A novel framework for assessing the robustness of genomic animal models applied to a wild marine fish

Joshua A. Thia[1,2,A], J. David Aguirre[2], James Hereward[3], Jennifer Evans[4], Libby Liggins[2], Cynthia Riginos[3,5,B], Katrina McGuigan[3,B]

[1] Bio21 Institute, School of BioSciences, The University of Melbourne, Melbourne, Australia

[2] School of Biological Sciences, University of Auckland, Auckland, New Zealand

[3] School of the Environment, The University of Queensland, Brisbane, Queensland, Australia

[4] Division of Ecology and Evolution, Research School of Biology, Australian National University, Acton, Canberra, Australia

[A] Corresponding author: josh.thia@live.com, joshua.thia@unimelb.edu.au

[B] Shared senior authors

ORCID NUMBERS

Joshua A Thia	https://orcid.org/0000-0001-9084-0959
J David Aguirre	https://orcid.org/0000-0001-7520-441X
James Hereward	https://orcid.org/0000-0002-6468-6342
Jennifer Evans	https://orcid.org/0009-0001-6682-4710
Libby Liggins	https://orcid.org/0000-0003-1143-2346
Cynthia Riginos	https://orcid.org/0000-0002-5485-4197
Katrina McGuigan	https://orcid.org/0000-0002-0525-2202

RUNNING TITLE

Genomic animal models for wild organisms

KEYWORDS

Geometric morphometrics, genomic animal model, heritability, population genomics, microhabitats, rocky intertidal shores, quantitative genomics

ABSTRACT

1. Heritable genetic versus non-heritable plastic phenotypic variation determine how selection may shape evolutionary responses. Population genomic studies typically aim to identify associations between loci (SNPs) and phenotypes, missing broader insight into heritable basis of polygenic traits with quantitative genomic approaches.
2. Genomic animal models can facilitate estimates of trait heritability from genomic data but can be challenging to apply to wild organisms. This is because the populations of many wild organisms are large, comprising many distantly related individuals, which makes the sampling close-kin unlikely. This in turn limits the variation of relatedness in a sample, reducing the efficacy of genomic animal models.
3. We developed a novel framework for applying genomic animal models to wild organisms where sampling close-kin may be near impossible. Our framework combines resampling of genetic markers with simulating different levels of heritability to infer the power and reliability of genomic animal models fitted to datasets lacking kin structure.
4. We use our framework to estimate heritability of head shape traits in a wild marine intertidal fish, *Bathygobius cocosensis* (Bleeker 1854), using ddRAD SNP genotypes and phenotypes measured using geometric morphometric methods. In its eastern Australian range, *B. cocosensis* is highly abundant, has high effective gene flow, and occupies diverse habitats and microhabitats.
5. Placing our observed genomic animal model results in the context of simulated results, we conclude that two of the head shape traits were heritable (estimated h^2 = 0.26 and 0.18). This suggests that head shape phenotypes could evolve through heritable genetic changes in *B. cocosensis*, although there is likely a large non-heritable component. We also found strong phenotypic differentiation (pairwise $P_{ST} \leq 0.33$) despite very weak genetic differentiation (pairwise $F_{ST} < 0.002$), suggesting that environmental factors might structure phenotypes amidst processes homogenising genetic variation.
6. Our study addresses key methodological gaps for using quantitative genomic approaches on wild organisms, offering a rare example of partitioned heritable and non-heritable contributions to phenotypic variation in a non-model wild marine fish.

This would have been difficult without our new framework to overcome the challenges of a dataset lacking kin structure.

INTRODUCTION

Phenotypic diversity within and between populations arises from the combination of genetic and plastic variation in measured traits. Partitioning the relative importance of (heritable) genetic versus (non-heritable) plastic components to phenotypes is a fundamental goal in evolutionary genetics because each can have different evolutionary outcomes (Hoffmann et al., 2005; McCairns & Bernatchez, 2012; Ørsted et al., 2018; Via & Lande, 1985). For adaptive traits, selection acting on heritable phenotypic variation results in evolutionary change, whereas selection acting on non-heritable phenotypic variation does not (Gienapp & Merilä, 2014; Hoekstra et al., 2004; McCairns & Bernatchez, 2012). In practice, quantifying these different components can be challenging because the factors influencing phenotypic variance and adaptive processes are often poorly understood and hard to characterise (Gienapp & Merilä, 2014; Merilä & Hendry, 2014). If we can understand their contributions to phenotypic variance, we may gain greater capacity to predict how organisms will respond to human-driven environmental change (Chevin et al., 2010; Hoffmann & Bridle, 2022; Merilä & Hendry, 2014).

Population genomic approaches have greatly advanced our ability to study the genetic basis of adaptive traits, particularly through genome-wide association tests that aim to identify correlations between genotypes and phenotypes, or genotypes and environments (Lasky et al., 2022; Santure & Garant, 2018). However, genome-wide associations can be limited by the availability of adequate reference genomes (Lowry, Hoban, Kelley, Lotterhos, Reed, Antolin, et al., 2017; Meirmans, 2015) or may be underpowered when traits are polygenic (Duntsch et al., 2020; Lasky et al., 2022; Santure & Garant, 2018). Attractive alternatives use genomic animal models, which combine classic quantitative genetic analysis with genomic estimates of inter-individual genomic relationships (Bérénos et al., 2014; Kruuk, 2004; Richards et al., 2023; Schmidt et al., 2023; Wilson et al., 2010). A key parameter for fitting animal models is the matrix, **A**, which describes the expected proportion of genes that are identical by descent variances between individuals. When armed with genomic data, it is possible to reconstruct **A** in the absence of a pedigree (Gervais et al., 2019; James et al., 2022; Porth et al., 2015; Schmitt et al., 2021) by directly estimating the proportion of genes that are identical by state between individuals (de los Campos et al., 2015; VanRaden, 2008). Nonetheless, genomic animal models can be underpowered when sample sizes are small and (or) there is little

variability in the genetic relationships of sampled individuals (Fraimout et al., 2024; Richards et al., 2023).

Another consideration is the choice of genetic markers used to construct **A**. In principle, **A** can be constructed from any number of genetic markers, which provides scope to conduct quantitative genomic analyses in non-model organisms without reference genomes using reduced representation methods (e.g., RADseq). The choice of genetic markers used to construct **A** can, however, influence estimated effect sizes. Enough genetic markers are needed to be in linkage disequilibrium with the focal trait and capture additive genetic relationships, but too many markers can make point estimates of heritability less precise and less reliable, for example, due to multicollinearity or through spurious covariances with phenotypes (de los Campos et al., 2015). Different combinations of loci with different allele frequencies may produce different estimates of additive genetic variances, which will also influence the point estimate of heritability (Stanton-Geddes et al., 2013; Wang et al., 2017). We therefore need to be circumspect when interpreting the results of genomic animal models in samples of wild-caught organisms. How much power is there to estimate a given level of heritability, and how reliable are these estimates, given a specific experimental design (the number and relationships of sampled individuals, as well as the linkage of genetic markers and their allele frequencies)?

Marine organisms with pelagic (planktonic) dispersal provide a context where quantitative genomic approaches can be both highly useful and challenging to implement (Gagnaire & Gaggiotti, 2016). In the classically 'panmictic marine metapopulation' (Waples, 1998), pelagic dispersal facilitating long-distance migration may favour the evolution of plastic responses to local conditions. Dispersal by ocean currents might cause individuals to settle in locations where they are locally maladapted, in which case, plasticity may help buffer against genotype−environment mismatches (Garant et al., 2007; Parsons, 1997; Scheiner, 2013). Nonetheless, the large effective population sizes of marine populations, coupled with strong and spatiotemporally heterogeneous selection, could also drive adaptation despite processes that lead to genetic homogeneity across populations (Babin et al., 2017; Rey et al., 2020a; Tepolt et al., 2021). The expected polygenicity of many traits in marine organisms means that quantitative genomic approaches could provide key insights into the availability of adaptive genetic variance (Babin et al., 2017; Fuller et al., 2020; Gagnaire & Gaggiotti, 2016; Rey et al.,

2020b), helping us understand the opportunity for selection to drive adaptive evolution. However, unless biotic and (or) abiotic factors reduce the spatial scale of larval dispersal and (or) reduce the genetic neighbourhood size (Christie et al., 2010; Hedgecock et al., 2007; Selwyn et al., 2016), given the panmictic nature of many wild marine populations, individuals will tend to be uniformly distantly related making the use of genomic animal models challenging (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016).

There are some examples of animal models being applied to wild marine organisms (DiBattista et al., 2009; Malvezzi et al., 2015; Salles et al., 2020), but these studies benefitted from the availability of close-kin in their datasets. Work in a marine population of ninespine stickleback (*Pungitius pungitius*) has clearly demonstrated the pitfalls of using genomic animal models on wild-caught individuals that are uniformly distantly related (Fraimout et al., 2024). Indeed, it may be very hard to estimate even moderate heritability from large samples in the absence of some close-kin and variance in relatedness among sampled individuals (Gagnaire & Gaggiotti, 2016). A knowledge gap exists in understanding how to implement, and then assess the effectiveness, of genomic animal models in systems where capturing variability in genetic relationships is nearly impossible. Whilst this challenge is relevant to any abundant species that is dispersive and has large population sizes, it is pervasive in marine taxa (Fraimout et al., 2024).

Marine organisms occupying rocky intertidal habitats have provided fundamental insights into how spatiotemporal environmental heterogeneity shapes the distribution of phenotypes in wild populations (Johnson & Black, 2008; Schmidt & Rand, 1999; Trussell, 1996). Across space, environmental variation can be hierarchically organised at large macrogeographic scales (kilometers to hundreds of kilometers) to very small microgeographic scales (meters to tens of meters). For example, the intensity of wave exposure and predation by crabs on *Littorina* snails exhibits both among- and within-patch variability, which leads to predictable distributions of shell phenotypes across macro- and microgeographic scales (Koch et al., 2021; Trussell, 1996; Westram et al., 2014). Because environmental differences can be so striking at small microgeographic scales, they are often analysed in the context of discretely categorized microhabitats (Helmuth & Hofmann, 2001; Schmidt & Rand, 1999; Sokolova et al., 2000). However, analysis of continuous microhabitat variation can provide a more nuanced understanding of how intertidal organisms respond to microgeographic variation in

their environment (Griffiths et al., 2003; Johnson & Black, 2008). Across time, environments can change on scales ranging from over an individual's lifetime to over many generations. For instance, gene expression analyses have demonstrated that *Mytilus californianus* mussels undergo several physiological state shifts in association with tidal fluctuations (Gracey et al., 2008), whereas at the other extreme, shell phenotypes in populations of *Littorina obtusata* snails shifted over a decade in response to a range expansion of a crab predator (Seeley, 1986; Trussell, 2000).

In this present study, we developed a new method to bring quantitative genomics into the wild and tested our approach on a marine intertidal fish, *Bathygobius cocosensis* (Bleeker, 1854: Perciformes; Gobiidae). *Bathygobius cocosensis* has a wide distribution across the Indo-Pacific. In Australia, *B. cocosensis* is a major resident species of rocky intertidal zones throughout tropical and temperate-subtropical waters (Griffiths, 2003b). Prior work in Australian populations of *B. cocosensis* has documented plasticity in different traits associated with life in the intertidal zone, including thermal physiology (da Silva et al., 2019), feeding preferences (Carbia et al., 2020), cognition (Carbia & Brown, 2019), and colour change (da Silva et al., 2020). Head shape has also emerged as a putative ecologically important trait that not only varies among populations at large macrogeographic scales (hundreds of kilometers) but also within populations across small microgeographic scales (tens of meters) (Malard et al., 2016; Thia et al., 2021). A pelagic larval duration of 20 to 25 days potentially facilitates long-distance dispersal in this fish (Thia et al., 2018), producing patterns of weak genetic structure across hundreds of kilometres along the east Australian coastline (Thia et al., 2021).

This study had three over-arching research questions. Our first question was methodological in nature: (1) How can we use genomic animal models to partition phenotypic variance in a non-model wild organism with no detectable kin structure? Our remaining two questions were biological in nature: (2) How are genetic and phenotypic variation structured over different spatiotemporal scales in *B. cocosensis*; and (3) What are the relative contributions of heritable versus non-heritable variation to head shape phenotypes in *B. cocosensis*? We addressed question 1 by developing a novel framework for applying genomic animal models to wild organism with demographic and life history characteristics that make them recalcitrant to quantitative genomic studies (Figure 1). This framework combined resampling

of genetic markers and simulations to place observed results in the context of expected results under different simulated levels of heritability. To address questions 2 and 3, we compiled a dataset of fish collected across multiple years (different cohorts), macrogeographic scales (two distant sites), microgeographic scales (continuous measures of tide pool microhabitats within a site), genotyped with reduced representation sequencing (ddRAD-seq), and measured for head shape traits using geometric morphometrics (Figure 2). We first addressed question 2 by contrasting genetic and phenotypic structure across spatiotemporally separated populations of *B. cocosensis*, and by testing for associations between head shape phenotypes and continuous microhabitat variables within populations. We then used our new framework developed in question 1 to address question 3, where we tested for heritability of head shape traits.

MATERIALS AND METHODS

Resampling-simulation framework for genomic animal models

Overview. In classic quantitative genetic theory, narrowsense heritability, h^2, is defined as:

$$h^2 = V_A / V_P = V_A / (V_A + V_E)$$

Here, V_A is the additive genetic variance, V_E is the environmental variance, and V_P is the phenotypic variance. In animal models, we can use the known additive genetic relationship matrix, **A**, to simultaneously solve for breeding values and V_A (Kruuk, 2004), given that:

$$g \sim \mathbf{A}V_A$$

The values of g are then used to represent the random intercepts for each individual in a mixed effects model of phenotypic variation:

$$y = \mathbf{X}b + g + e$$

Where **X** is a design matrix for fixed effects, b is a vector of fixed effect coefficients, and e is a vector of residual (environmental) deviations.

In traditional quantitative genetics studies, **A** would be obtained from a known pedigree of individuals. But it is also possible to estimate **A** using genomic data, such that **A** represents a

genomic relationship matrix. A genomic relationship matrix is a covariance matrix that describes the expected shared genetic variance between individuals (de los Campos et al., 2015; VanRaden, 2008). When estimating **A** from genomic data, we are expecting that at least some of the sampled genetic markers carry information about a trait's genetic variance, either directly as the quantitative trait (QT) loci controlling the trait, or indirectly as genetic markers linked to QT loci (de los Campos et al., 2015; VanRaden, 2008).

There are two challenges with estimating **A** with genomic data. The first challenge is related to genetic marker choice, which is a general concern when fitting genomic animal models. If the loci used to estimate **A** are poorly correlated with V_A, then estimates of V_A from genomic animal models will be biased. Typically, SNP sets should be thinned to reduce redundant information and reduce over-estimating genetic variance; conversely, too few markers might under-estimate the genetic variance (de los Campos et al., 2015). It might also be possible that some loci might exhibit spurious associations with phenotype, that could also lead to over-estimated genetic variance. The methods used to determine the final genetic marker set, such as sequencing methodology or filtering criteria, can have important implications on the estimation of **A**. It is therefore possible that different genetic marker sets can produce different estimates of **A**, leading to differences in the reliability to estimate, and power to detect, V_A and h^2 with genomic animal models.

The second challenge is a major concern in studies of wild organisms. Many wild organisms exist as large populations where most individuals are distantly related. In this context, it is very unlikely that a random sample of individuals from the population will contain any close-kin, and hence, there will be no variability in genomic relationships described by **A** (all individuals are uniformly related). This is problematic. Because **A** is effectively being correlated against phenotypic values to help partition out the putative genetic variance, if **A** is effectively invariable, then there is insufficient information to relate the phenotypic differences between individuals to the genetic variation between them. Thus, where **A** is comprised of uniformly related individuals, it is difficult to determine whether an estimated V_A of 0 (no heritability) is a reliable result, or instead is due to low power to detect heritable variation.

Here, we developed a novel framework to address these challenges, providing a means to assess the power and reliability of genomic animal models applied to wild organisms (Figure 1). This framework uses multiple resampled SNP sets to assess the variability of results in the *observed* dataset due to genetic marker choice. Next, *simulated* datasets are generated under different levels of heritability whilst maintaining the properties of the observed data (sampled individuals and genetic markers within SNP sets). These simulations then provide a context into which observed results can be placed to understand the expected estimates and their variability, given a simulated level of heritability and an experimental design.

Observed data. Our framework begins by defining a series of J randomly drawn SNP sets from the total pool of genetic markers that pass quality filtering criteria and are thinned to reduce redundancy. For each ith focal trait, and each jth SNP set, we estimate the corresponding **A** and use this to fit a genomic animal model on the observed phenotypes, with any covariates included as fixed effects. A corresponding null model is also fit without **A**. By iterating across J SNP sets, we obtain a range of estimates for genomic animal models fit to the same phenotype and fixed effects, but that vary in their **A**. We therefore obtain a range of estimated values for the observed narrowsense heritability, h^2_{ij}. We also obtain range of P-value for the animal vs null model comparison, P_{ij}. The spread of h^2_{ij} provides an inference of power and reliability of the heritability estimates based on genetic marker choice. The spread of P_{ij} provides an inference of power and the reliability of model comparison tests based on genetic marker choice. Central values are calculated for each ith trait as the medians of heritability estimates and model comparison tests across all J SNP sets, $\widetilde{h^2_i}$ and $\widetilde{P}_i$, respectively.

Simulated data. After drawing J SNP sets and fitting models on the observed data, simulations are then performed to obtain expectations under different simulated heritability levels. For each ith trait, and j^{th} SNP set, heritability at the hth level is simulated using the *family_sim_qtl* function from R's *genomalicious* package. This function performs a phenotype-from-genotype simulation using the theoretical relationship between V_{A}, the allele frequencies at a biallelic QT locus, p and q, and the locus-specific genetic effects for homozygote and heterozygote genotypes, a and d (Falconer, 1989):

$$V_{\text{A}} = 2pq[a + d(q - p)]^2$$

Detailed information on the underlying algorithm for the *family_sim_qtl* function can be found in the Supplementary Methods ('Simulating a quantitative trait using *family_sim_qtl*'). The key point is that V_A and V_E are simulated relative to the observed allele frequency distributions and genotypes in the observed sample of individuals. Simulated phenotypes will therefore be constrained to come from the same **A** as observed phenotypes but we do not need a pedigree to generate the genetic breeding values. For each ith trait, and j^{th} SNP set, a total of K nested QT loci sets are randomly drawn. Each kth QT loci set comprises M loci drawn from the total N loci in that SNP set. Each mth QT locus has an equal contribution to the trait.

Within the kth QT loci set, we obtain g, a vector of simulated genetic breeding values, and e, a vector of simulated environmental deviations, across individuals. The values in g are obtained by summing the locus-specific effects within individuals. The values in e are obtained by randomly drawing from a normal distribution, N(0, simulated V_E), using the *rnorm* function. Final simulated phenotypes per individual are obtained by summing their genetic breeding values and environmental deviations, and by adding any relevant fixed effects that were estimated from the observed data. Next, a genomic animal model and a null model are fitted to the simulated phenotypes, as per the observed data. However, **A** in these simulated genomic animal models is estimated in two different ways: *with* the QT loci and *without* the QT loci.

For each hth level of simulated h^2, we obtain a range of values for $h^2_{hijk(N)}$ and $P_{hijk(N)}$ as the simulated heritability estimate and model comparison P-value when using all N loci (with QT loci in **A**), and the respective values with $N - M$ loci, $h^2_{hijk(N-M)}$ and $P_{hijk(N-M)}$ (without QT loci in **A**). We calculate the central values within combinations of the ith trait and j^{th} SNP set by taking the median across the K QT loci sets: $\widetilde{h^2_{hij(N)}}$ and $\widetilde{P_{hij(N)}}$, and $\widetilde{h^2_{hij(N-M)}}$ and $\widetilde{P_{hij(N-M)}}$. The distribution of $\widetilde{h^2_{hij(N)}}$ and $\widetilde{P_{hij(N)}}$, and $\widetilde{h^2_{hij(N-M)}}$ and $\widetilde{P_{hij(N-M)}}$, provides an inference of how genetic marker choice might influence the power and reliability of heritability estimates and model comparison tests. Furthermore, comparing N vs N−M derived estimates provided an inference of how **A** calculated with vs without QT loci, respectively, might influence estimates. We then calculate central values for each ith trait by taking the medians across the

J SNP sets (median of medians): $\widetilde{h^2_{hi(N)}}$ and $\widetilde{P_{hi(N)}}$, and $\widetilde{h^2_{hi(N-M)}}$ and $\widetilde{P_{hi(N-M)}}$. Comparisons of simulated $\widetilde{h^2_{hi(N)}}$ and $\widetilde{P_{hi(N)}}$, and $\widetilde{h^2_{hi(N-M)}}$ and $\widetilde{P_{hi(N-M)}}$ to observed values provides an inference of power to detect a specific h^{th} level of heritability.

Sample collection and study sites

We sampled two environmentally distinct sites on Australia's east coast: Hastings Point (S28.36° E153.58°, New South Wales) and Heron Island (S23.44° E151.91°, Queensland), which respectively have temperate-subtropical and tropical climates. Hastings Point (Figure 2a) is a rocky headland with a topologically complex intertidal zone with discrete pools that *B. cocosensis* retreat to during low tide. Heron Island (Figure 2b) is situated ~80 km from the mainland at the southern limit of Great Barrier Reef coral islands, with shallow intertidal gradient that is absent of tide pools; *B. cocosensis* retreat to the subtidal areas during low tide.

Annual collections of *B. cocosensis* were made between 2016 and 2018. At Hastings Point, sampling occurred on the 25th of February 2016, 24th and 25th of February 2017, and 1st and 2nd of March 2018. At Heron Island, sampling occurred on 14th of January 2016 and 11th of January 2017. Fish were caught using hand nets and were euthanised with 500 mg/mL eugenol in seawater following ethically approved protocols (University of Sydney #2015/384; University of Queensland #SBS/221/15/HSF). Fish of all ages were included in this study, ranging from juveniles (~6 mm) to adults (~60 mm). In this study, we refer to a 'population' as fish collected from the same site (Hastings Point or Heron Island) in the same year (2016, 2017 and 2018). We use this spatiotemporal definition of populations for two reasons. First, these geographically distant sites are likely to experience different local conditions within each year. Second, the estimated lifespan of *B. cocosensis* is likely less than one year (Thia et al., 2021), suggesting that each yearly collection largely comprises a different cohort within each site. Where possible, we collected genotype, phenotype, and microhabitat data for each fish (see Figure 2c for sampling design).

Genotype data

Genome-wide SNP genotypes were obtained using a custom ddRAD procedure with the *Eco*RI and *Sph*I restriction enzymes (see Supplementary Methods, 'Custom ddRAD protocol').

The prepared ddRAD libraries were sequenced on an Illumina HighSeq X platform by AnnoRoad Gene Technology (Beijing, China). Reads were demultiplexed using the *process_radtags* function from *stacks2* v2.55 (Catchen et al., 2013; Rochette et al., 2019). Reads were quality trimmed using *fastp* v0.12.4 (Chen et al., 2018) and were truncated to a final length of 100 bp. There was a mean of 5,201,124 trimmed reads across samples, with a range of 10,928–21,942,052 reads.

RAD tags were assembled *de novo* using the *stacks2* functions *ustacks*, *cstacks*, *sstacks*, *tsv2bam*, and *gstacks*; the *populations* function was used to call SNPs. *Vcftools* v0.1.16 (Danecek et al., 2011) was used to perform an initial filtering to remove SNPs with minor allele counts (MAC) <3, genotype qualities <15, and missing data rates >50%. RAD tags were mapped against the *B. cocosensis* mitogenome (Evans et al., 2018) to identify sequences that were not derived from the nuclear genome: we used *seqtk* v1.3-r106 (Li, 2019) to reformat RAD tag sequences, *bwa mem* v0.7.17 (Li & Durbin, 2009) for mapping, and *samtools* v1.7 (Li et al., 2009) for filtering alignments.

SNPs were imported into R (R Core Team, 2022) for further filtering. We used an iterative procedure that considered the multifaceted nature of missing data: samples and loci can vary in their proportions of missing genotypes, and even loci within RAD tags may vary in missing data rates (Supplementary Figure S1). Please see the Supplementary Methods ('SNP filtering') for more details. SNPs derived from the mitogenome and samples with >50% missing data were removed first. We then imputed probable genotypes using the R package, *polyRAD* v2.0.0 (Clark et al., 2019), which allowed us retain SNP loci with missing genotype or low coverage. Next, we removed SNPs with missing data >40%, SNPs that were not observed in at least 5 individuals, and SNPs with evidence of heterozygosity excess based on Hardy-Weinberg equilibrium.

From the remaining SNPs, we then created 50 different SNP sets by randomly drawing one SNP locus per RAD tag. We used these different SNP sets as part of resampling simulation framework for heritability estimation (see Section 'Heritability of head shape phenotypes' below). To curate our final working SNP sets, we performed linkage analysis using the *LDna* v2.0 R package (Kemppainen et al., 2015). This allowed us to identify clusters of SNP loci in linkage disequilibrium (Supplementary Figure S2). We selected one SNP per linkage cluster

to retain independent loci in each SNP set. Genotypes were subjected to PCA (principal components analysis) using R's *prcomp* function to assess how removal of linked SNPs affected clustering of samples.

Our final working genotypic dataset comprised 542 samples: 97 and 242 fish from Hastings Point in 2016 and 2017 (respectively), and 100 and 103 fish from Heron Island in 2016 and 2017 (respectively). A total of 30,260 SNP loci across 4,404 RAD tags were distributed across 50 SNP sets (statistical properties illustrated in Supplementary Figure S3). A mean of 4,031 independent SNP loci (range of 3,877–4,131 SNP loci) were present in each SNP set.

Phenotype data

We used a set of six morphological landmarks to characterise head shape phenotypes (Figure 2d). Landmarks were placed on 2D images of fish heads using *tpsdig2* (Rohlf, 2015) and were imported into R for general Procrustes alignment using the *geomorph* package (Adams & Otárola-Castillo, 2013). It was not possible to place landmarks on very small fish because landmarks were hard to identify, and because small fish were prone to damage during preservation and handling. In total, 658 fish were phenotyped: 58, 189, and 316 from Hastings Point in 2016, 2017 and 2018 (respectively), and 46 and 49 from Heron Island in 2016 and 2017 (respectively). Imported landmarks were then subjected to PCA using R's *prcomp* function to derive three head shape traits, phenotypic PC1, PC2, and PC3, the explained 39.54%, 25.14%, and 15.11% of the total phenotypic variation, respectively.

Microhabitat data

We collected four microhabitat variables that described continuous variation in tide pool environments at Hastings Point: zone, rugosity, volume, and depth. Zone was measured as the time for tide pools to be inundated on the incoming tide (minutes). Pools in the low intertidal zone had shorter inundation times relative to pools in the high intertidal zone. Rugosity measured the structural complexity of tide pools. The longest axis for each focal tide pool was identified and two lengths were obtained using a tape measure (meters): the direct length was the distance end to end, whereas the conforming length tracked the topography of the substrate surface. The rugosity score was calculated as the ratio of the direct to conforming length (Griffiths et al., 2003; Willis et al., 2005). Volume of focal tide

pools was measured as the amount of water (litres) they contained at low tide. A bilge pump was used to empty tidepools; the water was stored and measured in 20 L buckets and returned to tidepools after measuring. Depth (cm) was measured as the distance from the deepest point of focal tide pools to the water surface at low tide. In total, we had measurements for 10 and 11 tide pools with at least four representative fish from the 2017 and 2018 Hastings Point collections, respectively; all tide pools sampled in 2017 were also sampled in 2018.

Macrogeographic structuring of genotypes and phenotypes

We calculated the statistics, F_{ST} and P_{ST}, for genetic and phenotypic differentiation, respectively, in R. We used the *fstat_calc* function from *genomalicious* (Thia, 2024) to calculate F_{ST} as per (Weir & Cockerham, 1984). We report F_{ST} for SNP set 1, but we found that estimates across SNP sets were highly congruent. We calculated P_{ST} using the method outlined in Chenoweth and Blows (2008), which was derived from Kremer et al. (1997) (see Supplementary Methods, 'P_{ST} calculations'). Random permutations of individuals among populations were used to test the significance of F_{ST} and P_{ST} estimates. The permuted *P*-value, P_{perm}, was calculated as the proportion of 100 permutations that produced a null statistic larger than the observed statistic.

In *B. cocosensis*, head shape changes with body size (age). Our collection comprised similar ranges of fish body size across populations, but distributions varied (Supplementary Figure S4). We fitted linear models to quantify the relationships between head shape and fish body size in each population. These models were used in our downstream heritability analyses (see Section 'Heritability of head shape phenotypes' below). Our models took the form:

$$y = u + \text{Population} + \text{Fish size} + \text{Population:Fish size} + e$$

$$y = u + \text{Population} + \text{Fish size} + e$$

Here, the response y is a vector of phenotypic PC scores, u is the mean, and e is the residual. 'Population' is the fixed categorical effect of population. 'Fish size' is the fixed effect of fish head size (log10 transformed centroid size, scaled to mean of 0 and variance of 1). 'Population:Fish size' is the interaction term for different effects of fish size per population. Models were fit using R's *lm* function. The full model containing the interaction term was

compared to the reduced model containing solely the main effects using R's *anova* function. The best model was retained for downstream analysis.

Microgeographic structuring of phenotypes

We tested the effects of tide pool microhabitat on the distribution of head shape phenotypes across the intertidal zone at Hastings Point in the 2017 and 2018 populations. There were 10 tide pools with ≥4 fish sampled in both populations: the mean number of fish per tide pool in 2017 was 16 fish with a range of 4 to 59 fish, and 27 fish with a range of 5 to 64 fish in 2018. We used these 10 resampled pools to fit separate linear mixed models using R's *glmmTMB* package (Brooks et al., 2017). We fit separate models for the Hastings Point 2017 and 2018 populations to avoid fitting many interaction terms and to examine the effects of microhabitat variables between years. Models for each annual population took the form:

$$y = u + \text{Fish size} + \text{Zone} + \text{Depth} + \text{Volume} + \text{Rugosity} + \text{Pool ID} + e$$

Here, y the vector of individual phenotypic PC scores, u is the mean, and e is the residual. 'Fish size' is the fixed effect of fish head size (log10 transformed centroid size, scaled to mean of 0 and variance of 1). 'Zone', 'Depth', 'Volume', and 'Rugosity' are also fixed effects for each tide pool's microhabitat (scaled to mean of 0 and variance of 1). 'Pool ID' models the random effect of tide pool. To assess collinearity among predictors, we estimated variance inflation factors (VIFs) for each model using R's *performance* package and the *check_collinearity* function (Lüdecke et al., 2021). For both the Hasting Point 2017 and 2018 models, all VIF point estimates were <5, suggesting that collinearity among predictors was not an issue (Supplementary Figure S5). We therefore retained all predictors in both models. We illustrate the associations between raw trait values and microhabitat variables in Supplementary Figure S6.

We also tested whether mean tide pool phenotypes were consistent between years. We calculated Spearman's rank correlation coefficients (r_s) of mean tide pool phenotypes between years, whilst controlling for differences in tide pool mean fish head size. If the effects of tide pool microhabitats had consistent influences on head shape phenotypes, we would expect tide pools to have similar ranks (positive r_s) between years. Partial correlation

coefficients were calculated using the *pcor.test* function from the *ppcor* R package (Kim, 2015).

Heritability of head shape phenotypes

Prior to fitting genomic animal models, we tested our capacity to identify pairs of close-kin in our dataset. Estimates of relatedness can deviate from expected values due to the Mendelian inheritance process, but also for technical reasons due to choice of genetic markers (Hill & Weir, 2011). For simplicity, we show results for three representative SNP sets. We performed these tests using the *family_sim_data* function from the R package, *genomalicious*, to generate 30 simulated pairs of unrelated individuals, half-cousins, cousins, half-siblings, siblings, and parent-offspring relationships using global allele frequencies for each locus per SNP set. We then calculated relatedness among these simulated individuals using the *A.mat* function from the *sommer* package (Covarrubias-Pazaran, 2016). The distribution of simulated relatedness estimates was compared to observed relatedness estimates with the *family_sim_compare* function from *genomalicious*. These results suggested that our sample was comprised mostly (if not exclusively) of distantly related individuals. Importantly, our relatedness simulations suggested that we had sufficient power to detect close-kin if they were present in our dataset, (Supplementary Figure S7).

We then applied our genomic animal model framework to infer the potential heritability of head shape phenotypes in wild-caught *B. cocosensis*. We first generated observed estimates by fitting genomic animal models to observed phenotypes across our 50 SNP sets. *Sommer*'s *A.matrix* function was used to construct **A** for each SNP set for all 542 genotyped fish. Of our genotyped fish, 298 were phenotyped: 40 and 167 from Hastings Point in 2016 and 2017 (respectively), and 46 and 45 from Heron Island in 2016 and 2017 (respectively). We fit genomic animal models on all 298 dually genotyped and phenotyped fish using *sommer*'s *mmer* function, with models taking one of the forms:

$$y = u + \text{Fish size} + \text{Population} + \text{Fish size:Population} + \text{Animal} + e$$

$$y = u + \text{Fish size} + \text{Population} + \text{Animal} + e$$

The first model was used for phenotypic PC1 and PC3 to capture variable relationships between head shape and body size observed in different populations (i.e., an interaction); the

second model was used for phenotypic PC2 where this relationship was consistent across populations (i.e., no interaction). These choices were based on our linear models of head shape across populations (see Results section 'Macrogeographic structuring of genotypes and phenotypes'). In these models, y is the vector of individual phenotypic PC scores, u is the mean, and e is the residual variation (V_R). 'Fish size' is the fixed effect of head size measured as the centroid size. 'Population' is the fixed effect for each population, a categorical factor that captures site and year specific influences on mean phenotype. 'Fish size:Population' is the interaction between these terms. 'Animal' is the random intercept of each fish, that is, their genetic breeding values, g.

Simulations using *genomalicous*' *family_sim_qtl* function were performed using simulated h^2 of 0, 0.05, 0.15, and 0.3 for each phenotypic PC and each of the 50 SP sets. For each SNP set, we simulated a total of 30 nested QT loci sets, each containing 100 randomly drawn loci to act as the simulated QT loci. We obtained simulated phenotypes by combining the genetic breeding values and residual deviations from simulations with fixed effect estimates of 'Size', 'Population', and 'Size:Population' (for phenotypic PC1 and PC3), which were derived from our observed data. Genomic animal models on simulated phenotypes were fit with the same form as the observed data, with **A** estimated with and without the simulated QT loci. Null models were also fit.

To visualise summarised estimates, we plotted the distributions of observed h^2_{hij} and P_{hij}, and simulated $\widetilde{h^2_{hij(N)}}$ and $\widetilde{P_{hij(N)}}$, and $\widetilde{h^2_{hij(N-M)}}$ and $\widetilde{P_{hij(N-M)}}$. We calculated the 2.5% and 97.5% percentiles to determine the bounds of the bulk of these distributions. We also calculated the percentage of model comparison median P-values that were significant across SNP sets, that is, $Pct(P_{ij} < 0.05)$, $Pct(\widetilde{P_{hij(N)}} < 0.05)$ and $Pct(\widetilde{P_{hij(N-M)}} < 0.05)$. This allowed us to assess the power and reliability to detect significant effects of additive genetic relationships in the observed, based on expectations from the simulated datasets.

We also performed validatory analyses within each spatiotemporally unique population. Our main analyses estimated **A** and fitted animal models across all individuals collected from Hastings Point and Heron Island in 2016 and 2017. We believe this is justifiable given the exceptionally weak genetic structure observed across these spatiotemporally separated populations – that is, our collection comprises a single effective population (see Results

section). Combining individuals across populations allowed us to achieve the largest possible sample size to maximise power. However, genetic heterogeneity within a dataset can potentially bias estimates of quantitative genetic parameters (Goudet et al., 2018; Kardos & Luikart, 2021; Muff et al., 2019; Weir & Goudet, 2017). We therefore separately estimated **A** and fitted animal models for within each of the four site-by-year populations. This allowed us to test whether analyses restricted within populations were consistent to those tests performed across populations. These validation analyses were only performed on the observed data.

RESULTS

Macrogeographic structuring of genotypes and phenotypes

Genetic structure among spatiotemporally separated populations of *B. cocosensis* was subtle. Although all values were significant based on permutation tests ($P_{perm} < 0.05$), F_{ST} values were very small, with a global F_{ST} of 0.003 and pairwise F_{ST} ranging from 0.001 to 0.003 (Supplementary Table S1). Phenotypic structure among populations was high in contrast to genetic structure. Global P_{ST} was 0.80, with pairwise P_{ST} ranging from 0.30 to 0.33 (Supplementary Table S2), and all values were significant based on permutation tests ($p_{perm} < 0.05$). This suggests that processes generating spatiotemporal patterns of head shape phenotypes are potentially greater than any forces (like migration) that might be homogenising genetic variation.

Phenotypic PC1 was associated with eye angle and size, and mouth size; it was best described by a full interaction model (model comparison: $F_{4,658} = 6.94$, $P < 0.001$) (Figure 3a & 3b). Phenotypic PC2 was associated with eye size and mouth angle; it was best described by a reduced additive model (model comparison: $F_{4,658} = 1.30$, $P = 0.27$) (Figure 3c & 3d). Phenotypic PC3 tended to be associated with eye angle and size; it was best described by a full interaction model (model comparison: $F_{4,658} = 7.97$, $P < 0.001$) (Figure 3e & 3f). These results overall suggest that there were clear differences in mean phenotype among the spatiotemporally separated populations, and that the relationships between some of our focal head shape traits with fish size varied among these populations.

Microgeographic structuring of phenotypes

Our mixed effect linear models provided some evidence that head shape phenotypes at Hastings Point may have some association with microgeographic variation in tidepool microhabitats. We only detected significant association with environment in the 2018 population (Figure 4; Table 1). Phenotypic PC1 scores were negatively associated with tide pool volume and positively associated with depth (Figure 4; Table 1; Supplementary Table S3). In other words, larger pools (decreasing PC1 scores) were associated with smaller mouths, and smaller less angled eyes, whereas deeper pools (increasing PC1 scores) were associated with larger mouths, and smaller more angled eyes (e.g., Figure 3a). Phenotypic PC2 did not exhibit any significant associations after correcting for multiple testing, but there was a marginally non-significant positive association with volume (Figure 4; Table 1; Supplementary Table S3). This perhaps reflected a weak tendency for larger pools (increasing PC2 scores) to contain fish with more angled mouths and smaller eyes (e.g., Figure 3c). Phenotypic PC3 was positively associated with zone and negatively associated with depth and rugosity (Figure 4; Table 1; Supplementary Table S3). This suggests that pools at higher tidal zone positions (increasing PC3 scores) were associated with smaller and more angled eyes, whereas deeper and more rugged pools (decreasing PC3 scores) were associated with larger and less angled eyes (e.g., Figure 3e).

Additional tests of the consistency in mean tide pool head shape suggested that only phenotypic PC2 was consistent between temporally separated populations. The Spearman rank correlations, r_s, for phenotypic PC1 and PC3 were $r_s = 0.32$ ($P = 0.40$) and $r_s = 0.001$ ($P = 1.00$), respectively, suggesting the rank of tide pools was not consistent between 2017 and 2018, after controlling for differences in head size. In contrast, for phenotypic PC2, $r_s = 0.77$ ($P < 0.015$), suggesting that tide pools with higher mean phenotypic scores were consistent between 2017 and 2018, even after controlling for differences in head size. These results of microgeographic structure in head shape may suggest some potential temporal variability in how fish respond to different microhabitat variables.

Heritability of head shape phenotypes

By applying our genomic animal model framework to wild-caught *B. cocosensis*, we were able to detect additive genetic variance for head shape described by phenotypic PC1 and PC2,

with median estimates of $\widetilde{h^2_{\text{PC1}}}$ = 0.26 and $\widetilde{h^2_{\text{PC2}}}$ = 0.18, respectively, across SNP sets (Table 2; Figure 5). There was little evidence for heritability in phenotypic PC3, with $\widetilde{h^2_{\text{PC3}}}$ < 0.0001 across SNP sets (Table 2; Figure 5). Despite evidence for heritable variation in phenotypic PC1 and PC2, comparisons between animal and null models tended to yield *P*-values > 0.05 (Table 2; Figure 5). The percentage of SNP sets that yielded significant model comparisons was low, with $Pct(P_{\text{PC1}} < 0.05)$ = 30% SNP sets for phenotypic PC1, and $Pct(P_{\text{PC2}} < 0.05)$ = 4% SNP sets for phenotypic PC2 (Table 2). Point estimates of observed heritability, h^2_{ij}, for the ith trait and the jth SNP set, and their uncertainty, are illustrated in Supplementary Figure S8.

The robustness of our observed heritability estimates and model comparison tests was assessed by comparing their distributions to simulated results. Uncertainty estimates for each simulated heritability point estimate, h^2_{hijk}, for each hth level of heritability, are illustrated in Supplementary Figures S9, S10, and S11. Our first key result was that simulations supported that presence of heritability of head shape in phenotypic PC1 and PC2, but not PC3. By constraining simulated phenotypes under different levels of heritability within our experimental design, we show that we had power to detect heritability at all simulated levels (h^2 at 0.05, 0.15, and 0.3), given the sampled individuals and the constructed **A** (Figure 5: compare purple vs dark green curves). The observed distribution of phenotypic PC1 heritability largely overlapped with simulated heritability of h^2 at 0.15 and 0.3 (Figure 5). The observed distribution of phenotypic PC2 heritability overlapped with simulated heritability largely overlapped with simulated h^2 of 0.05 and 0.15. In contrast, the observed distribution of phenotypic PC3 heritability largely overlapped with simulated h^2 of 0.

Our second key result is that our simulations could provide insight into how well our observed dataset might capture QT loci contributing to head shape phenotypes. We can see that the power to accurately estimate the true simulated heritability with our experimental design would have decreased if **A** calculation was performed without all QT loci (Figure 5: compare light vs dark green curves). These simulations therefore suggest that we might have captured some of the genetic markers in linkage with true QT loci to achieve the observed heritability estimates in phenotypic PC1 and PC2. However, the greater variance in our observed heritability estimates might suggests our current experimental design is missing

many linked genetic markers, and (or) the linked genetic markers we have assayed might not fully describe the additive genetic effects in phenotypic PC1 and 2.

Our third key result is that our simulations cast doubt on the value of model comparisons in testing whether estimated additive genetic variance was statistically significant. In principle, if **A** explains significant variance in phenotypes, we expect to obtain significant *P*-values when comparing the genomic animal models (with **A**) to null models (without **A**). In contrast, our results show that this expectation is often unrealised under our experimental design. Comparisons of simulated animal models versus null model comparisons also overwhelmingly yielded non-significant *P*-values ($P > 0.05$) (Supplementary Figure S12 and Table S4). The only simulated datasets where we did recover significant model comparisons was at simulated h^2 of 0.3 and when all QT loci were included in **A** matrix calculation (Supplementary Figure S12 and Table S4). The percentage of significant model comparisons in our simulated datasets at simulated h^2 of 0.3 with QT loci was 16%, 12% and 10% for phenotypic PC1, PC2, and PC3, respectively (Supplementary Table S4). The percentage of significant model comparisons in our observed data was 30%, 4% and 0% for phenotypic PC1, PC2, and PC3, respectively (Table 2). Our observed data therefore exceeded expectations for simulated h^2 of 0.3 for phenotypic PC1, whereas phenotypic PC2 and PC3 fell within expectations from simulated $h^2 < 0.3$, with respect to model comparison *P*-values. Our validatory analyses within each of the population samples showed similar signals of heritability as observed for the full data – with greater evidence for the presence of additive genetic variance in the phenotypic PC1 and PC2 traits (Supplementary Figure S13). The results from our validatory analyses therefore suggest that our fitting of animal models across populations is unlikely to have inflated estimates of heritability. However, as expected, these validatory analyses produced more variable results given the smaller sample sizes.

DISCUSSION

In this study, we developed a new framework using genomic animal models to study the quantitative genetics of wild organisms. This framework expands upon available methods by facilitating simulation of trait heritability from datasets lacking kin structure and variability in genomic relatedness. Our framework also leverages simulation across differently sampled

SNP sets to understand how genetic marker choice affects quantitative genetic estimation. Applying this framework to a marine intertidal fish, *B. cocosensis*, we detected additive genetic variance in some head shape traits, suggesting capacity for head shape evolution. We also detected strong patterns of spatiotemporal variation in head shape against a background of genome-wide genetic homogeneity.

A resampling simulation framework to assess the robustness of quantitative genetic estimates

Two features of an experimental design can affect the power and reliability of genomic animal models: the sampled individuals and genetic markers. Sampled individuals are a fixed feature of an experimental design in that they cannot (often) be recollected. In wild organisms with large population sizes and low probability of sampling close-kin, the constructed **A** (the genomic relationships matrix) is therefore unlikely to contain much variability in additive genetic relationships, with all individuals being uniformly distantly related (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016). In contrast, the final set of genetic markers used to construct **A** is a more fluid feature of an experimental design and is determined by the sequencing approach (whole-genome vs. reduced representation) and filtering steps. This impacts **A** because genetic markers vary in their information content with respect to reflecting the contribution of QT (quantitative trait) loci to the focal trait (de los Campos et al., 2015; Hill & Weir, 2011; Stanton-Geddes et al., 2013; Wang et al., 2017). In our framework, sampling of different SNP sets provides an assessment of how genetic marker choice affects **A** and hence generates different outcomes from genomic animal models. Other studies have adopted similar sampling approaches (de los Campos et al., 2015; Kumar et al., 2016; Wang et al., 2017). Our framework extends those approaches by including simulations of heritability on these resampled SNP sets, providing an assessment of the power and reliability of genomic animal models fitted with **A** generated from each SNP set. Results generated from simulations thus provide expectations for the observed dataset.

Simulation of phenotypes in our framework is underpinned by a novel take on a classic quantitative genetic formula from Falconer (1989): $V_A = 2pq[a + d(q - p)]^2$. By rearranging and iteratively solving this formula across different loci, we could directly simulate phenotypes from genotypes with a desired amount of heritability. This allows the simulated

phenotypes to be constrained by the allele frequency spectrum of a sample, but at the same time be simulated independently of the **A** constructed from those assayed genetic markers. This contrasts the common approach whereby phenotypes are simulated based on the structure of **A** (or some genetically inferred pedigree), as implemented, for example, in the *MCMCglmm* R package (Hadfield, 2010). Our phenotypes-from-genotypes method therefore circumvents issues that may arise when there is very low (or no) kin structure and when pedigrees cannot be inferred from the observed data. This provides an advantage when working with wild organisms that have demographic and life history characteristics that make it nearly (if not) impossible for random samples to include close-kin and variability in genomic relatedness.

Three key results emerged from the development of our framework and its application to wild *B. cocosensis*. First, our framework gave us more confidence that results obtained from our genomic animal models were robust. When **A** contains little variation in relatedness, heritability can be hard to detect and have large confidence intervals (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016). This was evident in both our observed and simulated datasets. For phenotypic PC1 and PC2, despite moderate point estimates of heritability, standard errors around these estimates in many SNP sets were large and approached zero. Our simulations also showed that even when there is no heritable variation, spurious covariance in some SNP sets could produce results suggestive of non-zero heritability. By placing our observed results in the context of simulated results, were able to infer what magnitude of heritability could be detected in our experiment. Our results suggested that our observed distribution of heritability estimates for phenotypic PC1 and PC2 were consistent with expectations of when there was true heritable variation in these traits. In contrast, the observed heritability estimate distribution for the third head shape trait, PC3, was consistent with expectations of no heritable variation.

Our second key result was that our simulations provided insight into how well our observed dataset might capture QT loci contributing to focal traits. Our simulations include an additional resampling step to generate nested QT loci sets within each SNP set. Phenotypes are generated by iterating over QT loci sets, fitting genomic animal models per QT loci set, and then summarising across QT loci sets to get a median effect per SNP set. Within simulated iterations, constructing **A** with and without the QT loci provided another angle to

understand how genomic animal model robustness was affected by genetic marker choice. Our simulations showed that power to detect heritability generally decreases when **A** is constructed without QT loci, which is consistent with previous work (de los Campos et al., 2015; Wang et al., 2017). Our simulations hint at the possibility that we may have captured some SNPs associated (in linkage with) causal QT loci. Our observed heritability estimate distributions would be unlikely if no QT loci had been included in our SNP sets. Understanding whether a dataset contains QT loci, or genetic markers in linkage with them, is especially relevant when using reduced representation sequencing in the absence of a reference genome. A major caveat of anonymous reduced representation markers is that it is often unclear whether they appropriately sample the genome and capture regions underpinning adaptation or focal traits (Hoban et al., 2016; Lowry, Hoban, Kelley, Lotterhos, Reed, & Antolin, 2017). Our framework may therefore alleviate some of these issues by helping assess whether observed results are consistent with a dataset that has assayed, or not assayed, candidate genetic markers. However, in this study, simulated **A** was constructed with or without all the simulated QT loci. Hence, an area of future work may be to explore how different numbers of QT loci used to construct **A** affects the robustness of genomic animal models.

Finally, our third key result was that model comparisons may not be useful for testing the significance of heritability in a sample of wild organisms lacking kin structure. The model comparison procedure involves fitting a genomic animal model with **A** (accounting for additive genetic covariances) and testing whether it explains more of the phenotypic variance than a null model fitted without **A**. Across our simulations of specific levels of non-zero heritable variation, we found that model comparisons repeatedly yielded *P*-values >0.05, even when simulated heritability was moderate. This suggests that null model comparisons are not reliable indicators of whether **A** explains significant phenotypic variance. However, we did find that the distribution of *P*-values across different SNP sets to be informative of the robustness of genomic animal models. Our simulations showed that although *P*-values may not be significant at a threshold of >0.05, the overall distribution of *P*-values across SNP sets shifted towards 0 (and away from 1) with increasing levels of simulated heritability. In our dataset from *B. cocosensis*, our observed *P*-value distributions for phenotypic PC1 and PC2 were also left shifted, whereas phenotypic PC3 exhibited a large

stack of *P*-values at 1. These results again provided confidence that there was non-zero additive genetic variance for PC1 and PC2 head shape traits. More broadly, our results provide a further demonstration of the utility of combining resampling and simulations in our framework for genomic animal models.

Spatiotemporally structured head shape phenotypes have a heritable component in *Bathygobius cocosensis*

This study, and others (Thia et al., 2018, 2021), demonstrate that *B. cocosensis* is a classically panmictic marine species (Waples, 1998). A long pelagic dispersal period (high gene flow) and large effective population sizes (low drift) likely maintains genetic homogeneity across space and time in the east Australian range of this fish (Thia et al., 2018, 2021). Such among-population genetic homogeneity might be expected to be associated with phenotypic homogeneity if head shape was largely genetically controlled and neutrally evolving (Thia et al., 2021; Whitlock, 1999). Instead, phenotypic structure emerged across macrogeographic scales and across different annual cohorts. Hastings Point (a rocky outcrop in temperate-subtropical climate) and Heron Island (a tropical coral island) represent distinct sites that differ on multiple abiotic (e.g., topology and climate) and biotic (local ecological communities) axes. Hence, we suggest that our results hint at the potential for spatiotemporally variable environments to drive phenotypic differentiation against processes that maintain genetic homogeneity. Examples of such phenotypic differentiation against high gene flow in other marine taxa include amphidromous waterfall climbing gobies (Moody et al., 2015), bass (Rey et al., 2020a), sticklebacks (Defaveri & Merilä, 2013), and European green crab (Tepolt & Palumbi, 2020).

Our study in *B. cocosensis* provides new insight into how head shape phenotypes are locally structured over microgeographic scales. Previous work at Hastings Point demonstrated that *B. cocosensis* found in pools from the high and low intertidal zones, which have marked environmental differences, also exhibited significantly different head shapes (Malard et al., 2016). Our present analyses of continuous microhabitat measures provided a finer resolution of how local factors might affect head shape, as well as their temporal variability. We observed significant associations between continuous measures of microhabitat (depth, volume, rugosity, and inundation time) and phenotypic variation across the intertidal zone at

Hastings Point. However, these associations with microhabitat were not always consistent. We found significant microhabitat associations for phenotypic PC1 and PC3 for the Hasting Point 2018 population, but not the 2017 population. But whilst tidepool ranks were significantly correlated between the 2017 and 2018 population for phenotypic PC2, they were not for PC1 and PC3. These data might suggest that processes structuring phenotypic variation on microgeographic scales are temporally variable, as has been reported in other systems (Mojica et al., 2012; Stratton & Bennington, 1998). For example, the effect of our *measured* microhabitat variables might be dependent on some other *unmeasured* variable that modulates how they affect head shape year-to-year.

If head shape is an important trait in *B. cocosensis*, local associations between head shape and microhabitat might reflect the role of selection in shaping patterns of phenotypic variation at microgeographic scales. Alternatively, these associations emerge through post-settlement plastic development of head shape in response to local conditions (e.g., food availability) or individuals seeking tidepools that match their phenotype. Rocky intertidal fish, including *B. cocosensis*, are known to exhibit homing behaviour to return to specific tidepools at low tide, or when experimentally displaced (Griffiths, 2003a; Malard et al., 2016; White & Brown, 2013). Experimentally dissecting how possible developmental processes, selection, and behavioural decisions might affect local phenotypic structure would represent a major undertaking. However, doing so would enable a deeper understanding of the how these different factors combine to shape local phenotypic patterns in rocky intertidal zones.

Using genomic animal models, we partitioned variation in head shape phenotypes into putatively heritable and non-heritable components. Here, our results indicated that some measures of head shape morphology have a heritable component ($\widetilde{h^2_{\mathrm{PC1}}} = 0.26$, $\widetilde{h^2_{\mathrm{PC2}}} = 0.18$, $\widetilde{h^2_{\mathrm{PC3}}} < 0.0001$). Observed estimates of h^2 for phenotypic PC1 and PC2 were in the lower range of values observed for morphological traits reported in other animal taxa (median of h^2 ~0.5) (Mousseau & Roff, 1987). Our findings suggest that patterns of head shape phenotypes could be shaped by selection, if selection acts on the heritable component of head shape. However, the long-distance dispersal potential of *B. cocosensis* means that different genotypes could disperse to quite different environments from their natal habitat patch. High dispersal, coupled with the inherent spatiotemporal variation in marine and

rocky intertidal habitats might make it difficult for selection to drive heritable changes in trait means. Instead, these conditions might favour the evolution of plasticity in *B. cocosensis* as a way to buffer against possible environmental uncertainty (Garant et al., 2007; Scheiner, 2013). Indeed, many other ecologically relevant traits have been shown to be plastic in *B. cocosensis* (Carbia et al., 2020; Carbia & Brown, 2019; da Silva et al., 2019, 2020). However, estimating the fitness benefits of plasticity will be difficult in this system where we must first determine whether head shape is plastic. Furthermore, while it is relatively straightforward to estimate the additive genetic variance in trait plasticity when close relatives experience alternative environments, it may be unattainable in systems such as *B. cocosensis* where close-kin are not sampled. Of course, even though head shape is known to have important ecological functions in other fishes (Burress et al., 2017; Moody et al., 2015; Schluter, 1993; Vera-Duarte et al., 2017), we acknowledge the need for validatory experiments to draw definitive conclusions about the adaptive role of this trait in *B. cocosensis*.

Concluding remarks

Our work on *B. cocosensis* provides a unique characterisation of the processes that structure genetic and phenotypic variation. This is the first study, to our knowledge, that fully combines measures of genetic, phenotypic, and environmental variation at different spatiotemporal scales in a marine intertidal fish. We also provide the first estimation of quantitative genetic parameters for a marine intertidal fish collected from the wild. The detected heritability in head shape is intriguing, given our hypothesis that this is an important trait in *B. cocosensis*. However, fully characterising the possible adaptive landscape of head shape phenotypes is the scope of future work.

Genomic animal models offer an attractive way to derive quantitative genetic estimates from wild populations in the absence of a pedigree (Bérénos et al., 2014; Duntsch et al., 2020). Although there is growing interest in using genomic animal models to study polygenic traits in wild organisms (Aase et al., 2022; Richards et al., 2023; Schmidt et al., 2023), their application to species with large and abundant populations has been unclear due to challenges in sampling close-kin pairs (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016). Our framework provides a novel application for using genomic animal models on such wild organisms. However, this framework could be applied to any system (kin structured or non-

kin structured samples), in studies using reduced representation or whole-genome sequencing data, and with any software to estimate quantitative genetic parameters and statistics.

We have included an R tutorial as an Appendix to assist interested researchers in understanding our framework and applying it to their focal system. We hope the developments made in this work help quantitative genomics go truly wild and encourage the molecular ecology community to not just hunt for candidate loci but to quantify the genetic variance of polygenic traits.

ACKNOWLEDGEMENTS

We thank I. Popovic, C. Da Silva, A. Matias, A. Mather, and D. Blower for field assistance. We thank Will Figueira for assisting in the logistics of the NSW collection. We thank Ary Hoffmann for valuable conversations around the evolution of plasticity, and Moshe Jasper for his useful insights on estimating relatedness from genomic data. Fish were collected under NSW DPI Permit P13/0046-1.1, and from QLD under the QLD Government Permit #174684. Funding was provided by The Herman Slade Foundation (HSF 13/14 to CR and LL) and The Ecological Society of Australia (Student Research Award 2015 to JAT). Major bioinformatic procedures were supported by the University of Queensland's High Performance Computing resources.

AUTHOR CONTRIBUTIONS

JAT conceptualised the novel genomic animal model framework, with input from JDA and KM. JAT, LL, KM, and CR conceptualised the original population genomic component of this study, with the addition of the quantitative genomic component developed by JAT, JDA, CR, and KM. JAT and JE led the collection of wild *B. cocosensis*, measurement of tide pool microhabitats, and measurement of head shape phenotypes. JAT led the molecular work (DNA extraction and ddRAD library preparation), and JH developed the ddRAD protocol and helped optimise it for *B. cocosensis*. JAT performed all the analyses in this study, created the

Appendix material, and archived all data and scripts. JDA, KM, and CR provided feedback on the analyses. JAT wrote the original version of this manuscript, with all authors providing feedback on the final version.

DATA AVAILABILITY

All data and scripts required to perform the analyses presented in this study are archived in a Dryad repository and will be released upon publication. All short-read Illumina ddRAD sequences have been uploaded to the GenBank BioProject, PRJNA595396. The associated GenBank SRA and BioSample accessions can be found in the Appendices. We have also included an R demonstration of our genomic animal model framework as an Appendix ZIP file.

REFERENCES

Aase, K., Jensen, H., & Muff, S. (2022). Genomic estimation of quantitative genetic parameters in wild admixed populations. *Methods in Ecology and Evolution*, *13*(5), 1014–1026. https://doi.org/10.1111/2041-210X.13810

Adams, D. C., & Otárola-Castillo, E. (2013). Geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, *4*, 393–399. https://doi.org/10.1111/2041-210X.12035

Babin, C., Gagnaire, P. A., Pavey, S. A., & Bernatchez, L. (2017). RAD-Seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American eel (*Anguilla rostrata*). *Genome Biology and Evolution*, *9*(11), 2974–2986. https://doi.org/10.1093/gbe/evx226

Bérénos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology*, *23*(14), 3434–3451. https://doi.org/https://doi.org/10.1111/mec.12827

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Burress, E. D., Holcomb, J. M., Tan, M., & Armbruster, J. W. (2017). Ecological diversification associated with the benthic-to-pelagic transition by North American minnows. *Journal of Evolutionary Biology*, *30*(3), 549–560.

Carbia, P. S., & Brown, C. (2019). Environmental enrichment influences spatial learning ability in captive-reared intertidal gobies (*Bathygobius cocosensis*). *Animal Cognition*. https://doi.org/10.1007/s10071-018-1225-8

Carbia, P. S., Brown, C., Park, J. M., Gaston, T. F., Raoult, V., & Williamson, J. E. (2020). Seasonal and developmental diet shifts in sympatric and allopatric intertidal gobies determined by stomach content and stable isotope analysis. *Journal of Fish Biology*.

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

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chevin, L. M., Lande, R., & Mace, G. M. (2010). Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLoS Biology*, *8*(4). https://doi.org/10.1371/journal.pbio.1000357

Christie, M. R., Johnson, D. W., Stallings, C. D., & Hixon, M. A. (2010). Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. *Molecular Ecology*, *19*, 1042–1057. https://doi.org/10.1111/j.1365-294X.2010.04524.x

Clark, L. V., Lipka, A. E., & Sacks, E. J. (2019). polyRAD: Genotype calling with uncertainty from sequencing data in polyploids and diploids. *G3: Genes, Genomes, Genetics*, *9*(3), 663–673. https://doi.org/10.1534/g3.118.200913

Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One*, *11*(6), e0156744. https://doi.org/10.1371/journal.pone.0156744

da Silva, C. R. B., Riginos, C., & Wilson, R. S. (2019). An intertidal fish shows thermal acclimation despite living in a rapidly fluctuating environment. *Journal of Comparative Physiology B*, *0*(0), 0. https://doi.org/10.1007/s00360-019-01212-0

da Silva, C. R. B., van den Berg, C. P., Condon, N. D., Riginos, C., Wilson, R. S., & Cheney, K. L. (2020). Intertidal gobies acclimate rate of luminance change for background matching with shifts in seasonal temperature. *Journal of Animal Ecology*, *89*(7), 1735–1746. https://doi.org/10.1111/1365-2656.13226

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

de los Campos, G., Sorensen, D., & Gianola, D. (2015). Genomic Heritability: What Is It? *PLoS Genetics*, *11*(5). https://doi.org/10.1371/journal.pgen.1005048

Defaveri, J., & Merilä, J. (2013). Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology*, *26*(8), 1700–1715. https://doi.org/10.1111/jeb.12168

DiBattista, J. D., Feldheim, K. A., Garant, D., Gruber, S. H., & Hendry, A. P. (2009). Evolutionary potential of a large marine vertebrate: Quantitative genetic parameters in a wild population. *Evolution*, *63*(4), 1051–1067. https://doi.org/10.1111/j.1558-5646.2008.00605.x

Duntsch, L., Tomotani, B. M., De Villemereuil, P., Brekke, P., Lee, K. D., Ewen, J. G., & Santure, A. W. (2020). Polygenic basis for adaptive morphological variation in a threatened Aotearoa | New Zealand bird, the hihi (*Notiomystis cincta*). *Proceedings of the Royal Society B: Biological Sciences*, *287*(1933). https://doi.org/10.1098/RSPB.2020.0948

Evans, J. L., Thia, J. A., Riginos, C., & Hereward, J. P. (2018). The complete mitochondrial genome of *Bathygobius cocosensis* (Perciformes, Gobiidae). *Mitochondrial DNA Part B: Resources*, *3*(1). https://doi.org/10.1080/23802359.2018.1437824

Falconer, D. S. (1989). Variance. In *Introduction to Quantitative Genetics* (3rd ed., p. 129). Logman Scientific and Technical.

Fraimout, A., Guillaume, F., Li, Z., Sillanpää, M. J., Rastas, P., & Merilä, J. (2024). Dissecting the genetic architecture of quantitative traits using genome-wide identity-by-descent sharing. *Molecular Ecology*, *33*(6), 1–15. https://doi.org/10.1111/mec.17299

Fuller, Z. L., Mocellin, V. J. L., Morris, L., Cantin, N., Shepherd, J., Sarre, L., Peng, J., Liao, Y., Pickrell, J., Andolfatto, P. L., Matz, M. V, Bay, L. K., & Przeworski, M. (2020). Population genetics of the coral *Acropora millepora*: Towards a genomic predictor of bleaching. *Science*, *369*(eaba4674). https://doi.org/10.1101/867754

Gagnaire, P. A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, *62*(6), 603–616. https://doi.org/10.1093/cz/zow088

Garant, D., Forde, S. E., & Hendry, A. P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, *21*(3), 434–443.

Gervais, L., Perrier, C., Bernard, M., Merlet, J., Pemberton, J. M., Pujol, B., & Quéméré, E. (2019). RAD-sequencing for estimating genomic relatedness matrix-based heritability in the wild: A case study in roe deer. *Molecular Ecology Resources*, *19*(5), 1205–1217. https://doi.org/10.1111/1755-0998.13031

Gienapp, P., & Merilä, J. (2014). Disentangling plastic and genetic changes in body mass of Siberian jays. *Journal of Evolutionary Biology*, *27*(9), 1849–1858. https://doi.org/10.1111/jeb.12438

Goudet, J., Kay, T., & Weir, B. S. (2018). How to estimate kinship. *Molecular Ecology*, *27*(20), 4121–4135. https://doi.org/10.1111/mec.14833

Gracey, A. Y., Chaney, M. L., Boomhower, J. P., Tyburczy, W. R., Connor, K., & Somero, G. N. (2008). Rhythms of Gene Expression in a Fluctuating Intertidal Environment. *Current Biology*, *18*(19), 1501–1507. https://doi.org/10.1016/j.cub.2008.08.049

Griffiths, S. P. (2003a). Homing behaviour of intertidal rockpool fishes in south-eastern New South Wales, Australia. *Australian Journal of Zoology*, *51*, 387–398. https://doi.org/10.1071/ZO02049

Griffiths, S. P. (2003b). Rockpool ichthyofaunas of temperate Australia: Species composition, residency and biogeographic patterns. *Estuarine, Coastal and Shelf Science*, *58*, 173–186. https://doi.org/10.1016/S0272-7714(03)00073-8

Griffiths, S. P., West, R. J., & Davis, A. R. (2003). Effects of intertidal elevation on the rockpool ichthyofaunas of temperate Australia. *Environmental Biology of Fishes*, *68*, 197–204.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22.

Hedgecock, D., Launey, S., Pudovkin, A. I., Naciri, Y., Lapègue, S., & Bonhomme, F. (2007). Small effective number of parents (N_b) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Marine Biology*, *150*, 1173–1182. https://doi.org/10.1007/s00227-006-0441-y

Helmuth, B. S. T., & Hofmann, G. E. (2001). Microhabitats, thermal heterogeneity, and patterns of physiological stress in the rocky intertidal zone. *Biological Bulletin*, *201*(3), 374–384. https://doi.org/10.2307/1543615

Hill, W. G., & Weir, B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, *93*(1), 47–64. https://doi.org/10.1017/S0016672310000480

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Hoekstra, H. E., Drumm, K. E., & Nachman, M. W. (2004). Ecological genetics of adaptive colour polymorphism in pocket mice: Geographic variation in selected and neutral genes. *Evolution*, *58*(6), 1329–1341.

Hoffmann, A. A., & Bridle, J. (2022). The dangers of irreversibility in an age of increased uncertainty: revisiting plasticity in invertebrates. *Oikos*, *2022*(4), 1–14. https://doi.org/10.1111/oik.08715

Hoffmann, A. A., Shirriffs, J., & Scott, M. (2005). Relative importance of plastic vs genetic factors in adaptive differentiation: Geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. *Functional Ecology*, *19*(2), 222–227. https://doi.org/10.1111/j.1365-2435.2005.00959.x

James, C., Pemberton, J. M., Navarro, P., & Knott, S. (2022). The impact of SNP density on quantitative genetic analyses of body size traits in a wild population of Soay sheep. *Ecology and Evolution*, *12*(12), 1–15. https://doi.org/10.1002/ece3.9639

Johnson, M. S., & Black, R. (2008). Adaptive responses of independent traits to the same environmental gradient in the intertidal snail *Bembicium vittatum*. *Heredity*, *101*(1), 83–91. https://doi.org/10.1038/hdy.2008.33

Kardos, M., & Luikart, G. (2021). The genetic architecture of fitness drives population viability during rapid environmental change. *American Naturalist*, *197*(5), 511–525. https://doi.org/10.1086/713469

Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., Somboon, P., Mahanta, J., & Walton, C. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Molecular Ecology Resources*, *15*(5), 1031–1045. https://doi.org/10.1111/1755-0998.12369

Kim, S. (2015). ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665.

Koch, E. L., Morales, H. E., Larsson, J., Westram, A. M., Faria, R., Lemmon, A. R., Lemmon, E. M., Johannesson, K., & Butlin, R. K. (2021). Genetic variation for adaptive traits is associated

with polymorphic inversions in *Littorina saxatilis*. *Evolution Letters*, *5*(3), 196–213. https://doi.org/10.1002/evl3.227

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the "animal model." *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1446), 873–890. https://doi.org/10.1098/rstb.2003.1437

Kumar, S. K., Feldman, M. W., Rehkopf, D. H., & Tuljapurkar, S. (2016). Limitations of GCTA as a solution to the missing heritability problem. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), E61–E70. https://doi.org/10.1073/pnas.1520109113

Lasky, J. R., Josephs, E. B., & Morris, G. P. (2022). Genotype–environment associations to reveal the molecular basis of environmental adaptation. *The Plant Cell*, 1–14. https://doi.org/10.1093/plcell/koac267

Li, H. (2019). *Toolkit for processing sequences in FASTA/Q formats*.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., & Antolin, M. F. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 142–152. https://doi.org/10.1111/1755-0998.12635

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular Ecology Resources*, *17*, 366–369. https://doi.org/10.1111/1755-0998.12677

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60). https://doi.org/10.21105/joss.03139

Malard, L. A., McGuigan, K., & Riginos, C. (2016). Site fidelity, size, and morphology may differ by tidal position for an intertidal fish, *Bathygobius cocosensis* (Perciformes-Gobiidae), in Eastern Australia. *PeerJ*, *4*, e2263.

Malvezzi, A. J., Murray, C. S., Feldheim, K. A., Dibattista, J. D., Garant, D., Gobler, C. J., Chapman, D. D., & Baumann, H. (2015). A quantitative genetic approach to assess the evolutionary potential of a coastal marine fish to ocean acidification. *Evolutionary Applications*, *8*(4), 352–362. https://doi.org/10.1111/eva.12248

McCairns, R. J. S., & Bernatchez, L. (2012). Plasticity and heritability of morphological variation within and between parapatric stickleback demes. *Journal of Evolutionary Biology*, *25*(6), 1097–1112. https://doi.org/10.1111/j.1420-9101.2012.02496.x

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*, 3223–3231. https://doi.org/10.1111/mec.13243

Merilä, J., & Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications*, *7*(1), 1–14. https://doi.org/10.1111/eva.12137

Mojica, J. P., Lee, Y. W., Willis, J. H., & Kelly, J. K. (2012). Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology*, *21*(15), 3718–3728. https://doi.org/10.1111/j.1365-294X.2012.05662.x

Moody, K. N., Hunter, S. N., Childress, M. J., Blob, R. W., Schoenfuss, H. L., Blum, M. J., & Ptacek, M. B. (2015). Local adaptation despite high gene flow in the waterfall-climbing Hawaiian goby, *Sicyopterus stimpsoni*. *Molecular Ecology*, *24*(3), 545–563. https://doi.org/10.1111/mec.13016

Mousseau, T. A., & Roff, D. A. (1987). Natural selection and the heritability of fitness components. *Heredity*, *59*, 181–197.

Muff, S., Niskanen, A. K., Saatoglu, D., Keller, L. F., & Jensen, H. (2019). Animal models with group-specific additive genetic variances: Extending genetic group models. *Genetics Selection Evolution*, *51*(1), 1–16. https://doi.org/10.1186/s12711-019-0449-7

Ørsted, M., Rohde, P. D., Hoffmann, A. A., Sørensen, P., & Kristensen, T. N. (2018). Environmental variation partitioned into separate heritable components. *Evolution*, *72*(1), 136–152. https://doi.org/10.1111/evo.13391

Parsons, K. E. (1997). Role of dispersal ability in the phenotypic differentiation and plasticity of two marine gastropods. *Oecologia*, *110*, 461–471.

Porth, I., Klápště, J., McKown, A. D., La Mantia, J., Guy, R. D., Ingvarsson, P. K., Hamelin, R., Mansfield, S. D., Ehlting, J., Douglas, C. J., & El-Kassaby, Y. A. (2015). Evolutionary quantitative genomics of *Populus trichocarpa*. *PLoS ONE*, *10*(11), 1–25. https://doi.org/10.1371/journal.pone.0142864

R Core Team. (2022). *R: A language and environment for statistical computing (v4.1.3)*.

Rey, C., Darnaude, A., Ferraton, F., Guinand, B., Bonhomme, F., Bierne, N., & Gagnaire, P.-A. (2020a). Within-Generation Polygenic Selection Shapes Fitness-Related Traits across Environments in Juvenile Sea Bream. *Genes*, *11*(4), 398. https://doi.org/10.3390/genes11040398

Rey, C., Darnaude, A., Ferraton, F., Guinand, B., Bonhomme, F., Bierne, N., & Gagnaire, P.-A. (2020b). Within-Generation Polygenic Selection Shapes Fitness-Related Traits across Environments in Juvenile Sea Bream. *Genes*, *11*(4), 398. https://doi.org/10.3390/genes11040398

Richards, T. J., McGuigan, K., Aguirre, J. D., Humanes, A., Bozec, Y. M., Mumby, P. J., & Riginos, C. (2023). Moving beyond heritability in the search for coral adaptive potential. *Global Change Biology*, *29*(14), 3869–3882. https://doi.org/10.1111/gcb.16719

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rohlf, F. J. (2015). *tpsDig* (2.22).

Salles, O. C., Almany, G. R., Berumen, M. L., Jones, G. P., Saenz-Agudelo, P., Srinivasan, M., Thorrold, S. R., Pujol, B., & Planes, S. (2020). Strong habitat and weak genetic effects shape the lifetime reproductive success in a wild clownfish population. *Ecology Letters*, *23*(2), 265–273. https://doi.org/10.1111/ele.13428

Santure, A. W., & Garant, D. (2018). Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources*, *18*(4), 729–738.

Scheiner, S. M. (2013). The genetics of phenotypic plasticity: XII: Temporal and spatial heterogeneity. *Ecology and Evolution*, *3*(13), 4596–4609. https://doi.org/10.1002/ece3.792

Schluter, D. (1993). Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. *Ecology*, *74*(3), 699–709.

Schmidt, P. S., & Rand, D. M. (1999). Intertidal microhabitat and selection at Mpi: interlocus contrasts in the northern acorn barnacle, Semibalanus glandula. *Evolution*, *53*(1), 135–146. https://doi.org/10.2307/2640926

Schmidt, T. L., Thia, J. A., & Hoffmann, A. A. (2023). How Can Genomics Help or Hinder Wildlife Conservation? *Annual Review of Animal Biosciences*, 1–24.

Schmitt, S., Tysklind, N., Hérault, B., & Heuertz, M. (2021). Topography drives microgeographic adaptations of closely related species in two tropical tree species complexes. *Molecular Ecology*, *30*(20), 5080–5093. https://doi.org/10.1111/mec.16116

Seeley, R. H. (1986). Intense natural selection caused a rapid morphological transition in a living marine snail. *Proceedings of the National Academy of Sciences*, *83*(18), 6897–6901. https://doi.org/10.1073/pnas.83.18.6897

Selwyn, J. D., Hogan, J. D., Downey-Wall, A. M., Gurski, L. M., Portnoy, D. S., & Heath, D. D. (2016). Kin-aggregations explain chaotic genetic patchiness, a commonly observed genetic pattern, in a marine fish. *PLoS ONE*, *11*(4), e0153381.

Sokolova, I. M., Granovitch, A. I., Berger, V. Ja., & Johannesson, K. (2000). Intraspecific physiological variability of the gastropod *Littorina saxatilis* related to the vertical shore

gradient in the White and North Seas. *Marine Biology*, *137*(2), 297–308. https://doi.org/10.1007/s002270000343

Stanton-Geddes, J., Yoder, J. B., Briskine, R., Young, N. D., & Tiffin, P. (2013). Estimating heritability using genomic data. *Methods in Ecology and Evolution*, *4*(12), 1151–1158. https://doi.org/10.1111/2041-210X.12129

Stratton, D. A., & Bennington, C. C. (1998). Fine-grained spatial and temporal variation in selection does not maintain genetic variation in *Erigeron annuus*. *Evolution*, *52*(3), 678–691.

Tepolt, C. K., Grosholz, E., de Rivera, C., & Ruiz, G. (2021). Balanced polymorphism fuels rapid selection in an invasive crab despite high gene flow and low genetic diversity. *Molecular Ecology*, 1–15. https://doi.org/10.1111/mec.16143

Tepolt, C. K., & Palumbi, S. R. (2020). Rapid Adaptation to Temperature via a Potential Genomic Island of Divergence in the Invasive Green Crab, *Carcinus maenas*. *Frontiers in Ecology and Evolution*, *8*, 1–14. https://doi.org/10.3389/fevo.2020.580701

Thia, J. A. (2024). *genomalicious*: serving up a smorgasbord of R functions for performing and teaching population genomic analyses. *BioRxiv*. https://doi.org/doi.org/10.1101/667337

Thia, J. A., McGuigan, K., Liggins, L., Figueira, W. F., Bird, C. E., Mather, A., Evans, J. L., & Riginos, C. (2021). Genetic and phenotypic variation exhibit both predictable and stochastic patterns across an intertidal fish metapopulation. *Molecular Ecology*, *January*, 1–23. https://doi.org/10.1111/mec.15829

Thia, J. A., Riginos, C., Liggins, L., Figueira, W. F., & McGuigan, K. (2018). Larval traits show temporally consistent constraints, but are decoupled from postsettlement juvenile growth, in an intertidal fish. *Journal of Animal Ecology*, *87*(5), 1353–1363. https://doi.org/10.1111/1365-2656.12844

Trussell, G. C. (1996). Phenotypic plasticity in an intertidal snail: The role of a common crab predator. *Evolution*, *50*(1), 448–454. https://doi.org/10.2307/2410815

Trussell, G. C. (2000). Phenotypic clines, plasticity, and morphological trade-offs in an intertidal snail. *Evolution*, *54*(1), 151–166. https://doi.org/10.1111/j.0014-3820.2000.tb00016.x

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Vera-Duarte, J., Bustos, C. A., & Landaeta, M. F. (2017). Diet and body shape changes of pāroko *Kelloggella disalvoi* (Gobiidae) from intertidal pools of Easter Island. *Journal of Fish Biology*, *91*(5), 1319–1336.

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505–522.

Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F., & Xiang, J. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genetics*, *18*(1). https://doi.org/10.1186/s12863-017-0507-5

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438–450.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Weir, B. S., & Goudet, J. (2017). A Unified Characterization of Population Structure and relatedness. *Genetics*, *206*, 2085–2103.

Westram, A., Galindo, J., Rosenblad, M. A., Grahame, J., Panova, M., & Butlin, R. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, 4603–4616. https://doi.org/10.1111/mec.12883

White, G. E., & Brown, C. (2013). Site fidelity and homing behaviour in intertidal fishes. *Marine Biology*, *160*, 1365–1372. https://doi.org/10.1007/s00227-013-2188-6

Whitlock, M. C. (1999). Neutral additive genetic variance in a metapopulation. *Genetics Research*, *74*(3), 215–221.

Willis, S. C., Winemiller, K. O., & Lopez-Fernandez, H. (2005). Habitat structural complexity and morphological diversity of fish assemblages in a Neotropical floodplain river. *Oecologia*, *142*, 284–295.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, *79*(1), 13–26. https://doi.org/10.1111/j.1365-2656.2009.01639.x

FIGURES AND TABLES

Table 1. Analysis of deviance tables for mixed effect linear models of head shape phenotypes among tide pools at Hastings Point in the 2017 and 2018 sampling years. Asterisks for *P*-values denote significance at a threshold of $P < 0.008$ (0.05/6 models) to account for multiple testing, and tildes indicate significance at the $P < 0.05$ threshold.

Trait	Population	Term	χ^2	DF	*P*-value
Pheno PC1	Hast17	Fish size	82.5	1	<0.001 *
		Zone	3.7	1	0.053
		Volume	1.2	1	0.276
		Depth	2.0	1	0.154
		Rugosity	1.4	1	0.236
	Hast18	Fish size	32.1	1	<0.001 *
		Zone	1.0	1	0.327
		Volume	14.4	1	<0.001 *
		Depth	20.6	1	<0.001 *
		Rugosity	0.5	1	0.491
Pheno PC2	Hast17	Fish size	172.3	1	<0.001 *
		Zone	3.6	1	0.057
		Volume	0.1	1	0.717
		Depth	0.0	1	0.832
		Rugosity	2.3	1	0.132
	Hast18	Fish size	142.7	1	<0.001 *
		Zone	0.9	1	0.343
		Volume	4.5	1	0.034 ~
		Depth	3.5	1	0.062
		Rugosity	3.7	1	0.054
Pheno PC3	Hast17	Fish size	25.8	1	<0.001 *
		Zone	1.4	1	0.234
		Volume	1.5	1	0.216
		Depth	0.5	1	0.488
		Rugosity	0.0	1	0.980
	Hast18	Fish size	21.1	1	<0.001 *
		Zone	8.3	1	0.004 *

Volume	0.0	1	0.985
Depth	10.7	1	0.001 *
Rugosity	7.3	1	0.007 *

Table 2. Results from genomic animal models of head shape traits in wild *Bathygobius cocosensis*. Estimates of V_A, V_E, h^2, and P-values of model comparisons (animal versus null model), are reported as the median across 50 SNP sets with 2.5% and 97.5% percentiles in parentheses. The proportion of model comparisons with p-values < 0.05 across those SNP sets is also reported. Subscripts denote the ith trait and j^{th} SNP set.

Trait	$\widetilde{V_{A(i)}}$	$\widetilde{V_{E(i)}}$	$\widetilde{h_i^2}$	$\widetilde{P_i}$	$Pct(P_{ij} < 0.05)$
Pheno PC1	0.0004 (<0.0001, 0.0008)	0.001 (0.0007, 0.002)	0.26 (0.03, 0.53)	0.16 (0.006, 0.89)	30%
Pheno PC2	0.0002 (<0.0001, 0.0003)	0.0007 (0.0005, 0.0009)	0.18 (<0.0001, 0.40)	0.37 (0.03, 1)	4%
Pheno PC3	<0.0001 (<0.0001, 0.0001)	0.001 (0.001, 0.001)	<0.0001 (<0.0001, 0.06)	1 (0.76, 1)	0%

Table 3. Heritability estimates from observed versus simulated data. Estimates of h^2 are presented as medians with 2.5% and 97.5% percentiles across 50 SNP sets. Simulated estimates for each heritability level are represented with respect to whether **A** was constructed with vs without QTLs. Subscripts denote the ith trait and j^{th} SNP set.

Trait	**Obs $\widetilde{h_i^2}$**	**Sim h^2 level**	**Sim A-matrix**	**Sim $\widetilde{h_{ij}^2}$**
Pheno PC1	0.26 (0.03, 0.53)	0	Without QTLs	0.01 (<0.01, 0.11)
			With QTLs	0.01 (<0.01, 0.11)
		0.05	Without QTLs	0.02 (<0.01, 0.1)
			With QTLs	0.07 (0.01, 0.16)
		0.15	Without QTLs	0.05 (<0.01, 0.13)
			With QTLs	0.2 (0.12, 0.27)
		0.3	Without QTLs	0.04 (<0.01, 0.13)
			With QTLs	0.34 (0.27, 0.45)
Pheno PC2	0.18 (0, 0.4)	0	Without QTLs	0.01 (<0.01, 0.07)
			With QTLs	0.01 (<0.01, 0.07)
		0.05	Without QTLs	<0.01 (<0.01, 0.07)
			With QTLs	0.05 (<0.01, 0.15)
		0.15	Without QTLs	0.02 (<0.01, 0.09)
			With QTLs	0.17 (0.1, 0.26)
		0.3	Without QTLs	0.04 (<0.01, 0.12)
			With QTLs	0.34 (0.24, 0.43)
Pheno PC3	0 (0, 0.06)	0	Without QTLs	<0.01 (<0.01, 0.06)
			With QTLs	<0.01 (<0.01, 0.06)
		0.05	Without QTLs	0.01 (<0.01, 0.09)
			With QTLs	0.05 (<0.01, 0.15)
		0.15	Without QTLs	0.02 (<0.01, 0.12)
			With QTLs	0.18 (0.07, 0.27)
		0.3	Without QTLs	0.05 (<0.01, 0.13)
			With QTLs	0.33 (0.23, 0.41)

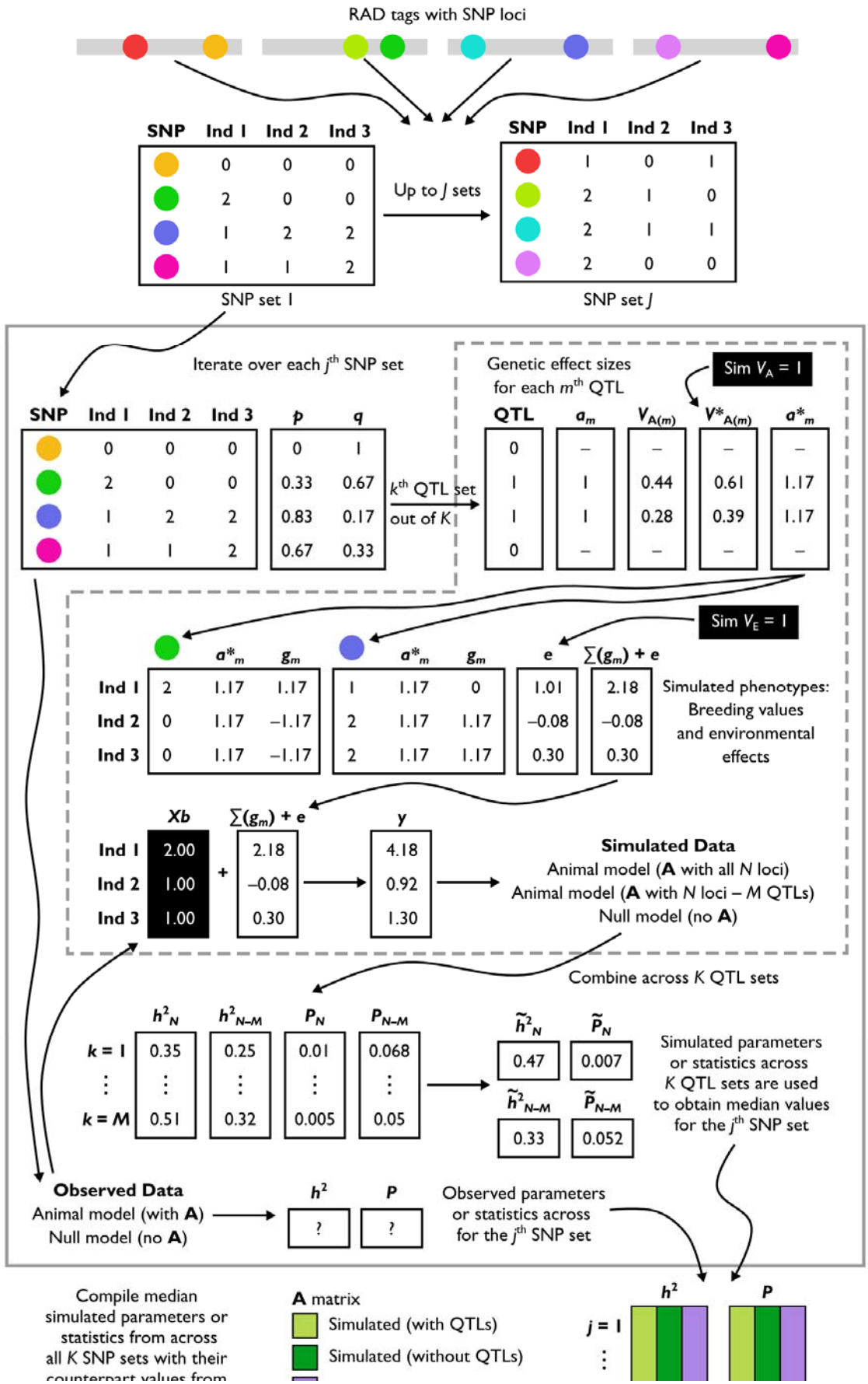
RAD tags with SNP loci
SNP Ind 1 Ind 2 Ind 3
0 0 0
2 0 0
1 2 2
1 1 2
SNP set 1
Up to J sets
SNP Ind 1 Ind 2 Ind 3
1 0 1
2 1 0
2 1 1
2 0 0
SNP set J
Iterate over each jth SNP set
SNP Ind 1 Ind 2 Ind 3 p q
0 0 0 0 1
2 0 0 0.33 0.67
1 2 2 0.83 0.17
1 1 2 0.67 0.33
k^th QTL set out of K
Genetic effect sizes for each mth QTL
Sim V_A = 1
QTL a_m V_A(m) V*_A(m) a*_m
0 – – – –
1 1 0.44 0.61 1.17
1 1 0.28 0.39 1.17
0 – – – –
Sim V_E = 1
a*_m g_m a*_m g_m e ∑(g_m) + e
Ind 1 2 1.17 1.17 1 1.17 0 1.01 2.18
Ind 2 0 1.17 –1.17 2 1.17 1.17 –0.08 –0.08
Ind 3 0 1.17 –1.17 2 1.17 1.17 0.30 0.30
Simulated phenotypes: Breeding values and environmental effects
Xb ∑(g_m) + e y
Ind 1 2.00 + 2.18 4.18
Ind 2 1.00 –0.08 0.92
Ind 3 1.00 0.30 1.30
Simulated Data
Animal model (A with all N loci)
Animal model (A with N loci – M QTLs)
Null model (no A)
Combine across K QTL sets
h²_N h²_N–M P_N P_N–M
k = 1 0.35 0.25 0.01 0.068
k = M 0.51 0.32 0.005 0.05
h̃²_N P̃_N
0.47 0.007
h̃²_N–M P̃_N–M
0.33 0.052
Simulated parameters or statistics across K QTL sets are used to obtain median values for the jth SNP set
Observed Data
Animal model (with A)
Null model (no A)
h² P
? ?
Observed parameters or statistics across for the jth SNP set
Compile median simulated parameters or statistics from across all K SNP sets with their counterpart values from
A matrix
Simulated (with QTLs)
Simulated (without QTLs)
h² P
j = 1

Figure 1. Conceptual schematic of our bootstrap simulation framework for genomic animal models. We start with a panel of markers (here, RAD tags with biallelic SNP loci) that pass filtering thresholds. We then generate a series of J randomly resampled SNP sets. Each jth SNP set is individually analysed as an observed and simulated dataset. For the jth observed dataset, we fit an animal model and a null model to derive parameters (e.g., h^2) and statistics (e.g., P-value for model comparison). For the jth simulated dataset, we generate a series of K randomly resampled QT loci sets from the N total loci, with M loci designated as simulated QT loci. We then iterate and perform simulations on each kth QT loci set. This involves calculating a locus-specific genetic values for each mth QT locus. We start by allocating a value of $a_m = 1$ to all QT loci. We then calculate a preliminary additive genetic variance for each mth QT locus, $V_{A(m)}$, simply from the allele frequencies at the locus. We then modify the additive genetic variances across QT loci based on a desired total simulated level (here, a value of 1), to adjust the locus-specific variances, $V^*_{A(m)}$. We then adjust the locus-specific genetic values, a^*_m, to fit with $V^*_{A(m)}$. Locus-specific genetic values are then used to derive locus-specific breeding values, g_m, which when summed, produce the total breeding value, $g = \sum g_m$. Each individual is assigned a randomly drawn environmental deviation, e, that is randomly drawn based on the desired level of environmental variance (here, a value of 1). The total simulated phenotype is given by $y = \mathbf{X}b + \sum g_m + e$, where $\mathbf{X}b$ represents the fixed effects obtained from the observed data ($\mathbf{X}$ is a design matrix, and b is the vector of partial regression coefficients). For the simulated phenotypes in the kth QT loci set, we fit three models: a genomic animal model with $\mathbf{A}$ calculated from all N loci, a genomic animal model with $\mathbf{A}$ calculated from N loci – M QT loci, and a null model. We then combine simulated parameters and statistics across the K QTL sets and calculate the median value for the jth SNP set. The observed point estimates and simulated median estimates for across all J SNP sets are combined.

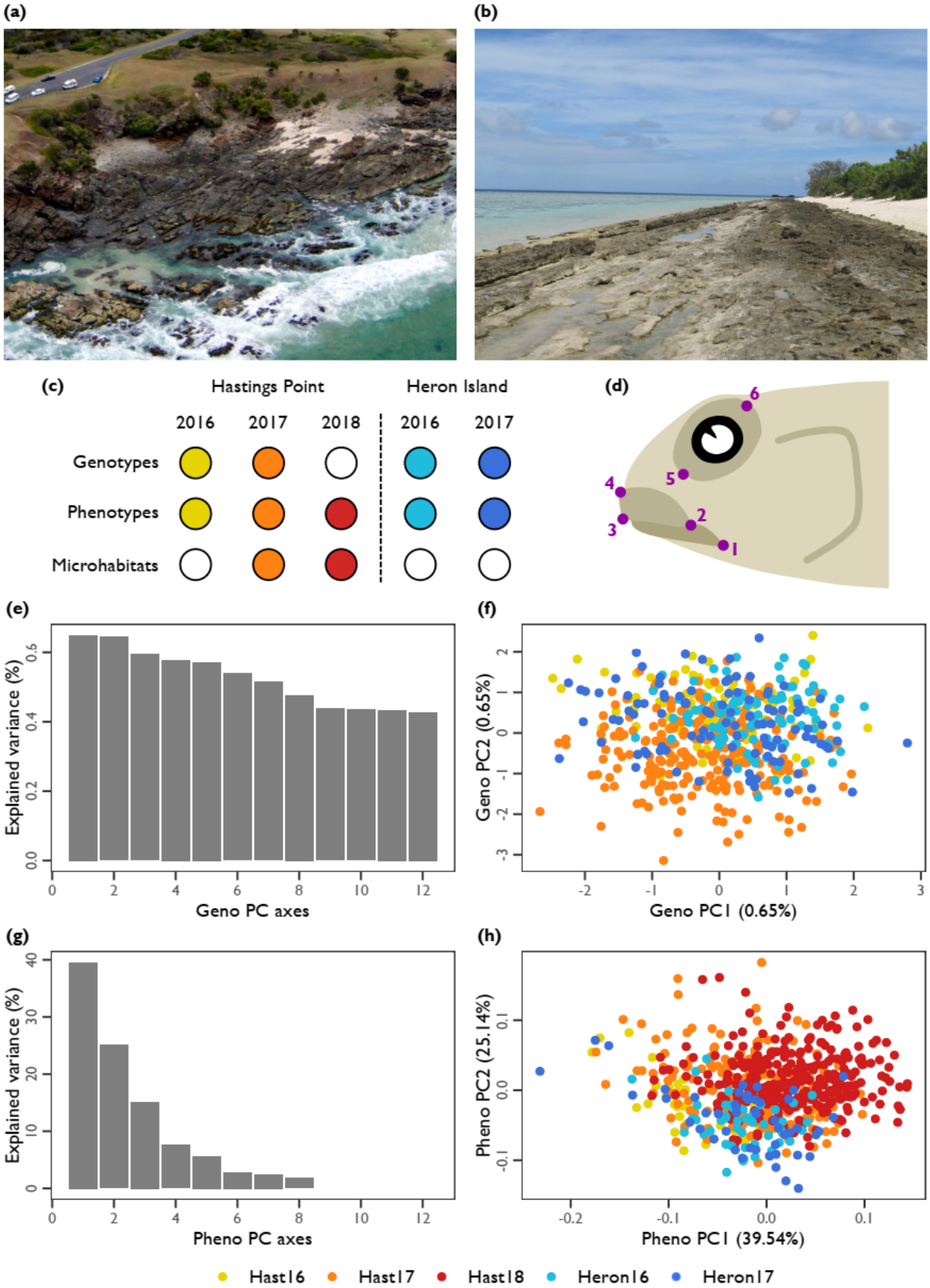
(a)
(b)
(c)
Hastings Point
Heron Island
2016
2017
2018
2016
2017
Genotypes
Phenotypes
Microhabitats
(d)
6
5
4
3
2
1
(e)
Explained variance (%)
Geno PC axes
(f)
Geno PC2 (0.65%)
Geno PC1 (0.65%)
(g)
Explained variance (%)
Pheno PC axes
(h)
Pheno PC2 (25.14%)
Pheno PC1 (39.54%)
Hast16
Hast17
Hast18
Heron16
Heron17

Figure 2. Schematic overview of our study on Australian populations of *Bathygobius cocosensis*. (a) Hastings Point (Photo credit: Prajwal Bhattaram). (b) Heron Island (Photo credit: Joshua Thia). (c) Sampling design of spatiotemporal populations from different sites, years, and data collected. Filled circles indicate collections for populations + data type; empty circles indicate the absence of these collections. (d) Head shape landmarks describing head shape morphology. (e, g) Screeplots of explained variance for principal component analyses (PCA) of (e) SNP genotypes and (g) head shape phenotypes: the *x*-axis represents the leading 12 PC axes and the *y*-axis represents the percent of explained variation. (f, h) Scatterplots of the leading two axes of PCAs of (f) SNP genotypes and (h) head shape phenotypes: the *x*-axis and *y*-axis represent PC1 and PC2, respectively, with percent of explained variation noted in axis labels; points represent individual fish, coloured by population (see legend).

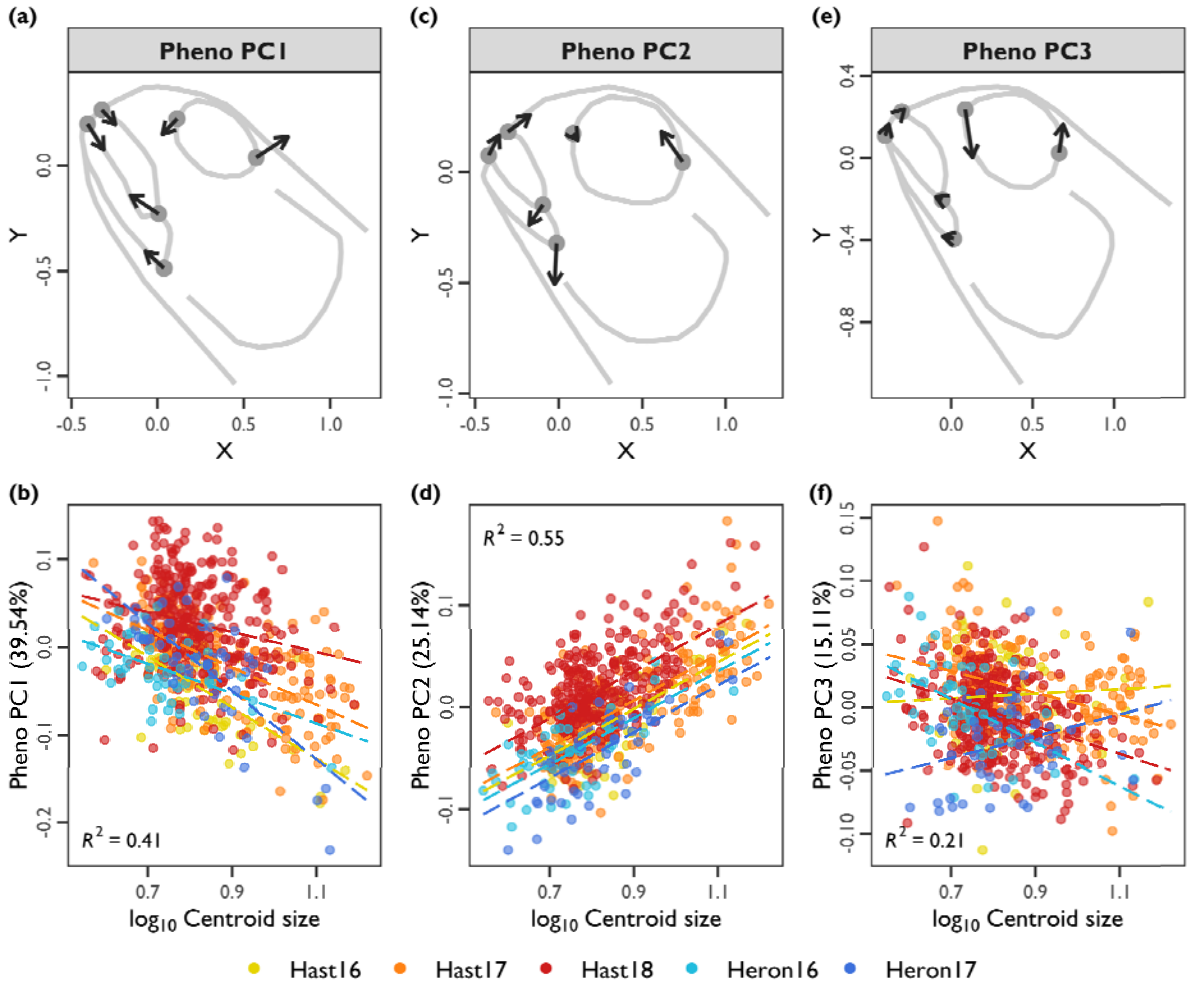

Figure 3. Phenotypic axes of head shape and their associations with fish size. (a, c, e) Head shape differences at the extremes of phenotypic axes. Each plot illustrates the head shape of lowest scoring fish on a focal axis. Arrows indicate the direction and magnitude of morphological landmark shifts from the lowest to highest scoring fish. (b, d, f) Associations between fish size, on the *x*-axes, and focal phenotypic axis, on the *y*-axes. Relative percentages of total are noted in the *y*-axis labels. Points represent individual fish, coloured by population (see legend). Dashed coloured lines represent the linear regression lines for each population. The coefficient of variation (R^2) for linear regressions are noted in each plot.

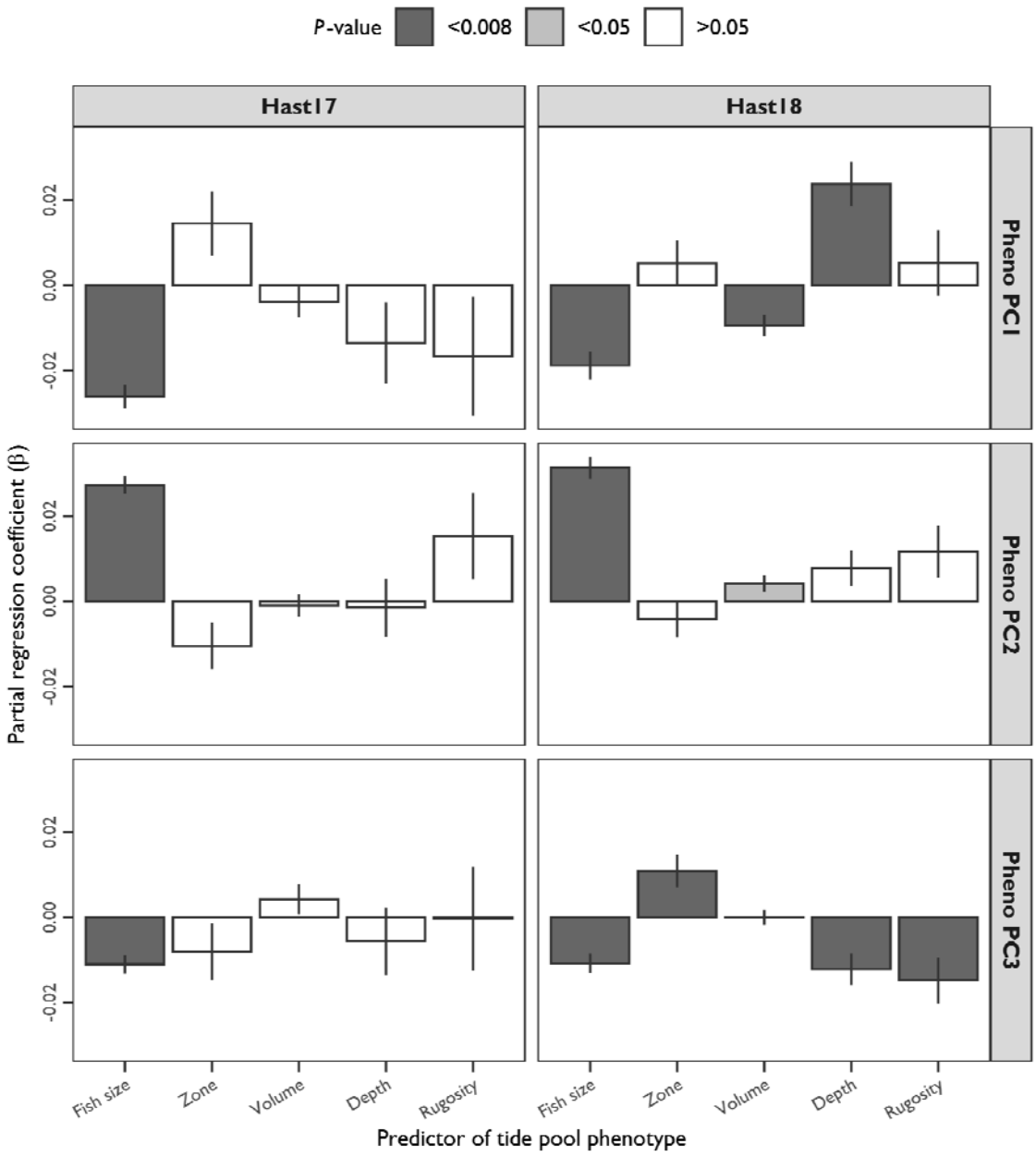

Figure 4. Partial regression coefficients for predictors of head shape phenotype in tide pools at Hastings Point. Predictors are organised on the *x*-axis, with estimated partial regression coefficients on the *y*-axis. Bars indicate the value of partial regression coefficient estimates with standard errors. Bar shading depicts the significance of each term as assessed through analysis of deviance at different *P*-value thresholds; note, <0.008 is the cutoff for multiple testing of 6 different models. Panels contain estimates for each phenotypic PC axis (rows) and population (columns). Also see Table 1 for full analysis of deviance tables, and Supplementary Table S3 for table of partial regression coefficient values.

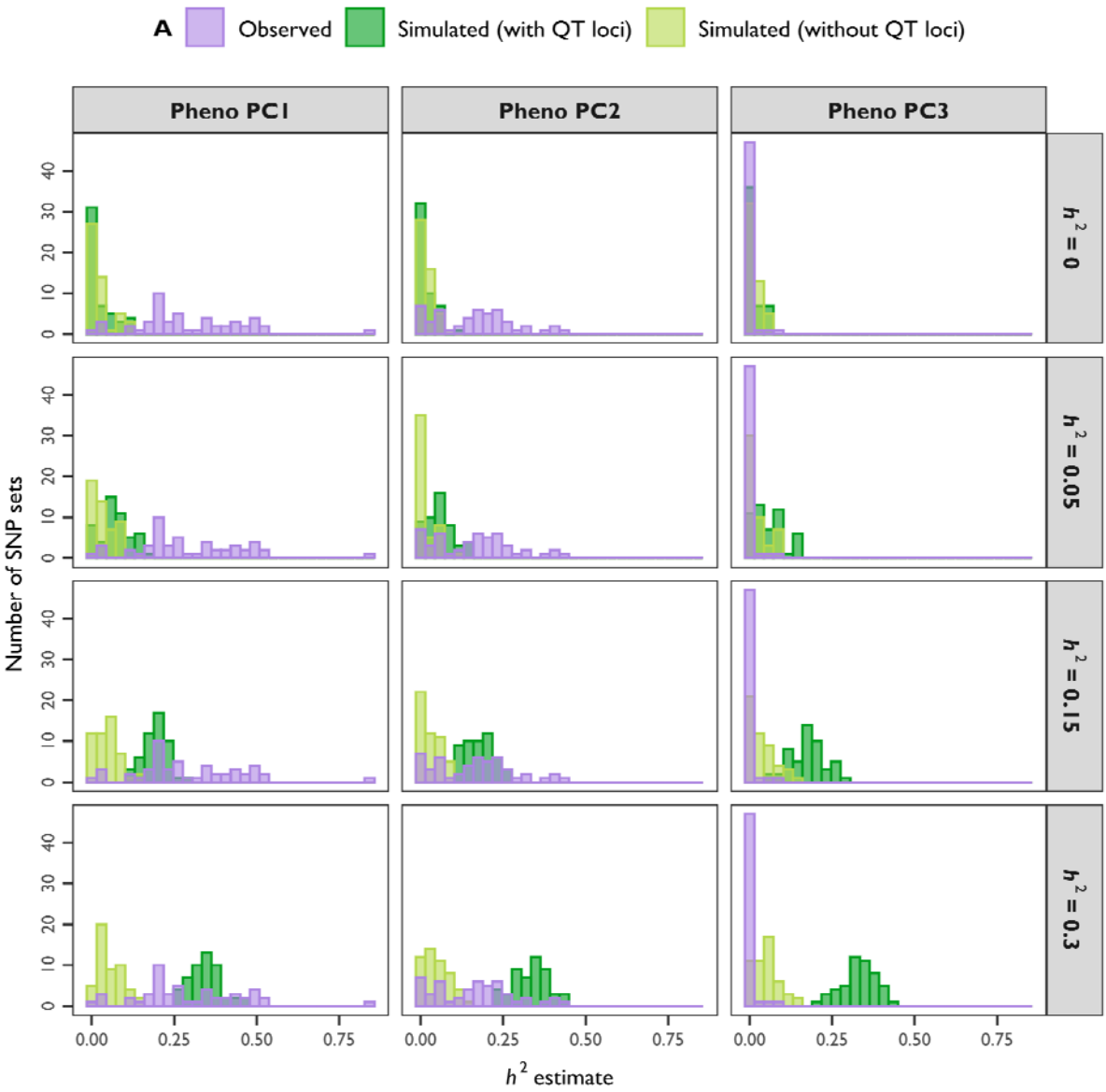

Figure 5. Frequency distributions of observed and simulated estimates of head shape phenotype heritability (h^2) over 50 SNP sets. The *x*-axis represents the estimated h^2 and the *y*-axis represents the number of SNP sets. Panels represent combinations of simulated h^2 values (rows: h^2 = 0, 0.05, 0.15, and 0.3) and head shape traits (columns: phenotypic PC1, PC2 and PC3). Frequency distributions are coloured based on whether they are derived from the observed dataset or simulated datasets with **A** estimated either with or without QT loci (see legend). For the observed datasets, values are plotted: the point estimated heritability for the ith trait and the jth SNP set. For the simulated datasets, values are plotted: the median estimated heritability across 50 QTL sets nested within the ith trait and the jth SNP set.

SUPPLEMENTARY RESULTS

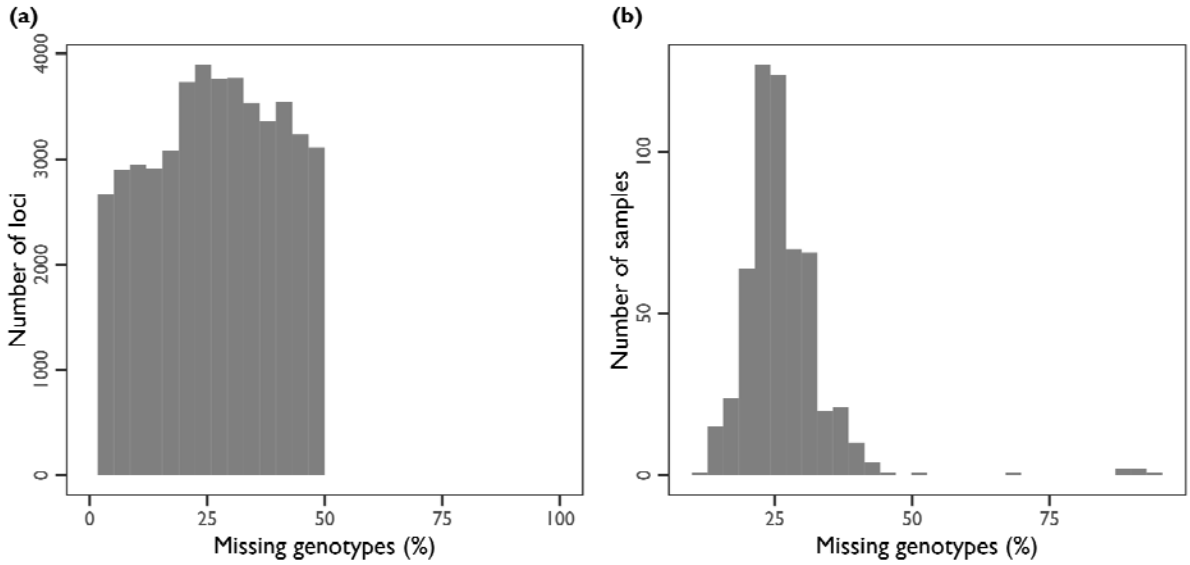

Figure S1. Missing data distributions following a preliminary filtering of SNP genotypes. The preliminary filtering included a minor allele count of 3, and missing data threshold of 50%. (a) Frequency distribution of the percent of missing genotype across SNPs. (b) Frequency distribution of the percent of missing genotypes across samples.

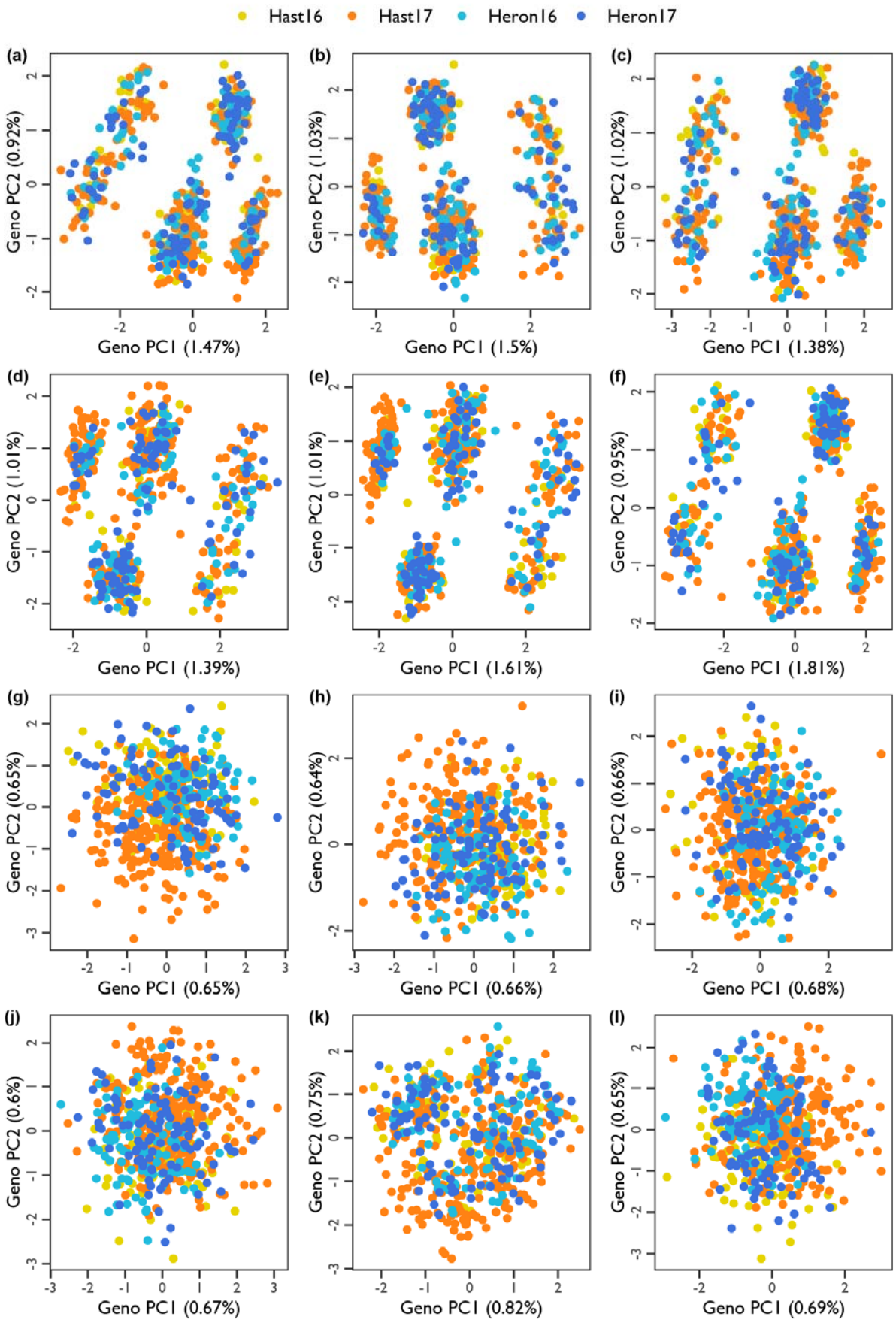
Hast16
Hast17
Heron16
Heron17
(a)
Geno PC2 (0.92%)
Geno PC1 (1.47%)
(b)
Geno PC2 (1.03%)
Geno PC1 (1.5%)
(c)
Geno PC2 (1.02%)
Geno PC1 (1.38%)
(d)
Geno PC2 (1.01%)
Geno PC1 (1.39%)
(e)
Geno PC2 (1.01%)
Geno PC1 (1.61%)
(f)
Geno PC2 (0.95%)
Geno PC1 (1.81%)
(g)
Geno PC2 (0.65%)
Geno PC1 (0.65%)
(h)
Geno PC2 (0.64%)
Geno PC1 (0.66%)
(i)
Geno PC2 (0.66%)
Geno PC1 (0.68%)
(j)
Geno PC2 (0.6%)
Geno PC1 (0.67%)
(k)
Geno PC2 (0.75%)
Geno PC1 (0.82%)
(l)
Geno PC2 (0.65%)
Geno PC1 (0.69%)

Figure S2. Principal components analysis (PCA) of SNP sets before and after removal of loci in linkage disequilibrium. Each plot depicts the projection of samples in genotypic PC1 (x-axis) and PC2 (y-axis) space, with percentage of explained variance in parentheses. Points are coloured by population (see legend). Results for six different SNP sets are depicted here. (a–f) SNP sets 1 to 6 before removing loci in linkage disequilibrium. There is evidence of an inversion captured on PC1 based on the three distinct clusters that might comprise one of 3 different haplotypes. Within these inversion haplotypes, there may also be additional haplotypic variation, as evidenced by further separation along PC2. (g–l) SNP sets 1 to 6 after removing loci in linkage disequilibrium. Samples have collapsed largely into single data clouds.

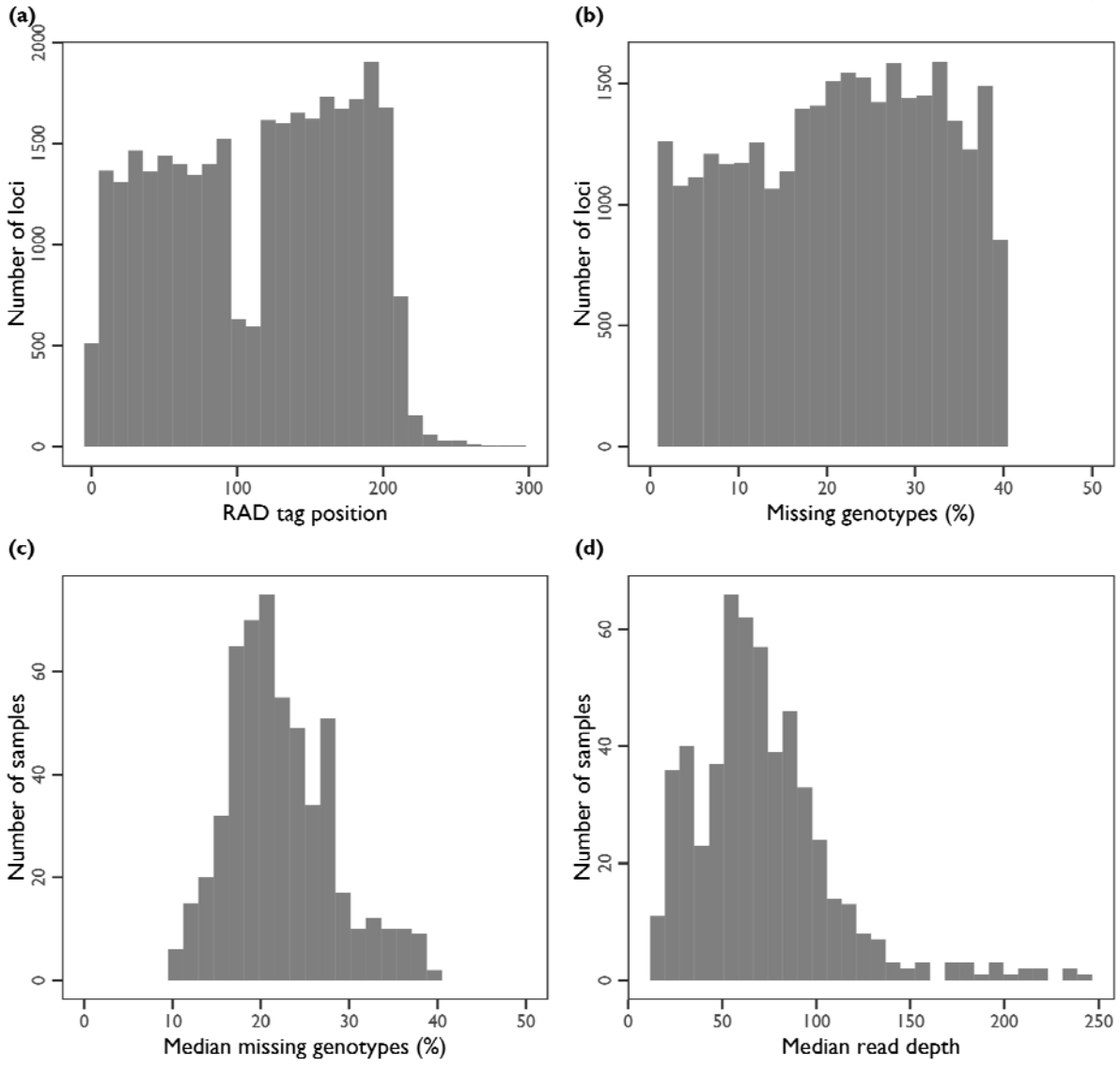

Figure S3. Summary statistics for our final 50 working SNP sets. A total of 30,260 SNP loci across 4,404 RAD tags were distributed across these SNP sets. A mean of 4,031 independent SNP loci (range of 3,877–4,131 SNP loci) were present in each SNP set. (a) Frequency distribution of SNP positions in RAD tags across all 30,260 SNP loci. The expected size of each RAD tag is roughly 200 bp, based on a *de novo* assembly from paired end reads trimmed to 100 bp in length. (b) Frequency distribution of the percent of imputed genotypes across 30,260 SNP loci. (c) Frequency distribution of the median percent of missing genotypes across samples, across SNP sets. (d) Frequency distribution of the median read depth across samples, across SNP sets.

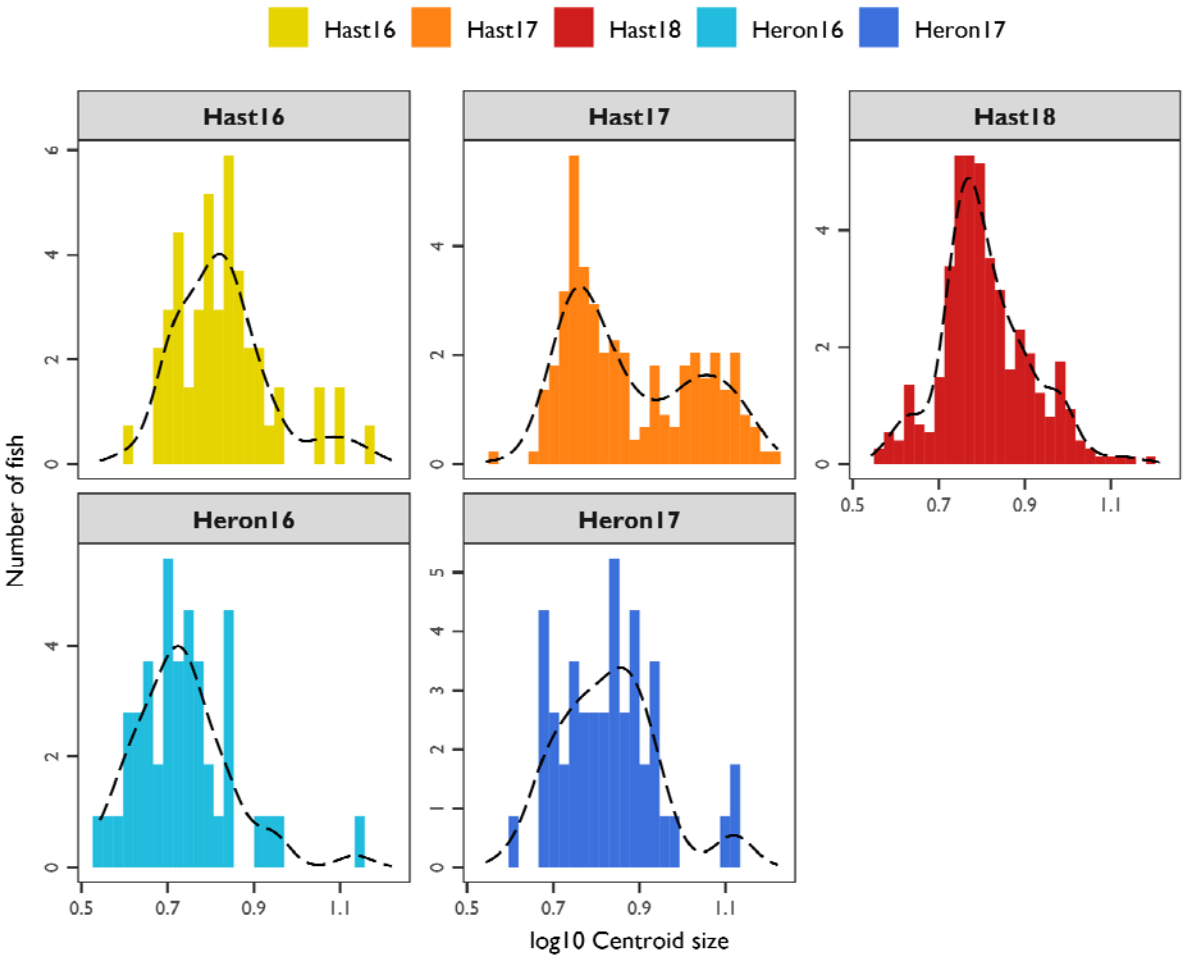

Figure S4. Size distributions of collected fish used in phenotypic analyses. The *x*-axis represents fish size measured as the log10 centroid size, derived from general Procrustes alignment of head shape morphological landmarks. The *y*-axis represents the number of fish. Bars depict the observed frequency distribution, whereas dashed lines depict the inferred probability density distribution. Panels separate populations and note that each panel has a different scale on the y-axis. Sample sizes for the populations are: Hast16 = 58, Hast17 = 189, Hast18 = 316, Heron16 = 46, and Heron17 = 49.

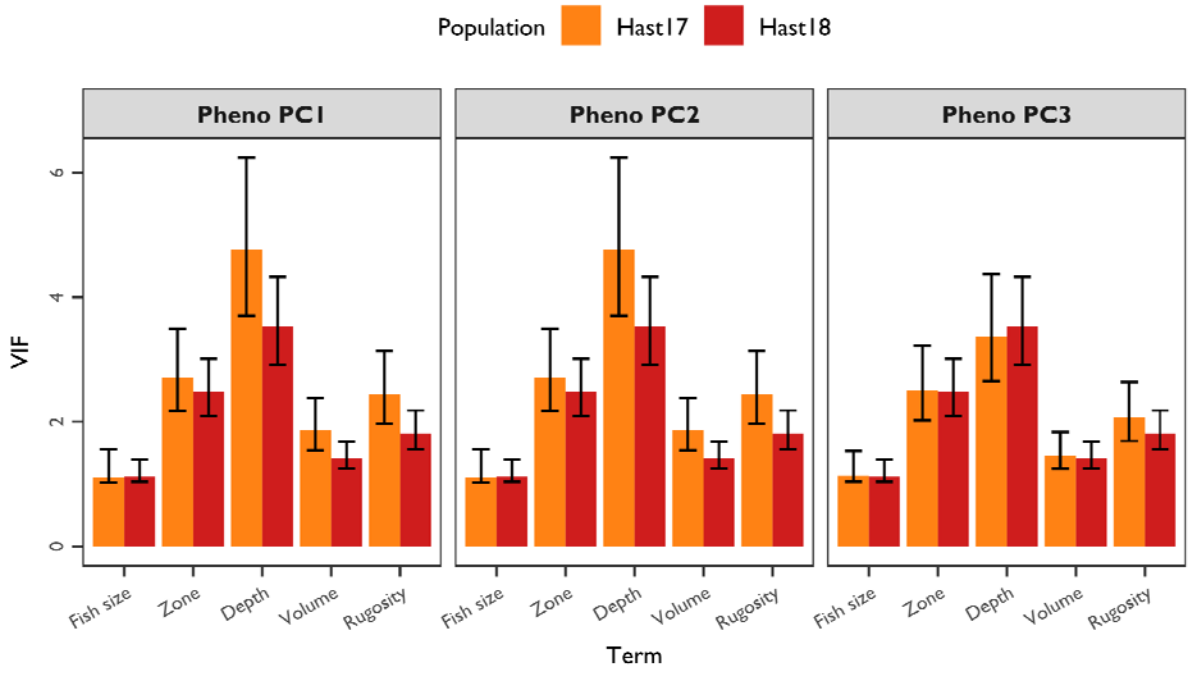

Figure S5. Variance inflation factors (VIFs) for models fitting mean head shape phenotype as a function of mean fish size and tide pool microhabitats. The *x*-axis represents the model terms, and the *y*-axis represents the VIF score. Bars represent the point estimates of VIFs, coloured by population (see legend). Error bars denote the 95% confidence intervals. Panels separate head shape traits.

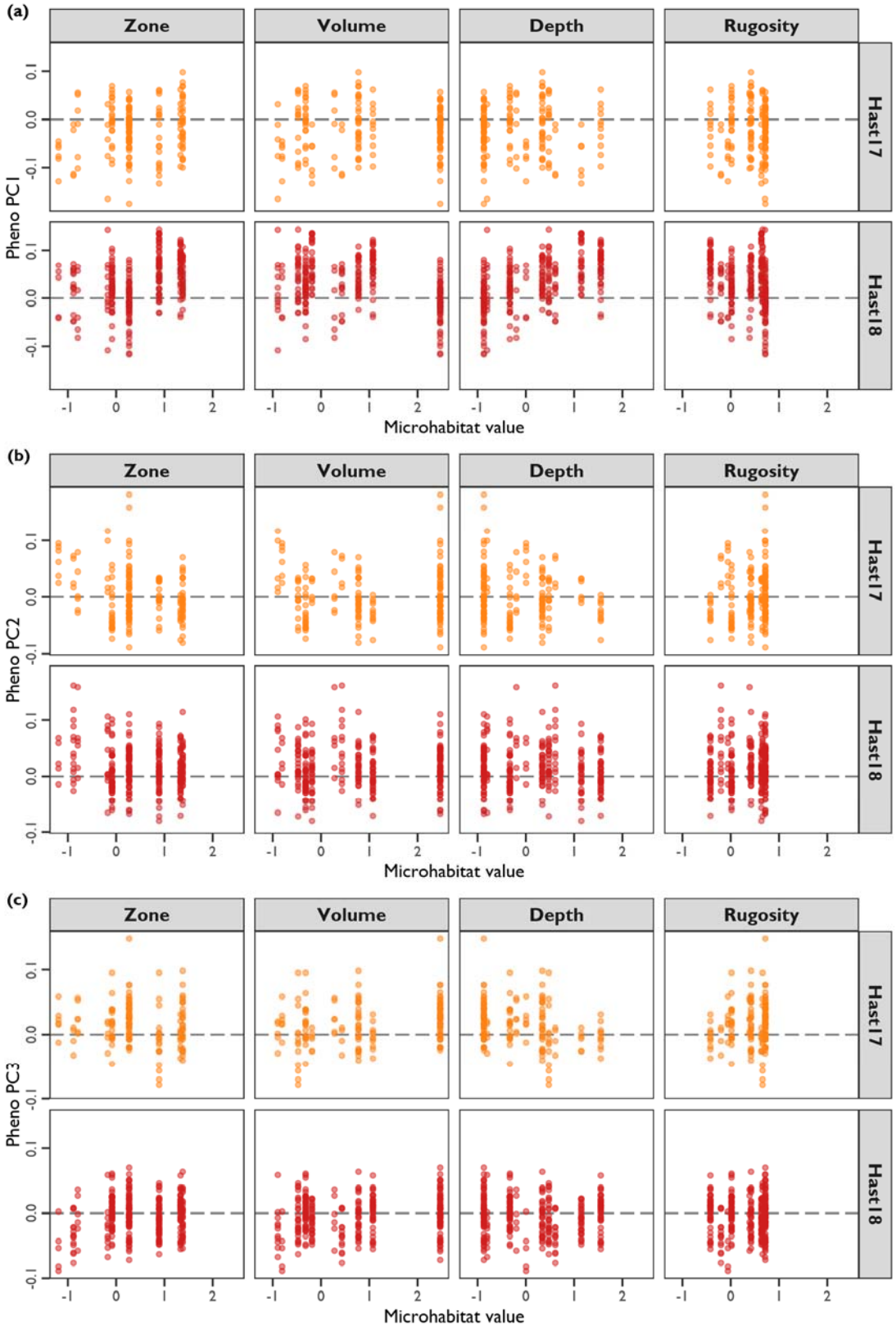

Figure S6. Associations between *B. cocosensis* head shape traits and tidepool microhabitat variables at Hastings Point. The *x*-axis represents the scaled and centered values of the

respective microhabitat variables. The *y*-axis represents the phenotypic (a) PC 1, (b) PC2, and (c) PC3. Panels represent combinations of population (rows: Hastings Point 2017 and 2018) and microhabitat (columns Zone, Volume, Depth, and Rugosity). The dashed grey line demarcates the value of zero for phenotypic traits.

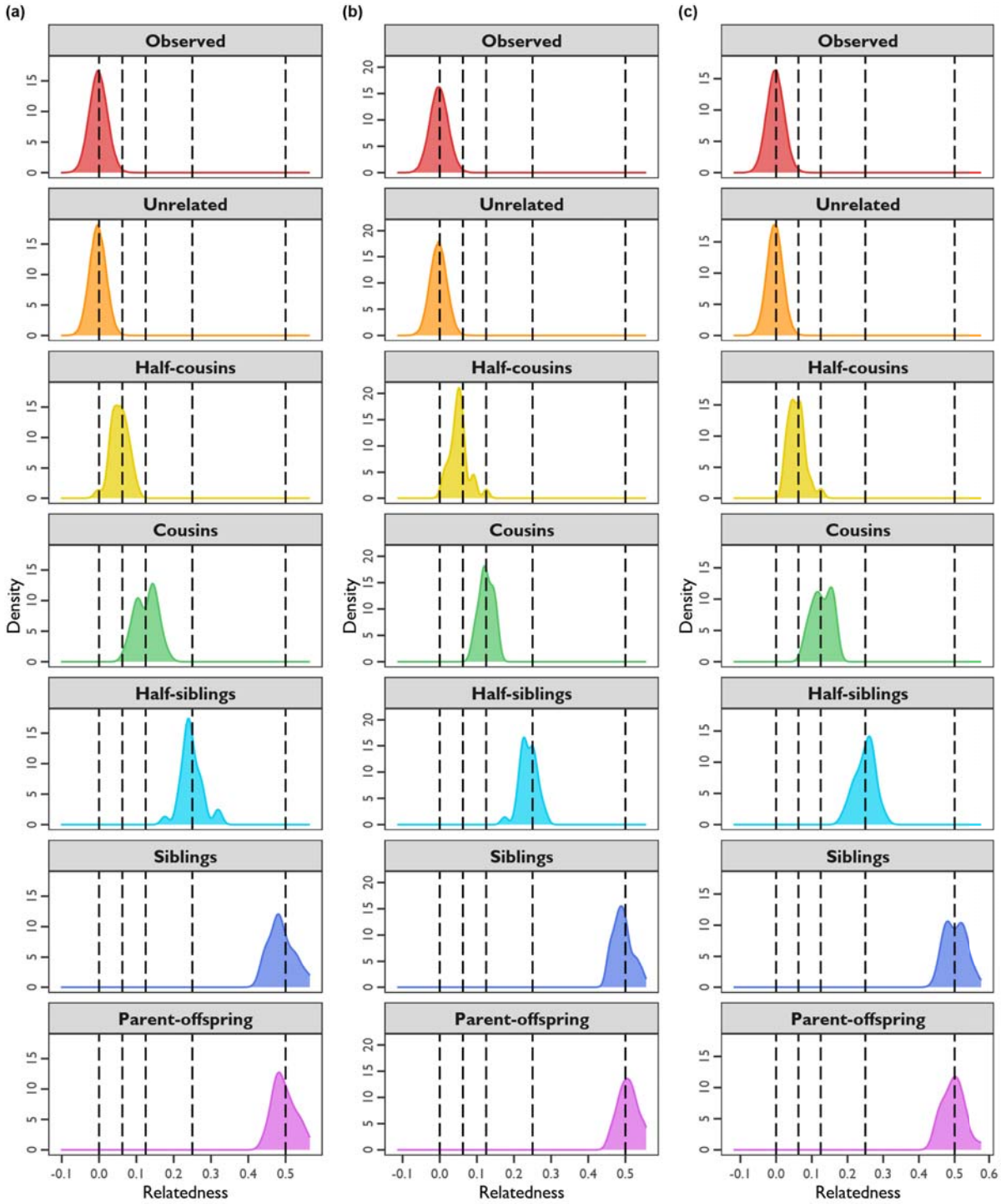

Figure S7. Simulated estimates of relatedness for close-kin relationships compared to observed estimates obtained from wild *B. cocosensis* in three representative SNP sets. The *x*-axis is the relatedness estimate, and the *y*-axis is the density. Curves represent the probability density distribution for relatedness estimates for different familial relationships (see legend). Dashed lines depict the theoretical relatedness, *r*, for unrelated individuals (r = 0), half-cousins (r = 0.0625), cousins (r = 0.125), half-siblings (r = 0.25) and siblings (r = 0.5), and

parent-offspring (r = 0.5). (a–c) Each panel depicts simulations for one of three representative SNP sets from our total 50 SNP sets.

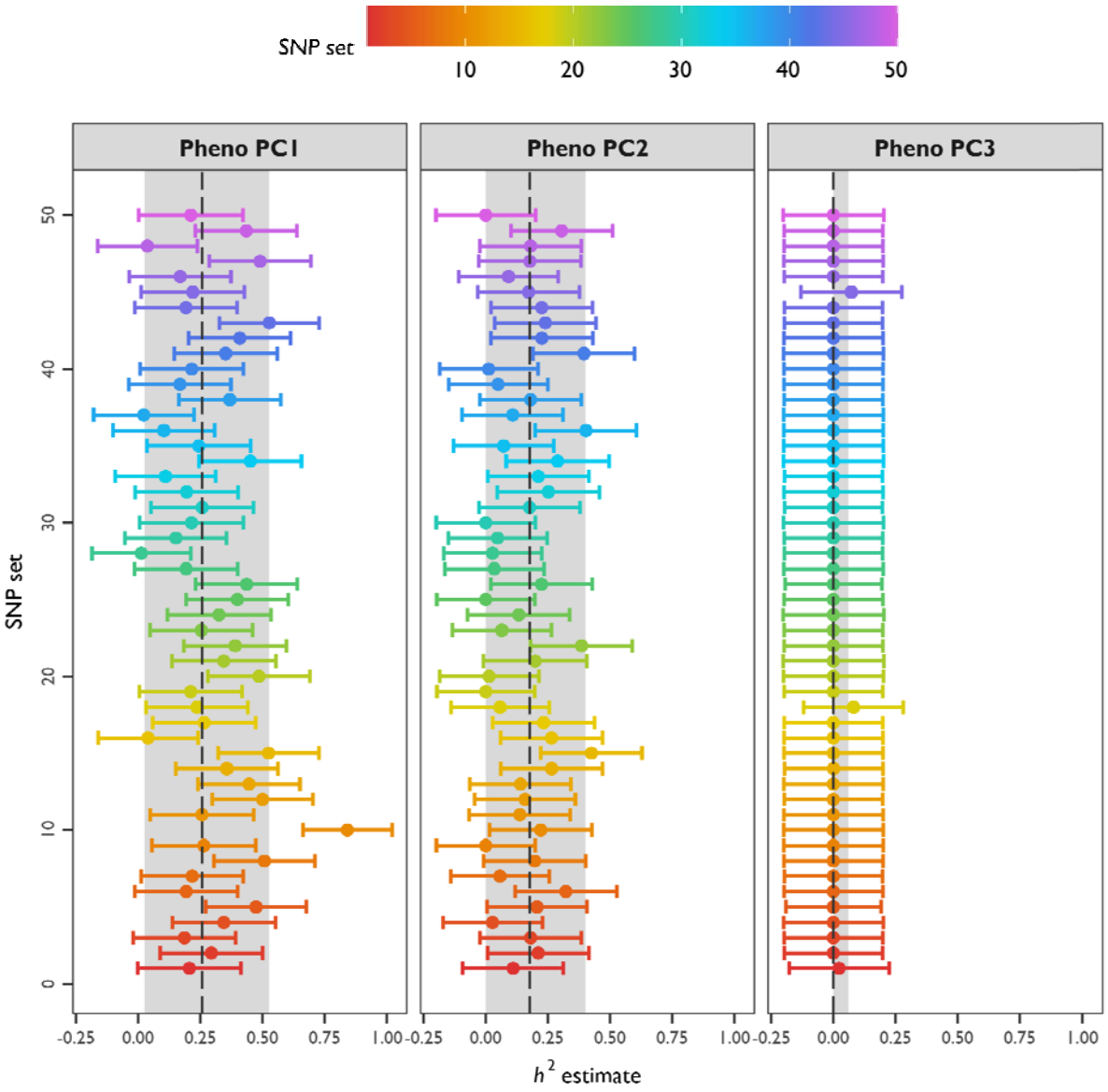

Figure S8. Observed heritability estimates for each SNP set and phenotypic PC. The *x*-axis represents the estimate of heritability derived from genomic animal models: . The *y*-axis represents the SNP set. SNP sets with the same *y*-axis position are directly comparable across panels. Panels contain data for each phenotypic PC. Points represent the point estimates of heritability for each SNP set with standard error bars. Dashed vertical lines demarcate the median of point heritability estimates across SNP sets. Grey vertical rectangles denote the 2.5% and 97.5% percentiles of point heritability estimates across SNP sets.

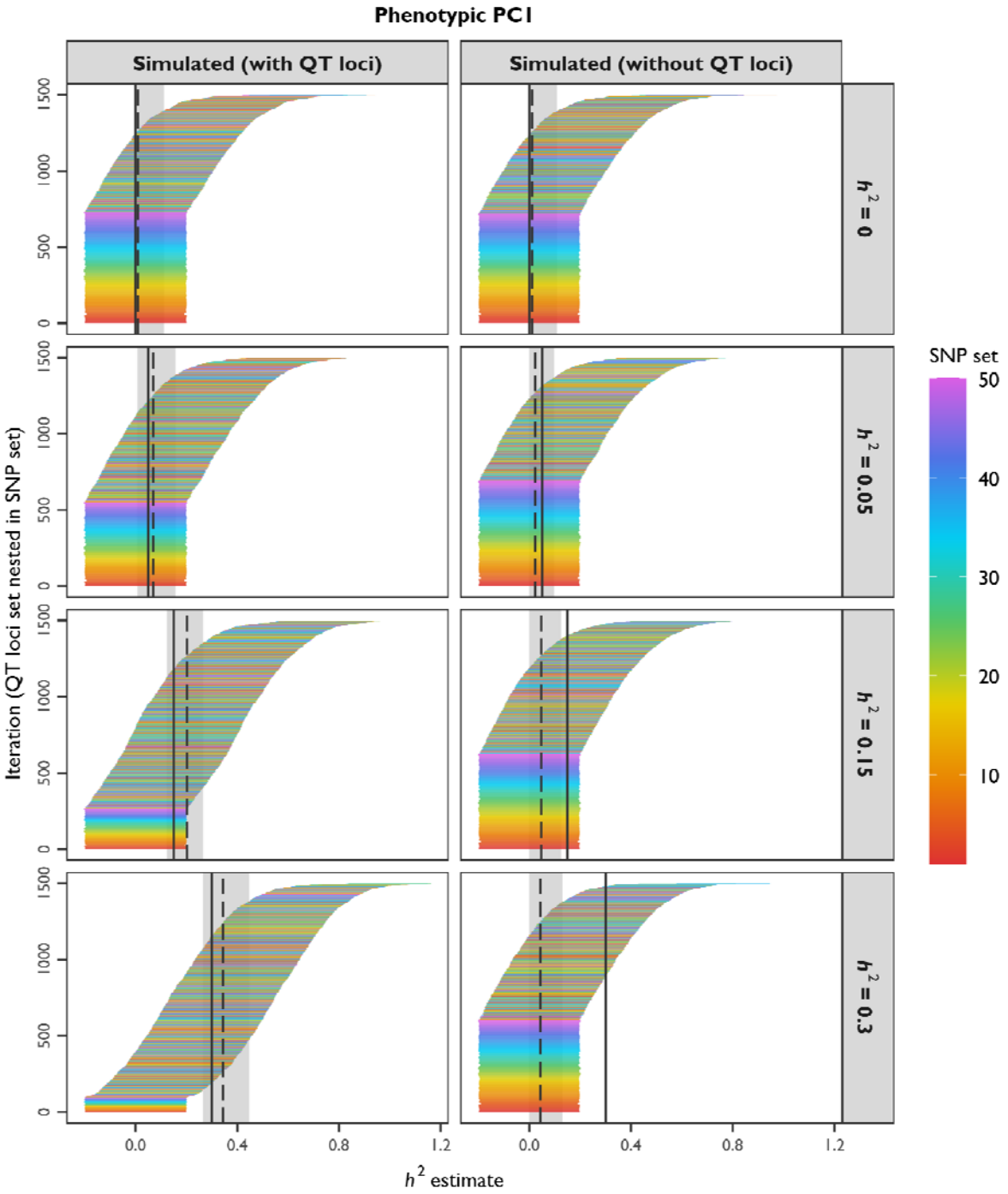

Figure S9. Simulated heritability estimates for each iteration of phenotypic PC1. The *x*-axis represents the heritability point estimates derived from genomic animal models fitted to simulated data: and _ . The *y*-axis represents the unique iterations: 30 QT loci sets nested in 50 SNP sets for a total of 1,500 iterations. Iterations have been ordered by effect size, so the identity of specific iterations is not consistent across panels. Panels contain data from different simulated heritability levels (rows) and genomic animal models fitted

with **A** constructed with and without QT loci (columns). Coloured horizontal lines represents the standard error of each iteration, with colour denoting the SNP set (see legend). Solid vertical lines demarcate the simulated heritability. The dashed vertical lines demarcate the median of medians of heritability point estimates: $\widetilde{h^2_{hi(N)}}$ and $\widetilde{h^2_{hi(N-M)}}$. Grey vertical rectangles denote the 2.5% and 97.5% percentiles for the medians of medians of heritability.

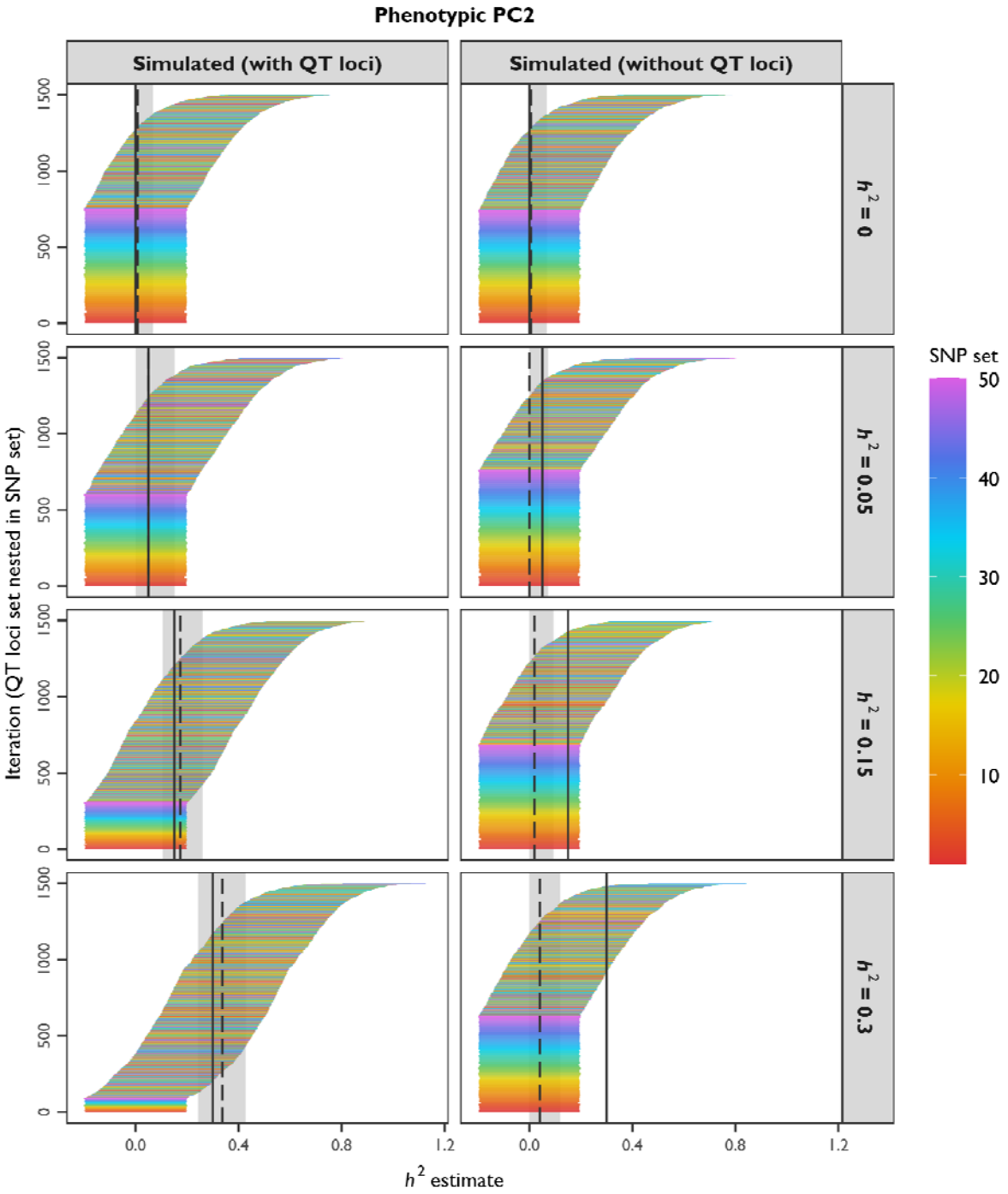

Figure S10. Simulated heritability estimates for each iteration of phenotypic PC2. The *x*-axis represents the heritability point estimates derived from genomic animal models fitted to simulated data: and _ . The *y*-axis represents the unique iterations: 30 QT loci sets nested in 50 SNP sets for a total of 1,500 iterations. Iterations have been ordered by effect size, so the identity of specific iterations is not consistent across panels. Panels contain data from different simulated heritability levels (rows) and genomic animal models fitted

with **A** constructed with and without QT loci (columns). Coloured horizontal lines represents the standard error of each iteration, with colour denoting the SNP set (see legend). Solid vertical lines demarcate the simulated heritability. The dashed vertical lines demarcate the median of medians of heritability point estimates: $\widetilde{h^2_{hi(N)}}$ and $\widetilde{h^2_{hi(N-M)}}$. Grey vertical rectangles denote the 2.5% and 97.5% percentiles for the medians of medians of heritability.

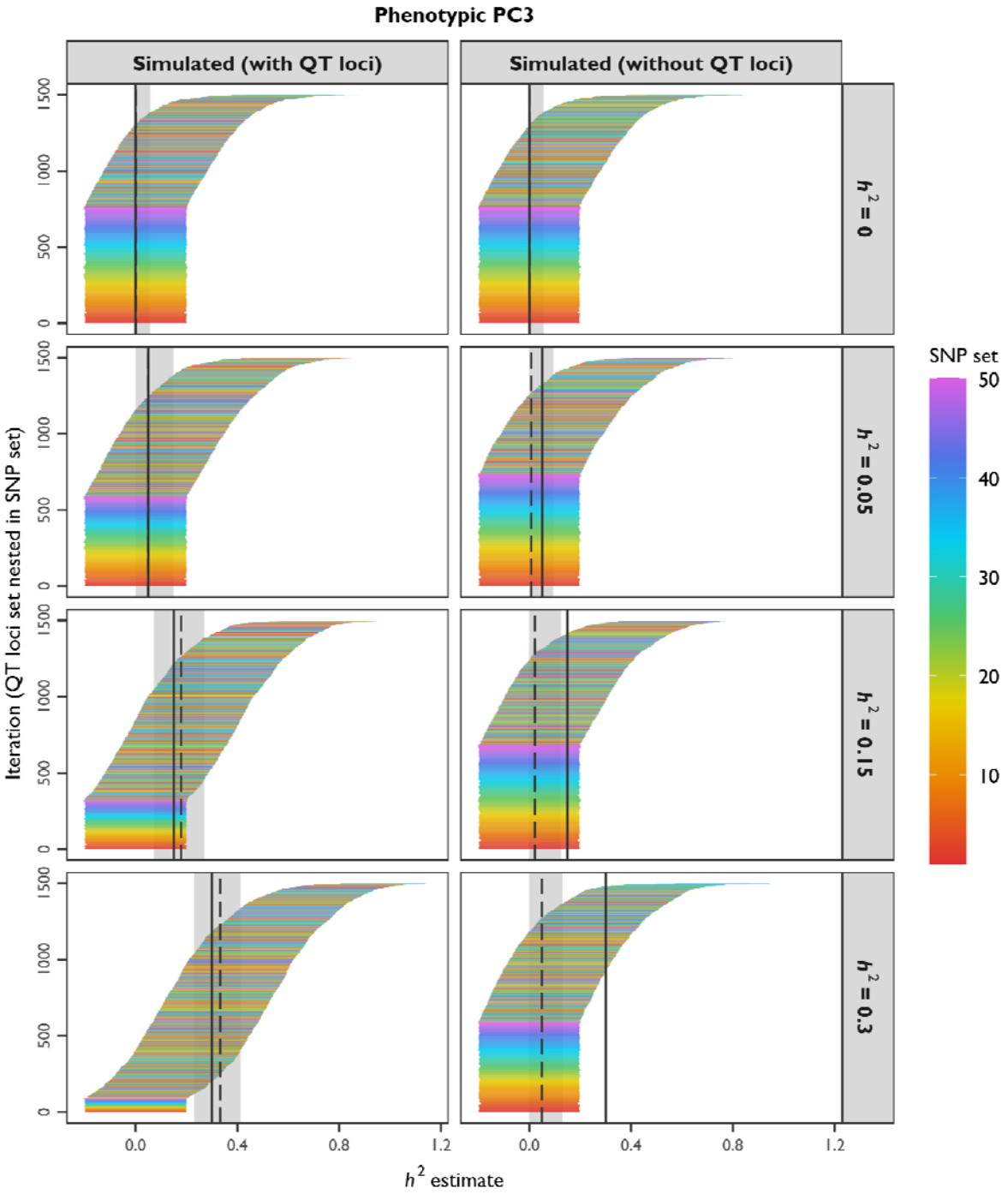

Figure S11. Simulated heritability estimates for each iteration of phenotypic PC3. The *x*-axis represents the heritability point estimates derived from genomic animal models fitted to simulated data: and _ . The *y*-axis represents the unique iterations: 30 QT loci sets nested in 50 SNP sets for a total of 1,500 iterations. Iterations have been ordered by effect size, so the identity of specific iterations is not consistent across panels. Panels contain data from different simulated heritability levels (rows) and genomic animal models fitted

with **A** constructed with and without QT loci (columns). Coloured horizontal lines represents the standard error of each iteration, with colour denoting the SNP set (see legend). Solid vertical lines demarcate the simulated heritability. The dashed vertical lines demarcate the median of medians of heritability point estimates: $\widetilde{h^2_{hi(N)}}$ and $\widetilde{h^2_{hi(N-M)}}$. Grey vertical rectangles denote the 2.5% and 97.5% percentiles for the medians of medians of heritability.

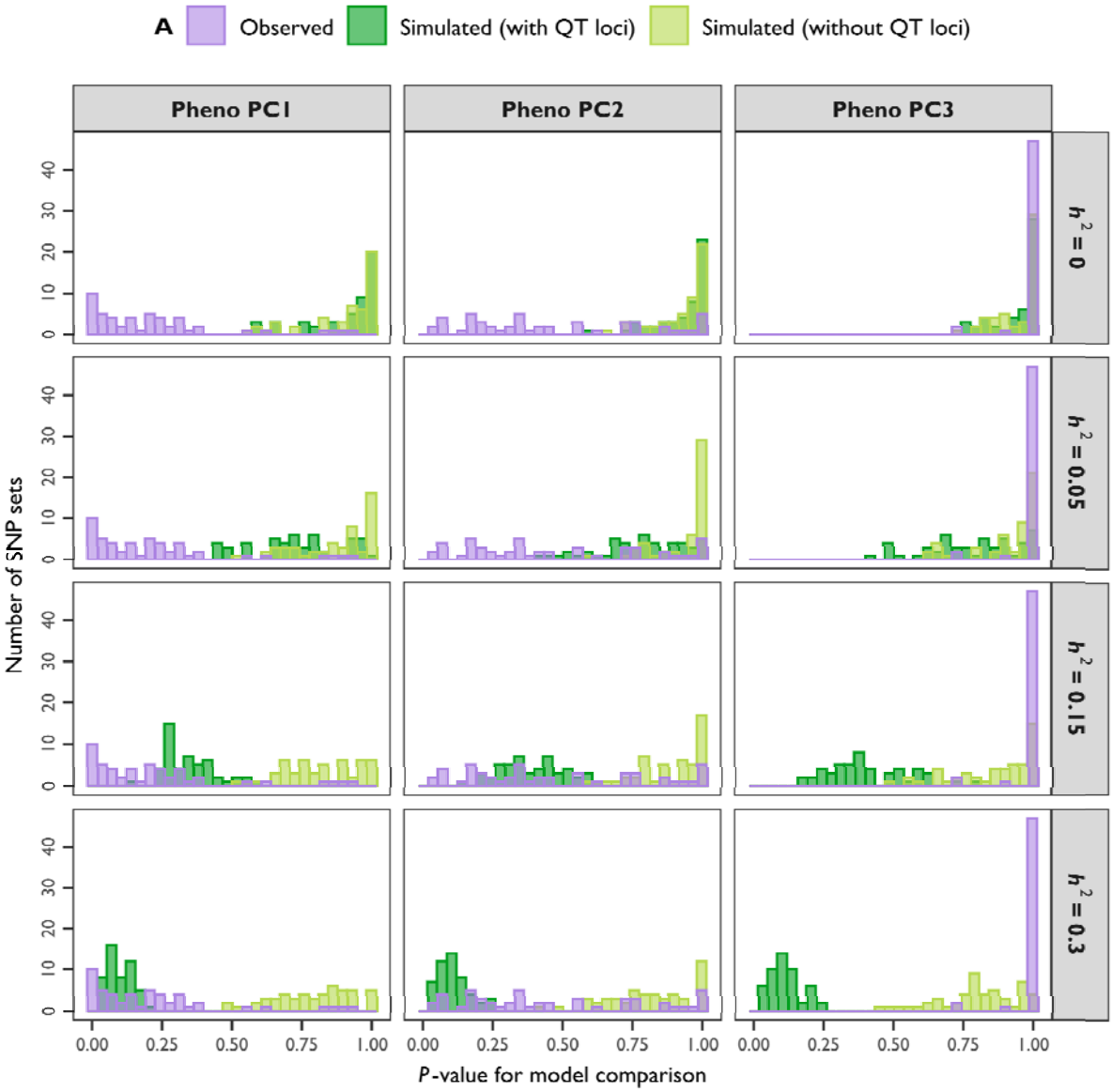

Figure S12. Frequency distributions of observed and simulated *P*-values of animal versus null model comparisons over 50 SNP sets. The *x*-axis represents the obtained *P*-values and the *y*-axis represents the number of SNP sets. Panels represent combinations of simulated h^2 values (rows: h^2 = 0, 0.05, 0.15, and 0.3) head shape traits (columns: phenotypic PC1, 2 and 3). Frequency distributions are coloured based on whether they are derived from simulated datasets with **A** with or without QT loci, or the observed dataset (see legend). For the observed datasets, values P_{ij} are plotted: the *P*-value for the ith trait and the jth SNP set. For the simulated datasets, values are plotted: the median *P*-value across 50 QT loci sets nested within the ith trait and the jth SNP set.

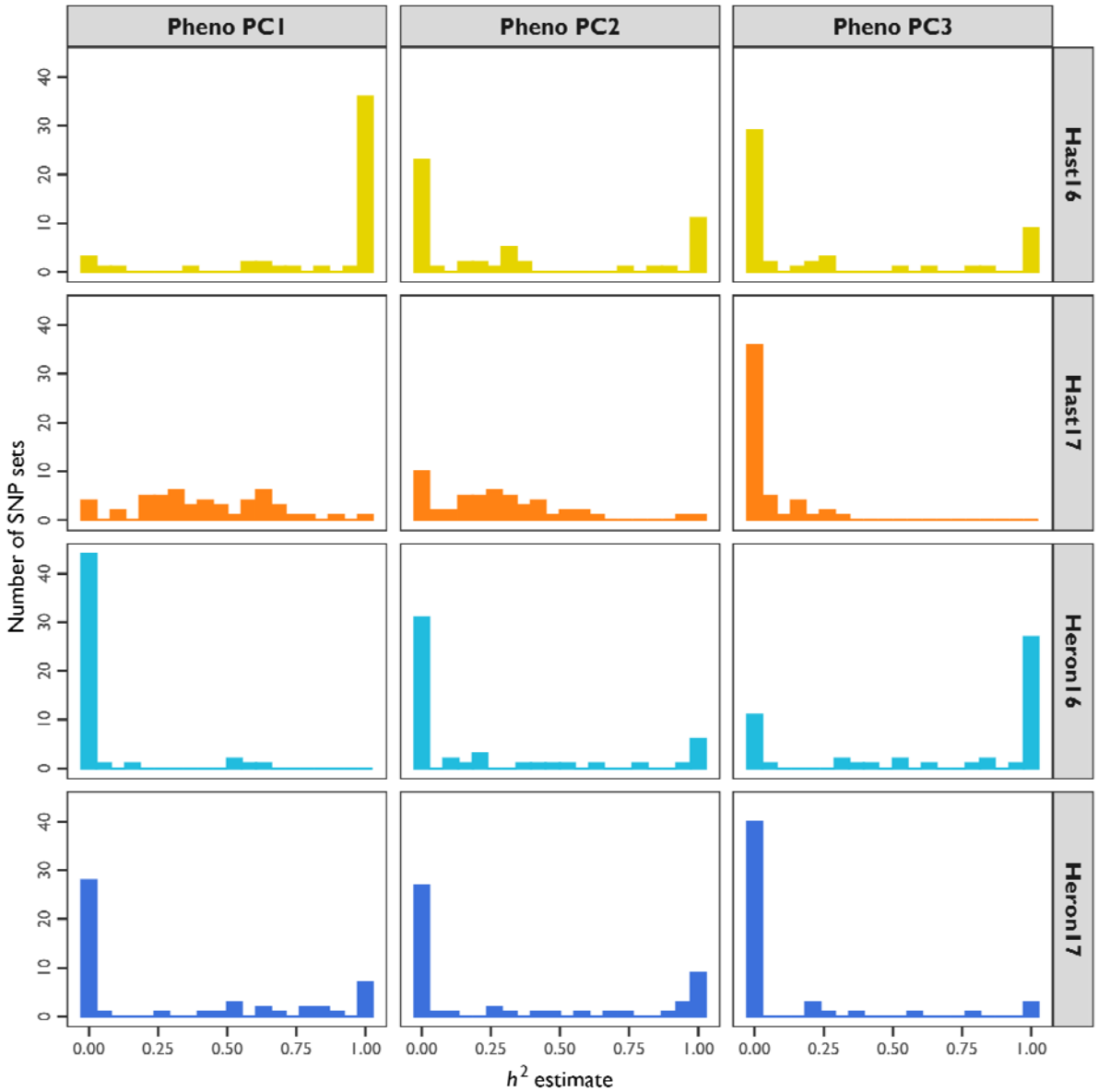

Figure S13. Frequency distributions of observed heritability from animal models fit within each population over 50 SNP sets. The *x*-axis represents the obtained h^2 values obtained for each population (rows: Hastings Point 2016, Hastings Point 2017, Heron Island 2016, and Heron Island 2017) and head shape traits (columns: phenotypic PC1, 2 and 3).

Table S1. Pairwise F_{ST}.

Pop1	Pop2	F_{ST}	P_{perm}
Hast16	Hast17	0.002	0.00
Hast16	Heron16	0.001	0.00
Hast16	Heron17	0.002	0.00
Hast17	Heron16	0.003	0.00
Hast17	Heron17	0.001	0.00
Heron16	Heron17	0.002	0.00

Table S2. Pairwise P_{ST}.

Pop1	Pop2	P_{ST}	P_{perm}
Hast16	Hast17	0.31	0.00
Hast16	Hast18	0.33	0.00
Hast16	Heron16	0.30	0.00
Hast16	Heron17	0.31	0.00
Hast17	Hast18	0.33	0.00
Hast17	Heron16	0.32	0.00
Hast17	Heron17	0.33	0.00
Hast18	Heron16	0.33	0.00
Hast18	Heron17	0.33	0.00
Heron16	Heron17	0.30	0.00

Table S3. Partial regression coefficients (*β*) and standard errors (SE) from mixed effect linear models of mean tide pool phenotype.

Population	Trait	Coefficient	*β*	SE
Hast17	PC1	Intercept	−0.010	0.006
		Fish size	−0.026	0.003
		Zone	0.015	0.008
		Volume	−0.004	0.004
		Depth	−0.013	0.009
		Rugosity	−0.017	0.014
	PC2	Intercept	0.025	0.004
		Fish size	−0.019	0.002
		Zone	0.005	0.005
		Volume	−0.009	0.003
		Depth	0.024	0.007
		Rugosity	0.005	0.010
	PC3	Intercept	−0.009	0.005
		Fish size	0.027	0.002
		Zone	−0.010	0.007
		Volume	−0.001	0.003
		Depth	−0.001	0.008
		Rugosity	0.015	0.012
Hast18	PC1	Intercept	0.016	0.004
		Fish size	0.031	0.003
		Zone	−0.004	0.005
		Volume	0.004	0.002
		Depth	0.008	0.005
		Rugosity	0.012	0.008
	PC2	Intercept	0.017	0.003
		Fish size	−0.011	0.003
		Zone	−0.008	0.004
		Volume	0.004	0.002
		Depth	−0.006	0.004
		Rugosity	0.000	0.006
	PC3	Intercept	−0.006	0.003
		Fish size	−0.011	0.002
		Zone	0.011	0.004

Volume	0.000	0.002
Depth	−0.012	0.004
Rugosity	−0.015	0.005

Table S4. Percentages of significant *P*-values in animal vs null model comparisons from the observed vs simulated data. The proportion of *P*-values was calculated across 50 SNP sets. Simulated estimates for each heritability level are represented with respect to whether **A** was constructed with vs without QTLs.

Trait	**Obs $Pct(P_{ij} < 0.05)$**	**Sim h^2 level**	**Sim A-matrix**	**Sim $Pct(\widetilde{P_{ij}} < 0.05)$**
Pheno PC1	0.3	0	Without QTLs	0
			With QTLs	0
		0.05	Without QTLs	0
			With QTLs	0
		0.15	Without QTLs	0
			With QTLs	0
		0.3	Without QTLs	0
			With QTLs	16%
Pheno PC2	0.04	0	Without QTLs	0
			With QTLs	0
		0.05	Without QTLs	0
			With QTLs	0
		0.15	Without QTLs	0
			With QTLs	0
		0.3	Without QTLs	0
			With QTLs	12%
Pheno PC3	0	0	Without QTLs	0
			With QTLs	0
		0.05	Without QTLs	0
			With QTLs	0
		0.15	Without QTLs	0
			With QTLs	0
		0.3	Without QTLs	0

With QTLs	10%	

SUPPLEMENTARY METHODS

Simulating a quantitative trait using *family_sim_qtl*

The *family_sim_qtl* function is available from R's *genomalicious* package (Thia 2024). This function simulates a quantitative trait using an iterative algorithm:

1. Specify a desired simulated additive variance, V_A^*
2. Determine which number of Q loci will be comprise the (QTLs) from N total loci.
3. Set the a = 1 and d = 0 initially: the locus-specific genetic effects for homozygotes and heterozygotes, respectively. It is assumed that all QTLs contribute equally to the trait.
4. For each mth QTL, calculate $V_{A(m)} = 2p_m q_m[a + d(q_m - p_m)]^2$. Because d = 0, this simplifies to $V_{A(m)} = 2p_m q_m[a]^2$.
5. We then obtain an initial estimate of the total additive variance by summing across all n QTLs, $V_A = \sum_{m=1}^{Q} V_{A(m)}$, but note that this will not be the desired simulated value.
6. To get the desired simulated additive variance, we must determine the rescaling factor to apply to each of the n QTLs. This scaling factor is simply V_A^* / V_A.
7. If $\vec{v}$ is a vector of the initial simulated V_A values for each of the n QTLs [e.g., $V_{A(1)}$, $V_{A(2)}$ ... $V_{A(n)}$], we can obtain their rescaled values as $\vec{v}^* = \vec{v} \times (V_A^* / V_A.)$, which is also a vector [$V_{A(1)}^*$, $V_{A(1)}^*$ ... $V_{A(1)}^*$]. If we sum these values, we will get the desired total additive variance, $V_A^* = \sum_{m=1}^{n} V_{A(m)}^*$.
8. Now that we have rescaled the locus-specific additive variances, we need to determine the rescaled genetic effects for each mth locus, a_m^*. We therefore rearrange

$V^*_{A(m)} = 2p_m q_m [a^*_m]^2$ to solve for $a^*_m = [V^*_{A(m)} / (2p_m q_m)]^{1/2}$. In principle, all values of a^*_m should be the same because they were all simulated with an initial value of $a = 1$, but if there is no additive genetic variance at the locus, $a^*_m = 0$.

9. Locus-specific breeding values are given by $g_m = x_{0(m)} a^*_m + x_{1(m)} d + -x_{2(m)} a^*_m$. Again, the effect of the heterozygous genotype is $d = 0$. The variables $x_{0(m)}$, $x_{1(m)}$, and $x_{2(m)}$, are dummy variables representing the reference allele homozygote, the heterozygote, and the alternate allele homozygote, respectively. These genotype dummy variables will take the value of 1 when true, and a value of 0 when false.
10. The final total breeding value is then the sum of all locus-specific breeding values, $g = \sum_{m=1}^{M} g_m$.
11. The phenotype for an individual, y, is then, $y = g + e$. Here, e is a randomly drawn environmental effect that is normally distributed following $N(0, V_E)$, and V_E is the desired environmental variance in the simulated trait.

The importance of this algorithm is that it generates phenotypes for each individual based on their multi-locus genotypes. This allowed us to maintain the exact additive relationship structure among observed individuals, but simulate different phenotypes based on some desired quantitative genetic properties (V_A, V_E and number of loci). This is useful in contexts where pedigrees cannot be easily generated, such as in our study where it is unlikely we sampled any close-kin in collection of *Bathygobius cocosensis*.

The schematic outlines the bootstrapping simulation framework used in this study. For simplicity, it depicts three individuals genotypes 4 RAD tags, each with 2 SNP loci. The bootstrapping component is achieved by generating J SNP sets, which randomly sample one SNP locus per RAD tag. For each jth SNP set, a total of K QTL sets are defined, with each kth QTL set containing M randomly chosen SNP loci from the total N SNP loci. In this schematic, $M = 2$. The simulation component is achieved by iterating over the K QTL sets to ultimately obtain a distribution of parameters or statistics, which are then summarised as the median for the jth SNP set. Simulated median values from each jth SNP set can then be aligned to observed values using the same j^{th} SNP set. Here, narrowsense heritability (h^2) and model comparison test (P-values) are shown.

In the schematic, three quantities are depicted in black boxes. These represent values specified by the researcher. Here, the additive genetic variance (V_A) and environmental variance (V_E) are set to a value of 1, and hence the total phenotypic variance (V_P) is 2, and narrowsense heritability of 0.5. Simulated genetic and environmental effects can be combined with covariates to shape the distribution of simulated phenotypes. These covariates ($\mathbf{X}b$) are represented by a design matrix and associated partial regression coefficients, which match that obtained from models on the observed data. Here, 'Ind 1' has a value of 2 added to its phenotype, whereas 'Ind 2' and 'Ind 3' have a value of 1 added to their phenotypes.

Custom ddRAD protocol

The protocol is designed for groups of 24 samples, and hence, can be scaled to a 96 well plate to maximize efficiency. Each sample is uniquely barcoded, and groups of 24 samples are pooled onto a single index.

Start with 40 µL 200ng DNA, normalised with a Qubit or Picogreen.

Double Digest

Add 10ul of the digestion mixture to the 40 µL of DNA: total of 50 µL.

Using a PCR-free thermocycler, run the digestion for: (1) 3 hours at 37°C, then (2) 20 minutes at 65°C.

Proceed directly to ligation (if you stop, the world will end!).

DIGESTION REACTION MIXTURE

Reagent	**×1**	**×25**
*Sph*I	0.5	12.5
*EcoR*1	0.5	12.5
Cutsmart Buffer	5.0	125.0
Water	4.0	100.0
TOTAL	10.0	250.0

Ligation

Add 1 µL of P1 barcode adaptor (at ~10 ng/µL) to each sample.

LIGATION REACTION MIXTURE

Reagent	**×1**	**×25**
P1 adapter (10 ng/µL) (barcode)		
P2 adapter (40 µM) (barcode)	1.0	25.0

10 mM ATP	6.5	162.5
T4 DNA Ligase	0.5	12.5
10× Cutsmart buffer	1.5	37.5
Water	4.5	12.5
TOTAL	14.0	250.0

Add 14 µL ligation reaction mixture to each sample. Therefore, the new sample volume is: 50 + 1 + 14 = 64 µL.

Using a PCR-free thermocycler, run the ligation for: (1) 3 hours at 22°C, then (2) 20 minutes at 65°C.

It is safe to pause the protocol here.

Pooling Samples

24 samples (each individually barcoded) can be pooled together. For a 96 well plate, this will result in 4 libraries. Each library is given a unique index. Filter columns are used for cleaning the ligates and pooling them. Each filter column will represent a pool of 24 samples:

1) In each sample, add equal volume of binding buffer (~65 µL) and equal volume of absolute ethanol (~65 µL)
2) Transfer the mix into filter column. Try to add 4 samples at a time, and then spin at high speed for 1 minute. Then discard the flow through. Repeat this step until all 24 samples are loaded into the column.
3) Add 500 µL wash buffer then spin at high speed for 1 min. Discard the flow-through. Repeat this wash step 2 times.
4) Place the filter column into a new tube and then spin for 2 minutes.
5) Place the filter column into a new clean tube again, and then add 40 µL elution buffer. Incubate at room temperature for 2 minutes (or more). Spin at high speed for 1 minute.

The 40 µL elution will be subjected to size selection.

It is safe to pause the protocol here.

Size Selection

Size select the cleaned ligates using Pippin Prep. A BluePippin cartridge has 5 lanes. Use one lane per index—so 4 lanes can be used for 96 samples to be efficient. The last lane can be used for the final size clean up. Size selection should be 420–700 bp (avoids overlap of forward and reverse reads, which are ~150 bp).

Size selection yields 40 µL size-selected ligate. Separate into 2 tubes containing 20 µL each.

It is safe to pause the protocol here.

PCR for Library Enrichment

For a single library, 2 × 20 µL aliquots are recovered from the Pippin prep. To each separately, add 7.6 µL of PCR reaction mixture and 1.2 µL each of forward and reverse primers (0.40 µM).

Therefore, the final volume for a PCR reaction should be: 20 + 7.6 + 1.2 + 1.2 = 30 µL.

In a thermocycler, run the PCR for: (1) 2 minutes 98°C; (2) 15 cycles of 30 seconds at 98°C, 20 seconds at 62°C, 30 seconds at 72°C; and (3) 5 minutes at 72°C.

It is safe to pause the protocol here.

PCR REACTION MIXTURE

Reagent	**×1**	**×2**
Forward primer (index)		
Reverse primer (index)		
Phusion 5× HF buffer	6.00	12.00
dNTP	0.60	1.20
Taq	0.24	0.48

Water	0.76	1.52
TOTAL	7.60	15.20

PCR Quality Check and Index Pooling

Check PCR products on Shimadzu (5 µL of the PCR product).

Check the concentration of the PCR product (PCR DNA concentration) on Qubit (1uL of the PCR product).

From Shimadzu calculate the proportion that of DNA that is library:

Library proportion = Library peak/(Library peak + Other peaks)

To get the library concentration:

Library concentration = PCR DNA concentration × Library proportion

Make equimolar pool of indices using the library concentration. Use the index with the smallest library concentration as the baseline. Ampure to 40ul (if larger), size select 350–700 on Pippin Prep, Ampure to 11 µL in 10 mM Tris, then check the concentration on a Qubit.

SNP filtering

Step 1. Pre-filtering of raw SNPs

Raw SNPs produced by *freebayes* were pre-filtered with *vcftools*:

```
vcftools --mac 2 --max-missing 0.5 --minGQ 20
```

These pre-filtered SNPs were imported into R with *genomalicious* for further filtering.

Step 2. Sample-level missingness, mitochondrial RAD tags, and imputation

Any RAD tags that mapped to the *Bathygobius cocosensis* mitochondrion were removed. Samples with >40% missing genotypes were removed: such samples contribute disproportionately to missing genotypes per locus. We used *polyRAD* v2.0.0 (Clark et al. 2019) to impute missing genotypes via Bayesian inference. Here, we the genotype dose as a more continuous description of genotype estimation and certainty: 0 indicates no alternate alleles (reference homozygote), 0.5 indicates a heterozygote, and 1 indicates two alternate alleles (alternate homozygote).

Step 3. Single SNP loci per RAD tag, locus-level missingness, and heterozygosity excess

We reduced dataset to a single SNP locus per RAD tag, whilst simultaneously aiming to reduce missingness. For each RAD tag, we identified those SNPs that were observed in a minimum of five individuals. From there, we identified all SNPs that had the minimal level of imputation (percent of missing genotypes). If only a single SNP remained, that SNP was used to represent the RAD tag. If multiple SNPs remained (tied for missingness), we randomly selected one SNP to represent the RAD tag. Any SNPs with ≥40% missing genotypes were removed.

SNPs with an excess of heterozygosity were then removed. Excess of heterozygosity may arise when paralogous genomic regions, or genomic regions with high similarity, are artefactually merged into a single RAD tag. The observed and expected heterozygosity was estimated per SNP locus. Deviations from the expected heterozygosity were tested using R's *pchisq* function with 1 DF and a two-tailed test. SNPs were removed when the observed heterozygosity was significantly larger than expectations ($P < 0.05$).

Step 4. Linkage

As a final filtering step, we analysed patterns of linkage disequilibrium among loci to remove non-independent SNPs. The *LDna* R package (Kemppainen et al. 2015) was used to construct linkage networks among SNPs using Pearson's correlation of genotypes. SNP linkage clusters were merged if they were connected by a minimum of 20 edges and if they had a correlation of 0.8. A PCA of SNP genotypes revealed that the first two PC axes captured several linkage groups in our dataset (Supplementary Figures S2). Most of the highest loading SNPs in the PCA came from a single linkage cluster identified by *LDna* (Supplementary Figures S3). One SNP per linkage cluster was selected to be included with non-linked SNPs in our final dataset.

P_{ST} calculations

We calculated global and pairwise P_{ST} using the multivariate method described in Chenoweth and Blows (2008), originally outlined in Kremer et al. (1997).

$$P_{ST} = \frac{\sum trace\{(\mathbf{B} + 2\mathbf{W})^{-1/2} \times \mathbf{B} \times (\mathbf{B} + 2\mathbf{W})^{-1/2}\}}{n}$$

In this formulation, *trace*{} is the diagonal of a matrix, and n is the number of associated eigenvalues (number of traits). **B** and **W** are the between and within subpopulation variance-covariance matrices, respectively, and were derived from a multivariate analysis of variance (MANOVA) fitting:

$$\mathbf{Y} = \mu + \text{Population} + \varepsilon$$

Here, **Y** is a matrix of traits (phenotypic axes PC1 to PC3), μ is the mean, 'Population' is a categorical factor for sampled population, and ε is the residuals. R's *manova* function was used to fit the model. The *summary* function was used to extract the sums-of-squares and cross-products matrices for the population and residual effects, which were divided by their degrees of freedom to obtain **B** and **W**.