Check for updates

Received: 26 June 2020 | Revised: 12 January 2021 | Accepted: 28 January 2021

DOI: 10.1111/mec.15829

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genetic and phenotypic variation exhibit both predictable and stochastic patterns across an intertidal fish metapopulation

Joshua A. Thia[1,2] | Katrina McGuigan[1] | Libby Liggins[3] | Will F. Figueira[4] | Christopher E. Bird[5] | Andrew Mather[1] | Jennifer L. Evans[1] | Cynthia Riginos[1]

[1]School of Biological Sciences, The University of Queensland, Saint Lucia, QLD, Australia

[2]School of BioSciences, The University of Melbourne, Melbourne, VIC., Australia

[3]School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

[4]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

[5]Department of Life Sciences, Texas A&M University Corpus Christi, Corpus Christi, TX, USA

Correspondence
Joshua A. Thia, School of Biological Sciences, The University of Queensland, Saint Lucia, Qld, Australia.
Email: josh.thia@live.com

Funding information
Harman Slade Foundation; Ecological Society of Australia; Goodman Foundation

Abstract

Interactions among selection, gene flow, and drift affect the trajectory of adaptive evolution. In natural populations, the direction and magnitude of these processes can be variable across different spatial, temporal, or ontogenetic scales. Consequently, variability in evolutionary processes affects the predictability or stochasticity of microevolutionary outcomes. We studied an intertidal fish, *Bathygobius cocosensis* (Bleeker, 1854), to understand how space, time, and life stage structure genetic and phenotypic variation in a species with potentially extensive dispersal and a complex life cycle (larval dispersal preceding benthic recruitment). We sampled juvenile and adult life stages, at three sites, over three years. Genome-wide SNPs uncovered a pattern of chaotic genetic patchiness, that is, weak-but-significant patchy spatial genetic structure that was variable through time and between life stages. Outlier locus analyses suggested that targets of spatially divergent selection were mostly temporally variable, though a significant number of spatial outlier loci were shared between life stages. Head shape, a putatively ecologically responsive (adaptive) phenotype in *B. cocosensis* also exhibited high temporal variability within sites. However, consistent spatial relationships between sites indicated that environmental similarities among sites may generate predictable phenotype distributions across space. Our study highlights the complex microevolutionary dynamics of marine systems, where consideration of multiple ecological dimensions can reveal both predictable and stochastic patterns in the distributions of genetic and phenotypic variation. Such considerations probably apply to species that possess short, complex life cycles, have large dispersal potential and fecundities, and that inhabit heterogeneous environments.

KEYWORDS
chaotic genetic patchiness, complex life cycles, geometric morphometrics, marine organisms, population genomics, spatiotemporal variation

1 | INTRODUCTION

The balance between gene flow and selection is fundamental to adaptive divergence among populations and species. When migration maintains a continual influx of unfit alleles that overwhelms local adaptive alleles, populations can exist in maladaptive states through a process of gene swamping (Garant et al., 2007; Nosil, 2009; Rolshausen et al., 2015). Yet amidst high gene flow, adaptive divergence may occur provided selection is strong (Dennenmoser et al., 2017; Hanski et al., 2010; Hoekstra et al., 2004; Moody et al., 2015; Schmidt & Rand, 1999). High gene flow may also favour the evolution of phenotypic plasticity to buffer genotype–environment mismatches in dispersive taxa (Scheiner, 1993; Sultan & Spencer, 2002). However, highly plastic strategies can preclude local

 © 2021 John Wiley & Sons Ltd

adaptation because locally fit phenotypes do not have a heritable genetic basis (Bulmer, 1972; Hanski et al., 2010; Sultan & Spencer, 2002). Therefore, predicting how adaptive traits evolve in high gene flow systems is challenging because a range of possible outcomes can manifest (Crispo, 2008; Garant et al., 2007).

General understanding of gene flow–selection interactions in natural populations is further complicated by heterogeneity of ecological systems and the specifics of species biology. For example, gene flow can oscillate through time, or be restricted to specific life stages, affecting the spatial distribution of genetic variation over time (Di Franco et al., 2012; Hogan et al., 2010, 2012; Moody et al., 2015; Schmidt & Rand, 1999; Watson et al., 2012). Spatiotemporal heterogeneity in selection dictates when and where certain alleles (or trait values) are adaptive (Aspi et al., 2003; Grant & Grant, 2002; Paccard et al., 2018; Stratton & Bennington, 1998). Moreover, as organisms progress through their ontogeny (their development trajectory), it is possible that the selective agents acting on them might also change. Such ontogenetic shifts in selection pressures can alter the fitness effects of alleles at different life stages (Diamond et al., 2019; Ebenman, 1992; Gagliano et al., 2007; Moran, 1994). Interpreting patterns of genetic variation at a single point in time, in a particular spatial context, or in a specific life stage therefore requires considerations of the variability and strength of gene flow and (or) selection. That is, are genetic patterns predictable across space, time, or life stage, and what factors determine when and where genetic differentiation emerges, persists, or is eroded?

Gene flow–selection interactions take place in the context of finite populations and are therefore additionally subject to the influence of genetic drift. Local adaptation through heritable genetic variation is less likely in smaller populations because stochastic sampling of the gene pool (high drift) reduces additive genetic variance (Blanquart et al., 2012; Whitlock, 1999). Turnover in local genetic composition due to drift might also lead to temporally variable genetic architectures underpinning adaptive traits because locally beneficial alleles and co-adapted complexes are randomly lost (Yeaman, 2015; Yeaman & Whitlock, 2011). Conversely, local adaptation in finite populations may benefit from gene flow because lost genetic variation can be replaced by migration (Blanquart et al., 2012). However, levels of local adaptation will always be reduced at equilibrium in smaller populations (Blanquart et al., 2012), which have lower critical migration thresholds for gene swamping (Yeaman & Otto, 2011; Yeaman & Whitlock, 2011). Therefore, when drift is high, genetic architectures comprising large-effect loci are more likely to contribute to adaptation and remain stable through time (Yeaman & Otto, 2011; Yeaman & Whitlock, 2011).

Given these complex interactions and theoretical expectations, empirical investigation of adaptive differentiation requires documenting which ecological dimensions (e.g., space, time, life stage) are important in structuring variation and their relevant scales (e.g., short or long geographic distances, year-to-year variability, juveniles versus adults). Quantifying variability in patterns of intraspecific variation across these dimensions can provide insight as to whether the processes driving microevolutionary change are predictable or stochastic. Such an approach might be particularly useful in marine systems. For many marine species dispersal is restricted to the early life stages, and demersal adults occupy patchy habitats, characteristic of metapopulations (Gaggiotti, 2017). Marine metapopulations are expected to have large effective migration rates (Jones et al., 2009; Waples, 1998), whereby large effective population sizes, and (or) extensive pelagic dispersal, maintain high genetic diversity and low differentiation among subpopulations. Yet many marine species exhibit weak, but statistically significant, genetic structuring; a pattern that is inconsistent with true panmixia (e.g., Gould & Dunlap, 2017; Hogan et al., 2012; Jackson et al., 2017; Moody et al., 2015). This weak-but-significant genetic structure can also be spatially uncorrelated and the magnitude of genetic structure may exhibit temporal fluctuations (Jackson et al., 2017; Selwyn et al., 2016; Toonen & Grosberg, 2011), a pattern described as "chaotic genetic patchiness" (sensu Johnson & Black, 1982).

Chaotic genetic patchiness is hypothesised to arise from variability in mechanisms that can be neutral or selective in nature. Neutral mechanisms typically refer to processes affecting dispersal trajectories, migration rates, relatedness of individuals within subpopulations, and genetic drift. For example, spatiotemporal variability in oceanic currents determines the presence, frequency, and intensity of kin aggregation, local retention, and asymmetric dispersal (Buston et al., 2009; Cuif et al., 2015; Gerlach et al., 2007; Selwyn et al., 2016; Watson et al., 2012; White et al., 2010; Yearsley et al., 2013). Furthermore, "sweepstakes reproduction" has been used to describe substantial genetic drift that can emerge from reproductive variance in marine organisms, a product of their high fecundity and stochastic juvenile mortality (Hedgecock, 1994; Hedgecock & Pudovkin, 2011). Whereas neutral mechanisms of chaotic genetic patchiness typically occur early in life, strong selective mechanisms can occur across all life stages for demersal marine taxa. Deterministic (nonrandom) mortality of new recruits with specific phenotypes and (or) alleles caused by selection may reduce gene flow, causing genetic structure to arise over spatial scales presumably smaller than a species' expected dispersal distance (Appelbaum et al., 2002; Johnson & Black, 1982; Marshall et al., 2010; Schmidt & Rand, 1999). Additionally, changes in selection through time or across life stages can cause shifts in genetic composition, for example, between years or over recruitment (Ciotti & Planes, 2019; Gould & Dunlap, 2017; Villacorta-Rath et al., 2018). Because the effects of neutral and selective mechanisms can vary across ontogeny in demersal marine species, comparisons between earlier life stages and recruited adults can provide insights into the potential roles of gene flow, selection, and drift (Christie et al., 2010; Pujolar et al., 2015; Schmidt & Rand, 1999, 2001; Toonen & Grosberg, 2011; Villacorta-Rath et al., 2018).

Whereas patterns of genome-wide variation are used to infer the contribution of neutral mechanisms to chaotic genetic patchiness (Table 1), insight into the action of selection comes from locus-specific analyses (Gould & Dunlap, 2017; Schmidt & Rand, 1999; Villacorta-Rath et al., 2018) and the observation of relevant phenotypic variation. Marine metapopulations are unlikely to be at equilibrium, that is, the point where genetic variation is stable due

TABLE 1 Three population genetic scenarios (panmixia, local retention, or chaotic genetic patchiness), associated genetic expectations for neutral genetic variation, and observations from *Bathygobius cocosensis*

Genetic expectation	Panmixia	Local retention	Chaotic genetic patchiness	References	Supporting evidence
Spatial genetic structure	None (✗)	High (✗)	Present, but F_{ST} is small and spatially uncorrelated (✓)	[1, 2, 3]	Figure 2
Temporal variability in spatial genetic structure	None (✗)	None (✗)	Present (✓)	[2, 3]	Figure 2
Life stage differences in spatial genetic structure	None (✗)	None (✗)	Present (✓)	[4, 5]	Figure 2
Divergence between years within a site	None (✗)	None (✗)	Present (✓)	[3, 6]	Figure 3
Divergence between life stages within a site	None: juveniles equally similar to local and nonlocal adults (✗)	None: juveniles are most similar to local adults and dissimilar from nonlocal adults (✗)	Present: unclear relationships between juveniles with local versus nonlocal adults (✓)	[6, 7]	Figure S2.1
Genetic diversity in adults >juveniles, due to sweepstakes reproduction	None (✗)	None (✗)	Present (✓)	[4, 7]	Figure S2.2

Note: A tick (✓) indicates an observation supported by our data in *B. cocosensis*, whereas a cross (✗) indicates lacking evidence.

[1]Gerlach et al. (2007), [2]Jackson et al. (2017), [3]Johnson and Black (1982), [4]Christie et al. (2010), [5]Moody et al. (2015), [6]Toonen and Grosberg (2011), [7]Hedgecock and Pudovkin (2011).

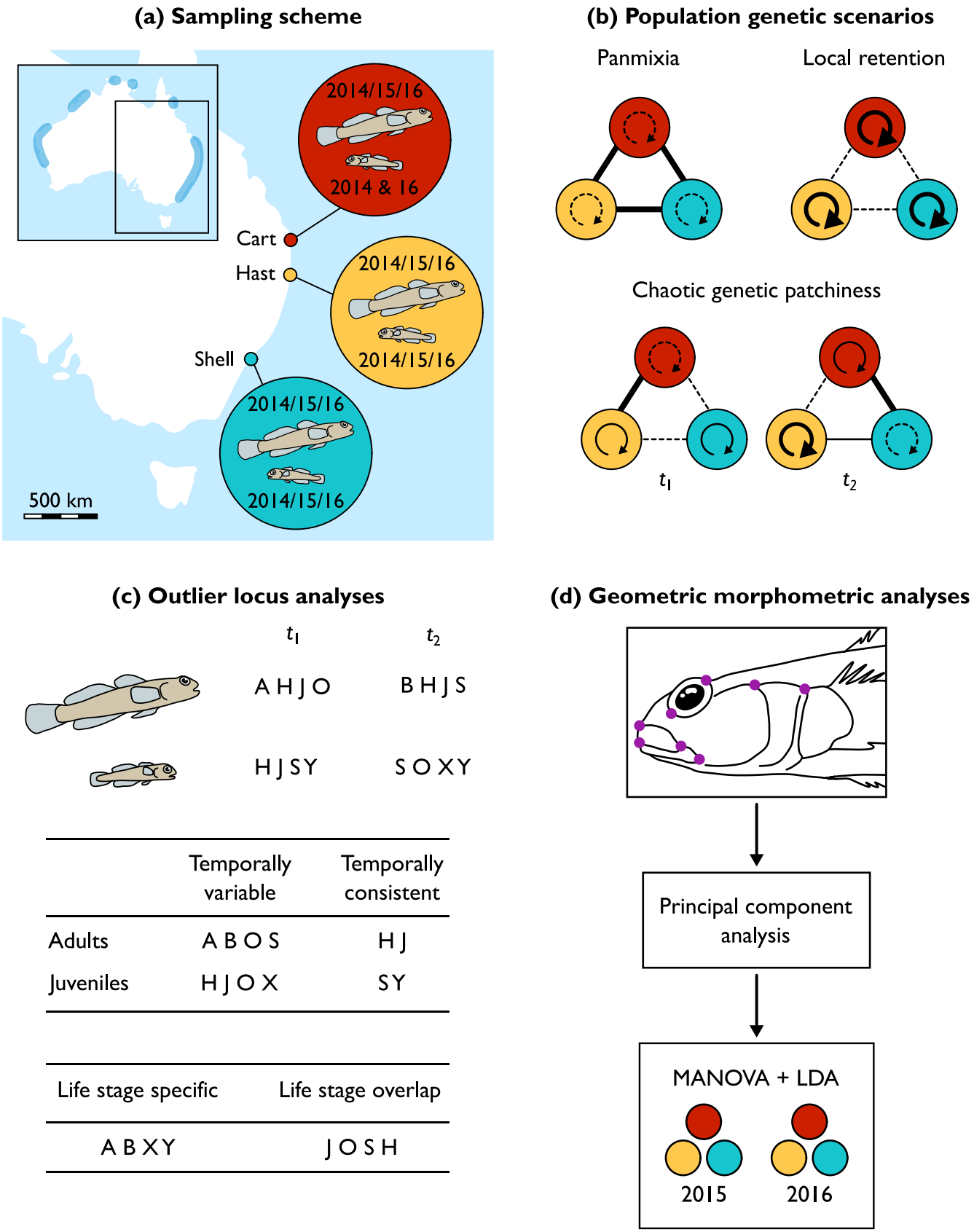
to the balance between drift, selection, and gene flow. Indeed, the potential non-equilibrium dynamics of marine metapopulations, caused by heterogeneous selective pressures, variable gene flow, and stochastic demography, probably pose challenges for most standard methods of identifying adaptive loci. Therefore, methods that relax equilibrium assumptions would help to infer the influence of selection on stochastic genetic patterns. Furthermore, even though the genetic architecture of adaptation can be temporally variable (Yeaman, 2015; Yeaman & Otto, 2011), recurrent observation of the same locus exhibiting extreme differentiation (relative to the genomic background) through time, or across life stages, would provide greater confidence that the locus is affected by selective processes.

Characterisation of phenotypic differences among subpopulations can also provide insights into divergent selection (Cohen & Dor, 2018; Galligan et al., 2012; Hoekstra et al., 2004). Weldon's studies of dispersive marine crustaceans are a classical example of such inference where he posited that differences in the mean and variance in morphological traits across different sites and life stages could be explained by deterministic mortality (Weldon, 1894, 1895). Even if the underlying genetic loci responding to selection cannot be identified, observation of phenotype–environment correlations amidst a homogenous genomic background provides strong evidence for the action of heritable and (or) plastic responses to local conditions. A compelling contemporaneous example is provided by Moody and colleagues, who showed that predictable habitat-specific morphologies can diverge in the presence of high, but stochastic, gene flow in the amphidromous goby, *Sicyopterus stimpsoni* (Moody et al., 2015, 2019). Hence, insight into how selection might generate spatiotemporal genetic patterns in marine metapopulations would benefit from a more holistic approach, one that combines temporally replicated sampling of different life stages, analysis of outlier loci, and measures of ecologically relevant phenotypes.

The overarching goal of our study was to assess the predictability in neutral and selective processes that affect the distribution of genetic and phenotypic variation in well-connected marine metapopulations. We focus on *Bathygobius cocosensis* (Bleeker, 1854), a demersal intertidal fish, as a representative of marine species with a complex life cycle, potentially high gene flow, and that occupies heterogeneous environments (Figure 1a). This fish has a wide distribution throughout the Indo-Pacific. In Australia, *B. cocosensis* inhabits rocky coasts throughout tropical and temperate waters (Figure 1a) (Atlas of Living Australia, 2018; Griffiths, 2003; Malard et al., 2016; da Silva et al., 2019; White et al., 2014). Large recruitment events occur in summer (January and February), but this species may breed and recruit throughout the year (Joshua Thia, personal observation). Eggs are laid and fertilised on the benthos. After hatching, larvae undergo a ~20–25 day pelagic phase before metamorphosing into juvenile fish and settling in intertidal and shallow subtidal habitats (da Silva, Wilson, et al., 2019; Thia et al., 2018).

Bathygobius cocosensis has large physiological tolerances (da Silva et al., 2019) and environmentally variable phenotypes (Carbia & Brown, 2019, 2020; Carbia et al., 2020; Malard et al., 2016); thus this species presents an interesting system within which to study

microevolutionary processes. We focus on head shape morphology as a phenotypic trait with potential ecological relevance (Figure 1d). In teleost fishes, head shape morphology is related to feeding and navigational adaptations (Burress et al., 2017; Schluter, 1993; Vera-Duarte et al., 2017). Additionally, prior work has shown microhabitat partitioning of head shape phenotypes in *B. cocosensis* (over tens of



	Temporally variable	Temporally consistent
Adults	A B O S	H J
Juveniles	H J O X	S Y

Life stage specific	Life stage overlap
A B X Y	J O S H

FIGURE 1 Study schematic. Pictorial representation of our study design quantifying genetic and phenotypic differentiation in *Bathygobius cocosensis*. (a) Spatial and temporal sampling on Australia's east coast: Cart = Point Cartwright; Hast = Hastings Point; and Shell = Shellharbour. Large and small fish icons represent adults and juveniles, and the numbers above and below them (respectively) represent collected sampling years. Inset: shaded areas illustrate recorded locations of *B. cocosensis* in Australia (Atlas of Living Australia, 2018) (b) Genome-wide pool-seq SNP data was used to discern between three population genetic scenarios (Table 1). Coloured circles represent subpopulations. Lines between subpopulations indicate gene exchange, whilst loops within subpopulations indicate local recruitment; intensity of these lines and loops indicates their magnitude. Different time points are represented by t_1 and t_2. (c) SNPs were used to test for genetic signatures of spatially divergent selection in different time points (t_1 and t_2), in adult or juvenile life stages (big and small fish). In this hypothetical example, capital letters (A, B, … etc.) represent loci identified as spatial outliers. Comparisons across time provided information on temporal consistency in outlier loci, whereas comparisons between life stages provided insight into which outlier loci might exhibit spatially divergent patterns across ontogeny. (d) Adult head shape morphology was used to quantify phenotypic variation in a geometric morphometric framework. Morphological landmarks (capturing variation in jaw shape, eye size and orientation, and position of the preopercula and opercula) were recorded, and variation in their position summarised in a principal component analysis (PCA). Individual fish scores on the PC axes were used as the response variables in a multivariate analysis of variance (MANOVA). Differences in head shape among sites and years were then visualised using linear discriminant analysis (LDA)

meters) (Malard et al., 2016). Our working hypothesis is that head shape is adaptive and responds to selection within and between subpopulations.

We used a combination of genomic and phenotypic data to: (i) infer the population genetic pattern characterising the *B. cocosensis* metapopulation (Figure 1b; Table 1), (ii) test whether spatial selection among subpopulations was predictable over time (Figure 1c; Table 2), (iii) evaluate the importance of presettlement (pelagic larval) versus post-settlement processes for structuring genetic variation (Figure 1b,c; Table 3), and (iv) determine whether patterns of post-settlement phenotypic differentiation were concordant with patterns of genetic variation (Figure 1d; Table 4). Our study represents one of the most comprehensive attempts to coanalyse multiple ecological dimensions (space, time, and life stage) and quantify their contributions to predictable or stochastic structuring of intraspecific variation in marine metapopulations.

2 | MATERIALS AND METHODS

2.1 | Sampling

We sampled three sites for *B. cocosensis*: Port Cartwright (S26.68° E153.14°, Queensland), Hastings Point (S28.36° E153.58°, New South Wales), and Shellharbour (S34.58° E150.87°, New South Wales), which span ~950 km and 8° latitude (Figure 1a). These locations occur in different climatic zones: Point Cartwright and Hastings Point are subtropical, whilst Shellharbour is more temperate. We cosampled juvenile (≤20 mm standard length) and adult (≥25 mm standard length) *B. cocosensis* at each site in three consecutive years: 2014, 2015, and 2016 (Figure 1a; Table S2.1). However, we were unable to obtain juvenile samples from Point Cartwright in 2015. Fish were collected using hand nets and euthanized with eugenol at 100 mg/L in seawater before being preserved in 100% ethanol.

Little is known about the demographic rates of *B. cocosensis*. We have observed gravid females with a standard length of ~25 mm (Joshua Thia, personal observation), which may be the lower size limit of reproductive maturity. The largest fish typically range from ~50 to 60 mm. In our study, we considered "juveniles" as fish <20 mm in standard length, and "adults" as fish >25 mm in standard length. Based on preliminary otolith data and other life history investigations (Thia et al., 2018), we inferred a growth rate of ~5 days post-settlement per 1 mm increase in standard length (Methods S1). This rate therefore suggests that *B. cocosensis* with standard lengths of 20, 25, and 60 mm might respectively be ~62, ~85, and ~250 days post-settlement. Thus, the largest fish in the population are probably <1 year old, meaning that local *B. cocosensis* subpopulations exhibit considerable year-to-year turnover (Table S1.1). Consequently, fish sampled in different years probably comprise different cohorts.

Note, we refer to our system as a "metapopulation", that is, a "patchy population" comprised of multiple structured subpopulations. However, we do not specifically imply that this system experiences classic metapopulation dynamics with extinction–colonisation cycles (i.e., Harrison & Hastings, 1996; Hanski, 1998). We consider a "subpopulation" as the group of fish corresponding to a specific site-by-year-by-life stage combination.

2.2 | Pooled ezRAD library preparation

Genomic DNA was extracted with the Omega E-Z 96 Tissue DNA Kit (Omega Bio-tek) following the Tissues and Mouse Tail protocol. We used the pooled ezRAD method (Toonen et al., 2013) to obtain genome-wide SNP data because it provided a cost-effective way to screen many individuals without the potential negative effects of PCR duplication (see Methods S1). We generated two library replicates (1 and 2) for each subpopulation, with $10 \leq n \leq 30$ fish for each library, with a mean of 22 and a standard deviation of 6.49, for a total of 366 fish (Table 5). The same individuals were pooled for each library replicate per subpopulation. Each library replicate was prepared from equal contributions of DNA per pooled fish, to a total of 500 ng DNA per library replicate. *MboI* and *Sau3AI* enzymes were used to digest the genomic DNA. Libraries were sequenced on 150 bp paired-end Illumina HiSeq4000 runs. In the initial sequencing run, some of the first library replicates yielded very low sequence coverage and required a second round of sequencing. For each subpopulation pool, this yielded three FASTQ file pairs: library 1 replicates sequenced twice, and library 2 replicates sequenced once.

TABLE 2 Genetic expectations for spatial outlier loci through time, and observations from *Bathygobius cocosensis*

Data		Temporally constant patterns			Temporally variable patterns		
Measured response[a]	**Evidence**	**Genetic expectation**	**Selective hypothesis**	**Alternative (neutral) hypothesis**	**Genetic expectation**	**Selective hypothesis**	**Alternative (neutral) hypothesis**
The identity of spatial outlier loci, within a life stage	Figure 4	Expect the same outlier loci to be recovered at each timepoint. (✗)	Selection acts on the same loci at all timepoints.	Stable demographic processes maintain anomalous locus-specific differentiation.	Expect different outlier loci to be recovered at each timepoint (✓)	Selection acts on different loci across timepoints. Or, traits under selection have a polygenic architecture and variance in other loci alters the genomic background and response of the focal locus between timepoints.	Stochastic demographic processes cause ephemeral anomalous locus-specific differentiation.
The allele frequencies at spatial outlier loci, within a life stage	Figure 5	Expect the spatial pattern of allele frequencies to be stable between timepoints (✗)	The direction of selection and the alleles it acts on is the same between timepoints.	Stable demographic processes maintain anomalous locus-specific allele frequencies differences.	Expect the spatial pattern of allele frequencies to be variable between timepoints (✓)	The direction of selection and the alleles it acts on varies across timepoints. Or, traits under selection have a polygenic architecture and variance in other loci alters the genomic background and response of the focal locus between timepoints. Or, genetic markers are in different phases with the causal variant, such that temporal variation in allele frequencies reflects haplotype composition shifts.	Stochastic demographic processes cause ephemeral anomalous locus-specific allele frequency differences.

Note: A tick (✓) indicates an observation supported by our data in *B. cocosensis*, whereas a cross (✗) indicates lacking evidence.

[a]Note that the temporal variability in selection is contingent on the timescale of observation. In our study of *B. cocosensis*, this timescale was three sampling years.

TABLE 3 Genetic expectations for spatial outlier loci between life stages, and observations from *Bathygobius cocosensis*

Measured response	Genetic expectation	Stable or variable[a]	Selective hypotheses	Alternative (neutral) hypotheses	Evidence
The identity of spatial outlier loci, between life stages	Spatial outlier loci are shared between juveniles and adults. **(✓)**	Stable	The same loci are under spatially divergent selection across life stages. Or, genetic differentiation due to spatially divergent selection in juveniles persists into adulthood (carry-over effect).	Genetic differentiation due to neutral processes in juveniles persists into adulthood (carry-over effect)	Figure 4
	Spatial outlier loci are present in juveniles but absent in adults. **(✓)**	Variable	Ontogenetic shift in selection causes reduced genetic differentiation (among sites) at a locus in the adult life stage. Or, genetic differentiation due to spatially divergent selection in juveniles is eroded by the accumulation of genetic variation over successive waves of recruitment.	Genetic differentiation due to neutral processes in juveniles is eroded by the accumulation of genetic variation over successive waves of recruitment.	Figure 4
	Spatial outlier loci are absent in juveniles but present in adults. **(✓)**	Variable	Ontogenetic shift in selection causes greater differentiation (among sites) at a locus in the adult life stage. Or, purifying selection at a locus occurs in both life stages. However, the strength of differentiation is not detectable until later in life when variance around local optima is decreased in the adult life stage.	Genetic differentiation among sites in the adult life stage is the result of stochastic nonadaptive processes.	Figure 4
The allele frequencies of spatial outlier loci, between life stages	Alleles frequencies within a site are the same in juveniles and adults. **(?)**	Stable	The same allele is locally adaptive in both life stages. Or, patterns of allele frequencies arising from spatially divergent selection in juveniles persists into adulthood (carry-over effect)	Patterns of allele frequencies arising from neutral processes in juveniles persists into adulthood (carry-over effect)	Figure 5
	Alleles frequencies within a site are different between juveniles and adults. **(✓)**	Variable	Ontogenetic shift in selection favours a different allele in each life stage. Or, genetic markers are in different phases with the causal variant, such that life stage variation in allele frequencies reflects different haplotype compositions in juveniles versus adults.	Differences in allele frequencies between life stages, within a site, are due to stochastic nonadaptive processes.	Figure 5

Note: A tick **(✓)** indicates a genetic observation supported by our data in *B. cocosensis*, and a question mark **(?)** indicates equivocal evidence.

[a]The variability in spatially divergent selection between life stages. "Stable" implies that the spatial pattern of selection is the same in juveniles and adults, which might entail: the same loci under selection, same direction of selection, and same allele effect sizes. "Variable" implies that the spatial pattern of selection differs, which might entail: different loci under selection, different direction of selection, and different allele effect sizes.

2.3 | *De novo* assembly and variant calling

Raw reads were initially screened for contaminant sequences using FASTQ_SCREEN (Wingett & Andrews, 2018) against a database containing phiX, human, mouse, *Drosophila serrata*, yeast, and *Escherichia coli* genomes obtained from NCBI (respective accessions: GCF_000819615, GCF_000001405, GCF_000001635, GCF_002093755, GCF_000146045, and GCF_000005845), as well as Illumina adapter sequences. The FASTQ files were then reordered using the PAIRFQ_LITE.PL script (Staton, 2016). Restriction sites (GATC) were trimmed using SEQTK (Li, 2019).

De novo assembly of subpopulations in this present study was performed in conjunction with additional subpopulations from a larger study, incorporating samples and genetic variation from a broader geographic extent of *B. cocosensis*' Australian range (see Methods S1). Quality trimming, RAD contig assembly, mapping, and variant calling was conducted in the DDOCENT pipeline (Puritz et al., 2014), which utilises a suite of independent programs: TRIMMOMATIC, RAINBOW, CD-HIT, and PEAR for assembly; BWA and SAMTOOLS for mapping; and FREEBAYES for variant calling (Bolger et al., 2014; Chong et al., 2012; Garrison & Marth, 2012; Li & Durbin, 2009; Li et al., 2009; Li & Godzik, 2006; Zhang et al., 2014). We briefly outline our parameter choices below but elaborate in the Methods S1.

DDOCENT was used to trim reads for poor quality bases and residual adapter sequences using default settings. To prevent nonindependent contributions, we selected the largest FASTQ file pair from each subpopulation pool for RAD contig assembly (Table S1.2),

TABLE 4 Joint phenotypic and genetic expectations to infer the role of spatially divergent selection, and observations from *Bathygobius cocosensis*

Phenotypic structure	Genetic structure	Correlation	Support[a]	Selective hypotheses: spatially divergent selection	Alternative (neutral) hypotheses: no spatially divergent selection	Evidence
None $P_{ST} = 0$	None $F_{ST} = 0$	Uncorrelated $cor(P_{ST}, F_{ST}) = 0$	(✗)	Gene swamping overwhelms the action of selection and leads to a homogenous distribution of *heritable* phenotypic variation.	Homogenising gene flow and lack of selection leads to a homogeneous distribution of *heritable* phenotypic variation.	–
Present $P_{ST} > 0$	Present $F_{ST} > 0$	Positively correlated $cor(P_{ST}, F_{ST}) > 0$	(✗)	Phenotypic variation (*heritable* or *plastic*) is spatially structured and is aligned with patterns of genetic divergence.	Lack of selection leads to spatial structuring of *heritable* phenotypic variation that tracks patterns of genetic variation.	–
Present $P_{ST} > 0$	None $F_{ST} = 0$	Uncorrelated $cor(P_{ST}, F_{ST}) = 0$	(✓)	Phenotypic variation (*heritable* or *plastic*) is spatially structured against a homogenous genetic background.	Locally stochastic environments cause spatial structuring of *plastic* phenotypic variation in the absence of selection and against a homogenous genetic background.	Figures 6 and S2.8
Present $P_{ST} > 0$	Present $F_{ST} > 0$	Uncorrelated $cor(P_{ST}, F_{ST}) = 0$	(✓)	Phenotypic variation (*heritable* or *plastic*) is spatially structured in a direction uncorrelated to the genetic background.	Locally stochastic environments cause spatial structuring of *plastic* phenotypic variation in the absence of selection and in a direction that is uncorrelated to the genetic background.	Figures 6 and S2.8

[a]Support for the joint phenotypic and genetic observations. A tick (✓) indicates an observation supported by our data in *B. cocosensis*, whereas a cross (✗) indicates lacking evidence.

specifying a depth of four reads in at least four subpopulation pools to retain a unique read sequence. Clustering of 99% similarity was selected to merge homologous sequences. Mapping of trimmed reads to the reference involved the BWA MEM algorithm (Li & Durbin, 2009) and we modified the DDOCENT script to specify the following parameters: match score (-A 1), mismatch penalty (-B 4), gap open penalty (-O 30), gap extension penalty (-E 10), and an unpaired mate penalty (-U 20). Variants were called using the DDOCENT default settings for FREEBAYES (Garrison & Marth, 2012), and the raw VCF file was filtered (filtering parse 1) using VCFTOOLS (Danecek et al., 2011) to remove indels (--remove-indels), keep only biallelic loci (--max-alleles 2), have a map score of 30 (--minQ), an average of one read per sample per locus (--min-meanDP 1), and no missing data (--max-missing 1). We removed all missing data because allele frequency imputation across replicates cannot be conducted on loci where zero reads were obtained in one of the replicates. More stringent filtering and imputation of allele frequencies was carried out in downstream analyses (described in the next section). Contigs with mitochondrial origin were identified by mapping reference contigs against the *B. cocosensis* mitogenome (Evans et al., 2018) (NBCI MG704838) using BOWTIE2 (Langmead & Salzberg, 2012) and retrieving hits with SAMTOOLS (Li et al., 2009).

2.4 | Post-assembly data filtering and allele frequency estimation

We imported the VCF file into R (R Core Team, 2018), merging read counts from the double sequencing runs of replicate library 1 for each subpopulation. The biallelic SNPs underwent a secondary filtering step (filtering parse 2). This began by removing SNPs with poor coverage. Within each subpopulation, reads across library replicates were summed for each locus. If a locus had a summed depth <40 reads in any subpopulation, it was excluded from further analyses to ensure a consistent set of loci across all subpopulations.

Pool-seq data encompasses two sources of sampling variance in allele frequencies: the first being variance due to subsampling the population, and the second being variance due to unequal DNA contributions of pooled individuals to the sequenced pool (Gautier et al., 2013; Hivert et al., 2018). Sampling variance due to unequal contributions can be reduced and quantified with replicate pool-seq libraries, providing greater confidence in the sample allele frequencies, relative to those estimated from singly sequenced pool-seq libraries (Gautier et al., 2013). To estimate allele frequencies, we used the algorithm developed by Gautier et al. (2013), implemented in POOLNE_ESTIM. Briefly, POOLNE_ESTIM uses a hierarchical Bayesian model to estimate the population reference allele frequency, herein denoted as p_{snp} (reported as π by POOLNE_ESTIM), in replicated pool-seq experiments. The estimate of p_{snp}, and its standard deviation, s_{snp}, are assumed to have beta prior distributions. Individual contributions to a pool are assumed to follow a Dirichlet distribution, and the experimental error is derived from the expected number of

TABLE 5 Sampling details for focal subpopulations

Life stage	Year	Month	Site	Replicate	*n*	n_e	SD (n_e)
Adults	2014	6	Cart	1	29	29	1
				2		29	1
		5	Hast	1	23	23	1
				2		23	1
		6	Shell	1	21	21	1
				2		21	1
	2015	11	Cart	1	13	13	1
				2		13	1
		9	Hast	1	20	20	1
				2		20	1
		7	Shell	1	16	16	1
				2		16	1
	2016	3	Cart	1	24	24	1
				2		24	0.961
		2	Hast	1	24	24	1
				2		24	1
		2	Shell	1	26	26	1
				2		26	0.995
Juveniles	2014	6	Cart	1	14	14	1
				2		14	1
		5	Hast	1	16	16	1
				2		16	1
		6	Shell	1	15	15	1
				2		15	1
Juveniles	2015	2	Hast	1	30	30	1
				2		30	1
		7	Shell	1	10	10	1
				2		10	1
	2016	3	Cart	1	30	30	1
				2		30	1
		2	Hast	1	30	30	1
				2		30	1
		2	Shell	1	25	25	1
				2		25	1

Life stage, adults or juveniles; **Year**, sampling year; **Month**, sampling month; **Site**, Point Cartwright (Cart), Hastings Point (Hast) or Shellharbour (Shell); **Library**, replicate pool-seq library; ***n***, experimental pool size, the true number of diploids; $\boldsymbol{n_e}$, effective pool size, estimated from POOLNE_ESTIM on filtered read data. **SD($\boldsymbol{n_e}$)**, standard deviation of n_e, estimated from POOLNE_ESTIM on filtered read data.

pooled diploids relative to the variance in individual contributions. Therefore, POOLNE_ESTIM also provides an estimate of the effective pool size, n_e, the number of equally contributing individuals to the pool. In our pool-seq data, n_e for all libraries exhibited very high concordance with the true number of pooled diploids (Table 5).

We formatted SNPs for imputation in POOLNE_ESTIM, and compiled output p_{snp} estimates, using R's GENOMALICIOUS (Thia & Riginos, 2019) functions POOLNE_ESTIM_INPUT( ) and POOLNE_ESTIM_OUTPUT( ), respectively. SNPs of mitochondrial origin, or that originated from contigs with >6 SNPs, or had a minor allele frequency (MAF) <0.05, were removed. We used a MAF <0.05 threshold because this represents the lowest detection limit in our smallest sample pools (Table 5): where *n* = 10 diploids, there are 20 possible alleles, and the occurrence of a single minor allele would comprise 5% of the sample. After randomly sampling one SNP per contig, we were left with a working data set of 1288 SNPs.

2.5 | F_{ST}- versus allele frequency-dependent analyses

After obtaining a working SNP set, we analysed our pool-seq data in two different ways: analyses that required F_{ST} estimates or allele frequencies. Genetic structure required F_{ST} estimates, whereas testing for genetic signatures of sweepstakes required allele frequencies. We used two outlier locus detection methods: OUTFLANK (Whitlock & Lotterhos, 2015) and PCADAPT (Luu et al., 2017), which required F_{ST} and allele frequency estimates, respectively.

To implement these different analyses on our replicated pool-seq data, we developed a probabilistic framework, one each for the F_{ST}- and allele frequency-dependent analyses (Figure S1.4). The differences between our frameworks for F_{ST} versus allele frequency analyses reflect the specificities of different available software for pool-seq data. However, the premise for each was the same: single pool-seq libraries provide single estimates of population variation, which contains both sampling and technical error, but replicate libraries of the same pool of individuals provide multiple independent estimates and can be used to assess the effects of pooling variance on downstream analyses.

For the F_{ST}-dependent analyses, R's POOLFSTAT package (Hivert et al., 2018) was used to estimate F_{ST}. This method decomposes the variance in allele counts as Q_1, the probability of identity-in-state within pools, and Q_2, the probability of identity-in-state between pools, such that:

$$F_{ST} = (Q_1 - Q_2)/(1 - Q_2)$$

This formulation is analogous to Weir and Cockerham's (1984) F_{ST}. Estimates of Q_1 and Q_2 are themselves derived from the distribution of read counts and pool sizes, and the expected allele frequencies, across samples.

POOLFSTAT does not accommodate replicate pool-seq data. Instead, we calculated F_{ST} for all possible combinations of replicate libraries among subpopulations for each comparison we made. For example, let $Pop_{x,y}$ represent a single pool-seq library sequenced from subpopulation, *x*, and library replicate, *y*. For a simple pairwise comparison between two subpopulations, there are four possible library replicate combinations: $Pop_{1,1}$ + $Pop_{2,1}$, $Pop_{1,2}$ + $Pop_{2,1}$, $Pop_{1,1}$ + $Pop_{2,2}$, and $Pop_{1,2}$ + $Pop_{2,2}$. Hence, the calculated F_{ST}-values reflect a range of estimates that are observable given the replicate pool-seq data. Further details are in the Methods S1.

For the allele frequency-dependent analyses, we used parametric bootstrapping to simulate the hypothetical multilocus allele frequency combinations that could have been observed in our study. The POOLNE_ESTIM imputed posterior mean allele frequencies (p_{snp}) and their standard deviations (s_{snp}) were used to generate a distribution of possible observable allele frequencies following a beta distribution. For each locus, we randomly drew 100 values; each *b* draw was combined across loci to create a single multilocus bootstrap replicate data set for each subpopulation. These parametric bootstraps provided insight into the range of test statistics that could be observed given uncertainty in allele frequency estimates. Further details are in the Methods S1.

2.6 | Genetic structure

Genome-wide SNPs were used to test the neutral genetic expectations for three population genetic scenarios: (i) panmixia, (ii) local retention, and (iii) chaotic genetic patchiness. These scenarios have contrasting expectations for spatial genetic structure, its temporal stability, and differences between life stages (Table 1). Importantly, we contrast the stable, predictable processes in panmixia and local retention against the variable, stochastic processes in chaotic genetic patchiness.

Panmixia by definition describes a randomly mating population with no genetic subdivision (Wright, 1949). *Bathygobius cocosensis* mate locally (Joshua Thia, personal observation), so we do not imply true panmixia. However, high gene flow and (or) large effective population sizes, and an absence of sweepstakes reproduction, might produce a pattern of genetic homogeneity among subpopulations through space and time, akin to that expected under panmixia. In contrast, under *local retention*, gene flow is reduced via locally recruited juveniles, allowing drift and local inbreeding to accumulate spatial genetic structure. A pattern of isolation-by-distance might emerge from locally constrained dispersal. Finally, under *chaotic genetic patchiness*, rates of gene flow vary through space and time, such that small-but-significant levels of ephemeral spatial genetic structure can emerge, and local gene pools exhibit temporal shifts in composition. Moreover, local juveniles might be significantly divergent from local adults due to sweepstake events or variable source contributions, which would create genetic structure between life stages.

A series of F_{ST} calculations were used to understand how genome-wide variation was structured across space, time, and life stage (Table 1). (i) Spatial genetic structure was determined by calculating F_{ST} among all sites (metapopulation level) and between pairs of sites (pairwise-site level), within each life stage, each year. This provided insights on whether genetic variation was homogenous or structured across space and its temporal and (or) life stage variability. (ii) Temporal genetic structure was determined by calculating F_{ST} within each site between pairs of years (pairwise-year level), for each life stage. These estimates allowed us to quantify temporal turnover in local gene pools, that is, whether genetic composition was stable through time. (iii) Life stage genetic structure was calculated as pairwise F_{ST} between a juvenile subpopulation from a focal site and each adult subpopulation from all three sites, within each year. These estimates afforded tests of which adult subpopulation(s) was (were) most genetically similar to a focal juvenile subpopulation.

We calculated F_{ST} using the COMPUTEFST( ) function from R's POOLFSTAT package (Hivert et al., 2018). For each comparison made, F_{ST} was calculated for all possible replicate library combinations among subpopulations (see above; also see Methods S1). Because we had two replicates for each subpopulation, pairwise comparisons yielded four F_{ST} estimates, whereas comparisons among three subpopulations yielded eight F_{ST} estimates. To test whether F_{ST} was nonzero, we evaluated whether the range of F_{ST} estimates across the library replicate combinations contained zero. There are two

caveats to this approach. Firstly, unlike individually sequenced samples, we could not use permutations to test the "significance" of F_{ST} with respect to random expectations. This is an inherent trade-off of pool-seq in that genotyping precision is sacrificed against capacity for increasing total sample size. We elected to use the pool-seq approach to enable increased spatial, temporal, and life stage replication. Secondly, the range of F_{ST}-values was used instead of confidence intervals because there were not enough datapoints for us to confidently characterise a distribution, especially in pairwise comparisons (n = 4). For any comparison, if at least one F_{ST}-value from any library replicate combination had a value ≤0, we classed the subpopulations in that comparison as undifferentiated. This approach is sensitive to the lowest F_{ST} estimated, but is probably overall more conservative, and is more robust than relying on estimates derived from single pool-seq libraries because replication provides an indication of the pooling variance and its effects on possible observable outcomes.

2.7 | Sweepstakes reproduction

We assessed signatures of sweepstakes by comparing the genetic diversity between adult–juvenile pairs in each site-by-year combination (Table 1). Under sweepstakes, new recruits at a site (juveniles) are expected to have less genetic diversity than adults because of skewed variance in reproductive success (Hedgecock et al., 2007; Hedgecock & Pudovkin, 2011). For adults and juveniles at the same site, in each year, where p_{snp} is the reference allele frequency at a locus, and $1 - p_{snp}$ is the frequency of the alternate allele, we calculated A_e (effective number of alleles) as:

$$A_e = 1/\left[p_{snp}^2 + (1 - p_{snp})^2\right]$$

We then calculated the difference in genetic diversity between adults and juveniles at each locus as:

$$\Delta A_e = \text{adult}\, A_e - \text{juvenile}\, A_e.$$

A_e = 2 when a biallelic locus has equally common alleles. We therefore used a one-tailed t-test to test the null hypothesis, $\Delta A_e \leq 0$, as no evidence of sweepstakes, versus the alternate hypothesis, $\Delta A_e > 0$, as evidence of sweepstakes, because ΔA_e will be positive if adults have more genetic diversity. We assumed loci were sufficiently unlinked to provide independent reflections of genome-wide variation but cannot appraise this assumption with pool-seq data.

These genetic diversity calculations were made using the 100 parametric bootstrap replicates of allele frequencies. For each b bootstrap we tested the support for sweepstakes across all 1288 loci. This gave 100 p-values for the null hypothesis, $p(\Delta A_e \leq 0)$. We then calculated, P_{sweep}, the proportion of bootstrapped p-values where the alternate hypothesis was supported, $p(\Delta A_e \leq 0) < 0.05$. Hence, as P_{sweep} approaches 1, there is greater evidence for genetic signatures of sweepstakes across the bootstrap replicates.

2.8 | Spatial outlier loci

We considered how variability in spatially divergent selection across time and life stages might structure genetic variation. Spatial selection is expected to increase allele frequency differences between sites. We were primarily interested in whether: (i) the same loci exhibited signatures of spatial differentiation across multiple years, within a life stage, which might indicate temporal predictability in targets of spatial selection; and (ii) the same loci exhibit signatures of spatial differentiation between post-settlement juveniles and recruited adults, which might indicate predictable targets of spatial selection between life stages.

Interpreting replicated outlier locus analyses can be challenging when considering the prevalence of highly non-equilibrium dynamics of marine metapopulations. Without clear knowledge regarding the adaptive role of a locus and its alleles (through experimentation), competing selective or stochastic neutral hypotheses offer reasonable explanations for spatiotemporal variability in outlier loci (Babin et al., 2017; Bourret et al., 2014). Nonetheless, we contend that shared outlier loci that have been identified independently in different sample sets (different years, or different life stages) provide the greatest support for those loci potentially affected by selective processes. We summarise genetic expectations for selective and competing neutral hypotheses in outlier locus analyses for: (i) temporal comparisons within life stages (Table 2), and (ii) comparisons between life stages (Table 3).

Two outlier locus methods were used to exploit available allele frequency and F_{ST} data. R's PCADAPT package was used to analyse allele frequencies. In PCADAPT's pool-seq implementation, singular value decomposition is performed on allele frequency matrices and SNPs with disproportionate contributions to all K (= n subpopulations − 1) right singular vectors are identified as outlier loci. PCADAPT is less sensitive to demographic effects causing co-dependence among subpopulations, which can generate false positives (Lotterhos & Whitlock, 2014). However, it does suffer from low power if the spatial pattern of selection does not correlate with genome-wide neutral genetic structure captured by the major PC axes, which may result in false negatives (Capblancq et al., 2018). We also used R's OUTFLANK package to analyse F_{ST} estimates. OUTFLANK identifies loci that deviate significantly from the expected neutral distribution of F_{ST}, which is χ^2 distributed with respect to K (= n subpopulations − 1). OUTFLANK is a highly conservative method and returns few false positives across many demographic scenarios (Luu et al., 2017; Whitlock & Lotterhos, 2015). Therefore, we expected that overlap between PCADAPT and OUTFLANK would provide high confidence regarding loci exhibiting truly significant spatial differentiation.

Outlier locus analyses were conducted in our probabilistic framework (Figure S1.4). For PCADAPT analyses, we obtained outlier loci for each b bootstrap replicate, whereas for OUTFLANK analyses, outlier loci were identified for each library replicate combination. For both analyses, a false discovery rate (FDR) of 0.20 was used for each bootstrap replicate (PCADAPT) or library replicate combination

(OUTFLANK). We then considered how many times a locus was identified as an outlier across replicates as a measure of support for that locus exhibiting spatially divergent patterns. For PCADAPT, loci identified as outliers across >90% bootstrap replicates were considered significant. OUTFLANK identified very few outlier loci overall, so we report those that were detected across ≥3 library replicate combinations. Loci needed to pass two filtering stages to be considered outlier loci in downstream analyses: the within replicate FDR threshold and multiple observations across replicates. Therefore, our outlier locus analyses are likely to be highly conservative. Further details are in the Supporting Information (Methods S1).

For the final set of outlier loci, we assessed overlap across time or life stage. Within life stages, we tested whether outlier loci were shared among years, which might indicate that the same loci are predictable targets of spatial selection (Table 2). Between life stages, we tested whether outlier loci were shared between adults and juveniles, which might indicate that spatial selection operates on the same loci across ontogeny (Table 3). We used randomised permutations to estimate null probabilities for our observed overlap in outlier loci (see Methods S1).

To evaluate whether any of our outlier loci were related to functional genetic variation, we identified outlier loci with the greatest overlap across years and life stages. The RAD contigs for the top four outlier loci were queried against the NCBI database using BLASTN (Altschul et al., 1990). One of these contigs, contig_878, was a single contiguous sequence (overlapping paired reads). The remaining contigs were scaffolded (nonoverlapping paired reads), so we queried the forward and reverse ends separately.

2.9 | Spatiotemporal phenotypic structure

We quantified the phenotypic differentiation among adult subpopulations to contrast spatiotemporal patterns in phenotypic variation to genetic variation (Figure 1d). Based on previous analyses of head shape variation within a single site and year (Malard et al., 2016), we expected head shape to exhibit adaptive variation among sites, due to either selection on heritable phenotypic variation, or adaptive phenotypic plasticity (Table 4). Adaptive plasticity occurs when individuals alter their phenotype in the direction of the local optima, which positively affects their fitness (Gibert et al., 2019). Alternatively, neutral processes could also generate temporal and spatial variation in head shape, for example, when correlations exist between neutral genetic and phenotypic variance (Whitlock, 1999) (Table 4). Additionally, locally stochastic environmental variation might lead to plastic responses that cause phenotypic variation to be spatially structured in a manner unassociated with local optima and in the absence of selection (Table 4).

All adult fish used in the genetic analyses were included in the morphometric analyses, except two fish from Point Cartwright (one each from 2015 and 2016) for which morphological characters could not be reliably measured. Although patterns of phenotypic variation across life stages might be informative about microevolutionary processes, the small size, and more delicate tissues of juvenile *B. cocosensis* resulted in undiscernible landmarks (Figure 1d) for many juvenile specimens. Consequently, we consider only adults in this study. Eight landmarks (Figure 1d) were used to characterise individual phenotypes and were placed on head images using TPSDIG264 (Rohlf, 2015). These landmarks captured variation in the mouth, eye, and operculum region.

Statistical analyses were conducted in R. Importation, general Procrustes alignment, and principal component analysis (PCA) of landmarks was performed using GEOMORPH (Adams et al., 2016; Adams & Otárola-Castillo, 2013). The first seven PC axes explained ~91% of variation in head shape and were retained for subsequent analysis. One sample, Bcoco849, was a multivariate outlier in PC morphospace, based on Mahalanobis distances of head shape PCs 1–7 (Figure S1.5) (see Methods S1). We excluded Bcoco849 from further analyses given its disproportionate divergence from the other samples.

To determine whether spatial phenotypic and genome-wide genetic differentiation were correlated, we estimated P_{ST} between site pairs and across the metapopulation. If selective processes structure head shape variation (via evolutionary or plastic responses), we expected head shape to diverge against a homogenous genetic background, or phenotypic differentiation to be uncorrelated to genetic differentiation (Table 4), with the expectation that $cor(P_{ST}, F_{ST}) = 0$. R's MANOVA( ) function was used to estimate the between and within subpopulation variance components in adult head shape phenotypes by fitting the model:

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

Where **Y** was a matrix of traits (head shape PCs 1–7), μ was the mean, SITE (a categorical factor) was the effect of sampling location (between subpopulation variance), and ε the error (within subpopulation variance). These variance components were used to calculate P_{ST} based on the approach described in Chenoweth and Blows (2008), originally outlined in Kremer et al. (1997). Permutations were used to test whether the observed P_{ST} values were significantly larger than those obtainable under random expectations. See Methods S1 for specifics on P_{ST} calculation and permutation tests.

Our main analysis of phenotypic structure in adult *B. cocosensis* involved a MANOVA to partition head shape variation into spatial and temporal components, followed by a linear discriminant analysis (LDA) (Rencher, 1998) to visualise the phenotypic variation (Figure 1d). Using R's MANOVA( ) function, we fit the model:

$$\mathbf{Y} = \mu + \text{SITE} + \text{YEAR} + \text{SITE:YEAR} + \varepsilon$$

The effect of sampling location (SITE) and year (YEAR) were both coded as categorical factors, and the site-by-year interaction (SITE:YEAR) was included to determine if spatial effects were temporally variable. Because the MANOVA revealed a significant site-by-year interaction, we focused our interpretation on this effect. We extracted the sums-of-squares and cross-products matrices for the SITE:YEAR interaction (**H**) and the residuals (**E**) to calculate the test matrix, **F** (= $\mathbf{E}^{-1}\mathbf{H}$) (Rencher, 1998). Singular value decomposition was used to decompose **F**, using R's SVD( ) function. The right singular vectors from this decomposition (**V** matrix) were used to transform the original

phenotype scores (head shape PCs 1–7) into LD axes that captured variation in head shape attributed to the SITE:YEAR effect. The means and standard errors for the first axis ($LD1_{SITE:YEAR}$) were calculated for each subpopulation to visualise spatiotemporal phenotypic structure.

Because shape can vary with size, we evaluated the role of body size on head shape variation to validate that spatiotemporal phenotypic structure was not a statistical artefact of sampling (details in Methods S1). Body size, measured as $\log_{10}$ standard length (mm) (Figure S1.6a), was significantly different among years, but there were no significant differences among sites or a site-by-year interaction (Figure S1.6b; Table S1.4). To confirm that variation in size was not causing the observed site-by-year phenotypic variation, we regressed $LD1_{SITE:YEAR}$ against body size. $LD1_{SITE:YEAR}$ scores were not significantly predicted by body size (ANOVA model: $F_{1,\ 118} = 2.406$, $p = 0.124$) (Table S1.5), indicating that sampling size bias between years did not affect our interpretations of phenotype on $LD1_{SITE:YEAR}$ scores.

3 | RESULTS

3.1 | Bioinformatics

After the first preliminary filtering step with VCFTOOLS, the mean number of reads mapped to our *de novo* RAD contig assembly was 14,825,021 reads per replicate (range: 1,677,409–101,634,531) and 29,650,042 per pool (range: 6,403,735–112,229,066). Our second filtering step in R removed SNPs that: (i) came from contigs with >6 SNPs; (ii) had a depth <40 reads in the combined replicates of our focal subpopulations; (iii) were of mitochondrial origin; and (iv) had a MAF <0.05. After selecting one random SNP per contig, we were left with 1288 SNPs that were present in all focal subpopulations. Read statistics for this working SNP set are reported in Table S1.3.

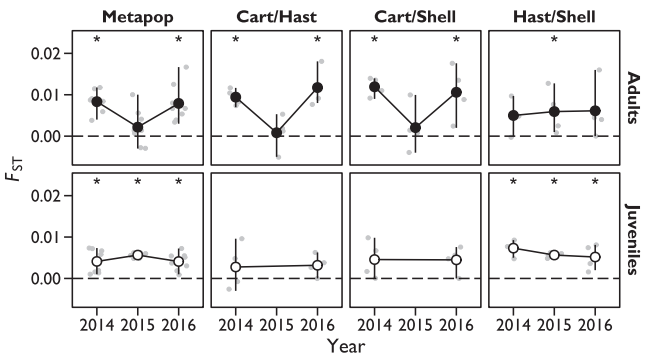


FIGURE 2 Spatial genetic structure. Genetic structure across the metapopulation (all three sites) or between site pairs. Each plot is for a different life stage (rows) in a focal year (columns). Large black or white points are the mean F_{ST} estimated across different pool-seq replicate combinations. Small grey points are the different F_{ST} estimates for each pool-seq replicate combination, with bars illustrating the minimum and maximum values obtained. An asterisk indicates those F_{ST} ranges that do not overlap zero. Dashed line demarcates $F_{ST} = 0$

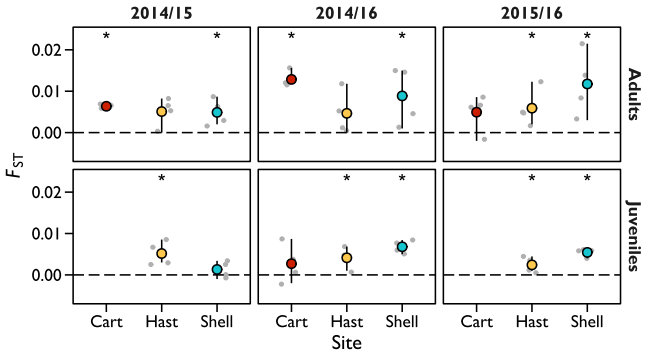


FIGURE 3 Temporal genetic structure. Genetic structure between year pairs. Each plot is for a different life stage (rows) in a focal year pair (columns). Large coloured points are the mean F_{ST} estimated across different pool-seq replicate combinations. Small grey points are the different F_{ST} estimates for each pool-seq replicate combination, with bars illustrating the minimum and maximum values obtained. An asterisk indicates those F_{ST} ranges that do not overlap zero. Dashed line demarcates $F_{ST} = 0$

3.2 | Genetic structure

Spatial genetic structure, within years, exhibited a range of $-0.003 \le F_{ST} \le 0.017$ across the metapopulation, whereas pairwise spatial estimates ranged from $-0.005 \le F_{ST} \le 0.018$ (Figure 2; Table S2.1). Temporal genetic differentiation between pairwise years, within sites, ranged from $-0.002 \le F_{ST} \le 0.021$ in adults, and $-0.002 \le F_{ST} \le 0.009$ in juveniles (Figure 3; Table S2.2). Life stage genetic differentiation between adult–juvenile pairs, within each year, ranged from $-0.0004 \le F_{ST} \le 0.015$ for local comparisons, and $-0.004 \le F_{ST} \le 0.016$ for non-local comparisons (Figures S2.1; Table S2.3).

Genetic differentiation was overall very weak ($F_{ST} \le 0.021$), and of comparable magnitude, across all ecological dimensions examined: space, time, and life stage. The metapopulation exhibited small fluctuations between weakly structured and panmictic states (Figure 2), and local sites underwent small changes in genetic composition year-to-year (Figure 3). Moreover, there was no clear indication that local adults were more likely to contribute disproportionately to local juveniles at a site, based on observations that non-local adults could be as (or more) genetically similar to focal juveniles within a site (Figure S2.1). Indeed, temporal changes in genetic differentiation between juveniles and different adult subpopulations (Figure S2.1) might indicate that variable source contributions play a role in shaping the genomic backgrounds within local subpopulations.

3.3 | Sweepstakes reproduction

Evidence of weak sweepstakes reproduction in *B. cocosensis* was found in some of our sampled adult–juvenile subpopulation pairs. In our probabilistic framework using parametrically bootstrapped allele frequencies, we measured support for sweepstakes using P_{sweep}, the proportion of bootstrap replicates where the null hypothesis, $\Delta A_e \le 0$, was false. When $P_{sweep} = 1$, all bootstrap simulations

support the alternative hypothesis, $\Delta A_e > 0$, evidencing sweepstakes. The cumulative distribution of mean ΔA_e values and their associated *p*-values across bootstrap simulations are respectively illustrated in Figures S2.2 and S2.3. P_{sweep} ranged from 0.00 to 0.47, which suggests that some hypothetical combinations of allele frequencies would support sweepstakes reproduction in some comparisons. Evidence for sweepstakes was greatest for Hastings Point 2015 (P_{sweep} = 0.47) and was weaker for Hastings Point 2014 (P_{sweep} = 0.26) and Shellharbour 2014 (P_{sweep} = 0.19). All other adult–juveniles pairs had $P_{sweep} \leq 0.05$, suggesting very little evidence of sweepstakes in those comparisons.

3.4 | Spatial outlier loci

The percentage of loci identified as spatial outliers by PCADAPT ranged from 0.23% (adults 2016) to 1.71% (juveniles 2015) of all analysed loci (Figure 4a). OUTFLANK only identified spatial outlier loci in 2016 juveniles, which comprised 0.23% of the total loci analysed (Figure 4a). Hence, we could only assess temporal and life stage overlap using PCADAPT results. All outlier loci detected by OUTFLANK were detected in PCADAPT (Figure 4c).

There was only one locus that exhibited temporal overlap (within life stages) in the PCADAPT analyses: a single locus between juveniles in 2015 and 2016 (Figure 4a). Estimation of null probabilities suggested that temporal overlap in juveniles was nonsignificantly different from random expectations when considering all three years at once (p = 0.160). However, when considering just the pairwise overlap between 2015 and 2016 in juveniles, our result was very significant relative to random expectations (p = 0.000). These results are consistent with selective hypotheses of temporally varying spatial selection (Table 2).

Life stage overlap of outlier loci (summing across years within life stages) was significantly different from random expectations for PCADAPT results (p = 0.000). Our results for outlier locus overlap between life stages covers the full gambit of possibilities: some loci are shared between life stages, some are unique to juveniles, and others unique to adults (Table 3). For those four loci shared between life stages (Figure 4b), a selective hypothesis would be that these loci are under divergent spatial selection in both juveniles and adults (Table 3). However, most loci were unique to their life stage, which might suggest that changes in spatial selection over ontogeny generates, or erases, differentiation at specific loci (Table 3).

We took the four overlapping outlier loci and plotted their allele frequencies across 100 bootstrap replicates (Figure 5). The locus, snp_878_26, was an outlier in three tests, whereas all other outlier loci were present in just two tests, snp_42180_291, snp_32744_49, and snp_26942_3000. Qualitative examination of allele frequencies yielded two observations. Firstly, within life stages, allele frequencies at these loci exhibit temporal fluctuations within sites and among sites. Secondly, there were no clear trends between life stages at the same site. Collectively, these observations imply that spatial selection might vary between life stages and over time, consistent with expectations of the selective mechanisms of chaotic genetic patchiness (Tables 2 and 3).

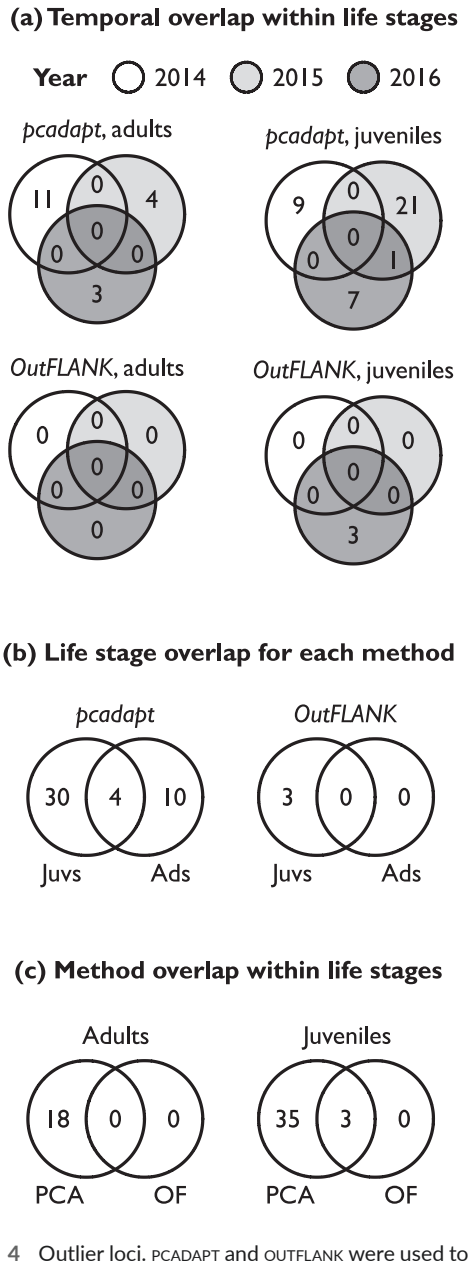


FIGURE 4 Outlier loci. PCADAPT and OUTFLANK were used to identify outliers among all three sites (Point Cartwright, Hastings Point, and Shellharbour), within each life stage. Note, juveniles from Point Cartwright were not sampled in 2015, so outliers for this year–life stage combination are only between Hastings Point and Shellharbour. The total number of outlier loci for each life stage, per method, included: 18 and 0 for adults, and 38 and three for juveniles, using PCADAPT or OUTFLANK (respectively). (a) Temporal overlap. Circles represent the outlier loci detected in each year. (b) Life stage overlap. Circles represent the number of unique outlier loci (summed across years) in adults ("Ads") and juveniles ("Juvs") for each method. (c) Method overlap. Circles represent the number of outlier loci (summed across years) detected by PCADAPT ("PCA") and OUTFLANK ("OF") within each life stage. (a–c) The number of overlapping loci is illustrated by values at the intersecting areas between circles, whereas nonoverlapping loci occur in the excluding areas of each circle

An alternative nonbiological hypothesis for variable spatial outlier loci is that bioinformatic artefacts are introduced during *de novo* assembly. For instance, different loci mapping to the same

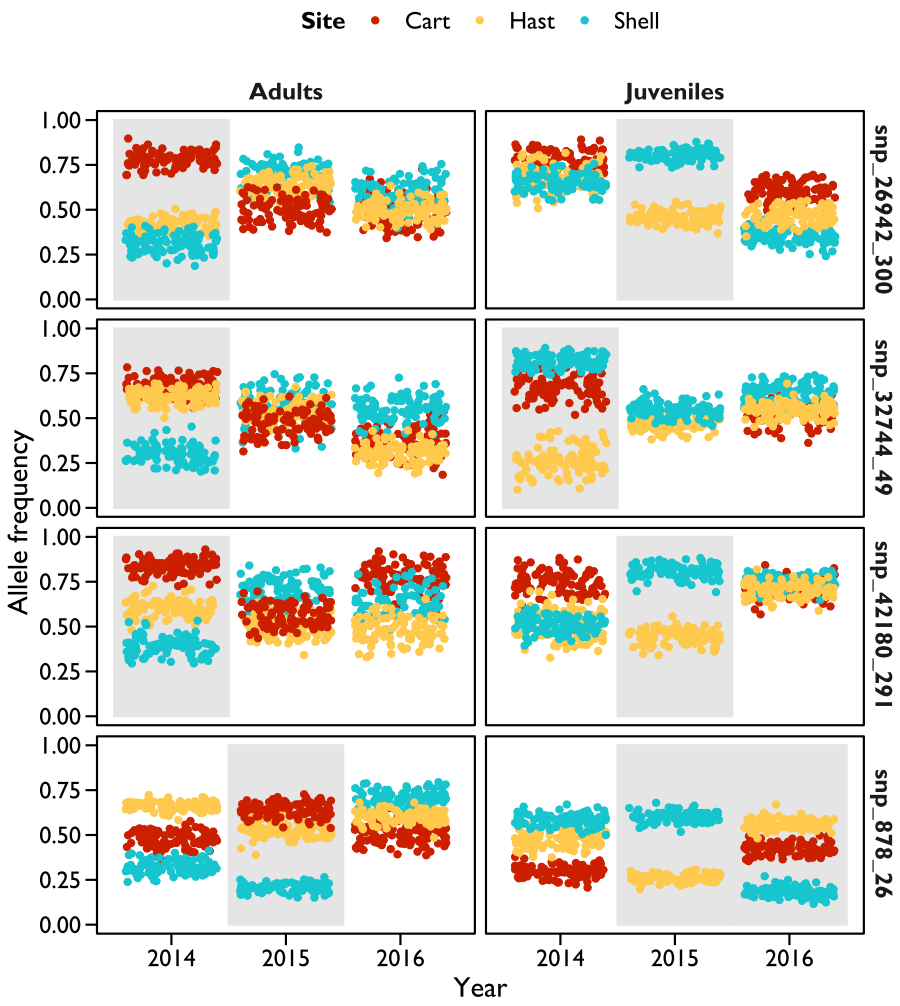


FIGURE 5 Bootstrap allele frequencies for four outlier loci with life stage overlap. Sampling year is on the *x*-axis with the reference allele frequency on the *y*-axis. Points represent parametric bootstrap allele frequencies estimated from pool-seq read counts. Points are coloured by site (see legend). Each plot is for a SNP locus (rows) and life stage (columns) combination. Grey shading indicates a year–life stage combination where a SNP was identified as a significant outlier locus among the sites. The SNPs presented are the top four loci with greatest overlap among adult and juvenile life stages

(misassembled) contig would inflate polymorphism by combining paralogous loci, which would make SNPs within such contigs sensitive to read counts from different paralogues. We therefore viewed the sequence alignments at the four overlapping outlier loci for adults in 2014 and 2015 (Figures S2.4–S2.7). For all RAD contigs from which the focal outlier loci reside, alignments did not give any indication of merged paralogues, which would be evidenced by high polymorphism and heterozygosity. Interestingly, we noted alignments at contig 878 exhibited two groups of reads: paired reads that mapped together within contig 878, and paired reads that were split between contig 878 and another contig. This observation implies that the outlier locus, snp_878_26, is associated with different RAD haplotypes (restriction site variation) in our sample of *B. cocosensis*.

Finally, to assess whether our top four overlapping outlier loci might be related to functional genetic variation, we ran a BLASTN search of the associated RAD contigs against the NCBI database. Whereas contig_878 and contig_42180 did not return any meaningful hits, there is evidence that contig_26942 and contig_32744 contain genic sequences. In contig_26942, the forward sequence had considerable homology to a cGMP phosphodiesterase in other fish across much of its length (Table S2.4), and a similar result was obtained for a shorter length of the reverse sequence (Table S2.5). The analysed locus, snp_26942_300, occurs in the reverse sequence, so it seems likely that this outlier locus is directly part of, or at least linked to, a cGMP phosphodiesterase gene in *B. cocosensis*. Additionally, in contig_32744, one of the top hits with the greatest length was for a fish toll-like receptor in the reverse sequence (Table S2.6). However, the outlier locus, snp_32744_49, occurred in the contig's forward sequence and the BLAST hits were more variable. Therefore, snp_32744_49 might be closely linked to a gene, but the evidence is equivocal.

3.5 | Spatiotemporal phenotypic structure

Estimates of spatial phenotypic structure in adult *B. cocosensis* head shape across the metapopulation were P_{ST} = 0.222 (in 2015) and 0.224 (in 2016) (Table S2.7). Pairwise comparisons among sites within a year ranged from 0.107–0.138, with the weakest differentiation between Hastings Point and Shellharbour, in both years (Table S2.7; Figure S2.8). There was no evidence that phenotypic differentiation was correlated with genome-wide genetic differentiation (P_{ST} and F_{ST} correlation, r = −0.09) (Figure S2.8).

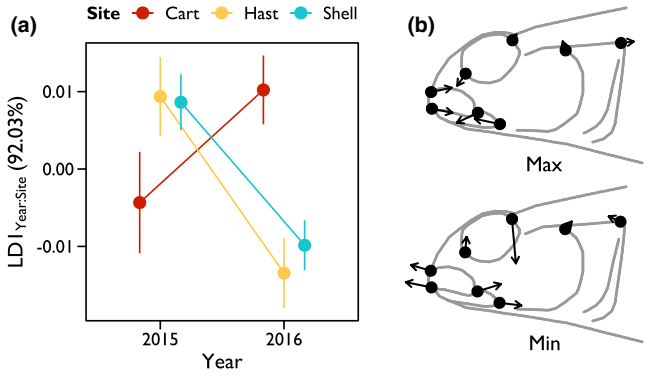


FIGURE 6 Spatiotemporal phenotypic structure among adult subpopulations. Phenotypic variation measured as the first axis in a linear discriminant analysis of site-by-year effects on adult head shape ($LD1_{SITE:YEAR}$). (a) Mean site scores for each year with standard error bars. (b) The shift in landmark coordinates from the mean head shape to the individuals with the maximum and minimum $LD1_{SITE:YEAR}$ scores. Vectors (scaled at 4×) indicate the direction and magnitude of landmark shifts

MANOVA of spatiotemporal variation in head shape revealed a significant site-by-year interaction (Λ_{Pillai} = 0.240, $F_{14, 218}$ = 2.125, p = 0.012; 4.12% of the total phenotypic variance; Table S2.8). Hence, spatial effects on head shape depended on sampling year. Most variation of the site-by-year effect was captured by a single linear combination of head shape PCs, $LD1_{SITE:YEAR}$ (92.03%) (Figure 6a). The $LD1_{SITE:YEAR}$ described variation in mouth and eye shapes, contrasting individuals possessing relatively short jaws and angled eyes with individuals possessing relatively long jaws and horizontally oriented eyes (Figure 6b). All sites exhibited shifts in mean $LD1_{SITE:YEAR}$ between years, with Port Cartwright changing in the opposite direction to Hastings Point and Shellharbour (Figure 6a). Furthermore, Hastings Point and Shellharbour were more similar to one another in their average $LD1_{SITE:YEAR}$ score and consistently had non-significant P_{ST} (with respect to random expectations). These results are surprising given greater geographic separation between Hastings Point and Shellharbour, and closer proximity of Hastings Point to Point Cartwright (Figure 1a).

Collectively, our results indicated that phenotypic differentiation was decoupled from both genetic differentiation and geographic distance among subpopulations. Given the putative ecological importance of head shape, these observations imply that selection might act on heritable and (or) plastic variation to structure phenotypes against a homogeneous or weakly differentiated genetic background (Table 4).

4 | DISCUSSION

Adaptive divergence in the presence of gene flow runs counter to expectations of genetic homogenisation in well-connected systems. Whereas stable conditions might be amendable to predictable outcomes from selection–gene flow interactions, ecological and physical processes that cause variability in selection, gene flow, and drift make it difficult to predict how biological variation is structured within species. We studied an intertidal fish, *Bathygobius cocosensis*, and interrogated three ecological dimensions that were hypothesised to exhibit variability in neutral and selective processes. Our study thus provides an empirical examination of how predictable versus stochastic processes acting across space, time, and (or) life stage might affect the distribution of genetic and phenotypic variation in marine metapopulations.

4.1 | The chaotic genetic patchiness of marine metapopulations

The demographic properties of metapopulations affect the distribution of neutral variation and the genomic background of adaptation. In our study of *B. cocosensis*, observations from genome-wide genetic variation were most consistent with a pattern of chaotic genetic patchiness (Table 1). Genetic differentiation (F_{ST}) across space, time, and life stage were of comparable magnitudes in *B. cocosensis*. Our study therefore adds to the growing consensus that patterns of chaotic genetic patchiness are prevalent in marine species (Jackson et al., 2017; Johnson & Black, 1982; Moody et al., 2015; Selwyn et al., 2016; Toonen & Grosberg, 2011; Villacorta-Rath et al., 2018).

Spatial genetic patterns did not exhibit consistent homogeneity, as per panmictic expectations, nor were spatial genetic patterns consistent with local retention (Table 1). Instead, all populations exhibited similar levels of genetic differentiation, despite being separated by hundreds of kilometers (Figure 2). It is possible that the prevailing, southward-bearing oceanographic force, the East Australian Current, helps to maintain connectivity of *B. cocosensis* along Australia's eastern coastline (Liggins et al., 2015; Murray-Jones & Ayre, 1997). Additionally, ephemeral, stochastic processes might generate transient spatial genetic structure among sites. Within sites, there was also evidence of small temporal shifts in genetic composition (Figure 3), which might be partly driven by variable source contributions each year (Figure S2.1), as has been observed in other marine invertebrate and fish species (Cuif et al., 2015; Harrison et al., 2020; Hogan et al., 2010; Lotterhos & Markel, 2012; Moody et al., 2015; Toonen & Grosberg, 2011). Sweepstakes reproduction can be an important contributor to genetic structure in marine metapopulations (Broquet et al., 2013; Hedgecock et al., 2007; Hedgecock & Pudovkin, 2011). Indeed, our examination of genetic diversity provided some limited evidence for signatures of sweepstakes in *B. cocosensis*, but only in a subset of site-by-year comparisons (Figures S2.2 and S2.3). Our results collectively imply that the spatial genetic structure among subpopulations of *B. cocosensis* waxes and wanes through time, probably shaped by temporal variation in the sources of new recruits and variance in reproductive success within and among subpopulations.

Chaotic genetic patchiness in marine metapopulations is characterised by small-but-significant genetic differences among subpopulations (Johnson & Black, 1982). In our study of *B. cocosensis*, the largest nonzero F_{ST}-value was 0.021, based on genome-wide pool-seq SNP data (Figures 2–3 and S2.1; Tables S2.1–S2.3). Other studies in marine organisms using individually sequenced SNP genotypes have observed similar levels of genetic structure among subpopulations. For

example, $F_{ST} \leq 0.015$ in *Lutjanus carponotatus* snapper (DiBattista et al., 2017), $F_{ST} \leq 0.021$ in *Siphamia tubifer* cardinalfish, and $F_{ST} \leq 0.027$ in settling *Jasus edwardsii* lobster (Villacorta-Rath et al., 2018). Although many studies have demonstrated such statistically significant genetic differentiation in diverse marine taxa, it is largely unclear whether this genetic differentiation translates into biologically significant differences. In other words, does patchy genetic structure also translate into patchy distributions of phenotypic variation, or substantial shifts in the genomic architecture of adaptation? Marine systems have historically played a pivotal role in generating hypotheses about the role of selection and dispersal in structuring the distribution of phenotypic variation across space (Weldon, 1894), through time (Thompson, 1897), or over ontogeny (Weldon, 1895), within species. Yet the microevolutionary implications of chaotic genetic patchiness are virtually unstudied in the literature, where studies have largely focused on characterising patterns of genetic variation. Quantifying the biological and ecological consequences of processes that produce chaotic genetic patterns would provide considerable insights into the dynamics of marine microevolution.

4.2 | Signatures of spatially divergent selection are variable, but some loci are consistent outliers among years and life stages

Whether selection is predictable, or variable, has important consequences for the direction of adaptive change within and among populations. We used outlier locus analyses to infer the action of spatially divergent selection in the *B. cocosensis* metapopulation for each year-by-life stage combination. Most outlier loci that we identified were unique to their sampling year (Figure 4a). These results are consistent with temporally variable, and ontogenetic shifts in, spatial selection (Figure 1c; Tables 2 and 3). Yet it is important to note that testing for variable selection using outlier locus analyses is potentially problematic in the presence of high genetic drift (discussed below). Our study occurred during an El Niño and La Niña oscillation in Australia, with the El Niño at its peak in 2015 and transitioning into the La Niña in 2016. These climatic dynamics probably had implications for dispersal and selection in the *B. cocosensis* metapopulation, potentially affecting the genetic composition of local subpopulations, the direction and magnitude of selection, and could explain why we recovered so few temporally consistent outlier loci. However, with only three sites sampled over three consecutive years, our study is unable to address the impact of climate oscillations on the *B. cocosensis* metapopulation adequately.

Despite considerable temporal variability in spatially divergent outlier loci, one locus was shared between 2015 and 2016 in our analyses of juvenile *B. cocosensis*, and three loci overlapped between life stages when summing across years. Therefore, some loci may be predictable targets of spatial selection over time and between life stages based on our replicated tests (Figure 4; Tables 2 and 3). Independent tests that converge on a set of overlapping loci provides greater confidence that observed "hits" are biologically important, relative to single point estimates with false discovery correction alone. For outlier loci with the greatest overlap in our study (Figure 5), BLAST searches using the associated RAD contigs suggested that at least one of these outlier loci, snp_26942_300, might be in a genic region, or tightly linked to one. However, apart from providing additional confidence that we might have recovered some biologically relevant outlier loci, we refrain from interpreting this possible functional genetic variation.

Our work in *B. cocosensis* and other marine genomic studies utilising replicated designs (Babin et al., 2017; Bourret et al., 2014; Gould & Dunlap, 2017; Villacorta-Rath et al., 2018) highlight pitfalls for inferring adaptive evolution from outlier locus analyses based on single timepoint samples. The outlier locus methods, PCADAPT and OUTFLANK, employed in our study perform well under many demographic scenarios (Luu et al., 2017; Whitlock & Lotterhos, 2015). However, stochasticity in dispersal and reproduction reduce power to discern between adaptive versus neutral factors that generate anomalous locus-specific differentiation (Babin et al., 2017; Hofer et al., 2009; Klopfstein et al., 2006). Additionally, stochastic neutral processes might change the genomic background of selection, which would be particularly problematic for polygenic adaptation. In polygenic traits, the effect size of individual loci is small, such that the genetic architecture of adaptation becomes less stable in the presence of high gene flow and drift (Babin et al., 2017; Bernatchez, 2016; Rey et al., 2020; Yeaman, 2015; Yeaman & Otto, 2011; Yeaman & Whitlock, 2011). Small effect loci and unstable genetic architectures are therefore less likely to yield replicable results in outlier locus analyses. Finally, when measured alleles are not under selection themselves but are linked to causal variants, it is possible that fluctuating allele frequencies are due to changes in haplotype composition, because measured alleles might exist in different phases with causal variants. These factors suggest that inferences of selection based on an outlier locus analysis from a single timepoint may be misleading. Indeed, in our study the stochasticity in both the identity and allele frequencies of outlier loci is challenging to reconcile (Figure 5; Tables 2 and 3), notwithstanding some overlapping results across independent sampling events.

When organisms possess complex life cycles, sampling multiple life stages is important for characterising when important genetic changes occur, and hence, inferring the timing of selective agents. For example, genetic variation among microhabitats at the *Mpi* locus in an intertidal barnacle shifts from being homogeneous to significantly differentiated from larval settlement to adulthood (Schmidt & Rand, 2001). In sea bream, post-settlement selection in brackish lagoon versus marine habitats causes polygenic differentiation during the recruitment of juveniles from a panmictic larval gene pool (Rey et al., 2020). Therefore, patterns of genetic variation might be subject to potential "carry-over" effects from processes acting in earlier life stages that cease later in life. Such phenomena have been investigated for morphological variation in waterfall climbing gobies (Diamond et al., 2019; Moody et al., 2015) and have been extensively studied with respect to the life histories of marine organisms (D'Alessandro et al., 2013; Gagliano et al., 2007; Hamilton et al.,

2008; Raventós & Macpherson, 2005; Shima & Swearer, 2010). Our study was not equipped to directly address the role of genetic carry-over effect across ontogeny in *B. cocosensis* because the year-round recruitment in this species makes it challenging to track discrete cohorts through time. However, extending such considerations to genetic variation might provide novel insights into ontogenetic shifts in selection and constraints to adaptation in marine environments (Marshall & Morgan, 2011; Rey et al., 2020). Certainly, sampling multiple ecological dimensions provides greater resolution on when and where selective processes operate, and hence their relevant scale, in marine metapopulations.

4.3 | Phenotypic variation is temporally variable, but exhibits spatial predictability

Head shape in *B. cocosensis* varies with microhabitat over small spatial scales within local sites (Malard et al., 2016) and is likely to be an ecologically important phenotype for this species. Despite temporal shifts of head shape phenotypes within each site, Hastings Point and Shellharbour were consistently more similar to one another than either site to Point Cartwright (Figure 6a and S2.8). This result was surprising given the spatial proximity (~1.68°) of Point Cartwright and Hastings Point (Figure 1a), the greater climatic disparity between Hastings Point and Shellharbour, and the potential north–south homogenising effect of the East Australian Current (Hoskin, 2000; Murray-Jones & Ayre, 1997; O'Kane et al., 2011). Moreover, spatial consistency among sites (Hastings Point and Shellharbour were always more similar) in their phenotypic differentiation (Figure 6a) contrasted the temporal fluctuations in genetic differentiation at outlier loci, where there was no clear spatial consistency (Figure 5). That phenotypic variation in *B. cocosensis* is unrelated to spatial distance and is heterogeneous through time suggests it exhibits "chaotic patchiness", analogous to Johnson and Black's (1982) original description of stochastic genetic patterns in marine species. Such spatiotemporally patchy phenotypic variation could arise from variable selection on heritable variation, or from phenotypic plasticity responding to environmental heterogeneity, or a combination of both, to produce "chaotic phenotypic patchiness".

Because phenotypes are the targets of selection, coanalysis of genetic and phenotypic variation can be more informative about the outcomes of gene flow–selection interactions than what might simply be deduced from outlier locus analyses alone (Cohen & Dor, 2018; Galligan et al., 2012; Garant et al., 2007; Hoekstra et al., 2004; Richardson et al., 2014) (Table 4). In our study, *B. cocosensis* eye shape and mouth length were the most spatiotemporally variable traits (Figure 6b). These traits have been shown to exhibit fine-scale phenotype–environment structuring across microhabitats within local subpopulations of *B. cocosensis* (Malard et al., 2016). In other fish species, eye and mouth morphology has been associated with intraspecific or between-ecotype variation in diet (Burress et al., 2017; Schluter, 1993; Vera-Duarte et al., 2017). Eye shape may also be important for navigating different environments in fishes (Caves et al., 2017). Our observation that significant phenotypic differentiation occurs against a background of low (but stochastic) genetic variation in *B. cocosensis* is compatible with other studies on marine or amphidromous fish. For instance, Moody et al. (2015) demonstrated predictable body shape differences among subpopulations of a waterfall climbing goby with respect to local environments, despite variable year-to-year source contributions into locally recruiting subpopulations. Rey et al. (2020) showed that body shape, growth, and condition factor exhibit habitat divergence in juvenile sea bream despite panmixia in the larval pool. In both Moody et al. (2015) and Rey et al. (2020), the phenotypes considered appeared to have a (partial) heritable genetic basis, but are likely to have appreciable environmental components as well.

The patchy distribution of head shape phenotypes in *B. cocosensis* that is uncorrelated to, and exceeds, measures of genetic differentiation implies that selection may influence phenotypic differentiation across the Australian metapopulation (Table 4). However, several unaccounted factors might also play a role in the observed phenotypic structure. We were unable to ascertain the sex of each adult fish, so cannot evaluate the potential effect of sexual dimorphisms and sex-biased collections on spatiotemporal variation in head shape. Furthermore, we are unable to assess how variation in local tidepool microhabitats sampled might affect patterns of spatiotemporal phenotypic variation. We aimed to reduce microhabitat variation by sampling across the intertidal gradient at each site. And although we endeavoured to resample the same tidepools, this was not always possible due to variation in local abundances across our study. Furthermore, local conditions are temporally variable, and we lack microhabitat measurements to perform a robust analysis of how variation in the intertidal environment within and between sites might impact the distribution of head shape phenotypes. Finally, we note that, despite our intriguing spatially consistent relationships among sites (Shellharbour and Hastings Point always more similar) and *a priori* expectations that head shape is ecologically important, the observed spatiotemporal differences might reflect neutral processes (Table 4). We are currently undertaking further investigations to understand how head shape phenotypes vary over ontogeny and to characterise the possible contributions of heritable and plastic variation. Such investigations will provide a greater understanding of the processes influencing the distribution of biological variation of *B. cocosensis* in Australia from settlement to adulthood.

4.4 | Caveats of pool-seq

Although sequencing libraries of pooled individuals is a highly cost-effective way to obtain genomic data from many populations, it does impose limitations to population genetic analyses (Futschik & Schlötterer, 2010; Schlötterer et al., 2014). Lack of information at the level of individuals precludes use of methods that require knowledge of genotypes, tests of Hardy-Weinberg equilibrium, or linkage disequilibria (Andrews et al., 2014; Cutler & Jensen, 2010; Schlötterer

et al., 2014). Furthermore, obtaining accurate allele frequency estimates and assessing population genetic structure is challenging when pool sizes and read depth are not substantial (Anderson et al., 2014; Lynch et al., 2014).

Our study of *B. cocosensis* used pool sizes of $10 \leq n \leq 30$ fish for each library, with a mean of 22 (±6.49 SD). Pool sizes of $n = 30$ diploids have been demonstrated to provide good allele frequency estimation (Gautier et al., 2013). However, our smaller sized pools might suffer from allele frequency inaccuracy and pooling error, although we note that mean and median read counts per locus were quite high (in the 100s) for many pools (Table S1.3). Moreover, through replicate sequencing of our pool-seq libraries, we found that the estimated effective contributions per pool (Gautier et al., 2013) exhibited strong concordance with the true number of pooled diploids (Table 5), so error due to variance in individual contributions is likely to be low in our analysed data set.

Indeed, our use of library replicates provided greater power and robustness relative other population genetic studies that have employed a pool-seq approach but only sequenced single libraries. This replication allowed us to devise probabilistic frameworks to understand how pooling variance affected our population genetic inferences from different replicate combinations and by simulating hypothetical parametric bootstrap data sets of allele frequencies that took estimation uncertainty into account. We believe our treatment of our replicate pool-seq data provides a rigorous interrogation of the (un)predictability of genetic patterns in *B. cocosensis*, given the limitations of pool-seq.

5 | CONCLUSION

Predicting how biological variation is structured in high gene flow systems is not straight forward because interactions between, and variability in, gene flow and selection can produce a diversity of evolutionary outcomes that are further modulated by drift. Our study in *B. cocosensis* demonstrates that biological variation is structured in a complex manner across space, time, and life stage. Our findings showcase the dynamic nature of microevolution in marine metapopulations. Variability in processes that modulate gene flow might cause marine metapopulations to fluctuate between structured and panmictic states at different spatiotemporal scales. Amidst this stochastic, but well mixed, genomic background, selection may still generate adaptive divergence. However, phenotypes and genotypes associated with high relative fitness may be subject to temporal variability. Such considerations are not just limited to marine systems but are highly applicable to any short-lived species that occupy heterogeneous environments, have complex life cycles, have high fecundity, and are highly dispersive. For any species fitting these criteria, point-in-time estimates should not be viewed as unequivocal insights into longer-term evolutionary processes. Instead, spatially, temporally, and life stage replicated data will provide the greatest insights into processes that predictably structure biological variation in the metapopulations of these species. Such information is especially imperative for conservation practitioners and environmental resource managers to make robust and informed decisions in a changing world.

ACKNOWLEDGEMENTS

We thank the Goodman Foundation (C.R. and L.L.) the Herman Slade Foundation (13/14, C.R. and L.L.) and the Ecological Society of Australia (Wiley Fundamental Ecology Award 2016, J.T.) for their generous funding. This study was conducted in accordance with animal ethics (#SBS/221/15/HSF and USyd #2015/834) and scientific collection permits (QLD Gov #174684 and NSW DPI #P13/0046-1.1). Thank you to the amazing Inkscape Project (2020) developers for providing superb graphic design freeware. We thank I. Popovic, A.M.A. Matias, K. Prata, K. Dunbar, D. Ortiz-Barrientos, O.E. Gaggiotti, three anonymous reviewers, and Subject Editor S. Rogers for providing us constructive feedback on this manuscript.

AUTHOR CONTRIBUTIONS

Conceptualisation and experimental design: Joshua Thia, Cynthia Riginos, Libby Liggins and Katrina McGuigan. Data collection: Joshua Thia, Jennifer Evans and Andrew Mather. Data analysis: Joshua Thia and Katrina McGuigan. Field and technical support: Christopher Bird and Will Figueira. First manuscript draft: Joshua Thia. Revisions of manuscript: all authors.

DATA AVAILABILITY STATEMENT

Raw sequencing reads are available on the Sequence Read Archive (NCBI BioProject PRJNA595396), with associated metadata archived in Genomic Observatories MetaDatabase (GEOME project ID 201, and expedition ID ark:/21547/DLY2). Data and scripts used in our analyses have been uploaded to Dryad (Thia, 2020: doi.org/10.5061/dryad.02v6wwq16).

ORCID

Joshua A. Thia https://orcid.org/0000-0001-9084-0959
Libby Liggins https://orcid.org/0000-0003-1143-2346
Christopher E. Bird https://orcid.org/0000-0003-0228-3318
Cynthia Riginos https://orcid.org/0000-0002-5485-4197

REFERENCES

Adams, D. C., Collyer, M., Kaliontzopoulou, A., & Sherratt, E. (2016). *Geomorph: Software for geometric morphometric analyses*. R package.

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.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Anderson, E. C., Skaug, H. J., & Barshis, D. J. (2014). Next-generation sequencing for molecular ecology: A caveat regarding pooled samples. *Molecular Ecology*, *23*(3), 502–512. https://doi.org/10.1111/mec.12609.

Andrews, K. R., Hohlenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G. (2014). Trade-offs and utility of alternative RADseq

methods: Reply to Puritz et al 2014. *Molecular Ecology*, *23*, 5943–5946. https://doi.org/10.1111/mec.12964

Appelbaum, L., Achituv, Y., & Mokady, O. (2002). Speciation and the establishment of zonation in an intertidal barnacle: Specific settlement vs. selection. *Molecular Ecology*, *11*(9), 1731–1737. https://doi.org/10.1046/j.1365-294X.2002.01560.x.

Aspi, J., Jäkäläniemi, A., Tuomi, J., & Siikamäki, P. (2003). Multilevel phenotypic selection on morphological characters in a metapopulation of *Silene tatarica*. *Evolution*, *57*(3), 509–517.

Atlas of Living Australia (2018). *Australia*, Bathygobius cocosensis: *Cocos frillgoby | atlas of living*. Retrieved from Atlas of Living Australia website: https://bie.ala.org.au.

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

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556.

Blanquart, F., Gandon, S., & Nuismer, S. L. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology*, *25*(7), 1351–1363. https://doi.org/10.1111/j.1420-9101.2012.02524.x.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bourret, V., Dionne, M., & Bernatchez, L. (2014). Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: Polygenic multilocus analysis surpasses genome scan. *Molecular Ecology*, *23*, 4444–4457. https://doi.org/10.1111/mec.12798.

Broquet, T., Viard, F., & Yearsley, J. M. (2013). Genetic drift and collective dispersal can result in chaotic genetic patchiness. *Evolution*, *67*, 1660–1675. https://doi.org/10.1111/j.1558-5646.2012.01826.x.

Bulmer, M. G. (1972). Multiple niche polymorphism. *The American Naturalist*, *106*(948), 254–257.

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

Buston, P. M., Fauvelot, C., Wong, M. Y. L., & Planes, S. (2009). Genetic relatedness in groups of the humbug damselfish *Dascyllus aruanus*: Small, similar-sized individuals may be close kin. *Molecular Ecology*, *18*(22), 4707–4715. https://doi.org/10.1111/j.1365-294X.2009.04383.x.

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233.

Carbia, P. S., & Brown, C. (2019). Environmental enrichment influences spatial learning ability in captive-reared intertidal gobies (*Bathygobius cocosensis*). *Animal Cognition*, *22*(1), 89–98. https://doi.org/10.1007/s10071-018-1225-8.

Carbia, P. S., & Brown, C. (2020). Seasonal variation of sexually dimorphic spatial learning implicates mating system in the intertidal Cocos Frillgoby (*Bathygobius cocosensis*). *Animal Cognition*, *23*(4), 621–628. https://doi.org/10.1007/s10071-020-01366-3.

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*, *97*(4), 1051–1062.

Caves, E. M., Sutton, T. T., & Johnsen, S. (2017). Visual acuity in ray-finned fishes correlates with eye size and habitat. *Journal of Experimental Biology*, *220*(9), 1586–1596.

Chenoweth, S. F., & Blows, M. W. (2008). Q_{ST} meets the G matrix: The dimensionality of adaptive divergence in multiple correlated quantitative traits. *Evolution*, *62*(6), 1437–1449. https://doi.org/10.1111/j.1558-5646.2008.00374.x.

Chong, Z., Ruan, J., & Wu, C. I. (2012). Rainbow: An integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, *28*(21), 2732–2737. https://doi.org/10.1093/bioinformatics/bts482.

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.

Ciotti, B. J., & Planes, S. (2019). Within-generation consequences of postsettlement mortality for trait composition in wild populations: An experimental test. *Ecology and Evolution*, *9*(5), 2550–2561.

Cohen, S. B., & Dor, R. (2018). Phenotypic divergence despite low genetic differentiation in house sparrow populations. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-017-18718-8.

Crispo, E. (2008). Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *Journal of Evolutionary Biology*, *21*(6), 1460–1469. https://doi.org/10.1111/j.1420-9101.2008.01592.x.

Cuif, M., Kaplan, D. M., Fauvelot, C., Lett, C., & Vigliola, L. (2015). Monthly variability of self-recruitment for a coral reef damselfish. *Coral Reefs*, *34*(3), 759–770.

Cutler, D. J., & Jensen, J. D. (2010). To pool, or not to pool? *Genetics*, *186*(1), 41–43. https://doi.org/10.1534/genetics.110.121012.

D'Alessandro, E. K., Sponaugle, S., & Cowen, R. K. (2013). Selective mortality during the larval and juvenile stages of snappers (Lutjanidae) and great barracuda *Sphyraena barracuda*. *Marine Ecology Progress Series*, *474*, 227–242.

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*, *189*, 385–398. https://doi.org/10.1007/s00360-019-01212-0.

da Silva, C. R. B., Wilson, R. S., & Riginos, C. (2019). Rapid larval growth is costly for post-metamorphic thermal performance in a Great Barrier Reef fish. *Coral Reefs*, *38*(5), 895–907.

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

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-Seq. *Molecular Ecology*, *26*(1), 25–42.

Di Franco, A., Gillanders, B. M., De Benedetto, G., Pennetta, A., De Leo, G. A., & Guidetti, P. (2012). Dispersal patterns of coastal fish: Implications for designing networks of marine protected areas. *PLoS One*, *7*(2), e31681.

Diamond, K. M., Lagarde, R., Schoenfuss, H. L., Walker, J. A., Ponton, D., Blob, R. W., & Réunion, L. (2019). Relationship of escape performance with predator regime and ontogeny in fishes. *Biological Journal of the Linnean Society*, *127*, 324–336.

DiBattista, J. D., Travers, M. J., Moore, G. I., Evans, R. D., Newman, S. J., Feng, M., Berry, O. (2017). Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia. *Molecular Ecology*, *26*(22), 6206–6223.

Ebenman, B. (1992). Evolution in organisms that change their niches during the life cycle. *The American Naturalist*, *139*(5), 990–1021.

Evans, J. L., Thia, J. A., Riginos, C., & Hereward, J. P. (2018). The complete mitochondrial genome of *Bathygobius cocosensis* (Perciformes, Gobiidae). *Mitochondrial DNA Part B: Resources*, *3*(1), 217–219. https://doi.org/10.1080/23802359.2018.1437824.

Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, *186*(1), 207–218. https://doi.org/10.1534/genetics.110.114397.

Gaggiotti, O. E. (2017). Metapopulations of marine species with larval dispersal: A counterpoint to Ilkka's glanville fritillary metapopulations. *Annales Zoologici Fennici*, *54*(1-4), 97–112.

Gagliano, M., McCormick, M. I., & Meekan, M. G. (2007). Survival against the odds: Ontogenetic changes in selective pressure mediate growth-mortality trade-offs in a marine fish. *Proceedings of the Royal Society B*, *274*(1618), 1575–1582.

Galligan, T. H., Donnellan, S. C., Sulloway, F. J., Fitch, A. J., Bertozzi, T., & Kleindorfer, S. (2012). Panmixia supports divergence with gene flow in Darwin's small ground finch, *Geospiza fuliginosa*, on Santa Cruz. *Galápagos Islands*. *Molecular Ecology*, *21*(9), 2106–2115. https://doi.org/10.1111/j.1365-294X.2012.05511.x.

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

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv*, *1207*, 3907.

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C., & Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool versus individual-based genotyping. *Molecular Ecology*, *22*, 3766–3779. https://doi.org/10.1111/mec.12360.

Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P., & Miller-Sims, V. (2007). Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(3), 858–863. https://doi.org/10.1073/pnas.0606777104.

Gibert, P., Debat, V., & Ghalambor, C. K. (2019). Phenotypic plasticity, global change, and the speed of adaptive evolution. *Current Opinion in Insect Science*, *35*, 34–40. https://doi.org/10.1016/j.cois.2019.06.007.

Gould, A. L., & Dunlap, P. V. (2017). Genomic analysis of a cardinalfish with larval homing potential reveals genetic admixture in the Okinawa Islands. *Molecular Ecology*, *26*(15), 3870–3882. https://doi.org/10.1111/mec.14169.

Grant, P. R., & Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's Finches. *Science*, *296*(5568), 707–711. https://doi.org/10.1126/science.1070315.

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.

Hamilton, S. L., Regetz, J., & Warner, R. R. (2008). Postsettlement survival linked to larval life in a marine fish. *Proceedings of the National Academy of Sciences*, *105*(5), 1561–1566.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, *396*(6706), 41.

Hanski, I., Mononen, T., & Ovaskainen, O. (2010). Eco-evolutionary metapopulation dynamics and the spatial scale of adaptation. *The American Naturalist*, *177*(1), 29–43.

Harrison, H. B., Bode, M., Williamson, D. H., Berumen, M. L., & Jones, G. P. (2020). A connectivity portfolio effect stabilizes marine reserve performance. *Proceedings of the National Academy of Sciences*, *117*(41), 25595–25600. https://doi.org/10.1073/pnas.1920580117.

Harrison, S., & Hastings, A. (1996). Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology & Evolution*, *11*(4), 180–183.

Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of marine organisms? In A. R. Beaumont (Ed.), *Genetics and evolution of aquatic organisms* (pp. 1222–1344). Chapman & Hall.

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.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: A review and commentary. *Bulletin of Marine Science*, *87*(4), 971–1002. https://doi.org/10.5343/bms.2010.1051.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, *210*(1), 315–330. https://doi.org/10.1534/genetics.118.300900.

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

Hofer, T., Ray, N., Wegmann, D., & Excoffier, L. (2009). Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics*, *73*(1), 95–108. https://doi.org/10.1111/j.1469-1809.2008.00489.x.

Hogan, J. D., Thiessen, R. J., & Heath, D. D. (2010). Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. *Marine Ecology Progress Series*, *417*, 263–275. https://doi.org/10.3354/meps08793.

Hogan, J. D., Thiessen, R. J., Sale, P. F., & Heath, D. D. (2012). Local retention, dispersal and fluctuating connectivity among populations of a coral reef fish. *Oecologia*, *168*(1), 61–71. https://doi.org/10.1007/s00442-011-2058-1.

Hoskin, M. G. (2000). Effects of the East Australian Current on the genetic structure of a direct developing muricid snail (*Bedeva hanleyi*, Angas): Variability within and among local populations. *Biological Journal of the Linnean Society*, *69*, 245–262. https://doi.org/10.1006/bijL.

Inkscape Project (2020). *Inkscape*. Retrieved from www.inkscape.org/en.

Jackson, T. M., Roegner, G. C., & O'Malley, K. G. (2017). Evidence for interannual variation in genetic structure of Dungeness crab (*Cancer magister*) along the California Current System. *Molecular Ecology*, *27*, 352–368.

Johnson, M. S., & Black, R. (1982). Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology*, *70*, 157–164. https://doi.org/10.1007/BF00397680.

Jones, G. P., Almany, G. R., Russ, G. R., Sale, P. F., Steneck, R. S., Van Oppen, M. J. H., & Willis, B. L. (2009). Larval retention and connectivity among populations of corals and reef fishes: History, advances and challenges. *Coral Reefs*, *28*, 307–325. https://doi.org/10.1007/s00338-009-0469-9.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*(3), 482–490. https://doi.org/10.1093/molbev/msj057.

Kremer, A., Zanetto, A., & Ducousso, A. (1997). Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics*, *145*, 1229–1241.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923.

Li, H. (2019). *Toolkit for processing sequences in FASTA/Q formats*. Retrieved from https://github.com/lh3/seqtk.

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

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

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659.

Liggins, L., Booth, D. J., Figueira, W. F., Treml, E. A., Tonk, L., Ridgway, T., & Riginos, C. (2015). Latitude-wide genetic patterns reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries of a tropical marine fish. *Ecography*, 1–13, *38*(12), 1212–1224. https://doi.org/10.1111/ecog.01398.

Lotterhos, K. E., & Markel, R. W. (2012). Oceanographic drivers of offspring abundance may increase or decrease reproductive variance in a temperate marine fish. *Molecular Ecology*, *21*(20), 5009–5026.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, *23*(9), 2178–2192. https://doi.org/10.1111/mec.12725.

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592.

Lynch, M., Bost, D., Wilson, S., Maruki, T., & Harrison, S. (2014). Population-genetic inference from pooled-sequencing data. *Genome Biology and Evolution*, *6*(5), 1210–1218. https://doi.org/10.1093/gbe/evu085.

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.

Marshall, D. J., Monro, K., Bode, M., Keough, M. J., & Swearer, S. (2010). Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters*, *13*, 128–140. https://doi.org/10.1111/j.1461-0248.2009.01408.x.

Marshall, D. J., & Morgan, S. G. (2011). Ecological and evolutionary consequences of linked life-history stages in the sea. *Current Biology*, *21*(18), R718–R725.

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

Moody, K. N., Wren, J. L. K., Kobayashi, D. R., Blum, M. J., Ptacek, M. B., Blob, R. W., Toonen, R. J., Schoenfuss, H. L., & Childress, M. J. (2019). Evidence of local adaptation in a waterfall-climbing Hawaiian goby fish derived from coupled biophysical modeling of larval dispersal and post-settlement selection. *BMC Evolutionary Biology*, *19*(1), 88.

Moran, N. A. (1994). Adaptation and constraint in the complex life cycles of animals. *Annual Review of Ecology and Systematics*, *25*, 573–600.

Murray-Jones, S. E., & Ayre, D. J. (1997). High levels of gene flow in the surf bivalve *Donax deltoides* (Bivalvia: Donacidae) on the east coast of Australia. *Marine Biology*, *128*(1), 83–89. https://doi.org/10.1007/s002270050071.

Nosil, P. (2009). Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution*, *63*(7), 1902–1912. https://doi.org/10.1111/j.1558-5646.2009.00671.x.

O'Kane, T. J., Oke, P. R., & Sandery, P. A. (2011). Predicting the East Australian Current. *Ocean Modelling*, *38*, 251–266. https://doi.org/10.1016/j.ocemod.2011.04.003.

Paccard, A., Wasserman, B. A., Hanson, D., Astorg, L., Durston, D., Kurland, S., Apgar, T. M., El-Sabaawi, R. W., Palkovacs, E. P., Hendry, A. P., & Barrett, R. D. H. (2018). Adaptation in temporally variable environments: Stickleback armor in periodically breaching bar-built estuaries. *Journal of Evolutionary Biology*, *31*(5), 735–752. https://doi.org/10.1111/jeb.13264.

Pujolar, J. M., Jacobsen, M. W., Bekkevold, D., Lobón-Cervià, J., Jónsson, B., Bernatchez, L., & Hansen, M. M. (2015). Signatures of natural selection between life cycle stages separated by metamorphosis in European eel. *BMC Genomics*, *16*(1), 1–15. https://doi.org/10.1186/s12864-015-1754-3.

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431. https://doi.org/10.7717/peerj.431.

R Core Team (2018). *R: A language and environment for statistical computing*. Retrieved from http://www.r-project.org.

Raventós, N., & Macpherson, E. (2005). Effect of pelagic larval growth and size at hatching on the post-settlement survivorship in two temperate labrid fishes of the genus Symphodus. *Marine Ecology Progress Series*, *285*, 205–211.

Rencher, A. C. (1998). *Multivariate statistical inference and applications*. Wiley.

Rey, C., Darnaude, A., Ferraton, F., Guinand, B., Bonhomme, F., Bierne, N., & Gagnaire, P.-A. (2020). Within-generation polygenic selection shapes fitness-related traits across environments in Juvenile Sea bream. *Genes*, *11*(4), 398.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution*, *29*(3), 165–176. https://doi.org/10.1016/j.tree.2014.01.002.

Rohlf, F. J. (2015). tpsDig. Retrieved from http://life.bio.sunysb.edu/morph.

Rolshausen, G., Muttalib, S., Kaeuffer, R., Oke, K. B., Hanson, D., & Hendry, A. P. (2015). When maladaptive gene flow does not increase selection. *Evolution*, *69*(9), 2289–2302.

Scheiner, S. M. (1993). Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics*, *24*, 35–68.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — Mining genome-wide polymorphism data without big funding. *Nature Publishing Group*, *15*(11), 749–763. https://doi.org/10.1038/nrg3803.

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

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

Schmidt, P. S., & Rand, D. M. (2001). Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of *Mpi* genotypes. *Evolution*, *55*(7), 1336–1344.

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.

Shima, J. S., & Swearer, S. E. (2010). The legacy of dispersal: Larval experience shapes persistence later in the life of a reef fish. *Journal of Animal Ecology*, *79*, 1308–1314. https://doi.org/10.1111/j.1365-2656.2010.01733.x.

Staton, E. (2016). *Sync paired-end FASTA/Q files and keep singleton reads*. GitHub. Retrieved from https://github.com/sestaton/Pairfq

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

Sultan, S. E., & Spencer, H. G. (2002). Metapopulation structure favors plasticity over local adaptation. *The American Naturalist*, *160*(2), 271–283.

Thia, J. A. (2020). Genetic and phenotypic variation in *Bathygobius cocosensis* from East Australia (2014–2016). *Dryad*, https://doi.org/10.5061/dryad.02v6wwq16.

Thia, J. A., & Riginos, C. (2019). *genomalicious*: Serving up a smorgasbord of R functions for population genomic analyses. *BioRxiv*, *667337*, https://doi.org/10.1101/667337.

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.

Thompson, H. (1897). On certain changes observed in the dimensions of parts of the carapace of *Carcinus mœnas*. *Proceedings of the Royal Society of London*, *60*, 195–198. https://doi.org/10.1098/rspl.1896.0037.

Toonen, R. J., & Grosberg, R. K. (2011). Causes of chaos: Spatial and temporal genetic heterogeneity in the intertidal anomuran crab

Petrolisthes cinctipes. In C. Held, S. Koenemann, & C. D. Schubart (Eds.), *Phylogeography and population genetics in Crustacea* (pp. 75–107). CRC Press.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ*, *1*, e203. https://doi.org/10.7717/peerj.203.

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

Villacorta-Rath, C., Souza, C. A., Murphy, N. P., Green, B. S., Gardner, C., & Strugnell, J. M. (2018). Temporal genetic patterns of diversity and structure evidence chaotic genetic patchiness in a spiny lobster. *Molecular Ecology*, *27*, 54–65.

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

Watson, J. R., Kendall, B. E., Siegel, D. A., & Mitarai, S. (2012). Changing seascapes, stochastic connectivity, and marine metapopulation dynamics. *The American Naturalist*, *180*(1), 99–112.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Weldon, W. F. R. (1894). II. On certain correlated variations in *Carcinus mænas*. *Proceedings of the Royal Society of London*, *54*, 318–328.

Weldon, W. F. R. (1895). An attempt to measure the death-rate due to the selective destruction of *Carcinus mænas* with respect to a particular dimension. *Proceedings of the Royal Society of London*, *57*, 360–379. https://doi.org/10.1098/rspl.1894.0165.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B*, *277*(1688), 1685–1694. https://doi.org/10.1098/rspb.2009.2214.

White, G. E., Hose, G. C., & Brown, C. (2015). Influence of rock-pool characteristics on the distribution and abundance of inter-tidal fishes. *Marine Ecology*, *36*(4), 1332–1344. https://doi.org/10.1111/maec.12232.

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

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist*, *186*(S1), S24–S36. https://doi.org/10.1086/682949.

Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, *7*, 1338.

Wright, S. (1949). The genetical structure of populations. *Annals of Human Genetics*, *15*(1), 323–354.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, *186*, S74–S89.

Yeaman, S., & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution: International Journal of Organic Evolution*, *65*(7), 2123–2129.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution: International Journal of Organic Evolution*, *65*(7), 1897–1911.

Yearsley, J. M., Viard, F., & Broquet, T. (2013). The effect of collective dispersal on the genetic structure of a subdivided population. *Evolution*, *67*(6), 1649–1659. https://doi.org/10.1111/evo.12111.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Thia JA, McGuigan K, Liggins L, et al. Genetic and phenotypic variation exhibit both predictable and stochastic patterns across an intertidal fish metapopulation. *Mol Ecol*. 2021;00:1–23. https://doi.org/10.1111/mec.15829