Adaptive introgression and *de novo* mutations increase access to novel fitness peaks on the fitness landscape during a vertebrate adaptive radiation

Austin H. Patton[1,2], Emilie J. Richards[1,2], Katelyn J. Gould[3], Logan K. Buie[3], Christopher H. Martin[1,2]

[1]*Museum of Vertebrate Zoology, University of California, Berkeley, CA*
[2]*Department of Integrative Biology, University of California, Berkeley, CA*
[3]*Department of Biology, University of North Carolina at Chapel Hill, NC*

Keywords: fitness landscape; adaptive radiation; introgression; standing genetic variation; *de novo* mutation

Abstract

Estimating the complex relationship between fitness and genotype or phenotype (i.e. the adaptive landscape) is one of the central goals of evolutionary biology. Empirical fitness landscapes have now been estimated for numerous systems, from phage to proteins to finches. However, the nature of adaptive walks connecting genotypes to organismal fitness, speciation, and novel ecological niches are still poorly understood. One outstanding system for addressing these connections is a recent adaptive radiation of ecologically and morphologically distinct pupfishes (a generalist, molluscivore, and scale-eater) endemic to San Salvador Island, Bahamas. Here, we leveraged whole-genome sequencing of 139 hybrids from two independent field fitness experiments to identify the genomic basis of fitness, visualize the first genotypic fitness networks in a vertebrate system, and infer the contributions of different sources of genetic variation to the accessibility of the fitness landscape. We identified 132 SNPs that were significantly associated with fitness in field enclosures, including six associated genes that were differentially expressed between specialists, and one gene (protein-lysine methyltransferase: METTL21E) misexpressed in hybrids, suggesting a potential intrinsic genetic incompatibility. We then constructed genotypic fitness networks from adaptive alleles and show that only introgressed and *de novo* variants, not standing genetic variation, increased the accessibility of genotypic fitness paths from generalist to specialists. Our results suggest that adaptive introgression and *de novo* variants provided key connections in adaptive walks necessary for crossing fitness valleys and triggering the evolution of novelty during adaptive radiation.

Significance

The fitness landscape describes the complex relationship between genotype or phenotype and fitness and is the conceptual bridge between micro- and macro-evolution. There are many studies of the phenotypic fitness landscapes across a range of systems from microbes to vertebrates, but studies of the genotypic fitness landscape remain limited to microbial systems and the fitness of individual proteins. Here, we conduct the first study investigating the genotypic fitness landscape for organismal fitness in an exceptionally young vertebrate adaptive radiation including a generalist algivore, molluscivore, and scale-eating pupfish. We show that genetic variation originating from adaptive introgression and de novo mutations exert a disproportionate influence on the fitness landscape, enabling the crossing of adaptive valleys and the evolution of novelty.

Introduction

First conceptualized by Sewell Wright in 1932, the adaptive landscape describes the complex relationship between genotype or phenotype and fitness (1). The landscape is a concept, metaphor, and an empirical measurement that exerts substantial influence over all evolutionary dynamics (2–6). Fitness landscapes were originally depicted as high-dimensional networks spanning genotypic space in which each genotype is associated with fitness (1). Simpson (7) later described the phenotypic evolution of populations through time on a rugged landscape, in which isolated clusters of fitness peaks represent 'adaptive zones' to which populations evolve from adjacent regions of low fitness (8). Lande and Arnold formalized the analysis of selection and estimation of phenotypic fitness landscapes (9–11), leading to empirical studies of fitness landscapes in numerous empirical systems (12–18). Fitness surfaces also provide a central component of speciation models and theory (19–21). Populations may speciate through a process of divergent selection (22), either on static landscapes or negative frequency-dependent disruptive selection (23–28).

A fundamental concept in fitness landscape theory is that not all genotypic trajectories are accessible (5, 29, 30); only those paths that monotonically increase in fitness at each mutational step are evolutionarily accessible, an expectation stemming from Fisher's fundamental theorem (31). These accessible genotypic trajectories can thus be considered potential adaptive walks under Fisher's geometric model (31, 32). Experimental studies of adaptive walks in nature are increasingly common, including the evolution of toxin resistance in monarch butterflies (33), metabolism and citrate exploitation in *E. coli* (34) (35), and host-shift in aphids (36).

Likewise, the accessibility of genotypic fitness networks has now been explored in numerous microbial systems, including the evolution of antibiotic resistance in *E. coli* (29), sensitivity to salinity in yeast (37), and Lenski's long-term evolution experiment (38). However, such studies are still limited to the investigation of the fitness of individual proteins in microbial systems well suited for experimental manipulation. Perhaps the closest study to a genotypic fitness landscape for a wild organism is that of Nosil et al. (39) who estimated fitness landscapes for *Timema* stick insects based on a field experiment. However, this study still focused it's investigation of the fitness landscape for a single gene underlying differential coloration between morphs. To our knowledge, no study has estimated genotypic fitness landscapes in a vertebrate system among multiple species as originally envisioned by Wright (1) despite the clear utility such a framework has in bridging micro- and macroevolutionary dynamics.

One promising system for estimating fitness landscapes is a recent adaptive radiation of *Cyprinodon* pupfishes (17, 18, 40, 41). This radiation is comprised of two endemic specialists, a molluscivore (durophage: *Cyprinodon brontotheroides*) and a scale-eater (lepidophage: *C. desquamator*), derived from a Caribbean-wide generalist ancestor (*C. variegatus*). These species all occur in sympatry in the hypersaline lakes of San Salvador Island, Bahamas (Fig 1A). Originating less than 10,000 years ago (based on geological age estimates for the lakes: (42)), the functional and trophic novelty harbored within this radiation is the product of exceptional rates of craniofacial morphological evolution (43–47). Furthermore, species boundaries appear to persist across multiple lake populations, despite persistent admixture among species (48, 49). We previously estimated fitness landscapes in these hypersaline lakes from two independent field experiments measuring the growth and survival of hybrids placed in field enclosures. Selection analyses revealed a temporally stable, multi-peaked adaptive landscape across lake populations,

year of study, and frequency or density manipulations (Fig. 1B: (17, 18, 41)). One of the strongest and most persistent trends across studies and treatments was the apparent isolation of hybrid phenotypes resembling the scale-eater by a large fitness valley from the other two species (17, 18). Hybrids resembling the generalist occupied a fitness peak and were isolated by a smaller fitness valley from hybrids resembling the molluscivore on a second region of high fitness. Understanding how an ancestral generalist population crossed these fitness valleys is key to understanding the origins of novelty in this system, particularly given the rare evolution of lepidophagy (scale-eating), a niche occupied by less than 0.3% of all fishes (50, 51).

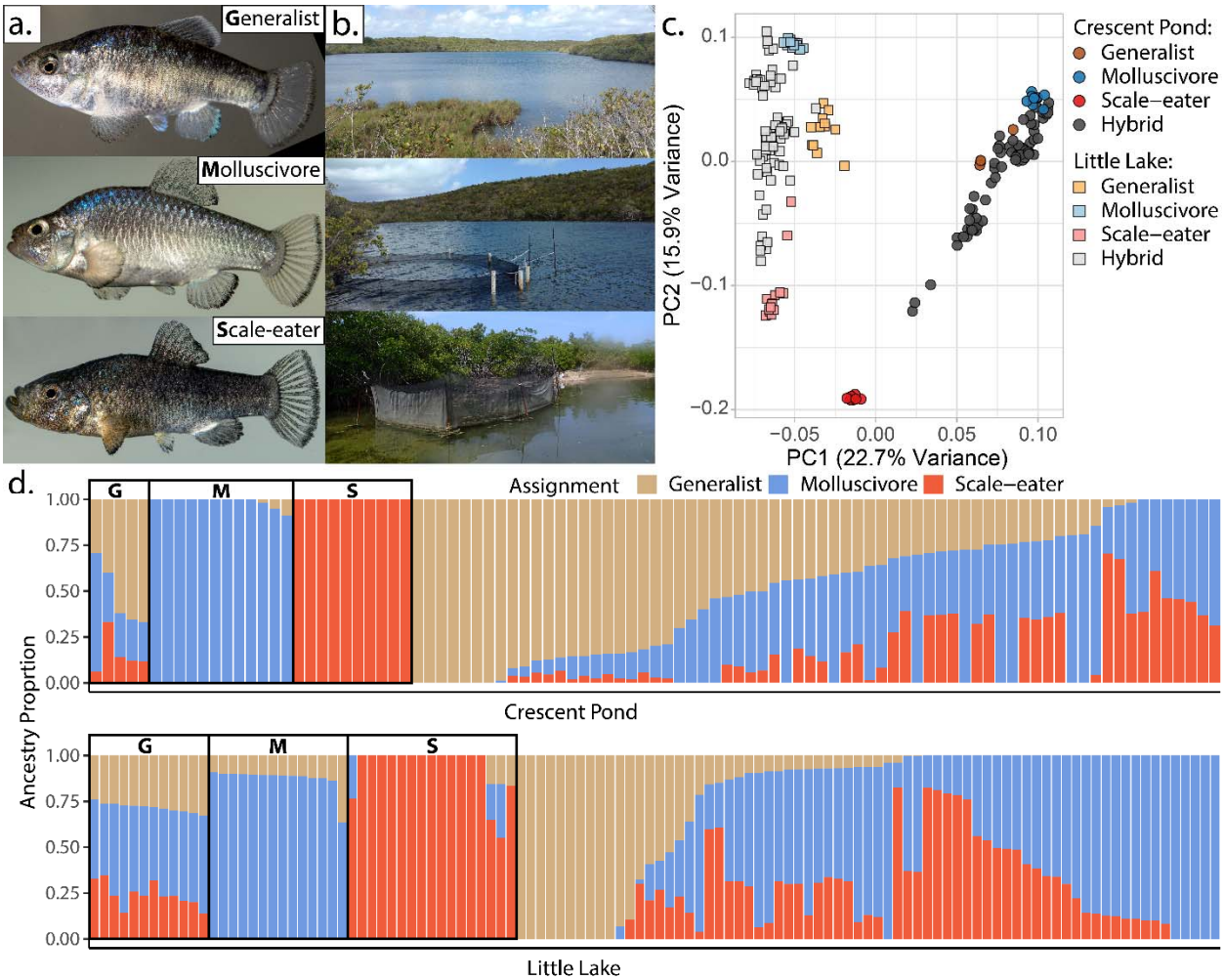

Figure 1. San Salvador Island Pupfish and their hybrids. a. From top to bottom: the generalist, *Cyprinodon variegatus*, the molluscivore *C. brontotheroides*, and the scale-eater *C. desquamator*. **b.** Representative images of experimental field enclosures. **c.** Principal component analysis of 1,129,771 SNPs genotyped in hybrids and the three parental species. **d.** ADMIXTURE analyses for Crescent Pond

(top) and Little Lake (bottom). G, M and S indicate individual samples of Generalists (G), Molluscivores (M), and Scale-eaters (S), respectively. Colors correspond to probability of assignment to one of three assumed populations/species (K = 3 in this analysis).

The relative contributions of standing genetic variation, *de novo* mutation, and introgression to the tempo and mode of adaptive radiation are now of central interest to the field of speciation genomics (52–57). Yet to date, no study has investigated the relative influence of each on the accessibility of the fitness landscape, a critical determinant of the dynamics of speciation. San Salvador Island pupfishes, like many others (58–62), appears to have benefitted from a complex interplay of abundant standing genetic variation, adaptive introgression from neighboring islands (Fig. 1: (49, 63)), and rare *de novo* mutations found only in the scale-eater. We have now identified many adaptive alleles showing evidence of hard selective sweeps in the two specialists, and these come from each of the three aforementioned sources of variation (49). Notably, both specialists harbor numerous introgressed loci that include genes of known craniofacial effects, implicating a role in adaptive divergence of the specialists (49, 63). Unlike adaptive introgression, hard selective sweeps on *de novo* mutation appear to have only taken place in the scale-eaters. Here, we leverage whole genome sequencing of 139 hybrids used in these field experiments to identify the genomic basis of fitness differences, infer genotypic fitness networks, and quantify the influence of each source of genetic variation on the accessibility of novel fitness peaks.

Results and Discussion

Population structure and ancestry association with fitness

We resequenced 139 hybrids (86 survivors, 56 deaths (Table S1)) from two independent field experiments across a total of six field enclosures and two lake populations (2011: two high-

density 4m diameter enclosures for three months: Crescent Pond n = 796; Little Lake = 875; 2014/2015; four high-density 3 m diameter enclosures: three months in Crescent Pond n = 923; eleven months in Little Lake = 842). We then characterized patterns of genetic variation among parental populations in each lake and their hybrids. Using 1,129,771 SNPs with an average coverage of 9.79X, we find that principal components analysis (PCA) of genetic variation strongly differentiated pupfishes sampled from Little Lake/Osprey Lake and Crescent Pond (PC1: 22.7% variance explained), as well as substantial differentiation among species within lake (PC2: 15.9% variance explained: Fig. 1D; Fig. S1-S2). These results are supported by ADMIXTURE analyses (64, 65) (Fig. 1E).

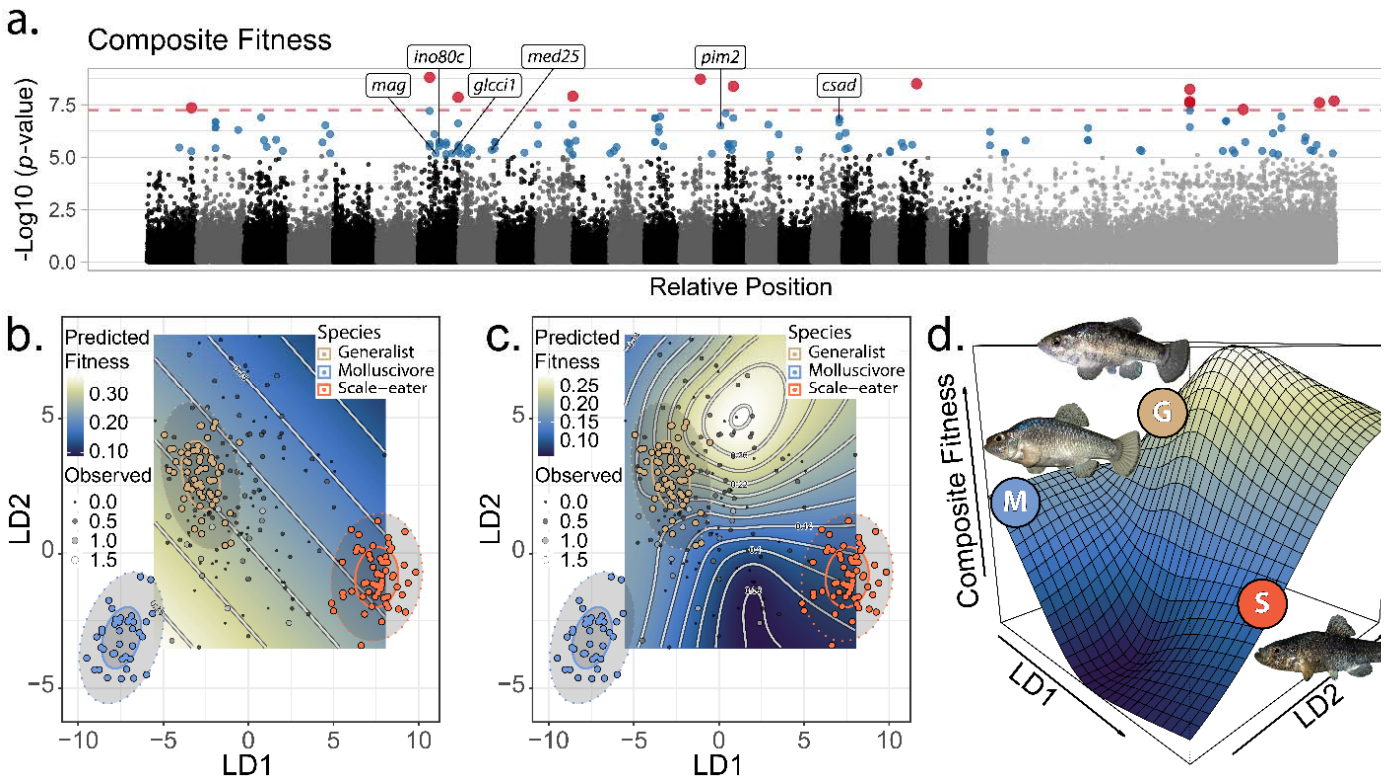

Figure 2. The genetic basis of fitness variation and improved inference of adaptive landscapes. a) Per-SNP -log10 *p*-values from a genome-wide association test with GEMMA for composite fitness (survival x growth). Lake and experiment were included as covariates in the linear mixed model. SNPs that were significant at FDR < 0.05 are indicated in blue; red SNPs above dashed red line cross the threshold for Bonferroni significance at α = 0.05. First twenty-four scaffolds are sorted from largest to smallest and remaining scaffolds were pooled. The six genes associated with composite fitness, strongly differentiated (F_{ST} > 0.95) and differentially expressed between specialists (66) are annotated. **b-c)** Best-

fit adaptive landscape for composite fitness using either morphology alone (**b:** flat surface with only directional selection) or morphology in combination with fitness associated SNPs (**c:** highly nonlinear surface). Best-fit model in C was a GAM including a thin plate spline for both LD axes, fixed effects of experiment and lake, and fixed effects of seven SNPs most strongly associated with fitness shown in red. **c)** Three-dimensional view of **C** with relative positions of the three parental species indicated.

Neither scale-eater nor molluscivore ancestry estimated by ADMIXTURE analysis was significantly associated with hybrid composite fitness (Fig. S3A, Table S2: Scale-eater P = 0.164; molluscivore P = 0.595), survival probability (Fig. S3B: Scale-eater p = 0.804; molluscivore P = 0.313), or growth (Fig. S3C: Scale-eater P = 0.283; molluscivore P = 0.328), while controlling for effects of lake and experiment. Similarly, neither PC1 nor PC2 was associated with composite fitness (Table S2). Furthermore, pairwise genetic distance did not predict pairwise morphological distance (Fig. S4). Therefore, in contrast to previous studies (67–69), in this system genome-wide ancestry is not associated with fitness, highlighting the complex nonlinear relationship between genotype, phenotype, and fitness within this nascent adaptive radiation. We must look to local ancestry to understand fitness relationships (e.g. (70)).

Genome-wide association mapping of fitness

From our linkage-thinned dataset we used a linear mixed model in GEMMA to identify 132 SNPs in regions that were strongly associated with composite fitness, including 13 which remained significant at the conservative Bonferroni threshold (Fig. 2A, Table S3: See supplement for results pertaining to other fitness measures [Table S4-S5; Figure S5-6]). Gene ontologies for these 132 fitness-associated regions were significantly enriched for synaptic signaling and chemical synaptic transmission (FDR rate < 0.01; Fig. S7; Table S5). Ontologies enriched at an FDR rate < 0.05 were related to signaling and regulation of cell communication. We did not identify any enrichment for ontologies related to craniofacial development which

have previously been identified to play a significant role in the adaptive divergence of these fishes (49, 63, 66, 71, 72). This suggests that fitness-associated regions in our field experiments capture other components of fitness in addition to the external phenotypes measured in previous studies. Surprisingly, our fitness associated regions did not overlap with regions showing significant evidence of a hard selective sweep (49). However, we did find six fitness-associated SNPs near genes differentially expressed between specialists in addition to one that was also misexpressed in F1 hybrids (66). This is particularly intriguing because misexpression in hybrids could indicate an intrinsic genetic incompatibility that would impact their fitness in field enclosures (66, 73).

Having identified fitness-associated regions, and the genes to which they are proximate, we then sought to identify whether these genes have been implicated in adaptive divergence of the specialists. Of the identified fitness-associated genes, five have previously been shown to include fixed divergent SNPs between specialists (*csad, glcci1, ino80c, mag, pim2*), and one has a fixed deletion between specialists (*med25*) (66). *med25* (Mediator Complex Subunit 25) is a craniofacial transcription factor associated with cleft palate in humans and zebrafish (74, 75). Four of the remaining genes shared across our study and McGirr et al. (66) are associated with growth and/or body size measurements in other fishes. For example, *csad* plays an important role in synthesizing taurine which is a rate-limiting enzyme affecting growth rate in parrotfishes (76), rainbow trout (77), and Japanese flounder (78). Second, *glcci1*, is associated with the body depth/length ratio in the large yellow croaker (79). Third, *ino80c* is associated with measures of body size in the Nile tilapia (80). The only gene associated with growth but not composite fitness, *mettl21e*, was also misexpressed in F1 hybrids between scale-eaters and molluscivores at eight days post-fertilization, and thus is a putative genetic incompatibility in this system (66).

Although it has not been associated with growth-related or body-size traits in fishes, it is associated with intramuscular fat deposition in cattle (81). Taken together, these findings support the interpretation that our fitness-associated SNPs are associated with unmeasured traits, particularly physiological growth rates, that affect fitness in our hybrid field experiments.

Fitness-associated SNPs improve inference of the adaptive landscape

We used linear discriminant axes and generalized additive modelling (GAM) to estimate phenotypic fitness landscapes for the sequenced hybrids on a two-dimensional morphospace indicating similarity to each of the three parental populations following previous studies (17, 18). We then tested whether the inclusion of the 13 genomic regions most strongly associated with fitness (red: Fig. 2A) in GAM models improved our inference of the underlying adaptive landscape. Models with fitness-associated SNPs were invariably favored over models with external morphology alone (ΔAICc > 8.6: Tables S8-S9). Morphology-only models predicted a flat linear landscape (Fig. 2B, Fig. S10); whereas including fitness-associated SNPs resulted in inferences of increased curvature and complexity of the fitness surface, despite our limited dataset of 139 hybrids relative to sample sizes in previous studies around 800 hybrids per enclosure.

To reduce complexity of the full model estimated from 31 morphological traits including all 13 fitness associated SNPs, we fit an additional model including only the seven SNPs that were significant in this model. This reduced model was the best-fit adaptive landscape, and was complex and characterized by a fitness peak near hybrids resembling generalist phenotypes separated by a small fitness valley from a second region of high fitness for molluscivore hybrid phenotypes. Scale-eater hybrid phenotypes again corresponded to a large fitness valley (Fig. 2B-

C: For results pertaining to growth or survival see Supplemental Materials – Fig. S10, Table S10-12). Compared to previous studies, the highest fitness optimum was shifted from molluscivore phenotypes to generalists. This suggests that fitness-associated SNPs increased the fitness of hybrids resembling generalists beyond expectations based on their morphology alone, consistent with the hypothesis that fitness-associated SNPs are associated with unmeasured traits affecting fitness. Indeed, visualization of genotypes on the fitness landscape support this interpretation: A common genotype (SNP haplotype of the seven fitness-associated SNPs) associated with high fitness was most frequent in hybrids resembling generalists and rare in hybrids resembling specialists (Fig. S11).

Trophic novelty is associated with isolation on the genotypic fitness landscape

We reconstructed genotypic fitness networks from 1,498 adaptive alleles previously identified in this system (e.g. Fig. 3). These were regions displaying significant evidence of hard selective sweep in both site frequency spectrum and LD-based methods, SweeD (82) and OmegaPlus (83), which contained fixed or nearly fixed SNPs (Fst > 0.95) differentiating trophic specialists across lakes (49). These fitness networks depict both hybrids and parental species in genotypic space, with nodes representing SNP haplotypes and edges connecting mutational neighbors (Fig. 3B). Genotype space is immense; using SNPs coded as homozygous reference, heterozygote, or homozygous alternate, the number of potential haplotypes is equal to $3^{\#\ SNPs\ in\ network}$. Thus, unlike experimental studies of individual proteins in *E. coli* (29, 38) or yeast (37), it is not possible for us to investigate the full breadth of genotypic space.

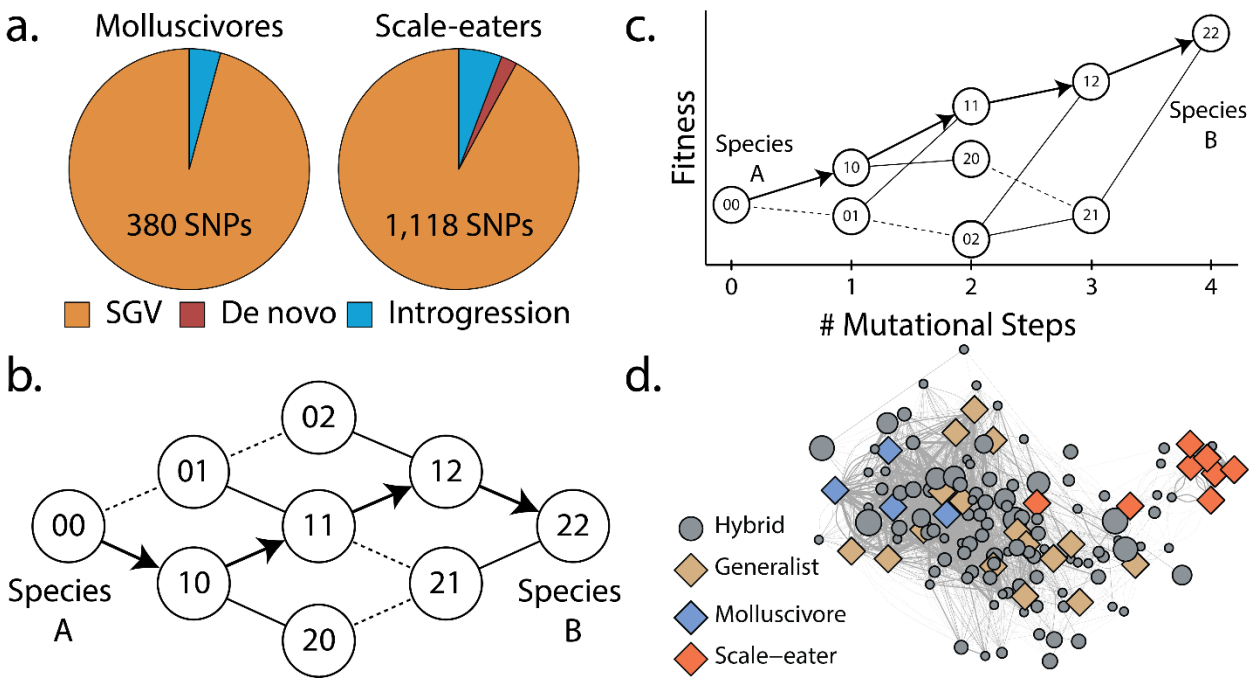

Figure 3. Genotypic fitness networks connecting San Salvador Island pupfishes. a) The majority of nearly fixed or fixed variants (F_{ST} > 0.95) experiencing hard selective sweeps (hereafter 'adaptive alleles') originated as standing genetic variation (SGV: Molluscivores = 96%, Scale-eaters = 92%), followed by introgression (Molluscivores = 4%, Scale-eaters = 6%), and lastly *de novo* mutation (Scale-eaters = 2%)(49). Pie charts shows adaptive alleles retained in our study for each species. **b)** Diagram illustrating genotypic fitness networks and adaptive walks in a pair of species for two SNPs concatenated into haplotypes. Species A and B are separated by four mutational steps. Dashed lines are inaccessible as they result in a decrease in fitness; this leaves a single possible accessible evolutionary trajectory between species A and B (indicated by bold arrows). Each node is associated with an empirical measure of hybrid fitness. Edges are always drawn as directed, from low to high fitness. **a)** The same network as in B, with fitness plotted on the Y-axis and number of mutational steps from species A to B on the X-axis. Dashed lines are inaccessible (decreasing in fitness), solid lines are accessible (increasing in fitness), and the only accessible path between species A and B is indicated by solid arrows. **d)** Genotypic network constructed using a random sample of ten SNPs, sampled from all SNPs shown in **a**. Each edge between nodes is up to five mutational steps away; edge width is proportional to mutational distance: wider edges connect closer haplotypes. Node size (for hybrids) is scaled proportional to mean composite fitness; larger circles correspond to fitter haplotypes.

Instead, to understand the distribution of parental species and their hybrids in genotypic space, we began by using a random sample of ten SNPs drawn from our set of candidate adaptive alleles in this system. Here, we plotted edges between nodes up to five mutational steps away and found that generalists and molluscivores are closer in the genotypic fitness network than either is to scale-eaters (Fig 4A), as expected based on their genetic distance. Most scale-eaters

appear quite isolated in genotypic space, separated from the generalist cluster of nodes by 12.6 ± 0.091 (mean +/- SE: $P < 0.001$) mutational steps and from molluscivores by 16.3 ± 0.060 steps ($P < 0.001$). In contrast, molluscivores were separated from generalists by 5.37 ± 0.103 steps ($P < 0.001$). Generalists show the greatest intrapopulation distances, separated from each other by 6.08 ± 0.088 steps ($P < 0.001$). In contrast, molluscivores exhibited the smallest intrapopulation distances, separated by 1.75 ± 0.021 steps ($P < 0.001$). Scale-eater intrapopulation distances were intermediate (4.71 ± 0.088 steps: $P < 0.001$).

Adaptive introgression and de novo *mutations increase accessibility of novel fitness peaks*

Evolutionary accessibility reflects the probability that one genotypic trajectory will be favored by natural selection over another, and thus can be related to the predictability of evolution (5, 29, 30, 84, 85). Genotypic trajectories with the minimum number of steps between two states that are characterized by monotonically increasing fitness at each step are considered to be the most accessible (5, 29, 30). That is, each mutational step is favored by natural selection and longer paths are less probable (30). This provides two useful metrics of fitness network accessibility: 1) The number of accessible paths and 2) the minimum length of accessible paths.

The adaptive radiation of pupfishes on San Salvador island originated within the last 10,000 years through a combination of selection on standing genetic variation, adaptive introgression, and *de novo* mutation (49). However, it is unclear how each source of genetic variation aided in the traversal of fitness paths or contributed to the colonization of novel fitness peaks. To address this question, we constructed genotypic fitness networks composed of each of the three main sources of adaptive genetic variation: standing genetic variation, introgression from one of four focal Caribbean generalist populations, or *de novo* mutations unique to San

Salvador Island. We also examined combinations of these categories to better reflect the stages of adaptive divergence. We compared sets of 1,000 random 5-SNP genotypic networks drawn from different sources of adaptive variation and identified and scored the lengths of accessible paths (i.e. those monotonically increasing in fitness) from generalist to each specialist. We asked whether source of genetic variation affected 1) the number of accessible paths or 2) the shortest length of accessible paths.

We discovered that both scale-eater and molluscivore specialist fitness peaks in genotypic networks constructed from introgressed or *de novo* adaptive mutations were significantly more accessible than networks constructed from standing genetic variation (Fig. 4C, Table S15). Random networks drawn from these sources of adaptive variation contained more accessible fitness paths from generalist to specialists than networks constructed from standing genetic variation, while controlling for differences in overall network size (Figs. S12-S13). Specifically, networks constructed from adaptive introgression had 126% more accessible paths than those constructed from standing variation for molluscivores (median: $P < 0.0001$: Fig. 4C), and 39% more for scale-eaters ($P < 0.0001$: Fig. 4C). Likewise, networks constructed from *de novo* mutations had 42% more accessible paths ($P < 0.0001$: Fig. 4C).

Consistent with past experimental studies in *E. coli* (29, 38) or yeast (37), accessible paths were quite rare; the vast majority of genotypic trajectories were inaccessible, i.e. they included one or more mutational steps that decreased in fitness from generalist to specialist. Compared to standing variation, accessible paths were also significantly shorter in introgressed (mean: 9% shorter in molluscivores, 23% shorter in scale-eaters) and *de novo* mutation networks (14% shorter in scale-eaters: Fig. 4D. Fig S12, Table S16). Thus, introgressed and *de novo* adaptive mutations present more possible trajectories and reduce the number of steps needed to

traverse between species and access novel trophic specialist fitness peaks (See Supplemental Materials Fig. S14-S15).

This result, of adaptive introgression and *de novo* mutation changing the topology of the fitness landscape such that novel fitness peaks are more accessible, is striking. Our results are consistent with a growing body of evidence that de novo and introgressed variation may contribute to rapid speciation (54, 60, 86). Indeed, adaptive introgression appears to be the main factor explaining the rarity of adaptive radiation of pupfishes restricted to San Salvador Island (49, 63). This finding is not unusual. For instance, hybridization and adaptive introgression has now been shown to have facilitated the adaptive radiation of cichlids in Lake Victoria (60) and Lake Malawi (87), *Heliconius* butterflies (88), and Darwin's finches (89) Likewise, *de novo* mutation (in concert with introgression and standing genetic variation) was involved in the adaptive radiation of two clades of wild tomatoes (58, 90). However, *de novo* mutation is still a rate-limiting step in adaptation and speciation (91).

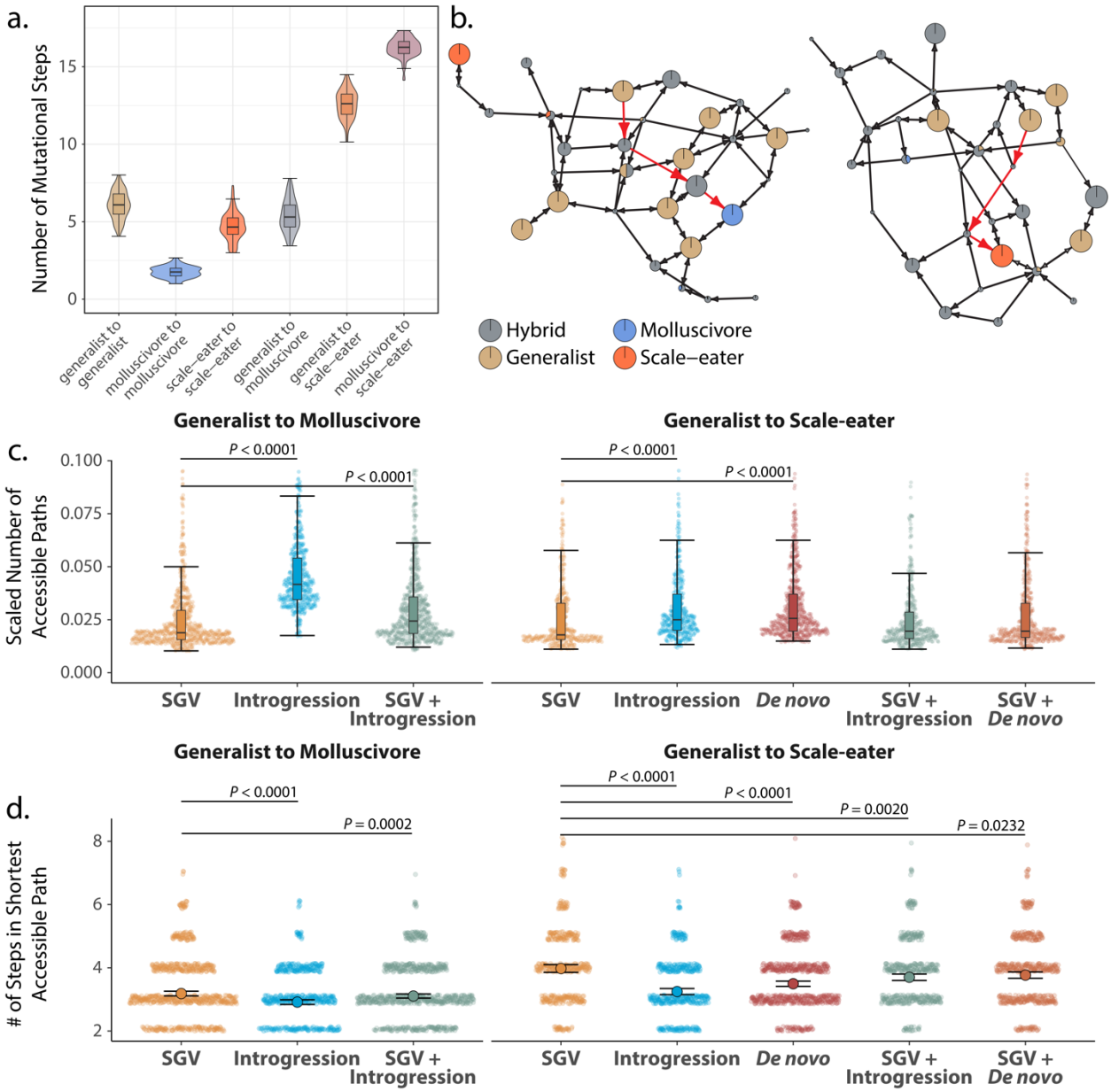

Figure 4. Adaptive introgression and *de novo* mutations increase access to specialist fitness peaks. a) Median number of mutational steps within or between species (e.g. Fig. 3C). All pairwise comparisons using Tukey's HSD test (after FDR correction) were significant. **b)** Example genotypic fitness networks constructed from five SNPs randomly drawn from introgressed adaptive loci for molluscivores (left) or scale-eaters (right). Edges connect nodes (haplotypes) that are separated by a single mutational step, arrows indicate the direction of mutational steps (directed towards higher fitness or parental haplotypes if not observed in hybrids), and red arrows indicate an example accessible trajectory. **c-d)** Accessibility of between-species genotypic trajectories in fitness landscapes constructed from adaptive loci originating from standing genetic variation (SGV), introgression, *de novo* mutation, or a combination. Due to the highly skewed nature of these distributions, pairwise significance among all network sets was determined using a nonparametric Kruskal-Wallis one-way analysis of variance following correction for multiple tests using an estimated False Discovery Rate. **C)** Proportion of accessible (i.e. monotonically increasing in fitness) paths per network, scaled by the total size of the network. The dashed line corresponds to the

median for SGV). **d)** Number of mutational steps in the shortest accessible path. Due to the discrete nature of these data, means are plotted as large circles, with two standard errors shown.

Conclusion

Here, we have analyzed the first multi-species genotypic fitness landscape based on hybrid fitness across multiple field experiments. We have shown that adaptive introgression and *de novo* mutation increase the accessibility of novel fitness peaks, thus promoting the evolution of novelty. Adaptation from standing genetic variation may initially be more rapid due to its higher frequency, initial availability, and potentially reduced genetic load within a population (91–93). However, our results suggest that once genetic variation arrives from introgression or arises *de novo* in a population, populations may colonize new adaptive zones due to the reduced number of mutational steps in accessible paths. Further, *de novo* mutations that undergo selective sweeps are also predicted to have larger effect sizes (94). These findings support an interpretation of adaptive radiation unfolding in stages (49, 95), first via standing genetic variation, next introgression from across the Caribbean, and finally *de novo* mutations.

Alternatively, our findings may help explain the relative isolation of scale-eaters on the fitness landscape and the rare evolution of scale-eating specialists despite comparable environments across the Caribbean with nearly identical levels of adaptive standing genetic variation: adaptive *de novo* mutations may enable the crossing of a large fitness valley via more accessible fitness pathways through genotype space, but it may also trap scale-eater populations on this peak through few accessible fitness pathways back to generalist peaks. That is, by increasing accessibility of the novel fitness peak, *de novo* mutations simultaneously reduce the accessibility of ‘back-stepping’ in genotypic space, as the population undergoes hill-climbing to the new adaptive optimum. Thus, individual specialists dispersing to neighboring islands may

rarely found new populations if their complement of *de novo* and introgressed adaptive loci limit viable fitness trajectories in their hybrid offspring.

Acknowledgements:

We thank Michelle St. John, David Tian, and Jacqueline Galvez for insightful discussion of the results; the Gerace Research Centre and Troy Day for logistical support; the government of the Bahamas for permission to collect, import, and export samples; the Vincent J. Coates Genomics Sequencing Center and Functional Genomics Laboratory at UC Berkeley for performing whole-genome library preparation and sequencing (supported by NIH S10 OD018174 Instrumentation Grant), and the University of California, Berkeley and University of North Carolina at Chapel Hill for computational resources. This work was funded by the National Science Foundation DEB CAREER grant #1749764, National Institutes of Health grant 5R01DE027052-02, the University of North Carolina at Chapel Hill, and the University of California, Berkeley to CHM, and graduate research funding from the Museum of Vertebrate Zoology to EJR. Genomic data are archived at the National Center for Biotechnology Information BioProject Database (Accessions:

PRJNA690558; PRJNA394148, PRJNA391309).

Methods:

Sampling

Our final dataset is comprised of 139 hybrid samples used in two separate field experiments (17, 18) on San Salvador island. Experiments were conducted in two lakes: Little Lake and Crescent Pond. All hybrids were measured for 32 external morphological traits (see below). Additionally, we include sequenced parental species of the generalist (N = 17), molluscivores (N = 27) and scale-eaters (N = 25)

sampled from these two lakes previously included in Richards et al. (49). Note that we treat samples from Little Lake and Osprey Lake as having come from the same population, as these two lakes comprise a single ecologically similar body of water, and fish from these populations are genetically undifferentiated (48, 49). For morphological analyses, we additionally measured samples of 60 generalists, 38 molluscivores, and 60 scale-eaters raised in the same laboratory common garden environment. A full list of samples are included in the supplement (Table S1).

Sequencing, genotyping and filtering

Raw reads from a combined set of 396 samples (see supplement) were first mapped to the *C. brontotheroides* reference genome (genome size = 1.16 Gb; scaffold N50 = 32 Mb) (49) using bwa-mem (v. 0.7.2). Duplicate reads were identified and reads using MarkDuplicates, and BAM indices were subsequently generated using the Picard software package (96). Samples were genotyped following Richards et al., (49) according to GATK best practices (97). Specifically, Single Nucleotide Polymorphisms (SNPs) were called and filtered using hard-filtering criteria in HaplotypeCaller. We used the following criteria in our filtering approach: QD < 2.0; QUAL < 20; FS < 60; MQRankSum < -12.5; ReadPosRankSum < -8 (97–99).

Following initial genotyping with GATK, we subsequently filtered our data further using VCFtools (100). Specifically, we filtered using the following flags: --maf 0.05; --min-alleles 2; --max-alleles 2; --min-meanDP 7; --max-meanDP 100; --max-missing 0.85. Indels were removed. To reduce non-independence among sites in our final dataset, we conservatively removed sites in strong linkage disequilibrium using plink v1.9 (--indep-pairwise 10['kb'] 50 0.5: (101)). Doing so resulted in the retention of 1,129,771 SNPs across 139 hybrid samples and the 69 samples from Richards et al (49). Unless other specified, these SNPs were used for all downstream analyses.

Hybrid fitness measures

Herein we use three measures of fitness: survival, growth, or a composite measure of the two. Survival was a binary trait indicating whether a fish survived or not during its exposure period in field enclosures. Growth was a continuous measure, defined as the proportional increase in Standard Length $\left(\frac{Final\ SL - Starting\ SL}{Starting\ SL}\right)$. Lastly, we defined composite fitness as survival * growth, similar to the metric used in (102), and analogous to composite fitness in Hereford (103) who used fecundity as their second fitness measure, rather than growth. Because composite fitness represents the most information rich metric of fitness, we report composite fitness results in the main text; results for growth and survival are included in the supplement.

Population genetic variation

To visualize genetic variation present in hybrids and across lakes (Crescent Pond and Little Lake), we first used a Principal Components Analysis (PCA) of genetic variation using plink v1.90 ((101), Fig. 1), plotting the first two principal component axes using R (version 3.6.3 (R Core team 2020). We then estimated admixture proportions in hybrids using ADMIXTURE v1.3.0 (64). Parental species found on San Salvador Island were substantially differentiated among Crescent Pond and Little Lake (48, 49); thus, independent ADMIXTURE analyses were conducted for each lake. Because we were primarily interested in admixture proportions of hybrids, we set K = 3 in these analyses, corresponding to the three parental species used in hybrid crosses. Using admixture proportions of hybrid individuals, we tested the hypothesis that ancestry predicts hybrid composite fitness in experimental field enclosures by fitting a generalized additive model including either 1) scale-eater ancestry or 2) molluscivore ancestry with fixed effects for experiment and lake. This was repeated for survival and growth separately.

Genome-wide association tests

To identify SNPs that were most strongly associated with fitness (survival, growth, or composite), we implemented the linear mixed model (LMM) approach in GEMMA (v. 0.98.1: (104)). This analysis

was repeated using each fitness measure as the response variable. To account for relatedness among samples we estimated the kinship matrix among all 139 hybrid samples, which in turn were used in downstream LMMs. To account for the potentially confounding effect of year/experiment and lake on estimated fitness measures, we included each as covariates in the LMMs. To ensure rare variants were not included in these analyses, we only included sites that had a minor allele frequency greater than 5% across all hybrids. A total of 933,520 SNPs were analyzed; 196,251 SNPS were excluded due to allele frequency change following removal of parental species. SNPs strongly associated with fitness were identified as 1) a False Discovery Rate (FDR: Benjamini and Hochberg 1995) less than 0.05, or a 2) *P*-value < 0.05 following a Bonferroni correction. We focus primarily on the sites identified by the conservative Bonferroni correction, however.

Gene ontology enrichment

We annotated sites that were significantly associated with fitness using snpEff (105) and the annotated *C. brontotheroides* reference genome (49). That is, we constructed a custom database within snpEff using the functional annotations originally produced by Richards et al. (49), and subsequently extracted information on the annotations and putative functional consequences of each variant.

Using genes identified for each SNP that is significantly associated with one of the fitness measures, we performed Gene Ontology (GO) enrichment analyses using ShinyGO v0.61 (106). For genes identified as being intergenic, we included both flanking genes. As in Richards et al. (2021), the gene symbol (abbreviation) database that had the greatest overlap with ours was that of the human database; thus, we tested for enrichment of biological process ontologies curated for human gene functions, based on annotations from Ensembl. Results are reported for biological processes that were significantly enriched with FDR < 0.05. We then compared this list of candidate loci to those identified in past studies of San Salvador Island pupfishes (49, 66, 107).

Morphometrics

We measured 31 external morphological traits for all 139 hybrids and 69 parental individuals from Crescent Pond (30 generalists, 18 molluscivores, and 30 scale-eaters) and 85 from Little Lake (30 generalists, 25 molluscivores, and 30 scale-eaters). We digitally landmarked dorsal and lateral photographs (both sides) of each lab-reared hybrid (pre-release) or parent using DLTdv8 (108). Measurements include 27 linear characters and three angles. For nearly all individuals, lateral measurements were collected from both lateral photographs and averaged. Morphological variables were size-corrected using the residuals of a log10(trait) – log10(standard length) regression standardized for selection analyses as outlined in the supplement. We used these 31 morphological traits to estimate two linear discriminant (LD) axes that best distinguished the generalist, molluscivore, and scale-eater using the LDA function in the mass package in R. We then used the resultant LD model to predict LD-scores for the 139 sequenced hybrids for later use in generalized additive models.

Estimation of adaptive landscapes

We fit generalized additive models (GAMs) using the mgcv package v. 1.8.28 (109) in R to estimate fitness landscapes relating the two discriminant axes (LD1-2) to fitness. All models included a thin-plate spline fitted for the two linear discriminant axes, and we included both lake and experiment in all models as fixed effects. Lake x experiment interaction terms were also included in some models. Models were ranked using the corrected Akaike Information Criterion (AICc) and were considered to be a significantly worse fit if ΔAICc > 4 from the best-fit model (110). The best-fit model from the above approach was in turn used to visualize fitness landscapes, plotting predicted values of fitness measures across LD space in R.

Using these results, we tested whether inclusion of SNPs that were strongly associated with fitness (i.e., those that surpassed the 0.05 Bonferroni threshold) improved estimation of fitness landscapes. We first extracted genotypes for the highly significant SNPs identified by GEMMA (13 for composite fitness, four for growth: see section *Fitness-genotype association test*), and coded these as either reference, single or double mutants using VCFtools (100). We then used the best-fit models

identified above and fit a range of models that included one or all SNPs. Individual fitness-associated SNPs were treated as ordered factors (i.e. transition from homozygous reference to heterozygote to homozygous alternate) and modeled using a factor-smooth in the generalized additive models. Note that factor "smooths" are effectively modeled as step-functions.

Estimation of genotypic fitness networks

We first estimated genotypic networks using sites previously shown to be highly divergent (F_{ST} > 0.95) and having undergone hard selective sweeps in specialists. To do so, we identified the SNPs in our un-thinned dataset overlapping with sites inferred to have undergone selective sweeps (49), resulting in 380 SNPs for molluscivores and 1,118 SNPs for scale-eaters. We subsequently constructed genotypic fitness networks in igraph v. 1.2.4.1 (111) following the procedure outlined in the supplement.

To visualize the high-dimensional genotypic fitness network, we randomly sampled ten adaptive loci 100 times and plotted haplotypes connected by edges if they were within five mutational steps of one another (Figure 3C). Then we calculated the mean distance between all species pairs (in number of mutational steps). We used pairwise Tukey's HSD tests to test whether inter-species distances differed.

Estimation of evolutionary accessibility

We tested whether the evolutionary accessibility of genotypic fitness trajectories through observed hybrid genotypes from generalist to each specialist species differed based on the source of genetic variation. We restricted our investigation to networks composed of adaptive loci as previously described (Figure 3A: Richards et al. 2021). This included a total of 380 SNPs in the molluscivores, and 1,118 in the scale-eaters. We further partitioned these SNPs by their respective sources: standing genetic variation (molluscivore N = 364; scale-eater N = 1,029), *de novo* mutation (scale-eater N = 24), or introgression (molluscivore N = 16; scale-eater N = 65), again using the assignments from Richards et al. (49). For analyses of trajectories between generalists and molluscivores, we included only SNPs found to

be sweeping in molluscivores; likewise, we included only SNPs sweeping in scale-eaters for analysis of trajectories between generalists and scale-eaters.

The full procedure for constructing genotypic fitness networks, identifying accessible paths, and quantifying accessibility is outlined in the supplement. Briefly, we randomly generated 1,000 datasets of five SNPs comprised of either 1) standing genetic variation, 2) adaptive introgression, 3) *de novo* mutation, 4) standing genetic variation + adaptive introgression, or 5) standing genetic variation + *de novo* mutation. Because different sets of sites are sweeping in each specialist, we conducted these analyses separately for each species. We then constructed genotypic networks, in which nodes are haplotypes of SNPs encoded in 012 format (0 = homozygous reference, 1 = heterozygote, 2 = homozygous alternate), and edges link mutational neighbors.

For each random network sample and for each measure of fitness we calculated 1) the minimum length of accessible paths between a random generalist and specialist sampled from our sequenced individuals, 2) the number of accessible paths between the same generalist and specialist pair, 3) the number of nodes and 4) the number of edges in the network. Larger networks will necessarily have a greater number of potential paths, including both accessible and inaccessible paths (Fig. S16), and we are interested in the relative proportion of accessible pathways, not just the raw number. For this reason and because network size differs among permutations, we divided the number of accessible paths in each random network sampled by the number of nodes, thus normalizing the counts and making them comparable across replicates and sources of variation. We assessed whether differences in these measures among the three sources of variation were significant using an ANOVA in R; due to the highly skewed nature of these distributions, post-hoc pairwise significance was assessed using a nonparametric Kruskal-Wallis one way analysis of variance in the agricolae package (112) in R.

Literature Cited

1. S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Sixth Int. Congr. Genet.*, 356–366 (1932).
2. S. Gavrilets, Evolution and speciation on holey adaptive landscapes. *Trends Ecol. Evol.* **12**, 307–312 (1997).
3. S. Gavrilets, "High-dimensional fitness landscapes and speciation" in *Evolution—the Extended Synthesis*, M. Pigliucci, G. B. Müller, Eds. (MIT Press, 2010), pp. 51–64.
4. E. Svensson, R. Calsbeek, *The adaptive landscape in evolutionary biology*, E. Svensson, R. Calsbeek, Eds. (Oxford University Press, 2012).
5. I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).
6. K. K. Fear, T. Price, The adaptive surface in ecology. *Oikos* **82**, 440–448 (1998).
7. G. G. Simpson, *Tempo and mode in evolution* (Columbia University Press, 1944).
8. S. Kauffman, S. Levin, Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
9. R. Lande, S. J. Arnold, The measurement of selection on correlated characters. *Evolution* **37**, 1226 (1983).
10. S. J. Arnold, M. E. Pfrender, A. G. Jones, "The adaptive landscape as a conceptual bridge between micro- and macroevolution" in *Microevolution Rate, Pattern, Process*, (Springer, Dordrecht, 2001), pp. 9–32.
11. S. J. Arnold, Performance surfaces and adaptive landscapes. *Integr. Comp. Biol.* **43**, 367–375 (2003).
12. D. Schluter, P. R. Grant, Determinants of morphological patterns in communities of Darwin's finches. *Am. Nat.* **123**, 175–196 (1984).
13. D. Schluter, Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 849–861 (1988).
14. A. P. Hendry, S. K. Huber, L. F. De León, A. Herrel, J. Podos, Disruptive selection in a bimodal population of Darwin's finches. *Proc. R. Soc. B Biol. Sci.* **276**, 753–759 (2009).
15. M. O. Beausoleil, *et al.*, Temporally varying disruptive selection in the medium ground finch (Geospiza fortis). *Proc. R. Soc. B Biol. Sci.* **286** (2019).
16. C. W. Benkman, Divergent selection drives the adaptive radiation of crossbills. *Evolution* **57**, 1176–1181 (2003).
17. C. H. Martin, P. C. Wainwright, Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science* **339**, 208–211 (2013).
18. C. H. Martin, K. J. Gould, Surprising spatiotemporal stability of a multi‐peak fitness landscape revealed by independent field experiments measuring hybrid fitness. *Evol. Lett.*, evl3.195 (2020).
19. S. Gavrilets, *Fitness landscapes and the origin of species* (Princeton University Press, 2004).
20. M. Turelli, N. H. Barton, J. A. Coyne, Theory and speciation. *Trends Ecol. Evol.* **16**, 330–343 (2001).
21. M. R. Servedio, J. W. Boughman, The role of sexual selection in local adaptation and speciation. *Annu. Rev. Ecol. Evol. Syst.* **48**, 85–109 (2017).
22. N. H. Barton, "Founder effect speciation" in *Speciation and Its Consequences*, D. Otte, J. A. Endler, Eds. (Sinauer, 1989), pp. 229–256.
23. U. Dieckmann, M. Doebeli, On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
24. D. I. Bolnick, Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution* **58**, 608–618 (2004).
25. F. J. Weissing, P. Edelaar, G. S. van Doorn, Adaptive speciation theory: A conceptual review. *Behav. Ecol. Sociobiol.* **65**, 461–480 (2011).
26. J. E. Cresswell, C. Galen, Frequency-dependent selection and adaptive surfaces for floral character combinations: the pollination of Polemonium viscosum. *Am. Nat.* **138**, 1342–1353 (1991).

27. S. P. Otto, M. R. Servedio, S. L. Nuismer, Frequency-dependent selection and the evolution of assortative mating. *Genetics* **179**, 2091–2112 (2008).
28. B. C. Haller, A. P. Hendry, Solving the paradox of stasis: Squashed stabilizing selection and the limits of detection. *Evolution* **68**, 483–500 (2014).
29. D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
30. J. Franke, A. Klözer, J. A. G. M. de Visser, J. Krug, Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* **7**, 1002134 (2011).
31. R. A. Fisher, *The genetical theory of natural selection* (Oxford University Press, 1930).
32. H. A. Orr, The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
33. M. Karageorgi, *et al.*, Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409–412 (2019).
34. F. Peng, *et al.*, Effects of beneficial mutations in pykF gene vary over time and across replicate populations in a long-term experiment with bacteria. *Mol. Biol. Evol.* **35**, 202–210 (2018).
35. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7899–7906 (2008).
36. K. S. Singh, *et al.*, The genetic architecture of a host shift: An adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. *Sci. Adv.* **6**, eaba1070 (2020).
37. C. Bank, S. Matuszewski, R. T. Hietpas, J. D. Jensen, On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14085–14090 (2016).
38. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
39. P. Nosil, *et al.*, Ecology shapes epistasis in a genotype–phenotype–fitness map for stick insect colour. *Nat. Ecol. Evol.* **4**, 1673–1684 (2020).
40. C. H. Martin, P. C. Wainwright, A remarkable species flock of Cyprinodon pupfishes endemic to San Salvador Island, Bahamas. *Bull. Peabody Museum Nat. Hist.* **54**, 231–241 (2013).
41. C. H. Martin, Context dependence in complex adaptive landscapes: frequency and trait-dependent selection surfaces within an adaptive radiation of Caribbean pupfishes. *Evolution* **70**, 1265–1282 (2016).
42. B. J. Turner, D. D. Duvernell, T. M. Bunt, M. G. Barton, Reproductive isolation among endemic pupfishes (Cyprinodon) on San Salvador Island, Bahamas: Microsatellite evidence. *Biol. J. Linn. Soc.* **95**, 566–582 (2008).
43. C. H. Martin, P. C. Wainwright, Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of cyprinodon pupfish. *Evolution* **65**, 2197–2212 (2011).
44. C. H. Martin, The cryptic origins of evolutionary novelty: 1000-fold faster trophic diversification rates without increased ecological opportunity or hybrid swarm. *Evolution* **70**, 2504–2519 (2016).
45. M. E. St. John, K. E. Dixon, C. H. Martin, Oral shelling within an adaptive radiation of pupfishes: Testing the adaptive function of a novel nasal protrusion and behavioural preference. *J. Fish Biol.* **97**, 163–171 (2020).
46. C. H. Martin, J. A. McGirr, E. J. Richards, M. E. St. John, How to investigate the origins of novelty: Insights gained from genetic, behavioral, and fitness perspectives. *Integr. Org. Biol.* **1** (2019).
47. M. E. St. John, R. Holzman, C. H. Martin, Rapid adaptive evolution of scale-eating kinematics to a novel ecological niche. *J. Exp. Biol.* **223** (2020).
48. C. H. Martin, L. C. Feinstein, Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Mol. Ecol.* **23**, 1846–1862 (2014).
49. E. J. Richards, *et al.*, A vertebrate adaptive radiation is assembled from an ancient and disjunct spatiotemporal landscape. *Proc. Natl. Acad. Sci.* **118**, e2011811118 (2021).
50. C. H. Martin, P. C. Wainwright, On the measurement of ecological novelty: Scale-eating pupfish are separated by 168 my from other scale-eating fishes. *PLoS One* **8**, e71164 (2013).

51. R. Froese, D. Pauldy, FishBase. *FishBase* (2021).
52. O. Seehausen, others, Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).
53. S. H. Martin, C. D. Jiggins, Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* **47**, 69–74 (2017).
54. D. A. Marques, J. I. Meier, O. Seehausen, A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
55. J. W. Poelstra, E. J. Richards, C. H. Martin, 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).
56. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
57. K. A. Thompson, M. M. Osmond, D. Schluter, Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* **3**, 129–141 (2019).
58. J. B. Pease, D. C. Haak, M. W. Hahn, L. C. Moyle, Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* **14**, e1002379 (2016).
59. E. J. Richards, J. W. Poelstra, C. H. Martin, Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evol. Lett.* **2**, 524–540 (2018).
60. J. I. Meier, *et al.*, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
61. M. D. McGee, *et al.*, The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**, 75–79 (2020).
62. I. Irisarri, *et al.*, Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**, 1–12 (2018).
63. E. J. Richards, C. H. Martin, Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* **13**, e1006919 (2017).
64. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
65. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 1–6 (2011).
66. J. A. McGirr, C. H. Martin, Ecological divergence in sympatry causes gene misexpression in hybrids. *Mol. Ecol.* **29**, 2707–2721 (2020).
67. M. E. Arnegard, *et al.*, Genetics of ecological divergence during speciation. *Nature* **511**, 307–311 (2014).
68. S. Wang, J. J. Hard, F. Utter, Genetic variation and fitness in salmonids. *Conserv. Genet.* **3**, 321–333 (2002).
69. R. Leimu, P. Mutikainen, J. Koricheva, M. Fischer, How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol.* **94**, 942–952 (2006).
70. D. Schluter, *et al.*, Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1914889118 (2021).
71. C. H. Martin, P. A. Erickson, C. T. Miller, The genetic architecture of novel trophic specialists: larger effect sizes are associated with exceptional oral jaw diversification in a pupfish adaptive radiation. *Mol. Ecol.* **26**, 624–638 (2017).
72. J. A. McGirr, C. H. Martin, Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies. *Evol. Lett.* **2**, 62–75 (2018).
73. J. Kulmuni, A. M. Westram, Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Mol. Ecol.* **26**, 3093–3103 (2017).
74. Y. Nakamura, *et al.*, Wwp2 is essential for palatogenesis mediated by the interaction between

Sox9 and mediator subunit 25. *Nat. Commun.* **2**, 1–10 (2011).
75. L. Mork, G. Crump, "Zebrafish craniofacial development: A Window into early patterning" in *Current Topics in Developmental Biology*, Y. Chai, Ed. (Academic Press Inc., 2015), pp. 235–269.
76. S. J. Lim, *et al.*, Taurine is an essential nutrient for juvenile parrot fish Oplegnathus fasciatus. *Aquaculture* **414–415**, 274–279 (2013).
77. T. G. Gaylord, A. M. Teague, F. T. Barrows, Taurine supplementation of all-plant protein diets for rainbow trout (Oncorhynchus mykiss). *J. World Aquac. Soc.* **37**, 509–517 (2006).
78. M. Yokoyama, T. Takeuchi, G. S. Park, J. Nakazoe, Hepatic cysteinesulphinate decarboxylase activity in fish. *Aquac. Res.* **32**, 216–220 (2001).
79. Z. Zhou, *et al.*, Genome-wide association study of growth and body-shape related traits in large yellow croaker (Larimichthys crocea) using ddRAD sequencing. *Mar. Biotechnol.* **21**, 655–670 (2019).
80. G. M. Yoshida, J. M. Yáñez, Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* **22**, 1–13 (2021).
81. L. F. S. Fonseca, *et al.*, Gene expression profiling and identification of hub genes in Nellore cattle with different marbling score levels. *Genomics* **112**, 873–879 (2020).
82. P. Pavlidis, D. Živković, A. Stamatakis, N. Alachiotis, SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
83. N. Alachiotis, A. Stamatakis, P. Pavlidis, OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
84. F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, S. J. Tans, Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
85. L. Ferretti, D. Weinreich, F. Tajima, G. Achaz, Evolutionary constraints in fitness landscapes. *Heredity (Edinb).* **121**, 466–481 (2018).
86. T. C. Nelson, *et al.*, Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). *PLoS Genet.* **17**, e1009095 (2021).
87. H. Svardal, *et al.*, Ancestral hybridization facilitated species diversification in the lake malawi cichlid fish adaptive radiation. *Mol. Biol. Evol.* **37**, 1100–1113 (2020).
88. M. Moest, *et al.*, Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. *PLoS Biol.* **18**, e3000597 (2020).
89. P. R. Grant, B. R. Grant, Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23216–23224 (2019).
90. M. Wu, J. L. Kostyun, M. W. Hahn, L. C. Moyle, Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol. Ecol.* **27**, 3301–3316 (2018).
91. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
92. J. Hermisson, P. S. Pennings, Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**, 700–716 (2017).
93. P. W. Hedrick, Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618 (2013).
94. S. Matuszewski, J. Hermisson, M. Kopp, Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* **200**, 1255–1274 (2015).
95. J. T. Streelman, P. D. Danley, The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**, 126–131 (2003).
96. Broad Institute, Picard Toolkit (2018).
97. M. A. Depristo, *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
98. R. Poplin, *et al.*, Scaling accurate genetic variant discovery to tens of thousands of samples.

bioRxiv, 201178 (2017).
99. C. D. Marsden, *et al.*, Diversity, differentiation, and linkage disequilibrium: Prospects for association mapping in the malaria vector anopheles arabiensis. *G3 Genes, Genomes, Genet.* **4**, 121–131 (2014).
100. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
101. S. Purcell, *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
102. C. T. DiVittorio, *et al.*, Natural selection maintains species despite frequent hybridization in the desert shrub Encelia. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33373–33383 (2021).
103. J. Hereford, A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**, 579–588 (2009).
104. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
105. P. Cingolani, *et al.*, Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3** (2012).
106. S. X. Ge, D. Jung, D. Jung, R. Yao, ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).
107. J. A. McGirr, C. H. Martin, Few fixed variants between trophic specialist pupfish species reveal candidate cis-regulatory alleles underlying rapid craniofacial divergence. *Mol. Biol. Evol.* **38**, 405–423 (2021).
108. T. L. Hedrick, Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspiration and Biomimetics* **3**, 034001 (2008).
109. S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **73**, 3–36 (2011).
110. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference* (Springer New York, 2002).
111. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1–9 (2006).
112. F. de Mendiburu, agricolae: Statistical procedures for agricultural research (2020).
113. T. Guillerme, dispRity: A modular R package for measuring disparity. *Methods Ecol. Evol.* **9**, 1755–1763 (2018).
114. B. J. Knaus, N. J. Grünwald, VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).
115. S. Der Yang, T. S. Lin, F. G. Liu, C. H. Liou, Influence of dietary phosphorus levels on growth, metabolic response and body composition of juvenile silver perch (Bidyanus bidyanus). *Aquaculture* **253**, 592–601 (2006).
116. J. M. Costa, *et al.*, Inadequate dietary phosphorus levels cause skeletal anomalies and alter osteocalcin gene expression in Zebrafish. *Int. J. Mol. Sci.* **19**, 364 (2018).
117. H. Yu, *et al.*, Effects of stocking density and dietary phosphorus levels on the growth performance, antioxidant capacity, and nitrogen and phosphorus emissions of juvenile blunt snout bream (Megalobrama amblycephala). *Aquac. Nutr.* **27**, 581–591 (2021).