ORIGINAL ARTICLE

Of bars and stripes: A Malawi cichlid hybrid cross provides insights into genetic modularity and evolution of modifier loci underlying colour pattern diversification

Jan Gerwin[1] 🄳 | Sabine Urban[1] | Axel Meyer[1] | Claudius F. Kratochwil[1,2] 🄳

[1]Department of Biology, University of Konstanz, Konstanz, Germany

[2]Institute of Biotechnology, HiLIFE, Helsinki, Finland

Correspondence
Claudius F. Kratochwil and Axel Meyer, Department of Biology, University of Konstanz, Konstanz, Germany.
Emails: claudius.kratochwil@helsinki.fi; axel.meyer@uni-konstanz.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 290977748 and 423396155; H2020 European Research Council, Grant/Award Number: 297300; Hector Fellow Academy, Grant/Award Number: 3000517; University of Konstanz

Abstract

Understanding the origins of phenotypic diversity among closely related species remains an important largely unsolved question in evolutionary biology. With over 800 species, Lake Malawi haplochromine cichlid fishes are a prominent example of extremely fast evolution of diversity including variation in colouration. Previously, a single major effect gene, *agrp2* (*asip2b*), has been linked to evolutionary losses and gains of horizontal stripe patterns in cichlids, but it remains unknown what causes more fine-scale variation in the number and continuity of the stripes. Also, the genetic basis of the most common colour pattern in African cichlids, vertical bars, and potential interactions between the two colour patterns remain unknown. Based on a hybrid cross of the horizontally striped Lake Malawi cichlid *Pseudotropheus cyaneorhabdos* and the vertically barred species *Chindongo demasoni* we investigated the genetic basis of both colour patterns. The distribution of phenotypes in the F_2 generation of the cross indicates that horizontal stripes and vertical bars are independently inherited patterns that are caused by two sets of genetic modules. While horizontal stripes are largely controlled by few major effect loci, vertical bars are a highly polygenic trait. Horizontal stripes show substantial variation in the F_2 generation that, interestingly, resemble naturally occurring phenotypes found in other Lake Malawi cichlid species. Quantitative trait loci (QTL) mapping of this cross reveals known (*agrp2*) and unknown loci underlying horizontal stripe patterns. These findings provide novel insights into the incremental fine-tuning of an adaptive trait that diversified through the evolution of additional modifier loci.

KEYWORDS
horizontal stripes, hybridzation, QTL mapping, standing genetic variation, transgressive segregation, vertical bars

1 | INTRODUCTION

Understanding the origin of phenotypic diversity during the formation of adaptive radiations remains a question that is only now being addressed with genomic and forward genetic approaches. Adaptive radiations are often characterized by shared genetic variation and ongoing hybridization that can both facilitate (Meier et al., 2017; Svardal et al., 2020) but in some cases also counteract morphological diversification (i.e., if species collapse; Mallet, 2007). The cichlid fishes inhabiting the East African Great Lake Malawi offer an example of an extremely young adaptive radiation (2–4 million years) with >800 genetically very similar (Malinsky et al., 2018; Meyer et al., 1990; Verheyen et al., 2003), but phenotypically highly diverse species. Axes of diversification include among others trophic morphology, body shape and colouration. Beyond diverse nuptial colourations, Lake Malawi cichlids greatly vary in melanic patterns (i.e., based on melanophores, melanin-bearing pigment cells; Figure 1; Hendrick et al., 2019; Liang, Gerwin, et al., 2020; Liang, Meyer, et al., 2020). The two most common colour patterns in African cichlids are vertical bars and horizontal stripes. Both patterns show meristic variation (i.e., number of stripes/bars) and differ in width, continuity, orientation and contrast. Horizontal stripe patterns consist of one or two distinct stripes, the dorsolateral stripe (DLS) and midlateral stripe (MLS) (Figure 1a) (Seehausen et al., 1999). Vertical bar patterns usually consist of six or more bands covering the distance between the dorsal fin and the ventral part of the body (Figure

1c, f). Although vertical bars and horizontal stripes were probably already present at the origin of the radiation (Barlow, 2000), there are a variety of combinations and modifications of these patterns. Therefore, one of the questions we wanted to address is how the diversity of melanic patterns in Lake Malawi cichlids (Figure 1) could have evolved within such a short evolutionary time.

Previous work suggests that the presence (or absence) of horizontal stripe patterns is controlled by regulatory alleles at a single major effect locus that affect *agouti related peptide 2* (*agrp2*) expression and thereby stripe absence (high *agrp2* expression) or presence (low *agrp2* expression) (Kratochwil et al., 2018). In Lake Victoria cichlids, the youngest major species flock that is less than 0.1 million years old (Meyer et al., 1990; Verheyen et al., 2003) stripes follow a Mendelian inheritance pattern. In contrast, a hybrid cross between a striped and a nonstriped Lake Malawi cichlid species shows that in Lake Malawi only around 50 percent of the variance is explained by this locus suggesting a more complex genetic basis (Kratochwil et al., 2018). Hence, it is possible that additional modifier loci were already present prior to or evolved within the older Lake Malawi radiation. These additional loci might then also explain the greater variation of stripe patterns found in Lake Malawi compared to the much younger Lake Victoria haplochromine cichlid radiation (Figure 1). Alternatively, diversity in colour patterns could be a result of reshuffling of ancestral standing genetic variation that entered Lake Malawi with the colonizing individuals and/or subsequent hybridization (Brawand et al., 2014; Svardal et al., 2020). The genetic
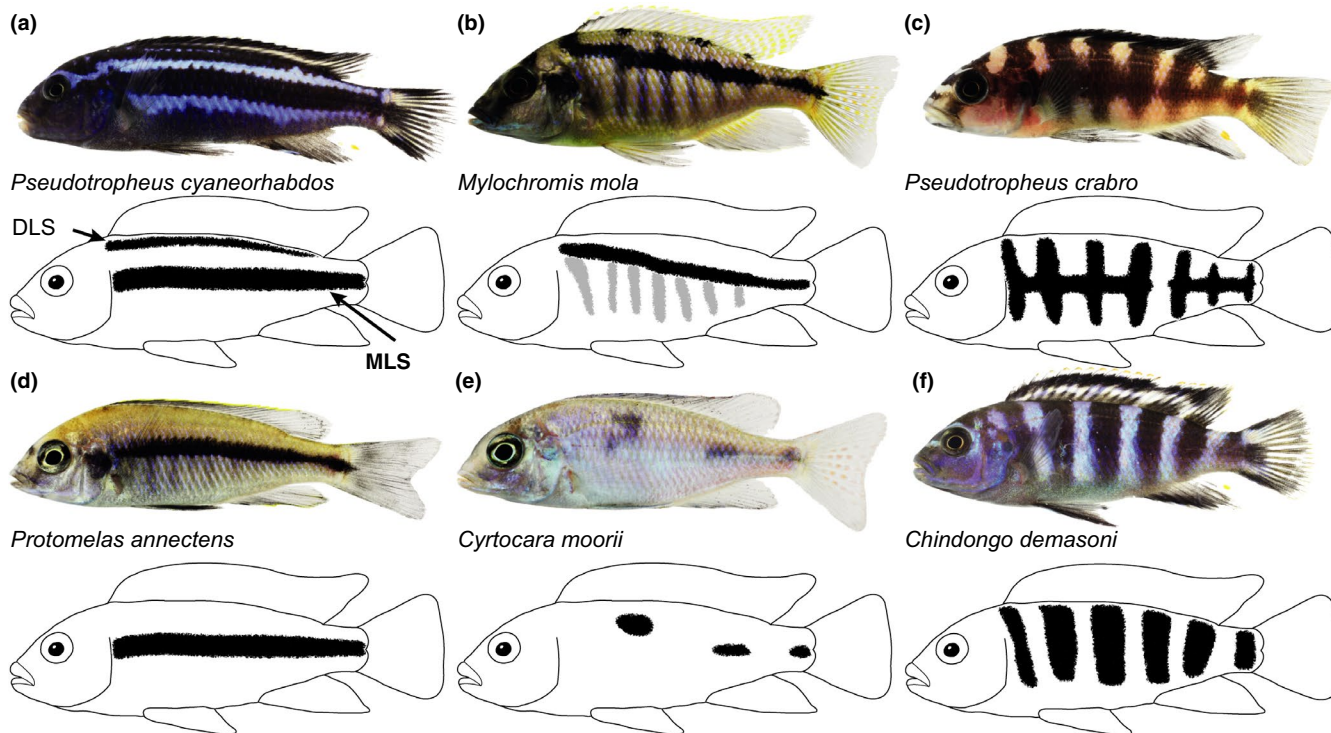

FIGURE 1 Diversity of melanic patterns in Lake Malawi haplochromine cichlid fishes. Each panel shows a photograph of a cichlid species from Lake Malawi and a schematic representation of its horizontal stripes and vertical bars. Stripes vary in several aspects including continuity (e.g., b and e), number (e.g., a and d), orientation (e.g., b and d), and contrast (e.g., b and e), bars mainly in number (e.g., c and f) and contrast (e.g., b and c). DLS, dorsolateral stripe; MLS, midlateral stripe

basis for vertical bar patterns in cichlids is still unknown, although we gained some initial insights into their development and the genes that might play a role in their development (Hendrick et al., 2019; Liang, Gerwin, et al., 2020; Liang, Meyer, et al., 2020).

A striking observation is the seemingly independent evolution of horizontal stripe and vertical bar patterns. The majority of cichlid fish species of the East African radiations (probably over 800 species in the three adaptive radiations of Lake Victoria, Malawi and Tanganyika and the riverine, nonendemic cichlids of East Africa) only displays either horizontal stripes (about one third of species) or vertical bars (most of the remaining two thirds of species). In the few species that show both colour patterns one of them is usually fainter than the other one (e.g., Figure 1b). This might be ultimately driven by a functional mismatch of the traits (as previously described for other traits, for example, Arnegard et al., 2014), as the two traits might constitute alternative and mutually exclusive strategies to hide from predators or prey (Seehausen et al., 1999) or mutually exclusive developmental trajectories. To better understand the evolution of these two traits in the haplochromine cichlid radiations of East Africa, here with a focus on Lake Malawi cichlids, it is important to examine how independent the two traits are in terms of their genetic basis and how this might have facilitated or constrained colour pattern divergence.

In an effort to gain novel insights into the genetic basis and evolution of stripe and bar patterns in cichlid fishes we performed a hybrid cross between the striped species *Pseudotropheus cyaneorhabdos* and the barred species *Chindongo demasoni* (previously: *Pseudotropheus demasoni*) from Lake Malawi (the same cross as used for single-locus association mapping in Kratochwil et al., 2018). One notable feature of cichlid fishes is that many can be bred easily in the laboratory, and that fertile crosses are possible among many species, in particular among the haplochromine cichlids of East Africa that make up the adaptive radiations of Lakes Malawi and Victoria. This permits breeding experiments such as the one we performed for this study. This cross allowed us to (i) identify the genetic architecture of both traits, (ii) to test if vertical bars and horizontal stripes are independently inherited traits, and (iii) to identify quantitative trait loci (QTL) that underlie the variation of stripe patterns found in Lake Malawi cichlid species.

2 | MATERIALS AND METHODS

2.1 | Experimental hybrid cross

Six *C. demasoni* females and one *P. cyaneorhabdos* male were set up in a 200 L aquarium until one of the females carried eggs in her mouth. The female was isolated in a separate tank until the offspring was released. The F_1 offspring was split into two identical 100 L tanks and raised until sexual maturity (between 6 and 9 months). The F_1 individuals were allowed to mate randomly to obtain individuals of the F_2 generation. Eggs were removed from the F_1 females after 6 to 10 days after fertilization and were reared in egg tumblers (Zissaqua

Zet-65). F_2 individuals were raised for 12 months in 100 L aquaria in groups of c. 20 individuals. Offspring of 33 broods was used in the final QTL analysis. We did not control for maternity nor paternity as the two pigmentation patterns of the parental species do not differ between the sexes. In this study we used individuals of the same hybrid cross as in Kratochwil et al. (2018) but at a later life stage that permitted a more refined phenotyping (12 months of age in this study compared to 3 months in the previous study).

2.2 | Photographs and visual phenotyping

F_2 individuals were photographed from the left and right side at the age of 12 months in a $25 \times 10 \times 5$ cm aquarium that was placed in a white photo shooting tent using a digital camera (Canon EOS 7D, 100 mm lens). The dorsolateral (DLS) and midlateral stripes (MLS) (Figure 1a) were analysed separately. Phenotyping of stripes was performed in three different ways. First, stripes were measured as a binary trait, where a completely or partially present stripe was scored as 1 and a completely missing stripe was scored as 0 (only used for Chi-square test). Second, stripes were measured as an ordinal phenotype. Here, stripes covering the whole anterior-posterior range of the potentially striped region were scored as 2, stripes that only covered parts of that region were scored as 1 and completely missing stripes were scored as 0 (only used to give an overview of the phenotypic distribution). Third, the proportion of the anterior-posterior range of the potentially striped region (the region that is striped in the striped parental species *P. cyaneorhabdos*) covered by melanic pigmentation was measured (used for QTL and Figure 2). Here, the stripes could potentially be scored in a range between 0% and 100%. To exclude the effect of vertical bars for the continuous stripe measurement, we measured both the potentially and actually striped regions between the vertical bars only (for a more detailed description see the supplementary text and Figure S1). Vertical bars were counted and classified into two categories: full bars (bars covering the whole dorsoventral range as seen in the parental species *C. demasoni*) and partial bars (bars not covering the whole range). Three different phenotype measures where obtained: number of full bars, number of partial bars and total bar number. All phenotypic measurements described above were scored on the left and right side. The average of both sides was used as the phenotype for QTL mapping.

2.3 | Phenotyping using patternize

Even though the measurement of stripe coverage takes into account the continuous distribution of stripe phenotypes in the F2 generation of our cross it is still limited regarding one important factor: it does not account for location-specific losses/gains of the stripe patterns. The R package patternize (Van Belleghem et al., 2018) enables extraction of colour patterns observed in photographs of different individuals of the F_2 generation and the parental species
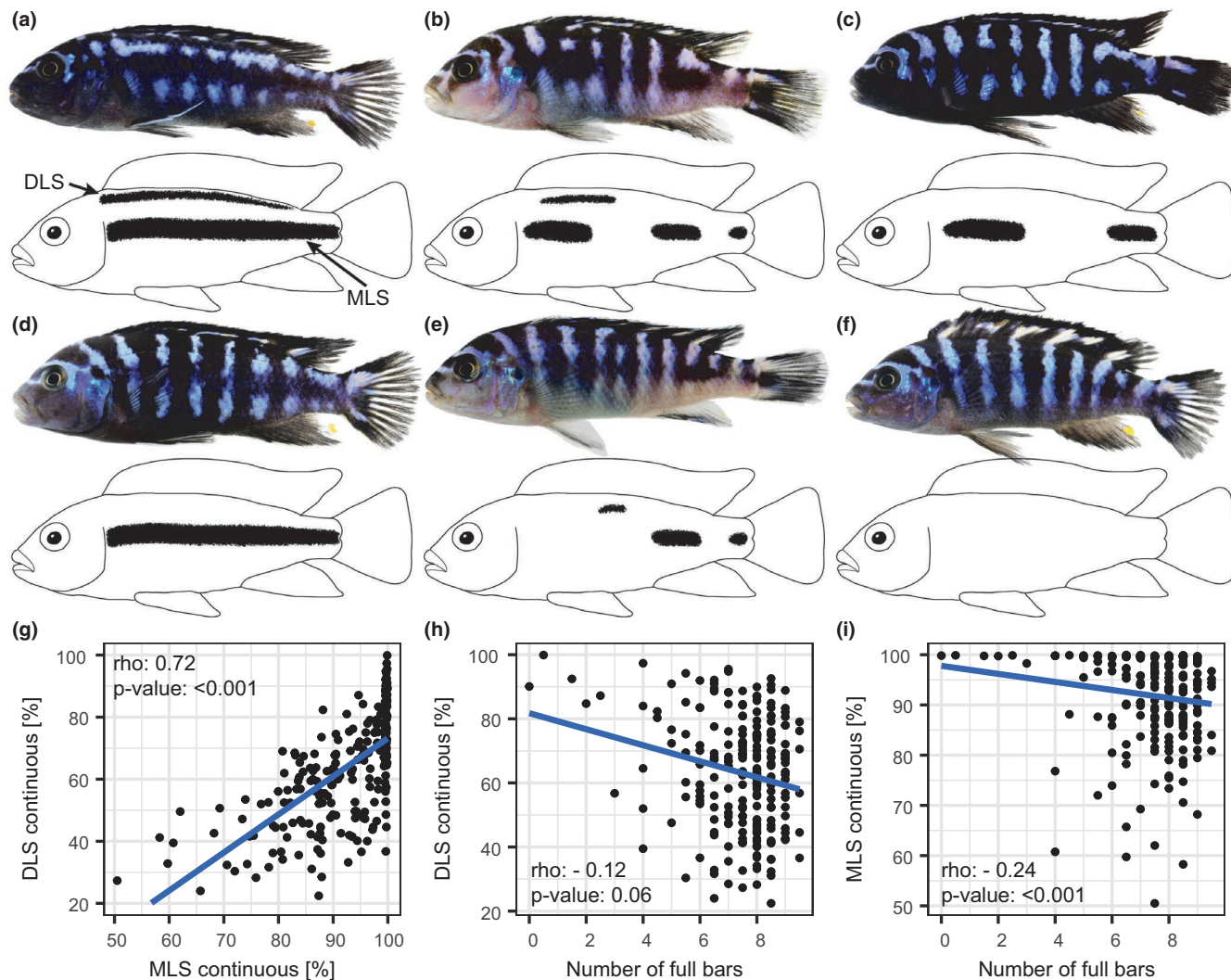
FIGURE 2 Horizontal stripes in the F_2 generation. Individuals of the F_2 generation show a range of expression of the horizontal stripe pattern: (a) complete pattern, (b, e) parts of the dorso- and midlateral stripe, (c) parts of the midlateral stripe, (d) midlateral stripe only, (f) no horizontal stripes. (g) The correlation of the continuous measurement of the DLS and the MLS. (h, i) The correlation between the number of full bars and the continuous measurements of the DLS and the MLS, respectively. The blue lines in panels g, h and i show linear regressions between the traits for visualization of the direction of the correlation. The results of Spearman's rank correlation test are shown in panels g, h and i. DLS, dorsolateral stripe; MLS, midlateral stripe

and to compare these colour patterns in a PCA. This analysis can be conducted for any given part of the colour pattern separately. Using the function patLanRGB the photographs were aligned according to fixed landmarks (Figure S1) previously determined on each photograph. Next, colour information was extracted from each photograph in windows of a predefined size (each of these windows represents a square of pixels). The colour information was translated into a binary matrix (1 = colour present, 0 = colour absent) that was then used for PC analysis. The colour information of all photographs used in our analysis was transformed to grey values (black and white photographs) using Adobe Photoshop CS6. Furthermore, the brightness levels of the photographs were adjusted using the levels tool in Photoshop. To correct for the different intensity of the colour pattern in different individuals, the darkest part of the colour pattern was set to be the darkest part of the photograph and the lightest

part of the fish to be the lightest part of the photograph. For our analysis, we created colour value matrices for each photograph of the left side of the F_2 individuals and for eight individuals of each of the parental species using the patLanRGB function (with parameters: RGB = c(0,0,0), colOffset = 0.35, res = 200). To create a morphospace of extreme values we first analysed the matrices of the parental individuals in a PCA (Figures S2 and S3). Using the R function predict we then projected the colour value matrices of the F_2 individuals into the parental morphospace. The resulting PC values were then used for QTL mapping.

For our analysis, we did not analyse the entire colour pattern, but limited the PC analysis to the colouration in our regions of interest. The maskOutline function allows the user to limit the analysis to a specific region. We wanted to specifically analyse the regions covered by the mid- and dorsolateral stripes and vertical

bars. Furthermore, we wanted to analyse different parts of the stripes separately. A major obstacle in the analysis was that horizontal stripes and vertical bars overlap in F_2 individuals that show both colour patterns. As the colour of stripes and bars is the same, it is not possible to reliably distinguish between black colouration contributed by a bar or by a stripe in the overlapping regions. For this reason, we limited the analysis with patternize to those regions where colour patterns in the parental species do not overlap. Consequently, the analysis for the presence and absence of bars was limited to the region between the mid- and dorsolateral stripe and the region ventrally of the midlateral stripe. The analysis of the mid- and dorsolateral stripe was only conducted in the regions between the vertical bars.

2.4 | DNA isolation and RAD sequencing

All F_2 individuals were fin clipped directly after taking the photographs for phenotyping. Fin clips were stored in 100% ethanol at –20°C until DNA isolation. About 0.25 cm² of fin tissue were used for DNA isolation using the Tissue DNA Purification Mini Spin Column Kit (Genaxxon Bioscience) following the manufacturer's instructions. Before the extraction, the tissue was incubated in TE buffer (500 mM Tris, 20 mM EDTA, 10 mM NaCl, pH 9.0) for 1 h at room temperature.

The isolated DNA was used to prepare double digest restriction site associated DNA (ddRAD) sequencing libraries (Franchini et al., 2017; Peterson et al., 2012). Briefly, double restriction enzyme digestion was performed on 1 µg of sample DNA using the enzymes PstI and MspI (New England Biolabs). Then P1 and P2 adapters were ligated to the digested DNA using T4 ligase (New England Biolabs). The Pippin Prep (Sage Science) was used to select DNA fragments of c. 450–550 bp and DNA libraries of 279 F_2 individuals were pooled in five sequencing pools. Each library pool contained one library of each of the parental individuals. Paired end sequencing (150 bp) was performed using the Illumina HiSeq X Ten platform (one library per lane) at the Beijing Genomics Institute, Hong Kong.

2.5 | Sequence processing, marker selection and linkage map construction

The sequences were processed using the STACKS pipeline (version 2.41) (Catchen et al., 2013). Briefly, PCR duplicates were removed using the clone_filter program. Using the inner and outer barcodes, the sequences were assigned to the corresponding individuals. Nonpaired reads were excluded from further analysis. Next, the reads were aligned to the *Maylandia zebra* genome (Conte & Kocher, 2015) using the Burrows-Wheeler Aligner (Li & Durbin, 2009) and sorted using samtools (Li et al., 2009). The program ref_map.pl (part of STACKS) was used for SNP calling and the populations program was run to generate population-level statistics.

Next, we used the program qtools.py (Kautt et al., 2020) for marker selection. Uninformative markers, markers missing in the parental samples and markers that are heterozygous in the parental samples were removed. The remaining 6,322 loci were further filtered and all loci that were present in less than 95% of the samples were removed.

For linkage map construction, the resulting file was imported into the program JoinMap 4 (Build: 4may07.4oct06), where identical loci and loci that significantly deviated from Hardy-Weinberg-Equilibrium were excluded from further analysis using the respective functions. The linkage map was constructed using the Haldane's mapping function under default settings.

2.6 | QTL mapping

QTL mapping was performed using the R package qtl (Broman et al., 2003). After loading the phenotype-genotype data frame using the function read.cross, the function jittermap was used to adjust the position of overlapping markers. Next, the functions calc.genoprob and sim.geno were used to calculate genotype probabilities (step = 0.1, error.prob = 0.00001) and to simulate genotypes between the observed markers (step = 0.1, n.draws = 64). QTL mapping was performed using the Haley-Knott method (Haley & Knott, 1992) in the function scanone. To identify significant QTL the genome wide significance threshold was calculated using the scanone function (n.perm = 1,000, 95th percentile LOD as threshold).

The function scanone works under the assumption of a single QTL and ignores the possibility of multiple QTL and possible epistatic effects between the loci. To find all loci associated with the phenotypes of interest and to investigate their potential interactions, we searched for further suggestive QTL by looking for those QTL only exceeding the chromosome wide threshold. Here, the significance threshold was calculated for each chromosome using the scanone function, but this time the desired chromosome was specified (n.perm = 1,000, chr = x, 90th percentile LOD as threshold to be sure to find all significant QTL). All identified QTL (exceeding genome and chromosome wide thresholds) were combined using the makeqtl function. The created QTL object was then analysed using the fitqtl function that calculates the percentage of variation explained (PVE) by the whole model and by each QTL. Nonsignificant QTL were dropped one by one and the analysis was repeated until all added QTL showed a significant effect in the multiple-QTL model.

QTL intervals were determined using the function bayesint of the R package qtl. The markers closest to the limits of the QTL intervals were used to identify a corresponding genomic region in the reference genome of the closely related species *Maylandia zebra* (Conte & Kocher, 2015). Names of known genes in the genomic regions were extracted using the biomart package in R. We identified genes that might be involved in colouration by filtering out those genes that had the terms "pigment" and "melan" as a part of their gene ontology (GO) description.

3 | RESULTS

3.1 | Independent inheritance of horizontal stripes and vertical bars

The first part of the study addresses the question if the two different colour patterns are inherited independently by conducting a hybrid cross. We analysed the presence of vertical bars and horizontal stripes in 230 adult individuals of the second offspring (F_2) generation of our hybrid cross between the striped species *Pseudotropheus cyaneorhabdos* and the barred species *Chindongo demasoni*. Both vertical bars and horizontal stripes can be found in individuals of the F_1 (Figure S4) and the F_2 generation (Figure 2). The average number of bars in the F_2 generation is ~8.5 bars (average between both sides of the fish) with a maximum of 10.5 bars and a minimum of one bar (Figure 3). Individuals of the barred parental species *C. demasoni* usually show either six or seven vertical bars. Not all bars in individuals of the F_2 generation fully cover the flank from dorsal to ventral as they do in *C. demasoni* (Figure 1f). Interestingly the inheritance of vertical bars significantly deviates from a 3:1 Mendelian ratio (Chi-square test, $\chi^2 = 76.6$, $df = 1$, $p < .001$) with all individuals showing some kind of bar pattern, suggesting a highly polygenic basis of the trait.

The distribution of horizontal stripes shows variation between the dorsolateral stripe (DLS, Figure 1a) and the midlateral stripe (MLS). The analysis of the ordinal measurements shows that 16.25% of the F_2 individuals lack the DLS completely, while only 1.25% of them do not show any parts of the MLS. Also, here both traits, DLS (Chi-square test, $\chi^2 = 87.7$, $df = 1$, $p < .001$) and MLS (Chi-square test, $\chi^2 = 560.7$, $df = 1$, $p < .001$) deviate from a 3:1 ratio that would have been expected if horizontal stripes had a simple Mendelian basis as shown before for the extremely young radiation of haplochromine cichlids of Lake Victoria (Henning et al., 2014). As stripes vary substantially in their continuity, we additionally scored the stripe phenotype as a continuous trait by calculating the proportion of the anterior-posterior range of the potentially striped region that is covered by melanic pigmentation (stripe coverage). To account for confounding effects of the vertical bars that cross the stripe region, we only measured the striped regions between the vertical bars. For the DLS, F_2 individuals showed a stripe coverage between 22% and 100%, for the MLS between 50% and 100% (Figure 2g–i).

To test for an association between the presence and absence of horizontal stripes and vertical bars we used the Fisher's exact test. We found no association between the presence of the MLS and bars (Fisher's exact test, $p = 1$) and between the presence of the DLS and bars (Fisher's exact test, $p = 1$). For a more detailed analyses we calculated Spearman's rank correlation coefficients to test for the correlation between the number of full bars and stripe coverage of the DLS and the MLS. We did not detect a significant correlation between the number of full bars and DLS ($rho = -0.12$, $p = .065$) (Figure 2h). In contrast, we found a significant negative correlation between the number of full bars and MLS ($rho = -0.24$, $p = .0002$) (Figure 2i). The weak correlation between the MLS and the number of full bars generally indicates an independent genetic basis of both phenotypes, but hints at the possible sharing of small effect loci. However, continuous measurements of DLS and MLS show a strong positive correlation ($rho = 0.72$, $p < .001$) (Figure 2g), possibly suggesting a shared genetic basis of the two horizontal stripes.

3.2 | QTL analysis of stripe coverage and bar number suggests multigenic basis of horizontal stripes and vertical bars

To further investigate the more complex genetic basis of stripe and bar patterns in Lake Malawi cichlids as well as to identify the loci associated with the traits, we performed QTL mapping. As we observed substantial variation in the phenotypes of both stripes we
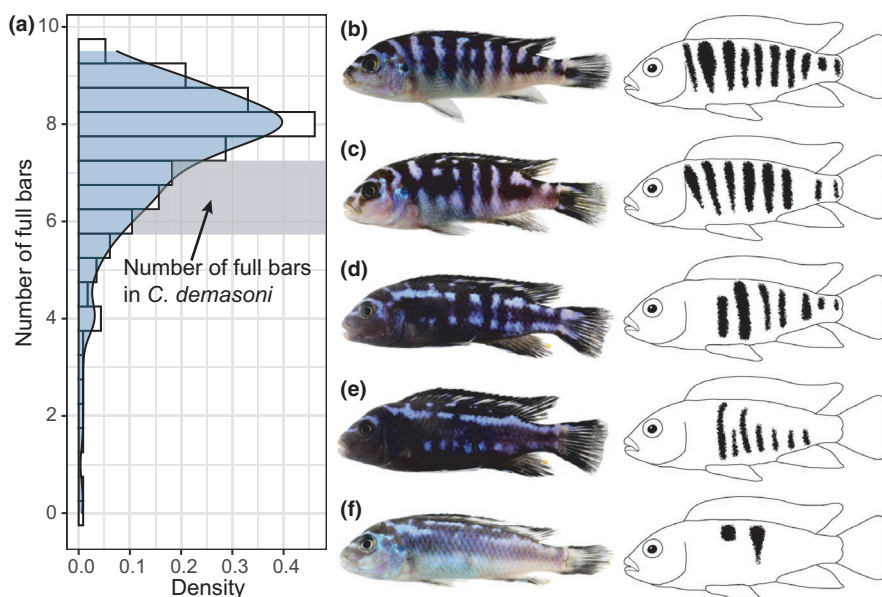

FIGURE 3 Vertical bars in the F_2 generation. Individuals of the F_2 generation show different numbers of vertical bars. (a) The distribution of the number of full bars in the F_2 generation. The grey area in a shows the phenotypic range of the barred parental species *C. demasoni*. (b–f) Individuals of the F2 generation with different bar numbers and schematic illustrations of their bar patterns

used measured stripe coverage as a more objective quantification instead of binary or ordinal measurements. To do so, we constructed a linkage map using quaddRAD markers from 230 F_2 individuals of the hybrid cross. The final linkage map contains 2770 markers arranged in 22 linkage groups corresponding to the 22 chromosomes of the high-quality reference genome of the closely related species *Maylandia zebra* (Conte & Kocher, 2015). The total length of the linkage map is 1297.8 cM with a mean marker distance of 0.47 cM. To identify all loci associated with vertical bar and horizontal stripe formation we tested for QTL exceeding the genome wide significance threshold (large effect QTL) as well as the chromosome-wide significance threshold (which we will refer to as suggestive QTL).

For the DLS, we found three QTL with log of the odds ratio (LOD) scores exceeding the genome wide LOD threshold on chromosomes 2 (LOD 5.6), 5 (LOD 4.6) and 18 (LOD 20.5) (Figure 4a). Additionally, we could identify four suggestive QTL on chromosomes 7, 17, 22 and 23. A multiple QTL model including all seven identified QTL explains 56.8% of the variation of the phenotype. The QTL on chromosome 18 that harbours the previously described major effect locus *agrp2* (Kratochwil et al., 2018) has the largest effect (24.4 percent variance explained [PVE]), followed by the QTL on chromosomes 2, 5 and 17 (~4 PVE). The remaining three QTL on chromosomes 7, 22 and 23 explain very little of the observed variation (<3 PVE) (Table S1, Figure S5).

For the MLS we found three significant QTL on chromosomes 2 (LOD 4.8), 10 (LOD 4) and 18 (LOD 12.2) (Figure 4b). Four suggestive QTL were found on chromosomes 3, 5, 15 and 17. A multiple QTL model including all identified QTL explains 42.2% of the observed variation of the MLS. Again, the QTL on chromosome 18 has the largest effect (14.1 PVE), while the QTL on chromosomes 2 and 10 show moderate effects (~3 PVE). The remaining four QTL explain only a small fraction of the variation (<3 PVE) (Table S1, Figure S6). The PVE of the QTL found on chromosome 18 is lower than previously reported with a single-locus model and using ordinal scoring of the MLS (52.7 PVE; Kratochwil et al., 2018). This difference is however mainly due to the phenotyping, as ordinal scoring of the MLS gives more comparable PVE for the QTL on chromosome 18 (31.8 PVE).

Next, we screened for QTL associated with the vertical bar patterns, a trait that had not been genetically mapped before. A limitation caused by the distribution of bar phenotypes in the F_2 mapping panel is that we could not map the presence/absence of the trait as all F_2 individuals had at least one bar (Figure 3). Also, due to meristic variation, a similar measurement of coverage as for the stripe phenotype is not possible. We hence focused on the number of bars as focal trait (Figure 3a). When analysing the total number of bars (including complete bars and bars that do not span the entire sides of the fish from dorsal to ventral) we could not identify any significant QTL. A more specific analysis of the number of complete bars revealed a single QTL with a LOD score of 5.3 (Figure 4c). We found five suggestive QTL on chromosomes 5, 7, 13, 14 and 18. A multiple QTL model including all six QTL explains 32% of the variation in the phenotype. In this case, the QTL located on chromosome 2 shows

the largest effect on the phenotype (6.8 PVE). All other QTL explain 5% or less of the variation (Table S1, Figure S7).

In summary, both of the horizontal stripes share the QTL with the largest effects on chromosomes 2 and 18 (the latter harbouring the previously associated *agrp2* locus [Table S3]) and two additional loci with smaller effects on chromosomes 5 and 17. The overlap of the QTL therefore probably contributes to the strong correlation of the DLS and MLS (Figure 2g). Moreover, each stripe shows individual and nonoverlapping QTL with smaller effect sizes (chromosomes 7, 22 and 23 for the DLS; chromosomes 3, 10 and 15 for the MLS) that might account for differences in the patterns. The analysis of the number of complete bars revealed one large effect QTL on chromosome 2 in a different location than the QTL associated with horizontal stripes suggesting that different genetic modules control the two melanic patterns. Three suggestive loci exclusive for vertical bars can be found on chromosomes 5, 13 and 14 and QTL that are shared between vertical bars and horizontal stripes could be found on chromosomes 7 and 18 (Figure 4c). The multiple QTL models explain a substantially higher percentage of variation for the two horizontal stripes (56.8 and 42.2 PVE for DLS and MLS, respectively) when compared to vertical bars (32.9 PVE).

3.3 | Distinct QTL control anterior-posterior variation of horizontal stripes

A striking observation from the phenotypic distribution of the F_2 individuals is the occurrence of intermediate phenotypes (Figure 2b–e) including more spotted or discontinuous stripe patterns that resemble naturally occurring phenotypes of other species (Figure 1c, d, e). This suggests that certain loci might regulate specific parts of the stripes and act on top of the previously described role of *agrp2* (Chr. 18) that merely acts as a more general inhibitor and on-off-switch of the whole pattern.

To test this hypothesis, we used a more fine-grained analysis of the spatial pattern using the R-package patternize (Van Belleghem et al., 2018). The software allows automatic alignment of images of the F_2 and parental individuals based on manually set morphometric landmarks followed by automatized extraction of colour information. Through PCA of the extracted colour information we generated PC values for each F_2 individual that could then be used for QTL mapping. This approach allowed us to analyse specific homologous regions of the respective colour patterns.

In a first step we compared this approach to the manual measurement (Figure 4). The results were similar and revealed two QTL with LODs exceeding the genome wide threshold on chromosomes 2 (LOD 7.2) and 18 (LOD 4.2) for the dorsolateral stripe (DLS). Contrary to the analysis of the visual phenotyping, the QTL on chromosome 2 shows the largest effect. Two suggestive QTL were found on chromosomes 22 and 23 (Figure 5a). A multiple QTL model including all four identified QTL explains 27.1% of the observed variation. The QTL on chromosome 2 has the largest effect on the phenotype (11 PVE), followed by the QTL on chromosome 18 (4.3 PVE). Each of the
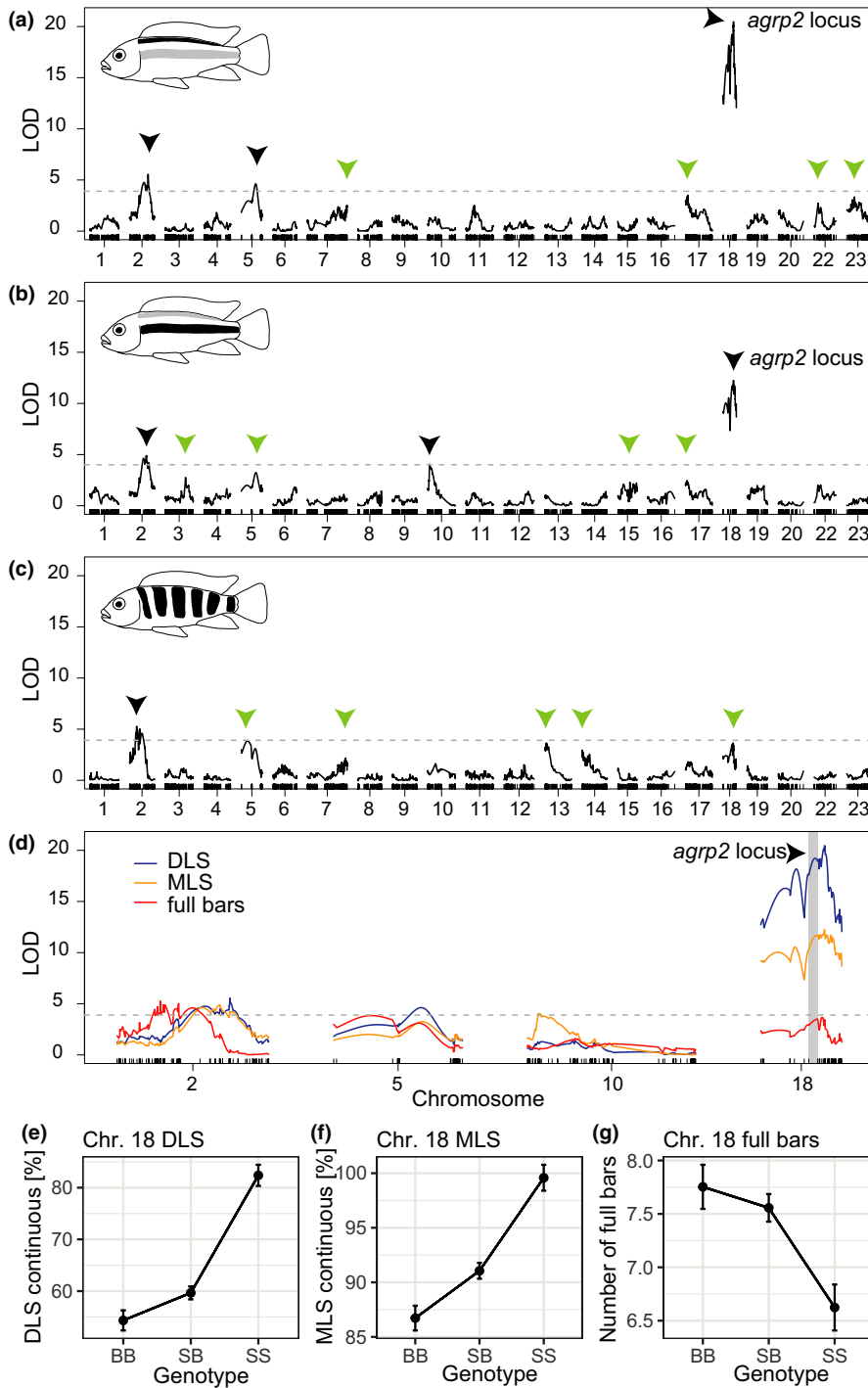
FIGURE 4 QTL mapping of visually phenotyped stripe and bar patterns in Lake Malawi cichlids. LOD scores across all 22 chromosomes for (a) the dorsolateral stripe, (b) the midlateral stripe and (c) complete vertical bars. (d) LOD scores across the chromosomes with QTL exceeding the genome wide significance threshold for all three phenotypes. The grey area in d indicates the position of the *agrp2* locus in the linkage map. Generally, grey dashed lines indicate genome wide significance thresholds and black arrows indicate QTL exceeding genome wide significance thresholds. Green arrows indicate suggestive QTL. The effect plots show the relationship between the genotype at the QTL on chromosome 18 and (e) the dorsolateral stripe, (f) the midlateral stripe and (g) the number of full bars. Points show the mean of the phenotype, error bars indicate the standard error. LOD, logarithm of the odds; BB, homozygous for allele of the barred parental species; SB, heterozygous; SS, homozygous for allele of the striped parental species; DLS, dorsolateral stripe; MLS, midlateral stripe

remaining four QTL explains less than 4% of the variation (Table S2). The peaks on chromosomes 2, 18, 22 and 23 are shared between the visual and patternize analysis.

The analysis of the midlateral stripe (MLS) revealed a single significant QTL on chromosome 18 (LOD 6.2). Additionally, we found four suggestive QTL on chromosomes 1, 7, 8 and 17 (Figure 5b). The multiple QTL model explains 28% of the observed variation. Again, the QTL on chromosome 18 shows the largest effect (10 PVE), while each of the remaining QTL explains less than 4% of the variation (Table S2). The peaks on chromosomes 17 and 18 are shared between the visual and patternize analysis (Figure 4).

To investigate the genetic basis of variation in the horizontal stripe pattern, we split each stripe into an anterior and a posterior part and performed separate QTL mapping for each of these parts. The detailed analysis of the DLS revealed that a large effect QTL on chromosome 2 is responsible for most of the variation observed in the anterior part (11.5 PVE) (Figure 5c), while the large effect QTL on chromosome 18 is responsible for the variation observed in the posterior part of the DLS (7.3 PVE) (Figure 5e). Similarly, the spatial analysis of the MLS showed that the largest effect QTL on chromosome 18 is responsible for 12% of the variation of the posterior part of the stripe, too (Figure 5d). No large effect QTL were found for the anterior part
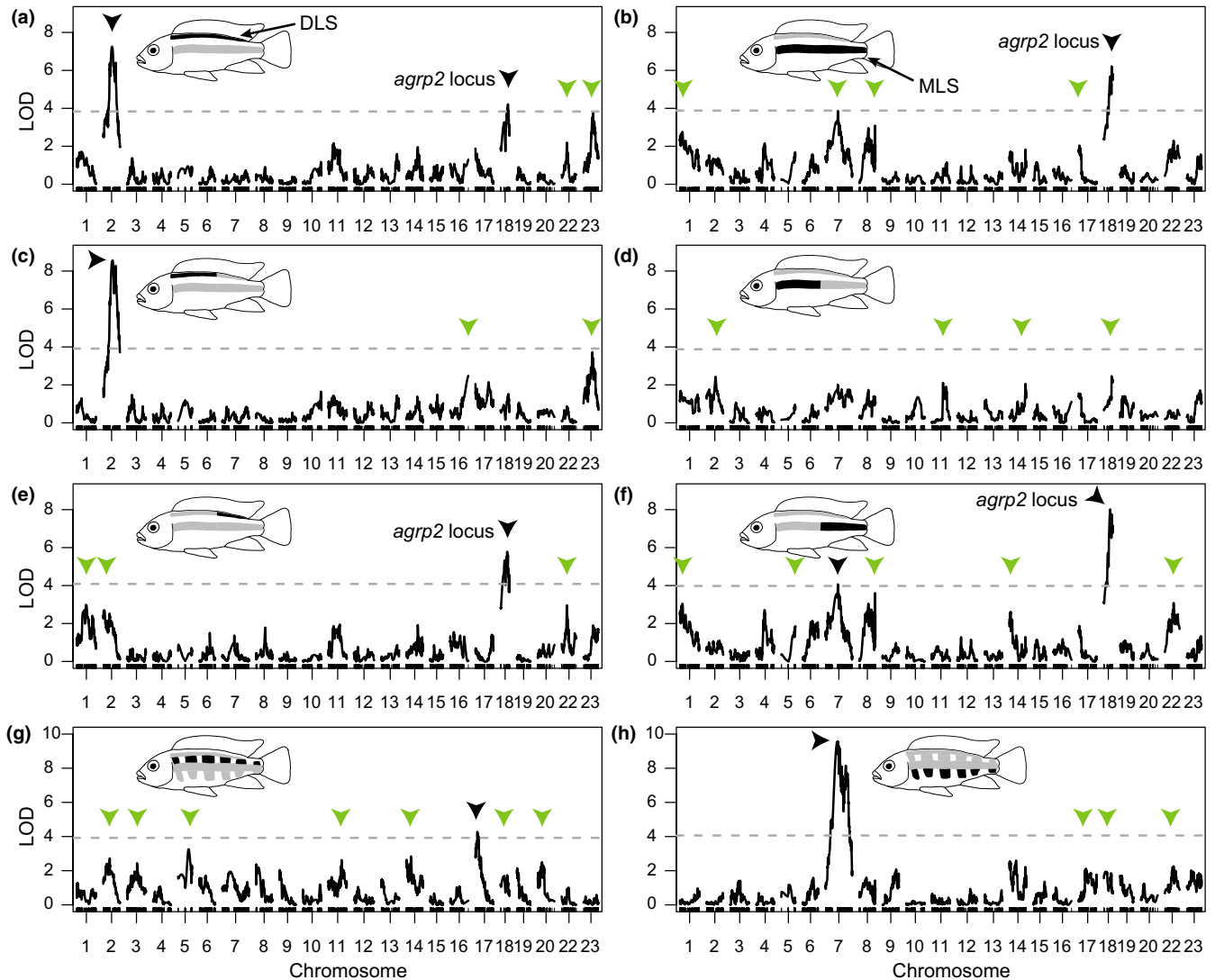
FIGURE 5 QTL mapping of stripe and bar patterns phenotyped using patternize. LOD scores across all 22 chromosomes for (a) the DLS, (b) the MLS, (c) the anterior DLS, (d) the anterior MLS, (e) the posterior DLS, (f) the posterior MLS, (g) the dorsal part of the vertical bars, (h) the ventral part of the vertical bars. Generally, grey dashed lines indicate genome wide significance thresholds. Black arrows indicate QTL exceeding genome wide significance thresholds. Green arrows indicate QTL exceeding chromosome wide significance thresholds

of the MLS (Figure 5f). Hence, we suggest that specific loci might be responsible for driving variation in specific parts of the stripe pattern.

3.4 | Spatial analysis of vertical bars reveals additional major effect QTL

Many species show variation in the intensity of melanic pigmentation along the dorsoventral axis, which includes bar patterns that often fade more ventrally — a phenotype that is also apparent in a substantial number of F_2 individuals (Figure 2b, e). Therefore, we investigated if there is a QTL associated with this dorsoventral patterning of vertical bars using patternize (Figure 5g and h). To avoid interference with the horizontal stripes we limited the analysis to the dorsal (between the DLS and the MLS) and the ventral region (ventral of the MLS). The analysis of the dorsal region revealed a

single significant QTL (LOD 4.3) on chromosome 17. Seven other suggestive QTL were found on chromosomes 2, 3, 5, 11, 14, 18 and 20 (Figure 5g). A multiple QTL model including all identified QTL explains 36.3% of the observed variation of the bar phenotype. The QTL on chromosome 17 shows the largest effect on bar presence (8.5 PVE), while the remaining QTL explain less than 4.2% of the variation. The analysis of the ventral region revealed one QTL with a LOD value exceeding the genome wide significance threshold on chromosome 7 (LOD 9.6). We found three suggestive QTL on chromosomes 17, 18 and 22 (Figure 5h). A multiple QTL model including all identified QTL explains 27.7% of the observed phenotypic variation. The QTL on chromosome 7 has the largest effect (14.9 PVE). Each of the remaining QTL explain less than 5% of the observed variation (Table S2). Therefore, similar to the results for the stripe patterns the in-depth analysis of the bar patterns suggests a spatial control of the dorsoventral pattern.

4 | DISCUSSION

We investigated the genetic basis of the two most common colour patterns in African cichlids, horizontal stripes and vertical bars, using a hybrid cross between the Lake Malawi species *P. cyaneorhabdos* and *C. demasoni*. We identified a multitude of genomic loci explaining the variation of the two colour patterns that we can see between the two parental species. Our results provide insights into the modularity of colour patterns as well as how more fine-scale variation of these patterns might be caused. Lastly, these results in combination with previous work gives, despite being speculative, insights into how the dynamics of colour pattern evolution might be constrained and facilitated by their genetic architecture.

4.1 | Modularity and independent evolution of stripe and bar patterns

Our first observation is that bar and stripe patterns are inherited independently. This is demonstrated by the independent assortment of the two traits in our hybrid cross. In case of a shared genetic basis of the two colour patterns, the traits would show a strong negative correlation in the F_2 generation as different alleles at the shared colour pattern loci would only allow the formation of either pattern. Furthermore, the QTL mapping does not show substantial overlap between the large effect QTL of bar and stripe patterns. These results suggest that the formation of vertical bars and horizontal stripes does not share the same genetic basis and that the traits thus constitute variational and developmental modules (sensu (Wagner et al., 2007)), allowing independent evolution, loss or gain of one colour pattern without affecting the other colour pattern. Clearly, both patterns are shaped by variation in number and properties (e.g., pigment content) of the same cell types (melanophores in the dark regions, and xanthophores and iridophores in the regions in between). Yet the coordination (and the loci that underlie this coordination) of the pigment cells that are shaping bar and stripe patterns seem to be largely independent as the two phenotypes freely segregate in the F_2 generation providing no evidence for a shared genetic basis. This is further supported by previous results from a CRISPR-Cas9 knockout of the major effect gene *agrp2* that specifically affects stripe but not bar formation (Kratochwil et al., 2018). The fact that only rarely species display both stripes and bars could therefore be explained by a functional mismatch and not by any sort of developmental constraint or pleiotropic effect (as individuals with both patterns are very common in the F_2 panel [Figure 2]). A plausible explanation for the fact that both traits rarely occur in the same species (with exceptions, as for example multiple species in the genus *Julidochromis*, *Neolamprologus buescheri*, *Haplochromis sauvagei*, etc.) is that both traits constitute alternative but not compatible dazzle camouflage or communication strategies (the barring pattern becomes more pronounced in courting males in particular) that have adaptive optima in species with specific behaviour (e.g., swimming speed), morphology (e.g., body shape) and environmental preferences (e.g., rocky, sandy,

vegetated, open-water environments) (Seehausen et al., 1999). However, more robust evidence for the adaptive function of the two colour patterns is still missing and needs further investigation. Due to the mostly uncoupled genetic basis of both traits, selection can effectively act on each trait separately. A similar example for such functional and genetic decoupling that greatly facilitated rates of diversification in cichlid fishes are pharyngeal and oral jaws (Ahi et al., 2019; Hulsey et al.,2006, Hulsey et al., 2016; Parsons et al., 2012). We observed a weak negative correlation between the number of full bars and the MLS (Figure 2i) that is probably caused by the shared QTL on chromosome 18 (Figure 4). This QTL has the largest effect on the presence of both of the horizontal stripes (PVE 14 for the MLS), but it only has slight effects the presence of full bars (3.7 PVE), causing a weak correlation of the phenotypes but still granting a mostly independent inheritance of the traits.

4.2 | The genetic basis of horizontal stripes: from major effect loci to incremental changes

Previous results have demonstrated a simple Mendelian basis for stripe patterns in Lake Victoria cichlids that are caused by cis-regulatory mutations of the *agrp2* gene and also identified the *agrp2* locus on chromosome 18 — using a single locus test — as a major effect locus in Lake Malawi (Henning et al., 2014; Kratochwil et al., 2018). Here, using a genome-wide QTL search we confirm that *agrp2* is indeed the major effect locus underlying the presence of stripe patterns (Table S3). On top of this, we have identified several loci that influence stripe patterns globally (Figure 4) and spatially (Figure 5). The fact that the identified QTL only explain around 50% of the observed variation (57% for the DLS, 42% for the MLS) indicates that there might be nongenetic factors that affect colour pattern formation. Interestingly, by comparison with the Lake Victoria cross between the striped species *Haplochromis sauvagei* and the nonstriped species *Pundamilia nyererei* (Henning et al., 2014), our Lake Malawi cross shows substantially more variation in the F_2 generation. This includes variation in the continuity, but also in the number of stripes. And while there is a strong positive correlation between the presence and relative coverage of the DLS and the MLS (Figure 2g), several individuals (~22%) had only one of the stripes. Interestingly, it was always the DLS that was missing (no individual only lacked the MLS), suggesting that one or more underlying alleles are specifically affecting the DLS, while others affect both. Indeed, QTL analysis of the visually phenotyped DLS and the MLS revealed that two major effect loci (chromosomes 2 and 18) were shared between the two stripes, while one major effect QTL was exclusively associated with each of the stripes (chromosome 5 for DLS, chromosome 10 for MLS) (Figure 4).

Stripe patterns in the F_2 panel did not only show meristic variation (i.e., in the number of stripes), but also variation in their continuity with some individuals having a more spotted or interrupted stripe and others only showing stripes in the anterior or posterior flank. Interestingly, many of these phenotypes (Figure 2) were never seen

in individuals of the parental species but had striking resemblance to stripe phenotypes of other cichlid species found in Lake Malawi (Figure 2). For example, we can find phenotypes that are like the stripe patterns of *Protomelas annectens* (DLS absent, MLS present; Figure 1d), *Cyrtocara moorii* (one anterior portion of the DLS, two posterior portions of the MLS; Figure 1e) or *Pseudotropheus crabro* (DLS absent, MLS present with small posterior gap; Figure 1c).

In the context of previous work these new results allow for interesting speculation on the evolutionary history of the loci underlying stripe formation in the haplochromine radiations of Lake Malawi and Lake Victoria. Based on these results we propose two interpretations explaining the differences of the genetic basis of gain and loss of horizontal stripes we found between the radiations of Lakes Malawi and Victoria. The genetic basis of stripes in Lake Victoria cichlids is well explained by regulatory evolution of two divergent haplotypes of a single intronic interval that evolved prior to the radiation and seems to be responsible for most of the variation in stripe patterns across the ~500 species of this radiation (Kratochwil et al., 2018; Urban et al., 2021). The simple, monogenic basis of horizontal stripes in Lake Victoria cichlids could be therefore the result of a bottleneck during the colonization of Lake Victoria (i.e., other stripe-affecting alleles did not enter the Lake Victoria radiation with the colonizers). The trait would have thereby evolved from being an oligogenic trait to a monogenic trait. Another potential explanation is that the trait was initially a monogenic trait and that this basis was maintained in Lake Victoria, but that additional stripe alleles only evolved in Lake Malawi. This second hypothesis is supported by previous results suggesting that the stripe-affecting alleles from Lake Victoria are older than the radiation itself and that stripe-associated alleles evolved de novo within Lake Malawi (Urban et al., 2021). The evolution of more loci affecting the formation of horizontal stripes in Lake Malawi might have been possible due to the older age of this radiation when compared to the Lake Victoria radiation (10,000–100,000 years for Lake Victoria Elmer et al., 2009; Meyer et al., 1990; Seehausen, 2002; Verheyen et al., 2003; Wagner et al., 2013); 800,000 years for Lake Malawi (Brawand et al., 2014; Meyer et al., 1990). In summary, stripes would have therefore continued to be a monogenic trait in Lake Victoria, while in Lake Malawi new modifier loci have evolved leading to a transition from a monogenic to an oligogenic trait. This hypothesis is supported by findings from studies on the genetic basis of colour patterns in the genus *Danio*. Here, it was shown that the colour pattern differences between a horizontally striped species (*D. quagga*) and a spotted species (*D. kyathit*) has a complex genetic basis, even though transitions from the striped phenotype to the spotted phenotype could be achieved by loss of function mutations in single genes in *D. rerio* (McCluskey et al., 2021). This example illustrates that phenotypic differences between divergent lineages are likely to be based on multiple genes even though the same phenotypic differences can potentially be caused by a simple genetic change, supporting the accumulation of additional modifier loci during the course of evolution.

Alternatively to the evolution of novel stripe modifying alleles within Lake Malawi, it is also possible that the evolution of stripe patterns in Lake Malawi was solely driven by selection on pre-existing genetic variation (including variation that was introduced by introgression and/or hybridization in ancestral lineages) (Loh et al., 2013; Svardal et al., 2020). Most of the ancestral, riverine species including for example, *Astatotilapia burtoni* and *Astatoreochromis alluaudi* have an interrupted stripe pattern that is much less pronounced than the stripes of most Malawi species and resembles some of the F_2 individuals (Figure 2b). Accordingly, the stripe phenotype diversity among cichlids in Lake Malawi might have been caused by incomplete lineage sorting of alleles (or even adaptive introgression) (Malinsky et al., 2018; Svardal et al., 2020) influencing stripe pattern formation (even if the phenotypes were not present ancestrally because of low allele frequencies) and selection acting on the diverse phenotypes caused by these different allele combinations (including the strongly striped species like *P. cyaneorhabdos*). This would be comparable to for example, the evolution of extreme body shapes from standing genetic variation via polygenic selection, as for example observed in cichlid crater lake radiations (Kautt et al., 2020; Malinsky et al., 2015). Furthermore, it would support the possible role of ancient variation in the formation of phenotypic diversity during rapid adaptation (Barrett & Schluter, 2008; Brawand et al., 2014; Irisarri et al., 2018; Meier et al., 2017; Svardal et al., 2020). Both mechanisms (evolution of novel modifier alleles and evolution of novel allele combinations) or a combination of both might therefore have contributed to the diversity of stripe pattern phenotypes that makes up the radiation of cichlids in Lake Malawi.

The higher number of QTL in our Malawi cichlid hybrid cross therefore might suggest an explanation for the higher phenotypic diversity in stripe patterns among Lake Malawi cichlids, as compared to Lake Victoria cichlids, that include variation in the number, orientation, continuity and contrast of the stripes. Our cross provides experimental evidence that different combinations of alleles from only one striped and one nonstriped species can indeed create a variety of naturally occurring phenotypes different from those observed in the parents (Figure 1). This supports the conclusion that novel combinations of alleles driven by hybridization and/or introgression could in fact have facilitated this diversification process in colour patterns (as it did here in our hybridization experiment in the laboratory). It has been suggested before that hybridization played a key role in the evolution of the cichlid radiations of Lakes Victoria (Meier et al., 2017), Tanganyika (Irisarri et al., 2018; Koblmuller et al., 2007; Salzburger et al., 2002) and Malawi (Genner & Turner, 2012; Joyce et al., 2011; Svardal et al., 2020) as it allows the reassembly of variants (Marques et al., 2019) that might ultimately shape the evolution of traits as we suggest here for colour patterns.

4.3 | Polygenic basis of vertical bars

Contrary to horizontal stripes, that were absent in some of the F_2 individuals, none of the 230 F_2 completely lacked bar patterns (Figure 3). This not only suggests a rather polygenic basis, but also supports the existence of several dominant alleles that are sufficient

in driving the formation of bars and ultimately override the recessive alleles that block bar formation resulting in a high robustness of the trait in the face of hybridization. As there is substantial variation in bar number, thickness and intensity, allele combinations might however result in variation of these characters. The QTL analysis showed that at least 11 different genomic loci are associated with bar number. The multiple-QTL model for the number of complete bars explained 32% of the observed variation, possibly indicating that here we were not able to detect all of the minor effect QTL due to phenotypic limitations or lack of power due too small sample size of the QTL cross or instead indicating nongenetic effects on pattern formation. A QTL study in subsequent backcross generations instead of the F_2 generation would be appropriate to better understand the effect of single QTL in a highly polygenic trait like vertical bars, as usually done when analysing the genetic basis of highly polygenic yield-relevant traits in commercial crops (Bernacchi et al., 1998; Huang et al., 2003; Pillen et al., 2003; Tanksley & Nelson, 1995). The analysis of backcross generations makes it easier to isolate certain alleles of one parental species in the genetic background that is mostly provided by the other parental species to better understand the effect of single QTL. The fact that we were only able to detect large effect QTL for the number of complete bars but not for the total number of bars might be explained by the higher number of individuals with extreme phenotypes when only counting the complete bars (more individuals with low bar numbers).

The barred parental species of our hybrid cross (*C. demasoni*) usually displays either six or seven complete vertical bars (Figure 1f). In contrast, the distribution of vertical bars in the F_2 generation is shifted towards a mean value of eight bars, with some individuals showing up to 11 bars (Figure 3a). This case of transgressive segregation is another potential source for novel phenotypes resulting from hybridization and one of many examples of transgressive segregation in hybrid crosses of Lake Malawi cichlids (Holzman & Hulsey, 2017; Husemann et al., 2017; Parsons et al.,2011). We speculate that the gap in the morphospace at the lower end of the bar distribution (Figure 3a) could hint at a developmental constraint during the formation of vertical bars preventing the formation of a vertical bar pattern with fewer than four bars (Maynard Smith et al., 1985). This would also explain why we see very few cichlid species with low bar numbers in the haplochromine radiations.

4.4 | The genetic and cellular basis of colour pattern formation

The large effect QTL on chromosome 18 encompasses the gene *agrp2* (*asip2b*) (Table S3) that previously has been associated with the presence of horizontal stripes in African cichlids (Kratochwil et al., 2018). Low levels of *agrp2* expression in the skin are associated with horizontal stripes in cichlids of the East African Great Lakes and knocking out the gene in a formerly non-striped species resulted in the formation of horizontal stripes (Kratochwil et al., 2018). Remarkably, *agrp2* expression in the skin does not differ between

striped and non-striped regions in African cichlids (Liang et al., 2021). Even though the association between low levels of *agrp2* and the presence of horizontal stripes seems evident, the exact mechanisms by which *agrp2* expression in the skin affects the formation of stripes remain unclear.

While there is no clear evidence for the exact molecular functions of *agrp2* during horizontal stripe formation it could be speculated that the gene affects colour pattern formation via the melanocortin system (Cal et al., 2017). The agouti gene family that *agrp2* is a member of, has been shown to act as a strong antagonist of different melanocortin receptors in mammals (Ollmann et al., 1997) and zebrafish (Zhang et al., 2010) which could affect chromatophore proliferation, pigment dispersion and pigment production (Cal et al., 2017). In zebrafish, *agrp2* expression is also necessary for the regulation of *pmch* and *pmchl* which are responsible for melanosome aggregation during background adaptation (Berman et al., 2009; Zhang et al., 2010). Both, *pmch* and *pmchl* were found in the intervals of different QTL detected in our study (*pmch* on chromosome 17 for DLS and MLS; *pmchl* on chromosome 7 for DLS and full bars [Table S3]). Further genes like *ednrba* (chromosome 18 for DLS, MLS and full bars) and *mitfa* (chromosome 5 for DLS) are not directly involved in melanin production and aggregation but affect pigment cell development (Nataf et al., 1996; Opdecamp et al., 1997). Previous studies emphasized the role of pigment cell development, migration and chromatophore interaction in the formation of horizontal stripes in zebrafish. In contrast to zebrafish, where melanophores (the melanin bearing pigment cells) are almost exclusively found in the dark regions of the colour pattern (Patterson et al., 2014), melanophores are numerous in the regions between bars and stripes in East African cichlids. Nevertheless, the contrast between dark and light regions is reached through a combination of differential melanophore density and differential melanosome aggregation in the melanophores (Hendrick et al., 2019; Liang, Gerwin, et al., 2020). Different melanophore densities between striped and nonstriped regions could be obtained by interactions between the different chromatophore types as seen in zebrafish (Frohnhofer et al., 2013).

Even though our study revealed multiple genes that might be involved in colour pattern formation in African cichlids, we still do not understand the exact molecular mechanisms and processes that connect these genes and the pattern formation. Further studies that build on the knowledge of colour pattern formation, involved genes in zebrafish as well as in-depth investigation of pigment cell interactions will be needed for a more comprehensive understanding.

4.5 | Conclusion

In this study, we found that the most common melanic patterns in East African cichlids, vertical bars and horizontal stripes, are two genetically independent modules. The QTL analysis revealed a multigenic basis for both of the colour patterns. Through a more detailed analysis of the stripe pattern we could identify QTL that specifically associate with certain portions of the stripe pattern. Our study

provides new insights into how genetic modularity and the evolution and combination of modifier loci might have driven the remarkable colouration diversification in the adaptive radiations of the East African Great Lake cichlids.

AUTHOR CONTRIBUTIONS

Jan Gerwin and Sabine Urban collected data. Jan Gerwin conducted the analysis. Jan Gerwin and Claudius F. Kratochwil wrote the manuscript. Sabine Urban and Axel Meyer contributed to the manuscript. Claudius F. Kratochwil and Axel Meyer designed and coordinated the study.

DATA AVAILABILITY STATEMENT

RAD-seq data is available through the NCBI Sequence Read Archive (PRJNA730667). The processed genotype and phenotype data are part of the Supporting Information (Table S4).

ORCID

Jan Gerwin https://orcid.org/0000-0001-7897-3302
Claudius F. Kratochwil https://orcid.org/0000-0002-5646-3114

REFERENCES

Ahi, E. P., Singh, P., Duenser, A., Gessl, W., & Sturmbauer, C. (2019). Divergence in larval jaw gene expression reflects differential trophic adaptation in haplochromine cichlids prior to foraging. *BMC Evolutionary Biology*, *19*(1), 150. https://doi.org/10.1186/s12862-019-1483-3

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., Bedford, N., Bergek, S., Chan, Y. F., Jones, F. C., Kingsley, D. M., Peichel, C. L., & Schluter, D. (2014). Genetics of ecological divergence during speciation. *Nature*, *511*(7509), 307–311. https://doi.org/10.1038/nature13301

Barlow, G. W. (2000). *The cichlid fishes - nature's grand experiment in evolution*. Perseus Publishing.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Berman, J. R., Skariah, G., Maro, G. S., Mignot, E., & Mourrain, P. (2009). Characterization of two melanin-concentrating hormone genes in zebrafish reveals evolutionary and physiological links with the mammalian MCH system. *The Journal of Comparative Neurology*, *517*(5), 695–710. https://doi.org/10.1002/cne.22171

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., & Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersiconhirsutum. *Theoretical and Applied Genetics*, *97*, 381–397. https://doi.org/10.1007/s001220050908

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

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Cal, L., Suarez-Bregua, P., Cerda-Reverter, J. M., Braasch, I., & Rotllant, J. (2017). Fish pigmentation and the melanocortin system. *Comparative Biochemistry and Physiology Part A Molecular Integrative Physiology*, *211*, 26–33. https://doi.org/10.1016/j.cbpa.2017.06.001

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Conte, M. A., & Kocher, T. D. (2015). An improved genome reference for the African cichlid, Metriaclima zebra. *BMC Genomics*, *16*, 724. https://doi.org/10.1186/s12864-015-1930-5

Elmer, K. R., Reggio, C., Wirth, T., Verheyen, E., Salzburger, W., & Meyer, A. (2009). Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. *Proceedings of the National Academy of Sciences*, *106*, 13404–13409. https://doi.org/10.1073/pnas.0902299106

Franchini, P., Monne Parera, D., Kautt, A. F., & Meyer, A. (2017). quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Molecular Ecology*, *26*(10), 2783–2795. https://doi.org/10.1111/mec.14077

Frohnhofer, H. G., Krauss, J., Maischein, H. M., & Nusslein-Volhard, C. (2013). Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. *Development*, *140*(14), 2997–3007. https://doi.org/10.1242/dev.096719

Genner, M. J., & Turner, G. F. (2012). Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*, *29*(1), 195–206. https://doi.org/10.1093/molbev/msr183

Haley, C. S., & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, *69*, 15–324. https://doi.org/10.1038/hdy.1992.131

Hendrick, L. A., Carter, G. A., Hilbrands, E. H., Heubel, B. P., Schilling, T. F., & Le Pabic, P. (2019). Bar, stripe and spot development in sand-dwelling cichlids from Lake Malawi. *Evodevo*, *10*, 18. https://doi.org/10.1186/s13227-019-0132-7

Henning, F., Lee, H. J., Franchini, P., & Meyer, A. (2014). Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Molecular Ecology*, *23*(21), 5224–5240. https://doi.org/10.1111/mec.12860

Holzman, R., & Hulsey, C. D. (2017). Mechanical Transgressive Segregation and the Rapid Origin of Trophic Novelty. *Scientific Reports*, *7*, 40306. https://doi.org/10.1038/srep40306

Huang, X. Q., Coster, H., Ganal, M. W., & Roder, M. S. (2003). Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics.*, *106*(8), 1379–1389. https://doi.org/10.1007/s00122-002-1179-7

Hulsey, C. D., Fraser, G. J., & Meyer, A. (2016). Biting into the genome to phenome map: developmental genetic modularity of cichlid fish dentitions. *Integrative and Comparative Biology*, *56*(3), 373–388. https://doi.org/10.1093/icb/icw059

Hulsey, C. D., León, F. J. G., & Rodiles-Hernández, R. (2006). Micro- and Macroevolutionary Decoupling of Cichlid Jaws: a Test of Liem's Key Innovation Hypothesis. *Evolution*, *60*, 2096–2109. https://doi.org/10.1111/j.0014-3820.2006.tb01847.x

Husemann, M., Tobler, M., McCauley, C., Ding, B., & Danley, P. D. (2017). Body shape differences in a pair of closely related Malawi cichlids and their hybrids: Effects of genetic variation, phenotypic plasticity, and transgressive segregation. *Ecology and Evolution*, 7(12), 4336–4346. https://doi.org/10.1002/ece3.2823

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., Fischer, C., Lemmon, A. R., Lemmon, E. M., Thallinger, G. G., Sturmbauer, C., & Meyer, A. (2018). Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications*, 9(1). https://doi.org/10.1038/s41467-018-05479-9

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), R108–109. https://doi.org/10.1016/j.cub.2010.11.029

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C. D., Franchini, P., Pippel, M., Myers, E. W., & Meyer, A. (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, 588(7836), 106–111. https://doi.org/10.1038/s41586-020-2845-0

Koblmuller, S., Duftner, N., Sefc, K. M., Aibara, M., Stipacek, M., Blanc, M., & Sturmbauer, C. (2007). Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika–the result of repeated introgressive hybridization. *BMC Evolutionary Biology*, 7, 7. https://doi.org/10.1186/1471-2148-7-7

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

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(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liang, Y., Gerwin, J., Meyer, A., & Kratochwil, C. F. (2020). Developmental and cellular basis of vertical bar color patterns in the east African cichlid fish *Haplochromis latifasciatus*. *Frontiers in Cell and Developmental Biology*, 8, 62. https://doi.org/10.3389/fcell.2020.00062.

Liang, Y., Grauvogl, M., Meyer, A., & Kratochwil, C. F. (2021). Functional conservation and divergence of color-pattern-related agouti family genes in teleost fishes. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 336(5), 443–450. https://doi.org/10.1002/jez.b.23041

Liang, Y., Meyer, A., & Kratochwil, C. F. (2020). Neural innervation as a potential trigger of morphological color change and sexual dimorphism in cichlid fish. *Scientific Reports*, 10(1), 12329. https://doi.org/10.1038/s41598-020-69239-w

Loh, Y.-H., Bezault, E., Muenzel, F. M., Roberts, R. B., Swofford, R., Barluenga, M., Kidd, C. E., Howe, A. E., Di Palma, F., Lindblad-Toh, K., Hey, J., Seehausen, O., Salzburger, W., Kocher, T. D., & Streelman, J. T. (2013). Origins of shared genetic variation in African cichlids. *Molecular Biology and Evolution*, 30(4), 906–917. https://doi.org/10.1093/molbev/mss326

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350(6267), 1493–1498.

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 & Evolution*, 2(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. https://doi.org/10.1038/nature05706

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, 34(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

McCluskey, B. M., Uji, S., Mancusi, J. L., Postlethwait, J. H., & Parichy, D. M. (2021). A complex genetic architecture in zebrafish relatives Danio quagga and D. kyathit underlies development of stripes and spots. *PLoS Genetics*, 17(4), e1009364. https://doi.org/10.1371/journal.pgen.1009364

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8, 14363. https://doi.org/10.1038/ncomms14363

Meyer, A., Kocher, T. D., Basasibwaki, P., & Wilson, A. C. (1990). Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, 347(6293), 550–553. https://doi.org/10.1038/347550a0

Nataf, V., Lecoin, L., Eichmann, A., & Le Douarin, N. M. (1996). Endothelin-B receptor is expressed by neural crest cells in the avian embryo. *Proceedings of the National Academy of Sciences*, 93, 9645–9650. https://doi.org/10.1073/pnas.93.18.9645

Ollmann, M., Wilson, B., Yang, Y., Kerns, J., Chen, Y., Gantz, I., & Barsh, G. (1997). Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science*, 278(5335), 135–138.

Opdecamp, K., Nakayama, A., Nguyen, M. T., Hodgkinson, C. A., Pavan, W. J., & Arnheiter, H. (1997). Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development*, 124, 2377–2386. https://doi.org/10.1242/dev.124.12.2377

Parsons, K. J., Marquez, E., & Albertson, R. C. (2012). Constraint and opportunity: the genetic basis and evolution of modularity in the cichlid mandible. *American Naturalist*, 179(1), 64–78. https://doi.org/10.1086/663200

Parsons, K. J., Son, Y. H., & Albertson, R. C. (2011). Hybridization Promotes Evolvability in African Cichlids: Connections Between Transgressive Segregation and Phenotypic Integration. *Evolutionary Biology*, 38(3), 306–315. https://doi.org/10.1007/s11692-011-9126-7

Patterson, L. B., Bain, E. J., & Parichy, D. M. (2014). Pigment cell interactions and differential xanthophore recruitment underlying zebrafish stripe reiteration and Danio pattern evolution. *Nature Communications*, 5, 5299. https://doi.org/10.1038/ncomms6299

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pillen, K., Zacharias, A., & Leon, J. (2003). Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 107(2), 340–352. https://doi.org/10.1007/s00122-003-1253-9

Salzburger, W., Baric, S., & Sturmbauer, C. (2002). Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, 11, 619–625. https://doi.org/10.1046/j.0962-1083.2001.01438.x.

Seehausen, O. (2002). Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14,600 year history for its cichlid species flock. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1490), 491–497. https://doi.org/10.1098/rspb.2001.1906

Seehausen, O., Mayhew, P., & van Alphen, J. (1999). Evolution of colour patterns in East African cichlid fish. *Journal of Evolutionary Biology*, 12, 514–534. https://doi.org/10.1046/j.1420-9101.1999.00055.x

Smith, J. M., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., & Wolpert, L. (1985). Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. *The Quarterly Review of Biology*, 60(3), 265–287. https://doi.org/10.1086/414425

Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., Genner, M. J., Turner, G. F., & Durbin, R. (2020). Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation. *Molecular Biology and Evolution*, *37*(4), 1100–1113. https://doi.org/10.1093/molbev/msz294

Tanksley, S. D., & Nelson, J. C. (1995). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, *92*, 191–203. https://doi.org/10.1007/BF00223376

Urban, S., Nater, A., Meyer, A., & Kratochwil, C. F. (2021). Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. *Molecular Biology and Evolution*, *38*(2), 465–477. https://doi.org/10.1093/molbev/msaa237

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, H., Hendrickx, F., Jiggins, C. D., McMillan, W. O., & Counterman, B. A. (2018). patternize: An R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, *9*(2), 390–398. https://doi.org/10.1111/2041-210X.12853

Verheyen, E., Salzburger, W., Snoeks, J., & Meyer, A. (2003). Origin of the Superflock of Cichlid Fishes from Lake Victoria, East Africa. *Science*, *300*(5617), 325–329.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, *22*(3), 787–798. https://doi.org/10.1111/mec.12023

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, *8*(12), 921–931. https://doi.org/10.1038/nrg2267

Zhang, C., Song, Y., Thompson, D. A., Madonna, M. A., Millhauser, G. L., Toro, S., Varga, Z., Westerfield, M., Gamse, J., Chen, W., & Cone, R. D. (2010). Pineal-specific agouti protein regulates teleost background adaptation. *Proceedings of the National Academy of Sciences*, *107*(47), 20164–20171. https://doi.org/10.1073/pnas.1014941107

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.