



HAL
open science

Molecular evolution through the joint lens of genomic and population processes

Guillaume Achaz

► **To cite this version:**

Guillaume Achaz. Molecular evolution through the joint lens of genomic and population processes. 2020, pp.100103. 10.24072/pci.evolbiol.100103 . hal-02873037

HAL Id: hal-02873037

<https://hal.science/hal-02873037>

Submitted on 18 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Molecular evolution through the joint lens of genomic and population processes.

Guillaume Achaz based on reviews by Benoit Nabholz and 1 anonymous reviewer

A recommendation of:

Fanny Pouyet and Kimberly J. Gilbert. **Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between (2020)**, *arXiv*, 1909.11490v4, ver. 4 peer-reviewed and recommended by Peer Community in Evolutionary Biology. <http://arxiv.org/abs/1909.11490>

Submitted: 26 September 2019, Recommended: 12 June 2020

Cite this recommendation as:

Guillaume Achaz (2020) Molecular evolution through the joint lens of genomic and population processes.. *Peer Community in Evolutionary Biology*, 100103. [10.24072/pci.evolbiol.100103](https://doi.org/10.24072/pci.evolbiol.100103)

Open Access

Published: 18 June 2020

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

In their perspective article, F Pouyet and KJ Gilbert (2020), propose an interesting overview of all the processes that sculpt patterns of molecular evolution. This well documented article covers most (if not all) important facets of the recurrent debate that has marked the history of molecular evolution: the relative importance of natural selection and neutral processes (i.e. genetic drift). I particularly enjoyed reading this review, that instead of taking a clear position on the debate, catalogs patiently every pieces of information that can help understand how patterns we observed at the genome level, can be understood from a selectionist point of view, from a neutralist one, and, to quote their title, from "everything in between". The review covers the classical objects of interest in population genetics (genetic drift, selection, demography and structure) but also describes several genomic processes (meiotic drive, linked selection, gene conversion and mutation processes) that obscure the interpretation of these population processes. The interplay between all these processes is very complex (to say the least) and have resulted in many cases in profound confusions while analyzing data. It is always very hard to fully acknowledge our ignorance and we have many times payed the price of model misspecifications. This review has the grand merit to improve our awareness in many directions. Being able to cover so many aspects of a wide topic, while expressing them simply and clearly, connecting concepts and observations from distant fields, is an amazing "tour de force". I believe this article constitutes an excellent up-to-date introduction to the questions and problems at stake in the field of molecular evolution and will certainly also help established researchers by providing them a stimulating overview supported with many relevant references.

References

[1] Pouyet F, Gilbert KJ (2020) Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between.

Revision round #2

2020-04-29

Dear authors,

The two referees and myself have carefully read the revised version of your manuscript. All the three of us agree and recognize the great improvement of this version, compared to the previous one. The ms really now conveys a more balanced and precise review of the interactions of the different processes at stake in the patterns of molecular diversity. Although the improvement of this new version is impressive, I believe that there is still some room for extra minor adjustments. Please read carefully the comments of the two reviewers on this new version and provide a point by point response.

I also would like to suggest on top several minor points that the authors may want to consider. These points can help clarifying or illustrating the scientific content of this version.

1) Effective population size - its first occurrence is, I think on p3. N_e is never defined and this lacks, as there are multiple ways to define and/or to understand N_e . I would encourage the authors to define N_e at its first occurrence, or to avoid this term entirely. I suspect that on p3 the authors meant some sort of "neutral" N_e (e.g. harmonic mean of population size), but some people use the term N_e to refer to "level of diversity" (measured through π or any other statistics). As (linked) selection will greatly affect the diversity, replacing "diversity" by " N_e " may lead to very confusing arguments. I know this is a long standing debate, getting back to at least Gillespie (and his articles on the meaning of N_e).

2) I am not a big fan of Figure 1 as it is now. Can the author work a bit to clarify it? I suspect that having more than 2 chromosomes/sequences per sample will be helpful.

3) For meiotic drive, may I suggest the reading of: Henikoff S, Ahmad K, Malik HS: The centromere paradox: stable inheritance with rapidly evolving DNA. Science 2001, 293:1098-1102.

4) If I am not too confused, there is a major difference between gBGC and selective sweep. In the former, there is no room for hitchhiking, so the neutral surrounding diversity is not affected. It may be worth mentioning it.

5) Finally, there is no Acknowledgments section in your article. Is this on purpose?

Preprint DOI: <http://arxiv.org/abs/1909.11490>

Reviewed by anonymous reviewer, 2020-04-07 15:19

The authors have done an excellent job responding to the prior round of review in my opinion, and I believe the revised manuscript is much improved. I have a few remaining mostly minor comments that I think would improve the manuscript, including a few minor but I believe essential revisions. However, in general I think the major concerns from the previous round have been addressed.

Minor but essential revisions:

1. There are still a few places where I think definitions need to be clarified.
 - First, I think that many of the demographic processes the authors discuss would often be considered special cases of genetic drift, as the patterns that arise under these demographic

scenarios are still the consequence of random sampling of individuals instead of fitness differences. Of course, a major thrust of this review is to outline the non-adaptive but still non-equilibrium processes that can cause departures from a standard neutral model. But I think an issue remains that the null model here is not carefully defined (e.g., on bottom of page 5 is isn't really clear what the authors mean by 'nor are they purely subject to drift'). It seems that the authors are thinking about something like a standard Wright Fisher model as the null, and then distinguishing between 'non-adaptive' and 'adaptive' departures from that null. This of course has been the subject of many, many papers, but the strength of this review is bringing together disparate topics that have not often been treated together.

- A second issue is using the word 'hitchhiking' to refer to all linked selection, which I think is non-standard in the field and may be confusing. Following Maynard Smith and Haigh (*Genet Res.* 1974), it is preferable to limit hitchhiking to the case of neutral alleles whose frequency changes due to a selective sweep (of some kind) and use the more general term linked selection (e.g. <https://www.nature.com/articles/nrg3425>) to refer to the general case.
2. Table 1. I am not sure the benefit of including the middle three columns in this table, since the answers are yes for all processes discussed in this paper, so it is not clear that listing this information repeatedly in the table is a valuable use of space.
 3. The paragraphs in sections 3 and 4 vary considerably in their depth and detail, particularly with regard to citations. I am impressed with how the authors have managed to pull together a huge amount of disparate topics for this review, but some of the value is a bit lost when certain sections (that have a very deep literature behind them) are shallowly referenced. In particular, sections 3.1 and 3.2 have been the subject of extensive empirical and theoretical work, as well as the topic of numerous reviews. Yet, in section 3.1 only three papers are cited, all empirical work on humans; similarly only two empirical papers are cited for section 3.2 In contrast, section 3.3 provides a deep context for further reading. I think, because many of the topics covered in this review have been discussed extensively before, it would be particularly valuable for the authors to provide more context of the previous literature. In this way, this review could almost serve as a one stop shop as a introduction to a bunch of related topics. This similarly comes up in the section on linked selection.

I also have some comments that I think would improve the manuscript, but largely involve stylistic concerns, so I list these solely for the author's consideration.

1. I think the introduction could even more strongly set up the central focus of the review, which I take to be "highligh[ing]...the major evolutionary processes that can change allele frequencies in ways that mimic signatures of selection." This is of course an important and perennially interesting topic, and throughout the review the authors cite a number of examples where consideration of non-adaptive forces have been important to correct interpretation of population genetic signatures. However in the introduction, the focus is at least in my reading more strongly on the question of 'what fraction of sites in the genome are free from selection', which isn't really strongly touched upon during the rest of the manuscript (although of course is related). This is largely a stylistic concern so I totally understand if the authors like the intro the way it is, I just think the review could have a more focused impact with greater emphasis on the goal of a summary / overview of all the ways data can trick us into thinking selection is acting when it isn't.
2. I still struggle with the way that linked selection is incorporated into this manuscript. It is certainly the case that some methods, especially for inference of demographic models, assume that sites are free of the influence of selection, as the authors discuss on the top of page 3. But in general, linked selection is not, in my opinion, a process that 'can change allele frequencies in ways that mimic signatures of selection' -- linked selection is a byproduct of selection itself. This is particularly the case for certain subsets of selection -- e.g., a completed sweep is generally invisible to detection in the absence of the signature left behind in linked neutral alleles. So I feel like linked selection really doesn't fit with the other topics that focus on 'things that are like selection but aren't', and including

it in this way is confusing. However, I get this may be a semantic argument that me and the authors simply disagree on, so I am not belaboring the point.

3. Personally, I would leave out the section on TEs, although I'm sure a TE biologist would disagree. To me this section doesn't really fit in the framework since the issue with TEs is generally that they are a hard-to-study kind of mutation, not that they are a process that mimics selection (although intra-genomic TE expansion can have its own complications of course).

Finally, as an aside, I hope the authors are holding up in these difficult times. Of course as a review paper revisions are not dependent on access to labs, but even so I realize it can even be harder to search for literature when off campus (depending on individual circumstances), and I am mindful of trying not to ask for too much revision at this stage. I hope the authors (and editor/other reviewers) will take my comments in that spirit and focus on what they see as the highest value improvements.

Reviewed by [Benoit Nabholz](#), 2020-04-12 15:09

I would like to congratulate the authors on their work. The manuscript is much improved. The manuscript is now more balanced, has more sections and appears more thorough. As an example, the number of citations has increase from 59 to 104. I think that most, if not all, readers will learn something out of this manuscript now. I think that it could be recommended by PCI Evol Biol.

I have few minor comments:

- 1) In the new section "Segregation distortion", the authors should consider to include Wolbachia that could drive fixation of mitochondrial haplotype by linkage (Hurst and Jiggins 2005).
- 2) P4. Temporal variation in selection could also exist in addition to spacial variation. See for example (Bergland et al. 2014; Wittmann et al. 2017).
- 3) P8. The two sentences "Additionally, the process of migration into populations or admixture among species can create an influx of novel genetic material. Even if fully neutral, the presence of such heterozygosity in the population leaves a signal indicative of either adaptive processes (e.g. balancing selection) or non-adaptive processes (e.g. secondary contact)." need references.
- 4) The section of "Gene surfing" still has too many references compare to the other section. P8, 16 references were included in three lines (ref 47 to 63). This is far more than for any other phenomenon described in the manuscript. The authors should make the effort to include more reference elsewhere of to reduce the number of references for this section.

Author's reply:

[Download author's reply \(PDF file\)](#)

Revision round #1

2019-11-22

Dear authors,

Two external reviewers and myself have carefully read and then evaluated your manuscript. All the three of us had mixed feelings about the current version of this review. Although, we all believe this is an important and timely topic, the current version may be too short to cover a large and deep topic of molecular evolution.

Although I cannot recommend the ms in its current version, I truly believe that there is the potential basis for a more satisfying version, with some extra work. Mainly, you will have to face two choices: (1) either elongate significantly the content (I am aware that most journal editors require to shorten article as much as possible, so that this recommendation may seem weird) and discuss more generally the interference between drift, selection, demography and/or structure (processes mostly acting at the population scale) together with molecular processes (such as BGC or for example meiotic drive). (2) Another possible choice is to refocus your review on a more specific point and to treat it more deeply.

On a personal note, I share your interest on BGC but believe that it belongs to a larger class of molecular processes on which you could include mutations (a topic on which several new observations and theory have been brought recently), meiotic drive, among others.

If you decide to send a revised version of the manuscript to PCI Evol Biol, please address all the points of the two reviewers and rework your article making an explicit choice between both options I have discussed above. I understand this decision may be a frustrating for you, but be assured that none of the reviewers nor myself have a negative opinion toward the general objective of your review. It is actually quite the opposite, we all have a strong interest in these questions.

Sincerely yours,

Guillaume Achaz

Preprint DOI: <http://arxiv.org/abs/1909.11490>

Reviewed by [Benoit Nabholz](#), 2019-11-08 21:35

In this article, the authors propose a point of view on the processes that must be considered to understand polymorphism patterns. They position their perspective in the context of the recent debate on the utility of the neutral theory to explain patterns of molecular evolution (Kern and Hahn 2018; Jensen et al. 2019). They also aim to “clarify the terminology” and “refine our definitions”.

I have a mixed opinion about this perspective. I agree that it is important to stress out that linkages, GC-biased gene conversion, and allele surfing could have an impact on alleles' fate and that these processes are not adaptive. Many previous articles had already indicated that these processes could mimic selection or limit our ability to detect positive selection. So, nothing is new in this manuscript. However, it has the strength to consider all these processes at once. Unfortunately, I found that the article has some approximations and ambiguities that make that it does not live up to its ambition to clarify the terminology. Moreover, the manuscript put too much emphasis on allele surfing compared to other demographic processes such as structure or bottleneck.

Below, I will elaborate on the points I have addressed in the paragraph above in order to help authors improve their manuscripts. In my opinion, it unfortunately requires a lot of work before to be recommended.

I apologized if my comments refer to page number and quotes of author's sentences but it would have been convenient to have numbering of the lines.

P2: “... of polymorphisms, but do not fall under the umbrella of selection, into a third category of non-adaptive evolution.” There is a problem with this sentence. Do you mean: “... a second category...”. Otherwise what is the second category? It is drift? But allele surfing is linked to drift.

P2: “...environmental context of polymorphisms” The term environmental is ambiguous here because it may refer to the ecological definition of the term. The reader understands much better that it is the genetic environment (i.e., linkage) to which the authors refer here. More generally, I don't like that the authors don't go straight to the point that they want to consider in the perspective: i) linkages, ii) allele surfing and iii) GC-

Biased gene conservation. This leaves ambiguities and some time, it seems that the authors beat around the bush.

P3: "... or cases where selection behaves in a stochastic manner (e.g. in finite populations)." I don't understand this statement. What do the authors mean by "selection that behaves in a stochastic manner"? Is it nearly neutral allele that behaves neutrally in a small population and is selected in a large population? This needs clarification.

P3: The whole paragraph on defining natural selection: I understand that this perspective is intended for non-specialists but this paragraph explains very basic concepts that are not useful in my opinion.

P3: Toward the end of the paragraph: Sexual selection is not the only form of selection that "are not necessarily advantageous at the population level". Many forms of selfish behaviors selected by positive selection can lead to a fixation of allele not advantageous at the population level (Hamilton 1970).

P4: "...but these other complicating processes to fully understand evolutionary biology and the generation and maintenance of genetic diversity across the genome and across populations and species." This is typically the kind of sentence that is not very clear at first but becomes clearer once you read the rest of the manuscript and understand that the 'complicating processes' are linkages, gBGC, and demography. Here, the authors also refer to variation across species (between-species divergence) but this is not addressed at all by the rest of the manuscript.

P4: "... their impacts on genetic diversity are often underappreciated (see [13, 14] and [15] for reply)." The fact that these processes are often acknowledged "by the vast majority of the community" actually reduces the interest of this perspective. I recommend the authors to find more examples of studies where these processes have been underappreciated to support their view that it is worth to make a manuscript. Here, the authors only refer to one example dating back to 2005/2006 that it is not even a genome-wide study (although, I agree that this example is relevant but I want to emphasize that the authors need more examples from genome-wide studies since the authors focus on the field of "evolutionary genomics"). As an example, the author could also consider the case of gBGC in the human HARs (Galtier and Duret 2007). But a literature review would strengthen the manuscript.

P6: "The former may simplify detecting the presence of selection but ...". I don't understand what the authors are referring to by "The former". Is it background selection? If so, I don't think that background selection simplify detecting the presence of selection. You should elaborate on that with some references to theoretical or empirical works.

In section 4.2 "The impact of demography" : the authors focus exclusively on gene surfing. One can feel that the authors have expertise on the subject and this section is interesting to read with many references. However, I don't think that this focus is justified. Indeed, other demographic scenario are known to mimic positive selection. This is particularly true for population bottlenecks (Thornton and Jensen 2007; Innan and Stephan 2003), but population structure (Tian et al. 2008) and possibly other intriguing demographic scenario such as sweep stack (Sargsyan and Wakeley 2008) could be considered as well. Similarly, gene flow could also be considered (Bierne et al. 2011) although I don't know if it should be defined as a demographic event. I have the same comment for table 1 where gene surfing has its own line together with genetic drift. However, gene surfing could be considered as a particular case of drift.

Section 4.3 "gene conversion" : Once again, this section is well written with abundant literature but too much emphasis is put on dBGC. The process is detailed whereas, to my knowledge, it has never been involved in the false-positive cases of natural selection. In contrast, gBGC has often been responsible for false positive (Duret and Galtier 2009; Ratnakumar Abhirami et al. 2010; Galtier et al. 2009).

A much more accurate reference for the existence of gBGC in birds is (Webster et al. 2006) rather than (Weber et al. 2014).

REFERENCES

- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* 20: 2044–2072.
- Duret L, Galtier N. 2009. Comment on " Human-specific gain of function in a developmental enhancer". *Science* 323: 714–714.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *TRENDS Genet* 23: 273–277.
- Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet* 25: 1–5.
- Hamilton WD. 1970. Selfish and Spiteful Behaviour in an Evolutionary Model. *Nature* 228: 1218–1220.
- Innan H, Stephan W. 2003. Distinguishing the hitchhiking and background selection models. *Genetics* 165: 2307–2312.
- Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* 73: 111–114.
- Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. *Mol Biol Evol* 35: 1366–1371.
- Ratnakumar Abhirami, Mousset Sylvain, Glémin Sylvain, Berglund Jonas, Galtier Nicolas, Duret Laurent, Webster Matthew T. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. *Philos Trans R Soc B Biol Sci* 365: 2571–2580.
- Sargsyan O, Wakeley J. 2008. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor Popul Biol* 74: 104–114.
- Thornton KR, Jensen JD. 2007. Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* 175: 737–750.
- Tian C, Gregersen PK, Seldin MF. 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 17: R143–R150.
- Weber CC, Nabholz B, Romiguier J, Ellegren H. 2014. Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol* 15: 542.
- Webster MT, Axelsson E, Ellegren H. 2006. Strong regional biases in nucleotide substitution in the chicken genome. *Mol Biol Evol* 23: 1203–1216.

Reviewed by anonymous reviewer, 2019-11-12 15:13

[Download the review \(PDF file\)](#)

Author's reply:

[Download author's reply \(PDF file\)](#)