Introgression from related species reveals fine-scale structure in an isolated population of mussels and causes patterns of genetic-environment associations

Marianne Elias based on peer reviews by Thomas Broquet 🆔 *and Tatiana Giraud*

Submitted: 28 December 2017, Recommended: 16 November 2018

Cite this recommendation as:

Published: 16 November 2018

Assessing population connectivity is central to understanding population dynamics, and is therefore of great importance in evolutionary biology and conservation biology. In the marine realm, the apparent absence of physical barriers, large population sizes and high dispersal capacities of most organisms often result in no detectable structure, thereby hindering inferences of population connectivity. In a review paper, Gagnaire et al. [1] propose several ideas to improve detection of population connectivity. Notably, using simulations they show that under certain circumstances introgression from one species into another may reveal cryptic population structure within that second species. The isolated Kerguelen archipelago in the south of Indian Ocean represents a typical situation where the structure of coastal marine organisms is expected to be difficult to detect. In an elegant genomic study, Fraïsse et al. [2] take advantage of introgression from foreign lineages to infer fine-grained population structure in a population of mussels around the Kerguelen archipelago, and investigate its association with environmental variables. Using a large panel of genome-wide markers (GBS) and applying a range of methods that unravel patterns of divergence and gene flow among lineages, they first find that the Kerguelen population is highly admixed, with a major genetic background corresponding to the southern mussel lineage *Mytilus platensis* introgressed by three northern lineages. By selecting a

panel of loci enriched in ancestry-informative SNPs (ie, SNPs highly differentiated among northern lineages) they then detect a fine-scale genetic structure around the Kerguelen archipelago, and identify a major connectivity break. They further show an associating between the genetic structure and environmental variables, particularly the presence of Macrocystis kelp, a marker of habitat exposure to waves (a feature repeatedly evidenced to be important for mussels). While such association pattern could lead to the interpretation that differentiated SNPs correspond to loci directly under selection or linked with such loci, and even be considered as support for adaptive introgression, Fraïsse et al. [2] convincingly show by performing simulations that the genetic-environment association detected can be entirely explained by dispersal barriers associated with environmental variables (habitat-associated connectivity). They also explain why the association is better detected by ancestry-informative SNPs as predicted by Gagnaire et al. [1]. In addition, intrinsic genetic incompatibilities, which reduce gene flow, tend to become trapped at ecotones due to ecological selection, even when loci causing genetic incompatibilities are unlinked with loci involved in adaption to local ecological conditions (Bierne et al. [3]'s coupling hypothesis), leading to correlations between environmental variables and loci not involved in local adaptation. Notably, in Fraïsse et al. [2]'s study, the association between the kelp and ancestry-informative alleles is not consistent throughout the archipelago, casting further doubt on the implication of these alleles in local adaptation. The study of Fraïsse et al. [2] is therefore an important contribution to evolutionary biology because 1) it provides an empirical demonstration that alleles of foreign origin can be pivotal to detect fine-scale connectivity patterns and 2) it represents a test case of Bierne et al. [3]'s coupling hypothesis, whereby introgressed alleles also enhance patterns of genetic-environment associations. Since genomic scan or GWAS approaches fail to clearly reveal loci involved in local adaptation, how can we disentangle environment-driven selection from intrinsic reproductive barriers and habitat-associated connectivity? A related question is whether we can reliably identify cases of adaptive introgression, which have increasingly been put forward as a mechanism involved in adaptation [4]. Unfortunately, there is no easy answer, and the safest way to go is to rely – where possible – on independent information [5], in particular functional studies of the detected loci, as is for example the case in the mimetic butterfly *Heliconius* literature (e. g., [6]) where several loci controlling colour pattern variation are well characterized.

References:

[1] Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evolutionary Applications, 8, 769–786. doi: [10.1111/eva.12288](https://dx.doi.org/ 10.1111/eva.12288)

[2] Fraïsse, C., Haguenauer, A., Gerard, K., Weber, A. A.-T., Bierne, N., & Chenuil, A. (2018). Fine-grained habitat-associated genetic connectivity in an admixed population of mussels in the small isolated Kerguelen Islands. bioRxiv, 239244, ver. 4 peer-reviewed and recommended by PCI Evol Biol. doi: [10.1101/239244](https://dx.doi.org/ 10.1101/239244)

[3] Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072. doi: [10.1111/j.1365-294X.2011.05080.x](https://dx.doi.org/ 10.1111/j.1365-294X.2011.05080.x)

[4] Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology, 22, 4606–4618. doi: [10.1111/mec.12415](https://dx.doi.org/ 10.1111/mec.12415)

[5] Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of Evolutionary Biology, 30, 1450–1477. doi: [10.1111/jeb.13047.](https://dx.doi.org/ 10.1111/jeb.13047)

[6] Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. A., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. Current Biology, 28, 1839–1845.e3. doi: [10.1016/j.cub.2018.04.072](https://dx.doi.org/10.1016/j.cub.2018.04.072)

Reviews

Evaluation round #3

DOI or URL of the preprint: https://doi.org/10.1101/239244
Version of the preprint: 3

Authors' reply, 24 October 2018

Download author's reply
Download tracked changes file

Decision by Marianne Elias, posted 24 October 2018

Fraisse et al.: recommendation pending minor changes

Dear Dr Fraisse,

I am pleased to announce that, after assessment of your revised version by the two referees, we will recommend your paper, pending minor revisions asked by Thomas Broquet. Both referees acknowledge major improvements in this version, so congratulations for the good job! Best regards,

Marianne Elias

Additional requirements of the managing board

We ask you to carefully verify that your manuscript complies with the following requirements (indicated in the 'How does it work?' section and in the code of conduct) and to modify your manuscript accordingly:

-Data must be available to readers after recommendation, either in the text or through an open data repository such as Zenodo, Dryad or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) must be available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.

-Details on experimental procedures must be available to readers in the text or as appendices.

-Authors must have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article."

This disclosure should be completed by a sentence indicating that some of the authors are PCI recommenders: "XY is one of the PCI Evol Biol recommenders."

Reviewed by Tatiana Giraud, 01 October 2018

The authors have done a great job in addressing the previous comments, with extensive work on the text and new analyses, and I can only congratulate them for this niece pice of work.

Reviewed by Thomas Broquet ⓘ**, 01 October 2018**

I have no further major criticisms or comments on this much clearer version of the manuscript. I list below a few minor points that caught my attention.

Most importantly, I struggled to understand the notions of ancient and recent incomplete lineage sorting (even with the explanations provided in the authors' responses to reviewers' comments). I think this point should be clarified.

L. 21: replace "connectivity" with "genetic connectivity" (?)

L. 23: "introgression clines confounded with local adaptation": revise wording (a pattern cannot be confounded with a process).

L. 39: "admixture variation": this verbal shortcut is unclear

L. 58-59: I'm not sure to completely follow the difference between points i) and ii)

L. 117-132: is this long résumé of the results really necessary in the introduction?

L. 141: comprises

L. 238: If this is not already in there (I did not check), I would add a figure presenting these different models in sup mat.

L. 372: the least supported

L. 467: wrong italicization

L. 581: "another caveat". Where is the first caveat?

Evaluation round #2

Authors' reply, 29 September 2018

Download author's reply
Download tracked changes file

Decision by **Marianne Elias**, posted 29 September 2018

Fraisse et al.: revisions requested

Dear Dr Fraïsse,

The same referees (Tatiana Giraud and Thomas Broquet) have evaluated the revised version of your manuscript. While they acknowledge improvements and consider the manuscript original and of general interest, there are still many concerns. I outline below the main points that need to be addressed carefully. The reviews provide more details on those, as well as additional comments, which should be addressed thoroughly.

- Focus of the manuscript: this has obviously improved since the first version, but only in some places. Notably, the introduction and the discussion of the manuscript are still very tuned on adaptation, as if the main goal of the paper were to investigate adaptive introgression (several examples are highlighted by the referees, and many more can be found throughout the text). As such, the current version looks a bit 'schizophrenic'. Introduction (and, to a lesser extent, discussion) and the questions should be entirely reframed to refocus the manuscript around the fine-scale genetic structure in the Kerguelen and the effect of introgression from foreign lineages (see the suggestions of both referees in this respect). Adaptive introgression can of course be mentioned (briefly in introduction), and discussed (in light of your results and literature on the topic in the discussion), but as pointed out at the previous round of reviews and here again, your data do not enable to test for adaptive introgression. Also, the manuscript lacks a general conclusion that extends beyond mussels.

- General clarity: as highlighted by both referees, the manuscript is complex, and lacks clarity in many instances. In addition to streamlining the text, improving general consistency and making sure appropriate vocabulary/phrasing is used, you should find a way to outline in a clear way 1) the hypotheses you want to test, 2) the data you have (you have different sets of samples and different sets of markers, which makes things hard to follow) and 3) the methods you implement for each hypothesis test. This could take the form of a kind of a synthetic section at the beginning of Material and Methods, and/or a figure, or something else, but it is essential that those methodological aspects are clarified.

- Tatiana Giraud raised an important point during the first round of reviews, which has not been satisfactorily addressed: please clearly explain (in the ms, not only in the reply to the referees) on what principle the methods used for that purpose disentangle ILS from introgression, so that the reader can assess how reliable the inference is.

- Another point raised by Tatiana Giraud relates to accounting for spatial distances in the environment-genotype analyses. This is done by using spatial coordinates, while mussel dispersal around the Kerguelen is likely more constrained by coastal distances. I don't know if even a crude measure of such distance can be incorporated in the analyses, but if not, the discussion should not be so affirmative that spatial structure does not have any effect.

- Although Thomas Broquet acknowledges the improvement provided by the new simulations to the link between fine-scale genetic structure and connectivity break (in the absence of local adaptation), he calls for caution for interpreting the results because of the lack of confidence intervals. Additionally, although the different analyses that enable to assess gene flow do indicate introgression from northern species (provided that these analyses are reliable), the patterns detected are somewhat different (e. g., introgression from galloprovincialis (Treemix) versus mostly edulis (Twisst) or both (dadi); see also Thomas Broquet's comment on the ancient gene flow inferred by dadi versus claims of secondary contact). Why do you think this is so?

A few minor comments, in addition to those mentioned in the review:
L. 53: should read 'thereby generally showing', or 'therefore they generally show'
L. 217: please briefly explain how KASPar works.
L. 263 and below: the 'delta' of dadi is not written the same way as in the results.
L. 342: you never mention how many demes you model.
L. 549: the % given are those of platensis, not edulis.
L. 559: should read 'provided'.
L. 585: 'in agreement' with...?
L. 787: 'our results are very promising that...' doesn't have a correct syntax.
Please check the text thoroughly for additional inconsistencies, typos etc that we may have missed.
Good luck with the revision,
Marianne Elias

Reviewed by Tatiana Giraud, 16 April 2018

This is a revised version of a study reporting genomic analyses of mussel populations for investigating introgression and adaptation. Overall, I am sorry to say that I have the same main concerns as with the previous version: the text is unclear and hard to read in many places, with even inappropriate or ambiguous wording in some cases (see below for some examples, some may seem details, but it is tiring to read a ms where you have to interpret many sentences to try to understand what the authors really mean, and science reports should be precise and exact). Furthermore, there are still too many confusing details while it is difficult for a reader not familiar to the system to get a global understanding. Most importantly, I am still not convinced by the interpretation that introgression is more likely than incomplete lineage sorting. No attempt has been made in

the manuscript to explain how the methods used can disentangle incomplete lineage sorting and introgression. The manuscript should explain, briefly and clearly, on what principle is based these methods and how they can reliably infer introgression when there is so much incomplete lineage sorting. Another important concern is that the whole text seem to assume there is local adaptation in the Kerguelen, while this can only be assessed by experiments; please state it local adaptation has really been demonstrated (by experiments and not by just genetic differentiation at a few markers, which most likely reflects gene flow more than adaptation), otherwise be more cautious in formulations. For example, one main goal of the study is stated to be to assess whether introgression can promote adaptation, but I do not think this question can be addressed with the data at hand. The results show fine scale genetic differentiation and heterogeneous introgression both in the genome and geographically, but this could result from neutral processes including barriers to dispersion, while adaptive explanations are still too much put forward as explanations for these patterns. I do think this study is sound and interesting for a broad audience in evolutionary biology, and it would be a pity if the manuscript would stay hard to read and not convincing enough. More specific comments are found below, some explaining more precisely the general comments above.

-L27: this is the question that you can indeed address the data, while I don't think you can address the question L21.

-L22: the sentence is not optimal: precisely because they are semi isolated species, Mytilus may not be a good model to assess the general importance of introgression.

-L33: what panel?

-L38: it would be important to compare the level of fine scale differentiation to those found in Northern species.

-L41: "we believe" is not scientific, evidence or rationale should be resented instead.

-L42: this appears at first sight hard to understand ("describe" is not the best term, see also L84), but actually after reading twice the manuscript, this is the best conclusion we can draw from the data: I would recommend to re-write the introduction in the abstract and the manuscript to further tone down adaptation and introduce better these ideas so that they will be clearer (eg L59, the idea is only introduced In a few words while adaptation is the focus of the whole following paragraph, that does not seem relevant to the study); similarly, the sentences L76-81 are unclear before having read the whole manuscript.

-L58: local adaptation precisely usually leaves no footprints genome wide because of gene flow, only a few genes are differentiated among populations, so an approach like the one used in the manuscript would likely not detect it.

-L66: "adaptation from hybridizing species" is an incorrect shortcut, it should be "adaptation allowed by gene flow from hybridizing species".

-A schema recapitulating species divergence on a map and history would be helpful.

-The figures should be either at the end or where they are cited, it is tiring to look for them in the text.

–L112-145: there are too many unneeded details and, more generally, too much focus is given to the marker Glu-5, just one historical marker that did not allow powerful inferences, you have much better power now, I would think Glu5 is just not relevant anymore.

-Logical links are not appropriately used in many places (e.g., accordingly, as such, indeed... are not appropriately used), it renders the text hard to read because we have to interpret each time what you mean.

-L131-132: I do not think it is relevant to introduce that adaptation has been suggested based on 3 allozymes....

-L135-136: similarly, I find misleading to place the introduction of differentiation at four markers within a paragraph on adaptation; genetic differentiation at a few markers cannot say anything on adaptation in my opinion, even if associated with environmental variables.

-L137 and elsewhere: "correlated" is wrongly used at several places; correlation is a precise statistical test, it should not be used for describing associations with non-quantitative variables . -L142: this is an example where the text seems to assume there is local adaptation while I am not convinced this is the case from what is presented. In addition, here and elsewhere, the formulation is too vague: "adaptation in Kerguelen populations"

seems to mean "local adaptation within the Kerguelen" while the sentence would mean "adaptation in the Kerguelen compared to Northern populations" (see also L195).

- L148 and L 617: examples where the text assumes there is local adaptation, while I am not convinced this is the case.

-L148-149: I do not think these are questions you can/did address, while the questions formulated at the end of introduction are very important for the reader to understand the manuscript. As it stands, one expects some kinds of results or experiments that never come and one has a hard time to understand the results because one has to guess the questions addressed. The question 1 is just not addressable with the data, the question 2 is not well formulated, in particular it is not that introgression "ease the investigation" but instead "reflect", this is very different. I would recommend the following questions: 1) is there genetic differentiation at fine scale in the Kerguelen? 2) is there association of genetic differentiation with environmental variables and/or geography? 3) can we detect footprints of introgression from Northern species and disentangle them from incomplete lineage sorting? 4) in this case, does introgression contribute to the fine scale genetic differentiation in the Kerguelen? 5) in this case, can we infer whether the heterogeneity in introgression is due to local adaptation or barriers to dispersal? (although not sure either you can address this last question).

-L164-167: "more" than what?

-L174: a map would be useful.

-L179: how can samples be "representative" of a population? Do you mean non-admixed?

-The Table and Figure legends are not clear enough: they should be understandable by themselves, while many abbreviations and population codes are not defined in the legends, genus names should not be abbreviated there.

-L195-196: you cannot address this question

-L206: delete "any" and "retained" . -L210: I do not understand why only 30 were highly differentiated as all have been retained for being differentiated.. do you mean a kind of threshold? What threshold?

-L239: "artificial chromosome" is not the appropriate term.

-L259 and L377: The "Chilean" mussels have never been introduced.

-P11 and P20: we really need to understand how introgression can be distinguished from incomplete lineage sorting: on what principle? Because this inference is at the core of your study, it is not sufficient to refer to the publication of the software, you have to convince here the reader that you have the power to reliably infer introgression. This issue should even be presented in introgression instead of adaptation.

-L278; BTGS and AIC are not defined.

-L365: the clade does not seem that divergent compared to others.

-L469: "adaptation in M. edulis": another example of unclear sentence where we have to guess what you mean: adaptation in M. edulis compared to other species as the sentence seems to say? Or local adaptation within M. edulis? Actually, most genes are "involved in adaptation", even housekeeping genes under purifying selection... in many other places, the term "selection" is used while you likely mean "positive selection" or "divergent selection" or "specific adaptation", this is different... same comment L532 "selection in the island": compared to elsewhere or local adaptation?

-L474: "from which they derive": incorrect wording: they do not derive from M. edulis but from a common ancestor, or I did not understand and the sentence is unclear.

-L476: "negligible": so is it really reliable?

-L499: awkward wording (a site cannot show genetic structure) . -L580 Looking at the map of the Kerguelen with the fractal coast, the genetic differentiation among sites does not seem that surprising, in contrast to what is said in the introduction about marine organisms? And it lets think that longitude is not the best variable to take into account geography in the analyses.. it should rather be the coastal rugged distance.

-L582 and elsewhere: replace "Macrocystis" by "Macrocystis presence" otherwise the sentences do not make sense. And actually, Macrocystis presence could also be linked to dispersal routes rather than indicating local adaptation of mussels?

-L621 and elsewhere : do not refer to panels without figure numbers.

-L621 and 641: they are not "control" (for which treatment?), they are background loci.

-L629: I would delete all reference to Glu5 or most of it, this sounds really anecdotal.

-L646-654: this should be introduction instead.

-L674: this has not been shown, stochastic processes could also lead to such heterogeneity, or argue why not.

-L679-680: not convincing, this can be incomplete lineage sorting.

-L688 and elsewhere: do not refer or figures and Tables in the discussion, the results should have been clear enough.

-L696-700: "relies on" would be more correct than "highlights".

-L705: there is no such thing as "most significant" with a given significance threshold in statistics... replace by "strongest structure".

-L714: I am not convinced the simple longitude analysis allows to say that you have controlled for spatial effects.

-L721: "substrata" is a Latin plural, use substratum or, better, substrate, as you used elsewhere.

-L765 and L773: I am not convinced this has been shown.

-L785: evidence is never plural.

-L788-L792: not sure this is the best conclusion given I am not convinced there is local adaptation.

-L794: a conclusion is missing that would explain the interest of the study for scientists not working on mussels.

Reviewed by **Thomas Broquet** ⓘ, 16 April 2018

Comments .

I had two main comments on the first version of this manuscript. The first one was criticizing the interpretation of the results in terms of local adaptation. This comment was partly due to the fact that I took phrasing such as "genetic-environmental association" for indications of local adaptation, but this whole aspect of the study is now clearer in the revised version.

The second comment called for an improved presentation of the introduction and discussion sections, and I find that this can still be improved in places (as detailed below).

I find that the new analyses presented in this revised manuscript help discussing the admixture vs incomplete lineage sorting hypotheses, and the newly performed simulations also help illustrating how the fine scale patterns observed could theoretically be produced in absence of local adaptation. However, these new analyses come at a cost: they add to an already complex set of analyses, and one of the new results (allelic frequency spectrum, dadi) calls for more caution in some of the interpretations. Hence below I suggest some places where the text could be streamlined and I also comment on the results from dadi. Finally, I have two new comments on aspects that I did not notice in the first round of review but that probably call for some clarification.

Overall, all comments on this revised version call for relatively minor text improvement, and I think that the main message (impact of reticulated evolution on present fine scale genetic structuring) is of interest to a wide audience.

1) It is difficult to follow what is called the "Southern lineage" (L. 659-660) given that most analyses focus on the genetic structure of Kerguelen mussels relative to Northern populations. Samples from the Southern Hemisphere were used in some of the analyses, but these samples are not described, and it is not clear why they were not included in the main analyses. The first mention of such samples (I think) is on line 259: "the Chilean mussels", but at this stage of the paper we don't know what these samples are and why they are used here and they were not used e.g. in the genetic network analysis. I don't mean that these samples should necessarily be included in these analyses, but their role in the analyses and interpretation could be clarified.

2) The genetic break observed between sites PAF and RdA is interesting, and its interpretation and comparison with simulations suggest that the scenario proposed in the paper is plausible at least. However, I suggest to be slightly more cautious in the interpretation for the following reasons: first, there are no confidence intervals on allelic frequency estimates in Fig. 4B, and it is expected that average frequency at all loci but four (empty

symbols) will be less variable that the average frequency at four loci (filled symbols). Hence the difference in variation around allelic frequencies is not a strong argument (contrary to L. 544). Second, there is another possible (not discussed) break around site "BOBO". I could not find this site on the map, but it seems that this break occurs within a really small geographic distance, so small that the barrier-to-dispersal scenario may not hold there (it could be random error around the allelic frequency estimate linked with sample size). So overall, I would generally more cautious in the text.

3) The dadi analysis suggests a scenario of ancient migration, not secondary contact. It still means that past admixture has played a role, but then the text needs to be clarified because it often refers to "secondary contact" (e.g. see simulation conditions), or "secondary admixture", "secondary genetic exchanges", etc. So you may want to clarify throughout whether your interpretation is that of a secondary contact, ancient migration, or unresolved past admixture that could be one or the other (or both), and why trust more one approach over the other.

Other specific comments:

L. 87: "local" where?

L. 112: are those mussels M. platensis?

L. 118: this sentence is really difficult to understand without reading your other published work (Gerard et al. 2015).

L. 139: "for many molluscs": add "including mussels" with ref. Otherwise, we wonder throughout the paper what do Macrocystis seaweeds have to do with mussels.

L. 149: point ii) has not been introduced so far, so I suppose it is going to be difficult to understand at this point of the text. This could be improved at the beginning of the intro.

L. 153: what handful of markers?

L. 154-171: long summary of the results. Since the text is long and complex, reductions would be welcome and this is one place where this could be done.

L. 219: seven loci out of 44, that seems like a lot of errors between the two sequencing methods. You could add somewhere that quantitative results for the comparison of the two genotyping methods (I mean in general, not for the Mytilus case) will be welcome.

L. 258: "we defined…" : difficult to understand why at this stage, and what are these Chilean samples. Same thing L. 341.

L. 353: what is a "bi-locus haploid genotypes at a barrier locus"?

L. 375: what are those samples and why were they not used in genetic network analyses? I don't mean that they should be, but that this aspect could be clarified in the text. Figure 1: "internal" and "external" denominations: not very intuitive.

L. 413: "three migration edges": on Figure 1 there are four. And 50% seems like a very permissive threshold. A 50% supported edge seems not significant. Can you justify this threshold? Or at least take it into account when discussing the results?

L. 453: this whole analysis says that incomplete lineage sorting is important to consider. It does not seem to bring any support to any of the scenarios tested, so it could probably be sent to entirely to sup mat (but keep the emphasis on ILS in text of course).

L. 527: mention here that these loci are ancestry-informative.

L. 544: isn't that expected just because the number of loci is not the same?

L. 544-546: see first comment above: where does that result come from? Where is the associated methods described (perhaps I just missed it)?

L. 566-571: already in the methods.

L. 588-591: I could not follow this sentence.

L. 592: "sharp". It looks sharp, but what is the uncertainty associated with these frequency estimates?

L. 601: predicted by what?

L. 604: what signal? Do you mean that there is an adaptation signal but that it is not visible?

L. 615: "surprisingly": why is it surprising? If you consider that you have two genetic backgrounds then it does not seem surprising that the allelic frequencies are roughly balanced when mixing the two types of individuals.

L. 617: local adaptation again seems to be the null hypothesis.

L. 659: I don't find clear where this is demonstrated.

L. 665-669: I could not follow this sentence.

L. 679: 51% is not "most". Plus this is 51% of 17%. That means essentially no support except for ILS. So I would be more cautious here.

L. 708: this part is difficult to follow. The local genetic structure (between sites within the island) does not depend on the global population size of the island (it depends on drift within local populations more or less connected by gene flow).

L. 733 what could be the effect of habitat on connectivity?

L. 765 to the end: a large section on local adaptation, whilst that hypothesis is not the most plausible. Yet it would be interesting to discuss the case of marker Glu in the light of the results of the present study. Is it or is it not affected by local adaptation?

Evaluation round #1

Authors' reply, 12 April 2018

Download author's reply
Download tracked changes file

Decision by **Marianne Elias**, posted 12 April 2018

Revise

Dear Dr Fraisse,

Your manuscript has now been assessed by two reviewers, Tatiana Giraud and Thomas Broquet. While both of them found the paper interesting, they also have important comments that need to be addressed before the paper can be reassessed for recommendation. Notably, Tatiana Giraud questions the ability of the methods used to discriminate between introgression and ILS, and Thomas Broquet is not convinced that the pattern of differentiation detected indicates local adaptation, rather than a physical barrier to dispersal. In line with this, Thomas Broquet also suggests changes in the focuses and structures of introduction and, to a lesser extent, discussion.

Both reviewers also have other minor points, particularly about clarifying the methods, which need to be addressed.

On the BioRxiv site, as noted by Thomas Broquet, the supplementary files do not correspond to the supplementary files mentioned in the MS (file S1, S2 and S3): a pdf file called 'Supplementary Information bioRxiv' and a xlsx file called 'Supplementary Information table'. This xlsx file cannot be opened.

Please make available all necessary information with a clear correspondance between the MS and the supplementary files, and please provide a correct table.

In addition, make sure that

-all the data are available to readers.

-all details of the quantitative analyses (e.g. data treatment and statistical scripts in R, bioinformatic pipelines scripts, etc.) are available to the readers, as appendices or supplementary online materials

Good luck with the manuscript revision,

Marianne Elias

Reviewed by Thomas Broquet ⓘ**, 10 January 2018**

This paper presents an empirical study of mussel populations (Mytilus spp) in the Kerguelen island, exploring the role of past admixture events on the local genetic structure of mussel populations around the island. Using a combination of genotyping-by-sequencing and target SNP genotyping, the authors dissect the complex divergence/admixture history of worldwide mussel lineages, infer the most likely origin of the Kerguelen population, and show that the local genetic structure is linked with past admixture events between native and non-native mussels.

How complex histories of reticulated evolution may affect current population genetic processes is a very interesting question that bears upon many important research topics (local adaptation, speciation in general and ecological speciation in particular, molecular analyses of connectivity…). The Mytilus complex is a good model species to tackle this question, and the dataset used here is adequate (the combination of worldwide and local SNP data is most interesting). The subset of SNPs used to look at the local genetic structure is small (33 loci), but this data set convincingly demonstrates that past admixture events have effects on the local micro-geographic structure. This result is, in my opinion, very interesting for a large audience in evolutionary biology. The paper is clearly written (although the methodology used is complex and could be better explained in places). For all these reasons I believe that this paper will make a very useful contribution in evolutionary biology.

I found, however, that there is perhaps room for improvement in two aspects (first concerning the interpretation of the results, and second how the introduction and discussion are structured). As detailed below, my comments call essentially for some clarification and perhaps a better structure of the presentation of scientific arguments and rationale.

1) The authors not only conclude that past admixture has effects on local genetic patterns but, more precisely, that it facilitated or at least impacted in some way local adaptation processes. The methods employed to reach this conclusion are thorough, and the investigation is interesting, but I don't think that the results are strong enough to reach a firm conclusion concerning local adaptation (yet).

Alternatively, as recognized by the authors in several places, the genetic break observed at four loci could be due to a genome-wide reduction in gene flow, with consequences visible only on a subset of markers that are linked with reproductive isolation genes (Fig. 3 in Gagnaire et al. 2015). In that case, the genetic data indicate a possible physical barrier to dispersal between sites RdA and PAF. This seems consistent with several observations:

i) the geographical structure of allelic frequencies at the four loci after the RdA/PAF break seems nearly linear, i.e. compatible e.g. with isolation-by-distance inside the Gulf of Morbihan. There is no obvious reason for local adaptation to result in such a clinal distribution, unless habitat conditions themselves vary along a geographic gradient. Can the authors comment on this?

ii) The Rda site, which shows the lowest foreign allele frequency, is not occupied by Macrocystis kelp. This seems to oppose directly the local adaptation hypothesis, which is based on the observation that foreign alleles are more frequent in habitats characterized by the absence of Macrocystis (text p.22 and Fig. 4B, grey dots). Moreover the authors mention p. 18 that "these two sites [i.e. RdA and PAF] differ at all five ecological variables". While I could not open Table S3 (format not recognized, for some reason), it seems from Fig. 4B that these two sites indeed differ in presence/absence of Macrocystis, but NOT in the direction predicted by the local adaptation hypothesis at most other sites . If this is true for the other environmental variables as well, then the fact that the two most differentiated sites differ in all ecological variables goes quite strongly against local adaptation.

iii) "water masses between Gulf of Morbihan and North coast do not mix well" (p. 18), suggesting that a physical barrier to dispersal may exist, even at such a small spatial scale.

iv) I did not find table S6 in sup mat, but from the text it seems that there is some genome-wide differentiation between several sites (e.g. PCu), compatible with short-scale spatial heterogeneity in neutral gene flow (this is not a strong argument, but it goes in the same direction: there could be some heterogeneity in dispersal-driven

11

connectivity between local sampling sites at the scale of the island).

On the other hand, obviously the hypothesis favoured by the authors (i.e. local adaptation: see end of abstract and end of introduction for instance) is also based on interesting observations. And perhaps that all of my points above are refuted simply because there are no loci involved in reproductive isolation (a point that I couldn't quite get clearly from the paper). While I am not sure that these observations are strong enough to conclude, I think that what is important here is the discussion of potential effects, rather than coming to a definitive conclusion. My following point deals with the structure of the paper.

2) In my view the main topic (and the strongest result) of this paper deals with the role of admixture effects on current population structuring, with possible effects on the detection of barriers to gene flow either due to physical barriers or local adaptation. These important topics are well discussed in the Discussion section, but their presentation in the Introduction could be improved. The Introduction starts with one general paragraph on local adaptation in marine environments, and then almost all of the intro details the Mytilus case, except for a few sentences on adaptive introgression near the end. I suggest that a more general presentation of the potential effects of admixture (role of ancestry-informative loci, difference with incomplete lineage sorting) is needed early in the introduction, including effects on signatures of connectivity breaks and links with (or even facilitation of) local adaptation. This would allow the introduction to end up with more formally hypothesis-driven objectives (e.g. no effect of past admixture vs detection of connectivity break vs local adaptation) - this is left to the judgment of the authors.

Accordingly, the Discussion could also better partition arguments for or against each of these hypotheses (surprisingly, the authors do just that in some instances, but generally in the Results section: e.g. see very good sentences such as the final sentence of the first paragraph of p. 17, and final sentence of first paragraph of p. 18). My point is perhaps best illustrated by the following example: The second-last sentence of the paper reads "Possibly these markers simply better reveals a genome-wide signal of habitat constrained connectivity". This statement contradicts the point apparently preferred throughout the whole article, namely that local adaptation is necessarily involved (e.g. last sentence of the abstract, last sentence of the intro…). I think that the discussion would be clearer if it would expose the arguments for each hypothesis, and then conclude on what the authors think is happening.

Other minor comments

Briefly explain why different filtering options were chosen for different analyses (end of last paragraph in supplementary methods)

In the same paragraph: What is a "maximum a posteriori genotype"?

Lisibility of figure S5A could be improved (red writings are difficult to see)

I could not find any of the supplementary tables.

Post-Scriptum: reviewing papers without line numbering is tiresome…

Reviewed by Tatiana Giraud, 10 January 2018

This study reports genomic analyses of mussel populations for investigating introgression and adaptation. Overall, I found the manuscript interesting, although unclear in many places, with even inappropriate wording in some cases (see below for some examples). There are too many confusing details while it is difficult for a reader not familiar to the system to get a global understanding. Other than that, my main concern is that I am not convinced by the interpretation that introgression is more likely than incomplete lineage sorting. These two phenomena are notoriously difficult to disentangle in closely related species as is the case here. Incomplete lineage sorting is actually inferred on P13, rendering even more confusing the interpretation of introgression. If the authors think the methods they used can reliably infer introgression and refute incomplete lineage sorting, they should explicit why and how.

Please find below more specific suggestions (lines were not numbered so it is not easy to refer to where the comments apply exactly):

-Abstract and elsewhere: "proto-edulis" should be "proto-M. edulis", but actually this view (as well as the "M.

edulis-derived" formulation) is biased and incorrect: a species divergence is symmetric and you do not know whether the ancestral species was more M. edulis-like or more M. platensis-like or even to any of the two.

-Abstract: "Southern lineage haS"

-Abstract and elsewhere: the number of contigs is uninformative without their size. Anyway there are too many details in the abstract; There should be less details and instead a more general conclusion at the end, explaining what this study brings as general insights beyond mussels.

-Abstract and elsewhere: "ancestry-informative SNPs or markers": could they be instead SNPs more affected by selection? And/or could this choice of highly differentiated markers bias the results?

- Abstract and elsewhere: "the Kerguelen is a divergent lineage": the Kerguelen is an island, not a lineage.

-P2: "evidence that local recruitment occurs in marine environments" does not "highlight the role of high population size", it only refutes "high levels of dispersal".

-P2: taxa cannot be "semi-differentiated" (differentiation is quantitative): they are either "partially-reproductively-isolated" or "weakly differentiated" or "with heterogeneous differentiation along their genomes"

-P2: "produce hybrid zones": replace "produce" by "display" or "show"

-P2: "encouraged to consider them as different taxonomic entities": who are "them"? And different from what? This does not seem contradictory to be "related to" M. edulis or M. galloprovincialis -P3: replace "unique" by "single"

-P3-4: there are too many details, dealing with just a few markers, while it is difficult for a reader not familiar with the system to understand the global picture. In particular, it is unclear whether the existence of "reproductive isolation genes" has been shown. One first question of the study could be whether the pattern found using just a few markers holds at the genome-wide level. In addition, it is not clear why incomplete lineage sorting is not considered here, while it seems a more parsimonious hypothesis.

-P4 and elsewhere: Macrocystis is mentioned in several places in relationship with local adaptation, but it is unclear what is meant here: competitive interactions? And why looking at only this organism presence regarding local adaptation? Aren't they any other important competitor, predators, parasites...?

-P4: how was introgression distinguished from incomplete lineage sorting in this previous study (Fraisse et al 2016)?

-P5 and elsewhere: KASpar, GEA, FCT... there are many undefined abbreviations all along the manuscript

-P5: explain better the "target enrichment sequencing": enriched in what and how?

-P5: explain if and how you could disentangle introgression from incomplete lineage sorting.

-P6 and elsewhere: FST is a mathematical index, it should be written with only the F in big capital and ST in subscript capitals

-P6 and elsewhere: "trossulus" should be "M. trossulus" and idem for other species names

-P6 and 7: "GBS individuals": replace by "GBS-typed individuals" (GBS is not a trait of individuals)

-P7: Fis should be written with the F in big capital and IS in subscript capitals

-P8: give the mean contig size and whether we know whether they are unlinked

-P9: explain better what is an RDA

-P10: M. platensis should be introduced from the beginning

-P11: replace "more" by "most"; not sure "edge" is the most appropriate word?

-P12: not sure "eventually" is the most appropriate word? "M. edulis-derived" is a wrong formulation: the bifurcating node between M. edulis and M. platensis is symmetric, there is no reason to call the ancestral node "M. edulis" (same comment P13).

-P13: replace "already involved" by "has been previously suggested to be involved"

-P16-17: there are many details that seems rather anecdotal, it is difficult to get an interesting take-home message on the Kerguelen structure results.

-P17: explain also briefly here what is a redundancy analysis

-P20 : not sure "sample" is the most appropriate word? "divergence between [mussel populations] in the two hemispheres"

-Discussion: it is too long, detailed, too much focused on the model. Try to highlight the main messages better and why your study is of general interest beyond the model species. Avoid citing Tables and Figures in the discussion (all results should have been presented in the result section already).