Our replies to the reviewers' comments (in blue) as well as the description of the changes made in the text are preceded by ">>>".

Editor comments:

This Preprint is worth a revision
by Yaniv Brandvain, 2017-08-29 08:16
Manuscript: https://doi.org/10.1101/161786
Decision & reviews
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 on 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.

>>> We thank the Editor as well as the reviewers for highlighting the broader interest of our data. We understand this MS needed extensive revisions to clarify many points. Thanks to the thoughtful suggestions of this first round of review we substantially modified the MS, with notably clearer take-home messages, more emphasis for alternative hypotheses, and highlighting when results are preliminary and deserve more attention and when they are well supported.

This manuscript was challenging to write because we present and discuss several results in the same paper: (i) the discovering of cryptic lineages in a non-model species and the description of hybrid zones, (ii) the easy identification of statistically strongly supported outlier loci despite a modest number of loci, (iii) the observation of a parallel pattern of divergence at some outliers between the Atlantic cline-shaped and the Mediterranean mosaic hybrid zones and (iv) the localization of these parallel-behaving outliers on a single chromosome (in a genome of 22 chromosomes). The reason we decided to present all these results at once was that points (i) and (ii) are standard observations we needed in order to address points (iii) and (iv) that arguably are the results of interest here.

- **About result (i) (cryptic taxa)**. This very common observation, especially in marine species (Knowlton 1993, Pante et al. 2015, Sheets et al. 2018), seems solid result to us. It mostly comes from the spatial distribution of the genetic clusters. We identified broadly distributed panmictic entities separated by abrupt genetic divides with co-existence in sympatry (Hossegor sample (site 6), SW France) or close parapatry (sea-lagoon samples in Sète (sites 15 & 16), Med Sea). This argument was used to describe emblematic hybrid zones in the pre-genomic era (Hewitt 1989) with a few allozyme loci (e.g. *Bombina* toads, *Podisma* grasshopper, *Mytilus* mussels etc… Barton and Hewitt 1985), but genome sequences have never contradicted these conclusions to our knowledge. We endeavored to make clearer in figures, tables and text that we observed no (detectable) departure from genetic panmixia over thousands of kilometers from UK to SW of France, from SW of France to the Alboran Sea (East of Gibraltar), and from Spain to Greece. The color code used in the map (Fig. 1) was not our interpretation of a blurry pattern: it does represent panmictic genetic clusters. In contrast the genetic differentiation between lineages is strong (F_{ST}>0.1), and an order of magnitude stronger than usually observed at the infra-specific level in marine species (review in Gagnaire et al. 2015). These two points are now better emphasized along with the coexistence of two lineages either in sympatry (site 6), or close parapatry (sites 15 & 16) with no sampled hybrids (*i.e.* first generation hybrids).

- About result (ii) (easy finding of statistically well-supported outlier loci). Again this is standard observation (Ahrens et al. 2018), although we agree the explanatory hypotheses are multifarious (Bierne et al. 2013). We are concerned by the false positive hypothesis, although it is not very clear what should be the null model when semi-permeable genetic barriers are involved. Despite a dedicated discussion paragraph (paragraph 1), we failed in the previous version to explain why outlier tests are still useful in hybrid zone studies. We endeavored to make clearer in the new version that outliers should not be over-interpreted and we explicitly recognized we should not expect to identify selected loci with so few markers in a standard infra-specific genome scan. We would nonetheless emphasize that we observed fixed differences (F_{ST}=1) between lagoon and marine habitats in the Med Sea while the average F_{ST} could be as low as 0.05 (in the case of the Bizerte lagoon, site 20), *i.e*. a noticeable signal. In addition, we have two lagoon replicates that confirm the same loci (those that map on the same chromosome), a point we did not sufficiently insist on. Although we may not totally exclude to be lucky enough to pick up outliers with so few loci, one can nonetheless recognize that the hybrid zone theory and literature provide an accurate expectation.

- About result iii (genetic parallelism). Our key result, we believe, is that we found parallelism in two different spatial and environmental contexts (geographic *vs*. habitat mosaic), a point that might deserve all attention. The signal of parallel differentiation is usually hardly questioned as it is known to be extremely difficult to obtain without some sort of selection. One can debate about the type of selection that produced the pattern (habitat-driven selection or differential introgression), but hardly deny the pattern. First, we made clearer that we have no preferred process to explain this pattern and we endeavored to make a balanced discussion of them all. Second, we used simulations to better highlight that parallelism (a contrast

between the genetic structure observed at outlier and non-outlier loci) can hardly be obtained by chance, even with few loci (see new Fig. 6). The pattern of parallel differentiation obviously requires the outliers to behave differently from the average genome and we would find it unfair to demand having genome sequences to confirm such robust result.

- About result (iv) (clustering on the same chromosome). We endeavored to better explain that outliers consistently map on the same chromosome and now provide as a sup file the results of the chromosomal position of every SNPs (that maps on other chromosomes). We acknowledge that presenting results with many fish genomes was confusing, though aiming to show the well-conserved synteny of fishes. We now only present the results obtained with the two syngnathids and better recognize that genome sequences would be needed to progress on the issue. Although we doubt the chromosomal island will be refuted, the main question is the contribution to the overall divergence of all the other loci that only affect a small genome region. The main risk could be to highlight this region/island just because it is large and easily found, while we have no solid arguments to refute many other small regions could also contribute to reproductive isolation. However this does not prevent us to highlight we find parallelism at this island in two alternative spatial and environmental contexts.

Although we failed to well present the spatial structure in the previous version of the MS, we hope it is now better addressed, so that our main conclusions will be better understood. The signal of parallel differentiation at the chromosomal island is strong (see new Fig. 6 with simulations included). The main surprise is that, contrary to other known cases, it is not linked with an obvious shared environmental variation. We failed to convince you with our first version, but, with due respect, we are not convinced either that alternative explanations were reasonably likely (e.g. strong drift). We hope our new version will convince you that our results are not preliminary on the questions we addressed, although we fully agree exciting follow-up are expected with genome sequences and lab crosses.

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, *27*(6), 1342–1356. doi:10.1111/mec.14549

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497–503. doi:10.1038/341497a0

Bierne, N., Gagnaire, P.-A., & David, Patrice, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, *59*(1), 72–86.

Gagnaire, P.-A., T. Broquet, D. Aurelle, F. Viard, A. Souissi, F. Bonhomme, S. Arnaud-Haond, and N. Bierne. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evol. Appl. 8:769:786.

Hewitt, G. M. (1989). The subdivision of species by hybrid zones. Speciation and its Consequences, 85-110.

Knowlton, N. (1993). Sibling Species in the Sea. *Annual Review of Ecology and Systematics*, *24*(1), 189–216. doi:10.1146/annurev.es.24.110193.001201

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., … Samadi, S. (2015). Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology*, *24*(3), 525–544. doi:10.1111/mec.13048

Sheets, E. A., Warner, P. A., & Palumbi, S. R. (2018). Accurate population genetic measurements require cryptic species identification in corals. *Coral Reefs*, *37*(2), 549–563. doi:10.1007/s00338-018-1679-9

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.

>>> We also concur that hypothesis driven science is more elegant. We acknowledge we had no predictions for this work, which initially was conducted to investigate the connectivity and population size of an endangered species. We did not expect to find a complex of semi-species and a chromosomal island of divergence, although it seems to be quite common, at least in the sea. However we obtained this very exciting result: it is as if we would have found a stickleback inversion, which usually differentiates freshwater and marine populations, to differentiate northern and southern populations elsewhere independently of the salinity. This is pure serendipity and quite fascinating result. We are really frustrated we failed to convey our excitement, and left referees and you with the feeling we were discussing preliminary results. As you stressed, a more question-driven framework is too challenging here, and would be artificial. However, we endeavored to better present previous knowledge about the study system (l. 114-129) as required by reviewers#3 and #4. We also substantially modified the discussion to open room for reasonable and alternative hypotheses following your and reviewer#1 relevant comments. Finally, in the new version of the manuscript, we aimed to justify each analysis and moved in Supplementary Information some analyses and redundant results, such as the Neighbor-Joining trees, according to your and reviewer#3 comments.

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).

>>> We apologize for this mistake (lagoon *vs*. maritime habitats in Fig. 1), but now corrected it in Figure 1. A careful attention was paid to clarify the new version of the manuscript and substantial changes were made in the new version of the manuscript to make it more linear.

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.
>>> We now define the JSFS where it should be (l. 222) and moved this part into a sup file as it is secondary to the main messages of this MS. Following relevant and thankful recommendations, a careful attention was paid to typos, and alternative hypotheses are now better developed to improve our manuscript. We hope these substantial changes will be more attractive and easier to follow.

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

Sorry for the time we took to make the revisions, reconciling my newborn and these revisions was a hard job! We took your concerns seriously and believe they have undoubtedly been positive to the manuscript clarity. We hope you will be interested to write a recommendation on it.

Best,
Florentine

Reviewed by anonymous reviewer, 2017-08-15 00:08
Reviewer#1 Comments to the Author:

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 artifacts 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.

>>> We thank the reviewer for highlighting the broader interest of our data and analysis. We acknowledge that we have sometimes pushed too strongly toward our favored hypotheses (e.g. secondary contact, intrinsic selection) and we endeavored to provide a much more balanced discussion of our results in the revised version, aiming to be less "conclusive" in the new version.

However, note that we did not over-interpret the main result: a parallel pattern of differentiation at some loci in a clinal and a mosaic hybrid zone. The existence of a large chromosomal region well explains that it was easy to find these loci, and the good synteny of fish genomes also supports this hypothesis. We regret this was not sufficient to convince reviewer#1 who would expect genome data. We made clearer that the hypothesis of a chromosomal island is secondary to our main message, *i.e.* genetic parallelism, which does not require genome-wide data. We substantially modified the discussion in agreement with reviewer#1 suggestions with more caution in our interpretations.

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.

>>> We did look for a relation between sex and loci with extreme level of differentiation (as well as all the other loci). Of all the 292 individuals genotyped, 168 individuals were sexed, a sampling that comprised individuals from the five lineages as well as the hybrid zones (Hossegor and Bizerte), with a balanced sex-ratio within each site. We did not find any association with sex in the present study. This point is now clearly stated in the new version of the manuscript (l.517).

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.

>>> The continuum between heterogeneous genomic divergence driven by local adaptation and speciation is recognized as a grey zone. The best evidence we have to start talking about speciation here is strong linkage disequilibria maintained in sympatry in the contact zone between the two Atlantic lineages (Hossegor, sample 6). According to Jiggins and Mallet (2000) this is fair evidence to start talking about species. As said Brandvain and Matute (2018), "the ultimate test of whether two diverged populations are, in fact, good species is whether they will maintain their

distinctness in sympatry". In the Mediterranean Sea we do not have sympatric populations, but observed close parapatry (sites 15 and 16). We agree the concept of ecotypes could be more appropriate in this case, although as fuzzy as species concepts are. We did not intend to be conclusive about the coupling hypothesis and we endeavored to tone down the interpretation and provide a more balanced discussion. Note that different –and basic- ecological variables were checked, such as salinity or temperature, to relate a putative ecological driver to each lineage or the Mediterranean lagoon and North Atlantic lineages, but were unsuccessful. We however reckon that other more subtle/less obvious ecological pressures -that we did not look at- could also be driving this pattern.

We endeavored to clarify these points in the new version of the manuscript, *i.e.* reproductive isolation *vs*. local adaptation with possible ecological drivers explicitly stated (paragraphs 3 and 4 of the discussion), though more subtle ecological characteristics may drive this pattern. We agree we may have ascertained our manuscript with strong inferences – a criticism also shared by the Editor- that we lightened up in the new version of the manuscript. For instance, we modified our titling as well as many points in the discussion to leave more room for alternative hypotheses.

Brandvain, Y., & Matute, D. R. (2018). When genes move, genomes collide. *PLOS Genetics*, *14*(4), e1007286. doi:10.1371/journal.pgen.1007286

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*(6), 250–255. doi:10.1016/S0169-5347(00)01873-5

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.

>>> We agree with the referee we pushed too strongly on our side here and removed the claim. Thanks.

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.

>>> We agree the three hypotheses are consistent with the data and tried to provide a balanced discussion of the three.

Line 605-607: Couldn't this also be explained by recent migration?

>>> This hypothesis that could indeed be suggested, is now exposed in the new version of the MS. We however do believe that we have strong arguments against it that we now develop in the new version. Let's imagine recent migration from the two diverged lineages explain their co-existence in sympatry in the Hossegor lagoon. We would have had the chance to sample first generation of migrants in this lagoon before they have time to hybridize. In addition to be a very unlikely event, we have evidence with the outliers that recombination occurred between the two

genetic backgrounds (see the genomic cline analysis and our discussion of it). As a consequence, it can hardly be recent migration alone, the genetic clusters need to be maintained despite hybridization.
We modified this section not only by notably including the hypothesis proposed by reviewer#1 (starting l.611) but also to make our point clearer (recombination and reproductive isolation).

Minor comments:

the "red" colour in Figure 5 comes across as pink (contrary to line 1201 but consistent with line 366).
>>> We apologized for this mistake. Substantial changes were made to this figure and we carefully checked the color consistency.

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?
>>> The initial figure 2 was substantially modified, and the initial panel C is now in supporting information 3. In Figure SI3, we made the five clusters more obviously delineated. Five bars differently colored depicted the five clusters identified below the Structure outplot, while stars depicted contact zones. Using this demarcation, we kept the geographical order, which, to our mind, helps in understanding the mosaic hybrid zone in the Mediterranean Sea.

lines 361-363: unclear wording
>>> We rephrased this section (l.436 -494).

Figure 4C: label Y-axes
>>> Done. Note that, according to reviewer#4 suggestions, panel C (of now Fig. 5) was substantially modified. Only the mapping and F_{ST} of the two closest species, *i.e. H. comes* and *Syngnathus scovelli,* were kept in the new version; the figure including the comparison of the seven fish is now a supplementary file (SI5).

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.
>>> Indeed three non-outlier loci does map to the unique chromosome as initially specified in the manuscript. We apologized for this mistake. As outlined by reviewer#1, we reckon that finding these non-outlier loci was not that relevant and deleted this information in the new version of the manuscript (main text in the results and Fig. 4 & 5).

Lines 371-385: this list of observations seems mainly anecdotal and could be cut
>>> We acknowledge that this section was long and confusing. We shortened it (from 14 to 8 lines) by detailing only the results of blasts against *Hippocampus*

comes and *Syngnathus scovelli*, the two closest related fishes. Blats against the seven well-assembled fish genome were moved in Supporting Information 5.

Line 392: was outlier classification done by human interpretation of the visual patterns or some stat? (human is fine, but please be clear)
>>> Outliers were statistically identified but they were initially classified according to the populations they characterize by visual inspection of the pattern, as they were not numerous and their spatial variation was clear. We largely modified this section, contrasting the six outliers mapping in the unique chromosome from the three other outliers. Note that following reviewer#2 suggestion, we added two outlier tests, reducing the number of outliers identified from 14 to nine.

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?
>>> We modified this figure (now Fig. 4) and reorganized its description in the main text to make it clearer (l.437-443). In the new figure, only the six outliers of genetic parallelism are shown. As also suggested by reviewer#2, we added a legend in which we inserted a map. Note also that, according to reviewer#2 suggestion, we made panel G (Fig. 6 in the new version of the manuscript) it's own figure.

Line 477: they aren't ALL mapping to the single chromosome (see also line 544)
>>> We modified these two sentences.

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.
>>> We deleted this entire section and inserted our main point -*i.e.* an efficient genotyping method in a depleted species constrained by its endangered status- in the results section as suggested by reviewer#1 (l.356-361). This method also allowed the use of long sequences from the transcriptomic population survey (Romiguier et al. 2016; 91 – 1628bp) to map the outliers identified to seven well-assembled fish. Mapping RAD-seq markers (c.a. 80 bp) would unfortunately have been less efficient for this analysis (*i.e.* mapping).

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

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.
>>> We unfortunately did not well understand what reviewer#1 meant here. Our point was to explain why the genomic island loci have lost the association with the

other loci differentiating the northern and southern genetic backgrounds. We concur that the genetic barrier may be in the ongoing non-equilibrium process of disintegrating by slow successive swamping at barrier loci and that the genomic island might be one of the first. We modified this sentence to include this possibility.

Reviewed by anonymous reviewer, 2017-08-13 02:45
Reviewer#2 Comments to the Author:

This is an interesting population study of the snouted seahorse in the Atlantic/Mediterranean. The study system in 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.
>>> We thank the reviewer for this very positive appraisal of our dataset.

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.
>>> We fully agree with the referee that the density of markers is largely insufficient for a genome scan at the infra-specific level. This is the reason why we frame our work into the hybrid zone framework.

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.
>>> Again, we fully agree with the referee and endeavored to make the point clearer in the new version. In addition, the existence of a large island of divergence, which can be an inversion, a 'supergene', or a coldspot of recombination (*i.e.* with strong LD), is easier to find with few markers. Indeed our outliers represent 3% of our markers and one chromosome is 4% of the seahorse genome. Mapping outliers in a unique chromosome is therefore in agreement with reviewer#2 comment. We modified Results (l.452-459) and Discussion (l.580-584) to make these points more obvious.

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.

>>> The LK analysis was done on a pairwise comparison to show that the overall observed F_{ST} distribution departs from a simple neutral model and the LK statistics was showed to be well suited to do this by Whitlock (2008). Following our extensive revision of the manuscript this analysis has been removed. Following reviewer#2's recommendation we now provide an additional outlier test with FLK, but OutFLANK did not prove useable with the data we have. We also have included an in-house simulation test following Fraïsse et al. (2014) approach. This latter test uses simulations and the prediction that false positives should not be the same loci in two pairwise comparisons. Using the F_{ST} - F_{ST} co-plot allows to evidence how the parallel pattern of differentiation observed at the genomic island loci is highly unlikely under the inferred demographic history of populations. Together with the FLK test, which also intends to account for the history of populations, these new tests reinforce our conclusions that the genomic island loci are under some sort of selection. Not only the lagoon-sea differentiation in the Mediterranean Sea (F_{ST} =1) and the north-south differentiation in the Atlantic Ocean (F_{ST} =0.8) are very strong at these loci, but the fact that they are the same loci in the two comparisons is nearly impossible to obtain under a neutral model.

Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene-Flow in a Mosaic Hybrid Zone: Is Local Introgression Adaptive? *Genetics*, *197*(3), 939–951. doi:10.1534/genetics.114.161380

Whitlock, M. C. (2008). Evolutionary inference from QST. *Molecular Ecology*, *17*(8), 1885–1896. doi:10.1111/j.1365-294X.2008.03712.x

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"?

>>> Yes, the genomic island is defined by all the SNPs mapping to the same chromosome (Fig. 4). We agree the terminology is not ideal but it is widely used in the literature. We now provide an explanation in the new version of the manuscript (l. 584).

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.

>>> As suggested by reviewer#2, we shortened the discussion from 249 lines to 233 lines. We notably removed the section dedicated to our genotyping method (Section 1- in the previous version of the manuscript) as suggested by reviewer#1. In addition, descriptions of specific locations are lightened up, but, when necessary, their locations are now specified.

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.

>>> Actually, habitat specialization, *i.e.* lagoon *vs*. sea, is likely driving genetic parallelism in the Mediterranean Sea. We thank reviewer#2 to underline this point and we apologize for the lack of precision in the previous version. We now clarified this point l. 666, thanks to reviewer#2 comment.

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.

>>> We apologize for the lack of clarity regarding Figure 5, a point also raised by reviewer#1. We substantially modified this figure (now Fig. 4) and kept only the outliers that both showed genetic parallelism and mapped to the unique chromosome (the initial A- category). A map was added as a legend, and graphical modifications now discriminate populations to highlight (as suggested by reviewer#1). According to reviewer#2 suggestion, we made the G-panel (now Fig. 6) 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.

>>> We thank the reviewer for her/his appraisal of the figures. By simplifying the figures with only outliers that both showed genetic parallelism and mapped to the unique chromosome, we hope figures are now easier to follow.

Reviewed by anonymous reviewer, 2017-08-18 00:45
Reviewer#3 Comments to the Author:

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.

>>> We apologize for the lack of clarity that likely confused reviewer#3. As explained above, describing a complex of cryptic semi-isolated taxa is a result of our work we needed before to describe genetic parallelism but that we didn't want to

discuss at length. As acknowledged by the Editor, it may be difficult in such context to fit the manuscript as a question-driven one. Our main result is pure serendipity, we acknowledge. We however endeavored to follow reviewer#3 suggestion in modifying the introduction to explicitly provide all results from previous work (Woodall et al. 2015 and Lopez et al. 2015; l. 114-129) and developed our predictions regarding this study system (l. 105-110; l.140-143), *i.e.* reproductively isolated cryptic lineages rather than spatial differentiation (Woodall et al. 2015).

López, A., M. Vera, M. Planas, and C. Bouza. 2015. Conservation Genetics of Threatened Hippocampus guttulatus in Vulnerable Habitats in NW Spain: Temporal and Spatial Stability of Wild Populations with Flexible Polygamous Mating System in Captivity. PLoS ONE 10:e0117538.

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.

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.

>>> We did not aim to analyze our multi-locus genotype dataset by different methods to be viewed as a "kitchen sink" of population genetics analyses, and are embarrassed if the MS was perceived as such. We used different methods developed with different statistical approaches to make solid assumptions on our results: we indeed thought interesting to cross check the outputs of a model-based approach with strong priors and hypotheses (HW equilibrium, no linkage between markers) with methods based either on distance without assumptions on data or on distance matrices among sets of individuals that fall within predefined population samples). However, according to reviewer#3 comment, a point also shared by the Editor, we acknowledge that these analyses may appear redundant as similar results were observed, so that some analyses were moved in the Supplementary file (SI3) in the new version of the manuscript. Moreover, in Materials and Methods, we clearly (and briefly) explained the purpose of each analysis. In addition, we reorganized and simplified the Results section aiming to improve its structure and its comprehension, as well as clarifying its link with genetic parallelism and hybrid zone dynamics.

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.

>>> We did not well understand referee's concern here. Most of our SNPs are synonymous mutations, and indirect selection at linked sites is only marginally stronger in coding regions than in intergenic regions that are also notoriously affected by linked selection. We nonetheless concur that "non-outlier" is a better terminology than "neutral", and modified the revised MS accordingly. We concur that every loci are affected by some kind of linked selection at varying rates, including the barrier effects of barrier loci (Barton and Bengtsson 1986), and that oulier loci could be understood as the most affected.

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

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.

>>> We modified our presentation of the results and hope that the consistency observed now becomes evident in the new version. We acknowledge that the order of the loci was not always consistent but the mapping on a single chromosome in genomes of >20 chromosomes was consistent. We acknowledge that presenting results with seven fish genomes was confusing and, we only kept the mapping against the two closest species to improve the clarity of the MS. The complete blats/blasts analysis is still available in the Supporting Information 5.

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.

>>> Our study was initially conducted to investigate marine connectivity in this endangered species, and we did not expect to find a complex of cryptic species with a chromosomal island of divergence. We endeavored to clarify our goal in the introduction and clearly stated results from previous work (paragraphs 2&3 of the Introduction).

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?

>>> Yes, site 6 is very interesting and that's why we conducted a genomic cline analysis on this particular site. However, such site of co-existence with no or few hybridization needs to be removed as soon as a method assumes random mating. In TreeMix, Pickrell and Pritchard (2012) clearly stated to remove such sample in the analysis. Though not clearly explained in the previous version of the manuscript, we

now clearly stated these reasons in the new version of the MS (l.279). Note that, according to reviewer#3 comment (point 2), NJ analyses are now removed.

Pickrell and Pritchard (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genetics 8:

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.
>>> The rationale of each analysis has been explained in the new version of the manuscript.

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.
>>> *H. guttulatus* has a very low genetic diversity in comparison to 75 non-model animal species (Romiguier et al. 2014). We now specified it l. 120.

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

Line 120: define the coupling hypothesis
>>> Done.

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.
>>> We used INTROGRESS only for raw data visualization and this has now been moved to the sup files. We did not conduct the genomic cline analysis with INTROGRESS because the alternative Barton's concordance method was better suited with the data we had.

Line 207: What evidence was used to support K=5 in the Structure analysis?
>>> A brief explanation of selecting K=5 is now included in the new Results section of the manuscript. Following recent warning messages (e.g. Janes et al. 2017), we do not think that a true value of K exists, and we simply stopped to further subdivide into new clusters when the additional clusters had no geographic meanings. Departure from Hardy-Weinberg equilibrium, unbalanced sample sizes, or strong linkage disequilibrium may lead to further errors in selecting the optimal K-value (Kalinowski 2011, Curby et al. 2013). Exploring different K-values and crosschecking the results with different methods freed of model-based algorithms - which are implemented with strong priors and hypotheses- was favored in selecting the optimal K. We explored different values of K (from 2 to 10) and increasing the K-value higher than five results in the distinction of Bizerte first, but then incoherent

results (no individuals completely assigned to any new cluster, but rather an equal admixture for each individual within a cluster). We crosschecked the outputs from this analysis with distance-based methods that make no assumptions on our data (here PCA and the visualization of the raw data) to validate K=5 in the Structure analysis. This is now stated in the MS (l.400).

Cubry, P., Bellis, F. D., Pot, D., Musoli, P., & Leroy, T. (2013). Global analysis of Coffea canephora Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genetic Resources and Crop Evolution, 60(2), 483–501. doi:10.1007/s10722-012-9851-5

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. *Molecular Ecology*, *26*(14), 3594–3602. doi:10.1111/mec.14187

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity, 106(4), 625–632. doi:10.1038/hdy.2010.95

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

>>> Of all fish genotyped – four *H. hippocampus* and 467 *H. guttulatus* identified based on morphological criteria-, 465 fish were correctly assigned to the target species. The number of misidentification (two fish, *i.e.* 0.4% of the whole dataset) is now clearly stated in the new version of the manuscript (l.340).

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.

>>> We tested this scenario in the Treemix analysis. In addition, note that we observed a geographic gradient in STRUCTURE ancestry rates from the Southwest of France to the UK, and this can reasonably be interpreted as an introgression gradient. STRUCTURE alone shouldn't be used to interpret directionality of gene flow but STRUCTURE + geography can.

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).

>>> As said above we found genetic homogeneity from Greece to Spain (2000 km) and from the Alboran Sea to the South-west of France (1900 km). We apologize if we weren't clear enough and we endeavored to make it clearer in the new version (e.g. l.600, l.651). A previous study with a few microsatellites also observed panmixia at a large spatial scales and three of the five genetic clusters we identified, but in this previous study only one lagoon was sampled in the Western Mediterranean Sea and a sea site in the Eastern basin (Woodall et al. 2015).

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.

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.
>>> We summarized Woodall et al. (2015) study (l.124-139). However, as mentioned above and underlined by the Editor, a question-driven framework is debatable in this context. We however endeavored to make our reasoning clearer along the new version of the manuscript.

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.

I also add an additional review I received that was not conducted via the PCI reviewer system (reviewer 4)
Reviewer#4 Comments to the Author:

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.
>>> We thank reviewer#4 for highlighting the broader interest of our data.

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.
>>> We substantially modified the introduction, clarified the results section and shortened the discussion. We hope these substantial changes will be more linear and attractive for the readers.

Questions: Whole argument about the clines was extremely contrived and confusing to follow.
>>> We may have misunderstood reviewer#4 comment, but, we are convinced it is the spatial organization of the different genetic backgrounds with the two traditionally described hybrid zones – clinal and mosaic- that explain the full story in *H. guttulatus*. We thus introduced hybrids zones first and modified this section to make it clearer. Analyzing clines was an important support in favor of different rate of introgression, contrasting convergent markers (Fig. 6) to other markers. We substantially modified Results and Discussion sections to clarify our point.

Difficult to follow the questions addressed in the manuscript
>>> This point was also shared by reviewer#3, but partially contradicted by the Editor (see the first and last comments of reviewer#3 but the second point of the Editor). We clarified our goal in the introduction (l.105-110; l.140-143).

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

>>> Surprisingly, we don't know a lot regarding *H. guttulatus* biology and we initially investigated this endangered species in a marine connectivity context as a follow-up of Woodall et al (2015) study. All biological characteristics are now clearly stated in the new version of the manuscript (l. 114-132).

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.

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.

>>> As explained above (see reviewer#4 comments 2, 3 and 5), we substantially modified our manuscript to make it easier to follow.

Minor comments: Line 124: Misspelled paratrically

>>> We corrected this mistake.

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

>>> We modified the single occurrence of "patchily" in the manuscript. A "patchy environment" is however the term dedicated to the environment observed in the Mediterranean Sea, *i.e.* a patchy fine-grained environment (Harrison and Rand, 1989).

Harrison RG, Rand DM (1989) Mosaic hybrid zone and the nature of species boundaries. In: Speciation and its Consequences (eds. Otte D, Endler JA), pp. 111–133. Sinauer Ass. Inc., Sunderland, MA.

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

>>> As recommended by reviewer#4, we detailed where (and how) we got the samples from the aquarium (l.177-180) in the new version of the manuscript. Briefly, the Mare Nostrum aquarium (France) kept captive-bred seahorse, sampled in three different sites and allowed us to fin-clipped them.

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

>>> We now stated that a *de novo* transciptome assembly was performed by Romiguier et al. (2014) and briefly explained the method they used (l.190-209).

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

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.

>>> SNPs were ascertained with samples from three populations in lagoon and sea habitat. We agree that ascertainment bias is a concern and this was the reason we dedicated a full analysis to it. We aimed to estimate the extend of ascertainment bias introduced in our dataset. We thus compared our dataset to Romiguier et al. (2014), assumed to be freed of any ascertainment bias. Our apprehension was to get less singletons that would also affect differently the genetic structure, but we did not (Fig. SI1 and l.352-355). In the present study, the obtained JSFS showed an even representation over the entire allele frequency range (Fig. SI1) between Romiguier et al. (2014) and our dataset, which demonstrates the suitability of the markers discovered to study demographic history of populations. Genetic structure was not affected and genetic homogeneity was still observed over large areas (F_{IS}=0), using either microsatellite data (Woodall et al. 2015) or our SNP-dataset.
We thank the reviewer to underline this point, which is now clarified in the new version of the manuscript (l. 352-355).

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.

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.
>>> We apologize for this omission now explicitly stated in l. 208.

Line 169-170: Clever! This is a useful measurement comparing this to other analyses, and this is not done in previous studies.
>>> We thank the reviewer for acknowledging this analysis.

Figure SI1: 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.
>>> Figure SI1 legend was modified and further explanations about the Joint-Site Frequency Spectrum were notably added aiming to clarify our purpose. How to read Figure SI1 is now clearly explained in the caption of the supplementary figure.

Along with the six *Hippocampus guttulatus* transcriptome assemblies, Romiguier et al. (2014) performed two *H. hippocampus* transcriptome assemblies (*i.e.* used as an outgroup), making possible the identification of the most parsimonious ancestral variant –and consequently the derived allele state. How we obtained the derived allele state is now clearly stated in the manuscript in l.220.

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

Line 179: Oriented = Polarized?
>>> Yes, changed.

Figure 1: need to define the difference between circle and triangles
>>> We apologize for this mistake now corrected. A legend with the same symbols used in the map was added in Figure 1. Note that, as suggested by reviewer#3 and the Editor, NJ trees were removed from the figure.

Figure 1: remove the little lines where you overlay a figure.
>>> Substantial changes were made in this figure following recommendations (see above), so that no little lines appear by now.

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.
>>> Following thankful recommendations, Figure 1 was substantially modified. NJ trees were removed and a legend was added. The caption clearly stated what each color corresponds to. Besides, a gradient from white to black would have been very confusing in differentiating the five lineages identified along with the two admixed populations, so that we used different strong color to differentiate lineages. Note that we do not aim to publish this paper in white and black, and would pay to publish it with colors.
Note that we modified the Mediterranean maritime lineage color from light green to dark green to hardly contrast the lineages.

Figure 2C: Need to add axis or add correlations between Figure 2B-2C as it is difficult to follow the big points.
>>> We substantially modified Figure 2 (now SI3) aiming to make the relation between these two panels more obvious.

Figure 4: Difficult to follow what the authors are trying to portray in figure 4C.
>>> Following thankful recommendations of reviewer#3, we substantially modified this Figure (now Fig. 5), so that we only kept the two closest species of *H. guttulatus*. We detailed our point in the new version of the manuscript from l.452 to l. 459.

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

>>> According to relevant reviewers recommendations, we split Figure 5 in Figures 5 and 6 and substantially modified them. In Figure 5, we illustrated that allelic frequencies are nearly identical between the North Atlantic and the Mediterranean lagoon lineages for the outliers mapping to the unique chromosome. Figure 6 (previously Fig. 5G) shows the parallelism in allele frequencies at these candidate loci. These markers appear to underlie the divergence between the North/South Atlantic on the one hand (along the y-axis) and the lagoon/marine sites in the Mediterranean on the other hand (along the x-axis), which is likely unexpected to occur by chance.

A short section was added in the new version of the manuscript aiming to clarify our point.

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.

>>> Following thankful recommendations from all reviewers, substantial changes were made to improve not only the figures but also their ordering.

Figure 1 is now focusing on the map of sampling, Neighbor Joining trees being removed. Figure 2 now illustrates only the PCA. Panels B and C were moved in Supporting Information 3 and a particular attention was paid to make the main point of the C panel clearer. Figure 3 shows the TreeMix analyses.

Once outliers identified, we looked for their frequencies (Fig. 4), their position along the genome (Fig. 5), and pattern (Fig. 6 in the new version of the manuscript) and further investigated the genomic cline at Hossegor (Fig. 7 in the new version of the manuscript).

We modified the results section to make it as fluent as possible.

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

>>> Reviewer #3 also raised this point. Site 6 corresponds to individuals from both North and South Atlantic lineages without any admixture. Adding this site in the trees would have reflected different and incorrect processes. Hossegor (site 6) would have been located between the North and South Atlantic lineages instead of being entirely part of each lineage. We thus removed this particular site in the Neighbor Joining trees (now removed anyway).