Title: Highly modular genomic architecture underlies combinatorial mechanism of speciation and adaptive radiation

Authors: Pooja Singh[1,2], Heidi Tschanz-Lischer[3,4], Kassandra Ford[1,2,5,6], Ehsan P. Ahi[7], Marcel P. Haesler[1,2] Salome Mwaiko[1,2], Joana I. Meier[1,2,8,9], David Marques[1,2,10], Remy Bruggman[3,4], Mary Kishe[11], and Ole Seehausen[1,2]

Affiliations:

[1]Aquatic Ecology & Evolution Division, Institute of Ecology and Evolution, University of Bern, Bern, CH-3012, Switzerland

[2]Department of Fish Ecology & Evolution, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Kastanienbaum, CH-6047, Switzerland

[3]Interfaculty Bioinformatics Unit – IBU, University of Bern, Bern, CH-3012, Switzerland

[4]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

[5]Current address: Department of Fisheries, Wildlife, & Conservation Biology, University of Minnesota – Twin Cities, Minnestoa, USA

[6]Current address: Bell Museum, College of Food, Agriculture, and Natural Resource Sciences, University of Minnesota – Twin Cities, Minnesota, USA

[7]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 9, 00014, Helsinki, Finland

[8]Current address: Department of Zoology, University of Cambridge, Cambridge, UK

[9]Current address: Tree of Life Programme, Wellcome Sanger Institute, Hinxton, UK

[10]Current address: Natural History Museum, Basel, CH-4051, Switzerland

[11]Tanzania Fisheries Research Institute (TAFIRI), 9750 Dar Es Salaam, Tanzania

Corresponding author: Pooja Singh (pooja.singh09@gmail.com)

Abstract

Adaptive radiation is a key modulator of biodiversity when new adaptive zones become available. Some of the largest and most diverse adaptive radiations have evolved from hybrid populations in the absence of geographic isolation. While phenotypic differentiation among their species is stunning, little is known about the genetics underpinning it. In Lake Victoria more than 500 cichlid fish species representing nearly as many different combinations of morphological and colouration phenotypes have evolved in 16,000 years. Analysing phenotypes and genomes of 107 of these species we show that the genetic basis of recurrently diverging traits associated with speciation and ecological adaptation is generally polygenic, often redundant, and dispersed across the genome. Up to 50% of alleles associated with variation in these traits were amassed during ancestral hybridisation, which may explain the low pleiotropy we find among traits. Such independent trait

modules act like Lego pieces, permitting many trait combinations to evolve by recombining a finite set of elements. We argue that modular trait architectures emerging from admixture variation are key to combinatorial speciation happening often, making adaptive radiation ultrafast and species rich.

Key words: speciation, polygenic adaptation, genetic architecture, pleiotropy, genetic redundancy, hybridisation

Main Text:

Adaptive radiation, where an ancestral lineage diversifies quickly into many eco-morphologically diverse species that can coexist, may have been an important and widespread source of biodiversity throughout the 3.5-billion-year history of Life on Earth[1]. Its occurrence is highly punctuated in space and time and despite its importance, knowledge of the biological factors that trigger, facilitate and constrain this process is incomplete. It is widely accepted that exceptional ecological opportunity is a prerequisite[2], but only some of many possible lineages seem to ever have responded to exceptional ecological opportunity with adaptive radiations. From several of the geologically young radiations we know the process can unfold in record time of just a few thousand years[3–6], consistent with the fossil evidence for long periods of relative stasis, punctuated by bursts of diversification in the history of life[7,8]. Such species radiations, far too fast and species rich for intrinsic reproductive isolation to have emerged, tend to be characterised by the coincidence of hybrid ancestry and rapid and simultaneous evolution of ecological and mating traits, as was predicted by Anderson & Stebbins 70 years ago[7]. While some inroads have been made linking rapid evolution to the availability of ancient admixture variation[9–14], we still understand little about the genetic basis that not only allows key traits to evolve rapidly, but to generate many different complex phenotypes from limited ancestral diversity, associated with many speciation events in short succession and in the absence of opportunity for geographical isolation.

Research has illuminated two prevalent genetic architectures underlying ecological speciation without geographic isolation. These architectures are either (1) simple and consist of a few genomic 'islands' of differentiation made up of either clustered small effect alleles or single loci of large effect[15,16] or (2) complex and involve genome-wide polygenic patterns of differentiation[17–19]. Each of these architectures imposes unique constraints on speciation in the face of gene flow. Under the scenario of simple concentrated trait architectures phenotypic differentiation may be maintained despite gene flow but speciation is improbable unless if the locus under divergent selection also causes reproductive isolation as a by-product, e.g. in the form of 'magic traits'[20,21] or through physical linkage to loci causing reproductive isolation[22–24]. When trait architectures are complex such that expression of divergent trait values requires many small to medium effect alleles distributed over multiple chromosomes it is more likely that some of these loci are linked to genes

affecting mating compatibility, but the strength of divergent selection per locus will likely be insufficient to counteract the homogenising effects of gene flow at any one locus[25,26]. Theory suggests that this constraint can be overcome either through the initially slow build-up over thousands of generations of weak polygenic LD at many loci across the genome, followed by rapid transition to two species around genomic 'tipping points'[27,28] or through hybridisation introducing many divergent haplotypes with multiple linked SNPs at once[29,30]. Neither of these theories have robustly been tested empirically in any adaptive radiation. Furthermore, most theoretical speciation-with-gene flow models were built to explain singular speciation events where the processes of accumulation of variation and building of LD restart after every speciation event and cannot easily be applied to adaptive radiations with tens or hundreds of speciation events in very short succession (but see[29,31]).

Adaptive radiations that produce many sympatric species, whether in cichlids, Darwin's Finches or Hawaiian Honeycreepers, is associated with high dimensionality of phenotypic, ecological, and environmental variation. Hence, selection can be diverse and multifarious, and the corresponding phenotypic response must also be diverse and multidimensional. Few theoretical and even fewer empirical studies have considered the genetic architecture that allows many successive or simultaneous speciation events when multiple environmental factors impose selection pressure on multiple traits, and there is not sufficient time between successive speciation events for recovery of genetic variation through mutation[30,32]. Here, we addressed this problem in the fastest and most species-rich cichlid adaptive radiation known by conducting radiation-wide genotype-phenotype association analyses of hundreds of individuals that represent 107 species and all but the smallest ecological guild in the Lake Victoria radiation (Fig. 1).

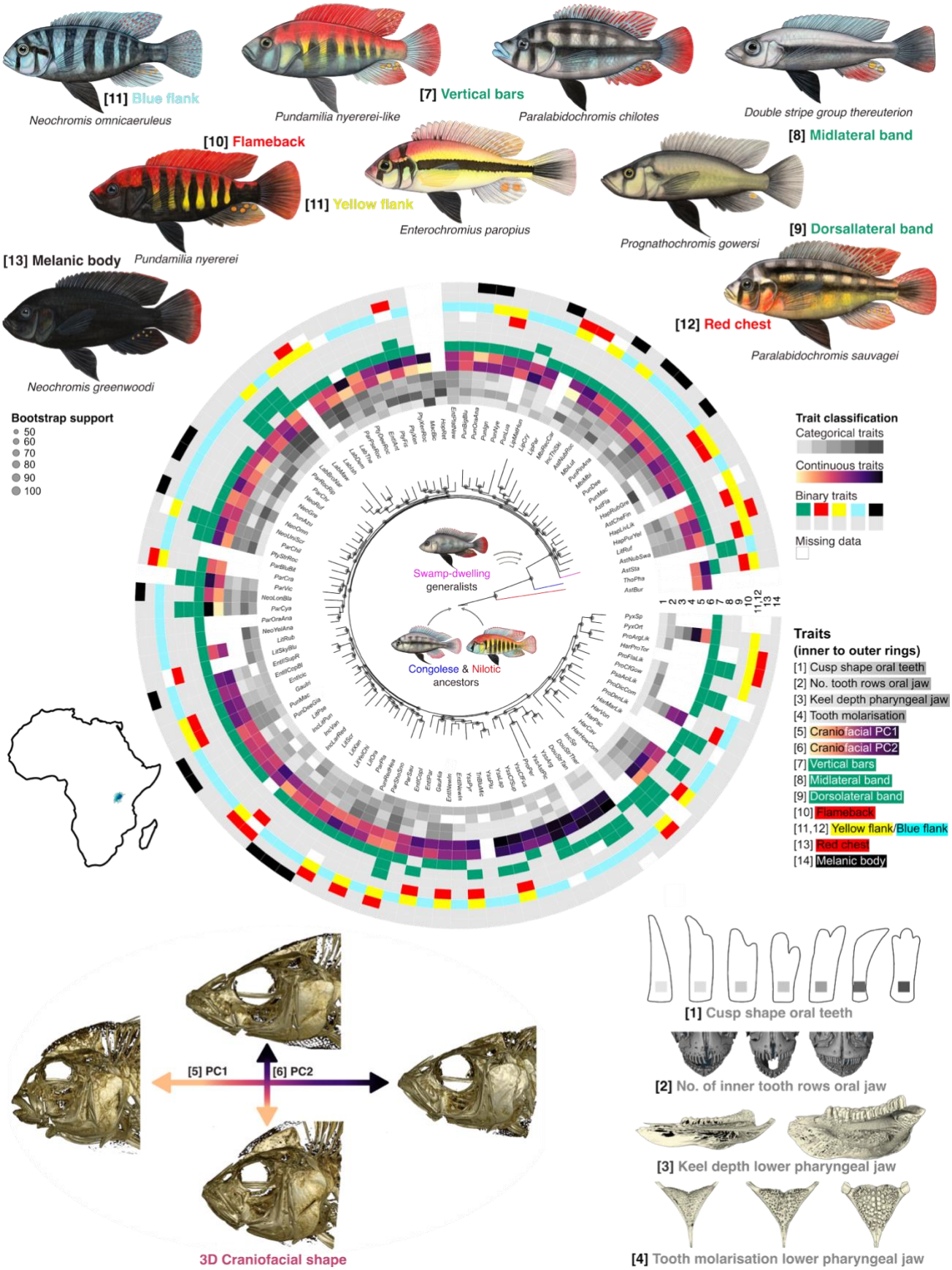
[11] Blue flank
Neochromis omnicaeruleus
Pundamilia nyererei-like
[7] Vertical bars
Paralabidochromis chilotes
Double stripe group thereuterion
[10] Flameback
[8] Midlateral band
[11] Yellow flank
Enterochromius paropius
Prognathochromis gowersi
[9] Dorsallateral band
[13] Melanic body
Pundamilia nyererei
[12] Red chest
Paralabidochromis sauvagei
Neochromis greenwoodi
Bootstrap support
50
60
70
80
90
100
Trait classification
Categorical traits
Continuous traits
Binary traits
Missing data
Swamp-dwelling generalists
Congolese & Nilotic ancestors
Traits
(inner to outer rings)
[1] Cusp shape oral teeth
[2] No. tooth rows oral jaw
[3] Keel depth pharyngeal jaw
[4] Tooth molarisation
[5] Craniofacial PC1
[6] Craniofacial PC2
[7] Vertical bars
[8] Midlateral band
[9] Dorsolateral band
[10] Flameback
[11,12] Yellow flank/Blue flank
[13] Red chest
[14] Melanic body
[5] PC1
[6] PC2
3D Craniofacial shape
[1] Cusp shape oral teeth
[2] No. of inner tooth rows oral jaw
[3] Keel depth lower pharyngeal jaw
[4] Tooth molarisation lower pharyngeal jaw

Fig. 1 Overview of ecological and mating trait evolution in Lake Victoria cichlids.
The Lake Victoria cichlid radiation encompasses ~500 known species (107 mostly sympatric species shown in the phylogeny) that represent unique phenotypic and genetic mosaics of recurrent ecomorphologies, melanic stripe patterning, and male nuptial colour motifs (illustrated schematically), assembled in <16,000 years through combinatorial mechanisms. Distribution of 14 core traits used in this study are shown in rows around the phylogeny in such a way that each column ray around the circle tree represents the unique set of trait combinations present in the species at the corresponding tip of the tree. Ecomorphological traits are numbered from 1 – 6: cusp shape of outer row frontal teeth [1] and number of inner tooth rows in oral jaws [2], lower pharyngeal jaw keel depth [3] and molarisation of lower pharyngeal jaw teeth [4], as well are 3-dimensional (3D) craniofacial shape PC1 [5] and PC2 [6]. Melanic stripe patterning traits are numbered from 7 – 9: vertical bars [7], midlateral band [8] and dorsolateral band [9], the latter only occurs in combination with [8]. Male nuptial colour motifs are numbered from 10 – 13: flameback [10], yellow or blue flank [11,12], red chest [13], and melanic body [14]. Traits 1 – 4 are categorical (see Fig. S1 for all categories), traits 5 and 6 are continuous, and traits 7 – 14 are coded as binary (presence/absence). Node bootstrap support and key ancestral admixture events (from 10, 11) are shown.

The phenotypic modules of adaptive radiation

If admixture variation from the hybrid swarm origin of the radiation was important to permitting exceptionally fast onset of speciation and adaptive radiation when the new lake formed, and if continued syngameon conditions in the emerging radiation were key for maintaining the momentum through many successive speciation events, we expect a pan-radiation genome panel to uncover the recycled component of the genetic architecture of key traits associated with repeated use of the same alleles for the same traits across the radiation, albeit in different trait combinations in different species[33]. We explored this expectation with a Radiation-Wide genotype-phenotype Association Study, denoted RWAS, by generating a long-read improved reference genome assembly and Iso-seq based gene annotation (Table S1), and genomes and phenotypes of 349 individuals from 107 different species (File S1). We focused on 14 key traits (Fig. 1) that make up three complementary trait complexes: (1) ecomorphology - we quantified craniofacial shape as continuous variation along major axes of a Principal Component Analysis based on 3D geometric morphometrics from high resolution CT scans; we categorised shape and organisation of teeth in the oral jaws, as well as pharyngeal jaw shape and tooth molarisation each into 5 to 7 categories (Fig. S1); (2) melanic stripe patterning - presence/absence of vertical bars, a midlateral band, and a dorsolateral band (only occurs in conjunction with midlateral band); (3) male nuptial colour motifs - we scored presence/absence of all known recurrent Lake Victoria cichlid male nuptial colour motifs[34,35]: blue flank versus yellow flank, flameback (red-dorsum), red

chest (red-ventrum), and melanic body. Craniofacial shape, pharyngeal and oral jaw shape and dentition are correlated with trophic ecology[36,37], melanic stripe patterning is thought to play a role in adaptation to different habitats and shoaling versus territorial behaviour[34] and male nuptial colouration is of critical importance in female mate choice, male territoriality and thus, species persistence in sympatry[38–41]. Together, these 14 traits capture major aspects of the interspecific variation that characterises the Lake Victoria cichlid radiation. To the extent that they may be freely recombined with each other, a very large number of unique species phenotypes could be generated just from this trait set. A few pairs of traits are correlated across the radiation but most of them are indeed evolutionarily independent of each other, with a mean absolute Spearman's ρ correlation coefficient of 0.27 across all traits (Fig. 2a). Noteworthy positive correlations exist between: yellow flank and flameback, whereas blue flank is exclusive of yellow flank and negatively correlated (but not exclusive) with flameback; midlateral and dorsolateral band, the latter does not occur without the former; lower pharyngeal jaw keel depth and tooth molarisation. Noteworthy negative correlations exist between craniofacial shape PC1 (elongation of head and snout and upwards/downwards orientation of mouth) and number of oral tooth rows (a high number of tooth rows is seen only in fish with downwards opening mouths); and between midlateral band and vertical bars. All of these recover well-known between-species trait associations[36,42,43]. That the other traits are only very mildly correlated across the phylogeny of this radiation is remarkable as we shall discuss later.

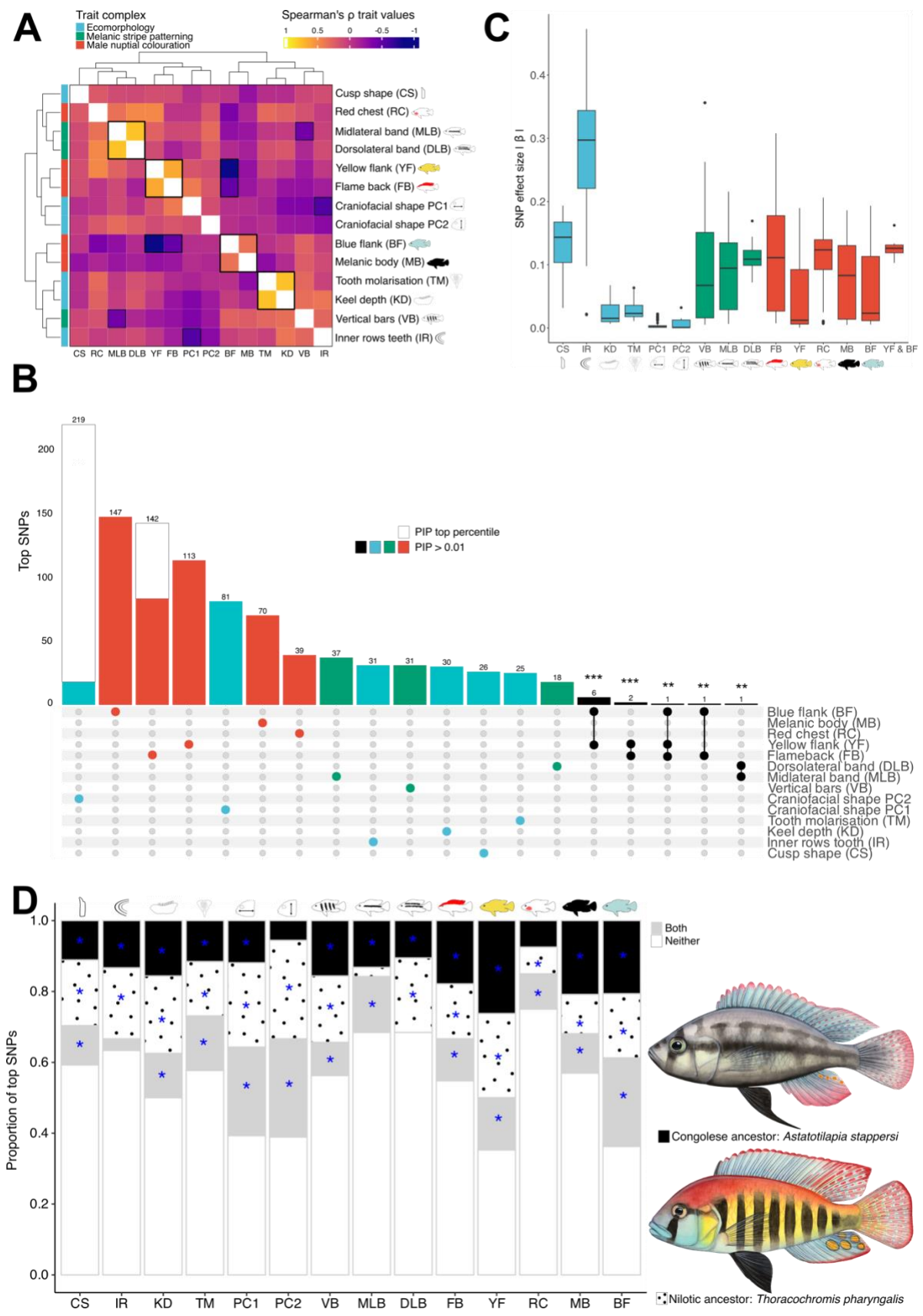
A
Trait complex
Ecomorphology
Melanic stripe patterning
Male nuptial colouration
Spearman's ρ trait values
1 0.5 0 -0.5 -1
Cusp shape (CS)
Red chest (RC)
Midlateral band (MLB)
Dorsolateral band (DLB)
Yellow flank (YF)
Flame back (FB)
Craniofacial shape PC1
Craniofacial shape PC2
Blue flank (BF)
Melanic body (MB)
Tooth molarisation (TM)
Keel depth (KD)
Vertical bars (VB)
Inner rows teeth (IR)
CS RC MLB DLB YF FB PC1 PC2 BF MB TM KD VB IR
C
SNP effect size | β |
CS IR KD TM PC1 PC2 VB MLB DLB FB YF RC MB BF YF & BF
B
Top SNPs
PIP top percentile
PIP > 0.01
219 147 142 113 81 70 39 37 31 31 30 26 25 18 6 2 1 1 1
Blue flank (BF)
Melanic body (MB)
Red chest (RC)
Yellow flank (YF)
Flameback (FB)
Dorsolateral band (DLB)
Midlateral band (MLB)
Vertical bars (VB)
Craniofacial shape PC2
Craniofacial shape PC1
Tooth molarisation (TM)
Keel depth (KD)
Inner rows tooth (IR)
Cusp shape (CS)
D
Proportion of top SNPs
Both
Neither
CS IR KD TM PC1 PC2 VB MLB DLB FB YF RC MB BF
Congolese ancestor: Astatotilapia stappersi
Nilotic ancestor: Thoracochromis pharyngalis

Fig. 2 Radiation-wide association analyses (RWAS) reveal modular nature of trait architectures and their admixture history.

(**a**) Spearman's ρ correlation of phenotypes across all species and traits in this study. Some pairs of traits show co-variation across species observed in nature (see text), but most traits do not covary strongly. (**b**) The genetic architecture of all traits is polygenic as demonstrated by the number of top RWAS SNPs significantly associated with each trait (PIP > 0.01 and top 99.9th or 99.99th percentile). Little overlap of top SNPs among traits is interpreted as low pleiotropy (Hypergeometric test *p* values denote significant overlaps: $p < 0.001$***, $p < 0.01$**). (**c**) The absolute effect size (|β|) distribution of top SNPs (PIP > 0.01) for each trait shows some medium to larger effect SNPs and many small effect SNPs. (**d**) Most trait-associated top RWAS SNPs (PIP > 0.01) are also found in the closest living relatives of the Congolese and Nilotic ancestors of the Lake Victoria region superflock. Their origins hence predate not only the Lake Victoria cichlid radiation but the entire Lake Victoria Region species flock which includes radiations in Lakes Edward, Albert and Kivu. We tested whether the number of SNPs found in the ancestral lineage categories is overrepresented (observed value exceeds 99th percentile of 100 iterations of randomly sampled SNPs across the genome; significant groups are indicated with an asterisk (*).

The genomic architecture of ecological and mating traits is highly modular

Our RWAS across 349 individuals revealed that all 14 key traits were polygenic but varied in the complexity of recurrently used variants ranging from 18 to 219 significantly associated SNPs with an effect on the trait (top percentile of PIP for each trait, as the commonly used PIP > 0.01 threshold does not account for complex trait architectures that are not recurrently used across the radiation, which is likely true for craniofacial PC1 and flameback) (Fig. 2b). Craniofacial shape and male nuptial colouration traits had the greatest number of top SNPs, followed by body patterning, dentition and pharyngeal jaw morphology. In an RWAS it is likely that we detect fewer top SNPs for more complex and redundant trait architectures, with many alleles of small effect, different subsets of which can be used in each species (i.e. less recurrent use across the radiation). This is probably why we find more top SNPs for male nuptial colour motifs than for most ecomorphological traits (Fig. 2b). This is especially true for craniofacial shape, summarised in our PC axes, is presumably a set of complex traits in itself (Fig. 1, Fig. S2). The effect sizes of top SNPs varied within and across ecological and mating trait complexes (Fig. 2c). As expected, the more quantitative traits such as craniofacial shape PCs, pharyngeal keel depth and tooth molarisation were entirely determined by alleles of small effect, while more categorical dentition traits such as the number of inner tooth rows in the oral jaw and cusp shape of outer row oral teeth had some of the largest effect sizes. Melanic stripe patterning and male nuptial colour motifs were characterised by small to medium effect alleles (Fig. 2b).

Surprisingly, there was little overlap of top SNPs across traits (Fig. 2b) and this overlap did not increase substantially when we considered overlaps of genes containing top SNPs (Fig. S3). This suggests low pleiotropy across ecological and mating traits such that variation in these traits can be largely independently inherited. Such modular architectures may have been important in allowing many different trait combinations to evolve in very little time, permitting the differentiation of this lineage into hundreds of different and phenotypically unique species. Those cases of pleiotropy that we did find are consistent with what is known from patterns of co-variation among species in nature and species crosses in the lab, particularly among nuptial colour motifs and among stripe pattern elements (Fig. 2a). For instance, the largest overlap were six top SNPs shared between Yellow flank (YF) and Blue flank (BF) (hypergeometric test *p-value* < 0.001), which are divergent and mutually exclusive male nuptial colours in the iconic red top and blue *Pundamilia* and *Pundamili*a-like sister species (Fig. 2b) and inversely correlated across the radiation (Spearman $\rho = -0.8$; Fig. 2a). These six SNPs that had opposite effect sizes on these traits (+β on BF vs -β on YF) mapped to *clul1*, a cone photoreceptor gene that is differentially expressed in light and dark-adapted retinas[44], and *pdfgrb*, a vertebrate colouration gene[45]. Male nuptial colouration and the ability to visually detect it are expected to be correlated[46]. Interestingly, we found *clul1* expressed in the eye transcriptomes of *P. pundamilia* that has blue flank male nuptial colour and blue-shifted colour perception to live in light-dark contrasting rock crevices in very shallow water[46]. This gene was not expressed in the eye transcriptome of its sister species *P. nyererei* that lives in deeper waters, has flameback and yellow-flank male nuptial colour and red-shifted colour perception[46]. It is possible that this SNP which sits in the intron of *clul1* may disrupt its expression. Other relevant cases of pleiotropy at the SNP and/or gene level were between Midlateral band (MLB) and Dorsolateral band (DLB) – the latter is never expressed in Lake Victoria cichlids without the former; Vertical bars (VB) and MLB – the occurrence of which are negatively correlated; and lower pharyngeal jaw tooth molarisation and keel depth, which are strongly positively correlated [36]. In summary, the few trait correlations that we observe (Fig. 2a) seem at least partly caused by pleiotropic gene action, but there is no evidence for pleiotropic gene action on other trait combinations. For downstream analyses only top RWAS SNPs with PIP > 0.01 as recommended by [47].

Ancient variation underlies key trait modules for adaptive radiation

Despite the extraordinarily young age of the radiation, Lake Victoria cichlid fish span over three trophic levels in the food web, have evolved highly specialised trophic adaptations and complex trophic species interactions as well as a large array of body colouration motifs and stripe patterns[36,48,49]. To better understand the role that resuspension of ancestral admixture variation may have played in enabling this process, we checked how many top RWAS SNPs with statistical effects on craniofacial shape, dentition, pharyngeal jaw morphology,

male nuptial colour motifs, and body stripe patterns could be mapped to the extant Congolese and Nilotic haplochromine species that are the closest living relatives of the species that hybridised ~150,000 years ago to give rise to the Lake Victoria region super-flock of radiations[10]. We found that at least ~30-50% of the top SNPs were likely present in at least one of the ancestral lineages (because they are present in their closest living relatives; Fig. 2d). Genetic variation underlying all traits except midlateral band had significantly higher contributions of Nilotic ancestor-derived variants than expected by chance (>99th percentile of 100 iterations of randomly sampled SNPs across the genome); the same was true for the Congolese ancestor for all traits with the exception of red chest and craniofacial shape PC2 traits. Interestingly, the midlateralband trait is found in the Congolese ancestors of this radiation but not in the Nilotic ancestor (Fig. 2d fish drawings), which is consistent with many of the alleles associated with this trait being significantly overrepresented in *Astatotilapia stappersi*, including alleles in the *agouti-related protein* (*agrp*) gene that has been shown before to control the presence/absence of midlateralband in cichlids[50]. Top SNPs that could not be assigned to either ancestor were not significantly overrepresented in the top SNPs for any trait, which implies that old (ancestral) Congolese and Nilotic SNPs are enriched among the top SNPs associated with key adaptive radiation traits. It is likely that some of the alleles that were carried into the ancestral hybrid swarm have since been lost in the closest living relatives of the ancestors that we sequenced, and we will have missed others in our limited sampling of these lineages compared to the large number of Lake Victoria cichlids that we sequenced. 30-50% of key SNPs derived from admixture is hence likely an underestimate. The low pleiotropy among traits that we found above would also be consistent with key SNPs for different traits being old, selection-tested variants derived from admixture variation.

Candidate genes controlling ecological and mating traits are dispersed across the genome

For all 14 traits we identified many genic as well as non-coding SNPs dispersed widely across the genome, including in well-known genes and previously unknown candidate genes (Fig. 3a shows results for five traits, see Fig. S4 for all 14 traits). Most strikingly, both **midlateral** and **dorsolateral band** shared an intergenic top SNP on chromosome 5 that was highly associated with their presence (MLB β=0.15, PIP = 0.74 and DLB β = 0.13, PIP = 0.99). Dorsolateral band never occurs in this radiation without the midlateral band, and this SNP may regulate this interaction. The top genic SNP for midlateral band with an effect size of β = 0.12 (PIP > 0.48) was in the *agrp* (*asip2*) gene that is known through QTL analysis and Crispr-cas9 as a genetic switch for midlateral band formation in cichlids[50], as mentioned above. A top genic SNPs for dorsallateral band was in *cyp27a1*, known as a vision gene[51] and the top SNP for **vertical bars** mapped to *sez6* that is interestingly associated with exploratory behaviour and cognition in mice[52].

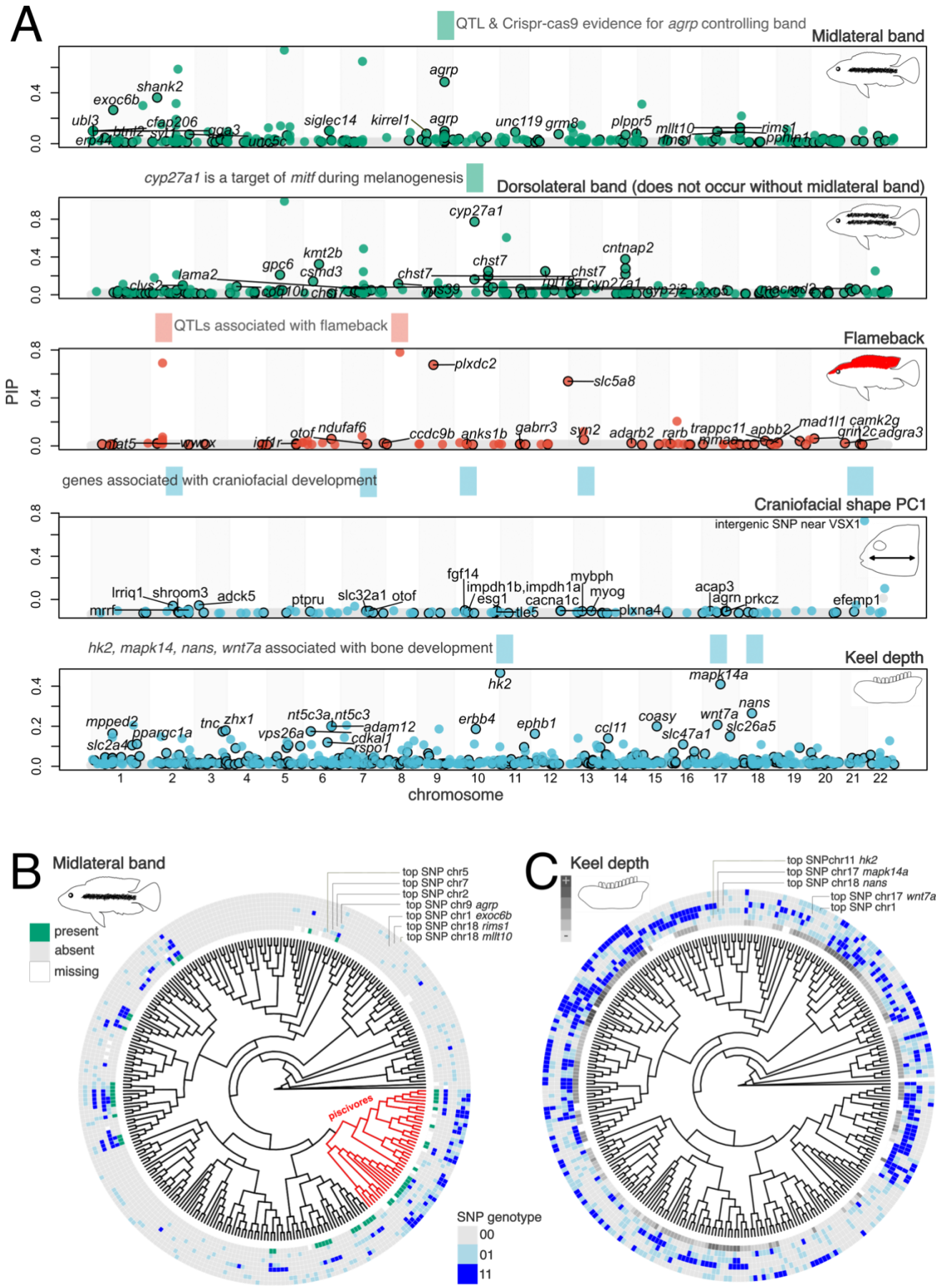
A
QTL & Crispr-cas9 evidence for agrp controlling band
Midlateral band
cyp27a1 is a target of mitf during melanogenesis
Dorsolateral band (does not occur without midlateral band)
QTLs associated with flameback
Flameback
genes associated with craniofacial development
Craniofacial shape PC1
intergenic SNP near VSX1
hk2, mapk14, nans, wnt7a associated with bone development
Keel depth
PIP
chromosome
B
Midlateral band
present
absent
missing
top SNP chr5
top SNP chr7
top SNP chr2
top SNP chr9 agrp
top SNP chr1 exoc6b
top SNP chr18 rims1
top SNP chr18 mllt10
piscivores
SNP genotype
00
01
11
C
Keel depth
top SNPchr11 hk2
top SNP chr17 mapk14a
top SNP chr18 nans
top SNP chr17 wnt7a
top SNP chr1

Fig. 3 Genomically dispersed and redundant trait architectures in Lake Victoria cichlids.
(**a**)Top RWAS SNPs repeatedly associated with key adaptive radiation and speciation traits in the Lake Victoria radiation are dispersed across the genome. Significant top SNPs ($PIP_{0.01}$) are highlighted as coloured circles and non-significant SNPs are denoted by grey circles (barely visible at the bottom of the Y axes). Colour of circles are categorised by trait complexes, coloured coded as in Fig. 2: blue – ecomorphology, green-melanic stripe pattern, red-male nuptial colour. Coloured circles with black outline are SNPs located in genes and coloured circles without black outline are SNPs located in intergenic regions. For each trait the top 20 genic SNPs are annotated with gene names - full list can be found in File S1. Functional evidence linking candidate genes to traits from previous genetic mapping and gene knock-out studies (cited in main text) for each trait is illustrated above each Manhattan plot (see Fig S4 for all 14 key traits). (**b**) Two examples of evolutionary genotypic redundancy in key traits: midlateralband and lower jaw keel depth, where different top RWAS SNPs/genes code for the same trait in different species and/or sub-clades in the radiation.

The four highest top SNPs for flameback include two **flameback** associated QTLs previously found in a cross of *P. nyererei*-like (a flameback species) and *P. pundamilia*-like (no flameback), with the region on chromosome 2 being a QTL for red dorsum and the region on chromosome 8 for red dorsal fin[53]. Candidate genes for the flameback male nuptial colour were vision (*slc5a8*[54]) and carotenoid/melanin/iridophore associated (*rarb*[55], *ahr*[56], *alkal2*[57]). The top flameback gene *plxdc2* has not been associated with pigmentation previously but plexins are implicated in melanocyte regulation[58]. Gene Ontology (GO) analysis of flameback top SNPs found enrichment ($q < 0.05$) of 'retinoic acid receptor signalling pathway' that regulates chromatophores[59], 'neurotransmitter secretion' that controls physiological colour change in vertebrates[60] and 'regulation of Toll signalling pathway' that is implicated in melanogenesis[61]. **Yellow flank** and **blue flank** are two reciprocally exclusive male nuptial colour motifs known to be repeatedly involved in speciation [35]. An intergenic top SNP on chromosome 2 for blue flnk overlaps with a QTL for yellow flank in a cross of *P. nyererei*-like (has yellow flank) and *P. pundamilia*-like (has blue flank)[53]. The strongest candidate gene for both these traits was *pdgfrb*, a gene thought to be involved in fish colouration as it evolved from a duplication of the proto-*kit/csf1r* gene when fish acquired chromatophores[45]. This may suggest that yellow and blue flank can sometimes be alternative allelic states of the same gene. Other skin/vision pigment-related candidates for yellow flank are *sox5*[62], *roraa*[63], *rarb*[55], and *clul1*[44]. Candidates for blue flank coloration are *clul1*[44] and *tead1*[64], the latter of which regulates *mitf*, a master regulator of melanocyte differentiation and animal pigmentation[65]. Blue flank top SNPs were also enriched for 'hippo signalling' that is involved in retinal differentiation[66] and melanogenesis[67]. Interestingly, a top genic SNP with positive effect on male **melanic body** colouration mapped to *nos1ap*, a gene associated with social status switches in cichlids[68]. Melanic body

had melanogenesis associated candidates *trpv4*[69], *wnt10b*[70], *adamts*[71]; and more broadly melanic body top SNPs were enriched for the 'Wnt signalling pathway' that is involved in the differentiation and proliferation of melanocytes, as well as the regulation of melanin production[72]. Top candidates for **red chest** male colouration were *smyd3*, a methyltransferase that regulates mesoderm development, where chromatophores are situated[73,74]. Overall, gene associations for male nuptial colour motifs were either involved in pigmentation or vision. The latter observation supports the idea that sexual selection by female mate choice and male-male competition contribute to speciation and species differentiation in Lake Victoria cichlids[38,42,75]. Male nuptial colouration can be associated with water depth and ambient light conditions in cases of speciation[46,76] and the evolution of female preferences for male nuptial colouration likely involves interactions between sexual selection[77], ecological adaptation in the visual system to water depth/light[46], frequency dependent selection by male-male competition[41] and possibly other components of behaviour that remain to be identified.

We found many significant SNPs/genes associated with craniofacial shape **PC1** and a few for PC2 (Fig. 3). The strongest predictor of a species' position on craniofacial shape PC1 (the bite force versus suction feeding continuum, with short vs longer heads and jaws) is a top SNP on chromosome 21 located near the *vsx* (alias *rinx1*) gene that is involved in craniofacial anomalies and modulating interpupillary distance in humans [78]. Interestingly, *lrriq* gene that is physically linked to *alx1* is an important locus for beak shape in Darwin's finches and great tits [79,80] and also one of the top candidate genes for craniofacial PC1 in our analysis (chromosome 2). Another candidate gene for PC1, *Efemp1* controls head length in zebrafish[81]. Other candidates known to be involved in cichlid bone/cartilage/muscle development were *fgf14*, *mybph*, *col8a2*[82] and the sex-specific development gene *foxo3*[83]. The intergenic top SNP for **PC2** on chromosome 20 overlaps with a QTL for preorbital depth in a cross between *Neochromis omnicaeruleus* and *P. nyererei*-like [84]. A promising candidate for shape variation along **PC2** (body depth, cheek depth and eye size continuum) is *roraa* that is involved in orofacial cleft phenotype in humans[85]. The top candidate genes for **cusp shape** were *macf1*, which controls tooth mineralisation in mice[86]; *lrp6* that interacts with the wnt pathway during tooth development in humans[87] and *prdm16* that is required for normal tooth development in mice[88]. *Ugt1a* was the strongest candidate for **number of inner tooth rows** in the upper oral jaw. This gene has not previously been associated with dentition. Another candidate gene, *fbp2*, overlaps with a QTL on chromosome 22 for this trait in a cross between *N. omnicaeruleus* (3 to 8 inner rows of teeth) and *P. nyererei*-like (2 inner rows of teeth) [84]. Lower pharyngeal jaw **keel depth** top SNPs mapped to many bone development genes such as *hk2*, *mapk14a*, *wnt7a* and *nans*. The latter is required for human bone development in several craniofacial regions, including the pharyngeal arch skeleton [89]. The top candidate gene for **pharyngeal tooth molarisation** was *vmp1*, which has not been previously associated with tooth development. However, another candidate for this trait was *pknox2*, a transcription factor that plays a crucial role in the early stages of tooth development in mammals [90].

Convergently evolved traits are genotypically redundant across the radiation

Convergent and parallel evolution of the same trait by the same gene has fascinated biologists, but one alternative to this model is the evolution of the same trait by different genes, denoted genotypic redundancy[91–93]. Genotypic redundancy through gene duplicates has been viewed as a substrate for evolutionary change and evolvability[94,95], but few empirical examples exist exploring genotypic redundancy among different genes, this is partly because redundancy is thought to be evolutionarily unstable[26] making it inherently challenging to map in natural systems. Furthermore, genotypic redundancy has rarely been considered in the context of speciation and adaptive radiation. To answer Kafri's[94] "why have two when one is enough?" question on the role of genotypic redundancy in evolution, we asked if redundancy exists in trait architectures in the Lake Victoria cichlid radiation. We mapped the genotypic evolution of top SNPs for midlateralband (Fig. 3b) and keel depth (Fig. 3c) across the phylogeny and discovered that non-piscivorous species with midlateralbands almost always have homozygous *agrp* alleles but *agrp* alleles are absent in the piscivore lineage many species of which also have a midlateralband (Fig. 3b). It is possible that the evolution of this trait is driven by *exoc6b* in the piscivores, which segregates almost perfectly with the presence of the midlateralband in heterozygous or homozygous state in piscivores. *Exoc6b* plays a crucial role in the transfer of melanin from melanocytes to keratinocytes in *in vitro* experiments[96]. Interestingly, Meier et al[11] showed that the Lake Victoria piscivores are slightly but significantly enriched for admixture variation of Nilotic ancestry. It is hence possible that genetic redundancy at this trait (and likely others) was generated by the combination of ancestral gene pools into the hybrid swarm ancestor of the radiation. We found similar patterns of genotypic redundancy among species with greater pharyngeal keel depths in the radiation where *hk2/mapk14a* and *nans/wnt7a* appear to function redundantly, and even additively in some species (Fig. 3c). Given the repeated cycles of hybridisation at the base of the Lake Victoria cichlid radiation[10,11], it is likely that genotypically redundant and additive alleles for the same traits arrived from different ancestral lineages from across the African continent (Fig. 2d). Such redundant genetic architectures may be a previously underappreciated factor that unconstrained the simultaneous and rapid evolution of many different species with different combinations of a limited set of similar traits in this adaptive radiation – which also offers one possible answer to Kafri's question[94].

Combinatorial speciation maintains interchromosomal linkage disequilibrium among dispersed trait modules

To test predictions of the process of combinatorial speciation that we hypothesise to be responsible for the rapid emergence of several hundred distinct species in record time, we combined RWAS information on the

genetic architecture and admixture history of recurrent allele use underlying major traits of adaptive radiation with estimates of the genomic landscape of speciation and species differentiation as inferred from genome wide divergence (F_{ST}) among closely related sympatric species (Fig. S6). A unique prediction of the hypothesis of adaptive radiation through combinatorial speciation (combinatorial adaptive radiation) is that each species pair in the radiation is characterised by unique patterns of linkage disequilibrium (LD) among different subsets of reused alleles[97]. Here we set out to perform the first test of this hypothesis in any adaptive radiation. We calculated F_{ST} landscapes and determined differentiation outlier loci (top 5% F_{ST}) for 12 different sympatric sister species pairs in the Lake Victoria radiation. We then annotated F_{ST} outlier loci with top SNPs from our radiation-wide RWAS to identify for each species pair the subset of 'speciation loci' (defined as loci displaying significantly stronger differentiation than the rest of the genome, in full sympatry) that are recurrent among speciation events in the radiation and tested for LD among these loci. Firstly, we found that in most sympatric sister species contrasts LD among the 'speciation loci' (top RWAS SNPs that are also top 5% F_{ST}) was significantly higher compared to their respective genomic background LD (wilcox.test $p < 0.001$) (Fig. 4a). There was a discernible build-up of LD in older species pairs such as *Pundamilia* on Makobe and Ruti islands compared to younger species pairs of *Pundamilia* on Python and Kissenda islands (Fig. 4a), that emerged much more recently through hybrid parallel speciation [98].

Next, we compared the distribution of these 'speciation loci' across the genome and LD between them for each species-pair (Fig. 4b, Fig. S7). We find that loci contributing to ecological adaptations and reproductive isolation are not physically linked in the genome but became coupled (presumably) during or shortly after speciation, with varying degrees of LD association among them. For example, between *P. nyererei* and *P. pundamilia* both trophic ecology (reef planktivore with fine pharyngeal bones and teeth vs crevice-dwelling insectivore with heavier pharyngeal bones and teeth) and male nuptial colour (flameback on yellow flank vs entirely blue flank) diverged, and this is reflected in the high LD between SNPs associated with these traits (Fig. 4b left). Similarly, *E. paropius* vs *E. coprologous* 'blue' are deeper water detritivores that diverged in male nuptial colouration (flameback on a light yellow flank vs blue flank with melanic underside), divergent melanic body patterning (present vs absent midlateral band), and subtle differences in trophic morphology (very fine pharyngeal bones and teeth vs slightly heaverier pharyngeal bones and teeth) which is reflected in LD among specific SNPs associated with these traits (Fig. 4b right). Linkage disequilibrium between trait complexes in full sympatry suggests non-random mating among individuals with these divergent traits (i.e. speciation). Interestingly, we found that different and unique subsets of top RWAS SNPs were diverging in different species pairs, even among phenotypically similar species pairs (see *Pundamilia* species pairs from Makobe and Ruti versus Python and Kissenda, Fig. 4b, S7), reinforcing the idea that genotype-phenotype mapping of trait architectures can be redundant in this radiation. The unique combination of LD patterns

among subsets of reused alleles in different species pairs that we observe provides support for the hypothesis of adaptive radiation through combinatorial speciation[97].

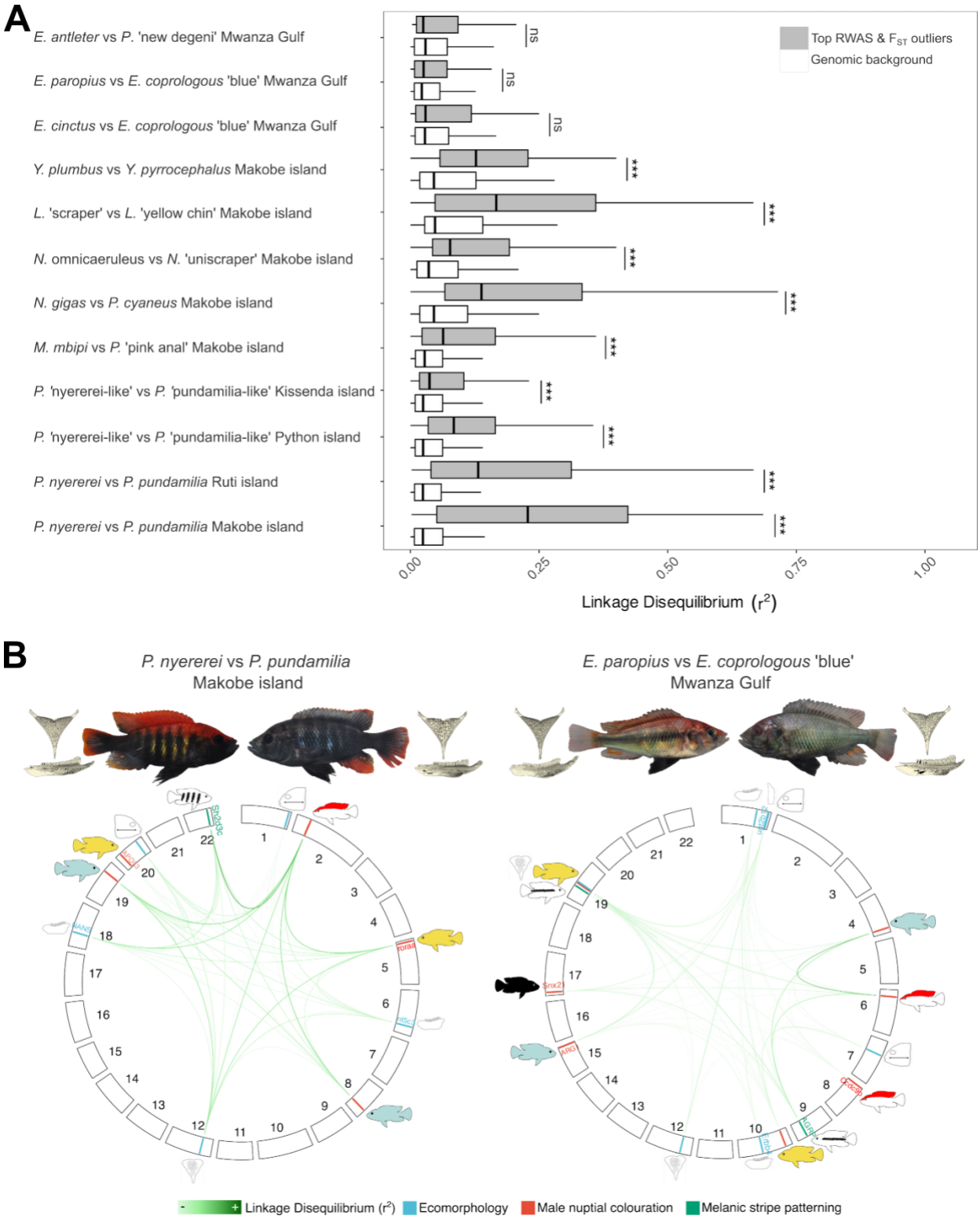

Fig. 4 Coupling of ecological and mating trait modules during combinatorial speciation.

(**a**) Linkage disequilibrium (LD; measured using r^2 statistic) among top RWAS SNPs ($PIP_{0.01}$) that are also top 5% F_{ST} outliers between sympatric sister species (grey) compared to the genomic background LD (white). Top SNP vs background LD comparisons are significantly higher (wilcox.test $p < 0.001$***) for all species pairs except demersal detritivores in the genus *Enterochromis*. (**b**) Genome-wide distribution of LD coupled alleles for two sympatric species pairs from (a), one with and one without significant differences in LD, illustrates unique pattern of high interchromosomal LD among SNPs associated with ecological and mating traits uniquely divergent among the species in each pair, consistent with expectation for combinatorial speciation. The plots for the remaining species pairs are shown in Fig. S7. Lines on the ideogram represent a top RWAS SNP ($PIP_{0.01}$ and top 5% F_{ST}); line colour denotes the trait complex the SNP was associated with in the RWAS: blue = ecomorphology, green = melanic stripe pattern, red = male nuptial colour motif as in Fig. 2; SNPs located in genes are annotated with gene names. Higher linkage disequilibrium (r^2) values are depicted by darker green colours – see figure legend with scale. Only SNPs on different chromosomes and SNPs more than 100 Kbp away on the same chromosome were included in the analysis to exclude effects of physical linkage in all LD analyses. Even in a in which the LD among all top SNPs is not significantly elevated over the genomic background of *Enterochromis* sister species, the highest LD patterns among top SNPs map to trait combinations that differentiate these species.

Discussion

Adaptive radiations have repeatedly punctuated the evolutionary history of biodiversity on Earth through bursts of diversification that seemingly defy expectations of gradual evolution[99]. In a classical article back in 1954 Anderson & Stebbins proposed that bursts of diversification can be enabled by hybridisation between species[7], a proposition that received support from botanists but long-standing criticism from zoologists[100]. Yet, cichlid fish adaptive radiations in East African lakes have presented a long-standing puzzle that zoologists failed to resolve: how did hundreds of ecologically and morphologically diverse species arise without geographic isolation and come to coexist in species rich assemblages within just a few thousand years? Here, we demonstrate that combinatorial speciation from genetic modules of hybrid ancestry was of great importance in this process. We found that countless successive events of rapid speciation in the fastest of these radiations were associated with the build-up and persistence in full sympatry of long-range linkage disequilibrium, presumably in response to multidimensional selection, across modular, polygenic, and redundant trait architectures between ancient genetic variants polymorphisms amongst which arose from hybridisation in the ancestry of this cichlid lineage. Combining ecological and mating trait modules in many different ways, like Lego pieces, or as Anderson & Stebbins put it, elements from assembly lines of several

car models co-produced in the same factory, made it possible for so many diverse species to evolve from a finite set of elements.

Our association analysis revealed that all key traits associated with ecological diversification and speciation in the LV cichlid fish adaptive radiation are polygenic and governed by a combination of some modest to larger effect alleles and many small effect alleles dispersed across the genome with considerable redundancy. This part of our findings is consistent with what is known about complex genetic architectures underlying most traits from quantitative genetics[101]. Our analysis further revealed that the genetic architecture of speciation/species divergence is physically unlinked and non-pleiotropic. It requires long range LD to build and be maintained between multiple traits each of which has polygenic, redundant and mixed effect size architecture. This is consistent with earlier studies of genome-wide divergence patterns in pairs of sympatric species at early stages of speciation[17,102]. Thus, our findings reconcile two initially disparate viewpoints on the genetic basis of quantitative traits and the genetic architecture of speciation[18,33]. The low pleiotropy and absence of physical linkage across adaptive radiation trait architectures may be surprising at first. However, pleiotropic or clustered (i.e. physically linked by inversion polymorphisms) architectures, while facilitating adaptation to correlational selection along environmental gradients, would constrain diversification beyond a finite number of ecotypes or species, as has been observed in some Crater Lake cichlids, stickleback, and to some extent in Darwin's finches[15,80,103,104]. Such architectures may also facilitate shared adaptation between lineages colonising the same environmental gradients in species-rich adaptive radiations, as has been demonstrated for large inversions linking co-adapted alleles in Lake Malawi cichlids that have been exchanged by hybridisation between clades sharing the same environment[105]. However, for large phenotypic and species complexity to evolve rapidly, with many species coexisting and sharing resources and mates in each habitat, it seems advantageous if genetic variants affect only one trait at a time and are independently inherited[106]. The modular trait architectures in both mating traits and ecological traits that we find in Lake Victoria cichlids may therefore be key to promoting diversification into many phenotypically diverse species within the same and in different environments, as modules can combine in countless different ways in different species[97] as demonstrated here. This is consistent with what we observe phenotypically in Lake Victoria cichlids, where many different combinations of essentially the same set of craniofacial traits, dentition traits, body stripe patterns, and male nuptial colour motifs exist (Fig. 1)[36]. While we focus on SNPs here, it is very likely that small inversions or other small structural variants contribute to modular trait architectures too (McGee et al[49]). We show that additive and redundant genetic variation in trait modules is derived (and enriched) from ancient admixture variation, which may explain the low pleiotropy among trait alleles. Low pleiotropy would be unexpected if many of these variants had emerged from *de-novo* mutation in the radiation. Equally important though is that our data do not uncover any evidence for physical linkage

among phenotypic speciation and adaptive radiation modules. Hence, the inversion polymorphisms that almost certainly segregate in this radiation[105,107] did not capture major genetic variants underlying mating and adaptation traits. Finally, the compensatory phenotypic effects of genetically redundant alleles add additional layers of evolvability in this adaptive radiation[33,91]. Simulations have shown that in the presence of admixture variation, polygenic and redundant trait architectures would promote recurrent speciation events and adaptive radiation [30].

Now that we have identified the genetic architecture of species differentiation in the fastest known adaptive radiation as building on recombinant use of ancient variation in polygenic and modular traits, questions that remain to be answered are the nature of the forces that generate and maintain long range linkage disequilibrium between these genetic modules in full sympatry? As important as the absence of strong genetic correlations between adaptive radiation traits is, the absence of strong ecological correlations between these traits, which we can also infer from our trait correlation matrix (Fig. 2A), is equally important. We argue this implies the absence of any consistent selection constraint on trait associations. We show that trait values at ecological and mating traits can be combined in many ways and that alleles associated with these two categories of traits in any specific sympatric species pair are in elevated long-range LD. We interpret this as a signal of non-random mating among sympatric species. In the presence of gene flow, the strength of divergent selection required to maintain LD increases with the number of loci involved which may be at least partly offset by non-random mating. Clearly, polygenic LD among multiple traits, including traits that are associated with mate choice, persists in sympatry without physical linkage. Future work will have to address how it arises from no LD in the first place. One possibility is initial build-up of LD among ecological and mating traits happens in geographical or ecological parapatry, and that the extent of assortative mating generated in this way becomes quickly sufficient for incipient species to persist as they reinvade each other's ranges and form sympatric species communities. Our findings reconcile quantitative genetics of traits with population genomics of speciation and have broad implications for the process of adaptive radiation and its constraints.

Data and Code availability

Pundamilia nyererei v3 genome assembly and annotation and the raw PacBio Hi-C and PacBio Iso-seq data is deposited to the NCBI SRA (BioProject PRJNA1231206). Individual resequencing data is sourced from [11,49,109] and all phenotyping information is in File S1. Scripts used in this study can be found at https://github.com/poojasingh09/2025_singh_radiation_gwas.

Author contribution

Pooja Singh: conception, lead all analyses and interpretation, and manuscript writing. Heidi Tschanz-Lischer: genome assembly and annotation. Kassandra Ford: 3D morphometrics data collection and analyses. Ehsan P. Ahi: candidate gene & GO analysis interpretation. Marcel Haesler: fish husbandry and tissue dissections. Salome Mwaiko: DNA and RNA extraction. Joana I. Meier: DNA extraction. David Marques: conception. Remy Bruggman: genome assembly support. Mary Kishe: fieldwork. Ole Seehausen: fish collection and identifications, conception, interpretation & contribution to writing.

Acknowledgements

We would like to acknowledge the support of the Tanzanian Fisheries Research Institute (TAFIRI) for allowing us access to study the amazing cichlid diversity in Lake Victoria. We are very grateful for the support of Mohammed Haluna and Mhoja Kayeba over the last three decades, without whose skilled boat driving on rocky reefs, fishing, and angling this research and the collection of Lake Victoria cichlids at our institutes in Switzerland would not be possible. We thank Arne Nolte for the DNA extraction protocol, Julie Johnson for fish illustrations, and the NGS facility at University of Bern for sequencing support. We acknowledge funding from EAWAG and the Swiss National Fund (SNF grants 163338 to OS) and University of Bern Initiator Grant to PS.

List of Supplementary Materials

Supporting data is made available in the online supplementary on an open-access basis for research use only.

Materials and Methods

Table S1

Figures S1 – S7

File S1

References

1. Neige, P. *Events of Increased Biodiversity*. (ISTE Press - Elsevier, 2015).
2. Losos, J. B. Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am Nat* **175**, 623–639 (2010).
3. Grant, P. R. Speciation and the adaptive radiation of Darwin's finches: the complex diversity of Darwin's finches may provide a key to the mystery of how intraspecific variation is transformed into interspecific variation. *Am Sci* **69**, 653–663 (1981).

4. Seehausen, O. Explosive speciation rates and unusual species richness in haplochromine cichlid fishes: Effects of sexual selection. *Adv Ecol Res* **31**, 237–274 (2000).
5. Danley, P. & Kocher, T. Speciation in rapidly diverging systems: lessons from Lake Malawi. *Mol Ecol* **10**, 1075–1086 (2001).
6. Poelstra, J. W., Richards, E. J. & Martin, C. H. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Mol Ecol* **27**, 4270–4288 (2018).
7. Anderson, E. & Stebbins Jr., G. L. Hybridization as an evolutionary stimulus. *Evolution (N Y)* **8**, 378–388 (1954).
8. Gaylord Simpson, G. *Evolution and Geography*. (Condon Lecure Publications, Singapore, 1953).
9. Martin, C. H. & Richards, E. J. The Paradox behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself Through an Early Burst? *Annu Rev Ecol Evol Syst* **50**, 569–593 (2019).
10. Meier, J. I. *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun* **8**, 14363 (2017).
11. Meier, J. I. *et al.* Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science (1979)* **381**, (2023).
12. Irisarri*, I. *et al.* Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun* **9**, 3159 (2018).
13. Singh, P. *et al.* Phylogenomics of trophically diverse cichlids disentangles processes driving adaptive radiation and repeated trophic transitions. *Ecol Evol* **12**, e9077 (2022).
14. Svardal, H. *et al.* Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol Biol Evol* **37**, 1100–1113 (2020).
15. Malinsky, M. *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (1979)* **350**, 1493–1498 (2015).
16. Nguyen, T. N., Repenning, M., Fontana, C. S. & Campagna, L. Genomic islands of speciation harbor genes underlying coloration differences in a pair of Neotropical seedeaters. *Evolution (N Y)* **78**, 1161–1173 (2024).
17. Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L. & Seehausen, O. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol Biol Evol* **35**, 1489–1506 (2018).
18. Seehausen, O. *et al.* Genomics and the origin of species. *Nat Rev Genet* **15**, 176–92 (2014).
19. Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M. & Reimchen, T. E. Genomic changes underlying repeated niche shifts in an adaptive radiation. *Evolution (N Y)* **76**, 1301–1319 (2022).

20. Servedio, M. R., Doorn, G. S. Van, Kopp, M., Frame, A. M. & Nosil, P. Magic traits in speciation: 'magic' but not rare? *Trends Ecol Evol* **26**, 389–397 (2011).
21. Kopp, M. *et al.* Mechanisms of assortative mating in speciation with gene flow: Connecting theory and empirical research. *American Naturalist* **191**, 1–20 (2018).
22. Felsenstein, J. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution (N Y)* **35**, 124 (1981).
23. Kathleen, G. F., Barnett, L. L., Blackman, B. K. & Willis, J. H. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature 2001 412:6850* **412**, 904–907 (2001).
24. Ferris, K. G., Barnett, L. L., Blackman, B. K. & Willis, J. H. The genetic architecture of local adaptation and reproductive isolation in sympatry within the Mimulus guttatus species complex. *Mol Ecol* **26**, 208–224 (2017).
25. Nosil, P., Harmon, L. J. & Seehausen, O. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24**, 145–156 (2009).
26. Yeaman, S. Local adaptation by alleles of small effect. *American Naturalist* **186**, S74–S89 (2015).
27. Barton, N. & Bengtssont, B. O. *The Barrier to Genetic Exchange between Hybridising Populations*. *Heredity* vol. 56 (1986).
28. Nosil, P., Feder, J. L., Flaxman, S. M. & Gompert, Z. Tipping points in the dynamics of speciation. *Nature Publishing Group* **1**, 1 (2017).
29. Kagawa, K. & Takimoto, G. Hybridization can promote adaptive radiation by means of transgressive segregation. (2017) doi:10.1111/ele.12891.
30. Kagawa, K. & Seehausen, O. The propagation of admixture-derived adaptive radiation potential. *Proceedings of the Royal Society B: Biological Sciences* **287**, (2020).
31. Gavrilets, S. & Vose, A. Dynamic patterns of adaptive radiation. *PNAS* **102**, 18040–18045 (2005).
32. Gavrilets, S. & Losos, J. B. Adaptive radiation: contrasting theory with data. *Science (1979)* **323**, 732–737 (2009).
33. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics 2020 21:12* **21**, 769–781 (2020).
34. Seehausen, O., Mayhew, P. J. & Van Alphen, J. J. M. Evolution of colour patterns in East African cichlid fish. *J Evol Biol* **12**, 514–534 (1999).
35. Seehausen, O. & Van Alphen, J. J. M. Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol Lett* **2**, 262–271 (1999).
36. Seehausen, O. *Lake Victoria Rock Cichlids*. (Verduyn Cichlids, 1996).

37. Barel, B., van Oijen, M., Witte, F. & Witte-Maas, E. An introduction to the taxonomy and morphology of the haplochromine cichlidae from Lake Victoria. *Netherlands Journal of Zoology* **27**, 333–380 (1976).
38. Selz, O. M., Pierotti, M. E. R., Maan, M. E., Schmid, C. & Seehausen, O. Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology* **25**, 612–626 (2014).
39. Dijkstra, P., Seehausen, O. & Groothuis, T. Direct male-male competition can facilitate invasion of new colour types in Lake Victoria cichlids. *Behav Ecol Sociobiol* **58**, 136–143 (2005).
40. Svensson, O., Woodhouse, K., Smith, A., Seehausen, O. & Turner, G. F. Sympatry and parapatry among rocky reef cichlids of Lake Victoria explained by female mating preferences. *J Evol Biol* **37**, 51–61 (2024).
41. Seehausen, O. & Schluter, D. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences* **271**, 1345–1353 (2004).
42. Seehausen, O., Witte, F., Van Alphen, J. J. M. & Bouton, N. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J Fish Biol* **53**, 37–55 (1998).
43. Seehausen, O., Van Alphen, J. J. & Witte, F. Can ancient colour polymorphisms explain why some cichlid lineages speciate rapidly under disruptive sexual selection? *Belg. J. Zoo!* **129**, 43–60 (1999).
44. Zhang, Q. *et al.* Comparative Analysis and Expression of CLUL1, a Cone Photoreceptor-Specific Gene. *Invest Ophthalmol Vis Sci* **44**, 4542–4549 (2003).
45. Braasch, I., Salzburger, W. & Meyer, A. Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. *Mol Biol Evol* **23**, 1192–1202 (2006).
46. Seehausen, O. *et al.* Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626 (2008).
47. Comeault, A. A. *et al.* Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in Timema cristinae. *American Naturalist* **183**, 711–727 (2014).
48. Greenwood, P. H. & Greenwood, P. H. The cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. *Bulletin of the British Museum (Natural History). Zoology. Supplement* **Supplement 6**, 1--134 (1974).
49. McGee, M. D. *et al.* The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**, 75–79 (2020).
50. Kratochwil, C. F. *et al.* Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science (1979)* **362**, 457–460 (2018).

51. Reem, R. E., Maeda, T., Maeda, A. & Pikuleva, I. A. Altered cholesterol homeostasis: ocular manifestations in CYP27A1 and CYP46A1 knockout mice. *Invest Ophthalmol Vis Sci* **51**, 6144–6144 (2010).
52. Gunnersen, J. M. *et al.* Sez-6 proteins affect dendritic arborization patterns and rxcitability of cortical pyramidal neurons. *Neuron* **56**, 621–639 (2007).
53. Feller, A. F., Haesler, M. P., Peichel, C. L. & Seehausen, O. Genetic architecture of a key reproductive isolation trait differs between sympatric and non-sympatric sister species of Lake Victoria cichlids. *Proceedings of the Royal Society B: Biological Sciences* **287**, (2020).
54. Ganapathy, V. *et al.* Sodium-coupled monocarboxylate transporters in normal tissues and in cancer. *AAPS Journal* **10**, 193–199 (2008).
55. Duester, G. Families of retinoid dehydrogenases regulating vitamin A function. *Eur J Biochem* **267**, 4315–4324 (2000).
56. Luecke, S. *et al.* The aryl hydrocarbon receptor (AHR), a novel regulator of human melanogenesis. *Pigment Cell Melanoma Res* **23**, 828–833 (2010).
57. Mo, E. S., Cheng, Q., Reshetnyak, A. V., Schlessinger, J. & Nicoli, S. Alk and Ltk ligands are essential for iridophore development in zebrafish mediated by the receptor tyrosine kinase Ltk. *Proc Natl Acad Sci U S A* **114**, 12027–12032 (2017).
58. Soong, J. & Scott, G. Plexin B1 inhibits MET through direct association and regulates Shp2 expression in melanocytes. *J Cell Sci* **126**, 688 (2013).
59. Chen, Q., Sato, K., Yokoi, H. & Suzuki, T. Developmental regulatory system of ocular-side-specific asymmetric pigmentation in flounder: Critical role of retinoic acid signaling. *J Exp Zool B Mol Dev Evol* **334**, 156–167 (2020).
60. Ligon, R. A. & Mccartney, K. L. Biochemical regulation of pigment motility in vertebrate chromatophores: a review of physiological color change mechanisms. *Curr Zool* **62**, 237 (2016).
61. Koike, S. *et al.* Toll-like receptors 2 and 3 enhance melanogenesis and melanosome transport in human melanocytes. *Pigment Cell Melanoma Res* **31**, 570–584 (2018).
62. Nagao, Y. *et al.* Sox5 Functions as a Fate Switch in Medaka Pigment Cell Development. *PLoS Genet* **10**, e1004246 (2014).
63. Akula, M. *et al.* Retinoic acid related orphan receptor α is a genetic modifier that rescues retinal degeneration in a mouse model of Stargardt disease and Dry AMD. *Gene Therapy 2024 31:7* **31**, 413–421 (2024).
64. Seberg, H. E., Van Otterloo, E. & Cornell, R. A. Beyond MITF: Multiple transcription factors directly regulate the cellular phenotype in melanocytes and melanoma. *Pigment Cell Melanoma Res* **30**, 454 (2017).

65. Mallarino, R. *et al.* Developmental mechanisms of stripe patterns in rodents. *Nature 2016 539:7630* **539**, 518–523 (2016).
66. Kim, J. E. *et al.* The role of the Hippo pathway in melanocytes and melanoma. (2013) doi:10.3389/fonc.2013.00123.
67. Asaoka, Y., Hata, S., Namae, M., Furutani-Seiki, M. & Nishina, H. The hippo pathway controls a switch between retinal progenitor cell proliferation and photoreceptor cell differentiation in zebrafish. *PLoS One* **9**, e97365 (2014).
68. Hilliard, A. T., Xie, D., Ma, Z., Snyder, M. P. & Fernald, R. D. Genome-wide effects of social status on DNA methylation in the brain of a cichlid fish, Astatotilapia burtoni. *BMC Genomics* **20**, (2019).
69. Zheng, J. *et al.* Mechanism for regulation of melanoma cell death via activation of thermo-TRPV4 and TRPV2. *J Oncol* **2019**, 7362875 (2019).
70. Ye, J. *et al.* Wnt10b promotes differentiation of mouse hair follicle melanocytes. *Int J Med Sci* **10**, 691–698 (2013).
71. Silver, D. L. *et al.* The Secreted Metalloprotease ADAMTS20 Is Required for Melanoblast Survival. *PLoS Genet* **4**, 1000003 (2008).
72. Min, Y., Li, Q., Yu, H. & Du, S. Examination of wnt signaling mediated melanin transport and shell color formation in Pacific oyster (Crassostrea gigas). *Mar Life Sci Technol* **6**, 488–501 (2024).
73. Mongera, A. & Nüsslein-Volhard, C. Scales of fish arise from mesoderm. *Current Biology* **23**, R338–R339 (2013).
74. Fittipaldi, R. *et al.* The lysine methylase SMYD3 modulates mesendodermal commitment during development. **10**, 1233 (2021).
75. Seehausen, O. & Van Alphen, J. J. M. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav Ecol Sociobiol* **42**, 1–8 (1998).
76. Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. Colour vision and speciation in Lake Victoria cichlids of the genus Pundamilia. *Mol Ecol* **14**, 4341–4353 (2005).
77. Maan, M. E. *et al.* Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid Pundamilia nyererei. *Proc R Soc Lond B Biol Sci* **271**, 2445–2452 (2004).
78. Mintz-Hittner, H. A., Semina, E. V., Frishman, L. J., Prager, T. C. & Murray, J. C. VSX1 (RINX) mutation with craniofacial anomalies, empty sella, corneal endothelial changes, and abnormal retinal and auditory bipolar cells. *Ophthalmology* **111**, 828–836 (2004).
79. Bosse, M. *et al.* Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science (1979)* **358**, 365–368 (2017).
80. Enbody, E. D. *et al.* Community-wide genome sequencing reveals 30 years of Darwin's finch evolution. *Science* **381**, eadf6218 (2023).

81. Gómez, A. E. *et al.* Dominant nonsense mutations in efemp1 alter vertebral and craniofacial characteristics in adult zebrafish. *bioRxiv* 2024.02.16.580611 (2024) doi:10.1101/2024.02.16.580611.
82. Singh, P., Ahi, E. P. & Sturmbauer, C. Gene coexpression networks reveal molecular interactions underlying cichlid jaw modularity. *BMC Ecol Evol* **21**, 1–17 (2021).
83. Böhne, A., Sengstag, T. & Salzburger, W. Comparative Transcriptomics in East African Cichlids Reveals Sex- and Species-Specific Expression and New Candidates for Sex Differentiation in Fishes. *Genome Biol Evol* **6**, 2567 (2014).
84. Feller, A. F. & Seehausen, O. Genetic architecture of adaptive radiation across two trophic levels. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220377 (2022).
85. Li, M. *et al.* Variant Analyses of Candidate Genes in Orofacial Clefts in Multi-Ethnic Populations. *Oral Dis* **28**, 1921 (2021).
86. Qiu, W. *et al.* MACF1 deficiency suppresses tooth mineralization through IGF1 mediated crosstalk between odontoblasts and ameloblasts. *Genes Dis* **11**, 101103 (2023).
87. Ockeloen, C. W. *et al.* Novel mutations in LRP6 highlight the role of WNT signaling in tooth agenesis. *Genet Med* **18**, 1158 (2016).
88. Bjork, B. C., Turbe-Doan, A., Prysak, M., Herron, B. J. & Beier, D. R. Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet* **19**, 774 (2009).
89. Van Karnebeek, C. D. M. *et al.* NANS-mediated synthesis of sialic acid is required for brain and skeletal development. *Nature Genetics 2016 48:7* **48**, 777–784 (2016).
90. Krivanek, J. *et al.* Dental cell type atlas reveals stem and differentiated cell types in mouse and human teeth. *Nature Communications 2020 11:1* **11**, 1–18 (2020).
91. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature 1997 388:6638* **388**, 167–171 (1997).
92. Láruson, Á. J., Yeaman, S. & Lotterhos, K. E. The Importance of Genetic Redundancy in Evolution. (2020) doi:10.1016/j.tree.2020.04.009.
93. Yeaman, S., Gerstein, A. C., Hodgins, K. A. & Whitlock, M. C. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genet* **14**, 1–25 (2018).
94. Kafri, R., Springer, M. & Pilpel, Y. Genetic Redundancy: New Tricks for Old Genes. *Cell* **136**, 389–392 (2009).
95. Bomblies, K. Evolution: Redundancy as an Opportunity for Innovation. *Current Biology* **20**, R320–R322 (2010).
96. Moreiras, H. *et al.* The exocyst is required for melanin exocytosis from melanocytes and transfer to keratinocytes. *Pigment Cell Melanoma Res* **33**, 366–371 (2020).

97. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol* **34**, 531–544 (2019).
98. Meier, J. I. *et al.* Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Mol Ecol* **26**, 123–141 (2017).
99. Simpson, G. G. *The Major Features of Evolution*. (Columbia University Press, Ney York, 1953).
100. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol Evol* **19**, 198–207 (2004).
101. Sella, G. & Barton, N. H. Thinking about the evolution of complex Traits in the era of genome-wide association studies. *Annu Rev Genomics Hum Genet* **20**, 461–493 (2019).
102. Marques, D. A. *et al.* Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet* **12**, e1005887 (2016).
103. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
104. Wolf, J. B. W. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nature Publishing Group* (2016) doi:10.1038/nrg.2016.133.
105. Blumer, L. M. *et al.* Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid radiation. *Science (1979)* **388**, eadr9961 (2025).
106. Orr, H. A. Adaptation and the cost of complexity. *Evolution* **54**, 13–20 (2000).
107. Feller, A. F., Ogi, V., Seehausen, O. & Meier, J. I. Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages. *Hydrobiologia* **848**, 3727–3745 (2021).
108. Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O. & Puebla, O. Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution 2019 3:4* **3**, 657–667 (2019).
109. Marques, D. A. *et al.* Genomes reveal age and demographic consequence of ultrafast adaptive radiation. *bioRxiv* 2025.03.07.638630 (2025) doi:10.1101/2025.03.07.638630.