Received: 22 April 2017 | Accepted: 2 May 2017

DOI: 10.1111/mec.14170

NEWS AND VIEWS

Perspective

WILEY MOLECULAR ECOLOGY

Finding genes and lineages under selection in speciation

Michael W. Hart | Vanessa Guerra

Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada

Correspondence
Michael W. Hart, Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada.
Email: mwhart@sfu.ca

What are the genes and traits that respond to selection and cause prezygotic reproductive isolation between species? This question has been hard to answer because genomes are large, the targets of selection may be scattered across the genome (Sabeti et al., 2007) and different genes may respond to the same selective pressure in different populations (Scheinfeldt et al., 2012). In this issue of *Molecular Ecology*, Weber et al. (2017) use a clever comparative approach and leading-edge transcriptomic methods to identify the species and genes under positive selection for divergence between brittle stars (the echinoderm class Ophiuroidea) in the *Ophioderma longicauda* species complex. They found convincing evidence of positive or diversifying selection acting on two genes encoding ion channels that form part of the signal transduction cascade within the sperm in response to pheromones. Evidence for selection was concentrated in genes from one species (called C5, with internal fertilization and female parental care of brooded juveniles and not in the other species (called C3, with more conventional broadcast spawning and planktonic development of embryos and larvae). That analysis greatly extends the range of taxa, life history traits and molecules that are associated with positive selection in speciation. It also illustrates some of the current limitations on the application of RNAseq methods in the search for the targets of selection in nonmodel organisms like brittle stars. From both points of view, the new work by Weber et al. (2017) has important implications for our understanding of speciation in the ocean.

KEYWORDS
fertilization, positive selection, RNAseq, speciation

The choice of organisms such as brittle stars (Figure 1) for that study may seem idiosyncratic, but is in fact key to many recent advances in this area of research on the evolution of reproductive isolation. The complex mating systems of vertebrates, arthropods and animal-pollinated plants include traits such as mate attraction, courtship, competition among males or female choice of mates, and the response to selection acting on those complex phenotypes might depend on hundreds or thousands of loci. By contrast, brittle stars and many other marine invertebrates lack complex behavioural interactions between males and females and instead have simple mating systems in which the reproductive compatibility between mates (and prezygotic isolation between species) depends on a few proteins involved in sperm chemoattraction to the egg, sperm binding to the egg coat, lysis of the egg coat and fusion of the gametes (Hirohashi et al., 2008). In these taxa, the search for genes that can account for the evolution of premating isolation under selection may be greatly simplified by focusing mainly on the analysis of genes expressed in gonads or gametes that mediate sperm–egg interactions during spawning and fertilization (Lessios, 2011).

The study of those molecular mating system traits and their response to selection in speciation has benefitted from two critically important methodological advances. First, for a few model organisms scattered across the animal tree of life (including some echinoderm species), the identity and expression of genes encoding gamete traits have been well studied by biochemists and cell biologists. Second, the advent of inexpensive, high-throughput DNA sequencing from

mRNA samples (called RNAseq) has democratized the analysis of sequence and expression variation to include organisms for which no genomic resources exist. Together those two advances have allowed molecular ecologists to extend genome-scale analyses to previously understudied lineages (such as *O. longicauda*) by *de novo* computational assembly of short sequence reads from RNAseq data, annotation of the assembled genes by sequence similarity to experimentally annotated genes of other organisms, comparison of orthologous genes from different individuals, populations or species, and analysis of their molecular evolution for signals of a response to selection on the expression or function of the gene product.

Weber et al. (2017) applied these tools to great effect. They found more than ten thousand orthologs (genes expressed in testes or ovaries of both species C3 and C5) and could annotate about half of those genes to infer gene function. This is remarkably successful considering that most annotations for echinoderm genes are based on sequence similarity to genes in the purple sea urchin, *Strongylocentrotus purpuratus*, which shares a most recent common ancestor with *O. longicauda* more than 400 million years ago.

Analysing thousands of genes for a signal of positive selection between species C3 and C5 is daunting. The most powerful methods for such analysis use models of codon evolution that compare the relative rate of silent or synonymous nucleotide changes (invisible to selection) to the rate of nonsynonymous changes that cause amino acid substitution (subject to selection; Cannarozzi & Schneider, 2012). If nonsynonymous rates of change are observed to be higher than rates of synonymous change, then some evolutionary process (i.e., selection) must have acted on nonsynonymous changes in order to cause those changes to accumulate over time at a rate higher than synonymous changes (if both types of mutation arise at the same rate). When such positive selection is associated with specific genes or protein functions, or with species that have particular phenotypic traits, then a plausible argument can be made for selection acting on those functions or traits (and not on other traits). However, fitting codon models to long alignments that include many sequences is computationally slow and requires careful consideration of model parameters for each gene and alignment. Manual curation of that process for thousands of genes is prohibitively slow. Weber et al. (2017) used a new, rapid method called PSGfinder to screen thousands of gene alignments to identify those with a relatively high frequency of nonsynonymous differences between species C3 and C5. They then applied codon model analyses to that subset of candidate genes that might show evidence of positive selection. The result was unexpected and clear: species C5 (in which internal fertilization has evolved) showed evidence of positive selection for amino acid substitutions in sperm-expressed genes that encode the sodium–proton exchanger (NHE) and the tetrameric potassium-selective cyclic nucleotide-gated channel (TetraKCNG; Figure 2).

Because the NHE and TetraKCNG proteins form part of the physiological response of sperm to the egg peptide pheromone (called speract in sea urchins, or asterosap in sea stars), Weber et al. (2017) concluded that selection on gamete interactions in

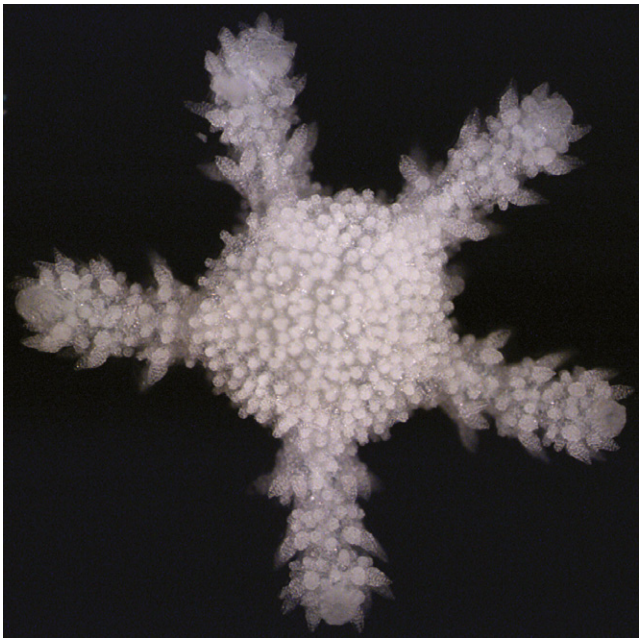

FIGURE 1 Juvenile brittle star, *Ophioderma longicauda* C5, from a population on the island of Crete. Photograph courtesy of Alexandra Weber

FIGURE 2 Predicted tertiary structure for part of the tetraKCNG sperm protein in *Ophioderma longicauda* C5. Most of the residues inferred to show positive selection are found in the series of three alpha helices (yellow, green) in the foreground. The structural model was estimated in SWISS-MODEL (Biasini et al., 2014) based on the amino acid sequence shown in Figure 2 of Weber et al. (2017); the ribbon image was rendered in Chimera (Pettersen et al., 2004)

Ophioderma longicauda acts on that early stage of the fertilization process. They also proposed that sperm chemoattraction to eggs may be the locus of selection for species specificity of fertilization among brittle stars in general. That result is surprising and novel because those genes and forms of gamete interaction have not previously been implicated in the evolution of reproductive isolation or speciation in echinoderms. That result is particularly interesting because NHE and TetraKCNG do not interact directly with eggs or pheromones but indirectly via a pheromone receptor in the sperm tail that did not show evidence of positive selection in codon model analyses. Because the evidence for selection on *NHE* and *TetraKCNG* was found in species C5 but not species C3, the authors also concluded that some evolutionarily derived feature of the fertilization ecology of that species (perhaps associated with sperm–egg interactions in the female brood chamber) may be the cause of selection on NHE and TetraKCNG function. That result is also surprising and exciting because other studies have typically shown stronger evidence for selection acting on gamete-recognition genes in echinoderms with broadcast spawning and planktonic fertilization (e.g., Patiño et al., 2016). Finally, because genes associated with other stages of sperm–egg interaction (including genes involved in sperm–egg binding) did not show evidence of a response to selection (in either species C3 or C5), the authors tentatively concluded that selection may not act on those traits and genes in *O. longicauda*. That result is notable because genes encoding the sperm acrosomal protein bindin and the egg bindin receptors evolve under positive selection in sea urchins and sea stars (Patiño et al., 2016; Stapper, Beerlie, & Levitan, 2015).

Those results point directly towards future areas of research on speciation in the *O. longicauda* species complex and make predictions about those future discoveries in brittle stars and other organisms. Sperm swimming speed or turning behaviour (or other sperm phenotypic traits) should show larger responses to pheromones from conspecific eggs (and reproductive isolation between species C3 and C5), and those responses should lead to higher conspecific fertilization success. Those effects are expected to be associated with amino acid substitutions at positively selected sites in *NHE* and *TetraKCNG* but not associated with other genetic polymorphisms, and those positively selected sites should co-evolve with polymorphisms in the gene encoding the egg coat pheromones (e.g., Stapper et al., 2015).

It is less clear whether those future studies should extend to variation in other sperm–egg recognition genes. Weber et al. (2017) concluded that other sperm–egg recognition genes do not show evidence of a response to selection, but the evidence may not be conclusive. Analysis of genes assembled from short-read RNAseq data has several limitations that have often been difficult to overcome, and some of those limitations are evident in the RNAseq assemblies from *O. longicauda*. First, some genes that encode repetitive protein domains may be difficult to assemble from short-read data if the repeat units are longer than the sequence reads, or if nucleotide sequences of repeats are strongly similar to each other. Many genes that encode cell surface recognition factors have such repetitive structures, including the sea star gene for asterosap; this repetitive structure may account for the absence of an asterosap ortholog in the ovary transcriptomes of *O. longicauda*.

Second, some genes (such as the contigs identified as *bindin* and the egg bindin receptor *EBR1* in *O. longicauda*) may be assembled as partial coding sequences rather than as complete genes. Analyses of partial coding sequences may be problematic if only the most evolutionarily conservative parts of the coding sequence can be assembled and annotated, or if the partial coding sequence is similar to other genes in the genome. Weber et al. (2017) noted that data for *bindin* in *O. longicauda* were limited to short sequences from the carboxyl end of the predicted protein, which encodes a highly conserved binding-site ligand, and did not include the complex repetitive region at the amino end of the protein that evolves under positive selection in sea stars (Patiño et al., 2016). Whether that repetitive region of *bindin* evolves under positive selection in brittle stars is unknown.

Third, the identification and annotation of partial coding sequences can be misled by apparently strong nucleotide sequence similarity to other parts of the genome. The partial coding sequence for the contig identified as *EBR1* in *O. longicauda* is a strong match to part of the complete *EBR1* gene in one sea urchin species (*Strongylocentrotus purpuratus*), but not to the complete *EBR1* gene in a second sea urchin species (*Mesocentrotus franciscanus*; Kamei & Glabe, 2003). The two sea urchin *EBR1* genes differ in the occurrence of a series of HYR-like repeats in *S. purpuratus* but not *M. franciscanus*; sequence similarity between the brittle star and *S. purpuratus* genes is limited to those shared HYR-like repeats; the brittle star gene lacks the series of paired CUB-TSP1 domains called "core EBR repeats" by Kamei and Glabe (2003) that define sea urchin and sea star *EBR1* genes; and instead the brittle star partial coding sequence is made up of several tandem series of HYR- and EGF-like domains, unlike other echinoderm *EBR1* genes. Those features suggest that the contig called *EBR1* in *O. longicauda*, and shown not to exhibit evidence of positive selection between species C3 and C5, is not an ortholog of other echinoderm *EBR1* genes and instead encodes some other protein (possibly hyalin). Like the results for the partial coding sequence of *bindin* in *O. longicauda*, that result suggests that the question whether *EBR1* evolves under positive selection in brittle stars (and co-evolves with *bindin*) is still unanswered.

Those difficulties in assembling and annotating other classes of gamete-recognition genes in *O. longicauda* do not represent errors by Weber et al. (2017), rather they exemplify some obstacles still to be overcome in the effort to understand which genes, species and modes of gamete interaction show evidence of a response to selection during speciation among nonmodel organisms. We face many of the same obstacles in our work on gamete-recognition genes assembled from RNAseq data for sea stars (e.g., Hart & Foster, 2013). Reduced error rates for long-read data, and improvements in assembler performance for repetitive regions of the genome, promise to overcome some of those limitations in the near future. The increased taxonomic and phylogenetic diversity of sequenced and annotated genomes will also greatly improve the quality and reliability of gene

identifications in *de novo* assemblies. The prospect of overcoming those obstacles should be a cause for cautious optimism: more molecular ecologists should consider following the lead of Weber et al. (2017) in using RNAseq data and *de novo* assembly methods to analyse speciation in groups like brittle stars that have compelling biological stories to tell but few genomic tools to tell them.

AUTHORS' CONTRIBUTION

M.W.H. and V.G. wrote this perspective.

REFERENCES

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., ... Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*, W252–W258.

Cannarozzi, G. M., & Schneider, A. (Eds.) (2012). *Codon Evolution*. Oxford, UK: Oxford University Press.

Hart, M. W., & Foster, A. (2013). Highly expressed genes in gonads of the bat star Patiria miniata: Gene ontology, expression differences, and gamete recognition loci. *Invertebrate Biology*, *132*, 241–250.

Hirohashi, N., Kamei, N., Kubo, H., Sawada, H., Matsumoto, M., & Hoshi, M. (2008). Egg and sperm recognition systems during fertilization. *Development Growth & Differentiation*, *50*, S221–S238.

Kamei, N., & Glabe, C. G. (2003). The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes and Development*, *17*, 2502–2507.

Lessios, H. A. (2011). Speciation genes in free-spawning marine invertebrates. *Integrative and Comparative Biology*, *51*, 456–465.

Patiño, S., Keever, C. C., Sunday, J. M., Popovic, I., Byrne, M., & Hart, M. W. (2016). Sperm bindin divergence under sexual selection and concerted evolution in sea stars. *Molecular Biology and Evolution*, *33*, 1988–2001.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605–1612.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*, 914–919.

Scheinfeldt, L. B., Soi, S., Thompson, S., Ranciaro, A., Woldesmeskel, D., Beggs, W., ... Tishkoff, S. A. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biology*, *13*, R1.

Stapper, A. P., Beerlie, P., & Levitan, D. R. (2015). Assortative mating drives linkage disequilibrium between sperm and egg recognition protein loci in the sea urchin *Strongylocentrotus purpuratus*. *Molecular Biology and Evolution*, *32*, 859–870.

Weber, A. A.-T., Abi-Rached, L., Galtier, N., Bernard, A., Montoya- Burgos, J. I., & Chenuil, A. (2017). Positive selection on sperm ion channels in a brooding brittle star: Consequence of life-history traits evolution. *Molecular Ecology*, *26*, 3744–3759.

How to cite this article: Hart MW, Guerra V. Finding genes and lineages under selection in speciation. *Mol Ecol*. 2017;26:3587–3590. https://doi.org/10.1111/mec.14170