

## Article

# Studying the genetic basis of speciation in high gene flow marine invertebrates

Grant H. POGSON\*

Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA

\*Address correspondence to Grant H. Pogson. E-mail: pogson@ucsc.edu.

Received on 19 February 2016; accepted on 16 August 2016

## Abstract

A growing number of genes responsible for reproductive incompatibilities between species (barrier loci) exhibit the signals of positive selection. However, the possibility that genes experiencing positive selection diverge early in speciation and commonly cause reproductive incompatibilities has not been systematically investigated on a genome-wide scale. Here, I outline a research program for studying the genetic basis of speciation in broadcast spawning marine invertebrates that uses a priori genome-wide information on a large, unbiased sample of genes tested for positive selection. A targeted sequence capture approach is proposed that scores single-nucleotide polymorphisms (SNPs) in widely separated species populations at an early stage of allopatric divergence. The targeted capture of both coding and non-coding sequences enables SNPs to be characterized at known locations across the genome and at genes with known selective or neutral histories. The neutral coding and non-coding SNPs provide robust background distributions for identifying  $F_{ST}$ -outliers within genes that can, in principle, identify specific mutations experiencing diversifying selection. If natural hybridization occurs between species, the neutral coding and non-coding SNPs can provide a neutral admixture model for genomic clines analyses aimed at finding genes exhibiting strong blocks to introgression. Strongylocentrotid sea urchins are used as a model system to outline the approach but it can be used for any group that has a complete reference genome available.

**Key words:** barrier loci, Dobzhansky–Muller incompatibilities,  $F_{ST}$ -outliers, genome scan, introgression, positive selection, sequence capture, speciation.

## Introduction

Understanding how reproductive barriers evolve between populations represents one of the most fundamental challenges in evolutionary biology (Coyne and Orr 2004; Gavrilets 2004). Recent studies of genome-wide patterns of divergence between closely related species have provided new insights into the identity of genes causing reproductive barriers (Nadeau et al. 2012; Nolte et al. 2013), the semi-permeable nature of species barriers (Rieseberg et al. 1999; The *Heliconius* Genome Consortium 2012), and the genetic architecture of species differences (Teeter et al. 2010; Arnegard et al. 2014). Following the discovery of “genomic islands of speciation” in *Anopheles* mosquitoes by Turner et al. (2005), heterogeneous patterns of genome-wide divergence have been described

in growing number of species pairs (reviewed by Sousa and Hey 2013; Seehausen et al. 2014). This heterogeneity was initially interpreted as reflecting highly variable levels of introgression across the genome (Via and West 2008; Nosil et al. 2009). However, the conditions producing genomic islands of divergence appear restrictive (Feder et al. 2012b; Flaxman et al. 2013) and recent re-analyses suggest that they might result from recent diversifying or background selection occurring during allopatric divergence (Cruickshank and Hahn 2014).

Many population genomic studies have embraced the concept of “ecological speciation” where diversifying selection represents the initial driver of speciation with, or without, ongoing gene flow (Schluter 2009; Nosil 2012). An important feature of ecological

speciation is the reduced fitness of intermediate phenotypes (or migrants) caused by a mismatch between genotype and the environment (i.e., extrinsic postzygotic isolation). The enthusiasm for ecological speciation has shifted attention away from investigating how species originate in allopatry without any marked changes in ecology, physiology, or morphology. The widespread presence of cryptic species in both terrestrial and marine environments (Knowlton 1993; Pfenninger and Schwenk 2007) shows that reproductive barriers commonly evolve between taxa that remain unrecognized until genetic identification. The evolutionary processes and genes responsible for establishing and maintaining ecologically similar species in communities remain largely unknown. In the absence of strong ecological divergence, it is possible that some intrinsic isolation evolves from neutral or non-adaptive processes such as the random loss or movement of duplicated genes (Lynch and Force 2000) or intragenomic conflict (Rice 1998; Crespi and Nosil 2013).

Progress toward understanding the genetic basis of speciation requires identifying the genes responsible for reproductive barriers between species, the selective (or neutral) processes responsible for their divergence and their order of appearance (particularly during the initial stage) (Coyne and Orr 2004; The Marie Curie SPECIATION Network 2012). A growing list of “speciation genes” or “barrier loci” has been discovered that cause reproductive incompatibilities between species (Noor and Feder 2006; Presgraves 2010; Rieseberg and Blackman 2010; Nosil and Schluter 2011). An interesting pattern emerging from the characterization of barrier loci is that they often exhibit signals of positive Darwinian selection driven by genomic conflict (e.g., Ting et al. 1998; Phadnis and Orr 2009; Tang and Presgraves 2009; Nolte et al. 2013). Irrespective of the form(s) of selection involved, the possibility that barrier genes commonly have histories of positive selection has not been systematically investigated on a genome-wide scale. Theory predicts that Dobzhansky–Muller incompatibilities (DMIs; Dobzhansky 1937; Muller 1940) should be more prevalent at genes experiencing accelerated rates of divergence (Navarro and Barton 2003; Coyne and Orr 2004; Payseur and Nachman 2005). However, it is possible that many positively selected genes (PSGs) do not contribute to reproductive barriers and some may even experience adaptive introgression (e.g., Ding et al. 2014; Huerta-Sánchez et al. 2014). A powerful approach for distinguishing between PSGs conferring strictly adaptive benefits and those causing reproductive incompatibilities has been to study their patterns of introgression in hybrid zones (Harrison 1990; Payseur 2010).

In broadcast spawning marine invertebrates, the possibility that positive selection at gamete recognition proteins (GRPs) plays an important role in establishing reproductive barriers between species has received considerable attention (see reviews by Swanson and Vacquier 2002; Clark et al. 2006; Palumbi 2009; Lessios 2011; Vacquier and Swanson 2011; Kosman and Levitan 2014). However, the forms of selection responsible for positive selection on GRPs remain largely unknown and their levels of polymorphism and patterns of divergence vary unpredictably among groups (Vacquier and Swanson 2011; Kosman and Levitan 2014). The importance of divergence at GRPs during the early stage of speciation is also not clear because GRPs in allopatric species often fail to exhibit positive selection whereas those in sympatric taxa commonly do (Palumbi 2009; Vacquier and Swanson 2011; however, see Hart et al. 2014). Furthermore, strong asymmetrical patterns of gamete compatibility are common between closely related species (Zigler et al. 2005) that are incapable of providing effective barriers to introgression (Lessios 2007). For the sea urchin sperm protein bindin, gamete

compatibility between species is correlated with amino acid divergence and Zigler et al. (2005) estimate that sister species can maintain compatibility for up to 5 million years. This is comparable to the mean time to speciation in many marine invertebrate groups (Coyne and Orr 2004), suggesting that more effective barriers must evolve earlier at other unknown genes or genetic elements.

The goal of this paper is to outline a research strategy for studying the genetic basis of speciation in broadcast spawning marine invertebrates that is based on a priori genome-wide knowledge of genes experiencing positive Darwinian selection. Although candidate gene approaches have been used to investigate genome-wide patterns of differentiation (e.g., Andrés et al. 2013; Hebert et al. 2013; Fraïsse et al. 2016), none has been based on a large, unbiased sample of genes with known selective (or neutral) histories. The advantage of this approach is that it provides testable predictions about the identities of genes that might diverge early in the speciation process and be more effectively blocked from introgressing upon secondary contact. This avoids the post hoc characterization of “genomic islands of divergence” that often fail to identify the presumed barrier loci due to unknown linkage associations with the selected gene(s). Strongylocentrotid sea urchins will be used as a representative group to highlight the approach, but it can be applied to any system provided a well-annotated reference genome is available.

## The Geography of Speciation in High Gene Flow Marine Invertebrates

The importance of geography in speciation has recently been de-emphasized and replaced with models based on a continuum of gene flow (e.g., Butlin et al. 2008; Fitzpatrick et al. 2009; Harrison 2012). This shift in perspective is necessary for nearshore marine invertebrates whose geographic ranges have undergone dramatic and dynamic changes over the Pleistocene caused by glacial cycles (Jansson and Dynesius 2002; Jacobs et al. 2004; Norris and Hull 2012). In the northeast Pacific region, the fossil record shows that nearshore species respond individually to changing ocean temperatures and sea levels, and that community composition is in a state of perpetual flux (reviewed by Valentine and Jablonski 1993; Lindberg and Lipps 1996). The dynamic and complex histories of species past distributions make it difficult to assign a strict geographic mode of speciation, to infer the selective agents causing adaptive divergence, or to assess the role played by reinforcement. The rapidity of environmental change over the latter part of the Pleistocene has apparently not resulted in increased rates of speciation or extinction (Jackson and Johnson 2000).

The high levels of gene flow that occur among populations of many marine invertebrates with long-lived pelagic larvae simplify studying some aspects of their speciation while complicating others. The weak or non-existent population structure commonly observed in these species renders models of “speciation-with-gene-flow” (Feder et al. 2012a; Via 2012) unlikely because extensive gene flow overpowers diversifying selection across most of the genome. If selection does overcome gene flow and cause local adaptation (Sanford and Kelly 2011) it is unclear whether this eventually results in the formation of isolating barriers. The initiation of speciation in high gene flow marine taxa must involve a period of allopatric divergence following long-distance colonization across or between ocean basins (Lindberg 1991; Vermeij 1991) or by major vicariance events such as the rising of the Isthmus of Panama (Lessios 2008). Under these scenarios, complications arise when attempting to identify the

**Table 1.** Complete genomes of marine invertebrate species with available gene models<sup>a</sup>

Phylum	Species name	Size (Mb)	GC (%)	Number of genes	Reference(s)
Placozoa	<i>Trichoplax adhaerens</i>	105.6	32.7	11,518	Srivastava et al. (2008)
Porifera	<i>Amphimedon queenslandica</i>	166.7	37.5	13,998	Srivastava et al. (2010)
Mollusca	<i>Crassostrea gigas</i>	557.7	35.3	32,261	Zhang et al. (2012)
Echinodermata	<i>Strongylocentrotus purpuratus</i>	814.0	36.9	21,092	Sodergren et al. (2006), Tu et al. (2012)
Cephalochordata	<i>Branchiostoma floridae</i>	521.9	41.2	28,627	Putnam et al. (2008)
Tunicata	<i>Oikopleura dioica</i>	70.5	39.8	17,212	Seo et al. (2001)
Tunicata	<i>Ciona intestinalis</i>	115.2	33.1	15,254	Dehal et al. (2002)
Brachiopoda	<i>Lingula anatina</i>	425.5	36.9	34,000	Luo et al. (2015)
Anthozoa	<i>Nematostella vectensis</i>	356.6	40.6	27,173	Putnam et al. (2007)
Cephalopoda	<i>Octopus bimaculoides</i>	2,282.8	36.1	32,819	Albertin et al. (2015)
Ctenophora	<i>Mnemiopsis leidyi</i>	155.9	38.9	16,548	Ryan et al. (2013)
Annelida	<i>Capitella teleta</i>	333.3	40.4	31,997	Simakov et al. (2013)
Mollusca	<i>Lottia gigantea</i>	359.5	33.3	23,827	Simakov et al. (2013)
Mollusca	<i>Aplysia californica</i>	927.3	42.0	21,312	<a href="https://www.broadinstitute.org/scientific-community/science/projects/mammals-models/vertebrates-invertebrates/aplysia/aplysia-genom">https://www.broadinstitute.org/scientific-community/science/projects/mammals-models/vertebrates-invertebrates/aplysia/aplysia-genom</a>
Priapulida	<i>Priapulus caudatus</i>	511.7	45.7	17,096	<a href="http://www.ncbi.nlm.nih.gov/genome/?term=Priapulus+caudatus">http://www.ncbi.nlm.nih.gov/genome/?term=Priapulus+caudatus</a>
Hemichordata	<i>Saccoglossus kowalevskii</i>	775.8	38.1	22,073	<a href="https://www.hgsc.bcm.edu/acorn-worm-genome-project">https://www.hgsc.bcm.edu/acorn-worm-genome-project</a>
Arthropoda	<i>Limulus polyphemus</i>	1,828.3	34.5	22,031	<a href="http://genome.wustl.edu/genomes/detail/limulus-polyphemus/">http://genome.wustl.edu/genomes/detail/limulus-polyphemus/</a>

<sup>a</sup>Compiled from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) in December 2015.

genes diverging early in the speciation continuum. Here, the challenge is to identify recently formed allopatric populations separated by large geographic distances (i.e., different ocean basins) that have undergone a small amount of genetic divergence. Although young allopatric populations may not eventually form new species, they do provide snapshots of the initial adaptive divergence following long-distance dispersal events, which is the starting point for how many marine species originate.

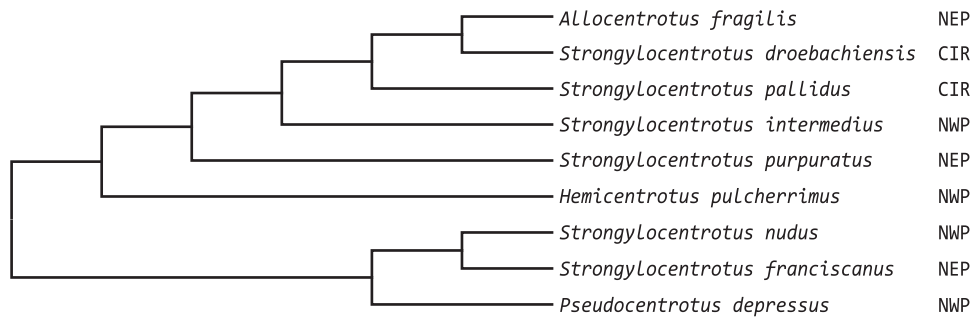
## The Targeted PSGs Strategy

Here, I propose a research strategy for testing the hypotheses that PSGs (i) diverge early among allopatric populations of a species and (ii) commonly evolve to become barrier loci. This strategy is based on knowing a priori the identity of genes experiencing positive selection across the genome in a group of closely related species. There are 4 requirements for this approach to be successful. First, a well-annotated reference genome must be available for what I call a “focal species”. This is clearly a stringent requirement that excludes virtually all species. However, the advantages provided by a reference genome for studying speciation are substantial (see below) and complete annotated genomes are now available for a growing number of marine invertebrate species (see Table 1). Second, a group of closely related species must be present whose genomes can be sequenced using the focal species’ genome as a reference. The generation of multi-species alignments of single-copy orthologs enables powerful and robust genome-wide tests for positive selection using programs like PAML (Yang 2007) or HyPhy (Kosakovsky Pond et al. 2005). A sufficient number of related taxa must be present (> 6) at an appropriate scale of divergence to provide adequate statistical power (Anisimova et al. 2001). Third, conspecific populations must exist at an early stage of divergence to allow the targeted capture and sequencing of population samples of all single-copy orthologs previously tested for positive selection. This will allow the identification of genes experiencing recent diversifying selection in

allopatry (i.e.,  $F_{ST}$ -outliers) and provide insights into the overlap between historical and contemporary selection. Finally, low levels of introgression must occur between some species that enable the identification of barrier loci that are blocked from moving between species. Ideally, multiple hybrid zones or regions experiencing introgression are present to test for the repeatability of introgression or blockage (Payseur 2010). Meeting all of these requirements is obviously extremely challenging. However, stronglycentrotid sea urchins represent one group where all of the above conditions are met and they will be used to illustrate some of the challenges of the approach described in this paper. Other groups with fairly well-resolved phylogenetic histories that could be studied in a similar manner include the genus *Octopus* (Söller et al. 2000; Guzik et al. 2005) and the genus *Crassostrea* (Yu and Li 2012; Trivedi et al. 2014).

## Genome-Wide Screens for Positive Darwinian Selection

Genome-wide tests for genes experiencing positive selection have been conducted on only a limited number of groups including *Drosophila* (*Drosophila* 12 Genomes Consortium 2007), mammals (Kosiol et al. 2008), birds (Künstner et al. 2010), ants (Roux et al. 2014), and sea urchins (Kober KM, Pogson GH, submitted for publication). Although a number of factors affect the success of genome-wide screens for positive selection, one of the most foundational is that a set of accurate gene models must be available. The well-annotated genome of the purple sea urchin *Strongylocentrotus purpuratus* provides a good example of this necessity. The *S. purpuratus* genome contains ~23,300 genes of which over 9,000 were manually annotated by an international consortium of over 200 members (Sodergren et al. 2006). The initial gene predictions were recently updated by Tu et al. (2012) who performed a deep transcriptome analysis on pooled RNA from 10 different embryonic stages, 6 different larval stages, and 6 adult tissues. These updates incorporated exons missing from the original gene models, excised



**Figure 1.** Species tree of the family Strongylocentrotidae constructed from 4-fold degenerate sites from 2,815 genes not experiencing positive selection (adapted from Kober and Bernardi 2013a). Trees reconstructed using maximum parsimony (MP), maximum likelihood (ML), and Bayesian approaches produced identical topologies. All nodes had MP bootstrap values of 100 and Bayesian posterior probabilities of 1. Abbreviations denote geographic locations of species distributions (NEP = Northeastern Pacific, NWP = Northwestern Pacific; CIR = circumpolar).

erroneously included exons, corrected inaccurate exon/intron boundaries, and removed overlapping gene models (Tu et al. 2012). Surprisingly, over half of the original manually curated gene models contained errors. These results highlight the need for direct experimental evidence (i.e., deep transcriptome sequencing) to validate computational gene predictions. The accuracy of the *S. purpuratus* gene models are on par with those from well-studied model species (humans, mice, *Drosophila*, and *Caenorhabditis elegans*) making sea urchins an excellent system for genome studies.

Using the genome of *S. purpuratus* as a reference, the complete genomes of 8 additional strongylocentrotid sea urchins have recently been obtained (Kober and Bernardi 2013a; Kober KM, Pogson GH, submitted for publication). Illumina paired-end reads (100 bp) were aligned directly to the reference genome, producing high coverage across single-copy protein-coding genes (mean = 58.2X). Although gene tree topologies were heterogeneous, the species tree was clearly resolved from 4-fold degenerate sites at 2,815 genes not experiencing positive selection (see Figure 1; Kober and Bernardi 2013a). By fitting PAML models M7 and M8 (Yang 2007) and applying a conservative false discovery rate of 5%, 1,008 of the 6,520 single-copy orthologs tested (15.5%) showed significant signals of positive selection. A further 824 candidate PSGs were found to exhibit positive selection along the 9 terminal branches of each gene tree (so-called “branch-sites” tests). Recent substitutions along these terminal branches represent prime candidates for barrier loci because of their potential negative epistatic effects (Orr 1995; Figure 2). Consistent with the action of repeated selective sweeps, the candidate PSGs had 68% higher rates of nonsynonymous substitution ( $d_N$ ) and 33% lower levels of heterozygosity than non-PSGs (both groups had similar rates of synonymous substitution,  $d_S$ ). As expected, positive selection was observed at innate immunity genes and reproductive proteins. However, it was also prevalent at membrane receptors, cell adhesion molecules, extracellular matrix proteins, and ion channels that have been implicated as targets of pathogens in other groups (notably mammals). The PSGs represent excellent a priori candidates that might experience diversifying selection in recently formed allopatric populations.

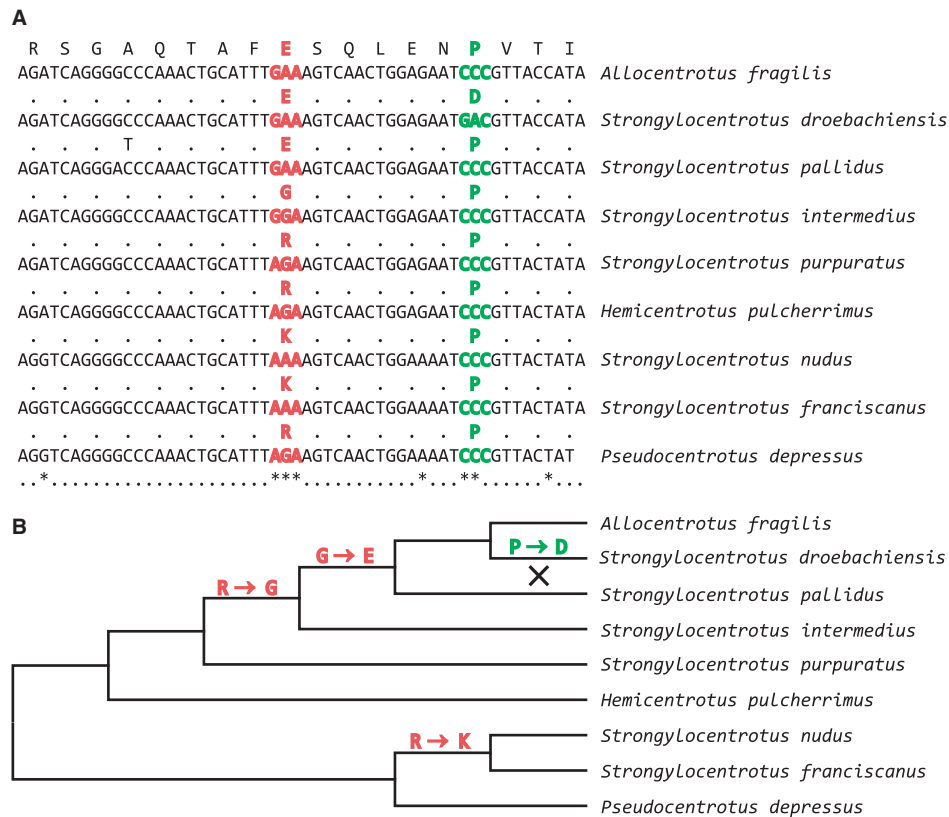
Caution must be exercised when interpreting genome-wide tests for positive selection because errors in sequencing, alignment, and annotation are known to generate significant numbers of false positives (Schneider et al. 2009; Fletcher and Yang 2010; Markova-Raina and Petrov 2011; Jordan and Goldman 2012). Although sequencing errors can be minimized by applying strict base quality metrics and high coverage cut-offs, mistakes in gene annotations

and alignments are more difficult to address. To minimize annotation and alignment errors, a conservative reference-based alignment method should be used that excludes all codons not present in the reference genome (caused by insertions) and removes all missing codons (caused by deletions) prior to testing for positive selection. This should largely avoid alignment errors caused by gains or losses of amino acids. In *Drosophila*, Markova-Raina and Petrov (2011) observed that alignment and annotation errors were common near exon borders and regions with insertions or deletions of amino acids that resulted in significant spurious signals of positive selection. Given these results, it is essential that all genome-wide screens for positive selection test for the possible enrichment of false positives in these regions. In sea urchins, Kober and Pogson (submitted for publication) subjected 3, 5, and 7 amino acid windows adjacent to exon borders and missing data regions for enrichment tests and obtained negative results. These results are important because they eliminate the possibility that the signals of positive selection are largely caused by annotation or alignment errors and provide further testament to the accuracy of the gene models.

### Targeted Genome Scans for $F_{ST}$ -Outliers

In many high gene flow marine invertebrate species inhabiting a single geographic region, it is problematic to test for  $F_{ST}$ -outliers because of the limited population genetic structure present. However,  $F_{ST}$ -outlier tests can be performed on marine invertebrate taxa that have broad geographic distributions ranges that may span different ocean basins. Here, populations often exhibit moderate levels of genetic differentiation when sampled at large spatial scales. For example, 2 strongylocentrotid species (*Strongylocentrotus droebachiensis* and *Strongylocentrotus pallidus*) have broad arctic-boreal distributions across the north Atlantic and north Pacific oceans (Mortensen 1943; Jensen 1974). No population structure has been described among populations of *S. droebachiensis* within the same geographic region but moderate genetic differentiation occurs between ocean basins (Palumbi and Wilson 1990; Addison and Hart 2004, 2005). Populations of *S. droebachiensis* from the northeast Pacific, northeast Atlantic, and northwest Atlantic exhibit significant pair-wise  $F_{ST}$ -values at the mitochondrial COI locus ranging from 0.213 to 0.325 (Addison and Hart 2005). For microsatellites, these same populations exhibit a mean overall  $F_{ST}$ -value of 0.087 (Addison and Hart 2004). Comparable studies have not been conducted on *S. pallidus* but its similar geographic distribution and pelagic larval duration to *S. droebachiensis* has likely resulted in





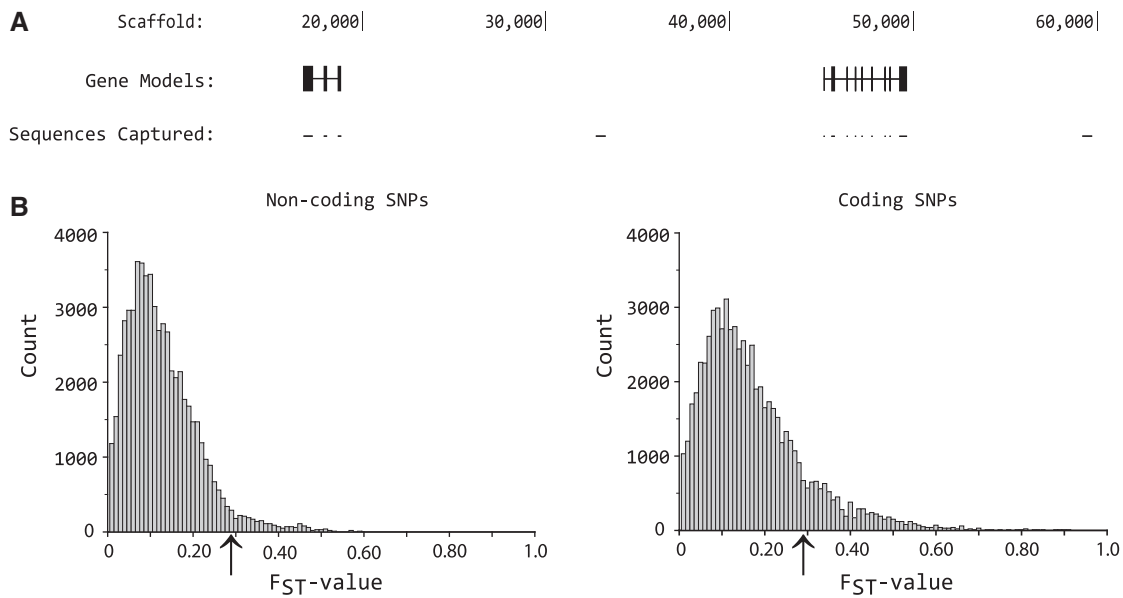
**Figure 2.** The distinction between historical and recent positive selection. (A) Alignment of portion of a hypothetical protein-coding sequence with one codon (red) showing historical positive selection and another (green) showing recent positive selection based on PAML (Yang 2007) sites tests and branch-sites tests, respectively. Historical and recent selections are not mutually exclusive. However, when both operate it is difficult to produce a significant branch-site test because the external branch needs to have a significantly elevated  $d_N/d_S$  ratio compared with the remainder of the tree. (B) Gene tree with locations of amino acid changes marked. Historical selection at the red codon cannot generate incompatibilities between *S. droebachiensis* and *S. pallidus* but recent selection at the green codon can (designated by the “X”).

similar levels of population divergence. Recently, genome-wide single-nucleotide polymorphisms (SNPs) were compared between mussel populations (*Mytilus trossulus* and *Mytilus edulis*) sampled from either side of the north Atlantic and were found to exhibit pair-wise  $F_{ST}$ -values ranging between 0.131 and 0.362 (Fraïsse et al. 2016). These results suggest that other marine invertebrates with similar broad distributions might be amenable for  $F_{ST}$ -outlier studies aimed at identifying recently diverged loci.

Excluding humans, the majority of genome-wide scans for  $F_{ST}$ -outliers have used randomly located SNPs, implicitly assuming that patterns of polymorphism across the genome reflect neutral and demographic processes (reviewed by Oleksyk et al. 2010; Haas and Payseur 2016). One of the long-standing problems faced by  $F_{ST}$ -outlier studies is the difficulty of accurately quantifying the neutral  $F_{ST}$  distribution. Some have questioned whether, in some species, a neutral background actually exists (Comeron 2014; Burri et al. 2015). The large effective population sizes of many marine invertebrates may have resulted in the pervasive influence of linked selection across the genome, similar to that observed in *Drosophila* (Begun et al. 2007; Langley et al. 2012; Comeron 2014). Targeting protein-coding genes for genome scans confounds this problem because these regions have the potential to directly experience diversifying selection, background selection, balancing selection, and even weak selection on patterns of synonymous codon usage (see Kober and Pogson 2013b).

With a reference genome and prior information on genes known to have experienced positive selection there are several ways to estimate a neutral  $F_{ST}$  distribution. One is to score SNPs at single-copy genes that do not exhibit positive selection (non-PSGs) or exhibit extremely low rates of nonsynonymous substitution. Another approach would be to score thousands of SNPs in non-coding, non-repetitive intergenic regions at large distances (e.g., >50 kb) from any known structural gene. Here, it would be possible to avoid highly conserved non-coding elements, transcription factor-binding sites, and to exclude all regions containing any PSGs. It would be worthwhile evaluating whether null distributions provided by non-PSG SNPs or those based on non-coding SNPs provide the most robust identification of  $F_{ST}$ -outliers. Because of differences in GC content, recombination rates may differ between non-coding and coding regions of the genome (Eyre-Walker 1993). The neutral  $F_{ST}$  distribution based on non-PSG SNPs might thus result in a more accurate detection of genes experiencing recent diversifying selection. Irrespective of this outcome, it will also be necessary to confirm the outlier status of SNPs using an absolute measure of divergence (such as  $d_{XY}$ ) to avoid the spurious inflation of  $F_{ST}$ -values caused by reduced levels of polymorphism (see Noor and Bennett 2009; Cruickshank and Hahn 2014).

As the cost of next-generation sequencing continues to fall, it is tempting to collect complete genome data for SNP identification and calling. However, because protein-coding genes are the focus of this



**Figure 3.** Identification of  $F_{ST}$ -outliers by the targeted sequence capture of non-coding and coding SNPs. (A) Two gene models are shown on a scaffold that are targeted for solution-based exon capture. Two non-coding regions  $>10$  kb from either gene are also captured (but could be much further away). (B) Hypothetical distributions of pairwise  $F_{ST}$ -values for genome-wide coding and non-coding SNPs are presented. The arrow designates the 5% false discovery rate cut-off estimated directly from the non-coding SNPs. Any genes or gene regions with  $F_{ST}$ -values beyond this threshold are candidates for recent diversifying selection. Null distributions for  $F_{ST}$ -outlier tests could also be generated from the subset of protein-coding genes that do not exhibit histories of positive selection.

research strategy, a more cost-effective approach would be to use high-throughput hybridization-based targeted sequence capture of exons (reviewed by Mamanova et al. 2010). Exon capture has proven to be a cost-effective genotype-by-sequencing method that, although widely used in human genomics, has not been widely applied to non-model species (see Jones and Good 2016). Many studies have shown that sequence capture using RNA “baits” generates longer contigs, produces higher quality SNP calls, and provides more even coverage than other genotype-by-sequencing methods like Rad-Seq (Baird et al. 2008) or variants thereof (see Tewhey et al. 2009; Harvey et al. 2013). Although some of the SNPs scored at non-PSGs or in non-coding regions may not be truly neutral, they are expected to provide a conservative null distribution because linked selection will act to inflate their variance and make  $F_{ST}$ -outlier calls more conservative.

$F_{ST}$ -outlier studies are susceptible to false positive signals of diversifying selection due to a variety of factors including population bottlenecks (Foll and Gaggiotti 2008), hierarchical population structure (Excoffier et al. 2009; Fourcade et al. 2013), cryptic hybrid zones (Bierne et al. 2011), and range expansions (Hofer et al. 2009). The proposed targeted capture of SNPs in both coding and non-coding regions of the genome offers a number of advantages over standard genome scans. First, the ability to detect  $F_{ST}$ -outliers should be significantly improved over the practice of scoring random SNPs in unknown locations. Rather than relying on linkage disequilibrium between a randomly located SNP and a selected gene, the protein-coding regions captured are potentially the direct targets of selection. Second, characterization of a large number of putatively neutral SNPs enables application of the empirical  $P$ -value approach described by Lotterhos and Whitlock (2014) where the significance of an  $F_{ST}$ -outlier is based on its quantile in the empirical distribution of the neutral SNPs (Figure 3). Lotterhos and Whitlock (2014) found that this empirical  $P$ -value approach significantly reduced false positives, even under non-equilibrium scenarios such as recent range expansions. Third, the  $F_{ST}$ -outliers in coding regions can be more easily

traced to specific mutations at genes with known selective histories and the impact of linked selection in these regions can be examined in detail (reviewed by Sousa and Hey 2013; Vatsiou et al. 2016). Fourth, alleles that have introgressed between species can be identified by reconstructing gene genealogies and by applying powerful new tests for detecting introgression (e.g., Geneva et al. 2015; Rosenzweig et al. 2016). Discovering that genes with histories of positive selection have successfully moved between species might provide evidence favoring adaptive introgression (Hedrick 2013). Finally, the neutral SNPs could provide robust insights into the species demographic histories (Excoffier et al. 2013; Liu and Fu 2015), which can provide important insights when interpreting the patterns of diversifying selection.

The  $F_{ST}$ -outliers detected among widely separated populations of marine invertebrates could provide important insights into the genes that first diverge following the colonization of new habitats by long-distance dispersal. A broad overlap between the  $F_{ST}$ -outlier genes and those known a priori to have experienced positive selection would confirm that diversifying selection is ongoing and that there is a tight coupling of historical and contemporary selection. This might be expected for genes experiencing sexual selection or conflict where relentless cycles of adaptation and counter-adaptation are expected. However, the link between selection at different time scales may be weak. If so, this might suggest that novel forms of selection are being imposed on the new allopatric population, such as that mediated by new suites of pathogens. However, the detection of diversifying selection at a locus does not provide any insight into whether it causes a genetic barrier between diverging lineages. This can only be assessed by performing controlled crosses or by examining the patterns of naturally occurring introgression.

### Genome Scans for Barrier Loci

Hybridization is common in the marine environment, occurring at similar levels to that described in terrestrial systems (reviewed by

Gardner 1997; Arnold and Fogarty 2009). In his review, Gardner (1997) observed that only 34 of the 108 documented cases of marine hybridization involved hybrid zones or hybrid swarms. The majority of hybridization events in the sea thus occur in the absence of distinct hybrid zones. Many hybridizing broadcast-spawners have broadly overlapping geographic ranges and heterospecific matings must occur randomly when breeding coincides in space and time. Such appears to be the case for *S. droebachiensis* and *S. pallidus*. The reproductive barriers between the 2 taxa appear fairly well established and  $F_1$  hybrids in nature are rare. For example, using morphological characters Vasseur (1952) identified 3 hybrids in a sample of 562 individuals (0.54%) from northern Norway. Although these observations suggest that gene flow between the two species is limited, even small amounts of backcrossing can result in significant introgression (Barton 2001). Low levels of asymmetric introgression has been recently observed from *S. pallidus* into *S. droebachiensis* at a small number of nuclear genes in populations from the northeastern Pacific region (Addison and Pogson 2009; Pujolar and Pogson 2011). The prevalence and identity of genes experiencing introgression across the genomes of this species pair is unknown.

The genomic clines method of Gompert and Buerkle (2009, 2010) can be used to study the patterns of introgression between diverged lineages to identify putative barrier loci. This approach uses multinomial regression to estimate the probability of a specific genotype at a marker given an observed level of genome-wide admixture (i.e., a hybrid index). The genomic clines method relies on estimating a model of neutral introgression. This is done by randomly permuting genotypes among loci within individuals, or by using genetic data from populations outside of the hybrid zone to estimate the probability of genotypes in the admixed population (see Gompert and Buerkle 2009). Similar to the problem faced by estimating the neutral  $F_{ST}$  distribution, using SNPs from protein-coding genes to estimate the neutral admixture model is inappropriate. However, the random permutation of non-PSG or non-coding SNP genotypes among markers within individuals should provide an accurate characterization of the neutral pattern of admixture. This in turn should increase the power for identifying barrier loci that exhibit significantly reduced levels of introgression between species.

A growing number of hybrid zone studies have reported highly variable degrees of hybridization and patterns of introgression in different geographic regions (e.g., Nolte et al. 2009; Teeter et al. 2010; Mandeville et al. 2015). Interpreting heterogeneous patterns of introgression is particularly challenging because it can result from geographic differences in a number of pre- or post-zygotic isolating mechanisms, changes in relative abundances, or from the presence of population-specific barriers. Since most broadcast spawning marine invertebrates lack strong mate choice, one of the most basic ways for patterns of introgression to vary among geographic regions are through differences in the degree of spatial or temporal isolation. If this information is available, it can be used to predict how introgression may differ between geographic regions. For example, in the northeast Pacific, the reproductive cycles of *S. droebachiensis* and *S. pallidus* overlap substantially (Leviton 1998) and  $F_1$  individuals produced from these populations in the laboratory are viable and fertile (Strathmann 1981). In the northeastern Atlantic, there is greater spawning asynchrony between *S. droebachiensis* and *S. pallidus*, with the former spawning earlier in the year (Vasseur 1952; Hagström and Lönning 1967). Therefore, in this system one might predict that, all else being equal, higher levels of introgression might occur in the northeast Pacific than the northeast Atlantic.

Another factor complicating patterns of introgression among populations is that the distributions of closely related species may vary geographically. Recent studies on Lake Victoria cichlids (Keller et al. 2013), Colorado River catostomid fishes (Mandeville et al. 2015), and blue mussels belonging to the genus *Mytilus* (Fraisé et al. 2016) have also illustrated how hybridization and introgression often needs to be assessed from a multi-species perspective. It is becoming increasingly clear that a geographical context is essential for properly interpreting patterns of introgression. In stronglylocentroid sea urchins from the northeast and northwest Atlantic, heterospecific matings are only possible between *S. droebachiensis* and *S. pallidus*. However, in the north Pacific different assemblages of species broadly co-occur along the Asian and North American coasts (see Figure 1). If reproductive barriers are incomplete among other species, the patterns of introgression might vary in different and unpredictable ways across the north Pacific.

## Some Limitations

The research strategy described here for evaluating the role of positive selection in producing barrier loci faces several important limitations. First, tests for positive selection based on  $d_N/d_S$  ratios are known to be conservative (Anisimova et al. 2001; Gharib and Robinson-Rechavi 2013; Lu and Guindon 2013). It is thus likely that many genes undergoing adaptive diversification fail to exhibit statistically significant signals of positive selection and their omission would result in an unknown number of false negatives. Less stringent criteria could be used to classify candidate barrier genes such as those exhibiting the highest rates of nonsynonymous substitution, irrespective of positive selection per se. Second, recently formed paralogous genes create challenges for genome assemblies and alignments and thus will be underrepresented in the tests for barrier loci that use conserved single-copy orthologs. The importance of new genes in forming reproductive barriers between species is largely unknown, but they appear to diverge quickly and often acquire new functions (Chen et al. 2013; Long et al. 2013). Finally, although candidate barrier genes may be successfully identified, the forms of selection responsible for their divergence may be difficult to discern and additional functional tests (and genetic crosses) are needed to confirm their direct involvement in establishing species barriers. In some cases, the identities of barrier loci might provide clear insights into a specific selective agent. For example, observing that innate immunity genes often experience positive selection and are strongly blocked from introgressing would implicate pathogens as the drivers of divergence. Distinguishing between pre- or post-zygotic barriers will also be challenging, but in some cases can also be discerned from gene identity. For example, a strong barrier at a GRP would clearly implicate a post-mating prezygotic barrier, but the form of selection responsible remains unknown.

Broadcast spawning marine invertebrates often possess extensive levels of genetic polymorphism that may complicate some of the analyses proposed here. Multi-species alignments of single-copy genes will contain large numbers of heterozygous mutations that must be filtered prior to testing for positive selection. This is because many mutations represent low frequency nonsynonymous changes (likely deleterious) that, if included, would lead to overestimation of  $d_N$  and  $d_N/d_S$  ratios and generate false signals of positive selection. High levels of heterozygosity can result in the loss of considerable data. For example, at the 6,520 single-copy genes studied by Kober and Pogson (submitted for publication) in sea urchins, codons containing 1.72 million high-quality SNPs were filtered prior to testing

for positive selection. High levels of standing variation may also lead to the confounding of divergence with polymorphism if only one individual is sequenced for each species. Sequencing the transcriptomes of several individuals per species might improve the identification of fixed differences and allow McDonald–Kreitman tests (McDonald and Kreitman 1991) to be performed that would complement the tests for selection based on  $d_N/d_S$  ratios.

Long-distance colonization events involving marine invertebrate populations with high levels of standing variation also leads to the prediction that soft sweeps (Hermisson and Pennings 2005; Pritchard et al. 2010) could be more common than hard sweeps. Since soft sweeps generate diminished signals of genetic hitchhiking, the ability to detect  $F_{ST}$ -outliers may be challenging especially in regions with high rates of recombination. Interpreting soft sweeps may be greatly improved if the biogeographical history of the group is known and if the beneficial alleles are still segregating in known ancestral populations. In the case of *S. droebachiensis* and *S. pallidus*, populations in across the north Atlantic are known to have originated from the north Pacific (Durham and MacNeil 1967; Addison and Hart 2005). It is thus possible that haplotypes experiencing sweeps in the north Atlantic may be identifiable in populations from the north Pacific.

## Concluding Remarks

Confirming that a diverse array of genes experiencing positive selection diverge early in the speciation process and commonly result in reproductive incompatibilities would represent an important step forward in understanding of the genetic basis of speciation. The power of the approach outlined here results from the targeted capture of both con-coding and coding SNPs in known locations and at genes with known selective (or neutral) histories. The neutral coding and non-coding SNPs provide more accurate estimations of the neutral  $F_{ST}$  distribution for  $F_{ST}$ -outlier tests and neutral admixture rates for the genomic clines analyses identifying barrier loci. It is also worth noting that negative results would be equally important to speciation research. Discovering that positive selection does not typically produce barrier loci would direct attention to genetic elements in non-coding regions responsible for establishing and maintaining species boundaries (e.g., Jones et al. 2012; Nadeau et al. 2012; Hebert et al. 2013). Moving beyond the post hoc characterization of genome-wide patterns of divergence to a hypothesis testing framework targeting known genes offers many advantages. It can provide direct and detailed insights into the identity of genes responsible for establishing species boundaries and the importance of gene flow in preventing or facilitating the adaptive divergence of lineages (Abbott et al. 2013). Although the ability to simultaneously study early divergence and patterns of introgression between young marine invertebrate species is clearly challenging, the approach outlined here might provide an exciting way forward.

## Acknowledgments

I would like to thank Nicolas Bierne for the opportunity of contributing to the Special Column: Population Genomics in the Sea. Helpful comments on the manuscript were provided by Nicolas Bierne and two anonymous reviewers.

## Funding

Partial funding for the work described on stronglycentrotid sea urchins was provided by the National Science Foundation (DEB-1011061).

## References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE et al., 2013. Hybridization and speciation. *J Evol Biol* 26:229–246.
- Addison JA, Hart MW, 2004. Analysis of population genetic structure of the green sea urchin *Strongylocentrotus droebachiensis* using microsatellites. *Mar Biol* 144:243–251.
- Addison JA, Hart MW, 2005. Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins *Strongylocentrotus droebachiensis*. *Evolution* 59:532–543.
- Addison JA, Pogson GH, 2009. Multiple gene genealogies reveal asymmetrical hybridization and introgression among stronglycentrotid sea urchins. *Mol Ecol* 18:1239–1251.
- Albertin CB, Simakov O, Mitros T, Yan Wang Z, Pungor JR et al., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524:220–224.
- Andrés JA, Larson EL, Bogdanowicz SM, Harrison RG, 2013. Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. *Genetics* 193:501–513.
- Anisimova M, Nielsen R, Yang Z, 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18:1585–1592.
- Arnegard ME, McGee MD, Matthews B, Marchinko KB, Conte GL et al., 2014. Genetics of ecological divergence during speciation. *Nature* 507:307–311.
- Arnold ML, Fogarty ND, 2009. Reticulate evolution and marine organisms: the final frontier? *Int J Mol Sci* 10:3836–3860.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376.
- Barton NH, 2001. The role of hybridization in evolution. *Mol Ecol* 10:551–568.
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP et al., 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5:e310.
- Bierne N, Welch J, Loire E, Bonhomme F, David P, 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* 20:2044–2072.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI et al., 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* 25:1656–1665.
- Butlin RK, Galindo J, Grahame JW, 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Phil Trans Roy Soc B Biol Sci* 363:2997–3007.
- Chen S, Krinsky BH, Long M, 2013. New genes as drivers of phenotypic evolution. *Nat Rev Genet* 14:645–660.
- Clark NL, Aagaard JE, Swanson WJ, 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11–22.
- Comeron J, 2014. Background selection as a baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genet* 10:e1004434.
- Coyne JA, Orr HA, 2004. *Speciation*. Sunderland: Sinauer Associates.
- Crespi B, Nosil P, 2013. Conflictual speciation: species formation via genomic conflict. *Trends Ecol Evol* 28:48–57.
- Cruikshank TE, Hahn MW, 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B et al., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298:2157–2167.
- Ding Q, Hu Y, Xu S, Wang J, Jin L, 2014. Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. *Mol Biol Evol* 31:683–695.



- Dobzhansky T, 1937. *Genetics and the Origin of Species*. New York: Columbia University Press.
- Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Durham JW, MacNeil FS, 1967. Cenozoic migrations of marine invertebrates through the Bering Strait region. In: Hopkins DM, editor. *The Bering Land Bridge*. Stanford: Stanford University Press, 326–349.
- Eyre-Walker A, 1993. Recombination and mammalian genome evolution. *Proc Roy Soc Lond B* 252:237–243.
- Excoffier L, Hofer T, Foll M, 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298.
- Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M, 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet* 9:e1003905.
- Feder JL, Egan SP, Nosil P, 2012a. The genomics of speciation-with-gene-flow. *Trends Genet* 28:342–350.
- Feder JL, Geiji R, Yeaman S, Nosil P, 2012b. Establishment of new mutations under divergence and genome hitchhiking. *Phil Trans Roy Soc B Biol Sci* 367:461–474.
- Fitzpatrick BM, Fordyce JA, Gavrillets S, 2009. Pattern, process and geographic modes of speciation. *J Evol Biol* 22:2342–2347.
- Flaxman SM, Feder JL, Nosil P, 2013. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution* 67:2577–2591.
- Fletcher W, Yang Z, 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol Biol Evol* 27:2257–2267.
- Foll M, Gaggiotti OE, 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993.
- Fourcade Y, Chaput-Bardy A, Secondi J, Fleurant C, Lemaire C, 2013. Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Mol Ecol* 22:2065–2073.
- Fraïsse C, Belkhir K, Welch JJ, Bierne N, 2016. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol Ecol* 25:269–286.
- Gardner JPA, 1997. Hybridization in the sea. *Adv Mar Biol* 31:1–77.
- Gavrillets S, 2004. *Fitness Landscapes and the Origin of Species*. Princeton: Princeton University Press.
- Geneva AJ, Muirhead CA, Kingan SB, Garrigan D, 2015. A new method to scan genomes for introgression in a secondary contact model. *PLoS ONE* 10:e0118621.
- Gharib WH, Robinson-Rechavi M, 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol* 30:1675–1686.
- Gompert Z, Buerkle CA, 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol Ecol* 18:1207–1224.
- Gompert Z, Buerkle CA, 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour* 10:378–384.
- Guzik MT, Norman MD, Crozier RH, 2005. Molecular phylogeny of the benthic shallow-water octopuses (Cephalopoda: Octopodinae). *Mol Phylogenet Evol* 37:235–248.
- Hagström BE, Lönning S, 1967. Experimental studies on *Strongylocentrotus droebachiensis* and *S. pallidus*. *Sarsia* 29:165–176.
- Haas RJ, Payseur BA, 2016. Fifteen years of genome-wide scans for selection: trends, lessons and unaddressed genetic sources for complication. *Mol Ecol* 25:5–23.
- Harrison RG, 1990. Hybrid zones: windows on evolutionary process. *Oxf Surv Evol Biol* 7:69–128.
- Harrison RG, 2012. The language of speciation. *Evolution* 66:3643–3657.
- Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT, 2013. Sequence capture versus restriction site associated DNA sequencing for phylogeography. *arXiv*:1312.6439.
- Hart MW, Sunday JM, Popovic I, Learning KJ, Konrad CM, 2014. Incipient speciation of sea star populations by adaptive gamete recognition coevolution. *Evolution* 68:1294–1305.
- Hebert FO, Renaut S, Bernatchez L, 2013. Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair *Coregonus clupeaformis*. *Mol Ecol* 22:4896–4914.
- Hedrick PW, 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol* 22:4606–4618.
- Hermisson J, Pennings PS, 2005. Soft sweeps: molecular population genetics of adaptation from standing variation. *Genetics* 169:2335–2352.
- Hofer T, Ray N, Wegmann D, Excoffier L, 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet* 73:95–108.
- Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM et al., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512:194–197.
- Jackson JBC, Johnson KG, 2000. Life in the last few million years. *Paleobiology* 26:221–234.
- Jacobs DK, Haney TA, Louie KD, 2004. Genes, diversity, and geologic process on the Pacific coast. *Annu Rev Earth Planet Sci* 32:601–652.
- Jansson R, Dynesius M, 2002. The fate of clades in a world of recurrent climate change: Milankovitch oscillations and evolution. *Annu Rev Ecol Syst* 33:741–777.
- Jensen M, 1974. The Strongylocentrotidae (Echinoidea), a morphological and systematic study. *Sarsia* 57:185–217.
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceci E et al., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
- Jones MR, Good JR, 2016. Targeted capture in evolutionary and ecological genomics. *Mol Ecol* 25:185–202.
- Jordan G, Goldman N, 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol Biol Evol* 29:1125–1139.
- Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM et al., 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol Ecol* 22:2848–2863.
- Knowlton N, 1993. Sibling species in the sea. *Annu Rev Ecol Syst* 24:189–216.
- Kober KM, Bernardi G, 2013a. Phylogenomics of Strongylocentrotid sea urchins. *BMC Evol Biol* 13:88.
- Kober KM, Pogson GH, 2013b. Genome-wide patterns of codon bias are shaped by natural selection in the purple sea urchin *Strongylocentrotus purpuratus*. *G3* 3:1069–1083.
- Kosakovsky Pond SL, Frost SD, Muse SV, 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
- Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD et al., 2008. Patterns of positive selection in six mammal genomes. *PLoS Genet* 4:e1000144.
- Kosman ET, Levitan DR, 2014. Sperm competition and the evolution of gametic compatibility in externally fertilizing taxa. *Mol Hum Reprod* 20:1190–1197.
- Künstner A, Wolf JB, Backström N, Whitney O, Balakrishnan CN et al., 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol* 19:266–276.
- Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR et al., 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192:533–598.
- Lessios HA, 2007. Reproductive isolation between species of sea urchins. *Bull Mar Sci* 81:191–208.
- Lessios HA, 2008. The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annu Rev Ecol Syst* 39:63–91.
- Lessios HA, 2011. Speciation genes in free-spawning marine invertebrates. *Integr Comp Biol* 51:456–465.

- Levitán DR, 1998. Does Bateman's principle apply to broadcast-spawning organisms? Egg traits influence *in situ* fertilization rates among congeneric sea urchins. *Evolution* 52:1043–1056.
- Lindberg DR, 1991. Marine biotic interchange between the northern and southern hemispheres. *Paleobiology* 17: 308–324.
- Lindberg DR, Lipps JH, 1996. Reading the chronicle of quaternary temperate rocky shore faunas. In: Jablonski D, Erwin DH, Lipps JH, editors. *Evolutionary Paleobiology*. Chicago: University of Chicago Press, 161–182.
- Liu X, Fu Y-X, 2015. Exploring population size changes using SNP frequency spectrum. *Nat Genet* 47:555–559.
- Long M, VanKuren NW, Chen S, Vibranovski MD, 2013. New gene evolution: little did we know. *Annu Rev Genet* 43:307–333.
- Lotterhos KE, Whitlock MC, 2014. Evaluation of demographic history and neutral parameterization on the performance of  $F_{ST}$  outlier tests. *Mol Ecol* 23:2178–2192.
- Lu A, Guindon S, 2013. Performance of standard and stochastic branch-sites models for detecting positive selection among coding sequences. *Mol Biol Evol* 31:484–495.
- Luo YJ, Takeuchi T, Koyanagi R, Yamada L, Kanda M et al., 2015. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature Commun* 6:8301.
- Lynch M, Force AG, 2000. The origin of interspecific genomic incompatibility via gene duplication. *Am Nat* 156:590–605.
- Mamanova L, Coffrey AJ, Scott CE, Kozarewa I, Turner EH et al., 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118.
- Mandeville EG, Parchman TL, McDonald DB, Buerkle CA, 2015. Highly variable reproductive isolation among pairs of *Catostomus* species. *Mol Ecol* 24:1856–1872.
- Markova-Raina P, Petrov D, 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res* 21:863–874.
- McDonald JH, Kreitman M, 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- Mortensen T, 1943. *A Monograph of the Echinoidea*. Vol III, Part 3. Copenhagen: CA Reitzel.
- Muller HJ, 1940. Bearing of the *Drosophila* work on systematics. In: Huxley JS editor. *The New Systematics*. Oxford: Clarendon Press, 185–268.
- Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK et al., 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Phil Trans Roy Soc B Biol Sci* 367:343–353.
- Navarro A, Barton NH, 2003. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* 300:321–324.
- Nolte AW, Gompert Z, Buerkle CA, 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol Ecol* 18:2615–2627.
- Nolte V, Pandey RV, Kofler R, Schlötterer C, 2013. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. *Genome Res* 23:99–110.
- Noor MAF, Bennett SM, 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103:439–444.
- Noor MAF, Feder JL, 2006. Speciation genetics: evolving approaches. *Nat Rev Genet* 7:851–861.
- Norris RD, Hull PM, 2012. The temporal dimension of marine speciation. *Evol Ecol* 26:393–415.
- Nosil P, 2012. *Ecological Speciation*. Oxford: Oxford University Press.
- Nosil P, Funk DJ, Ortíz-Barrionto D, 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402.
- Nosil P, Schluter D, 2011. The genes underlying the process of speciation. *Trends Ecol Evol* 26:160–167.
- Oleksyk TK, Smith MW, O'Brien SJ, 2010. Genome-wide scans for footprints of natural selection. *Phil Trans Roy Soc B Biol Sci* 365:185–205.
- Orr HA, 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139:1805–1813.
- Palumbi SR, 2009. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* 102:66–76.
- Palumbi SR, Wilson AC, 1990. Mitochondrial DNA diversity in the sea urchins *Strongylocentrotus purpuratus* and *S. droebachiensis*. *Evolution* 44:403–415.
- Payseur BA, 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol Ecol Resour* 10:806–820.
- Payseur BA, Nachman MW, 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biol J Linn Soc* 84:523–534.
- Pfenninger M, Schwenk K, 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evol Biol* 7:121.
- Phadnis N, Orr HA, 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323:376–379.
- Presgraves DC, 2010. The molecular evolutionary basis of species formation. *Nat Rev Genet* 11:175–180.
- Pritchard JK, Pickrell JK, Coop G, 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20:R208–R215.
- Pujolar JM, Pogson GH, 2011. Positive Darwinian selection in gamete recognition proteins of *Strongylocentrotus* sea urchins. *Mol Ecol* 20:4968–4982.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J et al., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.
- Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U et al., 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
- Rice WR, 1998. Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In: Howard DJ, Berlocher SH, editors. *Endless Forms: Species and Speciation*. New York: Oxford University Press, 261–270.
- Rieseberg LH, Blackman BK, 2010. Speciation genes in plants. *Ann Bot* 106:439–455.
- Rieseberg LH, Whitton J, Gardner K, 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152:713–727.
- Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW, 2016. Powerful methods for detecting introgressed regions from population genomic data. *Mol Ecol* 25:2387–2397.
- Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M et al., 2014. Patterns of positive selection in seven ant genomes. *Mol Biol Evol* 31:1661–1685.
- Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT et al., 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* 342:1242592.
- Sanford E, Kelly MW, 2011. Local adaptation in marine invertebrates. *Annu Rev Mar Sci* 3:509–535.
- Schluter D, 2009. Evidence for ecological speciation and its alternative. *Science* 323:737–741.
- Schneider A, Souvorov A, Sabath N, Landan G, Gonnet GH et al., 2009. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol Evol* 1:114–118.
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW et al., 2014. Genomics and the origin of species. *Nat Rev Genet* 15:176–192.
- Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P et al., 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* 493:526–531.
- Seo HC, Kube M, Edvardsen RB, Jensen MF, Beck A et al., 2001. Miniature genome in the marine chordate *Oikopleura dioica*. *Science* 294:2506.
- Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA et al., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–952.
- Söllner R, Warnke K, Saint-Paul U, Blohm D, 2000. Sequence divergence of mitochondrial DNA indicates cryptic biodiversity in *Octopus vulgaris* and supports the taxonomic distinctiveness of *Octopus mimus* (Cephalopoda: Octopodidae). *Mar Biol* 136:29–35.

- Sousa V, Hey J, 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet* 14:404–414.
- Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U et al., 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454:955–960.
- Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA et al., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466:720–726.
- Strathmann RR, 1981. On the barriers to hybridization between *Strongylocentrotus droebachiensis* (O.F. Müller) and *S. pallidus* (G.O. Sars). *J Exp Mar Biol Ecol* 55:39–47.
- Swanson WJ, Vacquier VD, 2002. Reproductive protein evolution. *Annu Rev Ecol Syst* 33:161–179.
- Tang SW, Presgraves DC, 2009. Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323:779–782.
- Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW et al., 2010. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64:472–485.
- Tewhey R, Nakano M, Wang X, Pabon-Pena C, Novak B et al., 2009. Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 10:R116.
- The Heliconius Genome Consortium, 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 48:94–98.
- The Marie Curie SPECIATION Network, 2012. What do we need to know about speciation? *Trends Ecol Evol* 27:27–39.
- Ting CT, Tsaour SC, Wu MI, Wu C-I, 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282:1501–1504.
- Trivedi S, Aloufi AA, Ansari AA, Ghosh SK, 2014. Molecular phylogeny of oysters belonging to the genus *Crassostrea* through DNA barcoding. *J Entomol Zool Stud* 2:308–313.
- Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EC, 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res* 22:2079–2087.
- Turner TL, Hahn MW, Nuzhdin SV, 2005. Genomic islands of divergence in *Anopheles gambiae*. *PLoS Biol* 3:1572–1578.
- Vacquier VD, Swanson WJ, 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb Perspect Biol* 2011 3:a002931.
- Valentine JW, Jablonski D, 1993. Fossil communities: compositional variation at many time scales. In: Ricklefs RE, Schluter D, editors. *Species Diversity in Ecological Communities*. Chicago: University of Chicago Press, 341–349.
- Vasseur E, 1952. Geographic variation in the Norwegian sea-urchins *Strongylocentrotus droebachiensis* and *S. pallidus*. *Evolution* 6:87–100.
- Vatsiou AI, Bazin E, Gaggiotti OE, 2016. Detection of selective sweeps in structured populations: a comparison of recent methods. *Mol Ecol* 25:89–103.
- Via S, 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil Trans Roy Soc B Biol Sci* 367:451–460.
- Via S, West J, 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol* 17:4334–4345.
- Vermeij G, 1991. Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology* 12:281–307.
- Yang Z, 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
- Yu H, Li Q, 2012. Complete mitochondrial DNA sequence of *Crassostrea nippona*: comparative and phylogenomic studies on seven commercial *Crassostrea* species. *Mol Biol Rep* 39:999–1009.
- Zhang G, Fang X, Guo X, Li L, Luo R et al., 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49–54.
- Zigler KS, McCartney MA, Levitan DR, Lessios HA, 2005. Sea urchin binding divergence predicts gamete compatibility. *Evolution* 59:2399–2404.

