



# Peer Community In Evolutionary Biology

## Genomic parallelism in adaptation to orthogonal environments in sea horses

**Yaniv Brandvain** based on peer reviews by 3 anonymous reviewers

Florentine Riquet, Cathy Liautard-Haag, Lucy Woodall, Carmen Bouza, Patrick Louisy, Bojan Hamer, Francisco Otero-Ferrer, Philippe Aublanc, Vickie Béduneau, Olivier Briard, Tahani El Ayari, Sandra Hochscheid, Khalid Belkhir, Sophie Arnaud-Haond, Pierre-Alexandre Gagnaire, Nicolas Bierne (2018) Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.1101/161786>

Submitted: 11 July 2017, Recommended: 07 September 2018

**Cite this recommendation as:**

Brandvain, Y. (2018) Genomic parallelism in adaptation to orthogonal environments in sea horses. *Peer Community in Evolutionary Biology*, 100056. [10.24072/pci.evolbiol.100056](https://doi.org/10.24072/pci.evolbiol.100056)

Published: 07 September 2018

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

---

Studies in speciation genomics have revealed that gene flow is quite common, and that despite this, species can maintain their distinct environmental adaptations. Although researchers are still elucidating the genomic mechanisms by which species maintain their adaptations in the face of gene flow, this often appears to involve few diverged genomic regions in otherwise largely undifferentiated genomes. In this preprint [1], Riquet and colleagues investigate the genetic structuring and patterns of parallel evolution in the long-snouted seahorse. Before investigating specific SNPs plausibly associated with adaptation, the authors first describe genome-wide population structure in the long-snouted seahorse. This species is split into five phenotypically similar, but genetically distinct populations. Two populations reside in the Atlantic Ocean and are geographically structured with one north of the Iberian peninsula and the other around the Iberian peninsula. Two other populations are found in the Mediterranean Sea and are structured by the environment as they correspond to marine and lagoon environments. The genetic clustering of lagoon populations in the Mediterranean, despite the substantial geographic distance between them is quite impressive, and worthy of further study. Finally, a fifth population resides in a lagoon-like habitat in the Black Sea. The authors then investigate patterns of extreme genomic differentiation among populations, and uncover a remarkable pattern of parallel differentiation in these populations. In an outlier scan, Riquet and colleagues find numerous SNPs in one genomic region that separates northern and southern Atlantic populations. Quiet surprisingly, this same genomic region appears to differentiate populations living in marine and lagoon habitats in the Mediterranean. The idea that parallel

patterns of genomic differentiation may underlie adaptation to differing environmental scenarios has not yet received much attention. This paper should change that. This paper is particularly impressive in that the authors uncovered this intriguing pattern with under three hundred SNPs. Future genome scale studies will uncover the genomic basis behind this unusual case of parallelism.

### **References:**

[1] Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Béduneau, V., Briard, O., El Ayari, T., Hochscheid, S. Belkhir, K., Arnaud-Haond, S., Gagnaire, P.-A., Bierne, N. (2018). Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. bioRxiv, 161786, ver. 4 recommended and peer-reviewed by PCI Evol Biol. doi: [10.1101/161786](<https://dx.doi.org/10.1101/161786>)

## **Reviews**

### **Evaluation round #1**

DOI or URL of the preprint: [10.1101/161786](https://doi.org/10.1101/161786)

Version of the preprint: 1

### **Authors' reply, 13 July 2018**

[Download author's reply](#)

[Download tracked changes file](#)

### **Decision by Yaniv Brandvain, posted 13 July 2018**

#### **Revise**

Thank you for submitting your preprint to PCI for evaluation. This is a fun process and I hope it improves your manuscript!

I have received feedback from four reviewers who have all found this manuscript to be interesting, but who have all agreed that substantial revision would dramatically improve this manuscript, and would make it publication quality. The reviews should come along with this note. Here is my high level summary and synthesis of their comments plus some of my own.

A major theme is that reviewers (particularly reviewers one and two) were impressed by the care in your analysis of such few markers, however they worried about potential alternative explanations of the data (see comments by reviewers one and three), and hoped for more markers. I agree with the reviewers that there are numerous explanations for the observed results (including strong drift in populations with low diversity, which may make outlier tests less reliable). For example are the treemix results evidence of selection on a few loci or introgression? I think that the current data are ambivalent. I think this paper could be better framed as preliminary results which highlight that this system deserves more attention.

Additionally, reviewer three hoped for a more question driven approach to this system. I am of two minds on this comment. First I agree with the reviewer that hypothesis driven science is quite powerful and makes for compelling writing, however it also feels to me that this manuscript is largely descriptive / exploratory. So I suggest splitting the difference and set up reasonable hypotheses when possible, but don't bend over backwards to make this paper into something it is not. I do agree with the reviewer that this can come across as the 'kitchen sink' so either justify why the difference in the information contained in eg STRUCTURE, PCA, or

trees is important to the major argument in the manuscript or be choosy about which belongs in the main text and move redundant analyses to the supp.

Finally, both myself and reviewer four found the paper difficult to follow at times. I suggest rewriting for clarity whenever possible, and (for example) making it clear who the Mediterranean lagoon populations are (I think triangles in Figure 1 meant lagoon, however I could not find this explicitly stated). A less severe version of this is shortening the joint site frequency spectrum to JSFS on line 182 but only defining it as such on line 302. Additionally the manuscript is littered with typos. A healthy dose of editing for typos, clarity of biological questions, and presentation of ideas would drastically improve this paper.

I hope to see a revised version of this manuscript shortly, and I hope that this process improves your paper and the publication process.

Best Yaniv

---

I also add an additional review I received that was not conducted via the PCI reviewer system (reviewer 4) Important contribution of the papers: Broadly, interested in how divergent lineages end up as species. People either think it is natural selection or incompatibilities and that can correlate with clines or mosaics in hybrid zones. Really cool and interesting questions the authors aim to address, but have some concerns.

Reviewer comments Major comments: Story should be linear, and the figures should follow, however throughout the manuscript it is difficult to follow the main story as it jumps around substantially.

Questions: Whole argument about the clines was extremely contrived and confusing to follow.

Difficult to follow the questions addressed in the manuscript

Need more biological background of why focus on the seahorses, for example. Broadly need more background.

Broadly it is difficult to follow the introduction, and the main questions the authors are interested in addressing in the manuscript. Currently it seems that the authors are interested in identifying the genetic variation association with locally adapting to the diverse environments (Lagoons and the sea for examples). I believe that starting with the broad and natural selection/genomic incompatibilities and the "mosaics" and "clines" makes this paper difficult to follow.

Minor comments: Line 124: Misspelled paratrically

Line 124: Would use a different word other than "Patchily" throughout the manuscript as it is difficult to follow.

Line 140: Need to explain in more detail where they get the samples from in the aquaria.

Line 155: Need to state what the reference transcriptome is that was aligned to, explain in more detail the paper that was referenced.

Concern with SNPs measurements, weird ascertainment issues. When developing SNPs from some transcriptomes you are likely to get more common SNPs, which would affect the structure. For example, it may be more common to get SNPs in the lagoon population.

After Line 168: Need to add a sentence about genotyping, we are guessing golden-gate assay was performed? If so I would expand on this in the main manuscript.

Line 169-170: Clever! This is a useful measurement comparing this to other analyses, and this is not done in previous studies.

Figure S11: Label axis standard, it is difficult to follow what is going on the supplemental figure. Also how did you obtained derived allele frequencies? It was not clear.

Line 179: Oriented = Polarized?

Figure 1: need to define the difference between circle and triangles

Figure 1: remove the little lines where you overlay a figure.

Figure 1 legend: need to include what each color corresponds to; also need to address what yellow corresponds to. Also color choice is not easy to interpret in black/white.

Figure 2C: Need to add axis or add correlations between Figure 2B-2C as it is difficult to follow the big points.

Figure 4: Difficult to follow what the authors are trying to portray in figure 4C.

Figure 5: Difficult to follow the main point of the Figure the authors are trying to convey.

Ordering of the Figures difficult to follow the main story. Should follow a general pattern. Something like Map of study populations first and then the PCA then the structure plot. Also what is the main point of Figure 2C? I would not even include this in the final paper.

Figure 1: Was population 6 dropped from the trees? Also difficult to follow where 20/24 samples were obtained?

## **Reviewed by anonymous reviewer 2, 14 November 2017**

The authors have presented some interesting data from a panel of 286 SNPs genotyped in individuals sampled from around the North Atlantic and Mediterranean. The data show some pronounced and intriguing patterns that indicate biologically interesting discordances and heterogeneous genomic patterns of divergence. I am generally quite skeptical of studies trying to find signatures of local adaptation and/or reproductive isolation with only a few hundred markers, but in this case the patterns of differentiation are quite clear, and are unlikely to be artefacts of the particular analytical method and outlier threshold. The authors present reasonable interpretations of these patterns and discuss several alternatives, but in many cases it is not possible to discount these alternatives. It is unfortunate that more markers and a better reference genome were not available, as finer-scale resolution would have allowed more conclusive insight into these patterns. However, the authors have done a good job within the constraints inherent to working on a non-model species and I think overall the manuscript has some broadly interesting results. This manuscript constitutes a good first step in this system, identifying an intriguing pattern that warrants further investigation, hopefully with a denser genome-wide approach coupled with a well assembled genome or linkage map. Generally, my biggest criticism of this paper is that the results are sometimes interpreted as seeming more "conclusive" than warranted by the data. Major comments- Did the authors check that the loci that are strongly differentiated in the putative genomic island are not associated with sex? Could it be that sampling was sex-biased in different regions? This can happen easily with some species of fish, depending on mating behaviour, etc.- Is this really speciation? Heterogeneous genomic divergence - yes. But unless reproductive isolation has been demonstrated, it seems premature to call this speciation rather than local adaptation (i.e. in the title, line 479, etc). While the divergence is uncorrelated to an obvious environmental variable (line 668), other more subtle/less obvious ecological pressures could be driving this pattern. I do agree that the data are consistent with the coupling hypothesis, but they are also consistent with other hypotheses. - Line 484-489: This is one possible interpretation of the data. Other equally reasonable interpretations could be made, and the data doesn't allow clear discrimination here. The three hypotheses outlined in lines 643-648 are all consistent with the data, so this seems like a strong statement. Try to refrain from making strong inferences without the data to back them up.- Line 605-607: Couldn't this also be explained by recent migration? Minor comments:- the "red" colour in Figure 5 comes across as pink (contrary to line 1201 but consistent with line 366).- line 340-341: figure 2c is said to have 5 clusters demarcated by vertical lines, but it looks more like 9. Does this need to be reordered to collapse some of the clusters?- lines 361-363: unclear wording- Figure 4C: label Y-axes- Line 370: I think there are more than two non-outlier loci are mapping to the chromosome (165, 178, 71 all appear twice). But I'm not clear what the significance of this is? Why is this interesting? Non-outliers are expected even within extreme genomic islands, so this doesn't really mean much.- Lines 371-385: this list of observations seems mainly anecdotal and could be cut- Line 392: was outlier classification done by human interpretation of the visual patterns or some stat? (human is fine, but please be clear)- Figure 5: I found the description one lines 394-399 quite confusing. If there was some way to graphically show which populations are discriminated by each category, for example drawing lines on a map indicating what each cluster delineates, it would be easier. Of course, that might be even more confusing?- Line 477: they aren't ALL mapping to the single chromosome (see also line 544)- Lines 491-511: this seems relatively uninteresting and could be in the results. RADseq would likely have provided much finer resolution on the chromosomal-scale patterns.- Line 588-597: This could also simply be

non-equilibrium conditions, as we have no way of knowing if this pattern is stable. Theory tends to rely on equilibrium solutions for tractability.

### **Reviewed by anonymous reviewer 3, 28 November 2017**

This is an interesting population study of the snouted seahorse in the Atlantic/Mediterranean. The study system is interesting because there is a clinal hybrid zone between the N and S Atlantic, while in the Mediterranean there appear to be locally adapting lineages to lagoon-like habitats from marine habitats. Overall they find evidence of a few SNPs that are likely co-located on a chromosome (e.g., a likely genomic island) that appear to underlie the divergence between the N/S Atlantic and the lagoon/marine sites in the Mediterranean, which is unexpected to occur by chance. This is shown most convincingly in Figure 5 G, which shows the parallelism in allele frequencies at these candidate loci. Although they do not have a genetic map, they go to great lengths to locate markers onto chromosomes of several related fish species, which strengthens the evidence for the putative genomic island. I only have a few comments and recommendations to improve the paper, which is my opinion is already a very thorough study.

1) The authors use a small number of high-quality SNPs (~286), which normally I would argue isn't very useful for understanding the genetic basis of adaptation. However, I think there must be very extensive LD in this study system, and it might be useful to make that point very clear (sorry if I missed it). Other investigators attempting to do genome scans with so few SNPs in a species with very low LD are unlikely to find such interesting results.

2) I'm not sure what the contribution of the LK statistic is to the study. Was it only for visualization? I believe the limitation of the LK statistic is that it assumes infinite sample sizes within populations, and the degrees of freedom may not equal the "effective" number of populations in the sample (although with pairwise comparisons maybe it is OK). I'm also not familiar with the fitting scheme, at least I haven't seen it used before in the literature. I'd recommend OutFLANK or FLK, but I'm not sure if there are enough SNPs to parameterize these models well.

3) I like how the outliers are categorized, but is unclear how the genomic island is defined here. Is the genomic island all the SNPs that map to the same chromosome in the other species. What is a "genomic island"?

4) The discussion is a little long... could cut some of the text about specific locations, which is hard to follow for someone unfamiliar with the study system. Putative ecological drivers of the parallelism are not even discussed- instead, they state "genetic parallelism is observed despite an apparent absence of a common selective pressure". For someone unfamiliar with the system, it might be helpful to defend this statement in the context of the actual ecology of these different locations.

I do like how the authors take care to discuss how different processes could result in the same pattern, however.

5) All the panels in Figure 5 (except for G) are hard to understand without reference to the map, and because (I think) the colors in the panels A-F correspond to locations while the colors in panel G correspond to how the SNPs group. Please add a legend for the colors of the populations for A-F - it might also be helpful to make panel G it's own figure.

6) Some of the colors are hard to tell apart for me - especially the reds, pinks, and oranges. Otherwise the figures are really nice.

### **Reviewed by anonymous reviewer 1, 28 November 2017**

Review of 'Parallel use of a shared genomic island of speciation in clinal and mosaic hybrid zones between cryptic seahorse lineages' This study describes population genetic patterns based on SNP data from the transcriptome of a species of seahorse. Widespread sampling throughout the range of this species across a range of divergent environments (ocean vs. lagoon) lay the groundwork for interesting questions to be

addressed with this data. However, the biggest problem with this study is that it currently lacks a question-driven framework. Instead of hypotheses or predictions that could be developed from previous work on this species and previously established biogeographic patterns of other species, results are briefly summarized in the final paragraph of the introduction. Largely stemming from the issue of not having a clear set of questions to guide us, the population genetic methods and results read somewhat like a “kitchen sink” paper in which many different analyses are pursued but without an obvious explanation of what new thing we are learning from each one. For example, the structure, PCA, and neighbor-joining trees appear redundant in that they all show similar genetic patterns and relationships among the sampled populations. It would be helpful to know what is uniquely gained from each of these analyses and how they directly relate to the interpretations of parallelism and hybrid zone dynamics. Similarly, the results section reads as disorganized and descriptive.

I also had some concerns about several of the analyses and subsequent interpretations. I don't have experience with transcriptomic datasets so I could be wrong, but what are the implications of the SNPs used in this study being located within the transcriptome with respect to interpretation of “neutral” demographic patterns? It doesn't seem appropriate to interpret non-outliers as representing neutral loci and therefore as reflecting neutral patterns when all loci are from functional regions. I also failed to understand the evidence for the “genomic island of speciation” and was not convinced about the usefulness of mapping SNPs to multiple distantly related fish genomes, given the lack of consistency across the different species.

Again I return to the need for a question- or hypothesis-driven framework because then the authors could explain what they were expecting (e.g., the 4 previously described lineages or high divergence at a subset of loci associated with ocean vs. lagoon environments, possibly in parallel, etc) and how their results matched or strayed from their expectations.

I was confused as to why the Hossegor site (site 6 on map) was removed from neighbor-joining and TREEMIX analyses. The authors state it was removed because individuals from both North and South genetic clusters were observed, but doesn't this make it an interesting site from the perspective of clinal variation and admixture?

In multiple cases (e.g., Joint Site-Frequency Spectrum, Bayescan) the description of the specific analytical method had too much detail and it would be useful if instead the authors spent more time discussing the rationale for using the method on this dataset and the specific question it addresses.

Other comments: Line 110: *H. guttulatus* has low diversity compared to what? Other closely related species? Other species in this biogeographical area? It would be interesting to know how unusual this species is in this sense.

Line 120: define the coupling hypothesis

Line 197-200: Unclear what purpose Introgress serves here. This is typically used in a genomic clines framework to scan for loci that show non-neutral patterns of introgression.

Line 207: What evidence was used to support  $K=5$  in the Structure analysis?

Line 298: How many individuals sampled were assigned to the target species *H. guttulatus* versus *H. hippocampus*?

Line 332-334: I don't think Structure should be used to interpret directionality of gene flow. This scenario could be tested in the Treemix analysis.

Line 467: The authors mention that this species shows ‘remarkable genetic homogeneity over large areas’. What is this based on? If this was the previously understood scenario for this species it should be explained in the introduction and would set up this study's results very nicely because it is NOT what this study found (but based on transcriptomic data, which I don't consider to represent neutral variation).

Line 554-556: Results from the Woodall et al 2015 study should be summarized in the introduction and used to frame the follow up questions of this study.