Pre-introduction introgression contributes to parallel differentiation and contrasting hybridisation outcomes between invasive and native marine mussels

Running title: (Un)parallel evolution in a marine invader

Iva Popovic[1*], Nicolas Bierne[2], Federico Gaiti[3,4], Miloš Tanurdžić[1] and Cynthia Riginos[1]

[1]School of Biological Sciences, University of Queensland, St Lucia, Queensland 4072, Australia

[2]Institut des Sciences de l'Evolution UMR 5554, Université de Montpellier, CNRS-IRD-EPHE-UM, France

[3]Weill Cornell Medicine, New York, NY, USA

[4]New York Genome Center, New York, NY, USA

*Corresponding Author:

Email: iva.popovic@uq.net.au

Tel: +61450658950

ORCID ID: 0000-0001-6582-4236

Email contact for other authors:

Nicolas Bierne: nicolas.bierne@umontpellier.fr

Federico Gaiti: feg2007@med.cornell.edu

Miloš Tanurdžić: m.tanurdzic@uq.edu.au

Cynthia Riginos: c.riginos@uq.edu.au

Acknowledgements

We would like to thank F. Viard, J. Thia, D. Matias, M. James, and D. Ortiz-Barrientos for valuable discussions and insight on the manuscript. The QRIScloud computing cluster at the University of Queensland provided computational infrastructure for this project. This work is supported by the Australian Biological Resources Study (ABRS) National Taxonomy Research Grant (grant number RF216-11), and awards from the National Sciences and Engineering Research Council of Canada (to I.P), the Society for the Study of Evolution (to I.P) and the University of Queensland (to I.P).

Pre-introduction introgression contributes to parallel differentiation and contrasting hybridisation outcomes between invasive and native marine mussels

Abstract

Non-native species experience novel selection pressures in introduced environments and may interbreed with native lineages. Species introductions therefore provide opportunities to investigate repeated patterns of adaptation and introgression across replicated contact zones. Here, we investigate genetic parallelism between multiple introduced populations of the invasive marine mussel, *Mytilus galloprovincialis*, in the absence (South Africa and California) and presence of hybridisation with a native congener (*Mytilus planulatus* in Batemans Bay and Sydney Harbour, Australia). Repeatability in post-introduction differentiation from native-range populations varied between genetically distinct Atlantic and Mediterranean lineages, with Atlantic-derived introductions displaying high differentiation (*maxFST*>0.4) and parallelism at outlier loci. Identification of long non-coding RNA transcripts (lncRNA) additionally allowed us to clarify that parallel responses are largely limited to protein-coding loci, with lncRNAs likely evolving under evolutionary constraints. Comparisons of independent hybrid zones revealed differential introgression most strongly in Batemans Bay, with an excess of *M. galloprovincialis* ancestry and resistance to introgression at loci differentiating parental lineages (*M. planulatus* and Atlantic *M. galloprovincialis*). Additionally, contigs putatively introgressed with divergent alleles from a closely related species, *Mytilus edulis,* showed stronger introgression asymmetries compared to genome-wide trends and also diverged in parallel in both Atlantic-derived introductions. These results suggest that divergent demographic histories experienced by introduced lineages, including pre-introduction introgression, influences contemporary admixture dynamics. Our findings build on previous investigations reporting contributions of historical introgression to intrinsic reproductive architectures shared between marine lineages and illustrate that interspecific introgression history can shape differentiation between colonising populations and their hybridisation with native congeners.

Key words: *Mytilus*, parallel evolution, introgression, hybrid zone, invasive species, genetic differentiation

Introduction

Biological invasions are leading threats to global biodiversity and central components of global change (Chown et al., 2015; Simberloff, 2013). The impacts of introduced species (including invasive species) on native communities can be diverse and complex, and subsequently difficult to predict (Oduor, 2013). Since the earliest studies on biological invasions (Baker, 1965, 1974), the literature now contains numerous examples of shared ecological traits contributing to increased survival and competitive abilities among introduced species (e.g., rapid growth rates and high fecundity through asexual reproduction; Richardson, 2011). There is also growing evidence that local adaptation is often a significant factor in the success of introduced populations (Colautti & Lau, 2015; Dlugosch & Parker, 2008; Prentis, Wilson, Dormontt, Richardson, & Lowe, 2008). Indeed, repeated evolution of ecologically important traits related to life history strategies, reproduction, growth and competition have been documented for introduced populations in both plants and animals (Colautti & Lau, 2015; Ferrero et al., 2015; Lee, 2016; Phillips, Brown, Webb, & Shine, 2006; K. G. Turner, Hufbauer, & Rieseberg, 2014; Vandepitte et al., 2014). We know relatively less, however, about how heritable genetic variation is redistributed or altered to yield adaptive responses following introduction (Bock et al., 2015). Because anthropogenic introductions are recent, selection on standing genetic variation from the native range is expected to be the primary genetic mechanism of adaptation in introduced populations (Prentis et al., 2008), thus increasing the likelihood that parallel genetic changes may accompany multiple independent introductions of the same species.

Genetic and genomic studies examining parallel evolutionary change in natural populations have shown that adaptation can be highly constrained and repeatable (Bolnick, Barrett, Oke, Rennison, & Stuart, 2018; Conte, Arnegard, Peichel, & Schluter, 2012; Elmer & Meyer, 2011). Replicated genetic changes across distinct lineages at various genomic scales are now documented, ranging from single amino acid polymorphisms (e.g., Hoekstra, Hirschmann, Bundey, Insel, & Crossland, 2006) to individual genes (e.g., Colosimo et al., 2005; Stern & Orgogozo, 2009) and large genomic regions (e.g., Mérot et al., 2018; Van Belleghem et al., 2018) including at the chromosomal scale (e.g., Riquet et al., 2019). Comparative evidence for parallel evolution has also been demonstrated among closely related populations (Jones et al., 2012; T. L. Turner, Bourne, Von Wettberg, Hu, & Nuzhdin, 2010), suggesting

that adaptation can be predictable even on timescales relevant for invasive species research (Marques et al., 2016). Most studies of genetic repeatability, however, are predicated on the assumptions that focal populations have evolved under similar selective pressures (Arendt & Reznick, 2008; Bolnick et al., 2018; Rennison, Delmore, Samuk, Owens, & Miller, 2020; Stern, 2013). Invasive species are likely to experience a variety of new ecological conditions under which they did not evolve (Colautti & Lau, 2015) and may occupy vastly diverse abiotic conditions and habitats in their introduced ranges (Guzinski, Ballenghien, Daguin-Thiébaut, Lévêque, & Viard, 2018; Tepolt, 2015). Invasive species are also likely to encounter congeneric species and gene flow may ensue if isolating barriers to reproduction are semi-permeable (Viard, David, & Darling, 2016; Viard, Riginos, & Bierne, 2020).

Introduced species adaptation frequently occurs in the context of gene flow between previously allopatric taxa (Bay, Taylor, & Schluter, 2019; Fitzpatrick et al., 2010). Such examples of very recent secondary contact can provide new insights into speciation genetics, especially how different genomic elements are shaped by selection and gene flow (Dlugosch, Anderson, Braasch, Cang, & Gillette, 2015; Ravinet et al., 2017). For instance, introgression from differentiated genetic backgrounds, such as native species, can contribute locally adapted alleles or new beneficial allelic combinations to introduced lineages (Bay & Ruegg, 2017; Huerta-Sánchez et al., 2014; Martin & Jiggins, 2017). Conversely, gene flow is expected to break down linkage disequilibria between sets of already well-adapted alleles, potentially impeding opportunities for post-introduction adaptation (Kirkpatrick & Barton, 2006; Nachman & Payseur, 2012; Sousa & Hey, 2013). Disassociation of locally adapted alleles can be slowed down, however, if beneficial alleles are tightly linked in low recombination regions, leading to heterogeneous genomic differentiation between populations in the presence of gene flow (Bierne, Roze, & Welch, 2013; Harrison & Larson, 2016). These low-recombining regions likely experience strong linked selection during periods of allopatry that can accelerate lineage sorting and promote ‘islands of differentiation’ between closely related lineages (Burri, 2017; Burri et al., 2015; Cruickshank & Hahn, 2014). Low-recombining regions are also the most likely to accumulate differences associated with reproductive barriers or local adaption loci that resist or delay introgression during secondary contacts (Barton & Bengtsson, 1986; Duranton et al., 2018; Ravinet et al., 2017; Riquet et al., 2019).

Islands of genomic differentiation are common in natural divergence-with-gene-flow situations and emerging results suggest they are also frequent when secondary contact is human-caused. Genomic islands in regions of low recombination have been implicated in maintaining interspecific differences through association with introgression barriers between native-introduced oysters (Gagnaire et al., 2018). Similarly, empirical evidence for parallel differentiation between replicated anthropogenic contact zones suggests that genome-wide admixture can lead to repeatable divergence when secondary contact is recent (Bay et al., 2019; Simon, Arbiol, et al., 2020). More broadly, theory suggests that interactions between concurrent divergent selection and gene flow may further constrain where differentiation can occur, thus favouring genomic 'reuse' or repeated differentiation in genomic regions (i.e. co-adapted alleles) that are less likely to be broken up by recombination (Samuk et al., 2017). Such genomic constraints may therefore increase the probability that the same loci are involved in maintaining differentiation between populations (Samuk et al., 2017), despite the likelihood of diverse selective pressure experienced by introduced populations in novel environments (Riquet et al., 2019; Tepolt, 2015).

Successful marine invasions provide exceptional opportunities to gain a better understanding of how gene flow and selection shape differentiation in introduced environments. Processes limiting adaptation rates in marine environments appear to be less severe than theory would predict (Dlugosch & Parker, 2008; Nei, Maruyama, & Chakraborty, 1975). In general, high fecundity and highly dispersive larvae promote successive introductions of large numbers of individuals through human-mediated vectors (e.g., ballast water discharge; Carlton, 1996). High propagule pressure is therefore a key mechanism suspected to help founding populations avoid the impacts of genetic drift and associated reductions in genetic diversity (Blackburn, Lockwood, & Cassey, 2015; Lockwood, Cassey, & Blackburn, 2005; Narum et al., 2017; Roman & Darling, 2007). Empirical studies in marine invasive species have supported these predictions, rarely demonstrating strong genetic bottlenecks in introduced populations (Bernardi, Azzurro, Golani, & Miller, 2016; Gagnaire et al., 2018; Riquet et al., 2019; Rius, Turon, Bernardi, Volckaert, & Viard, 2015), such that establishment could be aided by ample standing genetic variation available to selection (Viard et al., 2016). Furthermore, the combination of frequent and dense introductions that likely underpin successful marine invasions ensure multiple worldwide introductions

as (un)naturally replicated experiments of independent introduced populations occupying diverse coastal habitats alongside recent secondary contacts.

Despite these advantages, few studies have investigated genome-wide variation and differentiation in marine non-native species across both native and introduced ranges (e.g., Bernardi et al., 2016; Gagnaire et al., 2018; Guzinski et al., 2018; Riquet, Daguin-Thiébaut, Ballenghien, Bierne, & Viard, 2013; Rohfritsch et al., 2013; Saarman & Pogson, 2015; Tepolt & Palumbi, 2015). Inadequate resolution of population differentiation (e.g., Riquet et al., 2019) or complex demographic introduction histories (e.g., Rohfritsch et al., 2013), however, have precluded a detailed examination of whether introduced populations show parallel genetic responses to colonisation. Additionally, genomic studies of marine non-native species have either largely focused on genome scans of random genetic loci (Bernardi et al., 2016; Guzinski et al., 2018; Riquet et al., 2013; Saarman & Pogson, 2015; but see Tepolt & Palumbi, 2015), or have not discriminated among putative sources of functional variation (i.e. protein-coding and non-coding variation) driving differentiation between native and introduced populations (Tepolt & Palumbi, 2015). Thus, we still have little insight regarding the genomic components of potentially selectively favoured variation that may underlie differentiation in introduced environments.

Here, we capitalise on a uniquely replicated series of introductions of the invasive marine mussel *Mytilus galloprovincialis* to investigate the role of recent evolutionary changes in the absence and presence of gene flow with a native congener. Within the native range of *M. galloprovincialis,* introgression with its sister species, *Mytilus edulis,* in Atlantic Europe, has led to pronounced genetic differentiation (Fraïsse, Belkhir, Welch, & Bierne, 2016; Quesada, Wenne, & Skibinski, 1995) and partial reproductive isolation (El Ayari, El Menif, Hamer, Cahill, & Bierne, 2019) between Atlantic and Mediterranean *M. galloprovincialis* lineages. In the present study, we validate that each of these native lineages has been the source of multiple introductions in the northern and southern hemispheres (Fig. 1a). Moreover, there has been additional replication in the nature of receiving communities with respect to hybridisation: Atlantic *M. galloprovincialis* introduced into South Africa is not sympatric with any native congeners (Branch & Steffani, 2004; Daguin & Borsa, 2000; Grant & Cherry, 1985), whereas Atlantic *M. galloprovincialis* introduced into Australia hybridises with an endemic Australian taxon, *Mytilus planulatus* (Batemans Bay: Popovic, Matias, Bierne, & Riginos, 2020). Similarly, Mediterranean *M.*

149 *galloprovincialis* has displaced native *Mytilus trossulus* in southern California (Geller, 1999; Hilbish et
150 al., 2010; Rawson, Agrawal, & Hilbish, 1999) and only few hybrids are detected (Saarman & Pogson,
151 2015), whereas Mediterranean *M. galloprovincialis* introduced to Australia is admixed with *M. planulatus*
152 (Sydney Harbour: Popovic et al., 2020). The unintentional introduction history of *M. galloprovincialis*
153 provides an exceptional framework for investigating the reuse of standing genetic variation in the context
154 of adaptation to new environments with contrasting opportunities for gene flow with native species.
155
156 We use transcriptome-wide markers and focus on repeated patterns of genomic differentiation to
157 investigate whether introduced *M. galloprovincialis* in Australia and (non-hybridising) introductions in
158 California and South Africa have undergone parallel genetic change relative to their native genetic
159 backgrounds. We also quantify the extent of repeatability between independent admixture events in
160 Australia involving divergent *M. galloprovincialis* lineages that differ in colonisation histories.
161 Comparisons of allele frequency deviations from the average genomic admixture rate allowed us to
162 estimate genome-wide introgression patterns in each hybrid population and to determine whether loci
163 differentiating parental taxa deviate towards introduced *M. galloprovincialis* ancestry or resist
164 introgression (due to isolating barriers or selective processes). Because resolving the sources of
165 selectively favoured variation can improve our understanding of the genetic mechanisms underlying
166 adaptation (Prentis et al., 2008), we also examine whether differentiation and parallel responses involve
167 both protein-coding transcripts and major non-coding elements of the transcriptome, long non-coding
168 RNAs (lncRNAs). Non-coding elements of the transcriptome have been implicated as important sources
169 of genetic variation during invasion (Stapley, Santure, & Dennis, 2015) and there is evidence that higher
170 RNA structures are involved with thermal adaptation in bacteria and may have similar roles in eukaryotes
171 (de la Fuente, Valera, & Martínez-Guitarte, 2012; Gaiti, Calcino, Tanurdžić, & Degnan, 2017; Somero,
172 2018). Thus, lncRNAs may experience selection on ecological timescales relevant for studying
173 populations or congeners (Pang, Frith, & Mattick, 2006; Somero, 2018), but remain unexamined in the
174 context of biological invasions.
175
176 Our comparative investigation shows that repeatability of post-introduction differentiation and admixture
177 varies between genetically distinct introduced lineages, where historical introgression influences
178 contemporary evolution in independent introductions. The genetic data presented here represent the

first genomic investigation of multiple *M. galloprovincialis* introductions at a global scale and the first comparative investigation of lncRNA evolution in an invasive species to date.

Methods

Sample collection, RNA extraction and sequencing

We obtained published RNAseq data for three *Mytilus* species: *M. galloprovincialis* from its native (n=15) and introduced range (n=18) in Australia*, M. planulatus* (n=5) *and M. edulis* (n=3) from our previous investigation of *M. galloprovincialis* introductions in the southern hemisphere (Table 1; Popovic et al., 2020; BioProject ID: PRJNA560413). We combined these data with new samples from introduced *M. galloprovincialis* in California (n=5) and South Africa (n=5) (Fig. 1a). Mussels were collected from rocky intertidal or subtidal environments (Table 1). Individuals in California were sampled outside of the southern hybrid zone boundary with *Mytilus trossulus*, to avoid sampling introgressed individuals (Saarman & Pogson, 2015). All individuals were genotyped for the species diagnostic marker *Glu-5'* (Rawson et al., 1996) to confirm species identity. RNA was extracted from 10-20 mg of mantle tissue (preserved in RNAlater), using the RNeasy Plant Mini Kit (Qiagen, MD, USA) and following the Animal Tissues protocol. Strand specific cDNA libraries were constructed using the TruSeq stranded mRNA kit (Illumina) with average insert sizes of 250-300 bp. Paired-end libraries were sequenced across three lanes of an Illumina Hiseq2000 or across one lane of an Illumina Hiseq4000.

RNAseq data processing

We removed adapters from paired reads using Trimmomatic (v0.36) (Bolger, Lohse, & Usadel, 2014). We selected high quality reads using a Phred-scale average quality score of 20 within 4 bp sliding windows and a minimum size filter of 50 bp. Redundancy among high-coverage reads was reduced by digitally normalising each dataset using Trinity's insilico_read_normalization.pl script with a default kmer size of 25 and maximum read coverage of 50 (Grabherr et al., 2011). Overlapping paired reads were merged using FLASH v1.2.11 (Magoč & Salzberg, 2011) with a minimum overlap length of 10 bp. Read mapping, lncRNA identification and all downstream analyses were conducted against a *de novo M.*

galloprovincialis reference transcriptome obtained from a previously published study (Popovic et al., 2020 for *de novo* assembly details).

Identification of lncRNA transcripts

Putative lncRNA transcripts were identified following a published bioinformatic pipeline for lncRNA discovery (Gaiti et al., 2015). Transcripts are classified as either protein-coding or non-protein coding using a series of computational tools to perform stepwise filtering, annotation and isolation of transcripts satisfying lncRNA criteria based on (i) homology to known proteins, (i) the presence of signal peptides, (iii) transcript length, (iv) open reading frame length, and (v) non-coding validation to retain the most likely non-coding candidate transcripts as predicted lncRNAs (details in Supplementary Material).

Mapping and identification of SNPs

We performed additional filtering of the reference assembly prior to population genomic analyses: Transcripts showing high sequence similarity were clustered using Cd-Hit-Est (Fu, Niu, Zhu, Wu, & Li, 2012; Li & Godzik, 2006) with a minimum sequence identity of 95% of the shortest sequence. We removed contigs without a significant blastn hit to the *M. galloprovincialis* draft genome (e-value 10^{-4}; Murgarella et al., 2016). We also removed transcripts with significant blastn matches (e-value 10^{-3}) to the *M. galloprovincialis* male (Genbank reference: FJ890850.1, AY363687.2) and female (Genbank reference: FJ890849.1, AY497292.2) protein-coding and complete mitochondrial genomes. The resulting 'total' assembly of 143,093 nuclear sequences was used as a reference for Single Nucleotide Polymorphism (SNP) discovery (for analyses of parallelism) among four introduced and three native-range *M. galloprovincialis* populations, *M. planulatus*, and *M. edulis* (Table 1). For comparisons of lncRNA contributions to population structure, we removed all remaining lncRNA loci from this reference assembly, leaving 106,333 transcripts considered as the 'full' transcriptome assembly. RNAseq reads were mapped to each reference assembly (i.e. total, full and lncRNA) using Bowtie2 (Langmead & Salzberg, 2012). PCR duplicates were removed using Picard MarkDuplicates (http://picard.sourceforage.net) and indexed BAM files were created with SAMtools. SNPs were called using Freebayes (https://github.com/ekg/freebayes), and variants were filtered with VCFtools (Danecek

et al., 2011). We removed sites below a genotype quality of 30 and a minimum mean depth coverage of 10 reads, singletons (minor allele count=2 across all populations) and indel variants, and all variants with missing genotype data. Genotypes were statistically phased using beagle v4.1 (Browning & Browning, 2007).

Analyses of population differentiation: Full transcriptome and lncRNAs

We visualised variation among individual genotypes using Principal Components Analysis (PCA) on SNPs derived from the full and lncRNA assemblies. Because we previously determined that introduced *M. galloprovincialis* in Sydney Harbour and Batemans Bay are introgressed with *M. planulatus*, we included *M. planulatus* samples from Tasmania as a reference parental population, but excluded *M. edulis*. The full assembly resulted in 16,906 biallelic SNPs and we identified 471 biallelic SNPs for the lncRNA dataset following filtering. We calculated pairwise *FST* (Weir & Cockerham, 1984) for each dataset using the R package *hierfstat* (Goudet, 2005) and tested for significance using the *boot.ppfst* function (10,000 bootstrap replicates). Given low differentiation between the eastern and western Mediterranean populations (*FST*=0), we combined these populations for subsequent analyses.

To test the hypothesis that lncRNAs have a role in adaptation following introduction, we examined whether putative lncRNAs contribute to elevated differentiation among native and introduced populations relative to the full transcriptome. Alternatively, if lncRNA transcripts experience evolutionary constraints, we expected that transcripts would contribute significantly less to population structure compared to other expressed transcripts. To assess whether *FST* values were significantly different between full and lncRNA datasets, we randomly resampled 471 variants (i.e. number equivalent to lncRNA SNPs) from the full dataset 1000 times and calculated pairwise *FST* for each iteration to obtain a null distribution of estimates using custom scripts in R. We assessed whether the observed lncRNA *FST* values fell within the 5^{th} or 95^{th} quantiles of the null distribution to determine whether the lncRNAs contributed significantly to elevated or reduced population differentiation, correcting p-values for multiple comparisons at qvalue≤0.05 ('*qvalue*' R package; Dabney, Storey, & Warnes, 2010). To explore diversity signatures that may indicate selection acting on lncRNAs, we examined the density of polymorphic sites (proportion of segregating sites per contig averaged across contigs) and contrasted

observed and expected heterozygosity and allelic richness (using the smallest population size for rarefaction) for each dataset in *hierfstat*. *PopGenome* (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) allowed us to calculate per locus within-population nucleotide diversity and Tajima's D values (Tajima, 1989) to provide an indirect indication of the number of rare variants. In all subsequent analyses of genetic parallelism (subsections below), we asked whether lncRNAs were among loci showing atypical genetic differentiation following introduction.

Patterns of genetic differentiation: Replicated introduced populations

To test the hypothesis that *M. galloprovincialis* introductions have undergone parallel differentiation relative to their native genetic backgrounds, we identified highly differentiated loci between introduced and native populations: (i) Batemans Bay and South Africa against Atlantic *M. galloprovincialis* and (ii) California and Sydney Harbour against Mediterranean *M. galloprovincialis*. We calculated pairwise *FST* for each locus (Nei, 1987) using the function *basic.stats()* in *hierfstat* and converted negative *FST* values to zero. To focus analyses on long-term adaptive variation within individual contigs, we considered only SNPs that were shared between populations and also yielded the maximum *FST* value (*maxFST*) for each contig. If more than one shared SNP had a *maxFST* value, we retained a single SNP per contig to limit pseudoreplication of closely linked variants (e.g., Fraïsse et al., 2016). We also repeated parallelism analyses using the *maxFST* value for each contig, irrespective of which SNP yielded the maximum value. Because *FST* measures are contingent on within-population diversity and are influenced by variation in minor allele frequency and sampled population sizes (Berner, 2019), *FST* may not always reflect true allele frequency differences and may have decreased sensitivity at low differentiation levels, such as those characteristic of *Mytilus* populations; we therefore repeated analyses using absolute allele frequency differences between two samples (*deltaP*=|p1–p2|) to avoid assumptions associated with estimating *maxFST*.

If introduced populations diverge from their native genetic background in a repeatable manner or share the same history of colonisation (i.e. through stepping-stone migrations), we expected to see positive correlations in genetic differentiation between introduced-native populations, such that the most highly differentiated loci would overlap in different introduced locations. We also examined whether introduced

populations originating from divergent genetic backgrounds in California (Mediterranean origin) and South Africa (Atlantic origin) displayed similar differentiation patterns from native reference populations. For all joint comparisons, we used standardised major axis regression to estimate the strength of *maxFST* correlation. We estimated the slope and intercept (elevation) of the regressions using the SMATR R package (Warton, Duursma, Falster, & Taskinen, 2012). We categorised contigs as being highly differentiated in pairwise comparisons if they contained a *maxFST* value within the top 1% of the empirical distribution, whereby contigs within the 99th percentile in both distributions were considered shared outlier contigs.

To explore whether our results are robust to sampling artefacts associated with small sample sizes and because some of our joint inferences involved pairwise comparisons of the same native population (e.g., Fig. 2a, Fig. 2b), we undertook a permutation-based test of significance to ascertain the effect of sampling error on the *maxFST* distribution and the influence of small samples size on outlier detection. We took the total number of observed alternative allele counts per locus for each joint comparison and randomly resampled alleles within each locus and across all three populations (i.e. one native population and two introduced populations) of the same sizes as our empirical samples and recalculated *FST* and *deltaP*. This randomisation procedure (1000 permutations) allowed us to obtain the null distribution of summary statistics while (i) holding sample sizes and the number of loci sampled constant at the realised values and (ii) preserving the relationships across all three populations (i.e. one native population and two introduced populations). We recorded locus-specific estimates and used these values to generate marginal cumulative probability distributions to ascertain the probability of obtaining extreme values for some loci in joint comparisons due to random sampling effects.

Patterns of Introgression: Replicated contact zones

To evaluate genetic parallelism between independent admixture events in Australia, we quantified locus-specific introgression asymmetries as positive or negative deviations from the average genomic admixture rate, following the approach from Simon, Arbiol, et al. (2020). This allowed us to identify genomic regions either putatively resistant to introgression (positive deviations indicating an excess of introduced *M. galloprovincialis* ancestry) or permeable to introgression by heterospecific native alleles

(negative deviations indicating *M. planulatus* ancestry). For each admixed population, we calculated the expected allele frequency of an allele (F_{exp}) as a function of the average population ancestry for the focal admixed population and its frequency in each parental reference population;

$$F_{exp} = (f_{\_Introduced} * Q_{\_Introduced}) + (f_{\_Resident} * Q_{\_Resident});$$

where f is the frequency of the most common allele in the reference native-range population for the introduced species (i.e. $f_{\_Introduced}$ in Atlantic or Mediterranean *M. galloprovincialis*) and its corresponding frequency in the resident species (i.e. $f_{\_Resident}$ in native *M. planulatus*); and Q is the average admixture proportion in the admixed population corresponding to each putative parental ancestry component. Given an expected allele frequency, we can calculate the locus-specific deviation from the average genomic admixture rate:

$$D = F_{obs} - F_{exp};$$

where F_{obs} is the observed allele frequency in the admixed population.

We estimated genome-wide correlations in allelic deviations between independent contact zones to evaluate whether genome-wide departures from the genomic mean were similar between Batemans Bay and Sydney Harbour despite involving divergent *M. galloprovincialis* lineages. If admixture events have led to repeated genetic outcomes, we expected to see positive correlations between allelic deviations in separate localities. The repeatability of admixture can also be assessed by analysing introgression patterns independently within each admixture event. Specifically, if the degree of divergence between incipient species is inversely correlated to introgression rate (Duranton et al., 2018), we predicted that highly differentiated loci between native *M. planulatus* and introduced *M. galloprovincialis* lineages would display strong asymmetries in locus-specific introgression (either positive or negative deviations) relative to the average genomic admixture rate. Under the hypothesis of barriers to introgression or other sources of selection favouring either introduced or locally adapted alleles, genome-wide deviations and parental *FST* are expected to be correlated, whereby the most divergent loci show an excess of *M. galloprovincialis* ancestry (positive correlations) or tendencies towards resident ancestry (negative correlations). Under neutral expectations (no barriers to

introgression) or if highly differentiated loci deviate equally toward introduced or native ancestries (opposing deviations), we would not expect to see correlations with parental *FST*. For each admixed population, we compared allelic deviations against parental *FST* values (between *M. planulatus* and introduced *M. galloprovincialis* lineages) estimated for each SNP locus (*snpFST*). To account for non-independence of loci on the same contig, we fit the data with linear mixed-effects models that included contig identity as a nested random effect using the maximum likelihood method and the lme() function in the R package nlme (Pinheiro, Bates, DebRoy, Sarkar, & Team, 2013). Both parental *FST* and a Boolean predictor reflecting Mediterranean vs. Atlantic *M. galloprovincialis* differentiation (see below) were included as predictors along with their interaction term.

For these analyses, population allele frequencies were calculated using VCFtools. We estimated individual ancestry proportions with the program ADMIXTURE (Alexander, Novembre, & Lange, 2009) to obtain average admixture proportions (*Q*) for each population. For ADMIXTURE analyses, we retained positions with up to 20% missing data across all populations, which resulted in 84,150 SNPs across 4941 contigs. We considered four putative ancestries corresponding to Atlantic and Mediterranean *M. galloprovincialis* and *M. planulatus* as a reference parental population, and *M. edulis* as an outgroup taxon. We ran ADMIXTURE with 100 iterations and used the cross-validation procedure with 50 replicates for K=4 genetic clusters. Because both Batemans Bay and Sydney Harbour showed >99% average ancestry belonging to *M. planulatus* and a single *M. galloprovincialis* lineage, we only considered two population ancestries for calculating expected allele frequencies.

It is well established that Atlantic *M. galloprovincialis* and *M. edulis* have a history of secondary contact, and that differential introgression with *M. edulis* contributes significantly to intraspecific outliers and partial reproductive isolation between Mediterranean and Atlantic *M. galloprovincialis* lineages (El Ayari et al., 2019; Fraïsse et al., 2016). Because divergent *M. edulis* genetic elements have been detected in Batemans Bay through secondary admixture with introduced Atlantic *M. galloprovincialis* (Popovic et al., 2020), we explored whether SNPs mapping to contigs strongly differentiating Mediterranean and Atlantic *M. galloprovincialis* lineages (contigs with max*FST* ≥ 0.5) also showed elevated asymmetries in allelic deviations as the most likely candidates for reproductive barriers or local adaptation loci, where

388 locus identity with respect to *M. edulis* introgression was treated as an additional variable in linear mixed-
389 effect models (Boolean indicator for contigs with maxFST < 0.5 or ≥ 0.5).
390
391 **Results**
392
393 *Predicted lncRNA transcripts*
394
395 We predicted 44,096 putative lncRNA transcripts, representing 23.8% of the unfiltered reference *de*
396 *novo M. galloprovincialis* transcriptome. This proportion is similar to high percentages reported for a
397 marine sponge (Gaiti et al., 2015) and plants (garden pea, Kerr, Gaiti, Beveridge, & Tanurdzic, 2017;
398 fern, Atallah, Vitek, Gaiti, Tanurdzic, & Banks, 2018) using the same computational pipeline, although
399 this value is greater than the proportion of lncRNAs (14.6%) identified in the *M. galloprovincialis* digestive
400 gland transcriptome (Gerdol et al., 2014). Despite stringent filtering, our final lncRNA assembly may
401 include untranslated regions, other polyadenylated RNA transcripts that satisfied minimum length
402 requirements (i.e. small regulatory RNAs, tRNAs and rRNAs) or novel *Mytilus* coding peptides with no
403 similarity matches to protein databases.
404
405 *Genetic differentiation: Full transcriptome and lncRNA datasets*
406
407 Principal component analysis of 16,906 biallelic variants from the full assembly revealed genetic
408 separation between introduced populations in California and South Africa (Fig. 1c), with the first and
409 second PC axes explaining 6.51% and 5.69% variance among individual genotypes. This analysis also
410 confirmed genetic structure previously identified between hybridising populations from Batemans Bay
411 and Sydney Harbour, which showed intermediate placement between northern *M. galloprovincialis*
412 populations and Australian *M. planulatus* (Fig. 1c; Popovic et al., 2020). ADMIXTURE analyses
413 indicated shared ancestry between introduced *M. galloprovincialis* in California and Mediterranean
414 populations; individuals sampled in South Africa clustered closely with Atlantic mussels, validating
415 previous inferences regarding the origins of these introductions (Fig. 1b). PCA of genotypic variance
416 confirmed that individuals sampled in California are not hybrids with *Mytilus trossulus*, where
417 introgressed individuals would have shown divergent placement away from the Mediterranean cluster.

All populations displayed similar levels of observed heterozygosity, suggesting no reductions in genetic diversity compared to native populations that would indicate large bottleneck effects (Table S1, Table S2).

lncRNA variant identification and filtering resulted in 200 transcripts with 471 SNPs. Lower expression among lncRNAs compared to coding transcripts (e.g., Quinn & Chang, 2016) may have resulted in the removal of many variants due to low sequence coverage. Overall, PCA of lncRNA variants explained a lower proportion of variance across the first (5.92%) and second (5.10%) PC axes compared to the full transcriptome. For this dataset, only *M. planulatus* sampled in Tasmania and the admixed population in Batemans Bay showed genetic separation from all other populations which did not reflect clustering based on sampling region (Fig. 1c). Pairwise comparisons of population differentiation for the full dataset indicated *FST* values significantly different than zero for all population pairs, with the lowest values for comparisons between the Mediterranean and California (*FST*=0.003; Table S3). Compared to the full dataset, lncRNA variants returned lower *FST* values for most populations pairs (Table S3). Significant reductions in lncRNA *FST*, however, were only evident for some populations displaying high population structure in the full dataset (e.g., populations from divergent genetic backgrounds and comparisons with Tasmania; Table S3).

Considering all populations, we observed a lower density of polymorphic sites within lncRNAs (0.00397) compared to the full dataset (0.00453), although mean densities were not significantly different (Welch approximation t-test, P=0.07). There were no consistent differences between datasets in within-population observed or expected heterozygosity and allelic richness that would suggest impacts of purifying selection on lncRNAs. Similarly, significant reductions in lncRNA polymorphism (*Pi*, within-population nucleotide diversity averaged across loci) were only evident for the Atlantic population (P≤0.01, Table S1). lncRNAs did not show consistently lower Tajima's D values or diversity that would suggest stronger selective constraints or the possible action of balancing selection (i.e., positive Tajima's D; Table S4).

448 *Repeatability of post-introduction genetic differentiation*

449

450 Filtered datasets of native-introduced population pairs resulted in 11,906 (Sydney Harbour) and 10,088
451 SNPs (California) for Mediterranean-derived populations, and 11,183 (Batemans Bay) and 8,902 SNPs
452 (South Africa) for Atlantic-derived populations. Consistent with recent timescale of presumed *M.*
453 *galloprovincialis* introductions into Australia (since the earliest records of European contact <600 years
454 ago to 100 years ago with the advent of commercial shipping), *maxFST* distributions of shared loci
455 between introduced and native populations were skewed towards low values, with few loci being highly
456 differentiated between populations. Overall, introduced populations originating from Atlantic *M.*
457 *galloprovincialis* were more differentiated from their native-range and showed greater genome-wide
458 variance compared to Mediterranean-derived introduced populations. *maxFST* for individual contigs
459 ranged between 0-0.6 (mean=0.0397, SD=0.0651, n=1634) for South Africa and 0-0.696
460 (mean=0.0529, SD=0.0753; n=1779) for Batemans Bay. Introduced populations in California and
461 Sydney Harbour were less differentiated from Mediterranean *M. galloprovincialis*, with *maxFST* ranging
462 from 0-0.372 (mean=0.0325, SD=0.045, n=1729) in California and 0-0.375 (mean=0.0319, SD=0.0410,
463 n=1873) for Sydney Harbour.

464

465 Joint *maxFST* distributions revealed an effect of interspecific gene flow with *M. planulatus* on levels of
466 differentiation between introduced and native populations. Admixed populations showed greater
467 average *maxFST* against native-range populations from the Atlantic (Batemans Bay) (n=651 loci;
468 standardised major axis regression, elevation=0.00721, P=0.00253, Fig. 2a) and Mediterranean
469 (Sydney Harbour) (n=737, standardised major axis regression, elevation=0.00265, P=0.0115, Fig. 2b),
470 compared to non-hybridising populations. Joint distributions were positively correlated in both
471 comparisons. The *maxFST* correlation between Atlantic-derived populations was moderate, but
472 significant (R^2=0.532, P<0.0001; Fig. 2a), with highly differentiated contigs (*FST*>0.4) showing similar
473 trends in both introduced populations (labeled in Fig. 2a). Three contigs were classified as shared
474 outliers falling above the 99th percentile of both empirical distributions and had joint *maxFST* values
475 greater than expected by chance (P<0.001) based on random permutation tests that were conditional
476 upon empirical samples sizes (Fig. S1). Correlations between Mediterranean-derived populations were
477 also significant (R^2=0.539, P<0.0001; Fig. 2b), with two contigs classified as shared outliers showing

differentiation levels above the 99th percentile in both introductions, although only one outlier was significant based on the simulated marginal P-value threshold (P<0.001) in both comparisons (Fig. 2b, Fig. S1). Joint distributions of *deltaP* estimates and *maxFST* per contig provided concordant results in all comparisons (Fig. S2, Fig. S3).

To investigate whether genetically divergent introductions may lead to similar genetic outcomes in the absence of hybridisation with native congeners, we tested for correlations in genetic differentiation between non-introgressed populations in California and South Africa against respective native reference populations (Fig. 2c). Joint *maxFST* comparisons resulted in fewer shared loci (n=318), a weak genome-wide correlation (standardised major axis regression, elevation=0.0065, P=0.0043; R^2=0.129, P<0.0001) and no shared outliers (Fig. 2c). In all comparisons, lncRNAs were poorly differentiated and none were differentiation outliers (Fig. 2). Incomplete and unordered scaffolds in the *M. galloprovincialis* draft genome assembly (Murgarella et al., 2016) precluded inferences regarding the genomic positions and the possibility of linkage between shared outlier contigs.

Repeatability of introgression patterns between contact zones

Mussels from Batemans Bay showed average ancestry proportions belonging to 35.2% Atlantic *M. galloprovincialis* and 64.1% *M. planulatus*, whereas Sydney Harbour were composed of 73.3% Mediterranean *M. galloprovincialis* and 26.7% *M. planulatus* (Fig. 1b). Joint comparisons of allelic deviations revealed heterogeneity in locus-specific introgression in both admixed populations, with some SNPs showing concordant positive or negative deviations; but there was no strong evidence for parallelism in genome-wide introgression patterns (R^2=0.0003, P=0.02) (Fig. 3).

In Batemans Bay, parental *FST* (between *M. planulatus* and Atlantic *M. galloprovincialis*) predicted locus-specific allelic deviations in the direction of an excess of *M. galloprovincialis* ancestry (+D: slope=0.128, $F_{1,7421}$=190.180, P<0.0001, Fig. 4a) with positive genome-wide deviations from the average admixture rate (intercept=0.0264, $F_{1,7421}$=798.096, P<0.0001). Loci with *maxFST* ≥ 0.5 between Mediterranean and Atlantic populations had significantly higher deviation values (intercept=0.0463, $F_{1,7421}$=4.002, P=0.0455; but no difference in slope, P=0.817) (Fig. 4a). Additionally, among three contigs

(Contig15136, Contig34278, Contig33253) showing parallel genetic responses between Atlantic-derived populations (Fig. 2a), two are highly differentiated between parental Atlantic *M. galloprovincialis* and *M. planulatus* and show negative deviations in Batemans Bay, pointing to possible selection against introduced alleles (Fig. 4a). Consistent *M. edulis* introgression contributing to intraspecific outliers between *M. galloprovincialis* lineages (Fraïsse et al., 2016), SNPs with *maxFST* ≥ 0.5 between Mediterranean and Atlantic populations were largely fixed within *M. edulis* and show high allele frequency differentials (*deltaP*>0.6) for the *M. edulis* major allele (Fig. S4).

In contrast, allelic deviations in Sydney Harbour showed no effect of parental *FST* (between *M. planulatus* and Mediterranean *M. galloprovincialis*) (slope=-0.0181, $F_{1,8678}$=3.059, P=0.0803), although the mean genome-wide deviation was positive (intercept=0.0149, $F_{1,8678}$=226.816; P<0.0001) (Fig. 4b). Loci diagnostic of Mediterranean and Atlantic lineages showed significant negative deviations from the average admixture rate (intercept=-0.00344, $F_{1,8678}$=4.76994, P=0.0290), but there was no difference in slope compared to all loci (P=0.928) (Fig. 4b).

In both admixed populations, the average deviation across SNPs localised to lncRNA transcripts was not significantly different than zero (Batemans Bay: n=15 SNPs, mean=0.332, P=0.148; Sydney Harbour, n=15 SNPs, mean=0.0124, P=0.432). Consistent with intraspecific comparisons (native-introduced *M. galloprovincialis*), lncRNA variants were generally poorly differentiated between *M. planulatus* and native-range *M. galloprovincialis* (Fig. 4).

Discussion

Replicated species introductions provide opportunities to investigate repeated patterns of adaptation and local introgression across contact zones. In the present study, we clarify aspects of *M. galloprovincialis* introduction history, validating evidence for an Atlantic origin for South African *M. galloprovincialis* and Mediterranean-derived introductions in California (Daguin & Borsa, 2000) that complements our earlier results uncovering Atlantic and Mediterranean *M. galloprovincialis* admixture Australian mussels (Popovic et al., 2020). Select outlier loci showed repeated changes in allele frequencies from native reference populations, but there was limited evidence for strong genome-wide

correlated shifts in allele frequencies between introduced populations. These results suggest that parallelism affects parts of the genome, with differentiation occurring predominantly within protein-coding regions, as lncRNA loci were minimally differentiated in pairwise comparisons. Comparing two independent hybrid populations, loci did not show parallel shifts in genome-wide introgression patterns from neutral expectations, implying different admixture dynamics when introduced lineages are genetically distinct. Interestingly, we observed that some loci with putatively *M. edulis*-derived alleles were especially resistant to introgression in Batemans Bay. This result is consistent with pre-introduction introgression shaping contemporary admixture whereby local interspecies introgression reinforces isolation between populations within species (Duranton et al., 2018; Foote et al., 2019; Fraïsse et al., 2016; Simon, Fraïsse, et al., 2020, this issue). Together, our findings suggest that both selective processes and the demographic histories experienced by introduced lineages can influence the repeatability of post-introduction evolution.

Parallel differentiation among introduced populations at outlier loci

Replicated *M. galloprovincialis* introductions from the same native-range source lineage exhibited significant, albeit modest, genome-wide correlations (Fig. 2a, 2b), whereas introductions originating from genetically distinct lineages had much weaker correlations (Fig. 2c). Coincident with genome-wide evidence for parallel divergence, a small number of contigs were repeatedly involved in high levels of differentiation in Atlantic-derived populations (South Africa and Batemans Bay). Three contigs provided especially strong evidence for genetic parallelism following introduction, with high differentiation at the same nucleotide positions (*maxFST*>0.4) when considering joint outcomes across both population pairs in permutation tests (P<0.001) (Fig. 2a). Suggestively, Contig33253 returned a top blastn hit to the *M. galloprovincialis* hsp90-2 gene encoding heat shock protein 90 (HSP90); HSPs and other oxidative stress-related proteins have been shown to evolve under positive selection in *M. galloprovincialis* (Popovic & Riginos, 2020) and may also be targets of repeated selection within introduced environments or during long-distance transport experienced by migrant larvae in the ballast water of container ships.

Determining the evolutionary mechanisms by which parallel divergence arises across *M. galloprovincialis* introductions is challenging, as a number of processes may converge towards similar

568 differentiation patterns (Riquet et al., 2019). Parallel evolution is often interpreted as evidence for similar
569 selection acting on the same genomic regions (Bolnick et al., 2018; Butlin et al., 2014), where standing
570 genetic variation is adaptively reused (Bock et al., 2015; Jones et al., 2012). Genetic repeatability is also
571 modulated by species-specific genetic architecture, either because complexes of adaptive alleles are
572 protected in inversions (Van Belleghem et al., 2018) or because the differential effects of long-term
573 linked-selection are conserved across similar recombination landscapes between species (e.g.,
574 *Ficedula* flycatchers, Burri, 2017; Burri et al., 2015; *Helianthus* sunflowers; Renaut, Owens, &
575 Rieseberg, 2014; *Mimulus* monkeyflowers, Stankowski et al., 2019; *Heliconius* butterflies, Martin et al.,
576 2013). In the context of invasive species, background selection is less efficient at the earliest stages of
577 differentiation (Burri, 2017) (i.e. between incipient native and introduced populations) and has limited
578 power to generate differentiation in the absence of strong barriers to gene flow (Stankowski et al., 2019).
579 Consistent with this premise, genomic islands within low-recombining regions (those experiencing linked
580 selection) also resist introgression between marine populations (e.g., European sea bass, Duranton et
581 al., 2018). Similarly, parallel divergence in the absence of ecological similarities, implicates an influence
582 of recombination on both lineage sorting and introgression rate among other hybridising lineages (e.g.,
583 Pacific cupped oyster, Gagnaire et al., 2018; long snouted seahorse, Riquet et al., 2019). Although
584 genomic resources for *M. galloprovincialis* are insufficient to make inferences about whether loci
585 diverging in parallel coincide with low-recombining regions, our findings suggest that similar selective
586 pressures or genomic constraints linked to recombination landscapes or reproductive isolation may lead
587 to repeatable differentiation in some parts of the genome. In addition, such forces can shape
588 differentiation both in the absence or presence of interspecific gene flow with native taxa (Fig 2).
589
590 Differentiation estimates may be confounded by high allele frequency variance due to small sample
591 sizes that can affect population genetic diversity and *FST* distributions (Bierne et al., 2013; Hoban et al.,
592 2016). While false positive outliers are unlikely to contribute to parallel differentiation by chance (Fraïsse
593 et al., 2016; Riquet et al., 2019), permutation analyses allowed us to exclude the effects of small sample
594 sizes as a substantial influence on the observed patterns, indicating highly significant ($P<0.001$) outliers,
595 even when joint comparisons involved the same native population (Fig. 2a, Fig. 2b). Nevertheless, allele
596 frequency shifts in native reference lineages may have contributed in part to correlated differentiation
597 observed between introduced populations. Similarly, low genetic structure between native and

introduced populations that is maintained by high standing genetic diversity and possibly ongoing propagule pressure, precluded our ability to rule out the possibility that shared colonisation histories (stepping-stone introduction routes) have led to parallelism at some loci. Indeed, difficulty in determining the native origins of introduced populations appears to be a ubiquitous issue in marine invasion studies and raises significant challenges for interpreting signals of differentiation associated with introductions (Geller, Darling, & Carlton, 2010; Riquet et al., 2019; Viard et al., 2016).

Despite largely positive correlations across *M. galloprovincialis* introductions, many loci showed location-specific shifts in allele frequencies, with little overlap in many differentiated contigs for both Atlantic and Mediterranean-derived joint comparisons (e.g., *maxFST*≥0.2; Fig. 2a, Fig. 2b). Introduced populations likely experience diverse selective pressures imposed by location-specific abiotic conditions, diverse community assemblages and the presence of semi-isolated native lineages (Guzinski et al., 2018; Riquet et al., 2013). In such cases, parallelism is expected to be reduced when different phenotypes are selected (Bolnick et al., 2018; Thompson, Osmond, & Schluter, 2019). Confidence in how location-specific selection shapes evolution in *M. galloprovincialis* introductions, however, will require additional empirical investigations.

Locus-specific introgression patterns differ between hybrid populations

Contrasting introgression patterns between hybrid populations in Australia revealed that gene flow can remodel population structure between introduced and native taxa in different ways depending on the introduced genetic background (i.e. Atlantic or Mediterranean source lineages). Specifically, joint comparisons of allele frequency deviations in Batemans Bay and Sydney Harbour indicated heterogeneity in locus-specific introgression rates, with no evidence for genome-wide correlations (Fig. 3), highlighting that introgression outcomes are not especially repeatable when divergent genetic lineages are involved. Locus-specific deviations were significantly correlated with parental *FST* in Batemans Bay (Fig. 4a), suggesting that the degree of divergence between hybridising species influences locus-specific introgression rates. Specifically, loci showed genome-wide deviations towards an excess of introduced *M. galloprovincialis* alleles introgressing into the *M. planulatus* native genetic background (Fig. 4a). In contrast, there was no evident relationship for Sydney Harbour mussels,

628 although loci had a slight genomic tendency in the direction of *M. galloprovincialis* ancestry (Fig. 4b).
629 This result is unexpected as Sydney Harbour is a large and busy harbour likely subject to high propagule
630 pressures of non-native larvae compared to the small docks and embayments of Batemans Bay, where
631 we would expect *M. planulatus* to outnumber *M. galloprovincialis* (matching average ancestry
632 proportions of 35.2% Atlantic *M. galloprovincialis*: 64.1% *M. planulatus*); thus, the predominant direction
633 of introgression should be of native alleles into the introduced species background (i.e. negative
634 deviation; Currat, Ruedi, Petit, & Excoffier, 2008).
635
636 Overall, our results support the hypothesis that differential introgression at some loci mirrors the
637 divergence between historically allopatric taxa (Duranton et al., 2018). High rates of gene flow and
638 recent divergence times typifying many marine hybrid zones may promote extensive introgression and
639 divergence swamping between semi-isolated lineages, such that the history of divergence is only
640 preserved at the most highly selected loci (Bierne et al., 2013; Duranton et al., 2018; Riquet et al., 2019).
641 In such cases, extrinsic selection may favour either invading or native alleles, depending on each hybrid
642 zone environment. Alternatively, selection may be intrinsic, maintaining divergence in regions
643 associated with reproductive incompatibilities. In Australian hybrid zones, a lack of concordance in
644 admixture patterns among divergent introduced lineages is consistent with recent empirical work on *M.*
645 *galloprovincialis* introductions along the Atlantic coast of Europe (Simon, Arbiol, et al., 2020), where
646 parallel patterns in allele frequency deviations were only observed when the same parental lineages
647 were compared, whereas correlations were negative or non-significant when admixed populations
648 comprised different *M. galloprovincialis* lineages. Thus, our results and those of Simon, Arbiol, et al.
649 (2020), reveal that genomic backgrounds can strongly influence the outcomes of post-introduction
650 admixture even among very closely related taxa.
651
652 *Heterospecific alleles may drive genetic parallelism between Atlantic M. galloprovincialis introductions*
653
654 From our analyses of paired introductions, we find that a small number of contigs were highly
655 differentiated in both Atlantic-derived introduced populations (South Africa and Batemans Bay; Fig. 2a).
656 Among three contigs (Contig15136, Contig34278, Contig33253) showing the strongest evidence for
657 genetic parallelism between Atlantic *M. galloprovincialis*-derived populations, two appear to resist

introgression between Atlantic and Mediterranean *M. galloprovincialis* lineages (max*FST* ≥ 0.5, contigs in green; Fig. 2a). Additionally, contigs with alleles putatively shared with *Mytilus edulis* (Fig. 5; Fig. S4) were highly differentiated between parental Atlantic *M. galloprovincialis* and *M. planulatus* and showed negative deviations in Batemans Bay, pointing to possible selection against introduced genotypes (Fig. 4a). In Europe, local introgression from *M. edulis* into Atlantic *M. galloprovincialis* contributes significantly to differentiation at intraspecific outliers between Atlantic and Mediterranean *M. galloprovincialis* (Fraïsse et al., 2016). *M. edulis*-derived alleles also appear to resist introgression between Mediterranean and Atlantic *M. galloprovincialis* lineages across a zone of secondary contact in Algeria (El Ayari et al., 2019) and can explain repeated patterns of intraspecific differentiation in some populations throughout the native range (Bierne, David, Boudry, & Bonhomme, 2002; Fraïsse et al., 2016; Gosset & Bierne, 2013).

The influence of *M. edulis* introgression in naturally occurring native-range hybrid zones of Atlantic and Mediterranean *M. galloprovincialis* (Fraïsse et al., 2016) is echoed in our results among introduced populations. In Batemans Bay, *M. edulis* alleles have undoubtedly accompanied introduced Atlantic *M. galloprovincialis* (Popovic et al., 2020); when we restricted analyses of allelic deviations to variants within contigs strongly differentiating *M. galloprovincialis* lineages (contigs with max*FST* ≥ 0.5, analogous to surveys using ancestry informative loci), parental *FST* had a significant effect on allelic deviations, with elevated mean departures from the genomic admixture rate (intercept=0.0463, P=0.0455) compared to analyses considering all loci (Fig. 4a). This pattern of asymmetric introgression illustrates an intriguing aspect of divergence between Atlantic *M. galloprovincialis* and *M. planulatus,* suggesting that highly divergent *M. edulis* alleles involved in reproductive isolation between *M. galloprovincialis* lineages likely contribute to high locus-specific variation in introgression rates and may resist introgression with *M. planulatus*.

Consistently, we observed nearly fixed differences in two highly differentiated contigs in Atlantic-derived comparisons that are also putative *M. edulis* introgression outliers (Fig. 2a, Fig. 5). Contig15136 (encoding ribosomal protein L7a) was almost fixed for thymine (T) at position 965 in *M. edulis* (T) and in the Atlantic (T, except for one individual), while *M. galloprovincialis* and *M. planulatus* were fixed for cytosine (Fig. 5a). A similar pattern was observed in Contig34278, although the SNP frequency shift at

688 site 872 is associated with one of *M. edulis'* alleles (Fig. 5b). The high *maxFST* values between Atlantic
689 and South Africa (Fig. 2a) also likely involve hybridisation but of a different nature: mixed Atlantic and
690 Mediterranean ancestry proportions in one individual sampled in Cape Columbine, South Africa (Fig. 1),
691 points to the possibility that introgressed genetic variation originating from the Mediterranean native
692 range is likely to inflate *maxFST*, and could also reflect Mediterranean alleles resisting introgression
693 with Atlantic *M. galloprovincialis* (that are heterogeneously introgressed by *M. edulis*), such that similar
694 outcomes are evident in the presence (Batemans Bay) and absence (South Africa) of hybridisation with
695 native congeners. For Sydney Harbor mussels, where partial *M. edulis* ancestry is not involved, parental
696 divergence did not affect the departures of introgressing loci from neutral expectations (Fig. 4b).
697
698 From the circumstantial evidence that we have outlined, it appears likely that historical gene flow from
699 *M. edulis* into Atlantic *M. galloprovincialis* (predating anthropogenic introductions) affects evolutionary
700 dynamics in introduced mussel species, both in the context of hybridisation with the Australian *M.*
701 *planulatus* and the previously undocumented input of Mediterranean *M. galloprovincialis* to the invasive
702 South African population (formerly thought to be exclusively Atlantic *M. galloprovincialis*-derived; Daguin
703 & Borsa, 2000). Similar to the case of European port mussels, pre-introduction introgression with *M.*
704 *edulis* may have contributed to instances of repeatable admixture outcomes between introduced *M.*
705 *galloprovincialis* and the local genetic background (Simon, Arbiol, et al., 2020). However, given the small
706 number of genotyped individuals in this study, we cannot make strong claims regarding introgression
707 rates at individual loci. Sampling variance and genetic drift may lead to spurious allele frequency
708 deviations, especially for loci where divergence between hybridising species is low. Likewise, genetic
709 drift at individual loci may obscure parallel trends if only a few generations of backcrossing have
710 occurred since the initial admixture event and there has been insufficient time for introgression to
711 proceed (e.g., Gagnaire et al., 2018). Further investigations through simulations are warranted to better
712 understand the dynamics of allelic deviation statistics under a spectrum of divergence scenarios and
713 drift processes that may be important during invasive secondary contacts.
714
715 Based on genome-wide patterns, however, our findings in Batemans Bay implicate partial reproductive
716 isolation between Atlantic *M. galloprovincialis* and *M. planulatus*, which may also explain why we
717 observed different outcomes when Mediterranean *M. galloprovincialis* was considered. This results

718 suggests that different sets of incompatibility genes may be important for *M. planulatus* and divergent
719 *M. galloprovincialis* lineages. While several reproductive isolating mechanisms are documented among
720 several pairs of *Mytilus* congeners (i.e. asynchronous spawning, Gilg, Kirby, Sullivan, Knapp, & Hilbish,
721 2007; differences in habitat preference, Gosling & McGrath, 1990, assortative fertilization, Bierne et al.,
722 2002; Dobzhansky-Muller incompatibility substitutions at several loci, Bierne, Bonhomme, Boudry,
723 Szulkin, & David, 2006), we currently have no information on the possible isolating mechanisms
724 (ecological or genomic) between *M. planulatus* and other *Mytilus* species.
725
726 *Genetic differentiation: Long non-coding RNAs*
727
728 Distinguishing among the major transcriptomic components of genetic variation underlying
729 differentiation allowed us to characterise how lncRNAs shape population structure among native and
730 introduced *Mytilus* populations (Fig 1c). LncRNA transcripts contributed significantly less to population
731 genetic structure compared to variants derived from the full transcriptome (Fig. 1c; Table S1) and
732 showed no evidence of high differentiation in analyses of parallelism (Fig. 2) or between congeners (Fig.
733 4). These findings suggest that lncRNA evolution may be constrained within lineages, with expressed
734 protein-coding transcripts being more likely to dominate signatures of recent genetic differentiation.
735 Purifying selection on lncRNAs has been documented among *Drosophila* populations, and is consistent
736 with suspected lncRNA functions in conserved developmental regulatory networks (Gaiti, Degnan, &
737 Tanurdžić, 2018; Perry & Ulitsky, 2016). However, measures of diversity among lncRNAs in *M.*
738 *galloprovincialis* indicated no consistent reductions that would support stronger purifying selection or
739 balancing selection on these transcripts.
740
741 Despite notable progress in understanding the role of lncRNAs in speciation, the evolutionary processes
742 influencing lncRNA transcripts are poorly understood, with mixed evidence supporting both sequence
743 conservation and rapid evolution among lineages (Gaiti et al., 2018). lncRNAs frequently evolve faster
744 than protein coding sequences, as the primary sequence is not constrained by protein function (Hezroni
745 et al., 2015; Pang et al., 2006; Quinn & Chang, 2016; Ulitsky, 2016); however, alternative modes of
746 evolution, such as adaptive changes to secondary and tertiary structures may occur through non-
747 nucleotide substitution mechanisms that cannot be detected by SNP-based measures of genomic

differentiation as employed in this study. We acknowledge that lncRNA identification using the total assembly subjected to additional filtering (prior to population genomic analyses) resulted in a reduced number of lncRNA transcripts and polymorphisms. Thus, genotyping errors due to mapping biases or assembly errors in the absence of an annotated reference genome may have decreased detectable population structure and our results should be interpreted with some caution. For *Mytilus* mussels, we also cannot exclude the possibility that non-selective processes have contributed to patterns of lncRNA evolution and diversity observed in this study, including variation in mutation rate or evolutionary constraints linked to genomic synteny that do not operate in a population-specific way, or at least not within timescales relevant for biological invasions.

Conclusions

Here, we provide comparative evidence for parallel genetic changes in multiple *M. galloprovincialis* populations in the context of anthropogenic introductions and recent secondary contacts. Our study shows that species introductions, both in the presence and absence of gene flow with native congeners, may lead to similar genetic changes, especially for protein-coding loci. The extent of parallel differentiation and admixture patterns between contact zones, however, varied depending on the colonising *M. galloprovincialis* lineage. We observed stronger introgression asymmetries in contigs distinguishing Mediterranean and Atlantic *M. galloprovincialis* lineages compared to genome-wide trends in Batemans Bay, an observation consistent with *M. edulis* alleles influencing introgression barriers between introduced lineages (South Africa) and in secondary contact with native congeners (Batemans Bay). Genomic studies of marine invasive species and especially anthropogenic hybrid zones are still rare (Viard et al., 2016; Viard et al., 2020). The lack of strong reproductive isolation among *M. galloprovincialis* and *M. planulatus* provides a contrast to secondary contact in *Ciona* tunicates (Bouchemousse, Liautard-Haag, Bierne, & Viard, 2016; Viard et al., 2020) and oysters (Gagnaire et al., 2018) where reproductive isolation is strong. Thus, in introduced-native *Mytilus* hybrid zones there is the opportunity to observe the earliest stages of extensive admixture that may lead to consolidation into islands of divergence or to complete genomic swamping. Overall, our findings build on a growing number of investigations (e.g., Duranton et al., 2020; Fraïsse et al., 2016; Simon, Arbiol, et al., 2020; Simon, Fraïsse, et al., 2020) reporting significant impacts of historical introgression on the intrinsic genetic

778 differences between marine lineages, which may also influence the repeatability of recent evolutionary
779 change and the genetic outcomes of human-mediated admixture events.
780
781 **Data Archive**
782
783 RNAseq data is available on the NCBI Sequence Read Archive (SRA) (BioProject ID: PRJNA560413).
784 Raw RNAseq data for newly sequenced populations are deposited to the NCBI SRA (BioProject ID:
785 PRJNA679634). Reference transcriptomes, lncRNA reference assemblies, and genomic datasets that
786 support the findings of this study are openly available in the Dryad digital repository at
787 https://doi.org/10.5061/dryad.xsj3tx9d9 and the Genomic Observatories MetaDatabase (https://geome-
788 db.org).
789
790
791

792 **Figure Captions**

793

794 **Fig. 1. (a)** Sampling locations for native (circle) and introduced (triangle) *M. galloprovincialis* populations
795 and *M. planulatus*. Approximated distributions of introduced populations in Australia, California and
796 South Africa are denoted by red (Mediterranean) and purple (Atlantic) indicating known native source
797 lineages, and the *M. planulatus* (*MP*) range in Australia is shown in yellow. **(b)** ADMIXTURE results for
798 K=4 genetic clusters, including *M. edulis* (*ME*). Each bar represents an individual belonging to one or
799 more ancestral clusters, corresponding to different colours. **(c)** Principal components analysis (PCA) of
800 *M. galloprovincialis* individuals sampled from geographic locations coloured by geographic location,
801 where *M. planulatus* samples are included as a reference population. Top panel shows a PCA of 16,906
802 variants derived from the full transcriptome assembly and the bottom panel shows a PCA across 471
803 SNPs derived from lncRNA transcripts.

804

805 **Fig. 2.** Joint *maxFST* distributions for pairs of introduced *M. galloprovincialis* populations against native
806 reference populations from the **(a)** Atlantic (elevation=0.00721, P=0.00253; R^2=0.532, P<0.0001); and
807 **(b)** the Mediterranean (elevation=0.00265, P=0.0115; R^2=0.539, P<0.0001). **(c)** Joint *maxFST*
808 distributions for introduced *M. galloprovincialis* populations from different native origins in South Africa
809 and California (R^2=0.129, P=<0.0001). For all plots, the grey dotted line corresponds to the 1:1 slope
810 and the solid line denotes the major axis regression. The red dotted lines show the 99th percentile for
811 both distributions, above which contigs were considered outliers. Solid red lines indicate the probability
812 (P<0.001) of observing outliers due to chance based on permutation tests. Green points correspond to
813 contigs strongly differentiating Atlantic and Mediterranean *M. galloprovincialis* lineages (max*FST*≥0.5)
814 as putative loci with introgressed *M. edulis* alleles. None of these diagnostic loci were present in
815 comparisons involving Mediterranean populations (Fig. 2b, 2c). Black open circles indicate lncRNA
816 transcripts.

817

818 **Fig. 3.** Joint comparison of allelic frequency deviations from the average genomic admixture rate for
819 hybrid populations in Australia. Mean genome-wide deviations are shown by the dotted lines for
820 Batemans Bay (mean=0.00613) and Sydney Harbour (mean=-0.00320). Loci showing positive
821 deviations from the genomic average indicate an excess of *M. galloprovincialis* ancestry from the

Atlantic (purple) and Mediterranean (red) whereas negative deviations indicate ancestry towards *M. planulatus* (yellow).

Fig. 4. Locus-specific allele frequency deviations for admixed populations as a function of parental divergence (*snpFST*) between *M. planulatus* (*MP*) and *M. galloprovincialis* (*MG*) lineages from the Atlantic and Mediterranean. **(a)** Significant effects of parental *FST* on allelic deviations in Batemans Bay (grey; slope=0.128, P<0.0001) suggest that differential introgression is influenced by the degree of parental divergence, with loci showing positive genome-wide deviations (intercept=0.0264, P<0.0001) towards an excess of *M. galloprovincialis* ancestry. SNP variants within contigs differentiating *M. galloprovincialis* lineages (green; contigs with max*FST* ≥ 0.5) had significantly higher deviation values (intercept=0.0463, P=0.0455; but no difference in slope, P=0.817). **(b)** There was no effect of parental *FST* on allelic deviations in Sydney Harbour mussels (P=0.0803). The mean deviation for all loci was positive (grey; intercept=0.0149, P<0.0001), whereas diagnostic loci showed a slight negative departure from the average admixture rate (green; intercept=-0.00344, P=0.0290).

Fig. 5. Frequencies of polymorphic sites along **(a)** Contig15136 and **(b)** Contig34278. Graphs show allele frequencies in reference to the *Mytilus edulis* major allele. Focal SNPs with *maxFST* values for each contig and where signals of parallelism may be attributed to local introgression of *M. edulis* alleles are indicated.

843 **Literature Cited**

844

845 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in
846 unrelated individuals. *Genome research, 19*(9), 1655-1664.

847 Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned
848 about the genetics of adaptation? *Trends in ecology & evolution, 23*(1), 26-32.

849 Atallah, N. M., Vitek, O., Gaiti, F., Tanurdzic, M., & Banks, J. A. (2018). Sex determination in *Ceratopteris*
850 *richardii* is accompanied by transcriptome changes that drive epigenetic reprogramming of the
851 young gametophyte. *G3: Genes, Genomes, Genetics, 8*(7), 2205-2214.

852 Baker, H. G. (1965). Characteristics and modes of origin of weeds. *Characteristics and modes of origin*
853 *of weeds.*, 147-172.

854 Baker, H. G. (1974). The evolution of weeds. *Annual review of ecology and systematics, 5*(1), 1-24.

855 Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations.
856 *Heredity, 57*(3), 357-376.

857 Bay, R. A., & Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression?
858 *Proceedings of the Royal Society B: Biological Sciences, 284*(1850), 20162414.

859 Bay, R. A., Taylor, E. B., & Schluter, D. (2019). Parallel introgression and selection on introduced alleles
860 in a native species. *Molecular Ecology, 28*(11), 2802-2813.

861 Bernardi, G., Azzurro, E., Golani, D., & Miller, M. R. (2016). Genomic signatures of rapid adaptive
862 evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. *Molecular*
863 *Ecology, 25*(14), 3384-3396.

864 Berner, D. (2019). Allele frequency difference AFD–an intuitive alternative to FST for quantifying genetic
865 population differentiation. *Genes, 10*(4), 308.

866 Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the
867 dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*.
868 *Proceedings of the Royal Society B: Biological Sciences, 273*(1591), 1253-1260.

869 Bierne, N., David, P., Boudry, P., & Bonhomme, F. (2002). Assortative fertilization and selection at larval
870 stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *evolution, 56*(2), 292-298.

871 Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? why are FST outliers
872 sometimes so frequent? *Molecular Ecology, 22*(8), 2061-2064.

873 Blackburn, T. M., Lockwood, J. L., & Cassey, P. (2015). The influence of numbers on invasion success.
874 *Molecular Ecology, 24*(9), 1942-1953.
875 Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., . . . Rieseberg, L. H.
876 (2015). What we still don't know about invasion genetics. *Molecular Ecology, 24*(9), 2277-2297.
877 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence
878 data. *Bioinformatics, 30*(15), 2114-2120.
879 Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) parallel evolution.
880 *Annual Review of Ecology, Evolution, and Systematics, 49*, 303-330.
881 Bouchemousse, S., Liautard-Haag, C., Bierne, N., & Viard, F. (2016). Distinguishing contemporary
882 hybridization from past introgression with postgenomic ancestry-informative SNPs in strongly
883 differentiated *Ciona* species. *Molecular Ecology, 25*(21), 5527-5542.
884 Branch, G. M., & Steffani, C. N. (2004). Can we predict the effects of alien species? A case-history of
885 the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental*
886 *Marine Biology and Ecology, 300*(1-2), 189-215.
887 Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data
888 inference for whole-genome association studies by use of localized haplotype clustering. *The*
889 *American Journal of Human Genetics, 81*(5), 1084-1097.
890 Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection.
891 *Evolution Letters, 1*(3), 118-131.
892 Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., . . . Garamszegi, L. Z. (2015).
893 Linked selection and recombination rate variation drive the evolution of the genomic landscape
894 of differentiation across the speciation continuum of Ficedula flycatchers. *Genome research,*
895 *25*(11), 1656-1665.
896 Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., . . . Hollander, J. (2014).
897 Parallel evolution of local adaptation and reproductive isolation in the face of gene flow.
898 *evolution, 68*(4), 935-949.
899 Carlton, J. T. (1996). Pattern, process, and prediction in marine invasion ecology. *Biological*
900 *conservation, 78*(1-2), 97-106.
901 Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015).
902 Biological invasions, climate change and genomics. *Evolutionary Applications, 8*(1), 23-46.

903 Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: evidence for differentiation,
904 natural selection, and local adaptation. *Molecular Ecology, 24*(9), 1999-2017.
905 Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., . . .
906 Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of
907 ectodysplasin alleles. *Science, 307*(5717), 1928-1933.
908 Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism
909 and convergence in natural populations. *Proceedings of the Royal Society B: Biological*
910 *Sciences, 279*(1749), 5039-5047.
911 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are
912 due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23*(13), 3133-3157.
913 Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive
914 introgression by local genes. *Evolution: International Journal of Organic Evolution, 62*(8), 1908-
915 1920.
916 Dabney, A., Storey, J. D., & Warnes, G. (2010). qvalue: Q-value estimation for false discovery rate
917 control. *R package version, 1*(0).
918 Daguin, C., & Borsa, P. (2000). Genetic relationships of *Mytilus galloprovincialis* Lamarck populations
919 worldwide: evidence from nuclear-DNA markers. *Geological Society, London, Special*
920 *Publications, 177*(1), 389-397.
921 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011).
922 The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.
923 de la Fuente, M., Valera, S., & Martínez-Guitarte, J. L. (2012). ncRNAs and thermoregulation: a view in
924 prokaryotes and eukaryotes. *Febs Letters, 586*(23), 4061-4069.
925 Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the
926 details: genetic variation in introduced populations and its contributions to invasion. *Molecular*
927 *Ecology, 24*(9), 2095-2111.
928 Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation,
929 adaptive evolution, and the role of multiple introductions. *Molecular Ecology, 17*(1), 431-449.
930 Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and
931 remolding of genomic islands of differentiation in the European sea bass. *Nature*
932 *communications, 9*(1), 1-11.

933 Duranton, M., Allal, F., Valière, S., Bouchez, O., Bonhomme, F., & Gagnaire, P. A. (2020). The
934 contribution of ancient admixture to reproductive isolation between European sea bass
935 lineages. *Evolution Letters, 4*(3), 226-242.
936 El Ayari, T., El Menif, N. T., Hamer, B., Cahill, A. E., & Bierne, N. (2019). The hidden side of a major
937 marine biogeographic boundary: A wide mosaic hybrid zone at the Atlantic–Mediterranean
938 divide reveals the complex interaction between natural and genetic barriers in mussels.
939 *Heredity, 122*(6), 770-784.
940 Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism
941 and convergence. *Trends in ecology & evolution, 26*(6), 298-306.
942 Ferrero, V., Barrett, S. C., Castro, S., Caldeirinha, P., Navarro, L., Loureiro, J., & Rodríguez-Echeverría,
943 S. (2015). Invasion genetics of the Bermuda buttercup (*Oxalis pescaprae*): complex
944 intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native
945 and introduced populations. *Molecular Ecology, 24*(9), 2143-2155.
946 Fitzpatrick, B. M., Johnson, J. R., Kump, D. K., Smith, J. J., Voss, S. R., & Shaffer, H. B. (2010). Rapid
947 spread of invasive genes into a threatened native species. *Proceedings of the National*
948 *Academy of Sciences, 107*(8), 3606-3610.
949 Foote, A. D., Martin, M. D., Louis, M., Pacheco, G., Robertson, K. M., Sinding, M. H. S., . . . Ballance,
950 L. (2019). Killer whale genomes reveal a complex history of recurrent admixture and vicariance.
951 *Molecular Ecology, 28*(14), 3427-3444.
952 Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main
953 cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology, 25*(1),
954 269-286.
955 Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation
956 sequencing data. *Bioinformatics, 28*(23), 3150-3152.
957 Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., . . . Lapègue, S.
958 (2018). Analysis of genome-wide differentiation between native and introduced populations of
959 the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biology and*
960 *Evolution, 10*(9), 2518-2534.
961 Gaiti, F., Calcino, A. D., Tanurdžić, M., & Degnan, B. M. (2017). Origin and evolution of the metazoan
962 non-coding regulatory genome. *Developmental biology, 427*(2), 193-202.

963 Gaiti, F., Degnan, B. M., & Tanurdžić, M. (2018). Long non-coding regulatory RNAs in sponges and
964 insights into the origin of animal multicellularity. *RNA biology, 15*(6), 696-702.
965 Gaiti, F., Fernandez-Valverde, S. L., Nakanishi, N., Calcino, A. D., Yanai, I., Tanurdzic, M., & Degnan,
966 B. M. (2015). Dynamic and widespread lncRNA expression in a sponge and the origin of animal
967 complexity. *Molecular biology and evolution, 32*(9), 2367-2382.
968 Geller, J. B. (1999). Decline of a native mussel masked by sibling species invasion. *Conservation*
969 *Biology, 13*(3), 661-664.
970 Geller, J. B., Darling, J. A., & Carlton, J. T. (2010). Genetic perspectives on marine biological invasions.
971 *Annual review of marine science, 2*, 367-393.
972 Gerdol, M., De Moro, G., Manfrin, C., Milandri, A., Riccardi, E., Beran, A., . . . Pallavicini, A. (2014). RNA
973 sequencing and de novo assembly of the digestive gland transcriptome in *Mytilus*
974 *galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum*. *BMC*
975 *research notes, 7*(1), 722.
976 Gilg, M. R., Kirby, S. E., Sullivan, R., Knapp, L. W., & Hilbish, T. J. (2007). Dispersal vs. retention:
977 correspondence of species-specific reproductive cycles and settlement periods in a blue mussel
978 hybrid zone. *Marine Ecology Progress Series, 351*, 151-161.
979 Gosling, E., & McGrath, D. (1990). Genetic variability in exposed-shore mussels, *Mytilus spp*., along an
980 environmental gradient. *Marine Biology, 104*(3), 413-418.
981 Gosset, C., & Bierne, N. (2013). Differential introgression from a sister species explains high FST outlier
982 loci within a mussel species. *Journal of Evolutionary Biology, 26*(1), 14-26.
983 Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular*
984 *Ecology Notes, 5*(1), 184-186.
985 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Zeng, Q. (2011).
986 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature*
987 *biotechnology, 29*(7), 644-652.
988 Grant, W. S., & Cherry, M. I. (1985). *Mytilus galloprovincialis* Lmk. in southern Africa. *Journal of*
989 *Experimental Marine Biology and Ecology, 90*(2), 179-191.
990 Guzinski, J., Ballenghien, M., Daguin-Thiébaut, C., Lévêque, L., & Viard, F. (2018). Population genomics
991 of the introduced and cultivated Pacific kelp *Undaria pinnatifida*: Marinas—not farms—drive

992 regional connectivity and establishment in natural rocky reefs. *Evolutionary Applications, 11*(9),
993 1582-1597.
994 Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression,
995 and the origin and structure of hybrid zones. *Molecular Ecology, 25*(11), 2454-2466.
996 Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of
997 long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species.
998 *Cell reports, 11*(7), 1110-1122.
999 Hilbish, T. J., Brannock, P. M., Jones, K. R., Smith, A. B., Bullock, B. N., & Wethey, D. S. (2010).
1000 Historical changes in the distributions of invasive and endemic marine invertebrates are contrary
1001 to global warming predictions: the effects of decadal climate oscillations. *Journal of*
1002 *Biogeography, 37*(3), 423-431.
1003 Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . . Whitlock, M. C.
1004 (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future
1005 directions. *The American Naturalist, 188*(4), 379-397.
1006 Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino
1007 acid mutation contributes to adaptive beach mouse color pattern. *Science, 313*(5783), 101-104.
1008 Huerta-Sánchez, E., Jin, X., Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., . . . Ni, P. (2014).
1009 Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature,*
1010 *512*(7513), 194-197.
1011 Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . White, S. (2012).
1012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484*(7392), 55-61.
1013 Kerr, S. C., Gaiti, F., Beveridge, C. A., & Tanurdzic, M. (2017). De novo transcriptome assembly reveals
1014 high transcriptional complexity in *Pisum sativum* axillary buds and shows rapid changes in
1015 expression of diurnally regulated genes. *Bmc Genomics, 18*(1), 221.
1016 Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics,*
1017 *173*(1), 419-434.
1018 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods,*
1019 *9*(4), 357.
1020 Lee, C. E. (2016). Evolutionary mechanisms of habitat invasions, using the copepod *Eurytemora affinis*
1021 as a model system. *Evolutionary Applications, 9*(1), 248-270.

1022 Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or
1023 nucleotide sequences. *Bioinformatics, 22*(13), 1658-1659.
1024 Lockwood, J. L., Cassey, P., & Blackburn, T. (2005). The role of propagule pressure in explaining
1025 species invasions. *Trends in ecology & evolution, 20*(5), 223-228.
1026 Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome
1027 assemblies. *Bioinformatics, 27*(21), 2957-2963.
1028 Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016).
1029 Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics,*
1030 *12*(2), e1005887.
1031 Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., . . . Jiggins,
1032 C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies.
1033 *Genome research, 23*(11), 1817-1828.
1034 Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current*
1035 *Opinion in Genetics & Development, 47*, 69-74.
1036 Mérot, C., Berdan, E. L., Babin, C., Normandeau, E., Wellenreuther, M., & Bernatchez, L. (2018).
1037 Intercontinental karyotype–environment parallelism supports a role for a chromosomal inversion
1038 in local adaptation in a seaweed fly. *Proceedings of the Royal Society B, 285*(1881), 20180519.
1039 Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., & Canchaya, C. (2016). A first insight into
1040 the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PloS one, 11*(3), e0151561.
1041 Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical
1042 predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal*
1043 *Society B: Biological Sciences, 367*(1587), 409-421.
1044 Narum, S. R., Gallardo, P., Correa, C., Matala, A., Hasselman, D., Sutherland, B. J., & Bernatchez, L.
1045 (2017). Genomic patterns of diversity and divergence of two introduced salmonid species in
1046 Patagonia, South America. *Evolutionary Applications, 10*(4), 402-416.
1047 Nei, M. (1987). *Molecular evolutionary genetics*: Columbia university press.
1048 Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in
1049 populations. *evolution*, 1-10.
1050 Oduor, A. M. (2013). Evolutionary responses of native plant species to invasive plants: a review. *New*
1051 *Phytologist, 200*(4), 986-992.

1052 Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of
1053 conservation does not mean lack of function. *Trends in genetics, 22*(1), 1-5.
1054 Perry, R. B.-T., & Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem
1055 cells. *Development, 143*(21), 3882-3894.
1056 Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient
1057 Swiss army knife for population genomic analyses in R. *Molecular biology and evolution, 31*(7),
1058 1929-1936.
1059 Phillips, B. L., Brown, G. P., Webb, J. K., & Shine, R. (2006). Invasion and the evolution of speed in
1060 toads. *Nature, 439*(7078), 803-803.
1061 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2013). nlme: Linear and nonlinear mixed
1062 effects models. *R package version, 3*(1), 111.
1063 Popovic, I., Matias, A. M. A., Bierne, N., & Riginos, C. (2020). Twin introductions by independent invader
1064 mussel lineages are both associated with recent admixture with a native congener in Australia.
1065 *Evolutionary Applications, 13*(3), 515-532.
1066 Popovic, I., & Riginos, C. (2020). Comparative genomics reveals divergent thermal selection in warm-
1067 and cold-tolerant marine mussels. *Molecular Ecology, 29*(3), 519-535.
1068 Prentis, P. J., Wilson, J. R., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution
1069 in invasive species. *Trends in plant science, 13*(6), 288-294.
1070 Quesada, H., Wenne, R., & Skibinski, D. O. (1995). Differential introgression of mitochondrial DNA
1071 across species boundaries within the marine mussel genus *Mytilus*. *Proceedings of the Royal*
1072 *Society of London. Series B: Biological Sciences, 262*(1363), 51-56.
1073 Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function.
1074 *Nature Reviews Genetics, 17*(1), 47.
1075 Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., . . . Westram, A. (2017).
1076 Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow.
1077 *Journal of Evolutionary Biology, 30*(8), 1450-1477.
1078 Rawson, P., Agrawal, V., & Hilbish, T. (1999). Hybridization between the blue mussels *Mytilus*
1079 *galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited
1080 introgression. *Marine Biology, 134*(1), 201-211.

1081 Renaut, S., Owens, G. L., & Rieseberg, L. H. (2014). Shared selective pressure and local genomic
1082 landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology,*
1083 *23*(2), 311-324.
1084 Rennison, D. J., Delmore, K. E., Samuk, K., Owens, G. L., & Miller, S. E. (2020). Shared patterns of
1085 genome-wide differentiation are more strongly predicted by geography than by ecology. *The*
1086 *American Naturalist, 195*(2), 192-200.
1087 Richardson, D. M. (2011). Invasion science: the roads travelled and the roads ahead. *Fifty years of*
1088 *invasion ecology: The legacy of Charles Elton*, 397-401.
1089 Riquet, F., Daguin-Thiébaut, C., Ballenghien, M., Bierne, N., & Viard, F. (2013). Contrasting patterns of
1090 genome-wide polymorphism in the native and invasive range of the marine mollusc *Crepidula*
1091 *fornicata*. *Molecular Ecology, 22*(4), 1003-1018.
1092 Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., . . . Briard, O. (2019).
1093 Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid
1094 zones in a complex of cryptic seahorse lineages. *evolution, 73*(4), 817-835.
1095 Rius, M., Turon, X., Bernardi, G., Volckaert, F. A., & Viard, F. (2015). Marine invasion genetics: from
1096 spatio-temporal patterns to evolutionary outcomes. *Biological Invasions, 17*(3), 869-885.
1097 Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F., & Lapegue, S. (2013). Population
1098 genomics shed light on the demographic and adaptive histories of European invasion in the
1099 Pacific oyster, *Crassostrea gigas*. *Evolutionary Applications, 6*(7), 1064-1078.
1100 Roman, J., & Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions.
1101 *Trends in ecology & evolution, 22*(9), 454-464.
1102 Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels
1103 (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology, 24*(18), 4723-4738.
1104 Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene
1105 flow and selection interact to promote adaptive divergence in regions of low recombination.
1106 *Molecular Ecology, 26*(17), 4378-4390.
1107 Simberloff, D. (2013). Biological invasions: much progress plus several controversies. *Contributions to*
1108 *Science*, 7-16.

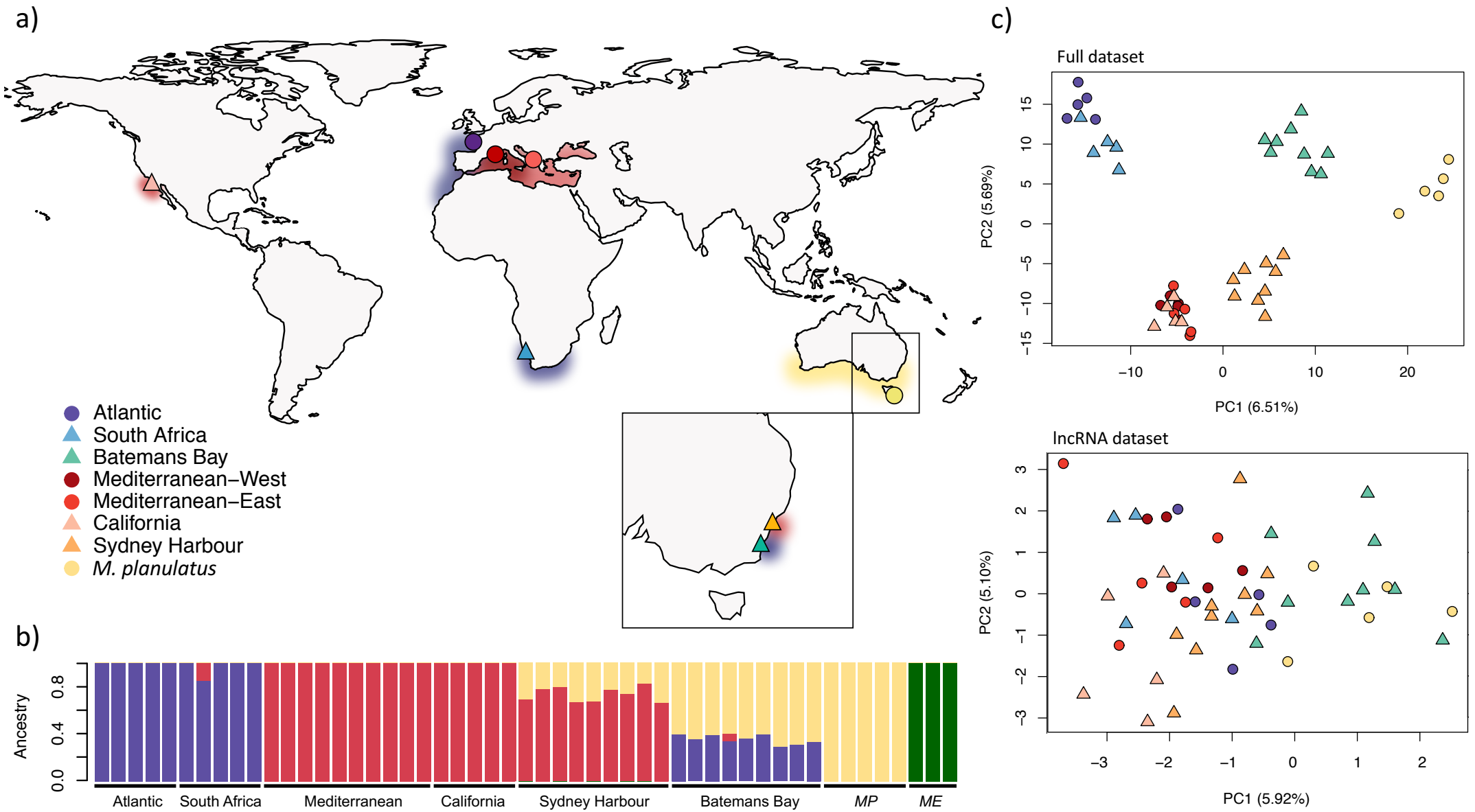
1109 Simon, A., Arbiol, C., Nielsen, E. E., Couteau, J., Sussarellu, R., Burgeot, T., . . . Robert, S. (2020).
1110 Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine
1111 mussels. *Evolutionary Applications, 13*(3), 575-599.
1112 Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2020).
1113 How do species barriers decay? Concordance and local introgression in mosaic hybrid zones
1114 of mussels. *Journal of Evolutionary Biology*.
1115 Somero, G. N. (2018). RNA thermosensors: how might animals exploit their regulatory potential?
1116 *Journal of Experimental Biology, 221*(4).
1117 Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling
1118 gene flow. *Nature Reviews Genetics, 14*(6), 404-414.
1119 Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019).
1120 Widespread selection and gene flow shape the genomic landscape during a radiation of
1121 monkeyflowers. *PLoS biology, 17*(7), e3000391.
1122 Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation
1123 may explain the genetic paradox of invasive species. *Molecular Ecology, 24*(9), 2241-2252.
1124 Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics, 14*(11), 751-
1125 764.
1126 Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science, 323*(5915), 746-751.
1127 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
1128 *Genetics, 123*(3), 585-595.
1129 Tepolt, C. K. (2015). Adaptation in marine invasion: a genetic perspective. *Biological Invasions, 17*(3),
1130 887-903.
1131 Tepolt, C. K., & Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive
1132 genome dynamics in a marine invader. *Molecular Ecology, 24*(16), 4145-4158.
1133 Thompson, K. A., Osmond, M. M., & Schluter, D. (2019). Parallel genetic evolution and speciation from
1134 standing variation. *Evolution Letters, 3*(2), 129-141.
1135 Turner, K. G., Hufbauer, R. A., & Rieseberg, L. H. (2014). Rapid evolution of an invasive weed. *New*
1136 *Phytologist, 202*(1), 309-321.

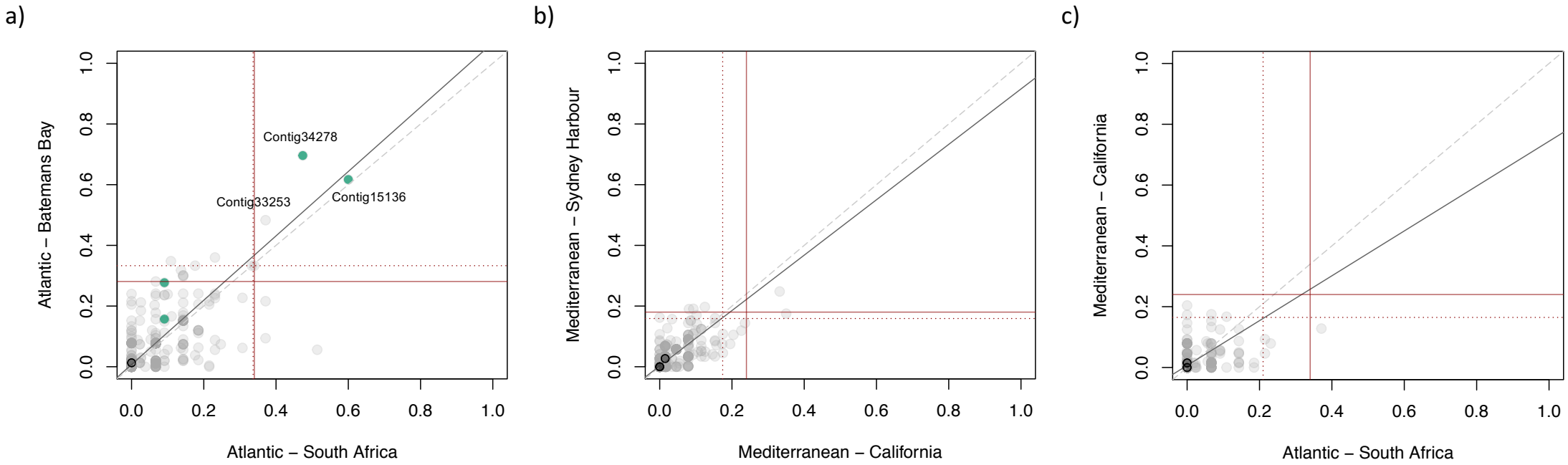
1137 Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population
1138 resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature genetics,*
1139 *42*(3), 260-263.
1140 Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding
1141 RNAs. *Nature Reviews Genetics, 17*(10), 601-614.
1142 Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., . . . Hendrickx, F.
1143 (2018). Evolution at two time frames: Polymorphisms from an ancient singular divergence event
1144 fuel contemporary parallel evolution. *PLoS Genetics, 14*(11), e1007796.
1145 Vandepitte, K., De Meyer, T., Helsen, K., Van Acker, K., Roldán-Ruiz, I., Mergeay, J., & Honnay, O.
1146 (2014). Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular*
1147 *Ecology, 23*(9), 2157-2164.
1148 Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: three lessons from
1149 the past, and the way forward. *Current zoology, 62*(6), 629-642.
1150 Viard, F., Riginos, C., & Bierne, N. (2020). Anthropogenic hybridization at sea: three evolutionary
1151 questions relevant to invasive species. *arXiv preprint arXiv:2006.05125*.
1152 Warton, D. I., Duursma, R. A., Falster, D. S., & Taskinen, S. (2012). smatr 3-an R package for estimation
1153 and inference about allometric lines. *Methods in Ecology and Evolution, 3*(2), 257-259.
1154 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure.
1155 *evolution*, 1358-1370.
1156

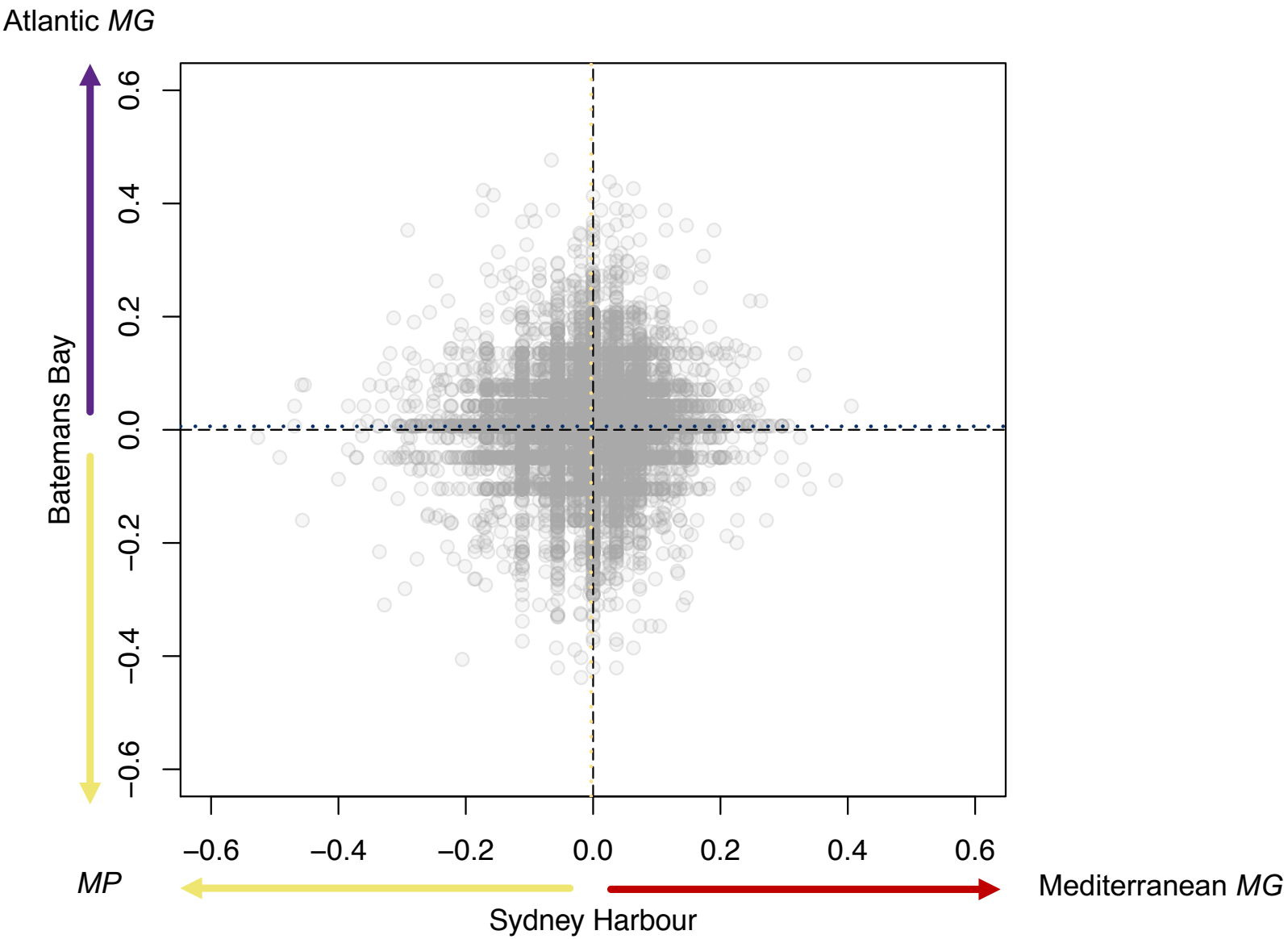
Table 1. Summary of sampling locations.

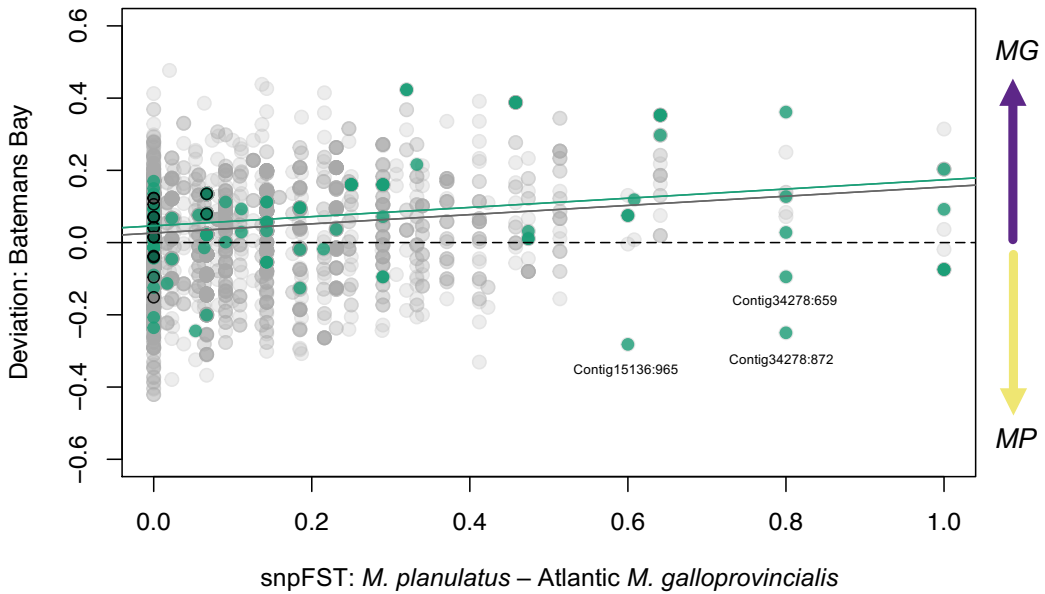
Taxon	Sampling Location	Range	Samples sequenced (RNAseq)
Mytilus edulis	Darling Marine Station, Maine, USA	Native	3
Mytilus galloprovincialis	Primel, France (Atlantic)	Native	5
	Crique des Issambles, France (Western Mediterranean)	Native	5
	Herceg Novi, Montenegro (Eastern Mediterranean)	Native	5
	Scripps Institute of Oceanography, California, USA	Introduced	5
	Cape Columbine, South Africa	Introduced	3
	Yzerfontein, South Africa	Introduced	2
	Sydney Harbour, New South Wales, Australia*	Introduced	9
	Batemans Bay, New South Wales, Australia*	Introduced	9
Mytilus planulatus	Spring Bay, Tasmania, Australia*	Native	5

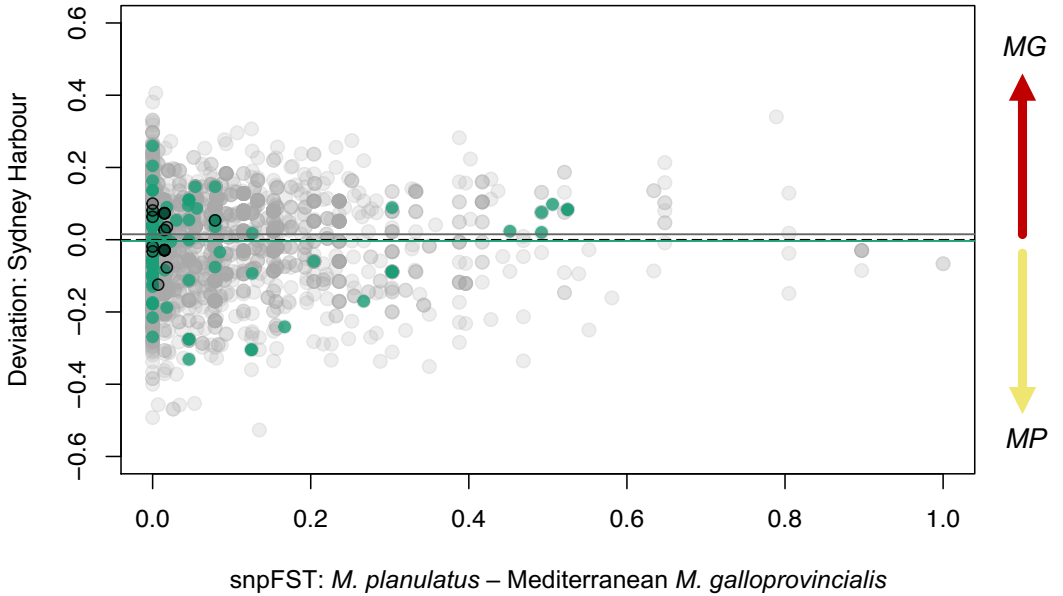
* Species identity and range for *Mytilus* samples is based on genome-wide analyses of species relationships (Popovic et al., 2020).

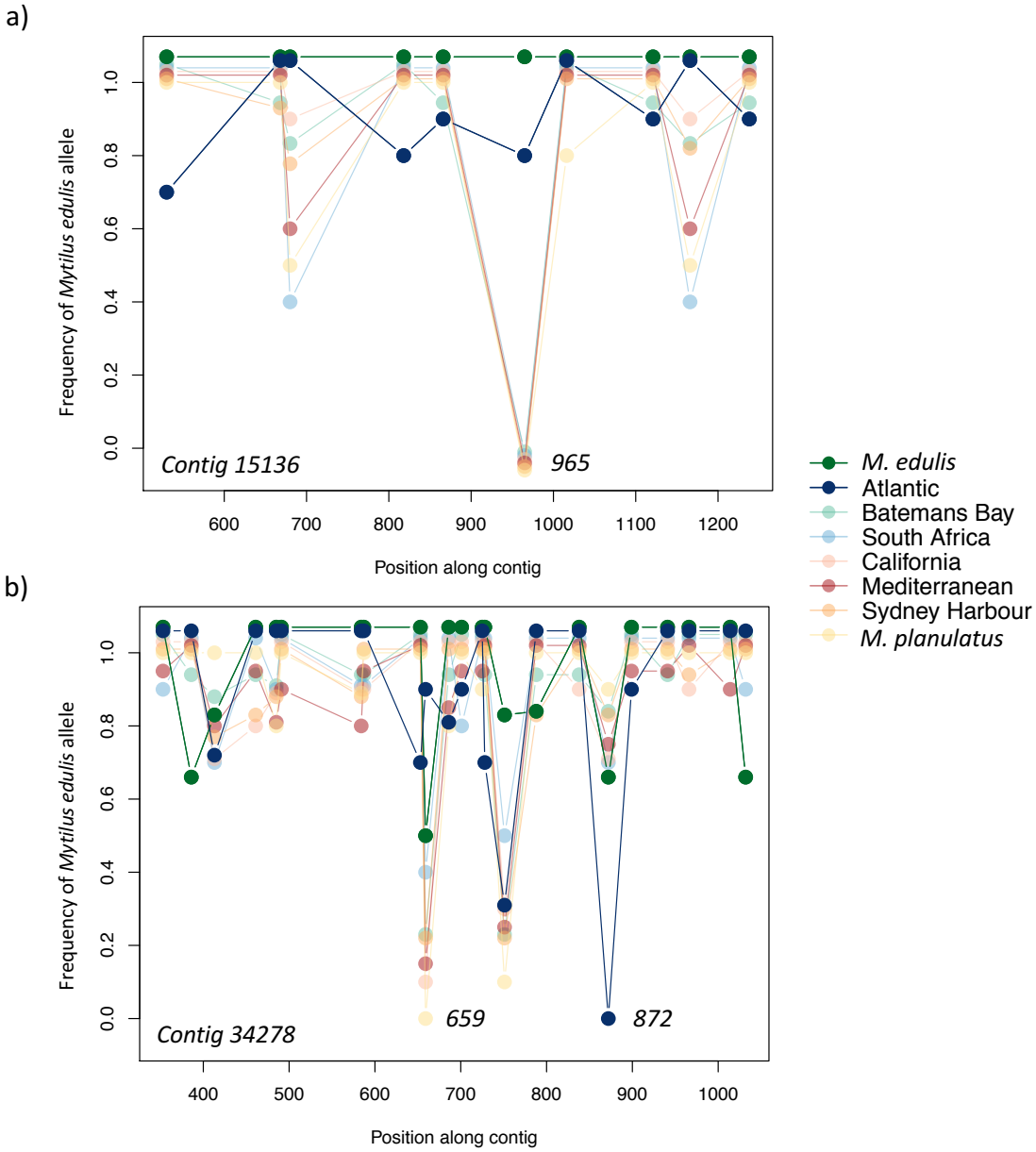

a)
Atlantic
South Africa
Batemans Bay
Mediterranean–West
Mediterranean–East
California
Sydney Harbour
M. planulatus
b)
Ancestry
0.0
0.4
0.8
Atlantic
South Africa
Mediterranean
California
Sydney Harbour
Batemans Bay
MP
ME
c)
Full dataset
PC2 (5.69%)
PC1 (6.51%)
lncRNA dataset
PC2 (5.10%)
PC1 (5.92%)


a)
b)
c)
Atlantic – Batemans Bay
Atlantic – South Africa
Contig34278
Contig33253
Contig15136
Mediterranean – Sydney Harbour
Mediterranean – California
Mediterranean – California
Atlantic – South Africa
0.0
0.2
0.4
0.6
0.8
1.0


Atlantic *MG*
Batemans Bay
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
MP
Mediterranean *MG*
Sydney Harbour


Deviation: Batemans Bay
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.0
0.2
0.4
0.6
0.8
1.0
Contig34278:659
Contig15136:965
Contig34278:872
MG
MP
snpFST: M. planulatus – Atlantic M. galloprovincialis


Deviation: Sydney Harbour
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.0
0.2
0.4
0.6
0.8
1.0
MG
MP
snpFST: M. planulatus – Mediterranean M. galloprovincialis


a)
Frequency of Mytilus edulis allele
0.0
0.2
0.4
0.6
0.8
1.0
Contig 15136
965
600
700
800
900
1000
1100
1200
Position along contig
b)
Frequency of Mytilus edulis allele
0.0
0.2
0.4
0.6
0.8
1.0
Contig 34278
659
872
400
500
600
700
800
900
1000
Position along contig
M. edulis
Atlantic
Batemans Bay
South Africa
California
Mediterranean
Sydney Harbour
M. planulatus