



Inferring rates of clonal versus sexual reproduction from population genetics data

Olivier J Hardy based on reviews by Stacy Krueger-Hadfield, Ludwig TRIEST and 1 anonymous reviewer

Open Access

A recommendation of:

Arnaud-Haond, Sophie, Stoeckel, Solenn, and Bailleul, Diane. **New insights into the population genetics of partially clonal organisms: when seagrass data meet theoretical expectations (2019), arXiv, 1902.10240, ver. 6 peer-reviewed and recommended by Peer Community in Evolutionary Biology.**

<https://arxiv.org/abs/1902.10240v6>

Submitted: 01 March 2019, Recommended: 30 October 2019

Cite this recommendation as:

Olivier J Hardy (2019) Inferring rates of clonal versus sexual reproduction from population genetics data. *Peer Community in Evolutionary Biology*, 100083.

[10.24072/pci.evolbiol.100083](https://doi.org/10.24072/pci.evolbiol.100083)

In partially clonal organisms, genetic markers are often used to characterize the genotypic diversity of populations and infer thereof the relative importance of clonal versus sexual reproduction. Most studies report a measure of genotypic diversity based on a ratio, R , of the number of distinct multilocus genotypes over the sample size, and qualitatively interpret high / low R as indicating the prevalence of sexual / clonal reproduction. However, a theoretical framework allowing to quantify the relative

Published: 4 November 2019

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/>

rates of clonal versus sexual reproduction from genotypic diversity is still lacking, except using temporal sampling. Moreover, R is intrinsically highly dependent on sample size and sample design, while alternative measures of genotypic diversity are more robust to sample size, like D^* , which is equivalent to the Gini-Simpson diversity index applied to multilocus genotypes. Another potential indicator of reproductive strategies is the inbreeding coefficient, F_{is} , because population genetics theory predicts that clonal reproduction should lead to negative F_{is} , at least when the sexual reproduction component occurs through random mating. Taking advantage of this prediction, Arnaud-Haond et al. [1] reanalysed genetic data from 165 populations of four partially clonal seagrass species sampled in a standardized way. They found positive correlations between F_{is} and both R and D^* within each species, reflecting variation in the relative rates of sexual versus clonal reproduction among populations. Moreover, the differences of mean genotypic diversity and F_{is} values among species were also consistent with their known differences in reproductive strategies. Arnaud-Haond et al. [1] also conclude that previous works based on the interpretation of R generally lead to underestimate the prevalence of clonality in seagrasses. Arnaud-Haond et al. [1] confirm experimentally that F_{is} merits to be interpreted more properly than usually done when inferring rates of clonal reproduction from population genetics data of species reproducing both sexually and clonally. An advantage of F_{is} is that it is much less affected by sample size than R , and thus should be more reliable when comparing studies differing in sample design. Hence, when the rate of clonal reproduction becomes significant, we expect $F_{is} < 0$ and $D^* < 1$. I expect these two indicators of clonality to be complementary because they rely on different consequences of clonality on pattern of genetic variation. Nevertheless, both measures can be affected by other factors. For example, null alleles, selfing or biparental inbreeding can pull F_{is} upwards, potentially eliminating the signature of clonal reproduction. Similarly, D^* (and other measures of genotypic diversity) can be low because the polymorphism of the genetic markers used is too limited or because sexual reproduction often occurs through selfing, eventually resulting in highly similar homozygous genotypes. The work of Arnaud-Haond et al. [1] shows that the populations genetics of partially clonal organisms should be better studied, an endeavour encompassed in a companion

paper using numerical simulations [2]. A further step that remains to be accomplished is to build a mathematical framework for developing estimators of rates of clonal versus sexual reproduction based on genotypic diversity.

References

- [1] Arnaud-Haond, S., Stoeckel, S., and Bailleul, D. (2019). New insights into the population genetics of partially clonal organisms: when seagrass data meet theoretical expectations. ArXiv:1902.10240 [q-Bio], v6 peer-reviewed and recommended by Peer Community in Evolutionary Biology. Retrieved from <http://arxiv.org/abs/1902.10240>
- [2] Stoeckel, S., Porro, B., and Arnaud-Haond, S. (2019). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: improving our estimation of clonal rates. ArXiv:1902.09365 [q-Bio], v4 peer-reviewed and recommended by Peer Community in Evolutionary Biology. Retrieved from <http://arxiv.org/abs/1902.09365>

Revision round #3

2019-09-17

I found the revision fine and I have only minor comments inserted in the pdf. Making the analysed datasets available (in dryad or as a table with summary statistics) would be welcome.

Olivier

Preprint DOI: <https://arxiv.org/abs/1902.10240>

Author's reply:

Dear colleague, we have now taken all comments into account and submitted to bioarXiv. Thanks again for your valuable contributions that helped a lot improving our manuscript. Sincerely Sophie Arnaud-Haond, on behalf of co-authors

Revision round #2

2019-07-17

Globally I found that the revision addressed adequately the concerns raised by the reviewers and I did not judge necessary to send them the new version. The objectives of the ms are better stated and the addition of Box 1 makes the ms more accessible. However, I still have two main concerns.

- The first was already mentioned in my previous revision (and by L Triest) but I'm not totally satisfied by the way it was addressed here. Basically, the ms insists on the importance of interpreting F_{is} as a proxy of clonality when $F_{is} < 0$. I agree as long as the occurrence of null alleles and/or selfing (or other inbred matings) is negligible and this was probably the case for the four studied species, but I don't think it can be assumed in general. I suggest to add a short paragraph in the discussion warning about these limits of F_{is} .
- The second point came to my mind when checking the 2007 article of Arnaud-Haond et al (Mol Ecol) and I'm sorry that I did not see this when reading the first version. As argued in the ms, R is a commonly used but problematic index of genotypic diversity because it is highly dependent on sample size, and this is the main argument the authors use to favour F_{is} as a more reliable index of the relative importance of clonality versus sexual reproduction. However, another commonly used index of genotypic diversity is based on Simpson index (the complement or the reciprocal). An advantage of Simpson index is that it can be estimated without sample size bias (estimator L), so I expect a positive correlation between L and c . $L=0$ in the absence of clonality (if marker polymorphism is high enough that the probability of obtaining the same MLG by independent sexual events is negligible) and $L=1$ if the population is made of a single clone. I therefore think it would be useful to compute L on the different datasets and assess whether it is also well correlated with F_{is} . Depending on the results, the author could recommend to consider both L and F_{is} to assess the importance of clonality ($F_{is} < 0$ and $L > 0$).

I made specific comments in the pdf version, including some other minor points (e.g. provide the equation defining R).

Additional requirements of the managing board: 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 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."

Preprint DOI: <https://arxiv.org/abs/1902.10240>

Author's reply:

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

Revision round #1

2019-05-23

Three external referees made constructive comments on the ms and expressed globally positive opinions of the interest of this work; I agree with them. However, referees raised different issues and did not all agree on the quality of the writing.

Therefore, I recommend a major revision. I've annotated the pdf file with my own comments, but having lines numbers would be useful for a next version to address specific comments. My major concerns are the following. I agree with the reviewer requesting to improve the introduction, in particular to better explain theoretical expectations regarding R , F_{is} and c . Some sentences are hard to follow. A central message of the ms is that F_{is} is a better proxy to evaluate the prevalence of clonality. However, F_{is} is itself prone to bias due to null alleles and it is strongly influenced by selfing or biparental inbreeding. Hence, the limitations of F_{is} must also be acknowledged in the discussion (+ abstract) and I personally think that alternative population genetics indices should still be developed to better reflect the balance between sexual and asexual reproduction.

Additional requirements of the managing board: 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."

Preprint DOI: <https://arxiv.org/abs/1902.10240>

Reviewed by **Ludwig TRIEST**, 2019-04-03 09:29

This paper concerns a meta analysis with a clear hypothesis to test. The meta analysis was rendered possible due to a standardized sampling design and this at multiple locations and for multiple species. The testing of an excess heterozygosity as a proxy for clonality within sites (negative F_{is}) is valid and was compared to a regularly used easy metric R (clonal richness). The idea that H and F_{is} are less influenced by sample size and design than R is conceptually correct. The manuscript is well-structured and clearly written. Below I will give a series of minor advices to improve wordings/sentences encountered during first reading.

Title: reflects the study

Abstract: corresponds to aims and main findings

The sentence Line 4 -5 should be rephrased (verb missing ? "However, the effect of clonality...")

Introduction Paragraph 1 Line 3: the drastic decline of what ? of species diversity ? Allele diversity? Gene diversity? Paragraph 2 Line 5 : One could add (A and H) after "genetic" and (ML and MLL) after "genotypic", for clarity towards reader. Paragraph 2 Line 8: i.e. the "multilocus" genotype ? Paragraph 2 Line 13: However, G naturally increases with sample size ...This is indeed an indirect effect of the increasing number of alleles (but not at all H). Paragraph 2 Line 15: ... the sample size of ramets (add " of ramets" for clarity) Paragraph 3 Line 4: ... interpretation of negative F_{is} values (heterozygous excess) "of remaining genets", when not Paragraph 4 Line 4:... possibly partly due to "lowered" sampling density ... (I suppose here you mean lower and not higher density ?) Paragraph 4 Line 6:... Values of genetic differentiation (F_{st}) "of populations when considering only genets" (I suppose you mean 'genets' and not 'ramets' here) Paragraph 4 Line 9-13: ... Eventually rephrase this sentence (had to read it 3 times before understanding it ; this because of the used wording '...maximum value that F_{st} can reach, ... in combination with concept R and H_o or F_{is}) Paragraph 5 Last sentence : ... and moving towards null (or slightly "negative" if heterozygosity deficiency occurs...). I suppose this must be slightly "positive" because refers to F_{is} . Paragraph 6 Aim 2 : ... does the genetic composition of natural meadows...What do you mean here exactly with "genetic composition" : MLG or gene diversity?

Materials and methods Studied species: Here, it would be interesting to mention the obligate outcrossing in case of sexual reproduction. Genetic data sets : Because an unequal number of polymorphic loci (7, 8 or 9 loci) was used, it would be interesting to demonstrate that the probability of identity (PI and PI of siblings) was similarly low or at least sufficiently low to ensure detection of MLG repeats across samples (sites) and species. The methods section is clearly written and complete.

Results Clonal richness R Paragraph 1, Line 1 : ...R increased regularly from ... Do you mean “gradually” instead of “regularly”? Relationship between R and Fis Paragraph 2, line 5 : When considering only the genet (i.e. no replicates).... I suppose you mean “ramets” instead of “genets” because the legend of Figure S3 includes the ramets with replicates ?

Discussion My main advice would be to write a paragraph on the sexual reproduction of the considered seagrasses and that one expects outcrossing behaviour. This is an important assumption because many other plants (also aquatic plants) do have a mixed reproduction system ranging from selfing, partially selfing to outcrossing. The findings of this met analysis should make clear that the expected lowered Fis and correlation to R might not be that straightforward as presented here for the four seagrasses. A moderate level of inbreeding, due to the pollination biology of a species, might blur the presented relationship.

Finding empirical data to model predictions Paragraph 5, last sentence is unclear : ... The results are even clearer for ...What exactly do you mean here: the difference between genets and ramets or the fact of having interquartile Fis values =>zero ? Implications for understanding Paragraph 2, line 6 : However, the prevalence of asexuality is associated with a diminution of the influence of drift....Bringing in “drift” here sounds confusing. Drift in small pops can cause both high and low Fis due to stochasticity ?

References, Figures and Tables are O.K.

Reviewed by [Stacy Krueger-Hadfield](#), 2019-04-29 02:55



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

Reviewed by anonymous reviewer, 2019-04-20 23:01

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

Author's reply:

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