Dear Drs. Castilho,

First of all, our apologies for such a long delay since the last revisions. We have invested this time in completing the datasets, particularly at the male mitogenome, among the *Macoma* species complex, enabling us to clarify the phylogeographical context in which we interpret the discordance in the genetic structure between the male and female mitogenomes, as requested.

Again, on behalf of all authors, we would like to thank you and the two reviewers for the very useful input to correct and improve our manuscript. In the following document, we will respond to each comment in blue.

Best regards,

Sabrina Le Cam and Eric Pante (corresponding authors)

Round #2

---

by Rita Castilho, 09 Sep 2022 11:21
Manuscript: **https://doi.org/10.1101/2022.02.28.479517** version Version 3

Another round of revision towards an improved manuscript

Dear authors,

I am sorry to have taken so much time to process this round of reviews. The summer is always a bit tricky to get these things rolling!
As you can read both reviewers have devoted time and attention to in-depth re-reviews. I think your manuscript merits the chance for another round of improvements, hopefully, the last...

Regarding reviewer 1 suggestions, can you particularly address the circularity of μ to get at Ne, since theta = 2Neμ where here Ne represents the relative abundance of individuals carrying M and F genomes. Also, model testing should probably be included to compare the population structure of the two genomes. Should you decide not to include it, you should at least justify it.

We removed these analyses and kept the discussions of Ne estimations to a minimum ; we looked at relative differences in effective population size between both markers using the Watterson/Nei theta ratio among genetically homogenous groups of samples.

We have attempted ABC analyses to explore demographic and historical parameter explaining the current state of the genetic diversity distribution at both mitogenomes but the lack of variability at *cox1m* in the southern sampling prevents us from confidently interpreting the results. We are currently gathering more mitochondrial and nuclear data for both

mitotypes (through targetted capture of all mt and nc genes involved in mitochondrial function), and hopefully we will be able to further test the process(es) underlying the discordance observed.

We have better addressed the discordance in population structure using cline analyses (including cline model selection with HZAR) to characterize the localization and the strength of the barriers to gene flow in M- and F- mitogenomes (details below).

Reviewer 2, has a lot to suggest, and I would highlight the possibility of estimating expansion time and address the observation that the analysis of diversity statistics is largely oriented to comparison under the expectation of neutral equilibrium, which does not seem to be the case.

Reviewer 2 proposed to clarify the evolutionary / phylogeographic framework in which to analyze and interpret our data. We have addressed this by adding samples from Pacific and Baltic lineages that were not represented in the previous version of the manuscript. This allowed us to more rigorously attempt to date the divergence between the southern and northern rubra lineages, and show that with the data at hand we cannot refute secondary contact between them. We detail these analyses below.

We also re-organized the discussion as to first interpret and discuss the results of the phylogeographic analysis and the non-equilibrium state of the analyzed populations, and then interpret patterns of genetic diversity in this context, in DUI species. Details are provided below.

Of course, there are many more instances deserving your attention and consideration, and I hope you can address them all in a justifiable and thoughtful manner.

With best wishes,

Rita

Reviews

Reviewed by John Wares, 02 Aug 2022 20:02

I was glad to read the latest version of "Discordant population structure..." by Le Cam et al. I note the update on taxonomy, which explains my confusion in the last version.

Thank you!

I still find it problematic how they are attempting to estimate μ to get at Ne, it is all a bit circular because of course theta = 2Neμ where here Ne represents the relative abundance of individuals carrying M and F genomes. So by estimating μ not from a distant relative - a concern I had previously - but by the greatest divergence within their data set, they are incorporating the population structure that they note varies between the two genomes. So,

avoiding the saturation problem of using phylogenetic substitution patterns but a very messy problem with trying to evaluate the ratio of Ne of the two mitogenomes using haplotype diversity. Effectively, the ratio of theta - which can be estimated with pi or with haplotype diversity - is sufficient and not much more can be gleaned from these data in that regard because the diversity itself is not representative of a single panmictic population.

The effective size ratio estimation between both mitogenomes proposed following methodology developed by Ladoukakis et al (2002) was removed from the manuscript, given the exposed flaws when using the maximum divergence to estimate μ.

As suggested, we have theta S ratio, as presented L254-257 and L385-L387.

It does appear that for their haplotype diversity and other contrasts they are using trimmed mitochondrial sequence from the male data, so that is useful. The population structure story overall, and how it varies between the two data partitions - is still quite interesting. However beyond the results themselves the authors spend a lot of time attempting 'post hoc' description of how the data fit different models, when that could be more rigorously tested. This is a valuable contribution, it feels like the authors are trying to extend the inference beyond what is feasible with the available data however.

We removed attempts to estimate Ne differences between cox1m and cox1f and kept our discussion focused on theta ratios. Additional analyses (mismatch distributions and genetic cline descriptions) were added.

Reviewed by Risto Väinölä, 24 Aug 2022 13:33

I am pleased to see the authors have taken practically all the suggestions into account and dealt with the problem of relating the results to previously published data, and the associated biases. This is exemplary.

Thank you!

Pending those corrections, in the previous review I did not touch much on the population genetic/statistical analyses of the data, and stressed that those should be considered in the evolutionary/phylogenetic framework or context where the variation is supposed to have been molded. As the main directions to proceed from now: There is space and need to clarify the implicated biogeographical background; on the other hand some of the presented analyses do not appear very illustrative or sensible in this framework.

As detailed below, we have followed your advice to provide information on the phylogeographic framework of this study by (1) adding the phylogenetic context (therefore adding key lineages to the analysis), (2) providing estimates of divergence between lineages, especially between the southern and northern rubra lineages (following Nikula et al 2007 – Evolution), (3) carry out genetic cline analyses for the study population structure, (4) add mismatch distribution analysis and theta ratios, and finally (5) homogenized diversity illustrations to match such approach. Also, the MK test initially proposed to investigate patterns of selection linked to DUI were deleted. We also re-organised the discussion as to

present the non-equilibrium state of populations prior to interpreting differences between cox1f and cox1m in light of this result.

Introduction, M & M

[There i a marked copy of the beginning [only] of the pdf text containing minor linguistic remarks/suggestions to clarify the message, as well as comments. To pick up some:

Changes suggested in the pdf were taken into account, thank you.

About the overall distribution of *M. balthica*, as far as I know the range in NW Atlantic does not extend to USA (now "Virginia"), but the previous "M.b." there are currently attributed to *M. petalum*.

The distribution of *M. balthica* was modified to exclude *M. petalum* and references were updated, thank you. L95-98

M&M, from the explanation on 154 ff, it remains unclear to me which sequences/taxa were used for the primer design of M. balthica M genome. Capt et al. (2020) see below?

We added this text to the M&M L155-159: "Both mitogenomes published in Capt et al. 2020 were used and they came from *M. b. rubra* specimens sampled in Aytre, France (clade b1b clade). We also used *cox1* sequences retrieved from transcriptomic data (Pante et al 2012) from specimens sampled in Aytré, Somme Bay (France, b1a clade) and Gdansk (Poland, b2_b3 clade)".

For the previous evidence of DUI in Macoma, reference is made (only) to Pante et al. (2017). That paper however presented only a hypothesis, no compelling evidence? There would be more appropriate and compelling subsequent papers to cite? Capt et al. (2020) https://doi.org/10.1038/s41598-020-57975-y is missed by mistake?

Indeed, thank you for picking this up! The reference to Capt et al was added L84.

Study context, scenarios and analyses.The context is about a hybrid zone / contact zone. It should be most important to explain or speculate more closely on the nature and history of the French contact zone in the beginning.

L188-L209: while we do not speculate on the nature of the hybrid zone at the beginning (ie the hybrid zone being the result of primary intergradation or secondary contact), we specify that we have gathered data to do so in the manuscript: we have added specimens from clades (*sensus* Nikula et al 2007) A (*M. petalum*), B (*M. balthica* rubra), C and D (*M. balthica* balthica from NW Atlantic and Pacific + Atlantic, respectively). This allowed us to estimate the divergence time between rubra and balthica following the molecular clock and method published in Nikula et al 2007, confirm the divergence times calculated in the aforementioned paper with *cox1f* and propose a divergence time based on both *cox1f* and *cox1m* for the rubra / balthica split. Based on these data, the Finistère / Cotentin hybrid zone is most likely a secondary contact zone.

On 97-99 two contact zones are mentioned. The nature of the North Sea-Baltic zone as an Atlantic-Pacific inter-subspecies contact might be somehow decipherable, but the nature of the other, focal zone is now hardly explained at the outset. So please make explicit that (i) there is an inter-subspecies zone (not dealt with) and an intra-subspecies zone analysed [suggestions in text].

The section L98-109 (now L97-101) was modified accordingly, deciphering between the inter- and intra- subspecific hybrid zones. As in the literature, we use hybrid zone to describe a pattern, a geographical location were two genetically differentiated lineages mix, and contact zone to imply that the hybrid zone was caused by secondary contact.

(ii) explain what is the hypothesis on the origin/nature of the latter zone [the major context for interpreting data].

We now explicitly state L106-113 that we are testing the hypothesis that the origin of the Gulf of Saint Malo hybrid zone is caused by secondary contact. We underline, here and in the discussion, that distinguishing primary from secondary intergradation is difficult and will require more data (e.g. test of specific isolation scenarios with genomic data).

(iii) present the results (also) in the conventional ways of transition zone analysis.

We have done so by (1) adding the phylogenetic context (therefore adding key lineages to the analysis), (2) providing estimates of divergence between the southern and northern rubra lineages (following Nikula et al 2007 – Evolution), (3) carry out genetic cline analyses for the study population structure, (4) add mismatch distribution analysis and theta ratios, and finally (5) homogenized diversity illustrations to match this approach.

In general when studying hybrid zones we would imply either a secondary or primary transition, and by reference to a “contact” zone we’d assume secondary contact is the working hypothesis. If not, please directly put forward the alternatives (you only return on this on ln 465), and if yes, please base your examination on this scenario, be explicit about it and, in any instance, use the standard presentations of presenting clinal variation and hybrid zones.

We now start by providing the phylogenetic and phylogeographic context, to provide the working hypothesis that we are working on a secondary contact zone. As suggested, we present clinal variation and compare the fit of our data to genetic cline models. More details are provided below.

Supposing a secondary contact framework, among the most essential phenomena shaping the variation would have been isolation in allopatry, and introgression/expansion after that. The analyses however focus discussing three other relevant phenomena, population size, mutation, and purifying selection, largely not putting them in the (spatio)temporal framework of the isolation and introgression history. Sorry to say but without that framework it is hard and frustrating to evaluate the meaning of and interest in the mutation patterns.

We hope that the changes made to the presentation of the context and working hypotheses, along with the new analyses will make the comparisons of genetic diversity more

meaningful. Also, the re-organization of the discussion (phylogeographic context discussed first) should provide the necessary framework to interpret the mutational patterns.

The isolation/contact scenario is necessarily tied to time, which the present treatment tries to cancel out from the analysis (ignoring rates and ages of events). While there are uncertainties, there do exist concepts of these ages ("post-glacial", "prior to last glacial cycle", "one or more cycles") and even published estimates of molecular rates (for F type), which, even if not taken literally, do give the order of magnitude and basis for evaluation of what time scale the observed divergences and the generation/loss of singleton mutations do represent.

As suggested, we carried out additional sequencing to be able to interpret our results in a spatio-temporal framework. To do so, we calculated the net divergence between lineages of *M. petalum*, *M. b. balthica* and *M. b. rubra* to estimate the upper and lower divergence times as done in Nikula et al 2007 (methods L188-L209, results: L260-275), for the *cox1f* and *cox1m* markers. With these new data, we tested the hypothesis that the divergence between the *M.b.rubra* and *M.b.balthica* lineages predates the LGM, suggesting a secondary contact following vicariance, and subsequent expansion + introgression.

There is also the scale of divergence between the M and F genomes, entering MK analysis, which is still not commented on. Finally, in conventional population genetics a mutation is a convenient 'neutral' unit of time, if calibration goes out of question.

As stated above, we removed the MK test which was initially included in the study to show that the *cox1m* is under selective pressures just as *cox1f* is (as is discussed in the DUI literature). Because of the ambiguity between the study of the hybrid zone per se, and the evolution of DUI and selective pressures on f and m mitogenomes, we agree with the reviewers that the MK unnecessarily complicates the ms.

Since the value of the results in documenting the discordance of variation (in space and time) in the transition zone (cline), I would very strongly also recommend using some conventional ways of presentation

1. The standard way is to plot frequencies and statistics against geographical (shoreline) distance. Please do this for haplogroup frequencies, and other relevant statistics when illustrative (e.g. diversity); separate panels on top of each other for F and M. [now there is only Fig 4 of using this presentation but for a very obscure statistic.

The figure 2 has been modified to represent the results of the cline analyses.

As suggested, we plotted haplogroup frequencies (as well as gene diversity, nucleotide diversity and the proportion of singletons) against geographical distance along the shore, from our northernmost sampling point (Figures 2 and 3). We fitted cline models using HZAR to estimate hybrid zone width and center. Figure 4 was removed.

The 'very obscure statistic' corresponds to the proportion of singletons, or proportion of haplotypes represented by a single individual. These rare variants are interesting in this context because they are expected to decrease towards the distributional range limit

(southern sampling sites), unless southern sites constitute a glacial refugium (see Becquet et al 2012). They are useful to discuss demography and singletons are used in the formulae of R2 from Ramos and Rozas, used here to statistically detect population expansion. We made this more explicit L.362-365.

2. If implying an isolation-contact cycle, state or speculate (present a hypothesis) on the correspondence, or lack thereof, of the mtDNA haplogroups and refugia (are these refugial lineages – as implied in earlier work).

This is now made explicit in the first part of the discussion (lines 411-422), as presented above: We work under an isolation followed by secondary contact cycle based on the observation that vicariance occurred prior to the last glacial maximum, that the separation (most likely) occurred along the coasts of the English Channel, that populations now meet in the Gulf of Saint Malo, resulting in mito-mitochondrial discordance. Based on the high genetic diversity observed between the southern range limit of the species and the Gulf of Saint Malo genetic break, we do not refute the hypothesis of Becquet et al 2012 that the Bay of Biscay is a Glacial Refugium where clades m1 (at *cox1m*) and b1b (*cox1f*) occur almost exclusively. We stress that this is a working hypothesis to be further tested using more samples and genes (see above our capture experiment).

3. Make the analysis/discussion of differences in introgression (mixing) of haplogroups following contact [or before it] a part of your treatment.

The discussion of mitochondrial introgression was moved to the first part of the Discussion section, L. 489-518

Comments/suggestions on the current analyses and illustrations (partly technical)

Considering the series of pie diagrams for M vs F haplotypes in Fig. 1 and Fig. 2, it is confusing that F is on the left in one fig and on the right in the other.

The positions of the markers were homogeneized between both figures.

As to the scales of these figures (conventionally, circle area corresponds to haplotype frequency), evidently in Fig. 1 the scale icon does not correspond to the scale in the actual network.

The scale of figure one was corrected.

In Fig. 2 the scale in turn is not according to area but to diameter, please check and harmonize.

The figure was modified accordingly

I appreciate the choice of not directly implying a (historical) correspondence between the major F & M haplogroups and to use “a neutral” indexing for M haplogroups. Nevertheless, it is kind of tedious compare patterns if using index 1 for the southern F types but index 3 for the southern M type, instead of starting consistently 1,2,3 from the south (now hard to read Fig. 1 and 3).

The male haplotypes were renamed from 1.2.3.4 from the south

The diagrams in Fig 1 (center panel) and Fig 3 are in a way non-standard. While they do express a sense of discrepancy, they are not self-explanatory, and (i) at least for me the procedure of constructing the Fig. 1 diagram remains unclear and there should be an explanation of this 'algorithm' in the legend.

The legend of figure 1 was modified as such:

" Figure 1: Median joining networks constructed from the haplotypes found among the 313 individuals distributed in 17 sampling sites for (A) the cox1f gene and (C) the cox1m gene. Haplotype associations among haplogroups are represented with a tanglegram (B): Haplotypes are sorted by haplogroup for each marker and a line represent the connection of the cox1m and cox1f haplotypes within one individual. The color of the line represents the population where it was sampled and bold lines and arrows highlight individuals sharing distant haplogroups"

And Fig. 3 too. (ii) the information of discrepancy in those two diagrams is largely the same (altho more detailed in Fig. 1); (iii) as regards the discrepancy of M-F haplotype distribution in individuals, the "analysis" in Fig. 3 is not very informative if it is not contrasted with any expectation (which evidently would depend on the geographical/contact context), and (iv) if intra-population correlations are not distinguished from those caused by the pooling of differentiated populations.

The figure 3 was removed.

Indeed in the Results l 296 ff intra-population disequilibria are reported, a most interesting finding, also as there is evidently the looming idea in the background of signals of incompatibility in the M-F associations. But, in reference to this and the previous paragraph, there seems to be nothing about these analyses in the Material and Methods section! Please detail the analyses, and theis justification there, as appropriate.

The Discussion contains much thoughtful consideration of the observations (statistics) and of the nature of the contact zone, including references to hypotheses on refugial history and introgression, very relevant and clever. I will not go to that, but return to the ways data are treated and presented in the first place, which make the discussion unnecessarily tedious.

We have attempted to reduce the result part in the discussion to ease the reading.

In general the treatment and analysis of diversity statistics (Tables and further, see below) is largely oriented in comparing them with the expectation under neutral equilibrium. Unfortunately in this instance such 'tests' can appear misplaced and irrelevant. This is because of the context, in which such an equilibrium null hypothesis is not reasonable: then it is waste of space to discuss (indirectly through the statistics) why results do not fit such models. And worse, when there is not enough statistical power the null hypotheses are accepted even if the geographical and haplotree patterns would point to the opposite. A null hypothesis is of a unique independent population in mutation-drift balance (a state which cannot be attained in a postglacial time scale) and not affected by gene flow/introgression.

But interpreting your populations in this context, and deviations in terms of mutation/Ne/selection, is insufficient or unnecessary, since gene flow and time (unbalance) are equally important actors. While the interest is in the clines and contact, it (only) makes sense to examine the data in that context from the start.

Agreed. We have modified the presentation and data analyses to fit to this, and previous comment and more thoroughly considered clines and contact between the different lineages

You want to keep the MK test since that is a commonly used test; yet as applied it is not very informative. Particularly, comparing outcomes of tests from northern and southern populations separately (Table) does not make sense given the relative time scales of M vs. F divergence compared to North vs. South. They probably represent >100 fold different time scales, and no meaningful difference can be thought to arise in the latest 1% of time. The more general aim of seeing whether selection affects differently in M vs. F also is not addressed by this test, but the outcome just tells there is selection on nonsyn sites, which is kind of trivial. I suppose the analysis of differences in strength/pattern of selection is topic of other studies, not only of the 500 bp here.

The MK test was ultimately removed from the analyses as explain above.

Further, the approach to the Nef/Nem ratio (215-234) appears a strange combination of misunderstanding and misinterpretation.

Considering the formulae, please note that this is (only) about the basic equilibrium expectation of IAM for haploid data, i.e. we expect diversity [theta] = Nu, and more complicated formulae here are just spurious. [the effective no. alleles is *by definition* ne = 1-1/[theta] (or 1-1/H). Substituting that is just to make the formula look more complicated, it is not needed!]. At equilibrium we expect the observed diversity H=theta and you would have Nf/Nm= Hf/Hm (or Nf/Nm=Hf/Hm*u/u).

(i) now note, we cannot assume any of the (individual or combined) populations to be in mutation-drift equilibrium, so the approach does not fit here for estimating Ne.

(ii) the justification to using the u/u term was that there would have been differences in the F vs M mutation rates (ln 225). I would not deny that (see below), but confusingly you do yourself, stating that there were no differences (254). Why then this game.

(iii) the reasoning for the equal mutation rates can hardly be followed, also does not make sense to me. Further the nature and implementation of “relative rate test” remains unclear here [did not check the github though]. [is it that in this single instance you take the deepest divergence in the network (“TMRCA”) as the sole measure of divergence and divergence rate, assuming it should be of same age in both.

We removed these analyses altogether and restrict our treatment to the ratio of theta, as was also recommended by reviewed 1 (addressed above).

Instead of spurious testing, it could be more reasonable to use some common “at face value” approaches to phylogeographic haplotype data in evaluating the diversity results:

“Both F&M haplotype networks involve three core haplotypes, each associated with a set of one-step satellites (singletons, mostly).”Without making explicit matches of FvsM cores, we see that the F cores are 3-9 mutations from each other, while M cores are 1-2 steps. Supposing that at least part of the inter-core differences represent refugial lineages, it seems that F lineages either evolved faster or are older. [yes there is an odd more distant M haplotype (isn’t it M. b. balthica?); whatever, involving it in age/rate comparisons should need a separate discussion/justification]. You may use other distance estimates, qualitatively the result should be the same. But using populations diversity estimates such as FST, they will be strongly affected be population mixing and introgression that confound the results. - This appears very different from the message in the Abstract?

This approach (distance between core haplotypes) was incoporated in our analyses when calculating the net distance (which accounts for both within-lineage variation and between-lineage differences) to estimate the timing of cladogenesis of different lineages. The phylogenetic tree of the *M. balthica* complex for the M mitogenome reveals that the more distant haplotype is part of the m4 lineage within *M. b. rubra*, as discussed in L.402–410.

Now that the phylogeographical context is clarified, population diversity estimates such as Fst combined with clinal analyses provided very interesting information of introgression and population mixing, in particular sex-specific patterns which is rather uncommon in marine invertebrates: It is discussed in L465-479 and 489-509

In phylogeography, it is a useful and justified approach to interpret the core-satellite star phylogenies as indication of population expansion after a bottleneck. That involves an extreme non-equilibrium situation. Then (as also implemented in mismatch analyses) you may roughly estimate expansion time (in mutation units) from the frequency of the satellites (~frequency of singletons). Conversely, if you have estimate of mutation rate, you can estimate expansion time. I recommend this approach. There should be many examples; Laakkonen et al (2013) BMC Evol Biol comes to mind (end of page 8, and 13-14). - Here, pooling the cores and satellites, you will have c 10% mutated haplotypes in F and 15-20% in M, implying that either effective (neutral) mutation rate is 1.5-2 fold in M, or M expansion was older. Clearly this is in contrast to the result from inter-core differences. (it is close to the ratio of singletons). Combine this with the geographical approach of plotting clines/statisics on shoreline distance, and pick up the peculiarity that introgression of the b2a to the Bay of Biscay was, exceptionally, mainly by single b2b haplotype (potential indication of selection].

We have carried out mismatch analyses which confirmed recent population expansion a Cox1m at least and adressed the potential effect of non- equilibrium state of the conclusions L579-586.

However, the relative frequency of core vs mutated haplotypes is roughly the same (23.5% in *cox1f* and 25% in *cox1m*) which indicates either a similar effective (neutral) mutation at both mitogenomes or, considering cladogenesis of southern and northern *M.b.rubra* clades was later in *cox1m* (30-50 ky) compared to *cox1f* (106-85 ky), this could be the result of a more recent expansion of *cox1m* “counter-balanced” by a higher mutation rate?