**Widespread introgression within the strongylocentrotid family of sea urchins challenges rapid speciation by evolution of gamete recognition**

\*\* 8,000 word limit (Abstract, Introduction, Materials and Methods, Results and Discussion) \*\*

Matthew R. Glasenapp\* and Grant H. Pogson\*

Department of Ecology and Evolutionary Biology, University of California, Santa Cruz

**Notes to Self:**

* Arguments about lack of postzygotic isolation are often made about hybrids raised only to metamorphosis, confirming that development is normal. But essentially nothing is known about hybrid survival or fecundity in the wild.
* “The fossil record confirms that species high larval dispersal have broader ranges, greater species durations, and slower rates of speciation than similar species with low dispersal (ref 13, Jablonski 1986). These studies support the early generalization that speciation in the sea is usually by gradual build-up of genomic incompatibility in allopatric populations, and that such genetic divergence of high-dispersal populations is minor and slow(14) (Palumbi 1992).”
* Ranges may be very misleading. Don’t have a solid handle on current ranges. Impossible to know ranges in the distant past over an evolutionary time scale

**Abstract** (No more than 250 words)

**Keywords** (Provide four to six keywords)

Introgression, hybridization, urchin, bindin

**1 Introduction**

**The Species Problem**

The origin of new species remains a central and elusive problem in evolutionary biology. A significant hurdle has been connecting the continuous processes of evolution, such as allele frequency change, to the origin of discrete groups in space and time. If a version of the biological species concept of Mayr and Dobzhansky is employed, the process of speciation can be understood as the evolution of reproductive isolating barriers between differentiated populations (Dobzhansky 1937; Coyne and Orr 2004). Speciation research to date has therefore focused on identifying reproductive barriers and the evolutionary processes behind their emergence.

**(The Genetics of ) Reproductive Isolating Barriers**

In many groups, the identity of isolating barriers, the genes involved, and the type(s) of selection behind their evolution remain unknown. The origin of reproductive isolation is primarily driven by natural selection and sexual selection, with genetic drift playing a less important role (Coyne and Orr, 2004). It has been well established that characters involved in diversifying selection often exhibit pleiotropic effects on reproductive compatibility and produce isolating barriers (citation), but the connection between genetics and reproductive isolation remains weak (citation). Little is known about the number and kinds of genes that contribute to speciation. Studying speciation therefore necessitates the identification of the reproductive barriers that initially reduced gene flow and the evolutionary forces behind their appearance. Identifying reproductive barriers and their order of occurrence is difficult in older species because barriers continue to accumulate after speciation and often, multiple isolating barriers work together to reduce gene flow (Coyne and Orr 2004). Current barriers to gene flow may therefore be misleading about which barriers were important in the early stages of speciation.

**Broadcast-Spawning Marine Invertebrate Speciation**

Broadcast-spawning marine invertebrates are a compelling group for speciation studies because their life history is drastically different from most animal speciation models. Speciation patterns and processes likely differ between taxa with different life history strategies, and most animal speciation research to date has focused on groups with low dispersal and fecundity (citation, is this still true?). Broadcast spawners represent a unique opportunity to test whether conclusions from speciation research in well studied terrestrial organisms hold for marine organisms with drastically different reproductive systems. Do speciation patterns group by life history or by evolutionary history? For example, it would be interesting to know if broadcast-spawning marine invertebrate speciation patterns more closely resemble closely related low dispersal animal taxa or distantly related seed plants with high fecundity and dispersal (Palumbi 1992).

Additionally, the diversity present within broadcast-spawning marine invertebrates is difficult for allopatric speciation theory to explain because their large population sizes, broad geographic ranges, and high dispersal rates would seemingly limit opportunities for population differentiation and evolution of reproductive isolating barriers (Palumbi 1994). The scale of population structure in broadcast spawning marine invertebrates can be on the order of thousands to tens of thousands of kilometers (Palumbi 1992). If speciation in these groups is dependent on long periods of time in allopatry, the current diversity of these groups may be unexpectedly high given their high level of gene flow (but see Palumbi 1994). Because of their highly dispersive planktonic larval stage, broadcast-spawners typically have large population sizes, large ranges, and massive fecundities, which should result in slowed rates of speciation. The fossil record supports this conclusion (Jablonski 1986), but species richness in marine environments if often high (Palumbi 1992).

Broadcast-spawning marine invertebrates are underrepresented in speciation research because they are often inaccessible, difficult to directly observe, are difficult to rear in the lab, and have long generation times that make experimental work challenging (Palumbi 1994; Lessios 2011). A notable exception has been the use of abalones, oysters, and sea urchins in studies of gamete compatibility. In broadcast-spawning marine invertebrates, there are no complex pre-mating behaviors such as courtship or copulation. The only potential barriers that could precede the interaction of gametes would be differences in habitat use or spawning time.

**Rapid Speciation by Evolution of Gamete Recognition**

The observation of species-specific fertilization in groups of broadcast spawners has promoted the idea that proteins involved in the binding and fusion of gametes may have been involved in speciation (Coyne and Orr 2004; Palumbi 1994; Palumbi 2009). A large area of research has been devoted to determining whether the interaction of gametes between external fertilizers may have established reproductive barriers between groups (Palumbi 1994; Lessios 2011). These proteins consist of the sperm protein bindin and its receptor EBR1 in sea urchins, and the sperm protein Lysin and its egg receptor VERL in abalone. The observation of rapid sequence divergence at these proteins in sea urchins and abalone (Palumbi 1992) has led to the hypothesis that diversifying selection acting at these loci may have facilitated speciation by establishing reproductive barriers via via intrinsic gametic incompatibility, the failure of heterospecific gametes to fertilize (Palumbi 2009; Lessios 2011). Several mathematical models have shown that both allopatric and sympatric speciation are theoretically possible when sexual conflict drives a coevolutionary chase between the sexes, leading to divergence between incipient species (Gavrilets 2000; Van Doorn et al., 2001; Gavrilets and Waxman 2002). Under this scenario, the GRPs involved would be considered speciation genes, defined by Nosil and Schluter (2011) as those genes whose divergence had a large relative effect on reproductive isolation before the completion of speciation. Sexual coevolution of GRPs in different directions could quickly lead to gametic isolation. This process may be driven by antagonistic sexual selection where the optimal mating rates of males and females differ. Such selection would act primarily on the egg receptor protein to avoid polyspermy, which could lead to evolution of the sperm protein to counter egg adaptation. However, other mechanisms are possible (see Palumbi 2009). Since divergence at GRPs could prevent interbreeding between differentiated populations via intrinsic gametic incompatibility, these loci could be responsible for initiating the speciation process and would therefore represent speciation genes.

**Critique of Rapid Speciation by Evolution of Gamete Recognition**

Current reproductive isolating barriers may be misleading about which barriers were historically important during the early stages of speciation as isolating barriers continue to accumulate after the speciation process is complete (Coyne and Orr 2004). Species-specific fertilization and rapid divergence of GRPs is also consistent with the action of reinforcement selection for stronger prezygotic isolation. In the case of reinforcement affecting GRPs in broadcast spawners, selection would likely act on egg receptor in females to prevent eggs from being fertilized by heterospecific sperm. This selection is not expected to act on the recognition protein in males because of the large amount of sperm produced by males (citations). However, the male protein may evolve in response to any changes produced by reinforcement selection acting on the female egg receptor. Allopatric taxa often fail to show diversifying selection at these loci (Vacquier and Swanson 2011), suggesting the action of reinforcement selection in sympatry following general divergence in allopatry. However, there is evidence of intraspecific sexual selection consistent with sexual conflict (Pujolar and Pogson, 2011; Levitan studies), and the behavior of selection at bindin and EBR1 in *Strongylocentrotus* is inconsistent with reinforcement (Pujolar and Pogson 2011). Additionally, alleles causing reinforcement only have a selective advantage within the contact zone between two populations. Selection may be swamped by gene flow from regions where the two species do not overlap. Therefore, it appears that divergence at GRPs has been produced by intraspecific sexual selection or sexual conflict, but it is unclear whether this divergence occurred before speciation was completed, and whether it had a large relative effect on reproductive isolation. GRP divergence may not have had a large effect on reducing gene flow relative to other genes or barriers early in the speciation process. Gametic incompatibilities may have evolved after speciation was complete as GRPs in different groups evolved independently via intraspecific processes.

**Hybridization and Introgression**

Hybridization occurs when reproductive barriers are incomplete and is now known to be much more common than previously thought, especially in animals (Mallet et al. 2016). It has been estimated that hybridization with a closely related species occurs in 25% of flowering plant and 10% of animal species (Mallet 2005). Reticulate evolution in marine systems may be as common as that of non-marine clades (Gardner 1997), but the difficulty in collecting and observing marine organisms has limited detection (Arnold and Fogarty 2009). Historically, hybridization has been thought of as an impeding force to speciation. While secondary contact and high levels of gene flow may prevent may result in hybrid swarm or the weakening of reproductive barriers, often the genetic differences distinguishing populations are maintained. Further differentiation can occur if hybrid fitness is low and there is selection for increased prezygotic isolation (Coyne and Orr, 2004). When hybrids and viable and fertile, backcrossing of hybrids to one or both parent species can lead to the introgression of genetic loci. Introgression has long been recognized as a significant evolutionary force in plants (Anderson and Hubricht 1938; Anderson and Stebbins 1954) but has not been appreciated in animals until recently (Hedrick 2013). Although introgression is now known to be common across the eukaryotic tree of life, it remains unknown how influential introgression has been to the evolutionary trajectories of hybridizing animal species (Taylor and Larson 2019). Introgression need not be problematic for the biological species concept or the study of speciation if we accept Coyne and Orr’s (2004) definition of “good species” as “those that maintain their distinctness in sympatry even if they occasionally hybridize with others.” Species are therefore characterized by considerable, but not necessarily complete, reproductive isolation.

**Strongylocentrotid Family of Urchins**

The sea urchin family Strongylocentrotidae represents a promising system for studying the evolution of reproductive isolation and characterizing patterns of introgression. Sea urchins are representative of many other marine species with their large effective population sizes, broad geographic distributions, and limited population structure. The strongylocentrotid family includes ten species, nine of which have whole-genome sequence data available. The family includes the purple urchin, *Strongylocentrotus purpuratus*, which has a well-annotated reference genome and a long history as a model organism for studies in embryology, development, fertilization, and gene regulation. The genome assembly of *S. purpuratus* (Spur\_5.0) is in its sixth major revision and extensive RNA sequencing has been done to verify gene model predictions (insert citations). The phylogeny of the group, resolved by Kober and Bernardi (2013), is characterized by two major clades and is consistent with a western Pacific common ancestor and two separate eastern Pacific invasions (Kober and Bernardi 2013). The western Pacific taxa include *P. depressus*, *M. nudus*, *H. pulcherrimus*, and *S. intermedius*. Five species co-occur in the eastern Pacific with overlapping geographic ranges, depth preferences, and spawning seasons: *S. droebachiensis*, *S. fragilis*, *S. pallidus*, *S. purpuratus*, *M. franciscanus*. Two of these species, *S. droebachiensis* and *S. pallidus* have expanded their ranges further, crossing the Bering Sea to colonize the Arctic Ocean and both the west and east Atlantic. These two species show little differentiation between Pacific and Atlantic Oceans (Palumbi and Kessing 1991), likely due to stepping-stone populations that facilitate gene flow.

**Urchin Hybridization**

It has been shown that sea urchin species usually don’t develop complete postzygotic isolation until at least five million years of separation (Zigler et al. 2005; Lessios 2007).Although viable hybrids of many species pairs have been produced and successfully backcrossed in the laboratory (Strathmann 1981, list viable crosses with citations), it had long been believed that the strongylocentrotid sea urchin species had not shared genetic material through introgression (Strathmann 1981; Lessios 2007) because hybrids are so rarely observed in nature (Vasseur 1952; Hagström and Lönning 1967). Indeed, sea urchin hybrids are rarely detected, but studies aiming to detect hybrids are rare, especially those incorporating molecular genetic data. Fertilizations between very distantly related taxa are possible in the lab. Taxa that produce viable and fertile hybrids in the lab may not hybridize naturally in the wild. However, natural hybridization may also be underestimated due to the inaccessibility of the marine environment and the fact that hybrids often morphologically resemble on of the two parent species (Strathmann 1981) rather than taking on an intermediate form. Indeed, despite the perceived absence of hybrid individuals in the wild, recent studies have indicated that asymmetric introgression from *S. pallidus* into *S. droebachiensis* has occurred in the Northeast Pacific (Addison and Hart 2005; Harper et al. 2007; Addison and Pogson 2009; Pujolar and Pogson 2011) and Northwest Atlantic (Addison and Hart 2005; Harper et al. 2007), although only a handful of nuclear and mitochondrial loci were analyzed. It remains unknown how common introgression is across the genomes of these two species and others. Hybrids of *S. droebachiensis* and *S. pallidus* resemble *S. pallidus* as larvae and *S. droebachiensis* as adults in morphology. Adult F1 hybrids could therefore easily be mistaken for *S. droebachiensis* if found in the field (Strathmann 1981). Therefore, first generation hybrids could be common but go undetected in the absence of genetic data.

**Introgression and GRP Speciation**

GRPs may be currently important in species recognition and maintaining species boundaries, but they are rarely complete barriers to strongylocentrotid introgression. In many groups, it has been shown that gamete compatibility does not scale with time, measured by COI divergence (Zigler et al., 2005), but rather is correlated with bindin divergence (Zigler and Lessios, 2003b; Zigler et al., 2005). However, in the Strongylocentrotus genus, gamete compatibility appears to scale with divergence time and not gamete compatibility, consistent with divergence at GRPs taking place after significant reproductive isolation had evolved rather than causing reproductive isolation (Palumbi 1992). Gametic compatibility is often asymmetric, which alone cannot prevent gene flow between incipient species, and species boundaries appear to be related to postzygotic barriers rather than intrinsic gametic incompatibilities (Addison and Pogson, 2009).

If the rapid evolution of gamete recognition proteins was important in restricting gene flow early in the speciation process, gene trees of gamete recognition proteins should be reciprocally monophyletic and show concordant gene tree topologies. It would be unexpected to find evidence of introgression between distantly related taxa. Conversely, observing discordant gene trees at GRPs and evidence of widespread introgression in extant species and ancestral lineages would be inconsistent with divergence at gamete recognition proteins causing speciation and would indicate that stronger isolating barriers likely existed early in the speciation process.

Here, we document widespread hybridization and introgression across the strongylocentrotid family, adding representation of marine invertebrates to the growing body of empirical evidence for hybridization and introgression in nature. Our findings are inconsistent with speciation via rapid evolution of gamete recognition proteins gamete recognition proteins in strongylocentrotid urchins. GRPs were likely not early diverging genes establishing reproductive isolation or causing speciation in this group.

**2 Materials and Methods**

2.1 Mitochondria Phylogenetics

All complete mitochondrial genome assemblies publicly available (Table S1) for the family on NCBI were downloaded and aligned with Clustal Omega v1.2.3 (Sievers & Higgins, 2018). A maximum likelihood tree with 1,000 bootstrap replicates was created using iqtree (Nguyen, Schmidt, von Haeseler, & Minh, 2015).

2.2 Whole Genome Resequencing

Raw sequencing reads for each member of the strongylocentrotid family were downloaded from the NCBI Sequence Read Archive in fastq format using fasterq-dump from the SRA toolkit (v2.11.2). Metadata for the downloaded accessions is available in Table S2. Read data included two individuals for both on *S. purpuratus* (SRR6281818, SRR7211988) and *H. pulcherrimus* (DRR107786, SRR5767283), and one individual from each of *S. fragilis* (SRR5767279), *S. droebachiensis* (SRR5767286), *S. pallidus* (SRR5767285), *S. intermedius* (SRR5767280), *M. nudus* (SRR5767281), *M. franciscanus* (SRR5767282), *P. depressus* (SRR5767284), and *P. lividus* (ERR5671699). *P. lividus* was included as a known outgroup to the strongylocentrotid family.

2.3 Data Pre-Processing

Data was pre-processed following [GATK’s Best Practices](https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-) workflow. This workflow is advantageous because the final BAM files produced retain the original sequencing reads and read information while preventing adapter sequences from contributing to reference genome alignment. Additionally, mapped BAM files can be restored back to the original fastq files or unmapped BAM (uBAM) format for re-analysis.

The downloaded fastq files were converted to unmapped BAM (uBAM) format using Picard FastqToSam (gatk4, v4.2.6.0). This was done to facilitate the adapter marking and trimming strategy. FastqToSam converts fastq files to BAM format, interleaving paired-end read files, adding read group information (flowcell ID and lane), and sorting by query name. All metadata from the fastq files are conserved.

Adapters were marked to prevent adapter sequences from impacting read mapping to the reference genome. Each BAM file/sample then had the location of adapter sequences marked using Picard MarkIlluminaAdapters (gatk4, v4.2.6.0). MarkIlluminaAdapters adds adapter trimming tags (XT:i:#) to a read record in the BAM file, in which “#” marks the 5' start position of the identified adapter sequence. A minimum of six bases are required for adapter sequences to be marked, and reads without adapter sequences remain untagged.

Reads with adapter sequences marked were then aligned to the S. purpuratus reference genome (Spur\_5.0) using a command piping Picard SamToFastq, bwa mem, and Pircard MergeBamAlignment [SamToFastq] | [BWA-MEM] | [MergeBamAlignment]. SamToFastq converts the unmapped BAM files with marked adapters back to fastq format, by default change the quality scores of bases marked as adapter sequences to two (phred) to prevent the adapter sequences from contributing to read alignment and scoring. Reads were then aligned to the S. purpuratus reference genome using bwa mem (v0.7.17), with the -M option set to mark shorter split hits as secondary. Finally, Picard MergeBamAlignment merges the raw unmapped BAM file with the mapped BAM file, restoring the original read information and base quality scores, and generates additional meta information based on the output from the aligner.

Duplicates were marked with Picard MarkDuplicates, which adds a sam bitwise flag of 0x0400 to read pairs identified as duplicates. The program also adds a DT tag to the SAM record indicating the type of duplicate, either library/PCR-generated duplicates (LB) or sequencing-platform artifact duplicates (SQ).

Variants were called using gatk HaplotypeCaller. VCF files for each species were merged using gatk CombineGVCFs and joint genotyping was performed using gatk GenotypeGVCFs. All SNPs within 5 base-pairs of an indel were filtered using bcftools filter. SNP and indel records were split into separate files using gatk SelectVariants and hard filtered following gatk recommendations. SNPS were filtered that had low quality (QUAL < 20), low map quality (MQ < 40), low quality by depth scores (QD < 2), high fisher strand scores (FS > 60), high stand odds ratios (SOR > 3), low mapping quality rank sum scores (MQRankSum < -12.5), and low read position rank sum scores (ReadPosRankSum < -8). The final call set had 53,163,120 high quality SNPs. 55% of the original SNPs were retained after filtering (43,241,306 SNPs removed). Indels were filtered that had low quality by depth scores (QD < 2), high fisher strand scores (FS > 200), or low read position rank sum scores (ReadPosRankSum < -20.0). The filtered SNP and indel vcf files were merged back together using Picard MergeVcfs to produce the final call set for downstream analysis.

2.4 Multiple Sequence Alignments

2.5 Phylogenetic Relationships and Concordance Factor Statistics

2.6 Tests for Introgression (D statistic, delta, collapse tests, test hypotheses with phylogenetic networks)

**3 Results**

2.x Mitochondria Phylogenetics

All assemblies were within 80 base pairs in length from one another. The shortest assembly was NC\_024177.1, an S. franciscanus assembly of 15,649 base pairs in length. The longest assembly was NC\_023773.1, a P. depressus assembly of 15,729 base pairs in length. There was variation in assembly length for most of the species. Species tended to not cluster by assembly length, suggesting assembly error, or conversely, a highly conserved mitochondrial genome with length polymorphisms present within and between species. Two *S. droebachiensis* assemblies were 15,710 base pairs, two *S. intermedius* assemblies were 15,705 in length, and two *S. pallidus* assemblies were 15,712. *S. purpuratus* and *S. fragilis* had just one assembly each.

The tree (Figure ?) indicates introgression or mitochondrial capture between two pairs of taxa, *P. depressus* and *M. franciscanus*, and *S. droebachiensis* and *S. pallidus*. Additionally, the positions of *S. purpuratus* and *S. intermedius* are swapped relative to the species tree, consistent with introgression between *S. purpuratus* and one or more of *S. pallidus*, *S. droebachiensis*, *S. fragilis*, the common ancestor of *S. droebachiensis* and *S. fragilis*, or the common ancestor of *S. droebachiensis*, *S. fragilis*, and *S. pallidus.* This is consistent with the low bootstrap support (70) for the placement of *S. purpuratus*. One *S. droebachiensis* accession from Svalbard (EU054306.1) clustered in an *S. pallidus* clade with high bootstrap support, indicating recent introgression from *S. pallidus* into *S. droebachiensis*. The grouping of *S. droebachiensis* and *S. pallidus* as sister taxa with high bootstrap support may indicate historical mitochondrial capture.

**4 Discussion**

**To Incorporate**

* “In many closely related plant species, cross-hybridization in the field appears to be limited by pollen-style interactions that inhibit the initiation of pollen tube growth, or slow it considerably. Such barriers between populations appear to be pleiotropic effects of genetic systems whose real function is to limit selfing (Palumbi 1992).”

**Mitochondrial DNA Introgression**

In the phylogenetic tree produced from alignments of available mitochondrial genome assemblies, *M. franciscanus* and *P. depressus* cluster together with high bootstrap, consistent with introgression between the two and corroborated by significant D statistic. This is unexpected given that S. franciscanus does not have an overlapping distribution with P. depressus, while M. nudus and P. depressus commonly co-occur together in Japan. In the tree, the positions of *S. purpuratus* and *S. intermedius* are swapped, consistent with introgression between S. purpuratus and some or all of *S. droebachiensis*, *S fragilis*, and *S. pallidus*. Introgression may have also occurred between S. purpuratus and the common ancestor of the three, or the common ancestor of *S. droebachiensis* and *S. fragilis*.

*S. droebachiensis* and *S. pallidus* are grouped together as sister taxa with high bootstrap support, indicating that mitochondrial capture may have occurred as the result of ancient introgression. Additionally, *S. droebachiensis* is not monophyletic; a Svalbard *S. droebachiensis* sample (EU054306.1) is more closely related to the *S. pallidus* sequences than the other Norway *S. droebachiensis* sample (NC\_009940.1). Interestingly, the geographically distant Friday Harbor and Norway S. droebachiensis samples cluster together, indicating that there may have been recent introgression of the S. pallidus mitochondrial genome into a Svalbard population of *S. droebachiensis*, or that this individual was a hybrid or backcross individual.

Mitochondrial and chloroplast DNA have been shown to introgress more readily than nuclear DNA (Coyne and Orr, 2004, p. 471). This may be because organelle genes mostly perform internal functions and are not involved in local adaptation to environmental differences, thereby making organellar genomes function fine in the genomic background of a close relative. Due to the lack of recombination in mitochondrial DNA, introgression may produce trees that appear paraphyletic even when the included species are not (Coyne and Orr, 2004).

**Urchin Ecology/Behavior**

How do species originate in allopatry without any marked changes in ecology, physiology, or morphology? Although it appears that the Strongylocentrotid sea urchin species lack significant differences in ecology, these differences may exist and may be undetectable aspects of resource use. Additionally, speciation events within the family need not proceed in the same way. Different pairs of taxa may have evolved different modes of reproductive isolation. The ranges that we see today may not provide any useful biogeographic information about the history of speciation (Coyne and Orr 2004, p. 118) given the old divergence times of this group.

**Reproductive Isolating Barriers**

Reproductive isolating barriers are required for producing and maintaining distinct groups in sympatry, but the barriers that currently maintain species boundaries may not have been historically important during speciation and are not necessarily the barriers that initially reduced gene flow or produced the species. When multiple reproductive barriers are present, the current relative importance or strength of different barriers may not be consistent with their historical importance. In trying to understand how new species arise, we want to determine which barriers were involved in the initial reduction of gene flow, that is, before reproductive isolation became complete, or nearly so. Problematically, gene flow is almost always prevented by more than one barrier (Coyne and Orr, 2004), and new reproductive isolating barriers continue to accumulate even after gene flow between taxa has ceased and speciation is considered complete.

Since reproductive barriers act sequentially over an organism’s lifetime, habitat isolation and temporal isolation could have a potentially stronger relative effect than gametic isolation if they were present. Even if gametic isolation has high absolute strength, its relative effect maybe be negligible if differences in microhabitat preference and spawning times prevent most heterospecific fertilizations. Both habitat preference and spawning time commonly overlap between sympatric species of sea urchins at present time. However, congeneric species of sea urchins often show depth zonation in areas of range overlap (Lessios 2007), and sperm in dilution ages rapidly, reducing the likelihood of encountering eggs when heterospecific individuals are not nearby (Lessios 2007; Pennington, 1985; Levitan, 1995). High fertilization rates may depend on hundreds of thousands of sperm per egg (Levitan et al., 1991), and it has been estimated that only 1% of the egg surface is fertilizable (Vogel et al., 1982).

Additionally, because reproductive barriers occur sequentially over the lifespan of the organism, pre-zygotic barriers will almost always be the current strongest barriers in good species, even in cases where postzygotic barriers were more important early in the process of speciation. Therefore, using current barriers to gene flow to reconstruct speciation history may lead to spurious results. Given that hybrids can be produced and backcrossed in the lab, there are likely additional barriers to gametic isolation, but due to the rarity of studies documenting natural hybridization, it is nearly impossible to evaluate extrinsic postzygotic isolation. Extrinsic postzygotic isolation in the form of ecological isolation could be important early on in speciation if offspring of crosses between individuals of differentiated populations are ecologically intermediate and suffer from reduced fitness. Adaptation to different communities of pathogens through different pathways could lead to negative pleiotropic effects in hybrids. Intrinsic postzygotic isolation is less likely to have been important during the early stages of speciation because genomes from different species still appear compatible in laboratory crosses. Hybrids tend to develop normally and even demonstrate hybrid vigor in some cases (citations). However, intrinsic postzygotic isolation sometimes does not appear until generations beyond the F1 if the alleles that cause intrinsic postzygotic isolation are partially recessive in hybrids (Coyne and Orr 2004, p. 68-69).

**Formation of Hybrids in Laboratory Crosses**

In studies of development, reciprocal crosses have been performed in the laboratory between the sand dollar *Dendraster excentricus* and both *S. purpuratus* and *S. franciscanus* (Moore 1933; Moore 1957; Brookbank 1969; Fujisawa 1993). Hybrid embryos of the cross between *D. excentricus* eggs and *S. purpuratus* sperm develop normally through gastrulation but then become abnormal and die within a few days (Flickinger 1957). However, embryos develop completely normally in the cross between *S. purpuratus* eggs and *D. excentricus* sperm and were kept as long as 12 days after fertilization, in which time they were healthy plutei (Flickinger 1957). Unfortunately, experimental crosses performed in the 20th century did not attempt to rear hybrid embryos past 12 days (Williamson and Boerboom 2012). Sea urchins and sand dollars fall under the same sub-class, Euechinoidea, but belong to different superorders. *Dendraster excentricus* belongs to Gnathostomata, while *Strongylocentrotus purpuratus* and *S. franciscanus* belong to Echinacea. Moore (1957) points out that the interordinal cross is more successful than the cross between *S. franciscanus* egg and *S. purpuratus* sperm.

*Heliocidaris crassispina* and *S. intermedius* hybrids are viable and have higher thermal tolerance than S. intermedius (Zhao 2021; Ding et al., 2007). Additionally, *S. intermedius* x *S. pulcherrimus* hybrids are used in aquaculture.

Another example involves the cross between an *S. purpuratus* egg and *M. franciscanus* sperm, which often results in vigorous larvae that often outlive those of conspecific crosses (Newman 1923). However, nonviable embryos are also produced in this cross, demonstrating both hybrid vigor and hybrid weakness. The reciprocal cross between *S. purpuratus* sperm and *M. franciscanus* egg results in abnormal development, and few embryos make it past the gastrula stage. Newman (1923) attributes the differences in developmental success to differences in the sizes of the gametes of the two species. The smaller *S. purpuratus sperm* under stimulates the *M. franciscanus* egg, while in the reciprocal cross, the larger *M. franciscanus* sperm overstimulates the smaller *S. purpuratus* egg. At the time of this study, the phylogenetic relationships of the family had not been resolved and the purple and red urchin were thought to be members of the same genus. However, this species pair diverged about 13-19 million years ago.

**Droe-pal case study**

* Hagstrom & Lonning (1967) report no obvious hybrids, and conclude that barriers to cross fertilization and chromosomal incompatibility of hybrids separate the species.

Strathmann (1981) studied barriers between *S. droebachiensis* and *S. pallidus* in the San Juan Islands, finding that laboratory crosses were viable in both directions, but that fertilization success was highly asymmetrical. *S. droebachiensis* eggs were much more susceptible to heterospecific crosses than *S.* pallidus eggs. From ten separate crosses (10 separate reciprocal crosses) from each direction, only four hybrids survived to the three-year mark when spawning was induced, and all were female. The hybrid females were able to be backcrossed with both pure *S. pallidus* and *S. droebachiensis*, although fertilization success was much higher in the F1 egg x *S. pallidus* sperm cross. Unfortunately, male hybrids were not tested.

Strathmann (1981) concluded that despite the ability to the species cross fertilize, the two species remain distinct over their shared range, and that introgression does not seem to be occurring. However, F1 hybrids and backcrosses may be overlooked because of the difficulty in distinguishing them by morphology (Strathmann 1981) and no genetic tests were performed. Based on the appearance of early-stage larvae of both species in single plankton tow samples, it was concluded that spawning times for the two species could be within a day of each other, although Strathmann (1981) indicates that he had not observed simultaneous spawning of mixed populations of the two species. Strathmann (1981) concluded that if a strong barrier to gene flow existed in the San Juan Islands, it would likely be small differences in space or time during spawning or substantially lower fitness of hybrids in nature relative to in the lab. The apparent weakness of prezygotic barriers such as habitat isolation, temporal isolation, and gametic incompatibility may emphasize the importance of postzygotic isolation from reduced fecundity or viability of hybrids in nature. However, not enough is known about microhabitat preference and spawning time across the two species’ ranges to rules these barriers out.

Although the two species are commonly seen together at depth, as depth increases, food becomes limiting and gonad size decreases. Individuals at high depth may not be spawning.

**Misc**

The youngest sister taxa in the *Strongylocentrotidae* family are *S. droebachiensis* and *S. fragilis*. If EBR1 was a speciation gene, we would expect to see a concordant gene tree at this locus. However, *S. droebachiensis* groups with *H. pulcherrimus* with high bootstrap support. Regardless of whether this discordance was produced by incomplete lineage sorting or introgression, this discordant pattern would be unexpected if EBR1 was a barrier locus. If speciation has proceeded by rapid evolution of gametic incompatibility, *S. droebachiensis* GRP alleles should be divergent from distantly related taxa and should not be compatible with ancestral alleles. Additionally, genes affecting reproductive isolation do not flow readily between species and should reflect species boundaries (Nosil and Schluter 2011). Each subsequent speciation event should act to increase gametic incompatibility with the ancestral state. If speciation of *S. pallidus* and the common ancestor of *S. droebachiensis* and *S. fragilis* was caused by divergence at GRPs, then *S. pallidus* should have remained isolated from the ancestral lineage and both descendants.

Given that GRPs are often evolving non-neutrally, incomplete lineage sorting of EBR1 alleles is unlikely, making introgression of EBR1 more plausible and implying that EBR1 is not a speciation gene or barrier locus.

It is peculiar that *S. droebachiensis* is the only taxa out of place in the tree. This may be because there has been sufficient time in the older speciation events for EBR1 to diverge and become reciprocally monophyletic. In the speciation event producing *S. droebachiensis* and *S. fragilis*, EBR1 could have only been an important speciation gene if adaptive introgression of EBR1 from *H. pulcherrimus* into a population of *S. droebachiensis* caused reproductive isolation and initiated speciation.

Reproductive isolation between *S. pallidus* and *S. droebachiensis* is incomplete, but the species remain distinct across their overlapping ranges, emphasizing the importance of postzygotic isolating mechanisms.

Sympatric speciation or speciation with gene flow cannot explain the observed patterns of introgression, as the phylogeny is resolved, and only non-sister taxa were tested for introgression. It is not possible to test for introgression between sister taxa with only a single genome per species. S*. droebachiensis* and *S. pallidus* are separated by two independent speciation events. If gamete compatibility scales with time only, it is likely that there has been introgression between sister taxa as well. Significant introgression between sister taxa would increase the proportion of concordant gene trees, thereby limiting the signal of introgression between nonsister taxa.

The two youngest speciation events that produced *S. droebachiensis* and *S. fragilis*, and *M. franciscanus* and *M. nudus* both feature discordant topolgoies at GRP loci. In the bindin tree, *M. nudus* groups with *P. depressus*. In EBR1, *S. droebachiensis* groups with divergent taxa.

**Echinometra**

In the *Echinometra* genus, there are several young species with strong gametic incompatibility. These species also show some habitat and temporal isolation; currently there is some degree of overlap in both habitat and spawning time. Therefore, gametic incompatibility is currently the strongest barrier, but it is not possible to rule out that habitat and temporal isolation were the barriers that initially reduced gene flow, and that the gametic incompatibility arose because of reinforcement. Indeed, there is strong evidence that reinforcement selection is occurring at gamete recognition proteins in this system, evidenced by reproductive character displacement in *E. oblonga* (Palumbi 2003). Sympatric populations of *E. Oblonga* and *E.s p. C* have more divergent bindin sequences than allopatric populations, and some allopatric populations shared alleles. Coyne and Orr (2004) point out that it is hard to see the selective advantage for reducing sperm binding in sympatry as a reduced ability to hybridize may lead to reduced fitness when hybrid fertilizations are viable and fertile. However, reinforcement could be acting on the egg surface proteins and drive bindin divergence through coevolution. In this case, range expansion following speciation could be masking the importance of habitat isolation during speciation.

**Conclusions and Further Work**

The long persistence of gametic compatibility between divergent taxa, phylogenetic discordance at both bindin and EBR1, and evidence of extensive introgression within the family is inconsistent with speciation via rapid evolution of gamete recognition proteins. If gametic incompatibilities did not prevent introgression following speciation, it is unlikely that this barrier alone would have reduced gene flow enough for speciation to occur. The fact that species boundaries have been maintained despite incomplete gametic incompatibilities emphasizes the importance of postzygotic isolation. Indeed, most identified speciation genes in animals are involved in postzygotic isolation rather than prezygotic isolation (Lessios 2011 via other sources.). More focus should also be given to other prezygotic isolating barriers such as habitat and temporal isolation. A potential test of the likelihood of gametic isolation in establishing reproductive barriers would be to look for allopatric populations of single species that show population-specific gamete preference. In addition, future surveys should look for hybrids using molecular diagnostics as hybrids may not be distinguishable by appearance.

The plausibility of rapid speciation via evolution of GRPs has been bolstered by the misconception that fertilization is a species-specific lock and key system, and that introgression has not occurred in this group. This may be true of other sea urchin genera such as *Echinometra*. Speciation via rapid evolution of GRPs is theoretically possible and may have occurred in other groups, but we find it unlikely that it has been the predominant mode of speciation within *Strongylocentrotidae*. This theory may receive undue attention because gamete recognition involves few genes. Other isolating barriers such as temporal isolation, microspatial habitat isolation, or postzygotic isolation likely evolved earlier and were more important in establishing reproductive isolation. Progress in understanding speciation and reproductive isolation in broadcast spawning marine invertebrates may be hindered by trying to infer one common mode of speciation across biologically different groups and different time periods. One pattern of speciation might not hold across different genera or independent speciation events.

**Acknowledgements**

UCSC Hummingbird HPC cluster for storage and computation.

**References** (single spaced)

**Data Accessibility and Benefit-Sharing**

All code used to produce the data and results in the paper is available at <https://github.com/matthewglasenapp/dissertation>.

**Data Accessibility Statement**

**Benefit-Sharing Statement**

**Author Contributions**

(Designed research, performed research, contributed new reagents or analytical tools, analyzed data, wrote paper)

**Conflicts of Interest**

**Tables and Figures** (with captions). Must be editable files. Single spacing for table and figure captions