Peer Community In Evolutionary Biology

Genomic structural variants involved in local adaptation of the European plaice

Maren Wellenreuther *based on peer reviews by 3 anonymous reviewers*

Alan Le Moan, Dorte Bekkevold, Jakob Hemmer-Hansen (2020) Evolution at two time-frames: ancient and common origin of two structural variants involved in local adaptation of the European plaice (Pleuronectes platessa). bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Evolutionary Biology.
https://doi.org/10.1101/662577

Submitted: 13 July 2019, Recommended: 03 April 2020

Cite this recommendation as:
Wellenreuther, M. (2020) Genomic structural variants involved in local adaptation of the European plaice. *Peer Community in Evolutionary Biology*, 100095. 10.24072/pci.evolbiol.100095

Published: 03 April 2020

Awareness has been growing that structural variants in the genome of species play a fundamental role in adaptive evolution and diversification [1]. Here, Le Moan and co-authors [2] report empirical genomic-wide SNP data on the European plaice (*Pleuronectes platessa*) across a major environmental transmission zone, ranging from the North Sea to the Baltic Sea. Regions of high linkage disequilibrium suggest the presence of two structural variants that appear to have evolved 220 kya. These two putative structural variants show weak signatures of isolation by distance when contrasted against the rest of the genome, but the frequency of the different putative structural variants appears to co-vary in some parts of the studied range with the environment, indicating the involvement of both selective and neutral processes. This study adds to the mounting body of evidence that structural genomic variants harbour significant information that allows species to respond and adapt to the local environmental context.

References:

[1] Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Molecular ecology, 28(6), 1203-1209. doi: [10.1111/mec.15066](https://dx.doi.org/10.1111/mec.15066)

[2] Le Moan, A. Bekkevold, D. & Hemmer-Hansen J. (2020). Evolution at two time-frames: ancient and common origin of two structural variants involved in local adaptation of the European plaice (*Pleuronectes platessa*). bioRxiv, 662577, ver. 5 peer-reviewed and recommended by PCI Evol Biol. doi: [10.1101/662577](https://dx.doi.org/10.1101/662577)

Reviews

Evaluation round #2

DOI or URL of the preprint: https://doi.org/10.1101/662577
Version of the preprint: 1

Authors' reply, 01 April 2020

Download author's reply
Download tracked changes file

Decision by Maren Wellenreuther, posted 09 March 2020

Evolution at two time frames: Genomic structural variants involved in local adaptation of the European plaice (Pleuronectes platessa)

Dear Alan,

Three referees evaluated your revised version and the responses provided. One of the referees made some suggestions for changes.

Once you have considered these suggestions (and the formatting requests below) and/or provided a response to justify not considering them, I will be happy to recommend your article for PCI Evol Biol.

Best regards

Maren Wellenreuther **Additional changes requested by the managing board**

Mandatory modifications

In order to reach a better referencing and greater visibility of your recommended preprint, we suggest you to do the following modifications:

(i) At the end of your MS you indicate "We will make available the raw data (Demultiplexed individuals fasta file) on NCBI SRA and the filtered data (the 3 vcf files including the overall dataset, the northern dataset and the southern dataset) on the DRYAD data repository"?. Please deposit these data and give the link to the corresponding deposit on DRYAD.

(ii) Authors must have no financial conflict of interest relating to the article. Your preprint must thereofre 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."

(iii) add the following sentence in the acknowledgements: "Version 4 of this preprint has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (https://doi.org/10.24072/pci.evolbiol.100095)"

 If you use bioRxiv to post your preprint, add this sentence also in a footnote of the 1st page of your pdf, it will be interpreted as a specific footnote section by bioRxiv.

Note that this DOI is not the DOI of your article, but the DOI of the recommendation text. The DOI of your article remains unchanged.

Doing so is very important because it would:

-indicate to readers that, unlike many other preprint in this server, your pre-print has been peer-reviewed and recommended. -make visible this information in Google Scholar search (which is quite important).

(iv) In addition, we suggest you to remove line numbering from the preprint.

Optional modifications

=> You can use templates (word docx template and a latex template) to format your preprint in a PCI style. This is optional. Here is the links of the templates:

https://peercommunityin.org/templates/

Please be careful to correctly update all text in these templates (doi, authors' names, address, title, date, recommender first name and family name ...). Please be careful to also choose the badge "Open Code" if appropriate (in addition to the "Open access", "Open data" and "Open Peer-Review" badges).

For word template, please be careful to correctly update all text in these templates (doi, authors' names, address, title, date, recommender first name and family name ...). Please be careful to choose the badges "Open Code" and "Open Data" only if appropriate (in addition to the "Open Access" and "Open Peer-Review" badges). If some of the reviewers are anonymous, indicate for example "Albert Ayler and two anonymous reviewer". Indicate in the "cite as" box the right version of your preprint. It is version 4.

For Latex template, main.tex and sample.bib should be filled. Please be careful to choose the badges "Open Code" and "Open Data" only if appropriate (in addition to the "Open Access" and "Open Peer-Review" badges). Preambule_xxx.tex should be modified (comment lines 115, 117) to select badges. If some of the reviewers are anonymous, indicate for example "Albert Ayler and two anonymous reviewer". In sample.bib, indicate the right version of your preprint. It is version 4.

We hope this is clear. Do not hesitate to ask any help if you need by contacting us at contact@evolbiol.peer-communityin.org

Reviewed by anonymous reviewer 1, 27 February 2020

After reading the revised manuscript and re-reading my comments, I think that this revision adequately addresses my concerns, and I'm happy to endorse the manuscript.

Reviewed by anonymous reviewer 3, 31 January 2020

The authors have made a good job of answering and fixing all my comments. I'm happy with this version. Congrats to the authors for a nice paper!

Reviewed by anonymous reviewer 2, 21 February 2020

Evolution at two time-frames: ancient and singular origin of two structural variants involved in local adaptation of the European plaice (Pleuronectes platessa)
Le Moan et al.

Overall, I think the manuscript has improved from the previous version. The objectives of the study are more clear and the flow/organization helps understand the context of the study better.

Title – The title wording is a bit strong. From the paper it is not definitive that both SVs are associated with local adaptation, and this has not be explicitly tested with environmental data. Also I'm not sure about the use of 'singular origin', given that it has not been confirmed that these are indeed SVs, and timing of SV evolution can be difficult given different dynamics of recombination and selection. While analyses suggest that timing of SVs are similar, the use of 'singular origin' seems a bit strong. Consider revising.

Abstract Line 12 – Are these confirmed to be SVs? Or are they putative SVs? This should be clarified in the Abstract/Introduction.

Introduction – Overall, I think the introduction is more balanced now with a better context for exploring population structure/history as well as structural variants.

Line 245-247 – Clarify wording here. Does this mean FST was calculated between the three genotype groups with just the SNP data from the SVs? I would remove the word "population" here, as it is confusing. Perhaps: "Using SNPs within each SV, pairwise FST was calculated between the three haplogroups identified by DAPC for each SV separately in hierfstat..." If that is not what was intended, then please clarify.

Line 252 – It was unclear initially what this FST represented. Move sentences (Lines 255-257) above to describe groups for calculating differentiation before discussing the quantreg R package (Lines 252-254).

Lines 259-260 – Were these visualized with heatmaps (what R package? Gplots?)? Also indicate that for each SNP "mean" pairwise LD between all loci along the chromosome was calculated.

Line 364-366 – This is useful to know. Provided this, I assume that a PCA using SNPs from both SVs, would produce 9 groups (and not 3 genotype groups) if they are independent? In some cases (translocation), combining data would should the same 3 genotype clusters.

Line 370 – Figure 2 – It would be useful to fit a line to these relationships.

Line 398-401 – Can this information about diversity in the SV provide information about the potential orientation of the SV? For example, if it is an inversion, is lower diversity expected in the rearranged orientation compared to the non-rearranged orientation?

Line 407 – Figure 3 – I'm not sure, but it might help to change the span parameter for loess in ggplot for these plots? It might track the changes across the chromosome better to do smoothing at a finer scale. I would indicate in the caption or text what span was used for the plots.

Line 403-405 – On Chromosome 21, it's possible that the loci in the other LD peak may actually be physically close to the primary SV. Perhaps an assembly of the plaice genome would help clarify this in the future. – I see later that this is addressed in the Discussion, but wonder if it could be somehow mentioned in the Results for clarity.

Line 427 – Indicate the name of these two genes here.

Line 436-438 – What are the time for haplogroups for each SV separately?

Line 519-520 – Indicate the date of this split here.

Line 575 – should this fsv19 subscript indicate "derived" as well?

Line 577 – So these are also 'derived' alleles? Not indicated, but states, the "same allele".. Based on Figure 3,S5 – it seems this is the case for SV21, but not sure that SV19 shows that same clear pattern. Allele frequencies in the North are not that different from allele frequencies in the North Sea/Kattegat.

Line 599-602 – I don't think this study identifies strong evidence of local adaptation associated with both structural variants. For example, previous studies found an association with salinity, but in this study, for SV 21, the frequency of derived allele in the Baltic is not different from other locations. This isn't clear from this sentence. In the case of SV 19, perhaps salinity could be a driver, as the Baltic has a higher frequency of the derived allele than other locations, as discussed. But without investigating the association with environment, it is not possible to determine whether these SVs are indeed associated with local adaptation to environment. And besides environmental features, like salinity, SVs can also be associated with life history variation, which may be the case here. Some caution to the interpretation could be added here.

Line 622 – the process "where" several ancient.

Line 626 – "repeatedly" rather than "repetitively"

Line 634 – I'm not sure about the use of 'singular origin' here. Perhaps just "suggesting these SVs evolved at similar times"? They likely didn't evolve at exactly the same time, which is what the term 'singular origin' would suggest to me.

Line 638 – In conclusions/perspectives, it's worth mentioning that confirmation that these are structural variants is still needed. And what methods would be used to do this. Long-read nanopore sequencing - to confirm that they are inversions and to identify exact breakpoints?

Download the review

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.1101/662577

Authors' reply, 14 January 2020

Download author's reply

Decision by Maren Wellenreuther, posted 28 September 2019

Revision: Evolution at two-time frames shape structural variants and population structure of European plaice (Pleuronectes platessa)

Evolution at two-time frames shape structural variants and population structure of European plaice (Pleuronectes platessa)
Alan Le Moan, Dorte Bekkevold & Jakob Hemmer-Hansen
https://doi.org/10.1101/662577 version 1

COMMENTS TO AUTHORS:

I now had three reviewers provide comments about the manuscript that has been submitted to PCI Evolutionary Biology entitled 'Evolution at two-time frames shape structural variants and population structure of European plaice (Pleuronectes platessa)'. This manuscript explores the population structure and variation in two structural variants (SVs) in the European plaice in the North and Baltic Sea. Previous work identified these two SVs on chromosome 19 and 21 and this study further explores this variation by incorporating additional sampling locations and attempting to date the SVs. The paper also investigates whether the SVs could have been introduced through introgression from another species that is known to hybridize with plaice. Overall, I think the paper is interesting, and this was also shared by the reviewers. Given the increasing awareness about the importance of SVs to population structure, I think this paper would be of interest to many readers.

However the reviewers have also highlighted a number of issues that need attention and have provided detailed and constructive comments below on how the manuscript could be improved. The major areas in need of attention are:

1. The general framework needs to be broadened, particularly in the Introduction, to include general details about why the authors investigate the population structure of plaice. Right now, the authors mostly focus in SVs in their Introduction.
2. All of the reviewers found that the Methods lacked critical detail and explanations (e.g. lack of details about sampling and dataset sizes). Please go over the comments that the reviewers have made on a point by point basis and clarify this section.
3. Reviewer 3 added some thoughts about the dating of the SVs, and possible problems with it. Further, it would also be good for the context and interpretation if the authors could provide a bit more detail about the genes that are located in the SVs, and to qualify statements like '...many of these were involved in ion transport' (how many?).
4. The reviewers felt that some of the statements were too vague, and were rather descriptive and lacked quantitative support, and I agree with that. I suggest the authors go over the manuscript again and qualify some of these (many-say how many, most of the times-how often?).

Sincerely,
Maren Wellenreuther

Additional requirements of the managing board:

Please ignore this message if you already took there requirements into consideration.
As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:
-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) 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) are 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 are available to readers in the text or as appendices.
-Authors 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." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders."

Reviewed by anonymous reviewer 3, 13 August 2019

I read "Evolution at two-time frames shape structural variants and population structure of European plaice (Pleuronectes platessa)" by Le Moan et al. I found the MS interesting and the findings novel and relevant. I think the methods used are mostly appropriate to lead to the conclusions the authors arrive (except for some minor confusing cases). On the downside, however, I found the MS difficult to read at places, some important details missing and many descriptive results that need quantitative support. I also feel the discussion can be consolidated and/or improved in some of the sections. This paper will be a nice contribution showing the relevance of Structural Variants in the evolution of divergence and demographic change

Below I include many comments. All of my comments are either minor details or are intended to add a more detailed explanation around the major issues I found in this MS:

1) Lack of details about sampling and dataset sizes etc.

2) Many different concepts are included without properly considering each of them in-depth or without consolidating a proposed mechanism to connect them. e,g, "edge effect", "founder effect", "IBD", "allele surfing".

3) Some inferences are rather descriptive and lack quantitative support

Comments:

L 9-11 At this early point in the MS, it is a bit unclear why SV "provide evidence" for evolution at two-time frames. I think in general this sentence is a bit confusing.

L 22 - What global distribution? If the plaice is not distributed globally, or at least its global distribution is not assessed here...

L 70-71 This argument needs to be developed a bit further and/or a reference included

L 83 This needs a reference, a combo later used in the text would work well here (e.g. Jones et al., 2012; Morales et al., 2018)

L 118 - "to examine multiple hydrographic gradients" It is not very clear which are these gradients. Would be good to mention them here.

L 118-123 Given that the evolutionary history / temporal framework of SV's is the highlighted aspect of the MS, I'm surprised this is not included in the goals here

L 145-146 "Most of THE northern limit of the plaice distribution was covered with this sampling design" What about the baltic distribution? How much of the Baltic distribution is included in this sampling?

L 184 Why a multiple-testing correction was not used for Hardy-Weinberg?

L 191 - 195 I suggest to add the final sample sizes for all these different datasets

L 211 - 222 I'm a bit confused behind the logic of "examine the demographic histories associated with the major population breaks identified in the overall dataset" aren't the largest breaks Icel, Norws and Katte? And then the most interesting to include Bals to date the Baltic colonisation? Maybe I'm being a bit dense and cannot fully understand what the authors are trying to teste here. Also, the results of the demographic modelling seem to have very little weight in the discussion later on, so it feels a bit forced.

In any case, I would like to see cartoons of the different demographic models tested in the SI

L 232 Why is this a haplotype allele frequency? Are not these markers for the entire chromosome? Thus, unlikely they represent a single haplotype?

L 248 I don't understand why the genomic architecture is invoked here? "The genomic architecture of differentiation was examined using SNP specific FST values"

This comes back in line 334. I think this is not about the genomic architecture of the SVs, given that SVs are a feature of the genomic architecture. This refers more to the genetic or spatial structure of SVs or something along those lines. I suspect the authors are misusing the definition of genomic architecture here.

L 255 Was LD calculated pairwise across the entire chromosome? With a sliding-window? Please clarify.
L 277-281 How these coordinates were defined? By eye? Please explain. Also why the authors did not use the other chromosomes to represent the genome-wide estimates? How likely/unlikely it is that these regions within the same chromosome are subject to some linked effects? E.g. How well the boundaries of the SV are defined and how long is the LD decay?
L 316-320 I do not fully understand what the authors are trying to say here. I find it difficult to appreciate a correlation within the blue dots with only 3 comparisons, also I cannot see the different effects between panel a and b of Fig. 2

Methods: there are some details that are missing in the methods when specific things are presented e.g. How many SNPs the authors started with, how many were filtered in each step and how many they end-up with. Also the final sample sizes of the different "datasets". Or how many genes were found where they say "and all genes with more than 80% mapping were recorded". Among other things. Would be good for the authors to double check the methods section and add all these small but important specific details. Particularly how many samples and markers go into each of the analyses.

Figures: the style for presenting panels as "(a)" or "A" and the way they are presented in the legends (e.g. before or after their corresponding explanation) varies between figures. Please consolidate in a single style.

L 356 "the SV21 FST was elevated only in pairwise comparisons including Nors/Katte ... (Fig. 3B)" But this Figure does not show comparisons with Katte
L 359-360 "The genome(-)wide differentiation outside... " Ideally this would be backed-up qualitatively, I think mean and sd Fst of SV and collinear will suffice.
L 360 "Several SNPs" How many? Quantitative support needed...
L 326-364 The decrease Fst / increase pi inference really relays on a visual pattern that to me is not immediately obvious. This also rests support to the statement in L 498-499. Some way of quantifying this pattern would add support to this statement

I think this section tends to be very descriptive and need to be backed-up with more quantitative data when possible

Just a minor note, Figure 3D The quality of the image is not enough to appreciate the differences between red and purple in the triangular LD plots. Either include a higher quality figure in next version or maybe change the colour palette a little bit.

L 408-410 I do not understand what the authors mean by "decoupled from the species' geographic distribution"
L 416 and L 459 "to our knowledge is one of the clearest" and "is the first example described " I personally consider this type of statements unnecessary, but if the authors feel are needed they definitely need at least some support from the literature
L 446-460 This section is very difficult to follow because it jumps between different demographic/evolutionary models without much structure or connection. Namely, "edge effects", "founder effects" and IBD. These have obvious connections to each other and to the observed pattern of genetic diversity, but I feel the discussion here does not do a good job at making these connections. I found the edge effect particularly confusing as at some point the author seems to imply that edge effect and founder effect are the same things?
L 462 "The two large SVs were polymorphic in most of the sites studied" I'm interested in entertaining the idea that the SV's are not in fact polymorphic but that individuals from different groups were combined in a single site. For example, there is not enough information in the Methods section to know how the sampling was conducted and where these individuals come from. Is it possible that individuals come from different micro-habitats or that there some cryptic variation in such a way that individuals could be grouped into some kind of ecotype with alternative SV's? i.e. the SV's are not really polymorphic.
This, of course, does not invalidate that SV's are polymorphic to the site level. Feel free to ignore this if it does not make sense given what is know about the ecology of the plaice
L 474 - 477 This edge effect is not super obvious by only looking at the pie-charts in Fig.3. I suggest that authors

to add a figure supporting this statement, e.g. correlation of frequency with latitude or something like that.
L 477-478 ”the ancestral haplotypes for SV19 disappeared” This is one more example of how the lack of detail about sampling does not allow to evaluate this type of statements. E.g. I cannot see if the sample size is lower/enough in this region, potentially preventing the authors to not see the less common haplotype.

L 491-494 It would be good to give an example of what the authors refer to here. A good citation would be Faria et al. 2019; Molecular Ecology; 28: 6, 1375-1393 where they show how to find inversions in non-model organisms with the type of approaches you talk about in this sentence.
L 539 - 540 I do not see where the evidence for 2.5 times more net divergence is coming from? I might just have missed it, though.
I think the section ”The evolution and maintenance of the structural variants” could be summarised a bit further given that it contains a lot of speculation. It is a nice discussion, it just feels a little bit too long.
L 584-585 I’m not sure what the authors talk about here, the first and the second? what? SV’s? This sentence is rather unclear.
L 588 Why waiting the entire MS to propose the SV’s are inversions? It is kind of irrelevant at this point and quite unexpected... I think the author should either not compromise and call them SV’s all throughout, or to bring this on earlier and justify it.
L 589 ”decoupled from geography” does not make sense as they might have a different geographic distribution, but they cannot be ”decoupled” from geography.

The ”Conclusions” section is very little of conclusions and a lot of future directions.

Reviewed by anonymous reviewer 1, 13 September 2019

This paper presents an analysis of population structure in the European plaice, using RAD-seq data and concentrating especially on the contribution of two previously detected large structural variants. The population genetic data suggest a genomic background of isolation by distance (strong correlation between Fst and geographic distance), from which the structural variants deviate. They make the case that the structural variants are likely to be ancestral polymorphisms, possibly maintained by balancing selection.

In my opinion, the main merits of the manuscript are that it uses standard methods in a sensible way, and that it is well written, especially the Introduction and Discussion sections.

I don’t have any major criticisms, but several minor comments and questions about framing, some details that are missing from the Methods section, and some of the Results that could be more clearly presented.

Minor comments

Title and conclusions (lines 583-590): The framing of the paper as about the ”evolution at two-time frames” (by the way, shouldn’t it be ”at two time-frames”?) strikes me at a somewhat odd choice, since it’s not that clear to me what this model entails in terms of testable predictions, and how it fits with the data. This becomes especially clear in the Conclusions section, that highlights this model to the exclusion of other results, but in very general terms. It is almost a necessity that any widespread polymorphism first is established in a population, then spreads, but what else does the model predict that fits, or doesn’t fit, with the plaice data? I acknowledge that this may just be an issue of how the model is presented, or a matter of personal taste.

Throughout the paper: I would advice against using bespoke abbreviations. ”SV” for ”structural variant” might be acceptable (though my preference would be to spell that out too), but I see no reason to for example abbreviate ”North Sea” to ”Nors”. This saves little space and impairs readability. I would also encourage spelling out isolation by distance, since the abbreviation ”IBD” unfortunately has two meanings in population genetics (even if it should be obvious from context that you aren’t talking about identity by descent).

Lines 38-39: The sentence says that because structural variants can harbour several genes, they may have functional consequences. This does not follow, so I guess the sentence is trying to say something else?

Lines 76-81: The idea of ”evolution at two time-frames” could be fleshed out in more detail.

Line 98: This is the first mention of the European plaice in the introduction. I recognize that this might be a matter of personal style, but in my opinion, the study system and the question to be addressed would benefit

from being introduced much earlier.

Line 114: The notion that structural variants explain most of the differentiation could be made more precise. How much, and what differentiation?

Line 160: Does ”DNA extractions were standardized” mean that samples were diluted?

Line 182: What does it mean that a SNP is ”present” in 80% of individuals? That 80% of individual had genotype calls for that SNP, or something else?

Line 184-185: Hardy-Weinberg equilibrium filtering of SNPs is problematic even in the best cases, and this threshold ($p < 0.05$, presumably uncorrected) is very low. How many variants were removed because of this, and is this really appropriate?

Lines 187-191: Could you explain in detail how and why ”size selection was slightly shifted”? This sentence is not clear to me. What happened, why did it happen, and how did you detect it (could the evidence be shown in a supplementary figure)?

Line 199: How did you choose which SNP to keep in each bin, at random? Given the strength of LD, is the 1 kbp limit enough to remove the effect of LD?

Line 207: The Mantel test is never discussed again. Is this an omission?

Lines 243-256 and elsewhere: Is ”genomic architecture” really a good term for diversity and differentiation?

Line 258: Could you give reference genome coordinates of the extracted sequences?

Line 262: What does 80% mapping mean precisely?

Lines 267-269 and 284-286: This procedure could be better explained. I also worry that the selection of random alleles might mean the creation of haplotypes that are not even present in the population. Is there a risk of getting the sequence wrong? Would statistical phasing help?

Lines 295-296: Is a strict molecular clock using human mutation rate really the best possible way to date this polymorphism? Are there no mutation rate estimates from fish available?

Lines 307-311: The observed heterozygosity may be a useful descriptive statistic, but its importance here is never really explained. Why are these results presented here?

Lines 318-323: and Figure 2: What, exactly, is ”expected under an IBD scenario”, and what is it that ”disappears” in panel B compared to panel A? From what I can tell, the relationship pictured in panel B is stronger than the one in A. This part of the paragraph, with multiple patterns that are lost under various removals is very hard for me to follow. It would also help to provide some context about what this is investigating. Could it be rephrased?

Line 321: The p-value can’t be precisely 0, can it?

Lines 375-377: Drawing any kind of conclusion about gene function based on 900 genes is highly questionable.

Lines 391-400: I’m afraid I do not understand this paragraph. Could it be rephrased?

Lines 500-502: I do not understand the sentence about fission/fusion of chromosomes. Could it be clarified?

Lines 575-581: The Discussion about what might be maintaining these polymorphisms is thoughtful and interesting, but it ends by mentioning process studied by the Kirkpatrick & Barton (2006) without going into detail. Could you the implications of this model and how it fits with the data?

Data availability: Will the data be made available in a standard repository? This is not mentioned in the manuscript. Depending on journal (or actually, even if the journal does not require it), a clear data availability statement would be nice.

Reviewed by anonymous reviewer 2, 11 September 2019

Download the review