

**Hybridization alters the shape of the genotypic fitness landscape, increasing
access to novel fitness peaks during adaptive radiation**

Austin H. Patton^{1,2}, Emilie J. Richards^{1,2}, Katelyn J. Gould³, Logan K. Buie³, Christopher H.
Martin^{1,2}

¹*Department of Integrative Biology, University of California, Berkeley, CA*

²*Museum of Vertebrate Zoology, University of California, Berkeley, CA*

³*Department of Biology, University of North Carolina, Chapel Hill, NC*

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

Correspondence: austinpatton@berkeley.edu, chmartin@berkeley.edu

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. However, adaptive walks connecting genotypes to organismal fitness, speciation, and novel ecological niches are still poorly understood and processes for surmounting fitness valleys remain controversial. One outstanding system for addressing these connections is a recent adaptive radiation of ecologically and morphologically novel pupfishes (a generalist, molluscivore, and scale-eater) endemic to San Salvador Island, Bahamas. We leveraged whole-genome sequencing of 139 hybrids from two independent field fitness experiments to identify the genomic basis of fitness, estimate genotypic fitness networks, and measure the accessibility of adaptive walks on the fitness landscape. We identified 132 SNPs that were significantly associated with fitness in field enclosures. Six out of the 13 regions most strongly associated with fitness contained differentially expressed genes and fixed SNPs between trophic specialists; one gene (*mettl21e*) was also misexpressed in lab-reared hybrids, suggesting a potential intrinsic genetic incompatibility. We then constructed genotypic fitness networks from adaptive alleles and show that scale-eating specialists are the most isolated of the three species on these networks. Intriguingly, introgressed and *de novo* variants reduced fitness landscape ruggedness as compared to standing variation, increasing the accessibility of genotypic fitness paths from generalist to specialists. Our results suggest that adaptive introgression and *de novo* mutations alter the shape of the fitness landscape, providing key connections in adaptive walks circumventing fitness valleys and triggering the evolution of novelty during adaptive radiation.

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, a 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 phenotypic evolution of populations through time on a rugged landscape, in which isolated clusters of fitness peaks represent ‘adaptive zones’ relative to 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 systems (12–18). Fitness surfaces are also central components of speciation models and theory (19–21).

A central focus of fitness landscape theory is the characterization of the shape of the fitness landscape. Theoretical and empirical studies frequently attempt to describe its topography, such as quantifying the number of fitness peaks, one component of landscape ruggedness that affects the predictability of evolution (5, 22–24). Importantly, the existence of multiple peaks and valleys on the fitness landscape implies epistasis for fitness, or non-additive effects on fitness resulting from genotypic interactions (23, 25–28). Fitness epistasis reduces the predictability of evolution because the resultant increase in the number of peaks increases the number of viable evolutionary outcomes (8, 29). Increasing fitness epistasis also increases landscape ruggedness, thus reducing the probability of converging on any one fitness peak and ultimately diversifying potential evolutionary outcomes (8, 30).

This leads to a fundamental concept in fitness landscape theory: Not all genotypic pathways are evolutionarily accessible (5, 26, 31–36). In large populations, paths through genotype space that monotonically increase in fitness at each mutational step are favored over alternatives with neutral or deleterious steps (37). These accessible genotypic paths can be considered adaptive walks under Fisher’s geometric model, by which adaptation proceeds towards a phenotypic optimum via additive mutations of small phenotypic effect (37, 38). On rugged landscapes as originally envisioned by Wright (23), greater numbers of peaks (i.e. the ruggedness) increase the mean length of potential adaptive walks to any one fitness optimum, while decreasing the length of accessible paths to the nearest peak. Ultimately, this leads to a decrease in the probability that any one fitness optimum is reached. Simultaneously, increasing landscape ruggedness decreases the length of adaptive walks to the nearest local optimum, owing to the corresponding increase in peak density.

There are a growing number of experimental studies of adaptive walks in nature, including the evolution of toxin resistance in monarch butterflies (39), alcohol tolerance in *Drosophila* (40, 41), and host-shift in aphids (42). Likewise, the accessibility of genotypic fitness networks has now been explored in numerous microbial systems, including the evolution of antibiotic resistance (31), metabolism (43), citrate exploitation (44), and glucose limitation in *E. coli* (45), and adaptation to salinity in yeast via evolution of heat shock protein *Hsp90* (22). However, these studies are still limited to the investigation of specific coding substitutions and their effects on fitness in laboratory environments. Nosil et al. (46) estimated genotypic fitness networks for *Timema* stick insects based on a field experiment. Similarly, this study focused on a single large-effect locus underlying dimorphic coloration between ecotypes. These studies represent significant advances, but extension of fitness landscape theory to empirical systems

including multiple species remains an underexplored area of future research at the intersection of micro- and macroevolution. Such studies can provide insight into the topography of fitness landscapes in natural systems, the accessibility of interspecific adaptive walks, and ultimately the predictability of evolution.

One promising system for estimating fitness landscapes is a recent adaptive radiation of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas (17, 18, 47, 48). This radiation is comprised of two trophic specialists, a molluscivore (durophage: *Cyprinodon brontotheroides*) and a scale-eater (lepidophage: *C. desquamator*), derived from a Caribbean-wide generalist (*C. variegatus*) which also coexists in the same habitats. These three species all occur in sympatry in the hypersaline lakes of San Salvador Island, Bahamas (Fig 1a). Found in the benthic littoral zone of each lake, all three species forage within the same benthic microhabitat; indeed, no habitat segregation has been observed in 14 years of field studies. Originating less than 10,000 years ago (based on geological age estimates for the lakes: (49)), the functional and trophic novelty harbored within this radiation is the product of exceptional rates of craniofacial morphological evolution (50–53). Furthermore, species boundaries persist across multiple lake populations, despite persistent admixture among species (54, 55). 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 (Fig. 1b). Selection analyses revealed a multi-peaked phenotypic fitness landscape that is stable across lake populations, year of study, and manipulation of the frequency of rare hybrid phenotypes (17, 18, 48)). One of the strongest and most persistent trends across studies and treatments was that hybrid phenotypes resembling the scale-eater were isolated in the lowest fitness region for both growth and survival relative to the other two species (17, 18). In contrast, hybrids resembling the generalist occupied

a fitness peak and were separated by a smaller fitness valley from hybrids resembling the molluscivore, which occurred on a second peak of higher fitness.

Evolutionary trajectories through regions of low fitness should be inaccessible to natural selection. How then did an ancestral generalist population cross these phenotypic fitness valleys to reach new fitness peaks and adapt to novel ecological niches? A growing theoretical and empirical literature on fitness landscapes has demonstrated the limited conditions for crossing fitness valleys (56–59). Fitness peaks and valleys in morphospace may result only from the reduction of the adaptive landscape to two phenotypic dimensions (60). Additional phenotypic and genotypic dimensions may reveal fitness ridges that entirely circumvent fitness valleys (48, 61, 62). Indeed, owing to nonlinearity in the association between phenotype and fitness (63, 64), even a single-peaked phenotypic fitness landscape may be underlaid by a multi-peaked genotypic fitness landscape (65, 66). In this respect, investigating the high-dimensional genotypic fitness landscape 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 (67).

Furthermore, the relative contributions of standing genetic variation, *de novo* mutations, and adaptive introgression to the tempo and mode of evolution are now of central interest to the field of speciation genomics (68–72). The three-dimensional adaptive landscape metaphor is often invoked to explain how the genetic, phenotypic, and ecological diversity introduced to populations by hybridization facilitates the colonization of neighboring fitness peaks that are unoccupied by either hybridizing species (73–75). However, extension of these ideas to more high-dimensional genotypic fitness landscapes remains underexplored. For instance, we have yet to learn how the appearance of novel adaptive genetic variation through introgressive

hybridization or *de novo* mutation alters epistatic interactions among loci, thus potentially altering the shape of the fitness landscape and the accessibility of interspecific adaptive walks.

The adaptive radiation of San Salvador Island pupfishes, like many others (76–80), appears to have originated from a complex interplay of abundant standing genetic variation, adaptive introgression from neighboring islands, and several *de novo* single-nucleotide mutations and deletions found only in the scale-eater (55, 81). Notably, both specialists harbor numerous introgressed SNPs showing evidence of hard selective sweeps in the regulatory regions of known craniofacial genes (55, 81). In contrast, hard selective sweeps of *de novo* mutations only appear in the scale-eating species, *C. desquamator*. Here, we leverage whole genome sequencing of 139 hybrids measured in field experiments to identify the genomic basis of fitness differences, infer genotypic fitness networks, summarize their topography, and quantify the accessibility of novel fitness peaks and the influence of each source of genetic variation on interspecific adaptive walks.

Results

Sample collection and genomic resequencing

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 3 m diameter enclosures exposed for three months: Crescent Pond $n = 796$; Little Lake $n = 875$ F2 hybrids (17); 2014/2015: four high-density 4 m diameter enclosures exposed for three months in Crescent Pond, $n = 923$ F4/F5 hybrids and eleven months in Little Lake, $n = 842$ F4/F5 hybrids (18)). We then characterized patterns of genetic variation among parental species in each lake and their lab-reared hybrids used in field experiments. We genotyped 1,129,771 SNPs with an average coverage of 9.79x per individual.

Population structure and ancestry associations with fitness

Principal components analysis (PCA) of genetic variation strongly differentiated pupfishes sampled from Little Lake/Osprey Lake and Crescent Pond (PC1: 22.7% variance explained) and among species within each lake (PC2: 15.9% variance explained: Fig. 1d, S1-S2). These results were supported by ADMIXTURE analyses (82, 83) (Fig. 1e). However, some hybrids were genotypically transgressive, falling outside the genotypic distributions of the three parental species (Fig. S2), leading ADMIXTURE to assign the third cluster to these hybrids, rather than generalists which often contain segregating variation found in trophic specialists (67). This pattern persisted in a supervised ADMIXTURE analysis, in which we assigned individuals from the three parental species *a priori* to their own population and estimated admixture proportions for the remaining hybrids (Fig. S3). Pairwise genetic distances did not predict pairwise morphological distances (Fig. S4).

We analyzed three measures of fitness (growth, survival, and their composite: see Methods and Supplement for details), but focus herein on composite fitness, which is equal to growth for survivors and zero for non-survivors. Growth could not be measured for tagged hybrids that died in field enclosures and thus were not recovered. Because reproductive success was not possible to quantify in field experiments (due to continuous egg-laying and very small, newly hatched fry), composite fitness included only measurements of growth and survival.

Interestingly, in no case were genome-wide patterns of parental ancestry in hybrids (estimated from unsupervised ADMIXTURE analyses) associated with hybrid composite fitness (generalist $P = 0.385$; scale-eater $P = 0.439$; molluscivore $P = 0.195$), growth (generalist $P = 0.119$; scale-eater $P = 0.283$; molluscivore $P = 0.328$), or survival probability (generalist $P =$

0.440; scale-eater $P = 0.804$; molluscivore $P = 0.313$) while controlling for effects of lake and experiment (Fig. S5, Table S2). Similar results were obtained when repeating these analyses using admixture proportions estimated from a supervised ADMIXTURE analysis (Table S3), using only samples from the second field experiment (Table S4), or using principal component axes estimated from genome-wide SNPs (Table S5: See Supplementary Results). Therefore, in contrast to previous studies (84–86), in this system genome-wide ancestry is not consistently 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. (87)).

Genome-wide association mapping of fitness

From our LD-pruned 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-corrected threshold (Fig. 2a, Table S6-S7; see supplement for results for survival and growth alone [Tables S8-S9; Figure S6]). Gene ontologies for these 132 fitness-associated regions were significantly enriched for synaptic signaling and chemical synaptic transmission [False discovery rate (FDR) rate < 0.01 ; Fig. S8; Table S7]. Ontologies enriched at an FDR rate < 0.05 were related to signaling and regulation of cell communication (for growth, see Fig. S8). 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 (55, 81, 88). This suggests that fitness-associated regions in our field experiments captured additional components of fitness beyond the external morphological phenotypes measured in previous studies.

We characterized whether genes in or near fitness-associated regions were implicated in adaptive divergence of the specialists. Surprisingly, no fitness-associated regions overlapped with regions showing significant evidence of a hard selective sweep (55). However, six fitness-associated genes were previously shown to contain either fixed divergent SNPs (*csad*, *glcci1*, *ino80c*, *mag*, *pim2*, *mettl21e*), or a fixed deletion between specialists (*med25*) (88). *Med25* (Mediator Complex Subunit 25) is a craniofacial transcription factor associated with cleft palate in humans and zebrafish (89, 90); a precursor of *mag* (Myelin Associated Glycoprotein) was also associated with the parallel evolution of the thick-lipped phenotype in Midas cichlids based on differential expression among morphs (91). Three of the six remaining fitness-associated genes containing divergent SNPs (88) were associated with growth and/or body size measurements in other fishes. First, *csad* plays an important role in synthesizing taurine which is a rate-limiting enzyme affecting growth rate in parrotfishes (92), rainbow trout (93), and Japanese flounder (94). Second, *glcci1* is associated with the body depth/length ratio in yellow croaker (95). Third, *ino80c* is associated with measures of body size in Nile tilapia (96). Finally, *mettl21e* was differentially expressed among specialists and 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 that may impact their fitness in field enclosures (88, 97). Although it has not been associated with growth or body size in fishes, *mettl21e* is associated with intramuscular fat deposition in cattle (98). Taken together, these findings support the interpretation that fitness-associated regions are associated with unmeasured traits, particularly physiological growth rate, or craniofacial shape in the case of the deletion in *med25*, that affect fitness in our hybrid field experiments. However, the fitness associated loci we identified appear not to have the subject of selective sweeps in either specialist.

Fitness-associated SNPs improve inference of the adaptive landscape

Fitness landscapes in past studies were estimated using slightly different sets of morphological traits; thus, to enable inclusion of all hybrids on a single fitness landscape, a single observer (AHP) remeasured all sequenced hybrids for 31 morphological traits (Fig. S9; Table S10). 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)(Fig. S10-11; Tables S11-S13). 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 including fitness-associated SNPs were invariably favored over models with external morphology alone ($\Delta AICc > 8.6$: Tables S14-S15). Morphology-only models predicted a flat fitness surface (Fig. 2b, S11; predictions restricted to observed hybrid morphospace). In contrast, models including fitness-associated SNPs predicted a complex and nonlinear fitness landscape, despite our limited dataset of 139 sequenced hybrids relative to samples in previous morphology-only studies of 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 most significant fitness-associated SNPs in the full model. This reduced model was the best-fit; the inferred adaptive landscape was complex and characterized by a fitness peak near hybrids resembling the generalist phenotype separated by a small fitness valley from a second region of high fitness for hybrids resembling the molluscivore phenotype. Hybrids resembling the scale-eater phenotype again occurred in a large fitness valley (Fig. 2b-2d: For results pertaining to growth or survival see Supplemental Materials – Fig. S11, Table S11-15). Each of these fitness

peaks and valleys were frequently recovered across 10,000 bootstrap replicates; landscapes inferred from bootstrap replicates were often more complex with increased curvature relative to inferences from our observed dataset (Figure S12). Thus, the fitness landscape estimated from our observed dataset appears robust to sampling uncertainty.

Compared to previous studies, the highest fitness optimum was shifted from the molluscivore to the generalist phenotype. 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 non-morphological traits affecting fitness. Indeed, visualization of observed haplotypes in hybrids across the fitness landscape supported this interpretation; one of the most common haplotypes was most frequent in hybrids resembling generalists near the peak of high fitness and rare in hybrids resembling either trophic specialist (Fig. S13). Regardless, this two-dimensional phenotypic fitness landscape did not reveal fitness ridges connecting generalists to specialists, further emphasizing the need to investigate the genotypic fitness landscape.

Trophic novelty is associated with isolation on the genotypic fitness network

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* mutations (55). 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 knowledge gap, we first sought to visualize genotypic fitness networks and gain insight into how isolated the three species are in genotypic space. Understanding the relative isolation of

each specialist from the generalist can reveal the relative accessibility of their respective adaptive walks on the genotypic fitness landscape.

To accomplish this, we reconstructed genotypic fitness networks from 1,498 candidate adaptive alleles previously identified in this system (e.g. Fig. 3a (55)). These regions displayed significant evidence of a hard selective sweep using both site frequency spectrum and LD-based methods, SweeD (99) and OmegaPlus (100), and contained fixed or nearly fixed SNPs ($F_{ST} > 0.95$) differentiating trophic specialists across lakes (55). Adaptive alleles were classified as standing variation, introgressed, or *de novo* mutations based on extensive sampling of focal and related *Cyprinodon* pupfish species across San Salvador Island and neighboring Caribbean islands, as well as North and South American outgroups (55). We note, however, that adaptive alleles designated as *de novo* on San Salvador Island may be segregating at low frequencies in other sampled populations or present in unsampled populations.

These fitness networks depict both hybrids and parental species in genotypic space, with nodes representing SNP haplotypes and edges connecting mutational neighbors (Fig. 3b). Genotypic space is immense; using SNPs coded as homozygous reference, heterozygote, or homozygous alternate, the number of potential haplotypes is equal to $3^{\# \text{ SNPs in network}}$. For instance, to construct a reduced network of 100 SNPs, there are a total of $3^{100} = 5.17 \times 10^{57}$ possible nodes. Thus, unlike experimental studies of individual proteins in haploid *E. coli* (31, 45) or yeast (22), it is not possible for us to investigate the full breadth of genotypic space.

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 (e.g. Fig. 3b) and found that generalists and molluscivores are closer on the genotypic fitness

network than either is to scale-eaters (Fig. 3c), 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 $\pm$ 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: $P < 0.001$).

Molluscivore genotypes are more accessible to generalists than scale-eater genotypes on the genotypic fitness landscape

The most accessible paths through genotypic fitness networks are characterized by monotonically increasing fitness at each mutational step and the smallest possible number of steps between two states (5, 31, 33). Furthermore, as described earlier, the accessibility of individual fitness peaks is predicted to be reduced on increasingly rugged fitness landscapes that are characterized by a greater number of fitness peaks (8, 30, 33, 101). This provides three useful metrics of evolutionary accessibility for genotypic trajectories: 1) the total number of accessible paths relative to network size (Fig. S14; Table S16), 2) the length of the shortest accessible paths, and 3) the number of fitness peaks (ruggedness). Here, we define peaks as genotypes with no fitter neighbors and within a single mutational step (32). With these three metrics, we can quantify the accessibility of interspecific genotypic pathways.

We used these measures of accessibility to ask: 1) whether molluscivore or scale-eater genotypes were more accessible to generalists on the fitness landscape (Fig. 4c-d), and 2) whether molluscivore and scale-eater genotypic fitness networks differed in their ruggedness,

characterized by peak number (Fig. 4e-g). These measures provide insight into the predictability of evolution and the role that epistasis plays in their evolution (8, 23, 29, 102).

We constructed 5,000 genotypic fitness networks from a random sample of five species-specific candidate adaptive SNPs (Fig. 3a) for either molluscivores or scale-eaters, requiring that at least one SNP of each source of genetic variation be present in the sample. We used odds ratios to compare the relative accessibility and ruggedness of molluscivore fitness networks compared to scale-eater networks (Fig. 4h). Thus, odds ratios (OR) greater than 1 imply summary statistics are greater for molluscivores than for scale-eaters.

We found that molluscivore genotypes were significantly more accessible to generalists on the fitness landscape than scale-eaters (Table S17); molluscivore networks had significantly more accessible paths [OR: (95% CI) = 2.095: (1.934, 2.274)] that were significantly shorter [OR and 95% CI = 0.253: (0.231, 0.277)]. Not only were molluscivore genotypes more accessible to generalists, but molluscivore fitness networks were significantly less rugged than scale-eater networks, comprised of fewer peaks [OR and 95% CI = 0.604: (0.575, 0.634)], and connected by significantly more accessible paths [OR and 95% CI = 1.514: (1.404, 1.635)], that contained fewer mutational steps [OR and 95% CI = 0.539: (0.500, 0.579)].

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

We further used our two metrics of accessibility and landscape ruggedness to ask how different sources of adaptive genetic variation may influence the topography of the fitness landscape, the traversal of fitness paths separating generalists from specialists, and ultimately colonization of novel fitness peaks. We constructed genotypic fitness networks limited to only one 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 all combinations of these three sources to better reflect the actual process of adaptive divergence originating from only standing genetic variation, then adaptive introgression plus standing genetic variation, and finally the refinement stage of *de novo* mutations (55).

We compared sets of 5,000 random 5-SNP genotypic networks drawn from different sources of adaptive variation (Fig. 4a) and compared the effect of each source of variation on measures of accessibility and landscape ruggedness relative to standing genetic variation. We treated standing variation as our basis for comparison because this is the source of genetic variation first available to natural selection (103).

We discovered that genotypic trajectories between generalists and either trophic specialist in genotypic fitness networks constructed from introgressed or *de novo* adaptive mutations were significantly more accessible than networks constructed from standing genetic variation (Fig. 5). Specifically, random networks that included alternate sources of adaptive variation contained significantly more accessible fitness paths from generalist to specialists than networks constructed from standing genetic variation alone, while controlling for differences in overall network size (Fig. 5a; Table S18). Furthermore, accessible paths between generalists and specialists in networks constructed from introgressed or *de novo* adaptive loci were significantly shorter in length (Fig. 5b). We recovered the same pattern whether constructing fitness networks from these sources of variation alone or in combination. These results held across all measures of fitness and for analyses repeated using only hybrids sampled from the second field experiment (Figs. S15-S16, Table S18-19).

Our finding of increased accessibility of interspecific genotypic trajectories suggests that fitness landscapes constructed from adaptive standing genetic variation alone are more rugged

than networks including adaptive loci originating from either introgression or *de novo* mutation. Quantification of landscape ruggedness supported this hypothesis in all cases (Fig. 5c; Table S18-S19). Additionally, increasing landscape ruggedness significantly decreased the length of accessible paths to the nearest local peak [glm(Min. Path Length ~ # of Peaks, family = “poisson”): $P < 0.0001$, $\beta = -0.088$, 95% CI = -0.095 – 0.081].

Scale-eater fitness genotypic fitness landscapes constructed from a combination of adaptive loci sourced from standing variation, introgression, and *de novo* mutations had significantly more accessible paths (scaled by network size) separating generalists from scale-eaters [OR and 95% CI = 1.879: (1.743, 2.041); LRT $P < 0.0001$; Fig. 5a] and these paths were significantly shorter in length compared to networks constructed from standing variation alone [OR and 95% CI = 0.876: (0.823, 0.932); LRT $P < 0.0001$; Fig. 5b]. The only exception to this pattern across all three fitness measures was for growth rate in genotypic fitness networks constructed for molluscivore adaptive loci; no significant difference was observed in the length of the shortest accessible path between networks constructed using standing variation alone or those constructed using introgressed alleles [OR and 95% CI = 0.994: (0.915, 1.079); LRT $P = 0.8826$; Table S18]. Interestingly, however, for networks constructed from standing variation and introgressed alleles, we again observed a significant reduction in length of the shortest accessible paths [OR and 95% CI = 0.897: (0.835, 0.962); LRT $P = 0.0050$; Table S18].

Discussion

We developed a new approach for estimating genotypic fitness landscapes for diploid organisms and applied it to a system in which phenotypic fitness landscapes have been extensively

investigated. We were able to address long-standing questions posed by fitness landscape theory in an empirical system and assess the extent to which the shape of the fitness landscape and accessibility of adaptive walks are contingent upon the source of adaptive genetic variation. We show that not only are scale-eaters more isolated than molluscivores from generalists on the fitness landscape, but that the scale-eater fitness landscape is more rugged than molluscivores. This indicates that epistasis is more pervasive on the scale-eater fitness landscape, leading to less predictable evolutionary outcomes and fewer accessible trajectories from generalist to scale-eater genotypes. Overall, we found that most genotypic trajectories were inaccessible and included one or more mutational steps that decreased in fitness from generalist to specialist. This finding is consistent with the patterns observed by Weinreich et al. (31), who constructed combinatorially complete fitness networks for five mutations contributing to antibiotic resistance in *E. coli* and found that only 18 of 120 possible genotypic trajectories were evolutionarily accessible. In contrast, Khan et al. (45) estimated that over half of all trajectories were accessible on a complete fitness landscape constructed using the first five adaptive mutations to fix in an experimental population of *E. coli*.

We also show that fitness landscapes are most rugged, and therefore epistasis is most pervasive, when constructed from standing genetic variation alone, ultimately leading to a reduction in the accessibility of fitness peaks on these landscapes (Fig. 5). This finding has significant implications for the predictability of evolution in the earliest stages of the speciation process. Adaptation from standing genetic variation is thought to initially be more rapid due to its initial availability and potentially reduced genetic load within a population (103–105). In contrast, we consistently found that networks constructed from a combination of adaptive standing variation, introgression, and *de novo* mutations reduced the ruggedness of fitness

landscapes and thus increased accessibility of interspecific evolutionary trajectories (Fig. 5). This would suggest that adaptive introgression or *de novo* mutations reduce the impacts of epistasis, resulting in a smoother fitness landscape with a greater number of accessible adaptive walks, facilitating the colonization of new adaptive zones. Future studies testing the generality of these findings will be invaluable for our understanding of the speciation process.

Furthermore, our results shed light on the classic problem of crossing fitness valleys on three-dimensional phenotypic fitness landscapes. We show that phenotypic fitness valleys may be circumvented by rare accessible paths on the genotypic fitness landscape. These results are consistent with increasing recognition that three-dimensional depictions of the fitness landscape may lead to incorrect intuitions about how populations evolve (3, 5, 106).

Our study represents a significant contribution to the growing body of work applying fitness landscape theory to empirical systems (39, 46, 107–109). Unlike previous studies that experimentally generated combinatorically complete fitness landscapes (22, 31, 45), we subsampled loci across the genome, enabling us to quantify aspects of the genotypic fitness landscape, despite the limitations imposed by large genome sizes and non-model vertebrates. One limitation of this approach is that subsampled fitness networks may not directly correspond to the full landscape (5, 110). For instance, a given subsampled fitness landscape may be present on multiple global, fully sampled fitness landscapes (110). Secondly, nodes (here, SNP haplotypes) can appear disconnected in a subsampled fitness landscape, but may be connected in the full fitness landscape (5). Nevertheless, given that there are more possible genotypes for a gene of 1,000 base-pairs than particles in the known universe (1, 111), nearly all empirical fitness landscapes must necessarily be subsampled at some scale.

Although inferences from subsampled fitness networks have their limitations, so too do those obtained from combinatorically complete fitness landscapes, which may themselves be misleading (102). By including mutations that are not segregating in natural populations, the shape of the “complete” fitness landscape and thus accessibility of fitness peaks may be quite different from what occurs in nature. The shape of fitness landscapes in nature is dictated by the “realized” epistasis that occurs among naturally segregating loci (102). Changes to “realized” epistasis induced by introgression or *de novo* mutations appears to be the mechanism altering the shape of the fitness landscape and thus accessibility of fitness peaks. Our findings that adaptive introgression and *de novo* mutations make fitness peaks more accessible points towards a pervasive role of epistasis in determining the predictability of evolution and the speciation process (5, 8, 22–24, 30).

In the present study we have taken snapshots of the fitness landscape from loci that have already undergone hard selective sweeps. Consequently, we cannot directly assess the influence of each adaptive allele on the fitness landscape through time as it increases in frequency. However, so far we have failed to detect evidence of frequency-dependent selection in this system after experimental manipulations, at least for morphological traits (18). Future experimental or simulation studies may track how novel adaptive alleles affect fitness landscape topography as they increase in frequency.

Conclusion

Our findings are consistent with a growing body of evidence that *de novo* and introgressed adaptive variation may contribute to rapid speciation and evolution towards novel fitness peaks (44, 70, 78, 112–116). We demonstrate that adaptive introgression smooths the fitness landscape

and increases the accessibility of fitness peaks. This provides an alternative mechanism to explain why hybridization appears to play such a pervasive role in adaptive radiation and speciation. There are many examples of hybridization promoting or inducing rapid speciation and adaptive radiation. Whether in Galapagos finches (117), African cichlids (77, 78, 113, 118, 119), or *Heliconius* butterflies (75, 120), hybridization has been shown to play a generative role in adaptive radiation and the evolution of novelty. One mechanism is the increased genotypic, phenotypic, and ecological diversity generated by hybridization in the form of transgressive phenotypes (74, 121–124). This diversity in turn facilitates the colonization of novel fitness peaks and ecological niches, particularly after colonization of a new environment rich in ecological opportunity (73, 74, 125). However, this model often assumes that the fitness landscape remains static after adaptive introgression. Here we show that adaptive introgression directly alters the shape of the fitness landscape, making novel fitness peaks more accessible to natural selection. Thus, hybridization not only generates genetic diversity, but this diversity can alter the shape of the fitness landscape, changing which genotypic combinations are favored by natural selection along with the adaptive walks that lead to them.

Acknowledgements:

We thank Michelle St. John, David Tian, Jacqueline Galvez, and Maria Fernanda Palominos for insightful discussion of the results; the Gerace Research Centre and Troy Day for logistical support; the BEST Commission and 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, graduate research funding from the Museum of Vertebrate Zoology to EJR, and an NSF Postdoctoral Research Fellowship in Biology under Grant No. 2109838 to AHP.

Data availability:

Genomic data are archived at the National Center for Biotechnology Information BioProject Database (Accessions: PRJNA690558; PRJNA394148, PRJNA391309). Sample metadata including morphological measurements and admixture proportions have been uploaded to Dryad: <https://doi.org/10.5061/dryad.0vt4b8h0m>.

Methods:

Sampling

Our final genomic dataset was 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 (N = 71) and Crescent Pond (N = 68). Hybrids used in the first field experiment (17) were comprised of F2 and backcrossed outbred juveniles resulting from crosses between all three species. Juveniles were raised for 2 months in the lab, individually tagged by injecting a stainless steel sequential coded wire tag (Northwest Marine Technologies, Inc.) into their left dorsal musculature, and photographed pre-release for morphometric analyses. Experimental field enclosures consisted of high- and low-density treatments; density was varied by the number of

tagged juveniles released into each enclosure. Hybrids in the second field experiment (18) were comprised of F4-F5 outbred juveniles resulting from crosses between all three species. Individuals were spawned, raised, tagged, and photographed in the same way prior to release. The second field experiment consisted of high- and low-frequency treatments of approximately equal densities. The frequency of rare transgressive hybrid phenotypes was manipulated between treatments in each lake, such that the high- and low-frequency treatments harbored an artificially increased and decreased frequency of transgressive phenotypes, respectively (18).

All hybrids were measured for 32 external morphological traits (see below). Additionally, we sequenced parental species of the generalist (N = 17), molluscivores (N = 27), and scale-eaters (N = 25) sampled from these two lakes and previously included in Richards et al. (55). Note that we treated samples from Little Lake and Osprey Lake as the same population because these two lakes are connected through a sand bar and fish from these populations are genetically undifferentiated (54, 55). For morphological analyses, we additionally measured samples of 60 generalists, 38 molluscivores, and 60 scale-eaters raised in the same laboratory common garden environment as the hybrids used in field experiments. A full list of samples is 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) (55) using bwa-mem (v. 0.7.2). Duplicate reads were identified using MarkDuplicates and BAM indices were subsequently generated using the Picard software package (126). Samples were genotyped following Richards et al. (55) according to GATK best practices (127). 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 (127–129).

Following initial genotyping with GATK, we subsequently filtered our data further using VCFtools (130). 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: (131)). This resulted in the retention of 1,129,771 SNPs across 139 hybrid samples and the 69 wild-caught samples from Richards et al (55). Unless otherwise specified, these SNPs were used for all downstream analyses.

Hybrid fitness measures

We used three proxies for fitness: survival, growth, or a composite measure of the two. Survival was a binary trait indicating whether a fish survived (i.e. a tagged fish was recovered) 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 (132), and analogous to composite fitness in Hereford (133) who used fecundity as their second fitness measure, rather than growth. Composite fitness is equal to growth for survivors and equals zero for non-survivors because growth could not be assessed for non-surviving individuals. 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 (131), 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 (82). Populations of each species were substantially differentiated between Crescent Pond and Little Lake (54, 55); 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. Composite fitness was analyzed using a tobit (zero-censored) model to account for zero-inflation using the *censReg* R package (134), survival was analyzed using a binomial model, and growth was analyzed using a gaussian model. We conducted additional ADMIXTURE analyses that either 1) were supervised, with generalist, molluscivore, and scale-eater parentals *a priori* assigned to one of three populations, with only hybrid ancestry proportions being estimated by admixture, or 2) using only samples from the second field experiment. The same linear models described above were subsequently repeated using these alternative admixture proportions.

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: (135)). 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 with 1) a False Discovery Rate (FDR: Benjamini and Hochberg 1995) less than 0.05, or a 2) P -value < 0.05 following Bonferroni correction. We focused 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 (136) and the annotated *C. brontotheroides* reference genome (55). We constructed a custom database within snpEff using the functional annotations originally produced by Richards et al. (55), and subsequently extracted information on the annotations and putative functional consequences of each variant.

Using genes identified for each SNP that was significantly associated with one of the fitness measures, we performed Gene Ontology (GO) enrichment analyses using ShinyGO v0.61 (137). 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 (55, 88, 138).

Morphometrics

We measured 31 external morphological traits for all 139 hybrids and 69 parental individuals from Crescent Pond (30 generalists, 19 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 (139). Measurements included 27 linear distances 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 $\log_{10}(\text{trait}) \sim \log_{10}(\text{standard length})$ 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 (140) in R to estimate fitness landscapes for the two discriminant axes (LD1-2) and fitness. All models included a thin-plate spline fit to the two linear discriminant axes and we included both lake and experiment in all models as fixed effects. Lake by experiment interaction terms were also included in some models. Models were ranked using the corrected Akaike Information Criterion for small sample sizes (AICc) and were considered to be a substantially worse fit to the data if $\Delta\text{AICc} > 4$ from the best-fit model (141). The best-fit model from the above approach was in turn used to visualize fitness landscapes, plotting predicted values of fitness measures on the two discriminant axes in R (Fig. 2).

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 only growth: see section *Fitness-genotype association test*), and coded these as either reference, single, or double mutants using VCFtools (130). 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.

To quantify whether the local features of the complete fitness landscape constructed using all morphological variables and the most strongly fitness-associated SNPs were robust to sampling uncertainty, we conducted a bootstrapping procedure for this model. Specifically, we resampled hybrids with replacement 10,000 times and refit the full model. We then calculated

the mean predicted composite fitness for each linear discriminant (LD) axis in slices across the fitness landscape, both for our observed dataset, and for each bootstrap replicate. Slices divided the fitness landscape into thirds for each LD axis. We then quantified the mean and standard deviation of the predicted composite fitness for each position along the other LD axis.

We quantified uncertainty (mean $\pm$ SD) around local features of the bootstrapped fitness landscapes as compared to the observed values of predicted fitness for the same ‘slice’ of the fitness landscape. We predicted values at the same 30 points along each LD axis. We then plotted the locations of parents along the x-axis (LD1 or LD2) to enable relation of features on the fitness landscape to parental phenotypic distributions.

Estimation of genotypic fitness networks

We first estimated genotypic networks using sites previously shown to be highly divergent ($F_{ST} > 0.95$) and showing significant evidence of a hard selective sweep in one of the trophic specialists (based on evidence from both SweeD and OmegaPlus: (55, 99, 100)). We identified the SNPs in our unpruned full dataset overlapping with sites inferred to have undergone selective sweeps (55), 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 (142) 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. The reduced number of adaptive SNPs sites in our dataset as compared to that of (55) is due primarily to the increased stringency of our filtering. 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. (55). 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 5,000 datasets of five SNPs comprised of either 1) standing genetic variation, 2) adaptive introgression, 3) *de novo* mutation (scale-eaters only), 4) standing genetic variation + adaptive introgression, 5) standing genetic variation + *de novo* mutation, or 6) standing genetic variation + adaptive introgression + *de novo* mutation (scale-eaters only). We additionally repeated this procedure using both classes of SNPs for molluscivores to determine whether genotypic trajectories separating generalists to molluscivores are more accessible than those between generalists and scale-eaters. 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. When determining whether a path was accessible or not, we only included paths for which each mutational step (i.e. each intervening haplotype) between generalist to specialist was observed in at least one hybrid sample.

With these networks, we sought to ask 1) whether molluscivores or scale-eaters are more accessible to generalists on their respective genotypic fitness landscapes, 2) whether the ruggedness of the genotypic fitness landscape varied among specialists, and 3) whether accessibility is contingent upon the source of genetic variation available to natural selection. For each random network sampled 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, 4) the number of edges in the network, 5) the number of peaks on the landscape (genotypes with no fitter neighbors (32)), 6) the distance of parental nodes to these peaks, and 7) the number of accessible paths separating them. Larger networks often have a greater number of potential paths, including both accessible and inaccessible paths (Fig. S14), and we were interested in the relative availability of accessible adaptive pathways. Consequently, we divided the number of accessible paths in each random network sampled by the number of nodes. Using our six summary statistics, we tested whether accessibility and landscape ruggedness differed between networks constructed from SGV/Introgression/*De novo* mutations (for scale-eaters) or SGV/Introgression (for molluscivores). To do so, we calculated the mean and standard error of each summary statistic across all 5,000 replicates. We then

modeled the association between each summary statistic and species using a logistic regression, whereby species was modeled as a binary response variable (i.e. scale-eater networks = 0, molluscivore networks = 1), with each measure of accessibility as the predictor. We arbitrarily treated scale-eater networks as the control, and using the estimated coefficients obtained an odds ratio (OR) that corresponds to the extent to which molluscivore networks either have increased (OR > 1) or decreased (OR < 1) accessibility measures relative to scale-eater networks. Significance was similarly assessed using a likelihood ratio test. Additional details on this procedure may be found in the supplement. Using the fitted logistic model, we conducted a likelihood ratio test to quantify significance. To explicitly test the hypothesis that increasing landscape ruggedness reduced the length of accessible paths to the nearest fitness peak, we fit a Poisson regression model in R in which the number of fitness peaks predicts the length of the shortest accessible path between any generalist or specialist node and any fitness peak on that landscape: $\text{glm}(\text{Min. Distance to Peak} \sim \text{Number of Peaks}, \text{family} = \text{"poisson"})$.

A similar procedure was used to assess whether measures of accessibility (scaled number of accessible paths, length of the shortest accessible path) and landscape ruggedness (number of peaks) differed within species among networks constructed from different sources of genetic variation. Here, networks constructed from SGV were treated as the control, to which all other networks were compared. For example, to test whether accessibility of the generalist-to-scale-eater paths are greater in networks constructed from *de novo* mutations than those from SGV, a logistic model was fitted wherein the response variable for SGV networks was assigned to be 0, and 1 for *de novo* networks. As before, significance was similarly assessed using a likelihood ratio test, but here *P*-values were corrected for multiple testing using the false discovery rate (143). We assessed whether differences in these measures among the two alternate generalist to

737 specialist trajectories in networks constructed from all three sources of variation were significant
738 using an ANOVA in R (144). Due to the highly skewed nature of these distributions, post-hoc
739 pairwise significance was assessed using a nonparametric Kruskal-Wallis one-way analysis of
740 variance in the agricolae package (145) in R.

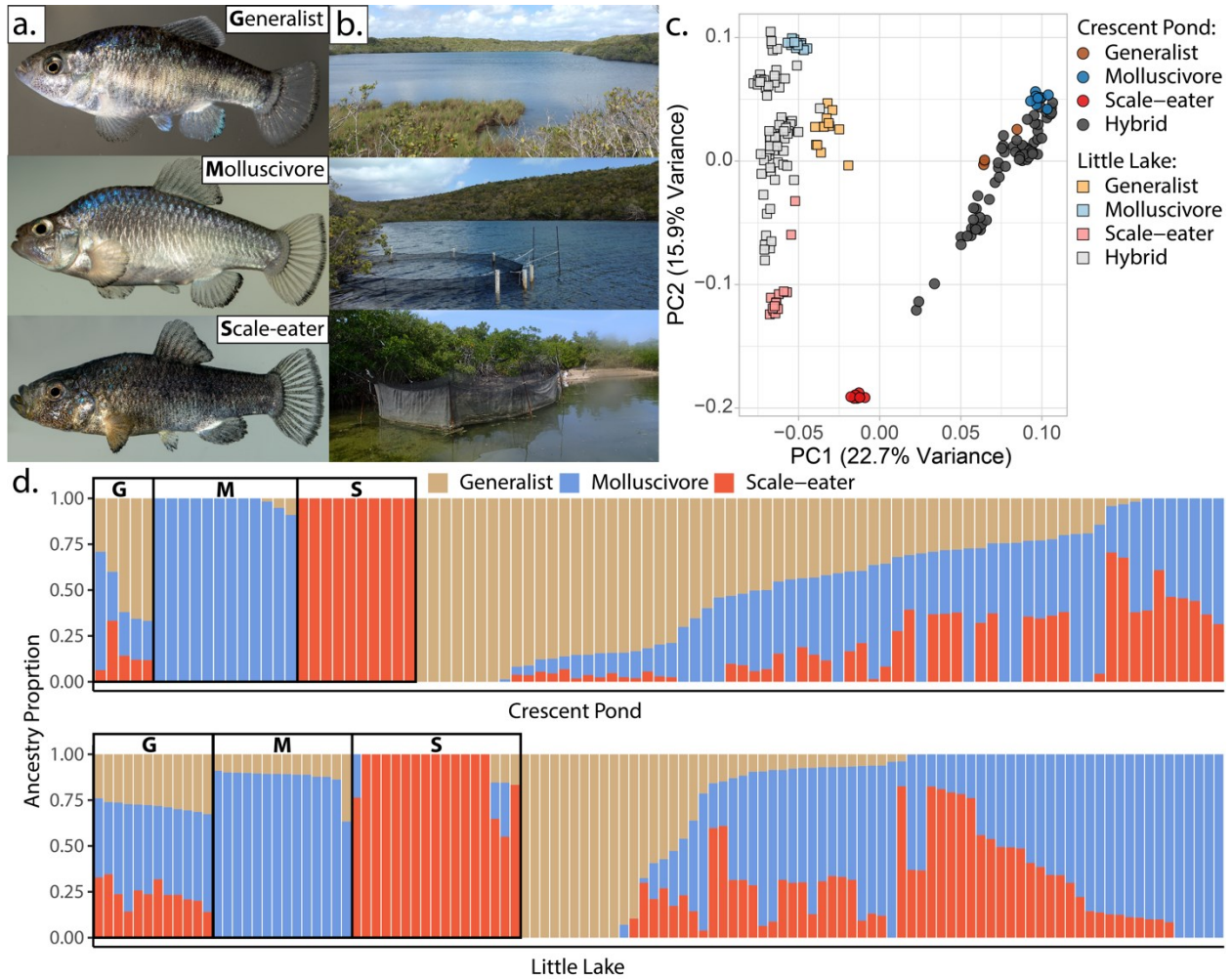


Figure 1. San Salvador Island pupfishes and their hybrids. **a.** From top to bottom: the generalist, *Cyprinodon variegatus*, the molluscivore *C. brontotheroides*, and the scale-eater *C. desquamator*. **b.** Representative images of experimental field enclosures. **c.** Principal component analysis of 1,129,771 LD-pruned SNPs genotyped in hybrids and the three parental species. **d.** Unsupervised 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, followed by all resequenced hybrid individuals from field experiments. Colors indicate ancestry proportions in each population ($K = 3$).

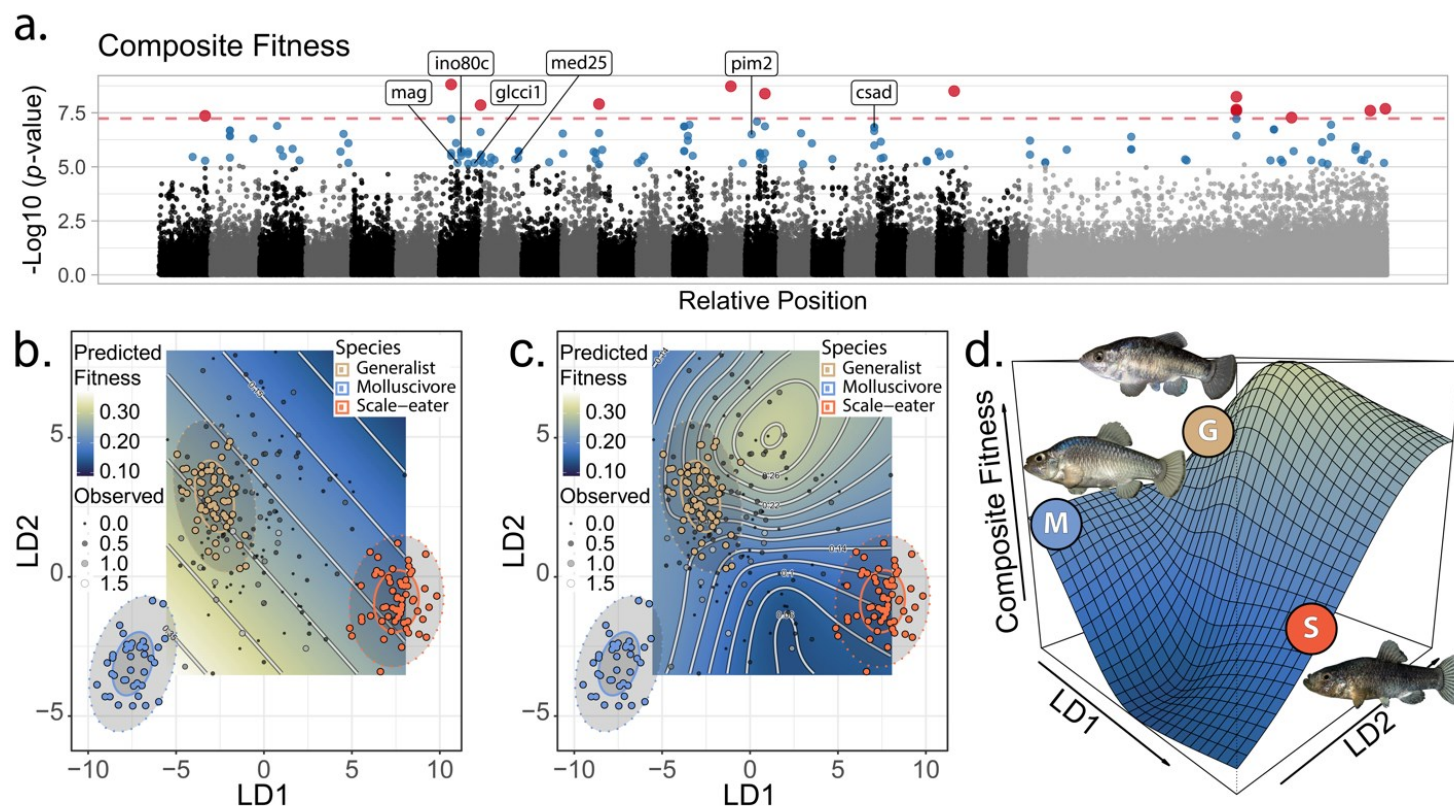


Figure 2. The genetic basis of fitness variation and improved inference of adaptive landscapes. **a)** Per-SNP $\log_{10} p$ -values from a genome-wide association test with GEMMA for composite fitness (survival x growth). Lake and experiment were included as covariates in the linear mixed model. SNPs that were significant at $FDR < 0.05$ are indicated in blue; red SNPs above dashed red line cross the threshold for Bonferroni significance at $\alpha = 0.05$. The first twenty-four scaffolds are sorted from largest to smallest and remaining scaffolds were pooled. The six genes associated with composite fitness which were both strongly differentiated ($F_{ST} > 0.95$) and differentially expressed between specialists (88) 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 generalized additive model (GAM) including a thin-plate spline for both LD axes, fixed effects of experiment and lake, and fixed effects of the seven (see supplementary methods) SNPs most strongly associated with fitness shown in red in panel **a**. **d)** Three-dimensional view of **c** with relative positions of the three parental phenotypes indicated.

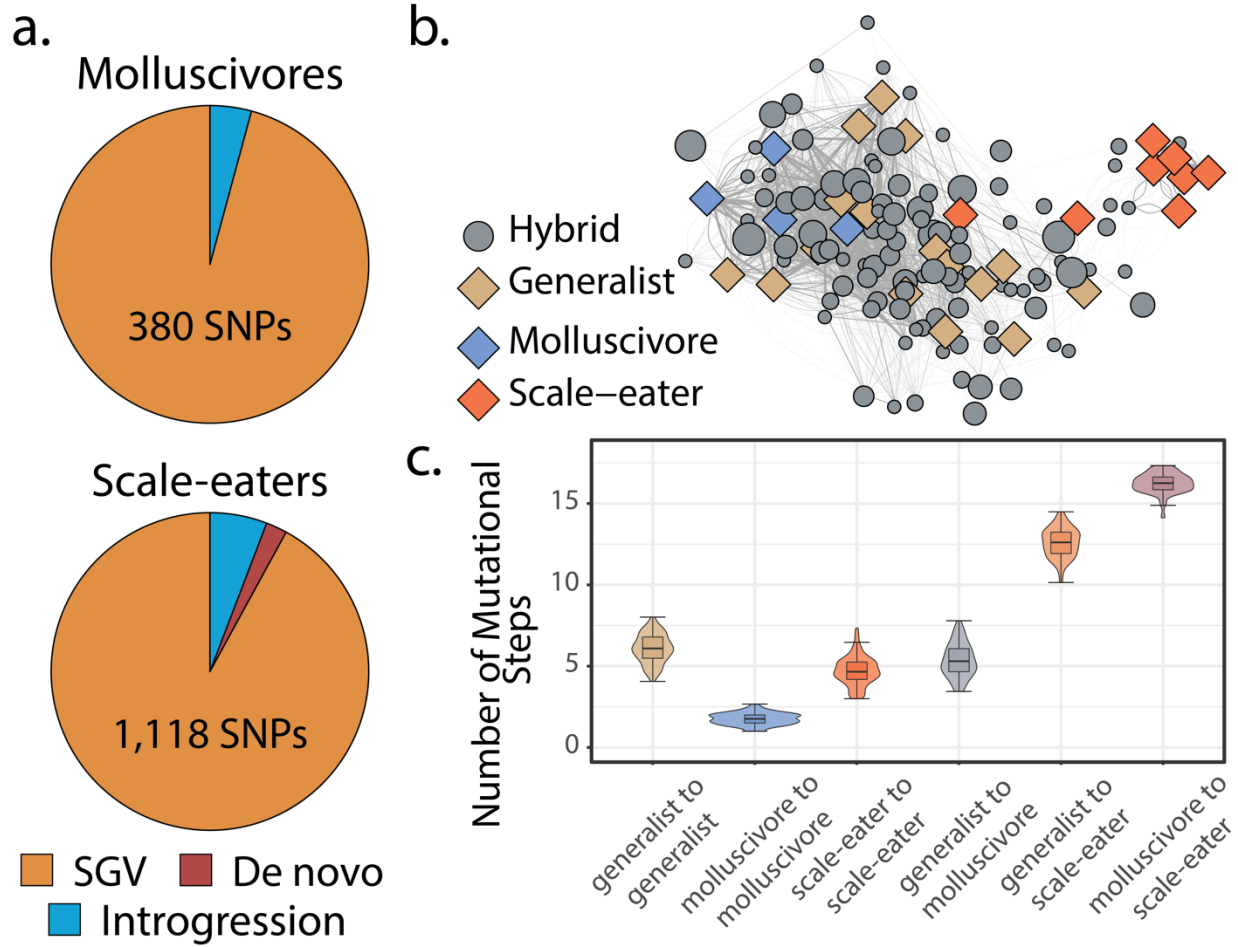


Figure 3. Scale-eaters are isolated on the fitness landscape. **a.** Most nearly fixed or fixed variants ($F_{ST} \geq 0.95$) experiencing hard selective sweeps (hereafter ‘adaptive alleles’) originated as standing genetic variation (SGV: molluscivores = 96%, scale-eaters = 92%), followed by introgression (molluscivores = 4%, scale-eaters = 6%), and *de novo* mutation (scale-eaters = 2%)(55). Pie charts show adaptive alleles retained in our study for each species; networks are constructed from either set of adaptive alleles. **b.** Genotypic network constructed from a random sample of ten SNPs, sampled from all SNPs shown in **a.** Each edge between nodes is up five mutational steps away; edge width is proportional to mutational distance: wider edges connect closer haplotypes; hybrid node size is proportional to fitness (larger nodes are of greater fitness value). **c.** Median number of mutational steps within or between species (e.g. Fig. 4a). All pairwise comparisons using Tukey’s HSD test (after FDR correction) were significant.

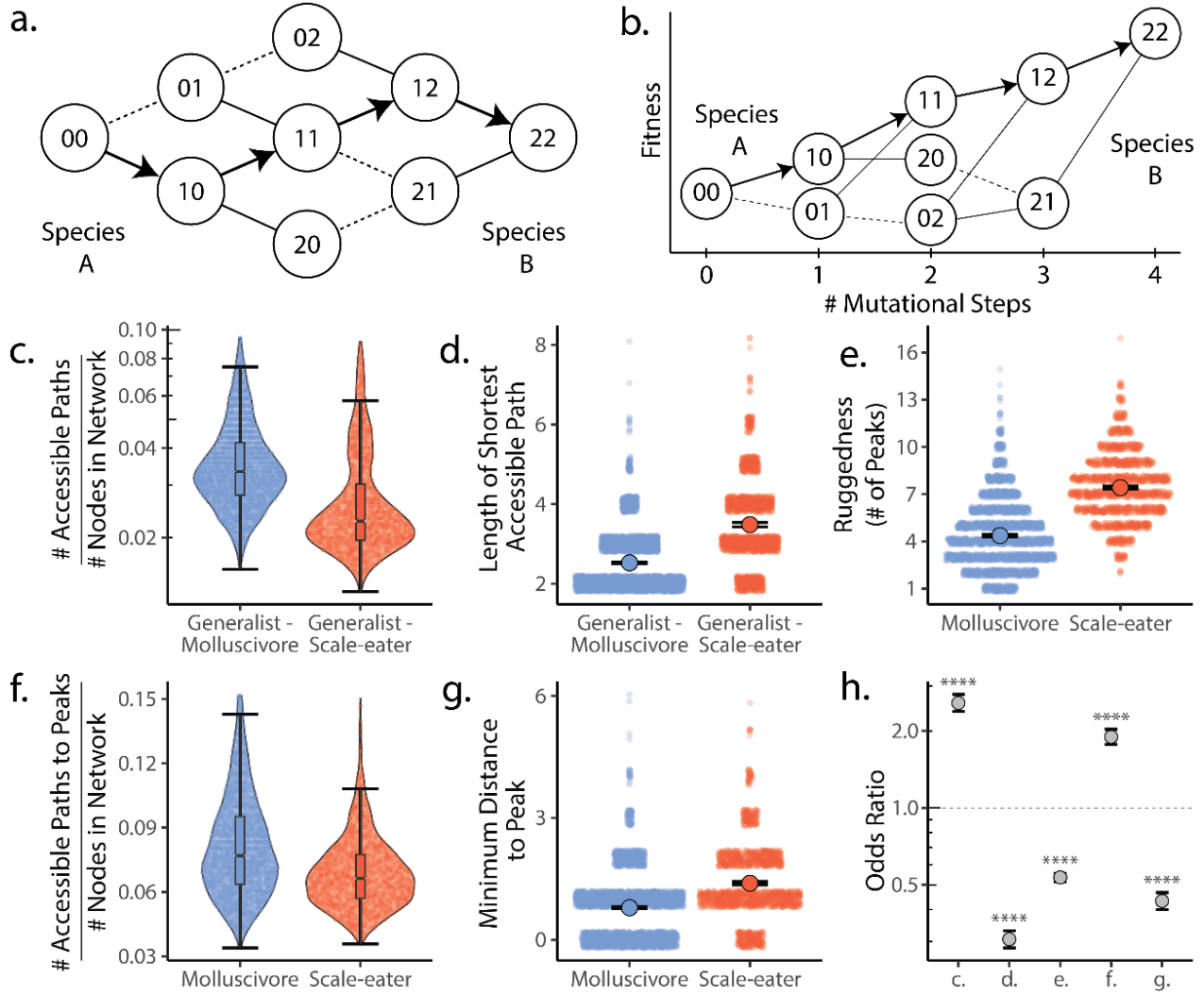


Figure 4. Molluscivore genotypes were more accessible to generalists on the genotypic fitness landscape than scale-eater genotypes. **a.** Diagram illustrating genotypic fitness networks and adaptive walks between species for a hypothetical two-SNP genotypic fitness landscape. Species A and B are separated by four mutational steps. Dashed lines indicate inaccessible paths that decrease in fitness leaving a single possible accessible evolutionary trajectory between species A and B (indicated by bold arrows). Each node in our study is associated with an empirical measure of hybrid fitness from field experiments (17, 18). Edges are always drawn as directed from low to high fitness. **b.** The same network as in (a.), with fitness plotted on the Y-axis and number of mutational steps from species A to B on the X-axis. The only accessible path between species A and B is indicated by solid arrows. **c.** Number of accessible paths between generalists and either specialist, scaled by network size. **d.** Length (# of nodes) of the shortest accessible paths. Means (large points) $\pm$ 2 standard errors are plotted. **e.** Ruggedness, as measured by the number of peaks (genotypes with no fitter neighbors within a single mutational step (32)). **f.** Number of accessible paths to peaks, scaled by network size. **g.** Length of the shortest accessible path to the nearest peak. **h.** Odds ratios (OR: Maximum Likelihood estimate and 95% CI) for each measure of accessibility (x-axis corresponds to panel letters); molluscivore networks have significantly greater summary statistics when OR > 1. Molluscivore genotypes are more accessible to generalists than scale-eater genotypes due to a significantly greater number of accessible paths separating them (c.) that are significantly shorter (d.). Molluscivore genotypic networks were also less rugged, i.e. they contained significantly fewer peaks (e.), each of which were in turn more accessible from the generalist genotypes (f., g.).

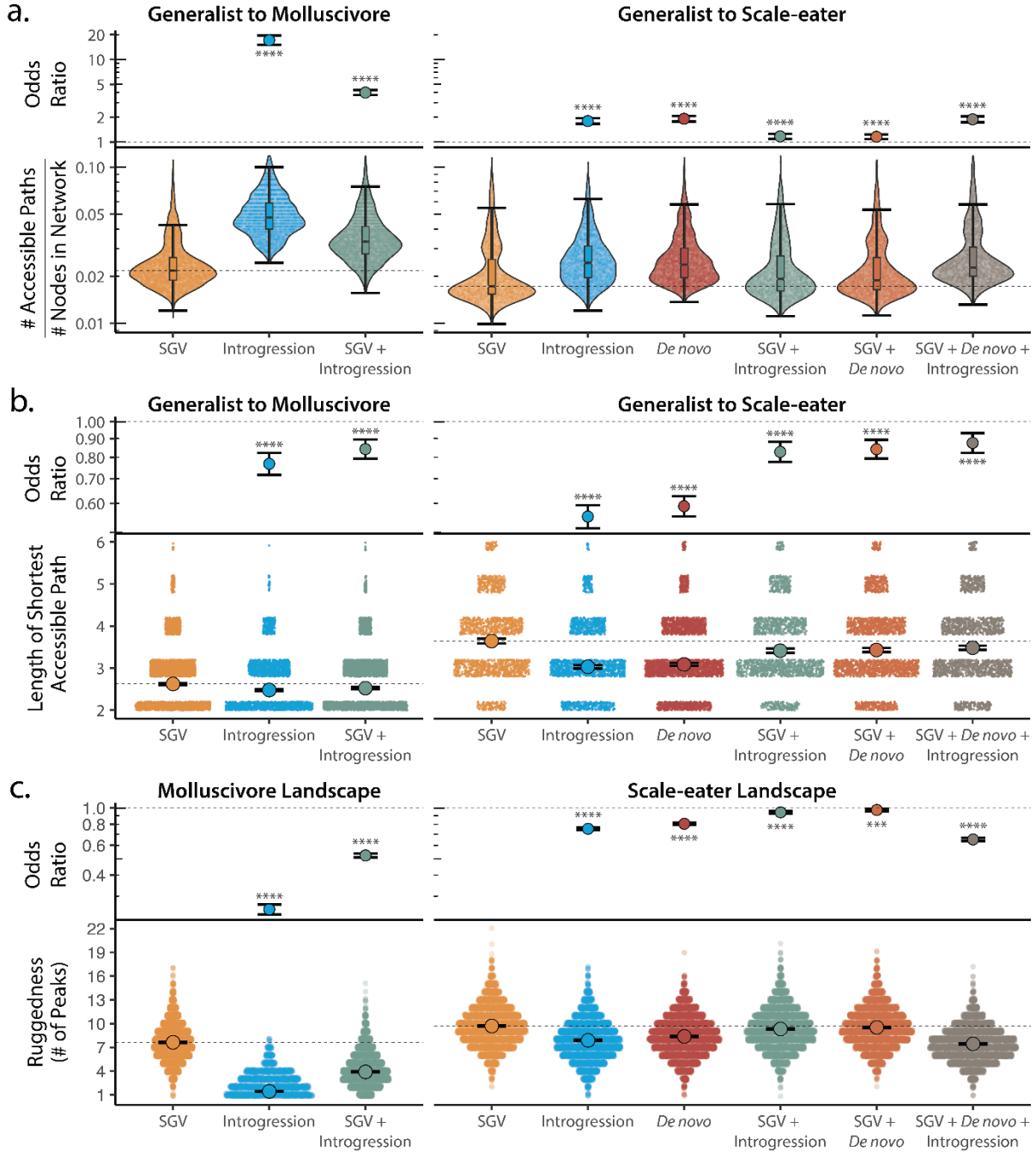


Figure 5. Adaptive introgression and *de novo* mutations increase access to specialist fitness peaks. Odds ratios (maximum likelihood estimate and 95% CI) indicate the effect of each source of variation on accessibility compared to networks estimated from standing variation alone. Asterisks denote significance ($p < 0.0001 = ****$, $< 0.001 = ***$). **a.** The number of accessible (i.e. monotonically increasing in fitness) paths per network, scaled by the size of the network (# of nodes in network). Significance was assessed using a likelihood ratio test, corrected for the false discovery rate (reported in Table S18). Dashed lines correspond to the median estimate for standing genetic variation to aid comparison to other sources of adaptive variation. **b.** Number of mutational steps in the shortest accessible path. Means are plotted as large circles, with two standard errors shown; dashed horizontal lines correspond to the mean for standing genetic variation. **c.** Ruggedness of molluscivore and scale-eater genotypic fitness networks constructed from each source of genetic variation measured by the number of peaks (genotypes with no fitter neighbors).

Literature Cited

1. S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Sixth Int. Congr. Genet.*, 356–366 (1932).
2. S. Gavrillets, Evolution and speciation on holey adaptive landscapes. *Trends Ecol. Evol.* **12**, 307–312 (1997).
3. S. Gavrillets, “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.* **4**, 530–544 (2020).
19. S. Gavrillets, *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. 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).
23. S. Wright, Evolution in Mendelian Populations. *Genetics* **16**, 97–159 (1931).

- 850 24. T. Aita, M. Iwakura, Y. Husimi, A cross-section of the fitness landscape of dihydrofolate
851 reductase. *Protein Eng. Des. Sel.* **14**, 633–638 (2001).
- 852 25. M. C. Whitlock, P. C. Phillips, F. B. G. Moore, S. J. Tonsor, Multiple fitness peaks and
853 epistasis. <https://doi.org/10.1146/annurev.es.26.110195.003125> **26**, 601–629 (1995).
- 854 26. F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, S. J. Tans, Empirical fitness landscapes
855 reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
- 856 27. F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, S. J. Tans, Reciprocal sign epistasis is a
857 necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144
858 (2011).
- 859 28. F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, S. J. Tans, Reciprocal sign epistasis is a
860 necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144
861 (2011).
- 862 29. J. Neidhart, I. G. Szendro, J. Krug, Adaptation in tunably rugged fitness landscapes: The
863 Rough Mount Fuji model. *Genetics* **198**, 699–721 (2014).
- 864 30. J. Neidhart, I. G. Szendro, J. Krug, Adaptation in tunably rugged fitness landscapes: The
865 Rough Mount Fuji model. *Genetics* **198**, 699–721 (2014).
- 866 31. D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, Darwinian evolution can
867 follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- 868 32. L. Ferretti, D. Weinreich, F. Tajima, G. Achaz, Evolutionary constraints in fitness
869 landscapes. *Heredity (Edinb)*. **121**, 466–481 (2018).
- 870 33. J. Franke, A. Klözer, J. A. G. M. de Visser, J. Krug, Evolutionary accessibility of
871 mutational pathways. *PLoS Comput. Biol.* **7**, 1002134 (2011).
- 872 34. J. A. G. M. De Visser, J. Krug, Empirical fitness landscapes and the predictability of
873 evolution. *Nat. Rev. Genet.* **2014** **15**, 480–490 (2014).
- 874 35. J. Maynard Smith, Natural selection and the concept of a protein space. *Nat.* **1970**
875 **225**, 563–564 (1970).
- 876 36. C. Brandon Ogbunugafor, A reflection on 50 years of John Maynard Smith’s “Protein
877 Space.” *Genetics* **214**, 749–754 (2020).
- 878 37. R. A. Fisher, *The genetical theory of natural selection* (Oxford University Press, 1930).
- 879 38. H. A. Orr, The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **6**, 119–127
880 (2005).
- 881 39. M. Karageorgi, *et al.*, Genome editing retraces the evolution of toxin resistance in the
882 monarch butterfly. *Nature* **574**, 409–412 (2019).
- 883 40. M. A. Siddiq, D. W. Loehlin, K. L. Montooth, J. W. Thornton, Experimental test and
884 refutation of a classic case of molecular adaptation in *Drosophila melanogaster*. *Nat. Ecol.*
885 *Evol.* **1**, 25 (2017).
- 886 41. M. A. Siddiq, J. W. Thornton, Fitness effects but no temperature-mediated balancing
887 selection at the polymorphic *Adh* gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.*
888 *U. S. A.* **116**, 21634–21640 (2019).
- 889 42. K. S. Singh, *et al.*, The genetic architecture of a host shift: An adaptive walk protected an
890 aphid and its endosymbiont from plant chemical defenses. *Sci. Adv.* **6**, eaba1070 (2020).
- 891 43. F. Peng, *et al.*, Effects of beneficial mutations in *pykF* gene vary over time and across
892 replicate populations in a long-term experiment with bacteria. *Mol. Biol. Evol.* **35**, 202–
893 210 (2018).
- 894 44. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution of a
895 key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci.*

- U. S. A. **105**, 7899–7906 (2008).
45. 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).
46. P. Nosil, *et al.*, Ecology shapes epistasis in a genotype–phenotype–fitness map for stick insect colour. *Nat. Ecol. Evol.* **4**, 1673–1684 (2020).
47. 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).
48. 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).
49. 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).
50. 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).
51. 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).
52. 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).
53. 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).
54. 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).
55. 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).
56. D. B. Weissman, M. W. Feldman, D. S. Fisher, The rate of fitness-valley crossing in sexual populations. *Genetics* **186**, 1389–1410 (2010).
57. D. B. Weissman, M. M. Desai, D. S. Fisher, M. W. Feldman, The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.* **75**, 286–300 (2009).
58. Y. Iwasa, F. Michor, M. A. Nowak, Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579 (2004).
59. A. F. Bitbol, D. J. Schwab, Quantifying the role of population subdivision in evolution on rugged fitness landscapes. *PLoS Comput. Biol.* **10**, e1003778 (2014).
60. A. Wagner, “High-dimensional adaptive landscapes facilitate evolutionary innovation” in *The Adaptive Landscape in Evolutionary Biology*, E. I. Svensson, R. Calsbeek, Eds. (Oxford University Press, 2012), pp. 271–282.
61. M. Conrad, The geometry of evolution. *Biosystems* **24**, 61–81 (1990).
62. A. C. Whibley, *et al.*, Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**, 963–966 (2006).
63. G. Martin, S. F. Elena, T. Lenormand, Distributions of epistasis in microbes fit predictions

- from a fitness landscape model. *Nat. Genet.* 2007 394 **39**, 555–560 (2007).
64. P. A. Gros, H. Le Nagard, O. Tenaillon, The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* **182**, 277–293 (2009).
65. S. Hwang, S. C. Park, J. Krug, Genotypic complexity of Fisher’s Geometric Model. *Genetics* **206**, 1049–1079 (2017).
66. S. C. Park, S. Hwang, J. Krug, Distribution of the number of fitness maxima in Fisher’s geometric model. *J. Phys. A Math. Theor.* **53**, 385601 (2020).
67. R. Froese, D. Pauly, FishBase. *FishBase* (2021).
68. O. Seehausen, others, Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).
69. S. H. Martin, C. D. Jiggins, Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* **47**, 69–74 (2017).
70. D. A. Marques, J. I. Meier, O. Seehausen, A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
71. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
72. K. A. Thompson, M. M. Osmond, D. Schluter, Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* **3**, 129–141 (2019).
73. J. Mallet, Hybrid speciation. *Nature* **446**, 279–83 (2007).
74. O. Seehausen, Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
75. C. Pardo-Diaz, *et al.*, Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**, e1002752 (2012).
76. 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).
77. 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).
78. J. I. Meier, *et al.*, Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
79. M. D. McGee, *et al.*, The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**, 75–79 (2020).
80. 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).
81. 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).
82. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
83. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 1–6 (2011).
84. M. E. Arnegard, *et al.*, Genetics of ecological divergence during speciation. *Nature* **511**, 307–311 (2014).
85. S. Wang, J. J. Hard, F. Utter, Genetic variation and fitness in salmonids. *Conserv. Genet.*

- 988 3, 321–333 (2002).
- 989 86. R. Leimu, P. Mutikainen, J. Koricheva, M. Fischer, How general are positive relationships
990 between plant population size, fitness and genetic variation? *J. Ecol.* **94**, 942–952 (2006).
- 991 87. D. Schluter, *et al.*, Fitness maps to a large-effect locus in introduced stickleback
992 populations. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1914889118 (2021).
- 993 88. J. A. McGirr, C. H. Martin, Ecological divergence in sympatry causes gene misexpression
994 in hybrids. *Mol. Ecol.* **29**, 2707–2721 (2020).
- 995 89. Y. Nakamura, *et al.*, Wwp2 is essential for palatogenesis mediated by the interaction
996 between Sox9 and mediator subunit 25. *Nat. Commun.* **2**, 1–10 (2011).
- 997 90. L. Mork, G. Crump, “Zebrafish craniofacial development: A Window into early
998 patterning” in *Current Topics in Developmental Biology*, Y. Chai, Ed. (Academic Press
999 Inc., 2015), pp. 235–269.
- 1000 91. T. Manousaki, *et al.*, Parsing parallel evolution: ecological divergence and differential
1001 gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from
1002 Nicaragua. *Mol. Ecol.* **22**, 650–669 (2013).
- 1003 92. S. J. Lim, *et al.*, Taurine is an essential nutrient for juvenile parrot fish *Oplegnathus*
1004 *fasciatus*. *Aquaculture* **414–415**, 274–279 (2013).
- 1005 93. T. G. Gaylord, A. M. Teague, F. T. Barrows, Taurine supplementation of all-plant protein
1006 diets for rainbow trout (*Oncorhynchus mykiss*). *J. World Aquac. Soc.* **37**, 509–517 (2006).
- 1007 94. M. Yokoyama, T. Takeuchi, G. S. Park, J. Nakazoe, Hepatic cysteinesulphinate
1008 decarboxylase activity in fish. *Aquac. Res.* **32**, 216–220 (2001).
- 1009 95. Z. Zhou, *et al.*, Genome-wide association study of growth and body-shape related traits in
1010 large yellow croaker (*Larimichthys crocea*) using ddRAD sequencing. *Mar. Biotechnol.*
1011 **21**, 655–670 (2019).
- 1012 96. G. M. Yoshida, J. M. Yáñez, Multi-trait GWAS using imputed high-density genotypes
1013 from whole-genome sequencing identifies genes associated with body traits in Nile tilapia.
1014 *BMC Genomics* **22**, 1–13 (2021).
- 1015 97. J. Kulmuni, A. M. Westram, Intrinsic incompatibilities evolving as a by-product of
1016 divergent ecological selection: Considering them in empirical studies on divergence with
1017 gene flow. *Mol. Ecol.* **26**, 3093–3103 (2017).
- 1018 98. L. F. S. Fonseca, *et al.*, Gene expression profiling and identification of hub genes in
1019 Nellore cattle with different marbling score levels. *Genomics* **112**, 873–879 (2020).
- 1020 99. P. Pavlidis, D. Živković, A. Stamatakis, N. Alachiotis, SweeD: Likelihood-based
1021 detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234
1022 (2013).
- 1023 100. N. Alachiotis, A. Stamatakis, P. Pavlidis, OmegaPlus: A scalable tool for rapid detection
1024 of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
- 1025 101. J. A. G. M. De Visser, J. Krug, Empirical fitness landscapes and the predictability of
1026 evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).
- 1027 102. M. C. Whitlock, P. C. Phillips, F. B. G. Moore, S. J. Tonsor, Multiple fitness peaks and
1028 epistasis. <https://doi.org/10.1146/annurev.es.26.110195.003125> **26**, 601–629 (1995).
- 1029 103. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol.*
1030 *Evol.* **23**, 38–44 (2008).
- 1031 104. J. Hermisson, P. S. Pennings, Soft sweeps and beyond: understanding the patterns and
1032 probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**, 700–
1033 716 (2017).

- 1034 105. P. W. Hedrick, Adaptive introgression in animals: examples and comparison to new
1035 mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–
1036 4618 (2013).
- 1037 106. J. Kaplan, The end of the adaptive landscape metaphor? *Biol. Philos.* 2008 235 **23**, 625–
1038 638 (2008).
- 1039 107. V. O. Pokusaeva, *et al.*, An experimental assay of the interactions of amino acids from
1040 orthologous sequences shaping a complex fitness landscape. *PLOS Genet.* **15**, e1008079
1041 (2019).
- 1042 108. L. I. Gong, M. A. Suchard, J. D. Bloom, Stability-mediated epistasis constrains the
1043 evolution of an influenza protein. *Elife* **2013** (2013).
- 1044 109. L. I. Gong, J. D. Bloom, Epistatically interacting substitutions are enriched during
1045 adaptive protein evolution. *PLOS Genet.* **10**, e1004328 (2014).
- 1046 110. F. Blanquart, T. Bataillon, Epistasis and the structure of fitness landscapes: Are
1047 experimental fitness landscapes compatible with fisher’s geometric model? *Genetics* **203**,
1048 847–862 (2016).
- 1049 111. I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, J. A. G. M. De Visser, Quantitative
1050 analyses of empirical fitness landscapes. *J. Stat. Mech. Theory Exp.*, P01005 (2013).
- 1051 112. T. C. Nelson, *et al.*, Ancient and recent introgression shape the evolutionary history of
1052 pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section
1053 *Erythranthe*). *PLoS Genet.* **17**, e1009095 (2021).
- 1054 113. H. Svardal, *et al.*, Ancestral hybridization facilitated species diversification in the lake
1055 malawi cichlid fish adaptive radiation. *Mol. Biol. Evol.* **37**, 1100–1113 (2020).
- 1056 114. S. Lamichhaney, *et al.*, Rapid hybrid speciation in Darwin’s Finches. *Science* **359**, 224–
1057 228 (2018).
- 1058 115. P. R. Grant, B. R. Grant, Hybridization increases population variation during adaptive
1059 radiation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23216–23224 (2019).
- 1060 116. N. B. Edelman, J. Mallet, Prevalence and adaptive impact of introgression. *Annu. Rev.*
1061 *Genet.* **55**, 265–283 (2021).
- 1062 117. P. R. Grant, B. R. Grant, Hybridization increases population variation during adaptive
1063 radiation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23216–23224 (2019).
- 1064 118. J. I. Meier, *et al.*, The coincidence of ecological opportunity with hybridization explains
1065 rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* **10** (2019).
- 1066 119. J. W. Poelstra, E. J. Richards, C. H. Martin, Speciation in sympatry with ongoing
1067 secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids.
1068 *Mol. Ecol.* **27**, 4270–4288 (2018).
- 1069 120. M. Moest, *et al.*, Selective sweeps on novel and introgressed variation shape mimicry loci
1070 in a butterfly adaptive radiation. *PLoS Biol.* **18**, e3000597 (2020).
- 1071 121. R. C. Lewontin, L. C. Birch, Hybridization as a Source of Variation for Adaptation to
1072 New Environments. *Evolution* **20**, 315–336 (1966).
- 1073 122. L. H. Rieseberg, M. A. Archer, R. K. Wayne, Transgressive segregation, adaptation and
1074 speciation. *Heredity (Edinb.)* **83**, 363 (1999).
- 1075 123. K. Kagawa, G. Takimoto, Hybridization can promote adaptive radiation by means of
1076 transgressive segregation. *Ecol. Lett.* **21**, 264–274 (2018).
- 1077 124. R. Abbott, *et al.*, Hybridization and speciation. *J. Evol. Biol.* **26**, 229–46 (2013).
- 1078 125. N. B. Edelman, J. Mallet, Prevalence and adaptive impact of introgression. *Annu. Rev.*
1079 *Genet.* **55**, 265–283 (2021).

126. Broad Institute, Picard Toolkit (2018).
127. M. A. Depristo, *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
128. R. Poplin, *et al.*, Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178 (2017).
129. 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).
130. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
131. 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).
132. 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).
133. J. Hereford, A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**, 579–588 (2009).
134. A. Henningsen, *censReg*: Censored regression (Tobit) models. (2020).
135. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
136. P. Cingolani, *et al.*, Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3** (2012).
137. 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).
138. 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).
139. T. L. Hedrick, Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspiration and Biomimetics* **3**, 034001 (2008).
140. 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).
141. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference* (Springer New York, 2002).
142. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1–9 (2006).
143. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, 289–300 (1995).
144. R. C. Team, R: A language and environment for statistical computing. (2019).
145. F. de Mendiburu, *agricolae*: Statistical procedures for agricultural research (2020).
146. T. Guillerme, *disPRity*: A modular R package for measuring disparity. *Methods Ecol. Evol.* **9**, 1755–1763 (2018).
147. 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).
148. 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*

- 1126 bidyanus). *Aquaculture* **253**, 592–601 (2006).
1127 149. J. M. Costa, *et al.*, Inadequate dietary phosphorus levels cause skeletal anomalies and alter
1128 osteocalcin gene expression in Zebrafish. *Int. J. Mol. Sci.* **19**, 364 (2018).
1129 150. H. Yu, *et al.*, Effects of stocking density and dietary phosphorus levels on the growth
1130 performance, antioxidant capacity, and nitrogen and phosphorus emissions of juvenile
1131 blunt snout bream (*Megalobrama amblycephala*). *Aquac. Nutr.* **27**, 581–591 (2021).
1132

