

## Review of MS MER-21-0253

Lyu and co-authors present an approach for jointly inferring the strength and timing of directional selection and gene flow, as well as the unknown allele–frequency trajectory underlying a time series of allele–frequency estimates at a focal genetic locus. This approach is timely and highly relevant in the current era of accumulating datasets derived from time–referenced ancient DNA samples. Joint inference of the strength and timing of selection and gene flow is key to our understanding of complex evolutionary processes including domestication, local adaptation, and speciation. The Bayesian inference framework introduced here extends a series of previous contributions all rooted in a hidden Markov model originally proposed by Bollback et al. (2008).

The approach by Lyu et al. is arguably the first one that couples the inference of gene flow to the inference of selection while accounting for genetic drift and still being computationally efficient. Jointly, these additions make the present paper a substantial and innovative contribution to the field.

The authors apply their approach to a time series of allele–frequency estimates at the Thyroid Stimulating Hormone Receptor in European domestic chicken. This dataset has been analysed in a previous study with a similar approach (Loog et al. 2017). The authors of the present study obtained results largely consistent with those of Loog et al. (2017), which validates their new approach.

The authors performed extensive simulations to assess the robustness of their inference approach to a number of confounding factors, including a mis–specified effective population size, a lack of historical information about the proportion of immigrant alleles, and small and uneven sample sizes over time. The novel approach is shown to overestimate the strength of selection and underestimate the time during which selection acted if the effective population size is assumed too high. Moreover, the authors found that a misspecification of the migration rate can distort the shape of the posterior distribution. The authors carefully discuss these limitations and hence nicely delineate the scope of their framework.

The authors motivate extensions of their approach to more complex models of gene flow than the continent–island model, and to the inference of selection at multiple independent loci. The manuscript is very well written, the notation is appropriate and I did not spot any mistakes in the equations (other than what are likely typos). I had some minor comments and two somewhat larger concerns, one about dominance in the simulation study, the other about the motivation of the four–allele model employed (see below). Both concerns can easily be fixed or discussed. My overall verdict is very positive and I highly recommend this manuscript for publication in MER.

### Major concerns

1. The authors assume a model of directional selection with dominance and let the dominance coefficient  $h$  vary between 0 and 1 (i.e. complete dominance to complete recessivity of the derived allele, but no under– or overdominance) in the formulation of the theory. However, the authors then fix  $h$  to 0.5 (no dominance) in their simulations and to 1 (complete recessivity) in their analyses of the derived TSHR allele in domestic chicken. While the latter decision is justified by empirical evidence, I wonder why the authors then did not also consider the case of  $h = 1$  in their simulations. More generically, I think the authors should replicate a subset of their simulations for a range of  $h$  values, including the extremes of  $h = 0$  and  $h = 1$ . Allele frequency trajectories under these two extremes show qualitatively different dynamics, especially close to the boundaries of 0 and 1. Whether these differences can be discerned by the proposed inference approach, and how estimates of  $h$  are confounded with estimate of the selection coefficient  $s$  if of high interest, but remains unexplored. If simulations to address this question are too expensive, the authors should at least thoroughly discuss the potential impact of their choices of  $h$  on their estimates of  $s$  and the timing of selection.
2. The authors distinguish four alleles in their one–locus model, even though there are only two functionally different categories of alleles. The second axis of differentiation contrasts alleles within the same functional category by whether they are derived from the resident or the continental population. The necessity for this second axis of differentiation is not immediately obvious, but becomes implicitly clear as one reads the theory: it encodes essential information to disentangle the effects of gene flow and selection on the frequency of functional allele categories. As this seems to be an essential detail of the design of the approach, I suggest the authors explicitly motivate the choice of a four–allele rather than a two–allele model. Such a motivation might be placed around what are currently lines 75–77, or alternatively lines 116–118.

## Minor comments

C: comment; Q: question; S: suggestion; R: request.

### Generic

- Parts that I think should be written in past tense (methods, what the authors did) are currently written in present tense. Examples include "run" instead of "ran" in line 242, "re-analyse" instead of "re-analysed" in line 339, and "simulate" instead of "simulated" in line 445. I propose the authors revise this issue in the entire text.

### Introduction

- [l.34] **R:** Please clarify what exactly is the undesirable feature in aDNA. I suspect the authors want to say that high computational costs for large evolutionary timescales become a strong limitation for the analysis of aDNA. The formulation may need to be improved.

### Materials and Methods

- [l.99] **Q:** Is there a particular reason for why  $s$  is scaled by  $2N$ , whereas  $m$  is scaled by  $4N$ ? I found this confusing. A corollary of the current choice of scaling is the factor of  $1/2$  multiplying  $\beta$  in Eq. (2), which could be omitted if  $\beta(t)$  were defined as  $2Nm$  rather than  $4Nm$  (if  $t \geq t_m$ ).
- [l.172] **S:** "Although it can be achieved by..." → "In principle, the transition probabilities can be obtained by..."
- [l.173–174] **S:** "...in Eq. (1) like Bollback et al. (2008) and He et al. (2020c), typically using a finite difference scheme, this requires..." → "...in Eq. (1) typically using a finite difference scheme (cf. Bollback et al. 2008; He et al. 2020c). However, this requires..."
- [l.178] **S:** "...in this work that only involves..." → "...in this work. The PMMH algorithm only involves..."
- [l.220] **Q:** Should it not say  $x_{1:K}$  instead of  $x_{1:K}^*$  on the RHS of the equation?

### Results

- [l.264–267] **S:** ", thereby assuming that the continent allele counts of the sample are unavailable at first three and seven sampling time points, respectively, for each..." → ". To study the impact of missing continent allele counts, we assumed that the continent allele counts of the sample are unavailable for the first three or seven sampling time points for each...". **S:** Omit the parentheses surrounding "as shown [...] Figure 1".
- [l.296] **S:** "that" → "for which"
- [l.301] **C:** "degrades" sounds somewhat strange here. How about "dilutes" or "blurs"?
- [l.302–303] **Q:** Should the two specific values of  $k_m$  not be reversed here, i.e.  $k_m = 360$  first and  $k_m = 90$  second, to agree with what is written in words?
- [l.321] **S:** "demonstrates" → "has"
- [l.368] **C:** The negative value for the time  $k_m$  at which migration starts may confuse the reader. Perhaps remind them of time being measured relative to 0 BC (cf. Table 2, where an analogous footnote might be appropriate).

### Discussion

- [l.445] **S:** Omit "further"
- [l.467] **S:** No period after "developed"; "See" → "see"; place parentheses only around "2018" rather than the entire reference.

### Figures and Tables

- Fig. 4 & 5: **S:** In the captions, replace "aquamarine" and "coral" by "green" and "orange" to simplify?
- Table 3: **R:** Typeset minus signs appropriately, i.e. wider than the currently used en-dash.