Dear Drs. Castilho, Väinölä and Wares,

On behalf of all authors, we would like to thank you 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 #1

by Rita Castilho, 29 Apr 2022 12:09
Manuscript: **https://doi.org/10.1101/2022.02.28.479517**

Space segregation of mitochondrial male and female lineages of an intertidal bivalve

Dear Authors,

Thank you for submitting your work for a recommendation at PCI Evol Biol.

Two reviewers and I have assessed your work. The recommendations from the reviewers are are-level and well balanced and while agreeing on several points, also point out distinctive aspects. We all agree that the data presented has interesting aspects: comparative DUI studies of geographical variation are scarce, and the chosen model allows pinpoint parallels and contrasts in other European marine taxa. This study will significantly contribute to the marine spatial pop gen and DUI literature if the issues raised by both reviewers are dealt with in detail.

Both reviewers' constructive concerns are valid and well-argued, and therefore should be fully rebutted.

You have written a very nice and broad introduction that will help the general reader be up and running on the DUI-system peculiarities. The most problematic points are methodological:

1. The reviewers and I agree, question the presentation (or the estimation) of some population genetic statistics in the context of the manuscript MK test, divergence rate and population size estimates). Authors must reconsider focusing the results on the sex-specific spatial breaks, which are the most robust part of this paper.

Thank you. As suggested, we removed divergence rates and population size estimates from the manuscript. We better present the MK test, which is meant to test for adaptive evolution of the male mitochondrial genome in comparison with the female mitochondrial genome. As this test is quite standard in papers focusing on the evolution of the DUI system, we propose to keep it in the manuscript.

2. Correctness and standardization of the taxonomic denomination (*L. b. rubra* vs *L. balthica*?) and the haplotype designation (see reviewer 1 comments on those issues).

Since the first submission, we came across the publication by Nielsen 2021 (doi: 10.11646/ZOOTAXA.5052.4.7) demonstrating the unavailability of the name *Limecola*

balthica and reinstating *Macoma balthica*. In addition, we standardized the denomination as to make clear that we work on the *M. balthica rubra* lineage. As explained below, we followed Risto Väinölä's recommendation to focus exclusively on this clade and remove the few individuals belonging to *M. balthica balthica*. We use the sub-species names *Macoma balthica rubra* and *Macoma balthica balthica* as needed.

As requested, haplotype designation has been modified to match previously published results on the biogeography of *M. balthica.* Lineage delimitation have been standardized to follow Nikula et al (2007). We also provide the correspondence with haplogroups reported in Becquet et al (2012).

3. Correspondence of the lineages b1, b2, b3 and d between the F vs. M genomes seems to constitute a major source of lack of clarity between this study and previous studies.

As stated in the previous answer, congruence with previous results has been made, and lineage/clade has been made for the F mitogenome. We decided not to follow the b1/b2/b3 naming scheme for the M dataset for several reasons. First, lineages at the M mitogenome do not conform exactly with those of the F mitogenome. Second, cox1m is a locus completely independent from the F mitogenome; although we can use the same naming system for cox1f and cox3f, we thought it would be more rigorous to use separate names (as one would for a mitochondrial and a nuclear locus). Finally, the genetic structure beyond southern *rubra* being still largely unknown, we thought that proposing a separate naming system for M haplogroups would ease the naming of clades in the future.

4. The particulars of the calibration ages used seem inappropriate and must be re-addressed.

The arguments brought forward by both reviewers are compelling and we removed the work on calibration and estimation of past population sizes.

There are many more points raised by the reviewers, but these stand out as the most relevant and somewhat impact the interpretation and conclusion of the study.

Thank you. We address those below.

Reviews

Reviewed by Risto Väinölä risto.vainola@helsinki.fi Väinölä, 25 Apr 2022 13:39

The manuscript describes contrasting population structures in distinct female and male inherited mitochondrial lineages of the intertidal bivalve Limecola balthica (Macoma balthica) along the West European coast. These are interesting data in several respects. While DUI has turned out to be unexpectedly widespread in bivalves, and levels of variability and evolutionary rates between lineages have been reported, comparative studies of geographical variation are rare, and the data here are unusual

and significant already per se. On the other hand, the L. balthica complex turns out to have a complicated biogeographical history, which finds both parallels and contrasts in other European marine taxa.

The potential of the system in elucidating the evolution of isolation mechanism between hybridizing lineages or taxa is presented as an argument for the importance and interest of the data, whereas at this point any inferences about this importance cannot be made.

References to the potential role in DUI in the establishment of reproductive barriers among genetically distinct lineages were reduced from the introduction L74-76. This point is mentioned in in the discussion as a research perspective L484-488 & L536-540 & L546-548.

I would see that the raw data on the haplotype diversity and on the discordant transition zones as such could be important results worth publishing. In addition the ms presents a series of standard (and non-standard) population genetic statistics on these data, with their supposed evolutionary interpretations (e.g. MK test, divergence rate and population size estimates…), which to me however appear largely misplaced, irrelevant or erroneous in this context, and their presentation should be reconsidered. The point is that interpreting these issues from the data are (should be) based on a model (scenario) of the genealogical and biogeographical history of the lineage variation (isolation and invasion events and their ages), but the scenarios underlying the current treatment seem to be ad hoc, unclear and likely erroneous; this is related to ignoring parts of the published record and hypotheses of mtDNA variation in Limecola.

We agree that references for previous results of the *M. balthica* phylogeography were lacking and have worked on re-situating the data according to previously-defined genetic lineages. Details will be provided upon responses to other comments concerning similar subject.

The variation of F mtDNA has been well explored by several research groups also previously. As noted in the introduction, there are two main lineages & taxa in Europe, L. b. rubra (b-lineage sensu Nikula et al.) and L. b. balthica (d lineage). L. b. rubra is thought to have been resident in Europe through most of the Pleistocene (2-3 Myr), whereas M. b. balthica = d lineage is a post-glacial invader from the Pacific (c. 10 ky). The variation dealt with in this study is only that of the rubra lineage at least as concerns the F genome [and that should be clearly stated, and even acknowledged in the taxonomic denomination of the study subject, L. b. rubra rather than L. balthica]. The haplogroups (F) I, II and III in this study correspond to H4, H3 and H1+2 of Becquet et al. (2012, 2013) and to the sublineages b3, b2, b1 of Nikula et al. (2007), and it would be fair to the reader to use uniform nomenclature for clarity. The genealogy of these lineages is probably most clearly depicted in Fig. 3 of Nikula et al (2007); while that tree is based on cox3 haplotypes, the identity of each (sub)lineage is also reported in terms of cox1 and thus cross-validated with the current and other reports (of Luttikhuizen, Becquet, Layton, etc.).

To correct for the lack of clarity in the lineage and haplotype designation, we have decided to follow the nomenclature used in Nikula et al. (2007). To do so, a phylogenetic tree was constructed using the data of the present study together with previous Fcox1 haplotypes for the female mitogenomes from Nikula et al. (2007) and Becquet et al (2012, 2013) and presented in supplementary materials (Fig S1). Thereafter, b1, b2, b3 and d lineage designations were used in the manuscript. A phylogenetic tree was also built with the male mitogenome data (Fig S2).

The main problem arising here and likely undermining a large part of the "inferences" in this paper is the interpretation of the correspondence of the lineages b1, b2, b3 and d between the F vs. M genomes. From the published record (of other research groups) the "Pacific" d lineage (or C group of Luttikhuizen) is common within the Baltic Sea (+Barents & White), but absent from the European Atlantic coast. In the Baltic Sea it is mixed with b1+2+3, but towards the north (the Umeå area) it becomes absolutely dominant (> 95 %, Nikula et al. Fig. 1; also Luttikhuizen 2003). The unexplained / unreported problem is that in this study, and in the authors' previous larger Baltic and Barents Sea material (Becquet et al 2012, 2013), is that the dominant Baltic/Barents d lineage is not recognized or reported at all. Why it is so, is not, but should, be very clearly reported, and implications of the discrepancy be explained.

The primer set for cox1f used in the present study is different from the ones used in Becquet et al 2012/2013 as to amplify a larger portion of the gene (313bp in Becquet et al 2012, 393 bp in Luttikhuisen et al 2003; a minimum of 479 bp in our case). Although they were designed as to capture as much diversity as possible, they failed to amplify cox1f some samples (mostly from Umeå), despite amplification at cox1m (we removed all individuals for which only one of the two makers was amplified). This *may* explain the absence of lineage d in our samples (along with, possibly, poor sample preservation and sampling biaises); we are currently working on this problem and decided (as suggested) to remove the Baltic samples from our study.

As requested by John Wares ("Was not clear where the M and F primers came from, are they Limecola specific or would they work in other bivalves?"), we added details on the primer design (L149-157): "Custom primers were designed in this study to (i) target homologous regions of the cox1 gene, (ii) encompass the genetic diversity of *M. b. rubra* and (but possibly not integrating the whole *M. b. balthica* diversity) and (iii) most importantly to be specific to either the male or the female mitogenome."

Now indeed in this study there are only a couple of F haplotypes from the inner Baltic Umeå sote, neither of them d lineage! (But possibly a larger number were sequenced but not reported, of females?) A critical question is: if the authors have failed to detect/report the dominant Baltic F d lineage (M. b. balthica), on what grounds it is assumed that they would also consistently miss the (unknown) M d lineage of M. b. balthica; and why that question is not raised?

The original idea in considering samples from the Baltic sites in the study was to consider the widest possible divergence range available in our data. Nevertheless, since a possible amplification bias might occur at cox1f and cox1m when studying *M. b. balthica*, we have chosen to remove these samples from the study (as stated above)

and focus on the *M. b. rubra* lineage. We are currently working on building a more robust dataset to document the *M. b. balthica* M and F genetic diversity by characterizing mitogenomes from Baltic specimens.

From genotypic data (Nikula et al. 2008) there are no significant interlocus / mitonuclear disequilibria in the Northern Baltic and there is no reason to think the M and F haplotypes should be associated. Now it should be noted that the putative M I haplogroup here is almost exclusively reported from the Baltic (save one site in North Sea, info in Fig vs. Table is unclear about which site actually), a pattern rather to be expected for a d lineage. Given the data available, it would seem a more reasonable hypothesis that haplogroup M I phylogeographically corresponds to the d lineage (M. b. balthica) rather than to F I / b3, and the two main core haplotypes within haplogroup M IIa rather correspond to F I & II of current data (=b2+b3). This should of course be easily checked from pure Pacific M. b. balthica data. That cannot be required, but if not verified, the basis for most other statistics comparing the F vs. M variation in this ms will be lost.

This is indeed very interesting. The samples from the Baltic in our study are probably d lineage at the M-mtDNA and b3 lineage at the F-mtDNA. While we currently do not have the data to verify that precisely, some clue might point in that direction. Indeed, while amplification success was lower in our Baltic samples in general, we have had several individuals for which cox1m amplified and not cox1f. This is currently being investigated.

If no credible data on the phylogeographic identity of the M haplogroups can be given, the situation should anyway be acknowledged. The option might remain then to leave out the three Baltic data points entirely (and thus the M I / putative d lineage) and restrict the report to the Atlantic/North Sea data and transition zones.

Thank you for this proposal; we have chosen to follow it and remove the Baltic data points to focus on the Atlantic/North Sea area.

Another confusing issue of interpreting genealogies appears in the estimation of M vs F mutation or divergence rates through a comparison to corresponding DUI sequences in a pair of rather distantly related clams, Donax. The logic of the procedure and even of comparisons involved remain unclear, and should be explained by depicting the genealogy (tree) of all sequences involved and the branch lengths estimated. (What is the age of DUI in Limecola/Macoma by the way?) But it should be immediately evident also that using calibration ages c. 100 times older than the ingroup branches is not a viable approach in general, and that the substitution model used K2p (implied with uniform rates) will not be appropriate for such calibration but would yield rates an order-of-magnitude off the point. Indeed the ms involves three separate estimations of the substitution model which provide vastly discordant results, from nearly uncorrected K2p uniform to an extreme rate-heterogeneous model (gamma parameter 0.11-0.15), and these models do not correspond logically to the depths (ages) of the genealogy from which they were inferred, and are in no way commented.

As both reviewers (rightfully) formulated concerns about using the Donax/Macoma split to calibrate BEAST demographic analyses (age and uncertainty around the calibration points, saturation at cox1 at Miocene divergence points) and as suggested by Rita Castilho, we removed these analyses from the paper. Rather, we keep the ratio of male and female effective population.

We added in the paper that in the formula proposed for NeF/NeM, the ratio µM/ µF is independent of the time calibration point and can be formulated as subsitutionsM/site/year / subsitutionsF/site/year, ie subsitutionsM/ subsitutionsF.

As a footnote, we speculate that DUI predates the diversification of Macoma as we have found it in Scrobicularia plana (Capt et al 2020) with similar mitogenomic features (such as an exuberant insertion within cox2m found in Scrobicularia, Macoma and Meretrix; Tassé et al in 2022). However, a phylogeny for the Tellinidae (or the Tellinoidea) is lacking so this topic remains open at this point.

At the same time, it goes unmentioned that there exist alternative calibration approaches in the literature and on a more relevant time-frame based on the main phylogeographic scenario of trans-arctic invasions (putting the d-vs-b lineage split at 2-3.5 Mya): any rate and age estimates should be also compared to those [or rather those could be used exclusively]. (Luttikhuizen went wrong here, and dismissed the alternative calibration points from Mytilus and Acropsis, which are congruent with the Macoma trans-Arctic scenario and time scale also).

Thank you. Concerning calibration points, we are currently stuck because we do not have a “pure pacific” individual in our dataset yet (ie an individual from the d cox1f lineage and d-equivalent at cox1m) to propose F and M clocks for the d/b split, which would be ideal. In the absence of such data, and given the potential problem brought by saturation at cox1 for older calibration points, we decided to remove this analysis altogether.

As noted, many of the statistics in the tables now would not be biologically meaningful, in reference to the confusion of the phylogeographic scenarios / genealogies discussed above, and it makes no sense to touch or elaborate on them until the data basis and relevance of comparisons have been reconsidered.

To summarize, the scope of the study has been refocused on *M. b. rubra* and the scenario used to interpret the evolutionary implications of the discordance in genetic patterns between the M- and F-mtDNA have been rooted on existing knowledge of the biogeographical history of the lineage variation. We hope that these changes will correct the issues you had addressed in this review.

Reviewed by John Wares, 17 Apr 2022 16:31

Review of Le Cam et al for PCI. Very interesting manuscript, well considered though I have some concerns they can address. They used M and F transmitted mtDNA to

understand variation in structure and Ne, higher divergence in male, basically higher mutation, and maybe relaxed selection. Overall, a super interesting system – and they suggest may lead to greater rates of barriers to gene flow arising, eg speciation. A great introduction, and interesting that they are doing this in my old friend Macoma (now Limecola), no I've never published work on L. balthica but have encountered it. Nice introduction to remind everybody of the many quirks of DUI as well as structure in L. balthica.

Thank you very much for these comments.

I'm going to point out here that the mechanisms they used for estimating mutation rate *μ* (1) should be moved up in the manuscript, as they refer to 'known' *μ* well before the contrasts with Donax are listed (both species are Tellinids) and (2) are somewhat problematic in my opinion, as the divergence of Limecola and Donax are considered to be on the order of 100mya, but mitochondrial COI is thought to saturate mutations at 3rd position sites for even Miocene divergences, an order of magnitude less. Thus, I agree that the M rates of mutation appear to be twice as high (likely) but suspect the rates themselves are not as useful as the authors would like; perhaps separating by codon position would be valuable for considering the rate at 1st/2nd.

We have removed the estimation of mutation rates for M and F markers based on the Macoma/Donax split and the subsequent Ne estimation based on skyline plot analysis.

The authors sexed each under microscope. Found that male mitochondria only found in gonad, though they sampled mantle for somatic. I guess this is possible in Tellinids unlike Mytilids. Was not clear where the M and F primers came from, are they Limecola specific or would they work in other bivalves?

As presented above, here are the details of the primer design strategy: “"Custom primers were designed in this study to (i) target homologous regions of the cox1 gene, (ii) encompass the genetic diversity of *M. b. rubra* and (but possibly not integrating the whole *M. b. balthica* diversity) and (iii) most importantly to be specific to either the male or the female mitogenome.” As Risto Väinölä pointed out, they might even be specific to *M. b. rubra* ; we are currently more fully exploring their applicability to other *Macoma* lineages.

Basic methods for sequencing, quality, haplotype frequency, networks, calculated H, pi, phiST, AMOVA, Taj D – and they estimated the ratio of Ne between the two using a peculiar older “effective number of alleles” Crow & Kimura – seems that a Hill number more appropriate?

Crow & Kimura was used as in Ladoukakis et al (2002), with a correction of the formula as F- and M- mtDNA exhibit different mutation rates. This approach is interesting as a ratio of Ne's is calculated rather than individual Ne's. As Ne is notoriously difficult to estimate, our initial approach was to compare Ne for male- and female-inherited markers with different methods. We have removed inferences of past demography based on skyline plots since our calibration point is probably too old to be accurate.

We are unsure of how to apply Hill numbers (effective number of species) here but are open to suggestions.

How did they estimate male μ twice as high this is not listed (line 227).

Indeed, this information comes from the substitution rate estimated using Donax, showing its was twice as high at M-mtDNA thant at F-mtDNA. This μ section has been moved upward in the manuscript and modified accordingly.

. Method (ii) listed is basically dependent on Hill number as Hd is inverse Simpson, modified a bit. That method does not rely on μ to my knowledge?

The total equation was added to be clearer. From equations (i) and (ii):

$$\frac{N_{eF}}{N_{eM}} = \frac{\frac{1}{1-HdF} - 1}{\frac{1}{1-HdM} - 1} x \frac{\mu_M}{\mu_F}$$

Line 236, not sure I like this as it is very indirect, we all know strong selection on COI (though on male COI not sure how it differs).

The skyline plots hmmm an indirect way of estimating Ne from gene tree shape, fine – but they will be confounded by data that deviate strongly from neutral expectations of course!

OK now they get at μ on line 247, estimated between Limecola and Donax, with a divergence time of 90-140 My – I don't like this much, as COI tends to saturate at Miocene divergence. Maybe good enough for nonsynonymous; no separation of 1st, 2nd, 3rd.

As formulated above:

As both reviewers (rightfully) formulated concerns about using the Donax/Macoma split to calibrate BEAST demographic analyses (age and uncertainty around the calibration points, saturation at cox1 at Miocene divergence points) and as suggested by Rita Castilho, we removed these analyses from the paper. Rather, we keep the ratio of male and female effective population.

We added in the paper that in the formula proposed for NeF/NeM, the ratio µM/ µF is independent of the time calibration point and can be formulated as subsitutionsM/site/year / subsitutionsF/site/year, ie subsitutionsM/ subsitutionsF.

As a footnote, we speculate that DUI predates the diversification of Macoma as we have found it in Scrobicularia plana (Capt et al 2020) with similar mitogenomic features (such as an exuberant insertion within cox2m found in Scrobicularia, Macoma and Meretrix; Tassé et al in 2022). However, a phylogeny for the Tellinidae (or the Tellinoidea) is lacking so this topic remains open at this point.

Before I read results, it seems structure will be straightorward to see if distinct between the two for the same animals. The diversity is what is harder to struggle with – what mechanisms lead to distinct levels or patterns of diversity? – in any case they assume the male mitotype originated prior to the TMRCA, I don't know how typical the 'reset' in DUI has happened in Tellinids, seems more variable in Mytilids.

Results:

Far more haplotypes in coxlm-long than coxlf, but *it is longer* – here they should compare cox1m; by the same measure there can be more divergence among haplotypes as the sequence is 200nt longer. However it is still more haplotypes for cox1m (the shorter version), so it is robust but that distinction has to be considered.

The" long" version of the cox1m was originally considered to build the haplotype network and capture the maximum information on diversity and divergence at this marker as possible. Yet for the sake of comparison and because the results are similar with the shorter (479 bp) the authors have chosen to only consider cox1m (short 479bp) in the manuscript (haplotypes/haplogroups frequencies). The results of the cox1m-long analysis are presented in supplementary material.

Line 290 divergence rates substituions per site (s/s) not standard notation, again I think the clock evidence should be presented earlier – does seem more divergence about 2 fold in cox1m but again my concerns above. The actual rate is not known because the time of divergence is not known with certainty but the 2 fold ratio I can buy.

Divergence within the M.B. rubra lineage is used to estimate the number of substitution/site. It is used to estimate the ratio NeF/NeM. Therefor the time calibration point is not used.

Super intriguing to get distinct patterns between the two sex-associated mitogenomes…also interesting to see some combinations rare or absent, perhaps genomic conflict; and those odd haplotype combinations were in the hybrid zone sampling sites – cool!

Huh but π is greater in the F (lines 334-342), and D is more strongly negative in M than F. I would agree from Table 1 it is a consistent trend to more negative D values, so more rare alleles, thus the higher number of haplotypes seen above.

Yes, very interesting point and to corroborate this remark, the proportion of singletons with each sampling site (n>10) was added to the manuscript. It clearly shows a higher number of singletons at *cox1m* in sites at the south of the Finistère peninsula (L315-318)

Lines 363-372. Rem we know strong purifying selection, but ancient divergence you get many fixed nonsyn mutations between M and F

I still really don't get the Ne approaches, relative μ seems higher in M but I think estimated problematically and that could influence this. The higher number of singletons/rare alleles will of course suggest "exponential increase" but could also be stronger constraints on M.

As stated above, this part was discarded from the manuscript

Discussion

As noted in lines 406-409, the Male-type mtDNA diverges more rapidly and is far more polymorphic, "with more haplotypes, more haplotypes represented by a single male, and more segregating sites" – exactly, more singletons and rare mutations, which is what Tajima's D has picked up. Perhaps the higher mutation rate (still under the significant constraints of COI, eg. See work by Dave Rand).

Yes, as developed in the discussion, selection is certainly purifying for both F and M mitogenome, but slightly relaxed in the latter case: as hypothesized by other authors, the M mitogenome is less solicited than the F mitogenome (arena hypothesis of Stewart et al 1996).

Overall because I think there are concerns both about the estimation of μ and that estimators of Ne will falter when there is evolution strongly affected by non-neutral processes, I don't put much weight behind the estimators of Ne. The sex-specific spatial breaks are the most robust part of this paper, with pairwise Fst often being larger because of the greater number of private alleles I suspect.

Thank you. We changed the presentation of Ne approaches as outlined above, and emphasize that these estimators should be interpreted with care.

Overall, I really like the paper and think with some consideration of these issues it would be a great contribution both to the marine spatial pop gen literature as well as the really fascinating DUI literature.

Thank you !