A novel framework for assessing the power of genomic animal models in wild non-model organisms

Joshua A. Thia[1,2,A], Cynthia Riginos[2], James Hereward[2], Libby Liggins[3], Moshe E Jasper[1], J. David Aguirre[3,B], Katrina McGuigan[2,B]

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

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

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

[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
Cynthia Riginos	https://orcid.org/0000-0002-5485-4197
James Hereward	https://orcid.org/0000-0002-6468-6342

Libby Liggins	https://orcid.org/0000-0003-1143-2346
Moshe E. Jasper	https://orcid.org/0000-0003-4541-3223
J David Aguirre	https://orcid.org/0000-0001-7520-441X
Katrina McGuigan	https://orcid.org/0000-0002-0525-2202

RUNNING TITLE

Genomic animal models for wild organisms

KEYWORDS

Genetic architecture, genomic animal model, marine intertidal fish, heritability, quantitative genomics, QTL effect sizes, simulations

ABSTRACT

1. The relative contributions of heritable versus non-heritable phenotypic variation determine how selection may shape evolutionary responses. While genome-wide association studies can provide some insight into large-effect loci underpinning phenotypes, they potentially miss much of the evolutionarily important variation in polygenic traits because many small-effect loci will contribute to heritable variation.
2. Genomic animal models can estimate trait heritability from genomic data but are often difficult to apply to wild organisms. Sampling groups of close kin

improves their power, yet this can be logistically challenging in species with large populations and high dispersal. Furthermore, many common approaches for power simulations rely on kin structure in the genomic relationship matrix, **A**, but such simulations on an unstructured **A** (comprised of unrelated individuals) are likely uninformative.

3. We developed a novel framework to generate simulated expectations from genomic animal models when applied to samples of unrelated wild organisms. A key innovation of this framework is the use of a 'phenotypes-from-genotypes' approach that allows quantitative traits to be simulated from genotypes, independently of **A**. This provides clear utility where kin structure is absent. We illustrated the use and flexibility of our phenotypes-from-genotypes simulation approach to explore how the performance of genomic animal models can be influenced by different genetic architectures (clumped vs even genomic distribution of QTLs, and different QTL effect sizes), levels of heritability, and experimental design (sampled individuals and genetic markers).
4. We then applied our framework to a study of head shape traits in a wild marine intertidal fish, *Bathygobius cocosensis* (Bleeker 1854), for which we acquired SNP genotypes with RADseq. Although our working sample of 297 fish contained no close kin, our framework helped validate the power of genomic animal models, allowing us to confidently infer heritable variation in one head shape trait (estimated h^2 = 0.17).

5. 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. Our novel 'phenotypes-from-genotypes' approach also provides a new method for simulating QTLs using observed population genetic data.

INTRODUCTION

Genome-wide association studies have greatly advanced our ability to study the genetic component of adaptive traits by identifying associations between genotypes and phenotypes (Lasky et al., 2022; Santure & Garant, 2018). However, genome-wide associations can be underpowered when traits are polygenic, with each locus having only a small effect on the trait (Lowry et al., 2017; Meirmans, 2015; Duntsch et al., 2020; Lasky et al., 2022; Santure & Garant, 2018). An attractive alternative is the genomic animal model, which combines traditional quantitative genetics with genomic estimates of shared allelic variation between individuals (Bérénos et al., 2014; Kruuk, 2004; Richards et al., 2023; Schmidt et al., 2023; Wilson et al., 2010). This approach can overcome the limitation of genome-wide associations because it estimates the cumulative contribution from all loci rather than seeking to characterise the effect of each individual locus. Despite this advantage, genomic animal models have rarely been applied to study adaptive trait genetics in wild organisms (Richards et al., 2023; Schmidt et al., 2023).

A key parameter for fitting animal models is the matrix, **A**, which quantifies the expected proportion of alleles that are identical-by-descent between individuals (de los Campos et al., 2015; Hill & Weir, 2011; VanRaden, 2008). Traditionally estimated from pedigrees, **A** can also be derived directly from genomic data, in which case it is referred to as the *genomic relationship matrix*, describing the realised proportion of shared allele between individuals (de los Campos et al., 2015; Hill & Weir, 2011; VanRaden, 2008). This approach reduces or eliminates the need for pedigrees in

quantitative genetic analyses (Gervais et al., 2019; James et al., 2022; Porth et al., 2015; Schmitt et al., 2021). Nonetheless, genomic animal models can be challenging to apply to wild organisms with large populations. The low likelihood of sampling closely related individuals in large populations results in low power to detect non-zero heritability (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016; Richards et al., 2023). In samples of unrelated individuals, achieving sufficient statistical power in genomic animal models requires large sample sizes; often thousands of individuals as is the case in human and agriculture studies (Niu et al., 2021; Yang et al., 2010). However, obtaining such numbers may be prohibitive in studies of wild organisms, where ethical, logistical, and financial constraints may limit the number of individuals that can be captured, phenotyped, and genotyped.

An additional consideration for estimating **A** is the selection of genetic markers that are used in its construction. Sufficient genetic markers are needed to ensure linkage disequilibrium with loci affecting the focal trait and to accurately capture additive genetic relationships between individuals (de los Campos et al., 2015). Whole-genome sequencing offers far greater genomic coverage than reduced-representation approaches (e.g., RADseq), increasing the likelihood of capturing causal variants or loci in strong linkage with them. Nevertheless, whole-genome data may need to be thinned to avoid oversampling loci, which can introduce noise into the estimation of **A**. This is because over-sampling of genetic markers may introduce noise, reducing the precision and reliability of heritability estimates through redundancy and (or) spurious associations with phenotypes (de los Campos et al.,

2015). There may be thresholds beyond which increasing marker density yields little improvement in heritability estimates (James et al., 2022; Wang et al., 2017). Moreover, different combinations of sampled genetic markers, with different allele frequencies, may yield different estimates of additive genetic variances, and will therefore influence conclusions about trait heritability (Stanton-Geddes et al., 2013; Wang et al., 2017). Consequently, genetic marker thinning and filtering can affect how well **A** captures the additive genetic variance in a focal trait.

Researchers studying wild organisms have typically assessed the power of animal models by essentially simulating genetic breeding values from the structure of **A**, via a genetically inferred pedigree, as in the *MCMCglmm* R package (Hadfield, 2010). In datasets with very low (or no) kin structure, however, the utility of this approach is limited because it is not possible to derive a meaningful genetically inferred pedigree from samples of unrelated individuals (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016). In other words, a method that allows the simulation of different heritability levels independently of **A** would be highly beneficial, for example, one that uses the genotypes directly. Moreover, researchers currently lack a clear framework for evaluating how aspects of experimental design influence the performance of genomic animal models. Such considerations include how the number of sampled individuals, their genomic relationships, and the choice of genetic markers affect the estimation of **A** and, ultimately, the power to estimate heritability in samples of wild organisms.

In this paper, we developed an analytical framework to allow researchers to ask: How much power is there to estimate a given level of heritability with a specific experimental design (the number of sampled individuals, their relatedness, and genetic marker choice)? This framework addresses the key knowledge gap of how to implement and then assess the effectiveness of genomic animal models for species where capturing a mix of close and distant kin is nearly impossible. We use an abundant, high-dispersal marine fish as an example. However, the challenge of applying genomic animal models to samples of uniformly unrelated individuals applies to any species with large population sizes, high dispersal, or without physical or behavioural mechanisms to keep kin in close proximity. These conditions are pervasive in marine taxa (Gagnaire & Gaggiotti, 2016), but also reflect population structures of many other taxa including: migratory birds with high dispersal capacity and low natal site fidelity (Popovic et al., 2019); mammal species characterized by low natal philopatry, large home ranges, and fission–fusion social structures (Beatty et al., 2020); insects with large effective population sizes and high dispersal potential (Endersby et al., 2006); and trees with extensive pollen flow that can homogenise genetic variation over large spatial scales (Bruxaux et al., 2024).

Our paper is divided into three sections. First, we outline our framework, which comprises a unique 'phenotypes-from-genotypes' simulation approach, and uses resampling of genetic markers to explore how the choice of genetic markers influences the performance of genomic animal models. This framework provides a set of simulated expectations against which the observed data can be interpreted.

Second, we showcase the flexibility of our phenotypes-from-genotypes approach to simulate quantitative traits under a range of different heritability levels and genetic architectures. We then use these simulated data to broadly explore how the robustness of heritability estimates from genomic animal models is affected by genetic architecture, relatedness among sampled individuals, and genetic marker choice (non-thinned vs thinned loci, and datasets capturing varying numbers of QTLs). Finally, we apply our framework to data from wild populations of a marine intertidal fish, *Bathygobius cocosensis* (Bleeker, 1854: Perciformes; Gobiidae). In this case study, our framework gave us confidence to conclude that head shape phenotypes in this fish may have a heritable component, despite our working sample of 297 individuals completely lacking any kin structure and variance in genetic relationships.

MATERIALS AND METHODS

Resampling-simulation framework for genomic animal models

Background. In animal models, we can use the known *additive genetic relationship matrix*, **A**, to simultaneously solve for breeding values and estimate the additive genetic variance, V_A (Kruuk, 2004). These genetic values can be combined with other measured fixed and random effect to decompose total phenotypic variance (de Villemereuil et al., 2018), such that:

$$y = c + \mathbf{X}b + \mathbf{Z}u + g + e \quad (1)$$

Where y is vector of phenotypes, c is the intercept, $\mathbf{X}$ is a design matrix for fixed effects, b is a vector of coefficients of these fixed effects, $\boldsymbol{Z}$ is a design matrix for other random effects, u is a vector of coefficients for these random effects, g is the random intercept for additive genetic breeding value, and e is a vector of residual (environmental) deviations. Modelling fixed effects and other random effects help account for additional sources of phenotypic variance, improving model fit and ensuring accurate estimation of V_A. However, partitioning out these additional sources of variance can inflate estimates of heritability by reducing residual V_E and thus the denominator of the heritability calculation (de Villemereuil et al., 2018; Wilson et al., 2010). When such effects are included in animal model fit, it is appropriate to also consider them in the calculation of the phenotypic variance, V_P, (de Villemereuil et al., 2018):

$$h^2 = \frac{V_A}{V_P} = \frac{V_A}{V_F + V_R + V_A + V_E} \tag{2}$$

Here, V_F and V_R are the phenotypic variance due to fixed effects and other random effects, respectively, and V_E is the environmental variance (typically estimated as the residual variance that cannot be ascribed to other modelled sources of variation).

Historically, $\mathbf{A}$ was typically obtained from a known pedigree of individuals. But it is also possible to estimate $\mathbf{A}$ using genetic markers, in which case it is referred to as the *genomic relationship matrix*, providing an estimate of the realised proportion of shared alleles between individuals that are identical-by-descent (de los Campos et al., 2015; VanRaden, 2008). This approach depends on the assumption that the genomic

relationship matrix has been constructed using (at least some) genetic markers that capture information about the trait's additive genetic variance, either directly through quantitative trait loci (QTLs), or indirectly through markers linked to QTLs (de los Campos et al., 2015; VanRaden, 2008). If the loci used to estimate **A** are not in strong linkage disequilibrium with the causal variants underlying the trait, then genomic animal models will underestimate V_A (de los Campos et al., 2015). The methods used to acquire SNPs, whole-genome vs reduced-presentation sequencing (Figure 1: Step 1), and quality filters applied to these SNPs (Figure 1: Step 2), may therefore impact on how accurately the estimate of V_A reflects the total additive genetic contribution to phenotypic variation (Gervais et al., 2019; Stanton-Geddes et al., 2013; Wang et al., 2017). There are two further general challenges to estimate of V_A from a genomic relationship matrix.

The first concern is related to the choice of which genetic markers to use when estimating **A**. While using all available (quality-filtered) SNPs to estimate **A** has the greatest opportunity to capture true QTLs, it can also lead to inaccurate estimates of V_A with large confidence intervals (de los Campos et al., 2015). This is because the inclusion of many redundant markers in high linkage with causal loci can upwardly bias the estimate of V_A; similarly, including more markers increases the risk of sampling markers that show spurious associations with the trait, again upwardly biasing the estimate of V_A (de los Campos et al., 2015). Thinning genetic markers may reduce these risks of upwardly biased estimates. However, thinning has greater tendency to produce downward biases on V_A estimation because there is a greater

likelihood of missing markers that are linked to causal QTLs (de los Campos et al., 2015). Depending on sampling of causal QTLs, redundant markers and spuriously associated markers, different sets of genetic markers can yield different estimates of **A**. One goal of this paper is to demonstrate how the distribution of estimates of V_A (and h^2 estimates) from multiple different sets of genetic markers (and hence different estimates of **A**) (Figure 1: Step 3) can support robust interpretation of this evolutionarily important parameter.

The second concern relates to power, specifically whether an estimated V_A of zero (no heritability) reflects a true lack of additive genetic variation or instead results from low power to detect it. Sample size is an important determinant of power and the precision of V_A estimates. In classical breeding designs, it is relatively straightforward to calculate the effective sample size and predict power (Klein, 1974). However, in genomic animal models, power is harder to quantify because it depends on the specific distribution of relatedness estimates, and the (often unknown) genetic architecture linking genotype to phenotype (James et al., 2022; Klein, 1974; Stanton-Geddes et al., 2013). When **A** reflects a range of closely and more distantly related individuals, the shared genotypic information supports accurate estimation of genetic effects on phenotype, and of V_A. In such cases, simulation of the genetic pedigree relationships can be used to determine power (Hadfield, 2010). However, in wild organisms, random samples typically lack close relatives, and the relationship structure is unknown or irregular. Low relatedness limits the replication of genotype–phenotype covariation that underlies the detection of heritable signal, that is,

reduced power, and also constrains the use of existing simulation tools for determining power (Fraimout et al., 2024). In many wild taxa, feasible sample sizes are often limited by ethical, logistical, and financial constraints. There is a clear need for methods to assess the power to detect non-zero V_A in genomic animal models using data from wild populations with limited or unknown relatedness structure.

Brief overview. In this paper, we present a novel 'resampling-simulation' framework for assessing the power of genomic animal models applied to wild organisms. The *resampling* component operates on two levels. This first level of resampling, thinning loci to generate multiple SNP sets, is optional. However, we show that it can be advantageous on several fronts: by providing insight into how marker choice affects estimates of heritability; by reducing spurious signals; and by producing distributions of heritability estimates that provide a more intuitive picture of parameter space in which they occupy. The second level of resampling, generating different QTL sets for simulation, is an essential mandatory step. This forms the core of the *simulation* component, which examines how traits with varying combinations of causal loci influence heritability estimates in genomic animal models.

The *simulation* component of our framework includes what we term a 'phenotypes-from-genotypes' approach. This innovative approach allows quantitative traits to be simulated directly from observed genotypes rather than from the structure of the estimated **A**. This is particularly valuable for samples of unrelated wild organisms,

where the lack of kin structure can render simulations based on **A** largely uninformative.

Conceptually, our resampling-simulation framework is similar to bootstrapping. We are effectively redrawing from available genetic markers to generate different combinations to be used in SNP sets and (or) QTL sets. Power is assessed by comparing observed results to simulated expectations and identifying the conditions under which the experimental design is likely (or unlikely) to detect non-zero heritability. The steps of our framework are outlined in Figure 1, and a detailed schematic is presented in the Supplementary Methods ('Schematic outline for the resampling-simulation framework'). The following subsections describe steps for unthinned genetic markers (a single SNP set), randomly thinned genetic markers (multiple SNP sets), and the general case where simulated QTLs are drawn at random.

Step 1: Acquire SNPs. Whole-genome sequencing or reduced representation sequencing can be used to acquire SNPs. Whole-genome sequencing is best because it captures genetic variation across the genome. However, reduced representation sequencing can be effective if the distribution of SNPs across the genome is dense enough to capture linkage with causal loci. Reduced-representation sequencing will often be the approach of choice for many wild organisms, as funding constraints can preclude large-scale whole-genome sequencing (Schmidt et al., 2023). For many non-model species, the absence of a reference genome also poses the challenge of

being unable to determine the genomic coverage or density of loci obtained through reduced-representation methods.

Step 2: Quality filter SNPs. As with any genomics study, filtering is essential for removing poor quality SNPs. However, the number of SNPs available for the construction of **A** is important. Too few SNPs will result in low power, so an optimal balance between removal and retention criteria is needed. Even moderate amounts of missing data may be tolerable because they can outweigh loss of power from removing too many loci (Attard et al., 2018).

Step 3: Specify the SNP set type. At least one SNP set is required to construct **A**. Two approaches can result in there being only a single realisation of **A** from a given data set: all SNPs are retained (not thinned), or by simply thinning the data once by sampling one SNP per linkage block (whole-genome sequencing) or contig (reduced representation sequencing). However, if SNPs are thinned and randomly selected as genetic markers, it is possible to generate many different marker combinations across linkage blocks or contigs. We recommend that studies using SNP thinning generate at least 20 SNP sets. Different combinations of genetic markers can capture additive genetic variance in different ways. By fitting genomic animal models with multiple SNP sets, distributions of quantitative genetic parameters can be obtained. These distributions help infer the plausible range of values that estimates may take.

When using reduced-representation data, we recommend dividing contigs into two sets: *focal contigs* and *reserved contigs*. *Focal contigs* are used in both the

construction of the *observed* **A**, and in *simulations* for simulating traits where QTLs are captured in the *observed* **A**. *Reserved contigs* are used only in *simulations* for simulating traits where QTLs are not captured in the *observed* **A**. This approach ensures consistency in SNP numbers across simulation scenarios and more realistically mimics the presence of unseen heritable variation in reduced-representation datasets (Figure 2c).

Step 4: Construct A-matrices for SNP sets. For each SNP set, we will have a corresponding observed **A** derived from those markers, and therefore there will be as many A-matrices as there are SNP sets.

Step 5: Fit models on the observed data. For each SNP set, we fit a genomic animal model using the **A** derived for that SNP set. A single null model is fit without any **A**. Comparison of each genomic animal model to the null model using a log-likelihood ratio test provides a test for the null hypothesis that there is no heritable phenotypic variation For multiple SNP sets, this will generate a range of observed estimates of heritability and LRT *P*-values. For a singular SNP set analysis, where there is only a single realisation of **A**, confidence intervals are the sampling variance around the observed heritability estimate. In this study, we used restricted maximum likelihood multivariate normal (REML-MVN) confidence intervals (Dugand et al., 2025; Houle & Meyer, 2015). When there are multiple SNP sets of randmoly sampled thinned SNPs, and therefore A-matrices, esimates of heritability and model comparison *P*-values, we derive statistics that describe the distribution of the *SNP set point estimates*. We

characterise the *global median* as the median across the SNP set point estimates of heritability and the confidence intervals of the observed heritabilty by the 2.5% and 97.5% percentiles of the *SNP set point estimates*.

Step 6: Simulate phenotypes. An algorithm for implementing these simulations is available via the *family_sim_qtl* function from the *genomalicious* R package (Thia, 2024). Detailed information on this underlying algorithm and our overall framework can be found in the Supplementary Methods ('Simulating a quantitative trait using *family_sim_qtl*'). Our phenotypes-from-genotypes simulation approach uses a rearrangement of a classic quantitative genetic formula (Falconer, 1989) that links V_A to the allele frequencies (p and q) and the homozygote and heterozygote genotype effects (a and d, respectively) at each mth biallelic QTL contributing to the trait. For genotypic effects, a is always greater than zero, and d is always zero, which is pure additive inheritance with no dominance effects. If there are M total QTLs, then:

$$V_A = \sum_{m=1}^{M} 2p_m q_m [a_m + d_m(q_m - p_m)]^2 \quad (3)$$

This flexible approach can accommodate any number of loci, effect sizes, and levels of heritability. A key point is that V_A is simulated relative to the observed allele frequency distributions and genotypes in the observed sample of individuals. In other words, simulation of additive genetic variance is performed independently of the inferred observed **A**.

To apply our phenotypes-from-genotypes approach, a set of nested QTLs are defined for each SNP set. For each QTL set, we randomly draw M SNPs to act as the

simulated causal loci. The purpose of resampling SNPs to simulate QTLs is to assess how different combinations of causal loci affect detection power for a given SNP set. In other words, how much does power vary for a given **A** under different configurations of underlying QTLs? For non-thinned SNP sets, all loci have the potential to be simulated QTLs. For thinned SNP sets, we suggest simulating two simple contrasting scenarios: one where **A** is fitted *without* any QTLs in the sampled SNP set, and one where **A** is fitted *with* all QTLs in the sampled SNP set (Figure 2). For the thinned 'without QTLs' scenario, the *M* SNPs for each QTL set are drawn from SNPs *not contained* in the focal SNP set. For the thinned 'with QTLs' scenario, the *M* SNPs for each QTL set are drawn from the SNP *contained* in the focal SNP set.

Next, we defined a range of known levels of heritability to be simulated. For each QTL nested in each SNP set, we used the phenotype-from-genotypes approach to simulate phenotypes for each known level of heritability. At least one of the levels of heritability needs to be zero, which effectively acts as the null hypothesis, with additional non-zero levels of heritability representing alternative hypotheses. We suggest that the range of levels of heritability captures a range of small, moderate, and larger levels of heritability, thereby capturing the spectrum of power required for fitting genomic animal models.

Step 7: Fit models on the simulated data. As with the observed phenotypes, we fitted genomic animal models and a corresponding null model to the simulated phenotypes for each combination of level of heritability, for each SNP set, and each

QTL set; for randomly thinned SNP sets, we did this separately for the 'without QTLs' and 'with QTLs' scenarios. This will produce a range of point heritability estimates and model comparison *P*-values for both the 'without QTLs' scenario and the 'with QTLs' scenario, in whichever SNP set strategy is used (non-thinned or thinned).

Estimates of the simulated heritability are summarised in hierarchical fashion for in the 'with QTLs' and 'without QTLs' scenarios. First, take the median estimate *within* each SNP set across all nested QTL sets, which are the *SNP set medians*. Then the median of those SNP set medians is the *global median* of simulated heritability. The *global median* represents the central heritability estimate derived from simulations across all SNP sets. The confidence intervals are the *global median* is obtained by calculating the 2.5% and 97.5% percentiles across the *SNP set medians*.

Step 8: Compare observed vs simulated results. Our simulations provide expectations of the estimates we might expect under different levels of heritability, given our experimental design (sampled individuals and genetic markers) and the specified genetic architecture simulated. Interpretation of observed and simulated results can be conceptually framed as addressing four key questions:

1. ***What is the distribution of estimates for simulated heritability of zero?*** *Simulated* heritability of zero represents the null hypothesis, defining the distribution of estimates expected if no heritable variation exists. While estimates should be near zero, genomic noise and spurious associations could produce positive values for heritability.

2. ***What levels of simulated non-zero heritability are distinguishable from zero?*** *Simulated* non-zero heritability provides the alternative hypothesis, defining the distribution of estimates expected if heritable variation exists. We are interested in how these non-zero *simulated* distributions compare to the null of zero *simulated* distribution. This provides insight into the power of our experimental design to detect different non-zero levels of heritability. Smaller levels of heritability may not be distinguishable if experimental designs are under-powered.
3. ***What are the expectations for simulated QTLs excluded vs included in the A-matrix?*** *Simulations* of the different A-matrix scenarios provide additional insight into power under different levels of heritability: the 'Without QTLs' scenario reflects a low-power scenario, whereas the 'with QTLs' reflects a high-power scenario. For a given level of known *simulated* heritability, we are interested in where the distribution of estimates lies for each scenario, and how different these two scenarios are from each other.
4. ***Where does the observed distribution of estimates sit in relation to simulated distributions?*** For the *observed* estimates, we first assess whether the distribution is clearly distinguishable from those *simulated* under zero heritability (Question 1), which would indicate that our design has power to detect non-zero heritability. If so, we then compare the *observed* estimates with *simulated* estimates for non-zero levels of heritability to better place the parameter space in which our *observed* estimates occupy (Question 2). Finally,

we compare *observed* estimates across different A-matrix scenarios (Question 3) to evaluate how well our design may captures additive genetic variance, and hence, power.

Visualisation offers a qualitative and intuitive means for comparing differences observed and simulated distributions. For thinned SNP sets, we complemented visualisations with quantitative measures to summarise results and test for differences between observed and simulated estimates:

- ***Percentage of significant log-likelihood ratio tests.*** The LRT *P*-values for the comparison of genomic animal models vs null models is a test for the hypothesis that inclusion of additive genetic breeding values provides a better description of phenotypic variance. We took the percentage of LRT *P*-values that were <0.05 for both the *observed* (multiple thinned SNP sets only) and *simulated* data as a measure of how often this hypothesis was supported.
- ***Percentage of observed estimates that are greater than simulated estimates.*** Each *observed* SNP set heritability point estimate was paired with the corresponding *simulated* SNP set median estimate for different known heritability levels in the 'without QTLs' and 'with QTLs' scenarios. We calculated the percentage of pairs where the *observed* estimate exceeded the *simulated* estimate, providing a measure of whether the *observed* data surpassed *simulated* expectations and helping to locate *observed* estimates in parameter space.

- ***Paired Wilcox signed-rank test.*** A one-sided paired Wilcox signed-rank test was used to test the hypothesis that *observed* SNP set heritability point estimates were greater than *simulated* SNP set median estimates. Tests were performed for each level of known simulated heritability and for the 'without QTLs' and 'with QTLs' scenarios separately. Bonferroni correction was used to adjust *P*-values for multiple testing.

*SLiM*ulated population

Overview. We simulate a population of individuals with heritable and non-heritable quantitative genetic traits as a proof-of-concept of our phenotypes-from-genotypes simulation approach (Steps 6 and 7). We use these analyses to demonstrate that our approach provides a flexible method for simulating traits, and to explore how different genetic architectures, levels of heritability, and experimental designs affect the performance of genomic animal models (Figures 3 and 4). We emphasise that these analyses *are not* a demonstration of our complete resampling-simulation framework; we will implement that below with data from wild populations.

SLiM model. We used *SLiM* v4.3 (Haller & Messer, 2023) to simulate a population of diploid individuals with a single-chromosome 1 Mb genome, evolving under pseudo-Wright-Fisher dynamics with an effective population of 10,000 individuals, for 40,000 generations, a mutation rate of $1e^{-7}$, and a recombination rate of $1e^{-5}$ (approximately 10 recombination events per meiosis). Although *SLiM* is typically used to simulate multiple populations under different scenarios, here, we exploited SLiM's utility for

individual-based simulation to generate a large, genetically variable population, and focus our interpretation on how different strategies for sampling from such a population might impact estimates of V_A and interpretation. After 40,000 generations, we randomly sampled 500 individuals from the population for further investigation; in this sample, 3% of sites were polymorphic, with a median minor allele frequency of 0.09, and a range of 0.0006 and 0.5. The sample size of unrelated individuals (500) was chosen to reflect an ambitious, but achievable number of samples that could be collected from wild organisms (Duntsch et al., 2020). In addition to randomly sampled individuals, we also generated 50 pairs of close kin under the following familial relationships: half-cousins, cousins, half-siblings, and siblings. These were also sampled at the end of 40,000 generations. See the Supplementary Methods section 'Generating close-kin samples in *SLiM*' for further details.

Simulating QTLs to explore the effects of genetic architecture and experimental design. In most cases, the genetic architecture of traits in non-model wild organisms is unknown. Although genome-wide SNPs are used to construct **A**, the true QTLs underlying a trait are unlikely to be evenly distributed across the genome. How does the performance of genomic animal models vary when QTLs are clustered in certain regions versus evenly spaced throughout the genome? Addressing this question can shed light on how different genetic architectures influence model performance. In addition, samples from wild populations often lack close relatives, limiting the power to detect heritable variation. How much might the inclusion of small groups of close

kin (e.g., sibling pairs) improve this power? Answering that question could help guide sampling strategies when collecting specimens from wild populations.

To explore these questions, we used our phenotypes-from-genotypes approach to generate traits for the *SLiM*ulated individuals in R. We considered two aspects of genetic architecture: (1) the genomic distribution of quantitative trait loci (QTLs), either *clumped*, where QTLs are concentrated in a small genomic region, or *even*, where QTLs are spaced across the genome; and (2) the distribution of QTL effect sizes, either *uniform*, where all loci contribute equally, or *variable*, where some loci have larger effects than others. These scenarios were chosen for both practical and biological reasons: uniform effect sizes are straightforward to simulate, but in many non-model systems, the true genomic distribution of QTL and their effect sizes are unknown. Exploring these contrasting scenarios allowed us to evaluate how assumptions about genetic architecture influence the performance of genomic animal models. (Realised values for these simulated traits are reported in Supplementary Table S1.)

We separately modelled quantitative traits under each combination of genomic distribution and effect sizes of QTLs for three levels of heritability: 0, 0.05, and 0.3. We focused on simulating no or low heritability because this is the range where low power is most likely to be confounded with a genuine lack of heritable variation in observed populations. These simulations allowed us to evaluate how different factors influence the power to detect true non-zero heritability and to assess the risk of

spurious signals that may lead to false interpretations of heritable variation in under-powered studies. By including simulations with simulated heritability of zero, we can quantify the potential for false associations, and the risk of misinterpreting observed estimates. Published estimates of heritability in wild populations suggest values of ~0.5 for morphological, behavioural, and physiological traits, and ~0.3 for life-history traits (Postma, 2014).

We only allowed SNPs with a minor allele frequency >0.01 to be simulated as QTLs, to avoid simulating on SNPs with near-zero genetic variance. Linkage disequilibrium was <500 bp in our synthetic genomes (Supplementary Figure S1), so we divided our genome into 2,000 linkage blocks of 500 bp. QTLs were drawn only from linkage blocks with multiple SNPs. This was to accommodate our downstream analysis where we replaced QTLs with non-causal loci from the same linkage block. To generate an even distribution of QTLs, we randomly drew one SNP locus approximately every 20th linkage block (~10,000 bp) for a total of 100 QTLs. To generate a clumped distribution of QTLs, we performed a pseudo-random sample of 100 SNP loci from a group of adjacent linkage blocks in a 10% region of the *SLiM*ulated genome to maintain the same frequencies of QTLs used in the even architecture. For the clumped architecture, multiple QTLs could come from the same linkage block. The uniform effect size scenario had all QTLs with an effect size of one. For the variable effect size scenario, QTL effect sizes (a in Equation 3) were drawn from a beta distribution, Beta(α = 0.5, β = 1), using R's *rbeta* function.

We examined how weak kin structure under different family relationships, along with marker choice for **A** construction, might influence estimates of heritability under the different genetic architectures. The original sample of random individuals (n = 500) was chosen to reflect an ambitious, but achievable number of samples that could be collected from wild organisms (Duntsch et al., 2020). However, in this random sample, some pairs of individuals exhibited evidence of being close kin. To define a sample with no or very distant kinship, we removed one member from pairs exhibiting genomic relatedness estimates >0.0625 (theoretical relatedness of half-cousins). This left us with a working sample of 371 unrelated individuals to use as a dataset with no close kin, which would still comprise a nontrivial sample size from a population of wild organisms. To create corresponding datasets with close kin, we removed the first 100 unrelated individuals from the random sample and substituted them with 50 pairs of individuals of (~27% of analysed individuals) with different familial relationships. This replacement procedure was performed separately for each familial relationship: siblings, half-siblings, cousins, or half-cousins.

We then generated a series of SNP sets (Step 3) to construct **A** (Step 4) and assess how marker choice affects heritability estimation from genomic animal models. The strategies included: (1) 'All SNPs', no thinning; (2) 'Just QTLs', thinning to include only QTLs; (3) 'Random SNPs', thinning to one random SNP per linkage block; (4) 'With QTLs', thinning to one SNP per linkage block, with preference for including QTLs if they occurred in the linkage block, otherwise, the SNP was drawn randomly; (5) 'Without QTLs', thinning to one SNP per linkage block, specifically excluding QTLs,

with one SNP drawn randomly. See Figure 2a for an illustration of these marker choice strategies. For the thinning strategies, we generated 80 unique SNP sets to be used with both clumped and even genetic architectures. **A** was constructed for each SNP set using the *A.mat* function in the R package *sommer* (Covarrubias-Pazaran, 2016). We used the REML-based *mmes* function from *sommer* to fit an observed model (Step 5) that only included a global intercept and the random intercept for additive genetic breeding values.

Wild populations of *Bathygobius cocosensis*

Species description. *Bathygobius cocosensis* (Bleeker, 1854: Perciformes; Gobiidae) is a marine intertidal fish with a wide distribution throughout the Indo-Pacific. In Australia, *B. cocosensis* is a major resident species of rocky intertidal zones throughout tropical and temperate-subtropical waters (Griffiths, 2003). 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).

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 3a) is a rocky headland with a topologically complex intertidal zone with discrete pools that *B. cocosensis* retreat to during low tide. Heron Island (Figure 3b) 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 2017. At Hastings Point, sampling occurred on the 25th of February 2016, and on the 24th and 25th of February 2017. 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). 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.

Phenotype data. We used a set of six morphological landmarks to characterise head shape phenotypes (Figure 3c). 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). Imported landmarks were then subjected to PCA using R's *prcomp* function to derive three head shape traits, phenotypic PC1, PC2, and PC3, that explained 39.54%, 25.14%, and 15.11% of the total phenotypic variation, respectively. We determined these axes to be the only ones that described significant variation in head shape phenotypes based on permutation tests performed with the *PCAtest* package (Camargo, 2022).

Framework: Steps 1 and 2. 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. Full details of the bioinformatic processing can be found in the Supplementary Methods ('Bioinformatic processing of *B. cocosensis* ddRAD data'). 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).

Framework: Steps 3 and 4. A total of 29,465 SNP loci were distributed across 4,220 RAD tags (statistical properties illustrated in Supplementary Figure S3, S4 and S5). A total of 80 SNP sets were generated following an approach to minimise redundancy (and thus the potential to over-estimate V_A). Specifically, each SNP set was independently produced by first randomly drawing 4,120 RAD tags, then randomly drawing one SNP per RAD tag. Therefore, 4,120 SNPs were used for the construction of **A** per SNP set. The non-sampled RAD tags within each SNP set were reserved for QTL simulation *without* QTLs (details below).

Our final dataset for head shape heritability analysis comprised 297 fish with both genotypic and phenotypic data: 40 and 166 from Hastings Point in 2016 and 2017 (respectively), and 46 and 45 from Heron Island in 2016 and 2017 (respectively). Across SNP sets, the mean relatedness was –0.002, with a minimum of –0.12 and maximum of 0.14. Therefore, by-and-large, there was very little evidence for close-kin in our dataset and most individuals were likely unrelated.

We quantified genetic structure among the spatiotemporally separated populations using the *fstat_calc* function from *genomalicious* to calculate F_{ST} as per (Weir & Cockerham, 1984). Although genetic structure was statistically significant (as assessed by permutation), F_{ST} values were very small: pairwise F_{ST} ranged from 0.001 to 0.004 (Supplementary Table S7). We therefore pooled fish from all populations together in our main analysis of heritability and also performed supplementary analyses on each population separately (see below).

Framework: Step 5. We fitted genomic animal models to the observed data using the *mmes* function from *sommer* for phenotypic PC1, PC2 and PC3. For each trait, we fitted the fixed effects of 'Fish size' (continuous measure of head size represented by the centroid size), 'Population' (categorical effect capturing site and year specific influences on mean phenotype), and 'Fish size:Population' (the interaction between these terms). We also included the random intercept for additive genetic breeding values.

Framework: Step 6. Simulations using the *family_sim_qtl* function from *genomalicious* were performed using simulated h^2 of 0, 0.05, 0.15, and 0.3 for each phenotypic PC and each of the 80 SNP sets. For each SNP set, we simulated a total of 5 nested QTL sets, each containing 100 randomly drawn loci with uniform effects. Simulated genetic breeding values and residual deviations were combined with fixed effect estimates of 'Population', 'Fish size', and 'Population:Fish size' derived from the simple linear models.

Framework: Step 7 and 8. Genomic animal models and corresponding null models were fit using *sommer*'s *mmes* function. Full details of the analysis of heritability can be found in the Supplementary Methods ('Analysis of heritability in *B. cocosensis*'). We compared the distributions of observed and simulated heritability estimates to infer confidence in our conclusions about the magnitude of heritable variation in *B. cocosensis* head shape.

RESULTS

*SLiM*ulated population

Key result 1: SNP sets including QTLs provide more accurate estimation of the true heritability. This result aligns with expectations, as the inclusion of causal loci ensures that the genomic relationship matrix (**A**) captures the component of phenotypic variation due to shared genetic effects (de los Campos et al., 2015). If we compare two extremes, using 'All SNPs' vs 'Just QTLs', we can see that accuracy and precision are dramatically improved when **A** is only constructed with causal loci (Figure 3). For thinned resampled datasets, including all QTLs in **A** ('With all QTLs') typically led to estimated heritability close to the simulated value, while constructing **A** with only some QTL ('Random SNPs', variable numbers of QTLs are randomly sampled as markers) typically resulted in under-estimation for moderate levels of heritability (Figure 4). The 'Without QTLs' strategy (in which QTLs are never among the sampled markers) generally produced estimates similar to those from the 'Random SNPs' strategy, suggesting that linkage in the 'Without QTLs' strategy may contain as much information about additive genetic variance as the variable mix of QTL and non-QTL markers in the 'Random SNPs' strategy.

Key result 2: There is a trade-off between not thinning vs thinning SNPs. Not thinning increases the likelihood of capturing QTLs (or closely linked markers) at the risk of increasing redundancy and noise (de los Campos et al., 2015). And in contrast, thinning reduces the likelihood of capturing QTLs (or closely linked markers) at the benefit of reducing redundancy and noise (de los Campos et al., 2015). Here, we

consider the estimates and uncertainty in the non-thinned 'All SNPs' strategy vs the thinned 'Random SNPs' strategy. We focus on these two scenarios because they are the most realistic marker choice strategies in quantitative genomic studies.

For the 'All SNPs' strategy, the REML-MVN 90% confidence intervals were very large, suggesting considerable uncertainty in heritability estimates (Figure 3). In some scenarios with low simulated heritability, point estimates were notably inflated (Figure 3: 'All SNPs' at h^2 = 0 and 0.05). Such uncertainty and overestimation may reflect the inclusion of markers in weak linkage with QTL or other sources of genomic noise (de los Campos et al., 2015). Although wide confidence intervals can be seen as 'conservative', interpretation is complicated when they span both very small values (sometimes overlapping zero) and unrealistically large heritability estimates. Moreover, a greater concern with not thinning SNPs is that including too many loci might introduce spurious effects that artificially inflate heritability estimates, potentially leading to false conclusions about the magnitude of heritable variation.

In contrast, the 'Random SNPs' strategy produced narrower confidence intervals but tended to underestimate heritability. Global median estimates across the 80 SNP sets typically fell within the REML–MVN confidence intervals of the 'All SNPs' strategy (contrast 'Random SNPs' in Figure 4 to 'All SNPs' in Figure 3), indicating that the 'Random SNPs' results are generally consistent with expectations with the 'All SNPs' results. A key advantage of resampling multiple thinned SNP sets is that it yields a distribution of estimates, providing a clearer picture of where the bulk of values lie

(Figure 4; Supplementary Figure S2). Although thinning might introduce a downward bias, it may also reduce noise and spurious signals by essentially averaging-out those biases over different combinations of genetic markers; this is beneficial because they lead to more truly conservative results.

Key results 3: Genomic animal models fit to samples of unrelated individuals performed similarly to samples including close-kin pairs. In our *SLiM*ulated population, we introduced only weak kin structure by including 50 pairs of individuals with different types of familial relationships. Although close kin made up approximately 27% of the 371 individuals in these datasets, this level of kin structure was insufficient to consistently improve the performance of heritability estimates relative to datasets composed entirely of unrelated individuals. We consider the implications of this further in the Discussion.

Key result 4: Genomic animal models fit under different genetic architectures performed similarly. Heritability estimates were comparable irrespective of whether the genetic architecture comprised QTLs with genomically clumped or even distribution, or whether the effect sizes of QTLs were uniform or variable (Figures 4 and 5). In our *SLiM*ulated dataset, genotyping was simulated to reflect a whole-genome sequencing approach, allowing us to identify linkage blocks (Supplementary Figure S1) and evenly space markers across the genome. Within the limits of our analysis, the results suggest that the genetic architecture may not substantially affect the performance of genomic animal models if markers are evenly spaced and

correspond well with the underlying linkage structure. The LRT *P*-values for model comparisons are reported in Supplementary Tables S2 to S6.

Wild populations of *Bathygobius cocosensis*

By applying our framework to wild-caught *B. cocosensis*, we were able to demonstrate that that our experimental design (sampled individuals and genetic markers) should be sufficiently powered to detect moderate levels of heritability if informative loci were included in the construction of **A** (Figure 6). Our *simulations* indicated that under a simulated heritability of zero, the expected distribution of heritability estimates would likely have a peak at (near) zero with a right tail (Figure 6, top row). With increasing levels of *simulated* heritability, the distributions of estimates under the 'with QTLs' scenario became increasingly distinguishable from the null of zero heritability, especially at the *simulated* levels of 0.15 and 0.3 (Figure 6, bottom two rows). In contrast, in the 'without QTLs' scenario, the distributions of heritability estimates could only be clearly distinguished at the highest *simulated* level of 0.3 (Figure 6, bottom row). Hence in the absence of informative loci used to construct **A**, it would be unlikely that we would have power to detect low levels of heritable variation.

For phenotypic PC1, the *observed* global median heritability estimate was 0.17, and the *observed* distribution of SNP set point estimates was clearly distinguishable from the expectation under *simulated* heritability of zero (Table 1; Figure 6). *Observed* model comparisons between genomic animal and null models were significant in

96% of SNP sets (that is, where the LRT *P*-values < 0.05). The *observed* distribution of SNP set point estimates were clearly different from all *simulated* levels in the 'without QTLs' scenario and fell within the range of estimates expected under a simulated level of 0.15 in the 'With QTLs' scenario (Table 2; Figure 6). These results support the presence of heritable variation in phenotypic PC1 and might indicate that we have captured some informative loci in our set of analysed genetic markers (see also Supplementary Figure S6, left panel).

For phenotypic PC2, the *observed* global median heritability estimate was 0.0, with SNP set point estimates of heritability being consistently lower than all *simulated* estimates in both the 'Without QTLs' and 'With QTLs' scenarios (Table 2; Figure 6). *Observed* model comparisons were significant in 0% of SNP sets. The *observed* distribution of estimates was markedly different to that expected under the null hypothesis of a *simulated* heritability of zero (Figure 6). Whereas both the *simulated* 'Without QTLs' and 'With QTLs' scenario produced a range of zero and non-zero SNP set median estimates (Figure 6), 91% of *observed* SNP set point estimates were exactly zero (see also Supplementary Figure S6, middle panel). These results support a lack of detectable heritability in phenotypic PC2.

For phenotypic PC3, the *observed* global median heritability estimate was 0.03. The bulk of *observed* SNP set median heritability estimates was centered near zero but was slightly higher *simulated* heritability of zero in the 'Without QTLs' and 'With QTLs' scenarios (Table 2; Figure 6). However, there was no support for *observed*

heritability being greater than the lowest *simulated* heritability of 0.05. *Observed* model comparisons were significant in 0% of SNP sets. These results might suggest the presence of heritable variation, but support is overall weak and may suggest low power to estimate heritability in phenotypic PC3 (see also Supplementary Figure S6, right panel).

Finally, we fitted separate genomic animal models for each population to evaluate their performance if we had not pooled our sampled individuals (Supplementary Figure S7). For phenotypic PC1, we observed clearer and more consistent signals of heritability in the Hastings Point populations, with many SNP sets yielding moderate heritability estimates. In contrast, the Heron Island populations showed far fewer SNP sets yielding moderate heritability for PC1, and a much higher proportion of near-zero estimates, suggesting a weaker signal of heritability. These may be partially due to sample size: Hastings Point 2017 had a much larger sample size (n = 166) relative to Hastings Point 2016 (n = 46), and Heron Island 2016 (n = 40) and Heron Island 2017 (n = 45). For phenotypic PC2 and PC3, patterns were more heterogeneous across all populations. While some SNP sets produced moderate to high estimates of heritability, most estimates were near zero, especially in the Heron Island populations. Notably, the distributions for phenotypic PC2 and PC3 were often broad or bimodal, highlighting the variability among SNP sets in describing additive genetic variance. These observations were broadly consistent with the results from the combined population analysis. However, they also underscore that heritability

estimates can vary considerably depending on the subset of individuals and the specific SNP set used.

DISCUSSION

A resampling-simulation framework for assessing the robustness of genomic animal models

Our framework provides a new approach for researchers to apply genomic animal models to wild organisms where a random sample of individuals is expected to have uniformly low relatedness. While simulations are a common tool for assessing power and bias in quantitative genetic analyses (e.g., Hadfield, 2010), a core innovation in our framework is our 'phenotypes-from-genotypes' approach that uses rearrangement of a classic quantitative genetic formula. Our framework provides several benefits that help overcome limitations associated with samples lacking close kin (only comprising distant relatives) (Fraimout et al., 2024; Gagnaire & Gaggiotti, 2016). Firstly, we can simulate phenotypes independently of the genomic relationship matrix, **A**. This is highly advantageous for studies of wild organisms that have demographic and life history characteristics that make it nearly (if not) impossible for random samples to include close kin. Secondly, phenotypes and their heritability are simulated relative to the sample allele frequency spectrum. This allows the simulations to be informed by, and grounded in, the genetic characteristics of the observed data. Thirdly, our approach is very flexible, allowing simulations under different genetic architectures, which could be used to explore the performance of

genomic animal models with different numbers of contributing loci and different effect sizes.

At the core of our framework is resampling markers to act as hypothetical QTLs in simulations. This is important for assessing how well an A-matrix, for a given SNP set and sample of individuals, captures additive genetic variance in a trait across a range of hypothetical causal loci, which may or may not be among, or linked to, the analysed markers. Our framework also includes the resampling of loci to generate a suite of randomly thinned SNP sets and associated A-matrices. Genetic markers vary in the information they carry about a trait's additive genetic variance (de los Campos et al., 2015; Hill & Weir, 2011; Stanton-Geddes et al., 2013; Wang et al., 2017). The SNPs available for analysis depend on factors such as sequencing approach, quality filtering, and thinning procedures. By fitting genomic animal models to a range of randomly thinned SNP sets, we can generate distributions of quantitative genetic estimates for different combinations of markers. This helps us understand how marker choice influences model performance and provides a more intuitive interpretation of the parameter space occupied by the observed data. Analysing multiple thinned SNP sets may also yield more conservative estimates, reducing the influence of spurious loci on model outputs.

Our study provides insight into a key question for **A** construction: to thin, or not to thin (de los Campos et al., 2015). Using all available genetic markers increases the probability that true QTLs (or linked loci) will be included in the construction of **A**.

This is more likely for whole-genome sequencing, but also possible for reduced representation sequencing that has achieved a high density of markers. However, a potential drawback of not thinning is that many uninformative genetic markers are also included in the construction of **A**. In our *SLiM*ulated population, the 'All SNPs' strategy, in contrast to 'Just QTLs', produced inaccurate, often inflated, estimates of heritability with large confidence intervals (Figure 3). In contrast, using thinned genetic markers reduces the probability that true QTLs (or closely linked loci) will be included in the construction of **A**. However, this may come with the benefit that less noise is introduced from oversampling genetic markers, which introduces both non-independence and compounds the effects of uninformative markers. In our *SLiM*ulated population, the resampled 'Random SNPs' thinning strategy produced conservative, downwardly biased heritability estimates when aggregating across SNP sets. This conservatism may be advantageous, as it reduces the risk of over-interpreting the presence of heritable variation.

We found that the performance of genomic animal models was comparable across the different genetic architectures simulated. In our *SLiM*ulated dataset, this included a clumped vs even genomic distribution of QTLs, and uniform vs variable QTL effect sizes. Prior work has demonstrated that at a certain density, sampling additional genetic markers provides little gain in the performance of genomic animal models (James et al., 2022; Stanton-Geddes et al., 2013; Wang et al., 2017). These data may suggest that for a given experimental design with genetic markers that align well with linkage structure, a trait's genetic architecture may not impact

the performance of genomic animal models. However, several aspects of genetic architecture were not explored in this study. We only considered 100 QTLs, a moderately polygenic trait, at the exclusion of oligogenic traits (a few QTLs) or monogenic traits (a single major gene). We also did not examine the effects of dominance. As we show in our application to a wild fish population, simulating the true or exact genetic architecture need not be the focus. The primary aim of our framework is to assess power, specifically, whether a given experimental design can detect non-zero heritability. While it is sensible to simulate a genetic architecture relevant to the trait, this is not essential. Our flexible simulation approach can also be used to explore expectations under different genetic architectures, as we partially demonstrated with *SLiM*ulated population, though this was not the focus of the wild fish example.

Our exploration of kin structure on genomic animal model performance provides additional insight into the challenges of applying quantitative genomic methods on wild organisms. In our *SLiM*ulated dataset, we tested different experimental designs with weak kin structure. This involved replacing our sample of 371 unrelated individuals with 50 pairs (100 individuals, 27%) of close kin with different familial relationships (siblings, half-siblings, cousins, half-cousins). However, we found that none of the kin-structured datasets improved heritability estimates over the dataset comprised of completely unrelated individuals, irrespective of genetic architecture and marker choice strategy. Our kin-structured datasets were simulated in a way to replicate a best-case scenario. It is possible to sample pairs of close kin in large wild

populations when sampling is intense, and when abiotic (or) biotic conditions favour spatiotemporal aggregation of relatives (Christie et al., 2010; Hedgecock et al., 2007; Selwyn et al., 2016). However, sampling entire family groups may be near impossible for many species. Yet, as has been shown under different experimental designs, phenotypic data from multiple related individuals, as well as sampling multiple different sets of such individuals, is necessary to increase power to estimate V_A (Fraimout et al., 2024; Klein, 1974). The absence of large sets of close kin will remain an unavoidable challenge when working with most wild organisms. However, our framework provides a means to explicitly assess power to detect non-zero heritability in any given sample of individuals, supporting more robust interpretation of heritability estimates with large confidence intervales, and expanding the scope for quantitative genomics in the wild.

Application of our framework to assess heritability estimates in a marine fish, *Bathygobius cocosensis*

In this case-study, we provide the first estimation of quantitative genetic parameters for a marine intertidal fish collected from the wild, *Bathygobius cocosensis*. Earlier work has suggested that head shape may be an ecologically important trait in *B. cocosensis* that appears to diverge among populations despite genetic homogeneity across large geographic scales (Malard et al., 2016; Thia et al., 2021). In this study, we analysed 297 fish collected from two locations in eastern Australia, Hastings Point and Heron Island, and in two years, 2016 and 2017. In this sample of fish, there was

no evidence of close kin, which is typical for marine organisms with high dispersal potential and large populations (Gagnaire & Gaggiotti, 2016).

Based on expectations obtained from our framework, we concluded that one axis of head shape morphology, phenotypic PC1, is likely moderately heritable, with a global median h^2 = 0.17 (95% CI: 0.10, 0.23). Simulated data suggested that our analysed SNPs may have captured some informative loci (likely those linked to causal variants). This conclusion is supported by the low probability of obtaining a heritability estimate of this magnitude in simulations without QTLs in the construction of **A** (Figure 6). We could not support rejection of the null hypothesis (h^2 = 0) for the other two head shape traits, phenotypic PC2 and PC3 (Table 2; Figure 6). If heritable variation does exist in these traits, it may be low and (or) our experimental design was underpowered to detect it. The variation in head shape explained by phenotypic PC2 (25.14%) and PC3 (15.11%) was much less than that for PC1 (39.54%), which may make it more difficult to quantify genetic variance with a less powerful experimental design.

Whilst phenotypic PC1 in *B. cocosensis* does appear to exhibit heritable variation, this heritability is on the lower end of the range typically reported for morphological traits in wild populations: mean h^2 = 0.56 (±0.035 SE), as reported in Postma (2014). However, those estimates were obtained across studies with access to pedigrees that were either behaviourally or genetically inferred (Postma, 2014). Extensive investigation in humans has highlighted the "missing heritability" where SNP-based

estimates (such as ours here) are lower than heritabilities estimated from other approaches, such as twin studies (Yang et al., 2010). Finally, many studies quantifying heritability may be upwardly inflated when the variance due to fixed effects accounted for when calculating V_P (the denominator in the calculation of h^2) (de Villemereuil et al., 2018; Wilson et al., 2010). Here, we included the estimated fixed effect variance in our estimate of V_P (as per de Villemereuil et al., 2018), and thus might expect our estimated heritability to be lower than studies that did not perform this correction.

Our main quantitative genomic analysis in *B. cocosensis* used fish aggregated from different spatiotemporal populations to increase sample size. We justified this because genetic structure was very low: pairwise F_{ST} ranged from 0.001 to 0.004, and PCA of SNPs in our final working dataset produced a single homogenous cloud of individuals. Nonetheless, combining individuals from different populations can bias heritability because of heterogeneous genetic architectures among different genetic backgrounds and linkage among sampled genetic markers (Goudet et al., 2018; Huang et al., 2025; Kardos & Luikart, 2021; Muff et al., 2019; Weir & Goudet, 2017). Solutions include accounting for the population structure by taking the eigenvectors of **A** (Yang et al., 2011), or deriving estimates of **A** that implicitly consider the structuring of allelic variation (Weir & Goudet, 2017), but these were not explored in this study. However, to an extent, our categorical fixed effect for the 'Population' effect may have partially acted as a proxy for genetic structure (Wolak et al., 2015). An unusual result in our analysis of *B. cocosensis* was that nearly all (91%) *observed*

estimates of heritability for phenotypic PC2 were exactly zero, which was not expected based on our *simulations* (which effectively treat the aggregated sample as coming from a single population). Such a phenomenon might occur when fixed effect groupings mirror the structuring of additive genetic variance (Wolak et al., 2015). This means that our estimate of heritability for phenotypic PC2 is likely highly conservative.

Concluding remarks

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 populations has been unclear due to challenges in sampling groups of close kin (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 (with or without kin structure), 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 to assist interested researchers in understanding our framework and applying it to their focal system using the tools available in *genomalicious*. We hope the developments made in this work help quantitative

genomics go truly wild and encourage the molecular ecology community to not only 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. 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 Hermon Slade Foundation (HSF 13/14 to CR and LL), The Ecological Society of Australia (Student Research Award 2015 and the Wiley Fundamental Ecology Award 2016 to JAT), and the Australian Research Council (IE240100151 to JAT). Major bioinformatic procedures were supported by the University of Queensland's High Performance Computing resources. Collections from Heron Island were supported by the University of Queensland's Heron Island Research Station.

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, CR, JDA and KM. JAT led the collection of wild *B. cocosensis* 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*. MJ developed the *SLiM* code for generating pairs of close kin. JAT performed the population genomic and quantitative genomic analyses in this study, created the Appendix material, and archived all data and scripts. CR, JDA, and KM provided feedback on the analyses. JAT wrote the original version of this manuscript, with all authors providing feedback on the final version.

OTHER SOFTWARE

In addition to the R packages listed in the main body of this manuscript, we also used the *data.table* (Dowle & Srinivasan, 2019), *doParallel* (Microsoft Corporation & Steve Weston, 2019), *doSNOW* (Microsoft Corporation & Weston, 2022a), foreach (Microsoft Corporation & Weston, 2022b), *ggpubr* (Kassambara, 2020), *polyRAD* (Clark et al., 2019), *tidyverse* (Wickham et al., 2019), and *WGCNA* (Langfelder & Horvath, 2008), to manipulate, analyse, and (or) visualise data. The software *fastsimcoal2* v27 was used to generate simulated data for the tutorial that accompanies this study (Excoffier et al., 2021).

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. The tutorial is also archived in the Dryad repository. 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest associated with this manuscript.

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

Attard, C. R. M., Beheregaray, L. B., & Möller, L. M. (2018). Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias. *Molecular Ecology Resources*, *18*(3), 381–390. https://doi.org/10.1111/1755-0998.12739

Beatty, W. S., Lemons, P. R., Sethi, S. A., Everett, J. P., Lewis, C. J., Lynn, R. J., Cook, G. M., Garlich-Miller, J. L., & Wenburg, J. K. (2020). Panmixia in a sea ice-associated marine mammal: Evaluating genetic structure of the Pacific walrus (*Odobenus*

rosmarus divergens) at multiple spatial scales. *Journal of Mammalogy*, *101*(3), 755–765. https://doi.org/10.1093/jmammal/gyaa050

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

Bruxaux, J., Zhao, W., Hall, D., Curtu, A. L., Androsiuk, P., Drouzas, A. D., Gailing, O., Konrad, H., Sullivan, A. R., Semerikov, V., & Wang, X. R. (2024). Scots pine – panmixia and the elusive signal of genetic adaptation. *New Phytologist*, *243*(3), 1231–1246. https://doi.org/10.1111/nph.19563

Camargo, A. (2022). PCAtest: testing the statistical significance of Principal Component Analysis in R. *PeerJ*, *10*. https://doi.org/10.7717/peerj.12967

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

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

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

de Villemereuil, P., Morrissey, M. B., Nakagawa, S., & Schielzeth, H. (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: issues and solutions. *Journal of Evolutionary Biology*, *31*(4), 621–632. https://doi.org/10.1111/jeb.13232

Dowle, M., & Srinivasan, A. (2019). *data.table: Extension of "data.frame."* https://cran.r-project.org/package=data.table

Dugand, R. J., David Aguirre, J., Hine, E., Blows, M. W., & Mcguigan, K. (2025). *The contribution of mutation and selection to multivariate quantitative genetic variance in an outbred population of* Drosophila serrata. https://doi.org/10.1073/pnas.2026217118/-/DCSupplemental

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

Endersby, N. M., Mckechnie, S. W., Ridland, P. M., & Weeks, A. R. (2006). Microsatellites reveal a lack of structure in Australian populations of the diamondback moth, *Plutella xylostella* (L.). *Molecular Ecology*, *15*(1), 107–118. https://doi.org/10.1111/j.1365-294X.2005.02789.x

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). Fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics*, *37*(24), 4882–4885. https://doi.org/10.1093/bioinformatics/btab468

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

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

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

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

Griffiths, S. P. (2003). 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

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

Haller, B. C., & Messer, P. W. (2023). SLiM 4: multispecies eco-evolutionary modeling. *The American Naturalist*, *201*(5), E127–E139.

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

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

Houle, D., & Meyer, K. (2015). Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *Journal of Evolutionary Biology*, *28*(8), 1542–1549. https://doi.org/10.1111/jeb.12674

Huang, J., Kleman, N., Basu, S., Shriver, M. D., & Zaidi, A. A. (2025). Interpreting SNP heritability in admixed populations. *Genetics*. https://doi.org/10.1093/genetics/iyaf100

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

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

Kassambara, A. (2020). *ggpubr: "ggplot2" Based Publication Ready Plots*.

Klein, T. W. (1974). Heritability and genetic correlation: statistical power, population comparisons, and sample size. *Behavior Genetics*, *4*(2), 171–189.

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

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

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

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.

Microsoft Corporation & Steve Weston. (2019). *doParallel: Foreach Parallel Adaptor for the "parallel" Package*.

Microsoft Corporation, & Weston, S. (2022a). *doSNOW: Foreach Parallel Adaptor for the "snow" Package*.

Microsoft Corporation, & Weston, S. (2022b). *foreach: Provides Foreach Looping Construct*.

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

Niu, Y., Chen, T., Wang, C., Chen, K., Shen, C., Chen, H., Zhu, S., Wu, Z., Zheng, T., Zhang, F., & Xu, J. (2021). Identification and allele mining of new candidate genes underlying rice grain weight and grain shape by genome-wide association study. *BMC Genomics*, *22*(1). https://doi.org/10.1186/s12864-021-07901-x

Popovic, I., Toews, D. P. L., Keever, C. C., St. Clair, C. T., Barbaree, B. A., Fernández, G., & Rourke, J. (2019). Genotyping-by-sequencing reveals genomic homogeneity among overwintering Pacific Dunlin (*Calidris alpina pacifica*) aggregations along the Pacific coast of North America. *Gerontologist*, *59*(4). https://doi.org/10.1093/condor/duz036

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

Postma, E. (2014). Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates? In A. Charmantier, D. Garant, & Kruuk Loeske EB (Eds.), *Quantitative Genetics in the Wild* (pp. 16–33). Oxford University Press. https://doi.org/doi.org/10.1093/acprof:oso/9780199674237.003.0002

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).

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

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

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.

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

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

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

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

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.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

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

Wolak, M. E., Roff, D. A., & Fairbairn, D. J. (2015). Are we underestimating the genetic variances of dimorphic traits? *Ecology and Evolution*, *5*(3), 590–597. https://doi.org/10.1002/ece3.1361

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Table 1. Results from genomic animal models of head shape traits in wild *Bathygobius cocosensis*. *Observed* quantitative genomic estimates include V_F, V_A, V_E, h^2, which are reported as the global medians across 80 SNP sets with 2.5% and 97.5% percentiles in parentheses. For the *simulated* data, h^2 is reported as global median heritability (for each known heritability and A-matrix scenario) across 80 SNP sets with 2.5% and 97.5% percentiles in parentheses. The percentage of log-likelihood ratio tests, %LRT, that yielded significant results ($P < 0.05$) in model comparisons between genomic animal models and null models are shown for both the *observed* and *simulated* datasets. Also shown: the percentage of *observed* point estimates that are greater than their corresponding *simulated* SNP set median estimates, %Obs>Sim; and the test for the hypothesis that these *observed* estimates are greater than their corresponding *simulated* estimates as the adjusted *P*-value for a paired Wilcox signed rank test, P_{WC}(Obs>Sim).

Trait	Obs V_F	Obs V_A	Obs V_E	Obs h^2	Obs %LRT	A-matrix	Sim level	Sim h^2	Sim %LRT	%Obs>Sim	P_{WC}(Obs>Sim
PC1	0.0024 (0.0023, 0.0024)	0.0006 (0.0003, 0.0008)	0.0005 (0.0003, 0.0007)	0.17 (0.1, 0.23)	96	With QTLs	0	0 (0, 0.19)	0	96	<0.001
							0.05	0.05 (0, 0.27)	0	81	<0.001
							0.15	0.17 (0, 0.33)	2	51	0.571
							0.3	0.34 (0.11, 0.55)	32	10	1
						Without QTLs	0	0.01 (0, 0.16)	0	98	<0.001
							0.05	0.01	0	92	<0.001

								(0, 0.25)			
							0.15	0.03 (0, 0.23)	0	88	<0.001
							0.3	0.06 (0, 0.22)	0	81	<0.001
PC2	0.0006 (0.0006, 0.0006)	0 (0, 0.0001)	0.0015 (0.0014, 0.0015)	0 (0, 0.06)	0	With QTLs	0	0 (0, 0.19)	0	8	1
							0.05	0.05 (0, 0.27)	0	2	1
							0.15	0.16 (0.01, 0.42)	5	1	1
							0.3	0.34 (0.11, 0.49)	29	1	1
						Without QTLs	0	0.01 (0, 0.19)	0	4	1
							0.05	0 (0, 0.17)	0	5	1
							0.15	0.03 (0, 0.2)	0	4	1
							0.3	0.04 (0, 0.27)	0	2	1
PC3	0.0003 (0.0003, 0.0003)	0 (0, 0.0002)	0.0009 (0.0007, 0.0009)	0.03 (0, 0.19)	0	With QTLs	0	0 (0, 0.21)	0	56	0.001
							0.05	0.08 (0, 0.29)	0	34	0.998
							0.15	0.15 (0, 0.36)	2	24	1

	0.3	0.32 (0.1, 0.5)	29	5	1
Without QTLs	0	0 (0, 0.14)	0	51	0.002
	0.05	0 (0, 0.22)	0	51	0.062
	0.15	0.02 (0, 0.25)	0	45	0.431
	0.3	0.05 (0, 0.27)	0	49	0.896

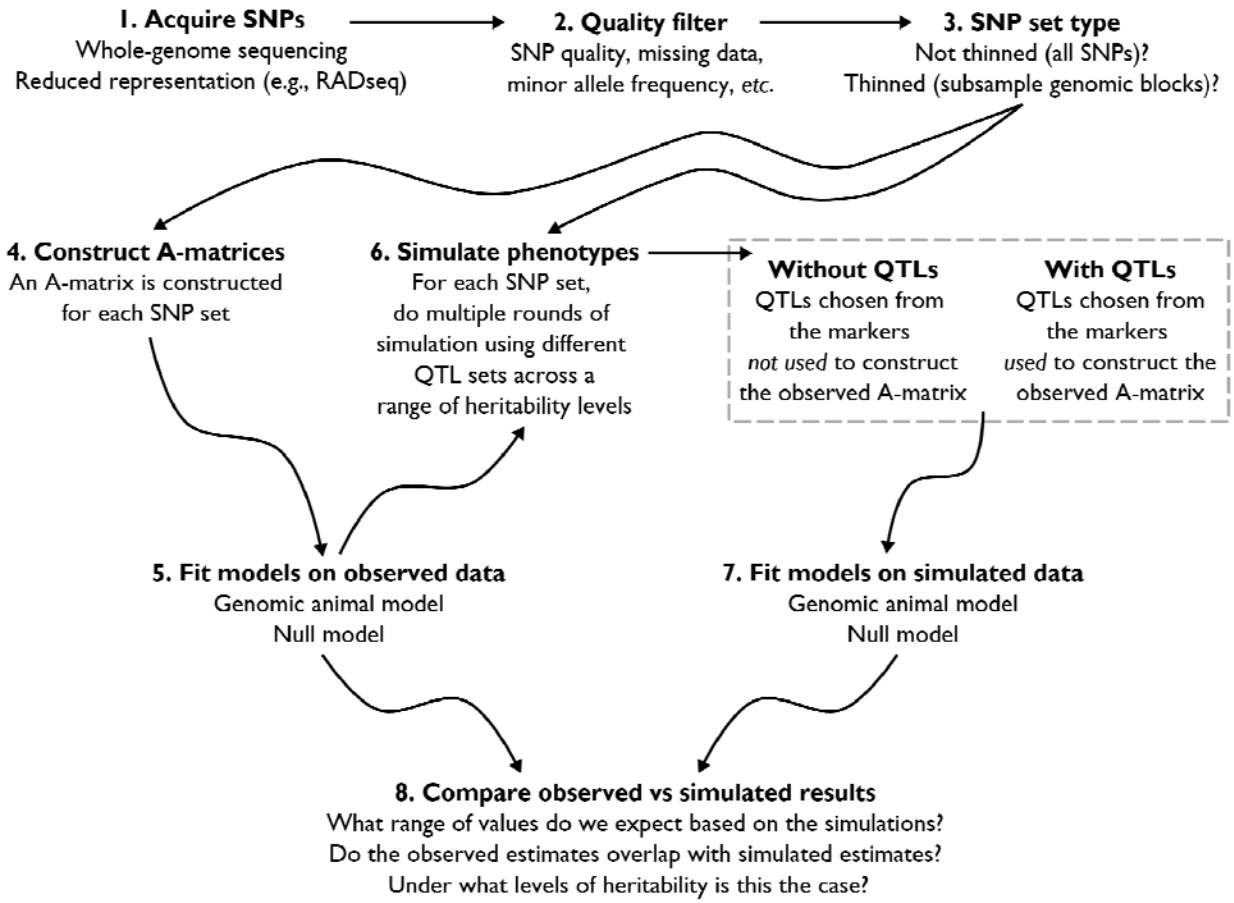

Figure 1. Conceptual verbal model of our resampling-simulation framework. Steps are as follows: (1) Acquire SNPs from across the genome. (2) Quality filter SNPs. (3) Specify the SNP set type: will all SNPs be used (not thinned) or will they be subsampled by linkage blocks (thinned)? (4) Construct A-matrices for the observed SNP sets. If using all SNPs, there will be a series of A-matrices. If a series of randomly resampled thinned SNP sets are uses, there will be a corresponding series of A-matrices. (5) Fit models for the observed data, predicting observed phenotypes with (genomic animal model) or without (null model) the observed A-matrix (not thinned) or series of A-matrices (thinned). (6) Simulate phenotypes for each SNP set. Simulation is performed by resampling a series of SNPs to

act as nested quantitative trait loci (QTLs) sets. (7) Fit models for the simulated data, predicting simulated phenotypes with (genomic animal model) or without (null model) the corresponding observed A-matrix (not thinned) or series of A-matrices (thinned). (8) Compare the observed results against the simulated results.

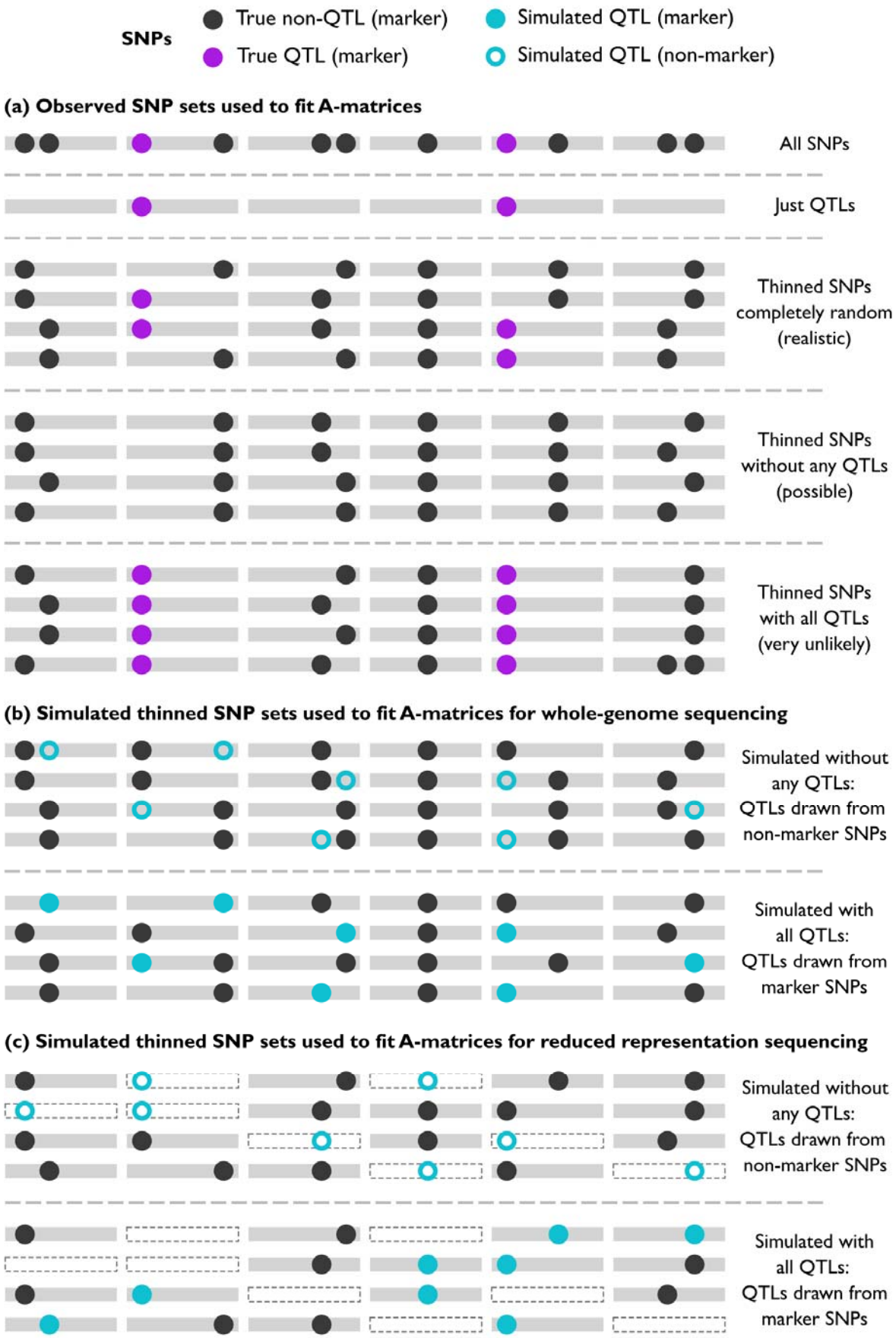
SNPs
True non-QTL (marker)
True QTL (marker)
Simulated QTL (marker)
Simulated QTL (non-marker)
(a) Observed SNP sets used to fit A-matrices
All SNPs
Just QTLs
Thinned SNPs completely random (realistic)
Thinned SNPs without any QTLs (possible)
Thinned SNPs with all QTLs (very unlikely)
(b) Simulated thinned SNP sets used to fit A-matrices for whole-genome sequencing
Simulated without any QTLs: QTLs drawn from non-marker SNPs
Simulated with all QTLs: QTLs drawn from marker SNPs
(c) Simulated thinned SNP sets used to fit A-matrices for reduced representation sequencing
Simulated without any QTLs: QTLs drawn from non-marker SNPs
Simulated with all QTLs: QTLs drawn from marker SNPs

Figure 2. Schematic of observed and simulated SNP sets.

(a) Different ways in which observed SNP sets can be defined and how this dictates how many biologically informative genetic markers are used in the construction of **A** (the genomic relationship matrix). Grey bars represent sequenced genomic regions: they could represent linkage blocks across the genome (*whole-genome sequencing*), or short contigs dispersed throughout the genome (*reduced-representation sequencing*). Black circles represent sampled marker loci that are true non-QTLs, and purple circles represent sampled marker loci that are true QTLs. Each row represents a different SNP set. In the 'All SNPs' scenario, all loci are used as genetic markers, so we can only have one possible SNP set; this is a strategy that could be used in any study. In the 'Just QTLs' scenario, only loci that are QTLs are used as genetic markers; this is a strategy that could be used in studies with candidate loci. The 'Thinned SNPs' scenarios represent cases where genetic markers are drawn randomly from genomic blocks. For the 'Thinned SNPs completely random' scenario, sampled genetic markers may or may not capture true QTLs, which is a realistic outcome in for studies with dense genotyping. For the 'Thinned SNPs without any QTLs' scenario, sampled genetic markers do not capture any true QTLs, which is a possible outcome in studies with sparse genotyping, or simply due to random chance. For the 'Thinned SNPs with all QTLs' scenario, sampled genetic markers completely capture all the true QTLs, which is a highly unlikely outcome for a thinned dataset if thinning occurs at random.

(b) Two contrasting approaches for simulating phenotypes with thinned SNP sets when using *whole-genome sequencing*. Observed SNPs are thinned to keep one locus per linkage

block, and we thus classify SNPs as marker loci used in construction of A-matrices and non-marker loci that are not used. In the 'Without any QTLs' simulations, only non-marker SNPs are simulated as QTLs. In the 'With all QTLs' simulations, only marker SNPs are simulated as QTLs. Black circles represent sampled marker loci that are not simulated as QTLs, solid blue circles represent sampled marker loci simulated as QTLs, open blue circles represent non-sampled SNPs that are simulated as QTLs but are in the same linkage block as sampled marker loci.

(c) Two contrasting approaches for simulating phenotypes with thinned SNP sets when using *reduced-representation sequencing*. This subplot is similar to (b), however, we suggest an additional step during SNP set generation in which SNPs are derived from two sets of contigs: (1) SNPs from 'focal' contigs are used in both the construction of A-matrices and the 'With all QTLs' simulations, and (2) SNPs from 'reserved' contigs are not used in the construction of A-matrices but are used in the 'Without any QTLs' simulations. Whereas in whole-genome sequencing, we aim to cover the entire genome, in reduced representation sequencing, only a fraction of the genome is covered. Hence, a more realistic scenario is that true QTLs are not captured by reduced representation contigs, but are in some region of the genome we have not been sequenced. The 'reserved' contigs, in essence, mimic the 'unseen' regions of the genome missed by reduced representation sequencing. Open bars with dashed outlines represent these 'reserved' contigs.

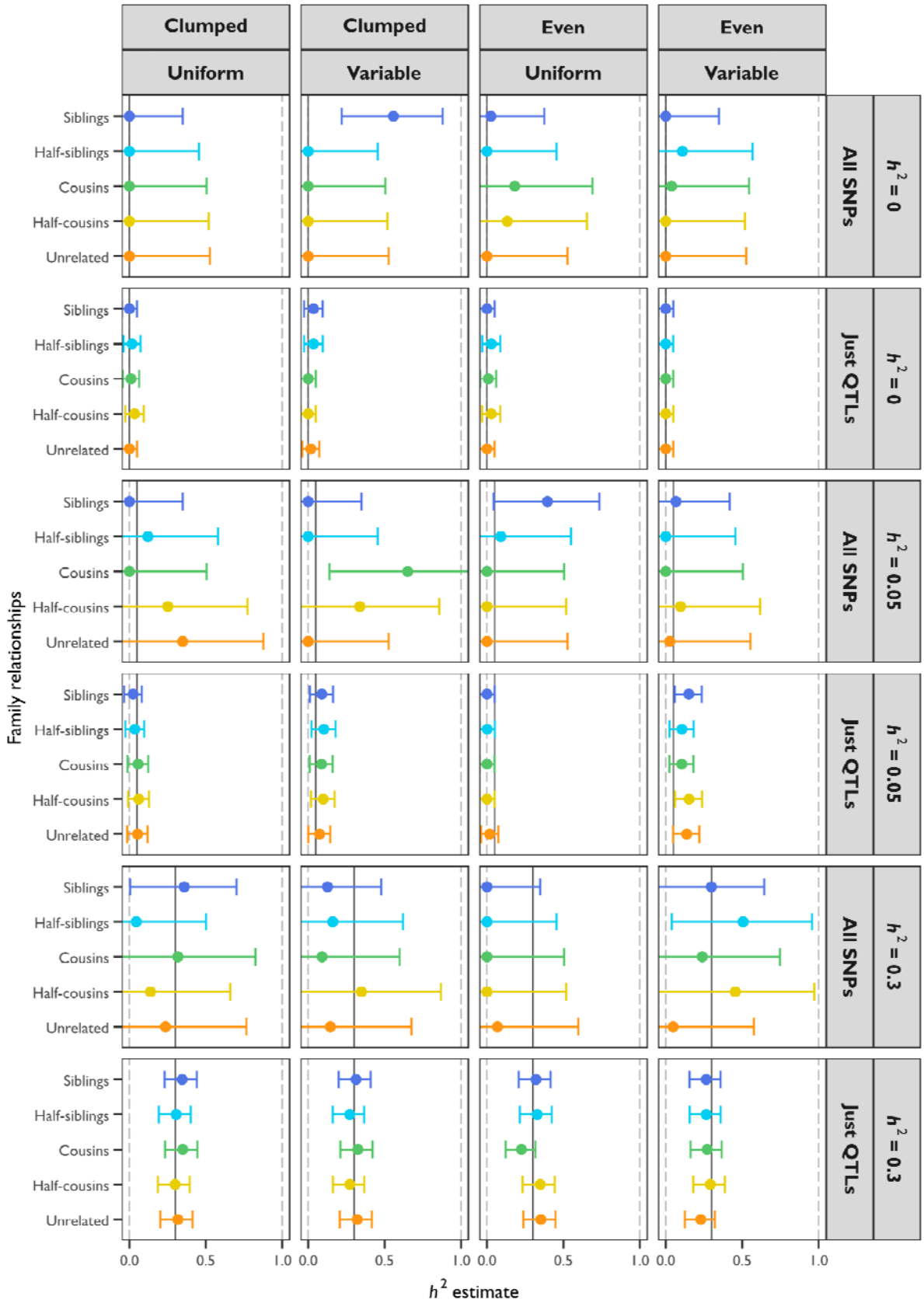
Clumped
Uniform
Clumped
Variable
Even
Uniform
Even
Variable
$h^2 = 0$
All SNPs
$h^2 = 0$
Just QTLs
$h^2 = 0.05$
All SNPs
$h^2 = 0.05$
Just QTLs
$h^2 = 0.3$
All SNPs
$h^2 = 0.3$
Just QTLs
Family relationships
Siblings
Half-siblings
Cousins
Half-cousins
Unrelated
0.0
0.5
1.0
h^2 estimate

Figure 3. *SLiM*ulated population heritability estimates when the true heritability ranged from 0 (top panels) to 0.3 (bottom panels), and **A** was estimated using the 'All SNPs' and 'Just QTLs' marker choice strategies for different forms of linkage among QTLs (clumped vs evenly distributed through genome) and effect sizes of QTL (uniform or variable). Within each panel, the *x*-axis represents the observed heritability estimate and the *y*-axis represents the different kin-structures analysed. Points represent the point heritability estimate with REML-MVN 90% confidence intervals. Solid vertical lines indicate the known level of heritability, and dashed vertical lines indicate the bounds of 0 and 1. Where the lower REML-MVN CI does not include zero, statistical support for non-zero heritability is inferred; where the lower CI is negative, the plot is truncated by the lower bound of the x-axis to aid visualisation of the values within the parameter space (Dugand et al., 2025). Panels represent different combinations of marker choice and known heritability (in rows) vs genomic distribution of QTLs and effects sizes (in columns).

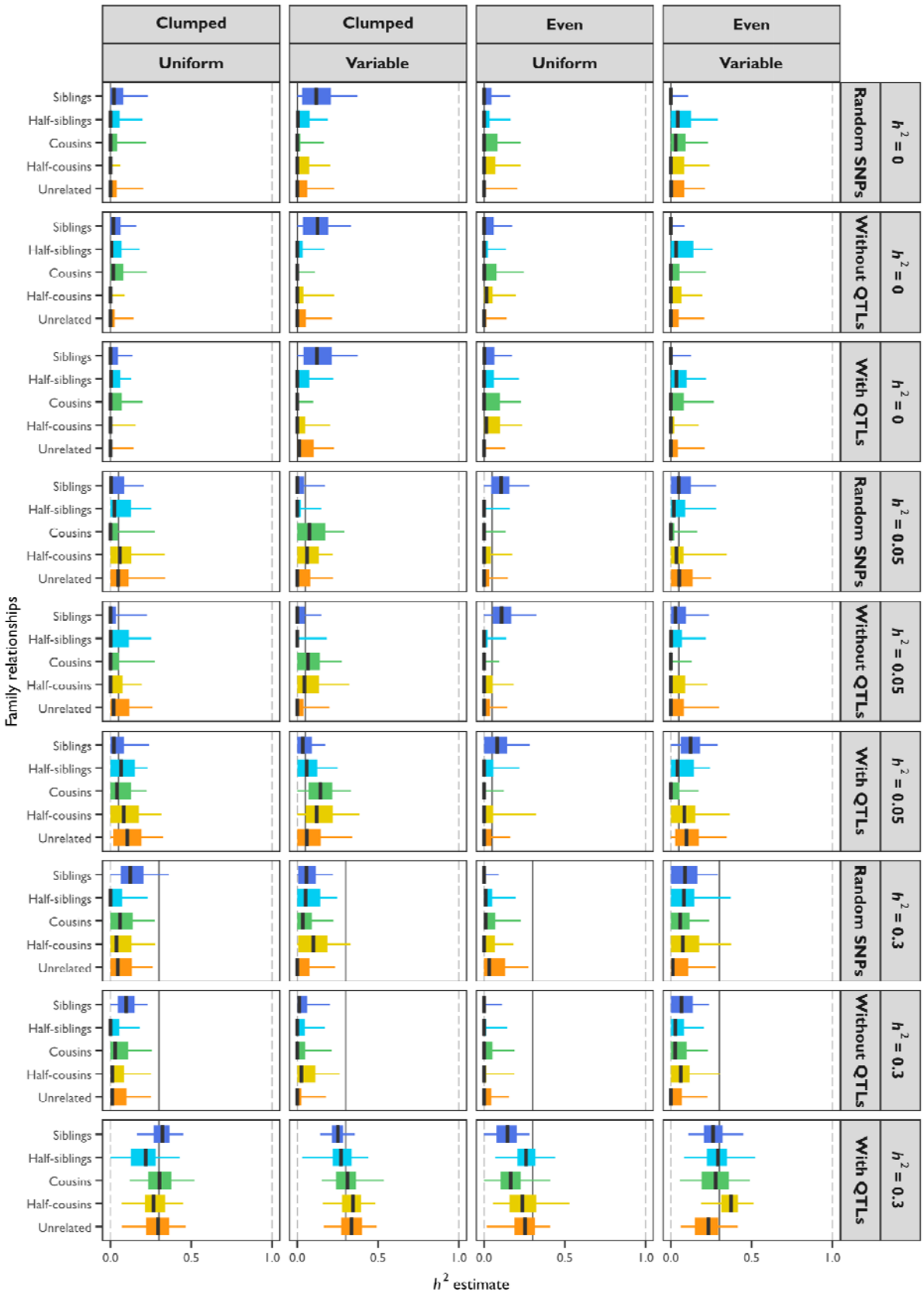
Clumped
Uniform
Clumped
Variable
Even
Uniform
Even
Variable
Random SNPs
Without QTLs
With QTLs
$h^2 = 0$
$h^2 = 0.05$
$h^2 = 0.3$
Siblings
Half-siblings
Cousins
Half-cousins
Unrelated
Family relationships
0.0
0.5
1.0
h^2 estimate

Figure 4. *SLiM*ulated population heritability estimates when the true heritability ranged from 0 (top panels) to 0.3 (bottom panels), and **A** was estimated using one of three resampled thinned marker choice strategies: 'Random SNPs', 'Without QTLs', and 'With QTLs' for different forms of linkage among QTLs (clumped vs evenly distributed through genome) and effect sizes of QTL (uniform or variable). The *x*-axis represents the observed heritability estimate and the *y*-axis represents the different kin-structures analysed. Boxes represent the 25%, 50% (median) and 75% percentiles of resampled heritability estimates, and whiskers represent the 2.5% and 97.5% percentiles. Solid vertical lines indicate the true level of heritability, and dashed vertical lines indicate the bounds of 0 and 1. Panels represent different combinations of marker choice and heritability (in rows) vs genomic distribution of QTLs and effects sizes (in columns).

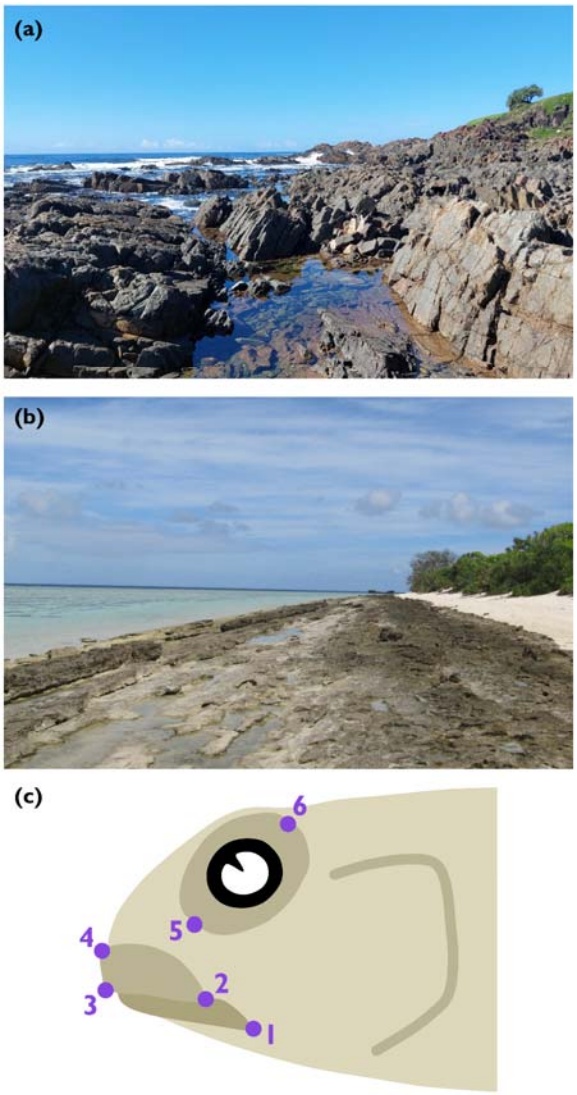

Figure 5. Design for our study of heritability of head shape phenotypes in Australian populations of *Bathygobius cocosensis*. (a) Hastings Point, a complex rocky intertidal habitat patch in (Photo credit: Joshua Thia). (b) Heron Island, a simple rocky intertidal habitat patch (Photo credit: Joshua Thia). (c) Morphological landmarks describing head shape variation in *B. cocosensis*.

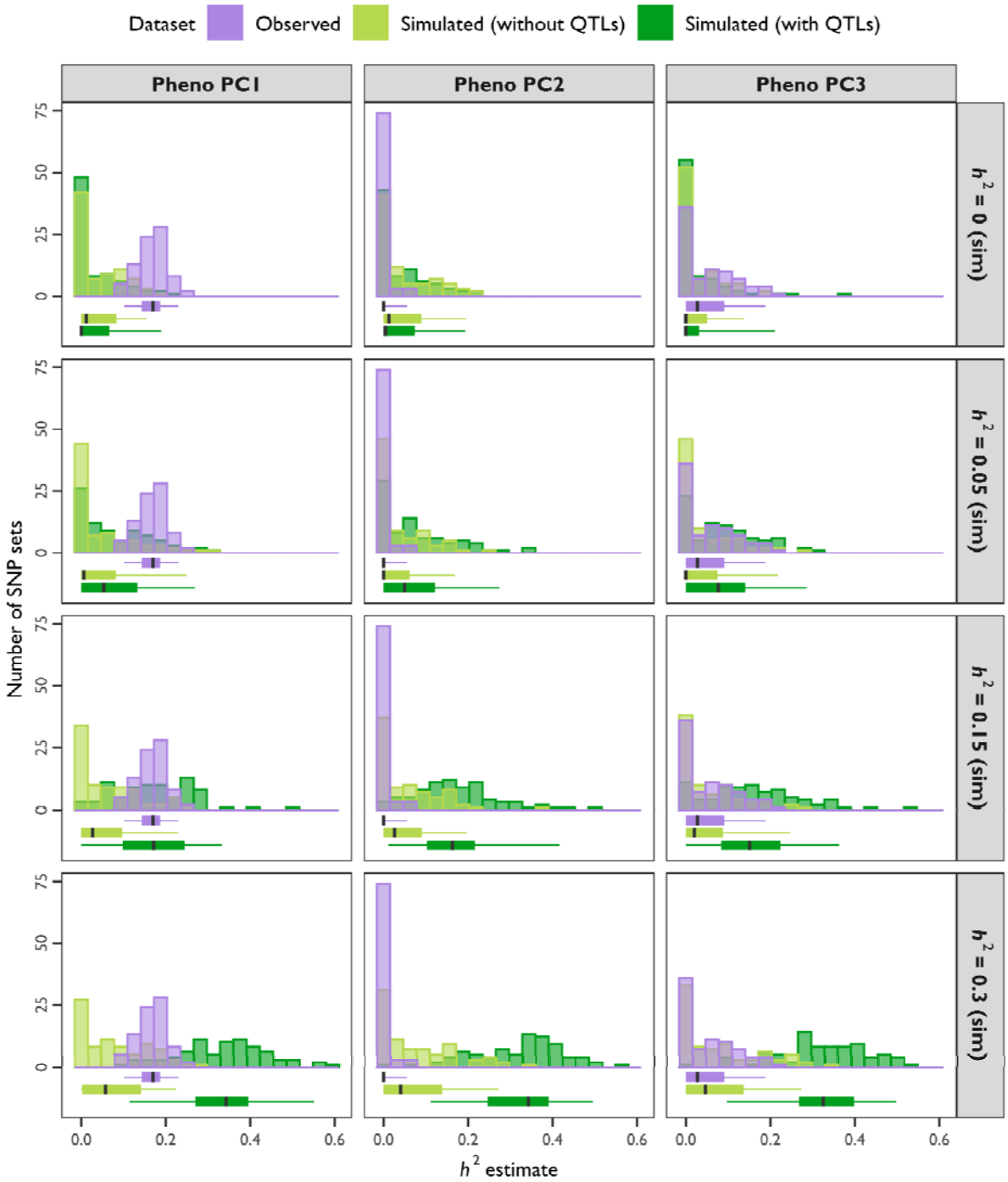

Figure 6. Application of our genomic framework to study the heritability of head shape phenotypes in Australian populations of *Bathygobius cocosensis*. The *x*-axis represents the estimated heirtabiltiy: for the *observed* data, this is the SNP set point estimate of heritability; for the *simulated* data, this is the SNP set median estimate of heritability (taken

across the 5 QTL sets nested in each SNP set). The *y*-axis represents the number of SNP sets. Bars depict the frequency distribution of estimates. Boxes below the *x*-axis delimit the 25% and 75% percentiles, the median is marked with a grey dash, and whiskers delimit the 2.5% and 97.5% percentiles as confidence intervals. Frequency distributions and boxplots are coloured by dataset (see legend). Panels represent different simulated levels of heritability (in rows) for each trait (in columns).