Summary
=======

In this manuscript, the authors propose a procedure to detect the signature of selection in genomic regions that potentially have a different demographic history compared to the rest of the genome. The genomic regions of interest here are inversions. After inferring region-specific demographic histories, the authors train a discriminant function based on simulated data. Simulations are done both under the neutral null models, and under specific scenarios of selection. The discriminant function is trained on the principal components of a set of summary statistics. Its task is to distinguish between the neutral vs. selection models. To account for the variation in effective recombination rate between the inversion breakpoints, the authors use an ad hoc method to break down the region-wide estimates of gene flow to smaller physical units. The approach is designed to cope with elevated background divergence in regions of low recombination, which makes it challenging to detect candidates of selection based on patterns of increased divergence. The authors conduct a simulation study with a single locus under selection to compare the performance of their approach to a classical Fst-outlier scan in terms of false positive and false discovery rate, and in terms of detection power. The approach based on discriminant functions performs best, but only if the signature from mulitple neutral SNPs in the neighbourhood of the site under selection is integrated, and if selection is strong. If only one neutral SNP is probed or if selection is weak, the performance is more similar to the Fst-outlier approach.
The novel approach is then applied to RADseq data from the Anopheles gambiae/arabiensis species complex. Two inversions on chromosome 2 are compared to the rest of the genome. For each inversion, the standard (S) and inverted (I) arrangements have been associated with wet and dry habitats in previous studies. Moroever, previous work seems to suggest that the I arrangements have introgressed from A. arab. into A. gam. The authors infer that one of the inversions is older than the split of the two species, and that selection has acted on the standard, rather than on the introgressing inverted arrangement. Conclusions about these inversions being involved in adaptation to both wet and dry conditions are made, although evidence for this seems far from compelling to me.

While the manuscript presents an interesting idea and tackles a very important issue, a combination of three concerns lead to my conclusion that the paper cannot be accepted in its current form: 1) Neither the extent of methodological innovation nor the biological insight gained from an application to A. gambiae seem to keep up with what I would consider appropriate for publication in PLoS Genetics; 2) There is a lack of understanding of the statistical and biological mechanisms behind some of the results, and a potential confusion of the signatures of selection and demography; 3) The manuscript needs substantial editorial revision; it contains redundant parts and the writing is often overly wordy or grammatically incorrect.

General comments
==========

- The approach proposed here is a compilation of methods to a single pipeline of analysis to detect the signature of selection. There is appeal to this undertaking and the rationale behind the different steps is plausible. However, the methods put together have individually been used in the literature before: inference of demographic history using site-frequency spectra; Fst-outlier tests to identify loci under selection; discriminant function analysis of PC-transformed summary statistics. My feeling is that the degree of innovation is therefore limited and might not quite keep up with what I would expect from a PLoS Genetics article.

- The central idea of this paper is to fit separate neutral null-models to different genomic regions (collinear vs. inversions), and to then identify candidates for targets of selection for regions of interest (inversions). The rationale is that inverted regions may have experienced a different demographic history due to reduced recombination rates and/or introgression. However, I am concerned that if selection was acting in the inverted regions, this will lead to a signature that is hard to separate from demography. In view of selection, inversions might essentially behave as a single super-locus. If this is the case, it is questionable whether identifying individual candidate sites is meaningful; many of the detected sites are likely not causal. I think the way forward would be to develop a more principle-based model that accounts for the effect of linked selection and demography jointly.

- Related to the previous issue, the step where the recombination rates are adjusted in bins along the inverted regions was dubious to me. Although I understand the rationale, I am nervous that this not only corrects for "demography", but also to some extent for the long-range effect of linked selection.

- If I understood correctly, only the DAPC part for inference of selection was tested in a simulation study. Although it is very important to know that this part seems to 'work', there is a lack of understanding of what creates the differences in performance between the Fst-outlier approach and DAPC (see next point). More fundamentally for this paper, however, I think that the performance (in terms of false discovery rate, false positive rate, and detection power) of the *entire* pipeline should be assessed, not only of the DAPC vs. Fst-outlier part.

- I missed an explanation for why the "DAPC-region" method had a much better detection power than the Fst-outlier approach. Is it because of the nature of DAPC per se? Or because Fst-outlier tests per definition are using Fst, whereas you allow the DAPC to also use pi (heterozygosity)? How would DAPC perform if you allowed it to use Fst only? On the other hand, would a heterozygosity-outlier scan perform better than an Fst-outlier scan? Similarly, why did the DAPC-region method outperform the DAPC-locus method in terms of

detection power? It is very interesting that the DAPC-region method had higher detection power at the same time as having lower false discovery rate. The partial explanation in the Discussion that the DAPC-region method is more robust in the face of noisy data would, if I understood correctly, explain a lower false-positive rate, but not necessarily a higher detection power.

- While the application to A. gambiae as a vector of malaria bears some biological relevance, I wonder if, to justify publication in PLoS Genetics, the identified candidate sites would have to be further studied. There would need to be at least an analysis of enrichment for functional terms. I also find it hard to believe that so many sites (regions) on 2La are individually under selection (Fig. 3). If this is an artefact of strong "linked" selection (many candidates may not be actual causal sites), it is questionable whether it is a good idea to look for individual outliers, rather than modelling an aggregate effect of linked selection for a given genomic region.

- The authors seem to find that selection occurred in favour of the 2L+a (S) arrangement rather than the introgressed 2La (I) arrangement. They conclude that this suggests that "the same region" has been involved in two episodes of adaptation, to wet and dry habitats. I could not follow this argument and was not convinced that the results support this conclusion. I found it confusing that the I arrangement is called an "adaptive" introgression from A. arab. into A. gam. although the results presented here suggest that it is actually not adaptive in A. gam., but rather deleterious.

- This study has only looked at scenarios of directional selection, either in favour of the S arrangement, or in favour of the I arrangement. In both cases, the nature of selection is defined only by the status of the inversion. While this seems to be supported by earlier literature showing correlations between dry and wet periods and the frequency of I and S arrangements, repsecitvely, I wonder whether selection could act in the form of underdominance. Selection against heterokaryotypes (S/I genotypes) in A. gam. might lead to elevated divergence between S and I in A. gam. Of course, underdominance would not necessarily lead to a stable situation in the long term, and it might be hard to explain the cyclic behaviour of I and S arrangements with dry and wet conditions.

- Similar to ABC, the DAPC approach relies on summary statistics. It is known that summary statistics are hardly ever sufficient for model comparison (e.g. Robert et al. 2011 PNAS). This issue should be mentioned.

Specific comments
==========

Notation: Each comment starts with an identifier of page and line number px.ly-z, where px is page x of the respective section, and ly-z refers to lines y to z on that page. Note that page numbering

starts anew from 0 for each section.

Title

I found the construction "demography-informed selection signatures" unfortunate, as the signature of selection is not really demography-informed itself. The title also promises to reveal two different roles for the two inversions. Having read the paper, I am not clear about what the difference actually is.

Abstract

The abstract is too wordy. See below for some specific comments.

l3: Add "in the face of gene flow" after "loci".

l4-6: That might not hold for very old and/or long inversions, where recombination has enough time to equilibrate diversities, except perhaps very close to the breakpoints.

l6-l9: This sentence is too long and too complicated. Please revise it.

l9: "We demonstrate the approach with an analysis of RAD..." -> "We apply our approach to RAD..."

l13: Insert comma after "inversions".

l13-14: "compared with prevalence of" -> "and". "Despite of both being adaptive..." -> "Although both are adaptive...". As mentioned above, I had a problem with these introgressions being called "adaptive", if it turns out that the introgressing arrangement 2La (I) is selected *against* in A. gam.

l16: A word is missing between "selected" and "within". "predated" -> "predates"

l20: I find the phrase "...exhibits much more selection signatures than..." unfortunate.

Author summary

l2-3: "With this new approach we found the..." -> "We found that the..."

l4: "...has involved in..." -> "...has been involved in..." (?)

l8: Insert "in inversions" after "...targets of selection".

Introduction
----

p1.l6-7: This sentence does not make sense to me.

p1.l9: "spectrums" -> "spectra"

p1.l10: Delete "among populations" (suggestion)

p1.l11: Delete "ecologically"

p1.l12-l13: Please improve the formulation ("The appeal of the approach also extends from...")

p1.l14: "becomes inherently" -> "is"

p1.l20-l25: Isn't the first characteristic just a consequence of the second one?

p1.l22-l25: This sentence is too long.

p1.l26-l29: In some sense, this rationale is strange, because the inversion could be considered as a single outlier compared to the rest of the genome, in which case the approaches criticised here might detect them. Of course, it is not straightforward to interprete the fact that a whole inversion pops out as an outlier. Insert closing round bracket after "FLK [13]".

p2.l7-8: "assign" -> "classify"; "into" -> "as"; "selection classes...function" -> "under selection"

p2.l8: "the newly develped" -> "our"

p2.l12: Delete "the"

p2.l13: Insert "long" after "7 Mb"

p2.l14: "defy" -> "complicate"

p2.l15: Delete "across the inversions"

p2.l18: "identify its role in..." -> "suggest that 2La is causally related to...". "..., as do similar trends observed in 2Rb" -> ". Similar trends are observed in 2Rb"

p2.l25: "...high divergence problem..." -> "...problem of high divergence..."

p2.l26: "Nevertheless, despite these challenges, as our analyses demonstrate, we are..." -> "Despite these challenges, we are..."

p2.l29-30: "We highlight...An. gambiae" -> "We discuss implications

four our understanding of rapid divergence in the malaria vector An. gambiae"

p2.l31: "for" -> "to". Delete "more generally"

p3.l1: "than" -> "compared to". Delete "because of the mosaic nature".

Results
----

p1.l8: "Radseq" -> "RADseq" (Applies to other places. I am not going to repeat myself.)

p1.l16: Delete "new"

p1.l17: Insert comma after "First"

p1.l18: "using" -> "from"; "spectrum" -> "spectra"

p1.l19: "inversion specific" -> "inversion-specific"; "gene flux rate" -> "the rate of gene flow"

p1.l20: "...age of inversion mutations" -> "...age of the inversions"

p1.l20-23: This sentence is too long and too complicated.

p1.l26: Delete "analyses"; delete "in the species".

p2.l9: Invert the order of the two numbers (smaller one first) (?).

p2.l10: "...which is about exchanging..." -> "...which corresponds to..."

p2.l9-10: Please check whether the two pairs of numbers (migration rate <> individuals) are consistent. A quick back-of-the-envelope calculation revealed some disagreement, but I did not double-check.

p2.l11-l12: Please remove the information about the physical position of the inversion and place it in the "Material and Methods" section.

p2.l12: "were distributed into one of" -> "formed"

p2.l15-18: Reformulate and shorted, e.g. to: "A comparison with molecular karyotyping results shows that the three clusters correspond to the three inversion genotypes: the inverted homokaryotypes (I/I), the heterokaryotypes (I/S), and the standard (non-inverted) homokaryotypes (S/S) (Fig. S4c, f)."

p2.l21: "gene flux" -> "gene flow" (?) (Applies to other places. I am not going to repeat myself.)

p2.l23–l24: Reformulate to "The inversion in An. gambiae was modeled as being introgressed from An. arabiensis after the split of the two species"

p2.l24: Delete "mutation"

p2.l25–26: "their introgression time" -> "the time of introgression"; "expected for detecting selection" -> "between the two karyotypes"

p2.l23–26: I am concerned about selection having a confounding impact on the inference of demographic parameters. In particular, the estimate of the coalescence time of S and I (T_IS) might have been inflated by divergent selection.

p2.l27: Insert "the" before "species"

p2.l26–30: If this scenario is true, then what maintained the S/I (ancestral) polymorphism during the 2.5 Ne generations in the species ancestral to A. arab. and A. gam.? Directional selection in favour of the S arrangement of 2La (as suggested by this study) might have lead to the extinction of the I arrangement. I therefore wonder if the mode of selection changed over time.

p3.l2: "inversion specific" -> "inversion-specific"

p3.l3: Delete "double"

p3.l4: "versus breaking points" -> "versus close to the breakpoints"

p3.l2–6: This procedure of adjusting local recombination rates seems very much ad hoc in nature. Without an explicit model incorporating both selection and demography, it is difficult to judge if it is appropriate.

p3.l8: Delete "e.g.," and the first "percentile"

p3.l6–12: Please report values of within-arrangement diversity and total diversity (i.e. the ingredients to calculating Fst) separately. Which of the two is responsible for high Fst?

p3.l22–23: Please make sure the statements about selection here and in "Material and Methods" (p5.l5–7) are congruent. As stated before, I wonder if the data could also be explained by a model of selection against hybrids (underdominance) instead of directional selection, or by temporally varying selection (driven, e.g., by alternating dry and wet periods).

p3.l29–30: Delete "which ...or not"

p3.l30: Insert "the" before "three". Why "therefore"?

p4.l1: "...with a comparable false-positive rates (~0.06) as the..."

-> "...with false-positive rates (~0.06) comparable to the..."

p4.l4-5: This seems important, and I would like to see a visualisation of the information contained in the various summary statistics (see previous comments related to this).

p4.l5: "of the variance" -> "to the variance"

p4.l6: "...statistics were transformed into..." -> "statistics had been transformed (in)to"

p4.l7 "assigned into" -> "assigned to"

p4.l8 Please check if a reference to "Fig 3a" is really intended here. Please label panels (a) and (b) in Fig. 3.

p4.l11-12: The number of significant SNPs seems very large, and it is hard to imagine that all of these are individual targets of selection.

Discussion

- p1.l13-14: "decipher" -> identify; Why can the new approach identify the branch where selection occurred, but Fst-outlier scans cannot? Is it because within-population (within-arrangement) diversity is reduced only in one population (arrangement type), but not in the other? In this case, Fst would not be expected to identify directionality, but a comparison of diversities (heterozygosities) would. That could give an advantage to the DAPC method, because it uses different information. Please report a summary of the simulated summary statistics used to train your discriminant classification function for each scenario of selection and each selection coefficient. For instance, you could plot each summary statistic as a function of the selection coefficient, for the different scenarios of selection.

- p1.l2-3: "selection signature" -> "a signature of selection". It was not quite clear to me what part of this sentence really describes the novelty. Is the emphasis on "inversion-specific coalescent expectations"?

- p1.l16-17: "has involved" -> "has been involved" (?). Delete ", respectively". "highlighted" -> "highlights". Again, I doubt that the results shown here allow for the interpretation that the inversion has been involved in adaptation to both wet and dry habitats.

- p1.l17-19: Given that selection seems to favour the S arrangement over the introgressing I arrangement, I do not think that the results highlight the importance of *adaptive* introgression. Rather, they ask for an explanation for why a deleterious

introverting variant is still segregating in A. gam.

- p1.l22-24: Delete "in genomic scan". I do not fully agree with this statement, because Fst (and, to some extent, D_XY) can be confounded by different modes of selection, as well as demography. Hence, they are not necessarily the "most effective way to detect targets of divergent selection in most cases".

- p1.l26-29: This sentence needs to be revised. Please avoid too many non-standard compound nouns.

- p2.l5: Not all ABC approaches apply a PC transformation to summary statistics. This step is not a defining property of ABC.

- p2.l6-8: ABC model choice has been done, too. Model choice/ comparison with both ABC and DAPC suffers from a lack of sufficient summary statistics. This remains a major issue and should be mentioned. See the ABC literature on this topic.

- p2.l8-13: This part creates the impression that using information from multiple summary statistics jointly is novel. However, this strategy has been widely used in the context of ABC, and the benefits and problems discussed in the relevant literature.

- p2.l16: Please revise this sentence.

- p2.l16-20: If I understand this correctly, the averaging strategy would reduce statistical noise, but that is not necessarily the main factor increasing detection power; it may contribute to a lower false positive rate, however.

- p2.l21-22: "1Kb" -> "1kb". Does this refer to regions outside the inversions? Please clarify.

- p3.l7-9: Surely, this must depend on the strength and mode of selection, the amount of gene flow, and the information extracted from the data.

- p3.l10: Insert "the" after "limited power in"

- p3.l11: "rise from the fact" -> "be"

- p3.l12: "origin time" -> "time of origin"

- p3.l13-15: It was not clear to me where this statement follows from.

- p3.l19: "given 2La's origin history" -> "given the inferred demographic history of 2La".

- p3.l30: The word "given" appears twice in the same line

- p3.l31: "Radtag" -> "RADtag"

- p4.l10-15: I found this a bare overstatement / speculation, with little support by the results reported.

- p4.l16: The word "that" appears three times in the same line

- p4.l17: What is meant by "linkage between co-adapted and maladapted genotypes"?

Material and Methods
----

p1.l9: What did you do with specimens identified as A. coluzzi? Were they excluded?

p1.l11: "part" -> "parts"

p1.l12: "integenig" -> "intergenic"

p1.l25: Please check formulation: "...were then ligated by part of an Illumina adaptor sequence..."

p2.l14 I did not understand the constraint "...present in [...] no more than two haplotypes per locus within each sample"

p2.l20-21: "Weir and Cockerham's F_ST (1984)" -> "Weir and Cockerham's [ref#] F_ST"

p2.l23: What is the reason for this thinning? A reduction of redundancy due to LD?

p2.l24-25: The formulation "Only SNPs that are present in [...] at least 80% of all individuals were included..." is not clear. Whether or not a site is a SNP is a property of the sample. It cannot be "present in a proportion of individuals". Are you referring to a minumum threshold for the minor allele frequency?

p3.l2-4: I am not an expert on these methods, but I found the directionality of this test interesting. Naïvely, I would have tested the clustering against the molecular methods, not the other way round.

p3.l6-7: "implemented in" -> "was done using"; End the sentence before "calculates", continue with "The program simulates the joint SFS across populations for each SNP under a user-specified demographic scenario, with parameter values drawn from prior distributions. It computes the composite likelihood of parameters across SNPs given the empirical joint SFS using a conditional maximization algorithm (ECM)."

p3.l13-14: Reformulate to "Single-nucleotide polymorphisms with more than two alleles across all four species were omitted, and the allele with the major frequency among all species was considered ancestral."

p3.l14-17: Please explain this better. "region specific" -> "region-specific"

p3.l22-23: "joint-SFS" -> "joint SFS"; "Chromosome" -> "chromosome"; "X chromosome" -> "the X chromosome"

p3.l24-26: This is unclear; please reformulate.

p3.l27: "were" -> "was"

p3.l29: "using" -> "assuming"

p4.l12-13: "...region has a larger SNP dataset" -> "...regions comprise many more SNPs"

p4.l14: "reduced recombination" -> "reduced recombination rate"; "The same introgression rate estimated..." -> "The introgression rate (m) as estimated..."

p4.l19: Insert "in msms" after "simulations"; "can then be" -> "were"

p4.l24-28: As mentioned above, I am worried that this step removes the region-wide signal of selection that might be informative about the past of the inversions. How exactly did you "adjust" the recombination rate to the "value that generated the empirical mean F_ST" for each segment?

p5.l1-2: Please give more details of this "comparison".

p5.l7-9: Insert "the" after "when". In what step sizes did you vary selection? Please state explicitly what fitness regime you used. As of now, it seems that you use directional selection in favour of either the S or I arrangement. But how is fitness parametrised? Is there dominance or none?

p5.l9: "Selected locus is..." -> "The selected locus is..."

p5.l12: Insert "the" after "contains"

p5.l13: "5Kb" -> "5kb"

p5.l14: "Radtags" -> "RADtags"; "9" -> "Nine"

p5.l16: "short" -> "long"

p5.l17: "according empirical Radtag distributions and" -> "region according to the location of RADtags, and"

p5.l25: Insert "the" before "false-positive rate" and "ratio"; "and" -> "divided by" (suggestion)

p5.l28: "levels" -> "numbers"

p5.l30: "are" -> "were"

p5.l22–p6.l6: I very much appreciate this simulation study. However, I think the performance of the *entire* procedure, including the steps where demography is inferred, should be tested.

References

Please revise the full list, paying attention to the following:

- Correct journal names (all first letters uppercase); use either full names or abbreviations, according to the journal's guidelines

- Use em-dash instead of en-dash to denote page ranges

- Italicise all genus and species names

- Be consistent with usage of upper- vs. lowercase letters, according to the journal's guidelines

Figure Legends

Figure 1:

l3: "procedures" -> "Procedures"

l4–5: "in collinear..." -> "In collinear...". Move "at time T_div" to after "diverged"; "expension" -> "expansion"

l6–7: "...species have constant gene flow since divergence" -> "species have experienced gene flow at a constant rate m since divergence"; "in regions..." -> "In regions..."; insert "original" before "arrangement"; "split with" -> "split from"

l11–12: Delete "(m)"; "the same as that" -> ""identical to the migration rate"; "maintain the same in..." -> "are shared between..."; "inversion specific" -> "inversion-specific"

Figure 2:

l15: "of" -> "for"; "region" -> "regions", delete "scan"; "dots" -> "Dots"

l17: Insert "from bottom to top," after "respectively,"

l18: "measures" -> "(F_ST)"

l19: "empirical" -> "Empirical"

l21: End sentence with full stop instead of semi-colon.

Figure 3

l27: "either" -> "having"; delete "lineage" after "S"

Figures

Figure 1:

(b) "Collinear region demographic history" -> "Demographic history of collinear regions"

(c) "Inversion region demographic history" -> "Demographic history of inversion regions"

Figure 2:

(c) and (d) Adjust font size of "Regions"

Figure S1:

Please add numerical scale to the colour index for drought. Please add a scale and the direction of North to the map.

Figure S2:

Please label the axes.

Figure S3:

Ditto.

Supporting Information Captions

l7: "Color" -> "Colors"

l9: Ditto

l11: "Right" -> "right"

l13: "each cluster" -> "the clusters are"

Tables

Please be consistent in the usage of the "," to separate triples of digits.

Supporting text

"Radseq" -> "RADseq"; "Chromosome 2" -> "chromosome 2"; It was not clear what is meant by "One SNP per locus with at least 1000bp between them..."; "When PCA were..." -> "PCA was..."; "...to identify inversion associated selected regions rather than populations..." -> "...to identify selected regions associated with inversions, rather than populations..."; "Fig. 1a,d" -> "Fig. S4a,d" (?); "match nicely to inverted" -> "match nicely with inverted"; "Fig. 1c,f" -> "S4c,f" (?); References: Please italicise species names.