographical ancestry of Lake Malawi's cichlid fish diversity. *Biology Letters* 11.

Genner MJ, Turner GF. 2005. The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries* 6:1–34.

Grant PR, Grant BR. 2008. How and why species multiply: The radiation of Darwin's finches. Princeton: Princeton University Press.

Greenwood PH. 1974. The cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. *Bulletin of the British Museum (Natural History)*:1–134.

Hashem, S., K. Kawai, P. S. K. Fatsi, A. Kodama, H. Saito. 2020. Genetic relationships of cichlid fishes from Lake Malawi based on mitochondrial DNA sequences. *Limnology* 21:151–163.

Henning F, Machado-Schiaffino G, Baumgarten L, Meyer A. 2017. Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. *Evolution* 71:1297–1312.

Henning F, Meyer A. 2014. The evolutionary genomics of cichlid fishes: Explosive speciation and adaptation in the postgenomic Era. *Annual Review of Genomics and Human Genetics* 15:417–441.

Hulsey CD, Alfaro ME, Zheng J, Meyer A, Holzman R. 2019. Pleiotropic jaw morphology links the evolution of mechanical modularity and functional feeding convergence in Lake Malawi cichlids. *Proceedings of the Royal Society B: Biological Sciences* 286.

Hulsey CD, Mims MC, Parnell NF, Streelman JT. 2010. Comparative rates of lower jaw diversification in cichlid adaptive radiations. *Journal of Evolutionary Biology* 23:1456–1467.

Hulsey CD, Roberts RJ, Loh YHE, Rupp MF, Streelman JT. 2013. Lake Malawi cichlid evolution along a benthic/limnetic axis. *Ecology and Evolution* 3:2262–2272.

Hulsey CD, Zheng J, Faircloth BC, Meyer A, Alfaro ME. 2017. Phylogenomic analysis of Lake Malawi cichlid fishes: Further evidence that the three-stage model of diversification does not fit. *Molecular Phylogenetics and Evolution* 114:40–48.

Hulsey CD, Zheng J, Holzman R, Alfaro ME, Olave M, Meyer A. 2018. Phylogenomics of a putatively convergent novelty: Did hypertrophied lips evolve once or repeatedly in Lake Malawi cichlid fishes? *BMC Evolutionary Biology* 18.

Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, Franchini P, Fischer C, Lemmon AR, Lemmon EM, Thallinger GG, Sturmbauer C, Meyer A. 2018. Phylogenomics uncovers hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications* 9:3159.

Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology* 21:526.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587–589.

Kassambara A. 2020. ggpubr: “ggplot2” Based Publication Ready Plots. *R package version 0.4.0*.

Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Härer A, Hulsey CD, Franchini P, et al. 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588:106–111.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR. 1993. Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetics and Evolution* 2:158–165.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF. 1995. Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Molecular Phylogenetics and Evolution* 4:420–432.

Konings A. 2007. Malawi cichlids in their natural habitat. 4th ed. El Paso: Oliver and Boyd.

Kratochwil CF, Liang Y, Gerwin J, Woltering JM, Urban S, Henning F, Machado-Schiaffino G, Hulsey CD, Meyer A. 2018. Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science* 362:457–460.

Lemmon EM, Lemmon AR. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44:99–121.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Long C, Kubatko L. 2018. The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology* 67:770–785.

Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279:2115–2118.

Machado-Schiaffino G, Henning F, Meyer A. 2014. Species-specific differences in adaptive phenotypic plasticity in an ecologically relevant trophic trait: Hypertrophic lips in Midas cichlid fishes. *Evolution* 68:2086–2091.

Machado-Schiaffino G, Kautt AF, Torres-Dowdall J, Baumgarten L, Henning F, Meyer A. 2017. Incipient speciation driven by hypertrophied lips in Midas cichlid fishes? *Molecular Ecology* 26:2348–2362.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 21:584–595.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution* 2:1940–1955.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* 20:229–237.

Manousaki T, Hull PM, Kusche H, MacHado-Schiaffino G, Franchini P, Harrod C, Elmer KR, Meyer A. 2013. Parsing parallel evolution: Ecological divergence and differential gene

expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology* 22:650–669.

Marques DA, Taylor JS, Jones FC, di Palma F, Kingsley DM, Reimchen TE. 2017. Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLoS Biology* 15.

Meier J, Marques D, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* 8:14363.

Meyer A. 1993. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution* 8:279–284.

Meyer A. 1994. Shortcomings of the cytochrome b gene as a molecular marker. *Trends in Ecology and Evolution* 9:278–280.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347:550–553.

Meyer A, Montero CM, Spreinat A. 1994. Evolutionary history of the cichlid fish species flocks of the East African great lakes inferred from molecular phylogenetic data. *Advances in Limnology* 44:409–425.

Meyer A, Montero CM, Spreinat A. 1996. Molecular phylogenetic inferences about the evolutionary history of the East African cichlid fish radiations. In: Johnson TC, Odada EO, Whittaker KT, editors. The limnology, climatology and paleoclimatology of the East African lakes. Toronto: Gordon and Breach Scientific Publishers. p. 303–323.

Meyer BS, Matschiner M, Salzburger W. 2017. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic Biology* 66:531–550.

Mims MC, Hulsey CD, Fitzpatrick BM, Streelman JT. 2010. Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Molecular Ecology* 19:940–951.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30:1188–1195.

Moran P, Kornfield I. 1993. Retention of an ancestral polymorphism in the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution* 10:1024–1029.

Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology* 22:2362–2368.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.

Oliver MK, Arnegard ME. 2010. A new genus for Melanochromis labrosus, a problematic Lake Malawi cichlid with hypertrophied lips (Teleostei: Cichlidae). *Ichthyol. Explor. Freshwaters* 21:209–232.

Omland KE. 1999. The assumptions and challenges of ancestral state reconstructions. *Systematic Biology* 48:604–611.

Ortiz E. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502:228–231.

Rahman SR, Cnaani J, Kinch LN, Grishin NV, Hines HM. 2021. A combined RAD-Seq and WGS approach reveals the genomic basis of yellow color variation in bumble bee *Bombus terrestris*. *Scientific reports* 11.

Reaz R, Bayzid MS, Rahman MS. 2014. Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. *PLoS ONE* 9.

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223.

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, Sharp BJ. 1983. A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *South African Journal of Zoology* 18:149–310.

Riley, SPD, Shaffer, HB, Voss, SR, Fitzpatrick, BM. 2003. Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications* 13:1263–1275.

Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, Indermaur A, Malinsky M, Ricci V, Kahmen A, Jentoft S. 2021. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature* 589:76–81.

Rundle HD, Nagel L, Boughman JW, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* 287:306–308.

Salisbury BA, Kim J. 2001. Ancestral state estimation and taxon sampling density. *Systematic Biology* 50:557–564.

Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology* 18:169–185.

Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology* 5.

Salzburger W, Meyer A. 2004. The species flocks of East African cichlid fishes: Recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* 91:277–290.

Scherz MD, Masonick P, Meyer A, Hulsey CD. 2022. Between a rock and a hard polytomy: phylogenomics of the rock-dwelling mbuna cichlids of Lake Malawi. *Systematic Biology* in press.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology and Evolution* 19:198–207.

Seehausen O, van Alphen JJM, Witte F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277:1808–1811.

Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution* 31.

Smith ML, Hahn MW. 2021. New approaches for inferring phylogenies in the presence of paralogs. *Trends in Genetics* 37.

Snoeks J. 2004. The cichlid diversity of Lake Malawi/Nyasa/Niassa: Identification, distribution and taxonomy. El Paso: Cichlid Press.

Stauffer JR, Bowers NJ, Kocher TD, Mckaye KR. 1996. Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malaŵi. *Copeia* 1:203–208.

Steinke D Salzburger W, Braasch I, Meyer A. 2006. Many genes in fish have species-specific asymmetric rates of molecular evolution. *BMC Genomics* 7:20.

Stern DL. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* 14.

Stiassny MLJ and Meyer A. 1999. Cichlids of the African Rift Lakes. *Scientific American* 280:64–69.

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

Swofford DL. 2002. PAUP* Phylogenetic Analysis Using Parsimony * (and other methods). version 4.0. *Sinauer Associates, Sunderland, Massachusetts*.

Tan A, Abecasis GR, Kang HM. 2015. Unified representation of genetic variants. *Bioinformatics* 31:2202–2204.

Taylor, SA, Larson, EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution* 3:170–177.

Thomson RC, Wang IJ, Johnson JR. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Molecular Ecology* 19:2184–2195.

Torres-Dowdall J, Meyer A. 2021. Sympatric and allopatric diversification in the adaptive radiations of Midas cichlids in Nicaraguan Lakes. In: Abate ME, Noakes DLG, editors. The Behavior, Ecology and Evolution of Cichlid Fishes. Fish & Fisheries Series, vol 40. Dordrecht: Springer. p. 175–216.

Townsend TM, Alegre RE, Kelley ST, Wiens JJ, Reeder TW. 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Molecular Phylogenetics and Evolution* 47:129–142.

Trewavas E. 1935. A synopsis of the cichlid fishes of Lake Nyasa. *Annals and Magazine of Natural History* 16:65–118.

Wright NA, Steadman DW, Witt CC, Estes JA. 2016. Predictable evolution toward flightlessness in volant island birds. *PNAS* 113:2021.

Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31:3081–3092.

Yu L, Luan PT, Jin W, Ryder OA, Chemnick LG, Davis HA, Zhang YP. 2011. Phylogenetic utility of nuclear introns in interfamilial relationships of Caniformia (order Carnivora). *Systematic Biology* 60:175–187.

Figure Legends

Figure 1.

Phylogenetic reconstruction and our ability to infer convergence could be impacted heavily by genomic introgression. Illustrated are three hypothetical scenarios showing the extent that introgression could affect tree inference in the Lake Malawi cichlid radiation. In each panel, the true evolutionary history (the topology of which remains the same across A–C) is depicted on the left and the resulting phylogenetic reconstruction influenced by different degrees of hybridization is shown on the right. Taxa possessing hypertrophied lips are indicated with stars. Red lines denote prior hybridization events and blue bars represent possible origins of the hypertrophied lip phenotype. Incongruence between the true history and the reconstructed tree is shown with gray lines. In (A), phylogenetic resolution is completely obscured by factors such as widespread hybridization and incomplete lineage sorting resulting in the failure to recover conspecifics as monophyletic groups. Convergence of the hypertrophied lip phenotype cannot be tested with bifurcating phylogeny-based comparative methods under this scenario. In (B), some phylogenetic structure is detectable, but clarity is insufficient to resolve many relationships among hypertrophied lip species. Convergence is difficult to examine due to introgression and the inability to recover conspecifics as monophyletic. (C) Despite some gene flow, phylogenetic structure is clearly resolved, conspecifics are recovered as monophyletic, and there is a clear indication that hypertrophied lips are independently derived in disparate lineages.

Figure 2.

Species tree reconstruction of cichlids from African Lake Malawi based on maximum likelihood analysis in IQ-TREE of 1,107,249 non-coding SNPs. Ultrafast bootstrap support values are displayed for each branch. Taxa possessing hypertrophied lips, denoted in bold, are recovered in largely two positions in the phylogeny. Notably, intraspecific sampling of these species recovered monophyletic groupings. Pie-diagrams represent the probability that the ancestral condition for select nodes was hypertrophied lips (indicated in blue). Node A: MRCA of all Lake Malawi hypertrophied species, node B: MRCA of the rock-dwelling mbuna (green shading) radiation, node C: MRCA of the sand-dwelling non-mbuna (sand shading), node D: MCRA of the non-mbuna hypertrophied lip fauna. The five hypertrophied Lake Malawi species sampled are pictured to the right of the phylogeny.

Figure 3.

Relationships of Lake Malawi species based on analysis with SVDquartets. The reconstructions based on non-coding versus coding SNPs are compared. Dotted lines are used to connect species that exhibit slightly divergent phylogenetic relationships based on the two data types. Bootstrap values are indicated for each branch. Node A: MRCA of all Lake Malawi hypertrophied species, node B: MRCA of the rock-dwelling mbuna radiation (green shading), node C: MRCA of the sand-dwelling haplochromine non-mbuna (sand shading), node D: MCRA of the non-mbuna hypertrophied lip fauna.

Figure 4.

Expectations based on the literature of the resolving power of coding (cds) versus unthinned and thinned non-coding (non-cds) SNPs across different nodal depths are depicted. The entire non-coding dataset contained ~1.19 million SNPs. The thinned non-coding dataset was down-sampled to the same size as the coding dataset (~50k SNPs). The different datasets are color-coded in all panels (cds = red, unthinned non-cds = blue, and thinned cds = purple). In the left panels, the y-axis corresponds to ultrafast bootstrap values from the IQ-TREE analyses, and in the right panels, bootstrap support values from the SVDquartet analyses are shown with boxplots. Deep nodes refer to those in our phylogenies that have 10 or more nested terminal branches and shallow nodes are those that join fewer than ten tips. Ranges of support are also displayed exclusively for nodes within the clade of hypertrophied lip non-mbuna (see Node D Figures 1 and 2). Significance was assessed using pairwise t-tests of mean support values (α=0.05).

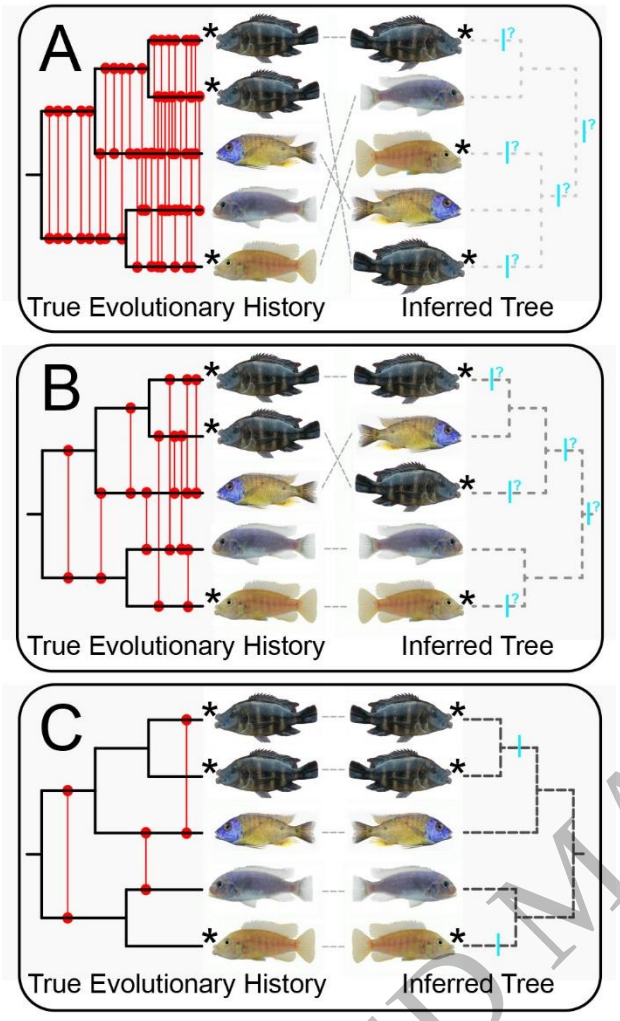

Figure 1.

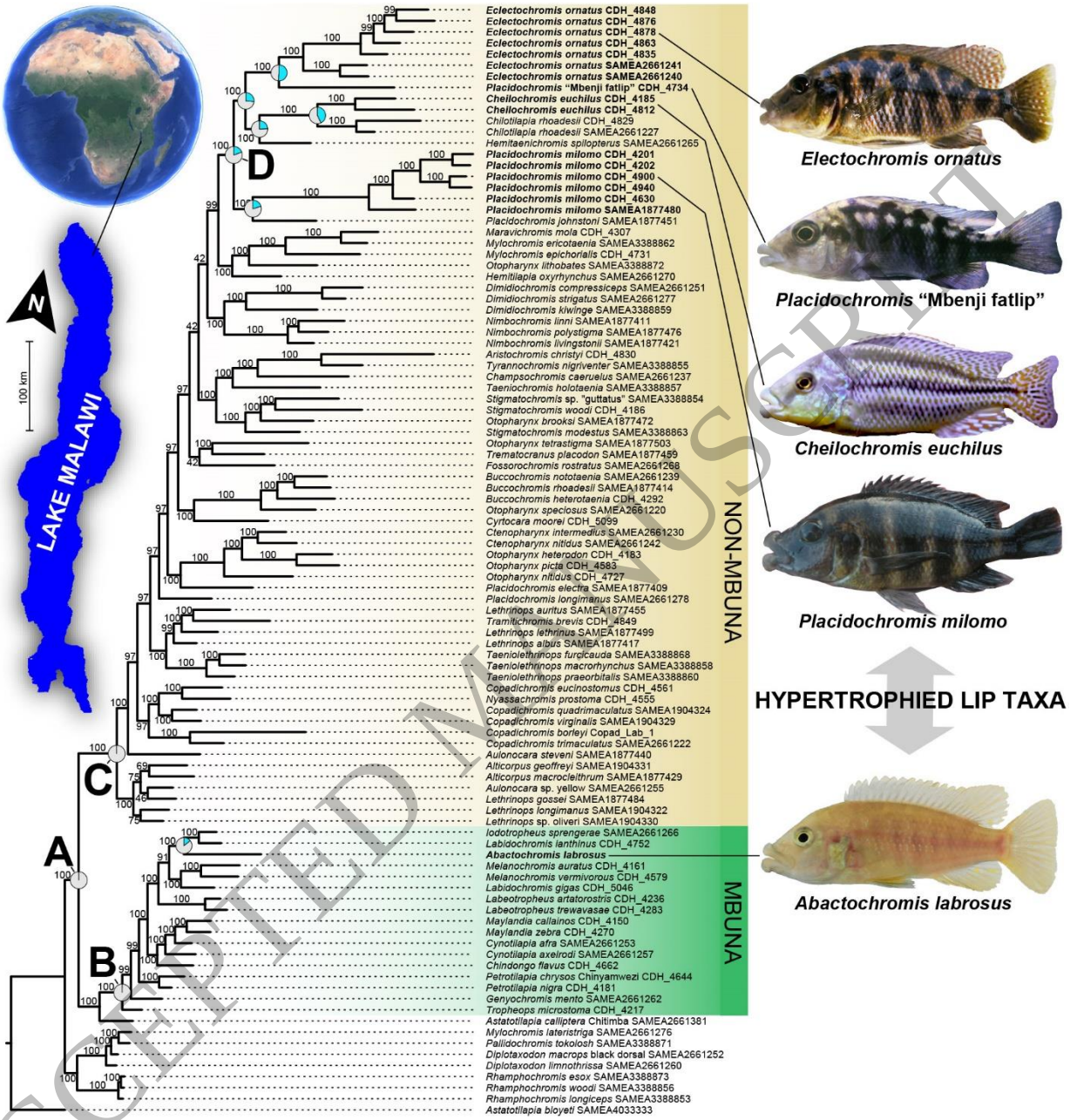

Figure 2.

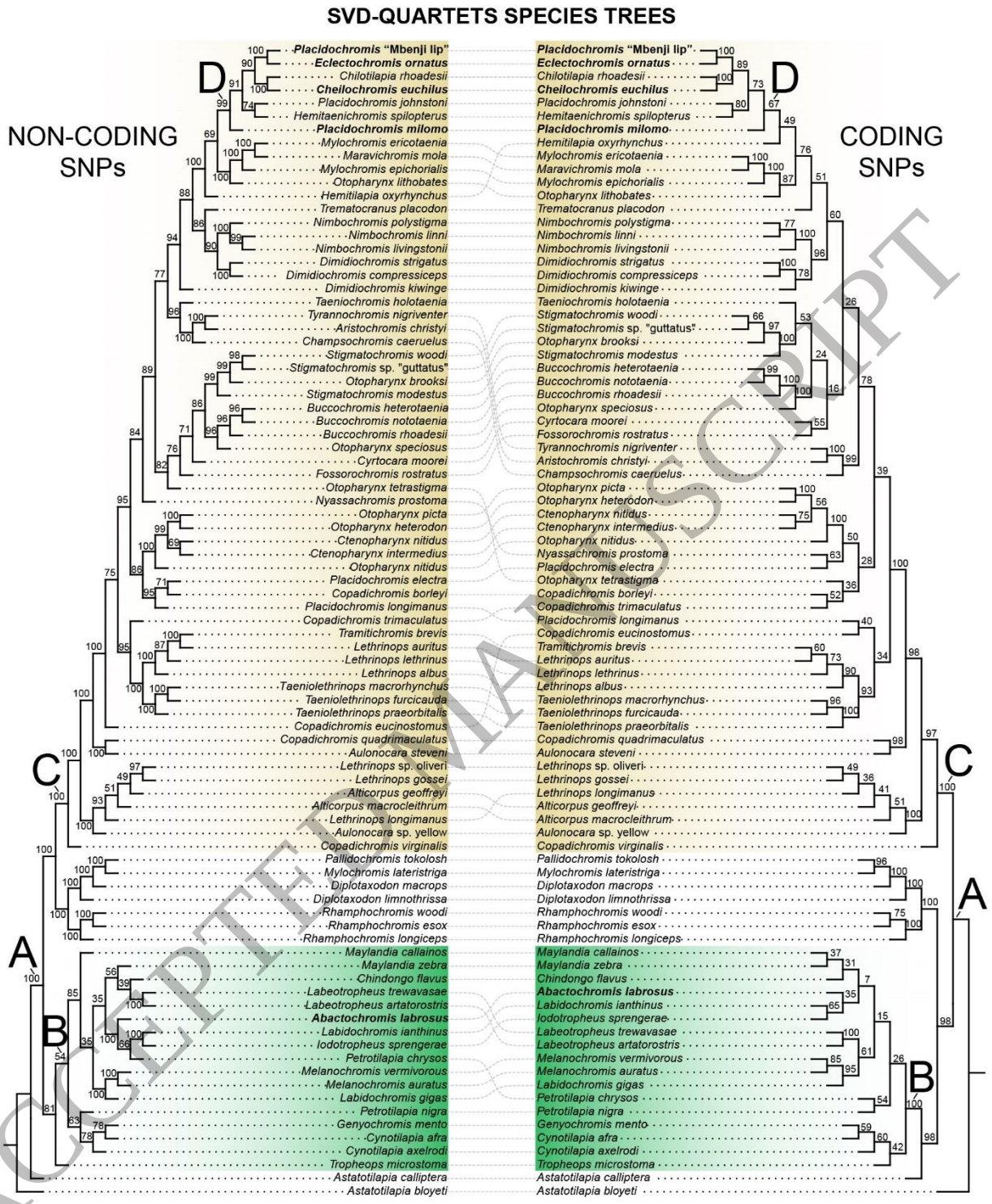

Figure 3.

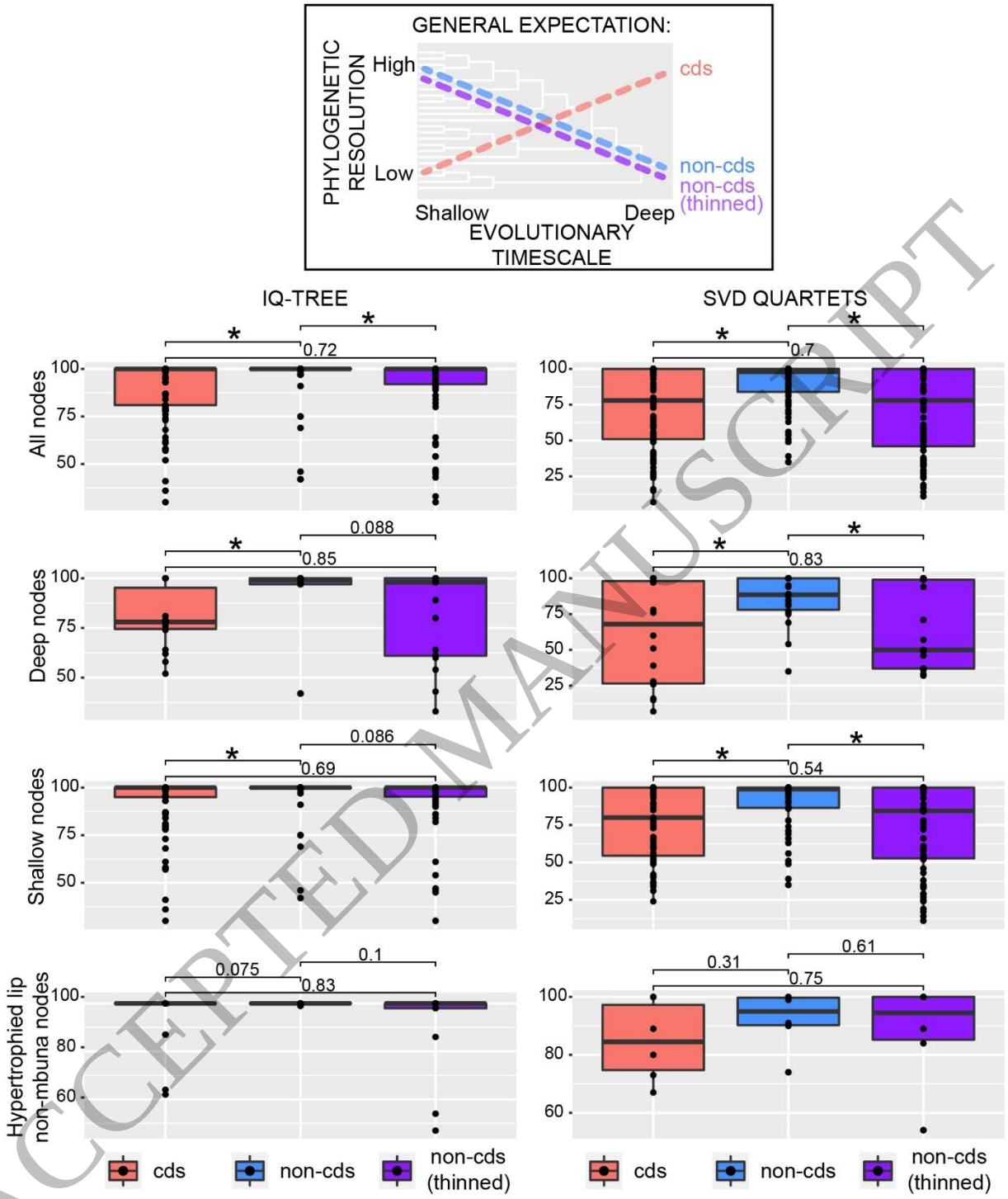

Figure 4.