Myriam Heuertz, 2019-05-13 16:46
Manuscript: https://arxiv.org/abs/1902.09365

Decision of "revise" for "The discernible and hidden effects of clonality..."

Dear authors,

We have now received three reviewer reports for your manuscript. Two of the reviewers found that the paper represents a significant contribution for evolutionary biologists interested in clonally reproducing organisms; the third reviewer was unable to assess the relevance of the paper. The reviewers made a series of suggestions which I invite you to take into account before your paper can be reconsidered for recommendation by PCI. My decision on this version of your manuscript is thus "revise".

Comment: We are thankful for your reviews and comments that will improve the clarity and readability of our manuscript. We appreciate your constructive proposals. We did modifications to address most of the points reported by reviewers. We changed Figure 2 and 4, and supplementary figures so those ones better plot distributions, using violin plots with varying y-axis scaling to better picture identifiable (and un-) signals. We also reduced the discussion part of one page. Please find the detailed list of actions and when relevant comments/answer here below.

Two reviewers pointed out the need for a clearer definition of research concepts and a clearer framing of research questions: please make sure all concepts are defined, including in the abstract of the manuscript. Clonality should be defined as a form of reproduction/multiplication at the first use of the term. The definition of the rate of clonal multiplication rests on the concepts of reference time frame, reference population and reference individual or assessment unit. Defining those concepts across (partially) clonal organisms is not straightforward but the topic requires to be addressed in the paper, to clarify and justify the definitions of these concepts used in the simulation approach you develop. Please see the reviewer reports for comments and suggestions on how to improve the manuscript. Please also consider my additional comments below.

Action: to address this shared and important point (still again discussed in depth at the 2019 AGA conference in Portland https://www.theaga.org/program.htm), we added a Box (Box 1) in the early introduction, summarizing definitions and concepts and the corresponding seminal references".

p.10 Can you be more specific about/ justify the "known shapes of curves" used to assess the relationship between c and genotypic descriptors?

Action: we agree that this part of the sentence was unclear and removed it, as the sentence that follows much more clearly explains, alone, our approach. We also added a sentence to explain better the selection of a number of theoretical curves to be compared to the simulation based ones. Our approach remains an empirical tentative to fit a parametric known-shaped curve as mentioned in the next sentence.

p.14: The relationship between R and c is not illustrated in Figure 3; it is in Figure 1. The dependence of beta on population size for a given level of clonality is not illustrated. Can you improve this?

Action: We modified the mention of the Figure and modified the Figure 4 to include Pareto

p. 15, bottom: you give the example of multi-locus genotypes present in small strictly clonal populations. I assume you are giving the value at equilibrium. Please link this result to the evolution of R over time (Fig S2).

Action: Done

…It is not clear how long it takes for the main MLG to establish, i.e., what is the impact of drift, and what is the impact of somatic mutation in this pattern. This example illustrates that R, which reflects MLGs, is not necessarily a very good statistic to reflect the diversity of a population with a high level of clonality. In this case, R appears to represent a higher estimate of diversity compared to the (more intuitive) number of clonal lineages present in the population because R confounds the number of MLGs and their origin (MLLs) (taking this logic to the extreme, the more markers you genotype, the higher will be R because the absolute number of somatic mutations will increase). Now that your genetic data allow sorting MLGs into MLLs, would it be beneficial to pull apart the roles of drift and of somatic mutation in the diversity pattern? I would assume that the role of drift is especially marked in small populations, whereas somatic mutation has the same effect for any population size (and the longer the time, the more of them accumulate). The number of MLLs represents information that is not much exploited in this manuscript, and it would be interesting to assess its usefulness in the context of realistic sampling.

Answer: The genotypic diversity R can be computed based on the number of MLGs as well as based on the number of MLLs (as was done in several articles including MLLs delineation, for example Arnaud-Haond et al., 2007-Journal of Biogeography or Becheler et al., 2010 and 2014), and thus far the differences this may generate in computed R values is negligible compared to the effect of sampling density (Arnaud-Haond et al., 2007 –Molecular Ecology-). This is the reason why in the simulation exercise we underwent, we defined a fixed maximum number of alleles (a finite allele model) and did not keep the information of lineages along generations, which would have represented more than tens of tera-octets of data (this was not realistic even with the computation resources we had access to). We however see the interest of quantifying this compared effect of mutation/drift effect on the basis of simulation (which would be more rigorous empirically made comparisons). At this stage however it implies repeating entirely the simulations (preferably with an infinite allele model, and more importantly investigating the effect of different mutation rates on this relative influence) finding a challenging computational way of keeping track of clonal lineages, which we feel would represent another work and article per se.

Figure 2: please verify the number of generations in the figure (500) vs. the legend (10,000).

Action: We modified Figure 2 and 4. We accordingly changed their legends.

P. 17 and Figure 2. You state that equilibrium is reached in tens to hundreds of generations, but the evolution of parameter values through time is not illustrated for the early time frame. I think

it might be insightful to zoom into what happens in the first tens/hundreds of generations; this would be pertinent for some organisms such as trees that display clonal reproduction.

Action: We changed Figure 2 to improve clarity, especially on those first generations.

In line with the prior reflection: you assess subsampling effects on the estimates of genotypic and genetic indices when your simulations have reached equilibrium. How realistic are such conditions for a real life population and the real life situation of the population geneticist sampling the population? I understand you are interested in equilibrium conditions to derive the relationships between c and indicators, but the choice of equilibrium conditions for assessing subsampling effects should be at least discussed.

Comment: We agree that's an important question and we discussed this point in the paragraph "Evolution of parameters when moving towards equilibrium", based on Figure 2 showing now better that for genotypic parameter the equilibrium is reached after rather small number of generations, a situation that applies also to genetic parameters except in case of strict or nearly strict clonality. We hope the new figure 2 better help illustrating the paragraph and thus may help answering this question. We however acknowledge the time slot even small before reaching equilibrium may modify the results of subsampling and following your question, we investigated the effect of subsampling during those first generations.

Action: We thus included in supplementary materials a combination of sampling effects out of equilibrium to address this point (see Fig 2 and Fig S2). The effect of subsampling out of equilibrium is either negligible (particularly for genotypic parameters), or exacerbate the subsampling bias, as some parameters like variance of FIS become hardly informative about rates of clonality. We commented on those newly shown results in the end of the result section where this exercise is now presented.

As a last comment: the discussion is long and you might be able to reduce its length without losing much information: see also reviewer reports.

Action: we reduce the discussion length.

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. Done
-Details on experimental procedures are available to readers in the text or as appendices. Done
-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.”
Done

Reviews

Reviewed by anonymous reviewer, 2019-04-10 16:06

I am unable to say whether this paper has merit. However, I can say that it would be easier to review than it was if the matters it discusses were defined precisely. A first example is "clonality" itself. Authors should be sensitive to the fact that in another context, "clonality" is the collision probability associated with pairs of rearrangements in the adaptive human immune system. After you say what clonality is, why, intuitively, do only high values of {\bf \it c} influence genetic description of {\bf \it R} and {\bf \beta}? More generally, this reviewer's task could have been helped by statement of a precise mechanism by which observations are generated. Otherwise, conclusions are qualitative at best. Here are some items of concern.

Action: see answer to the first comment of the editor, and Box 1

There is discussion of a machine learning approach to a 12-class problem in classification. What, exactly, were the 12 classes? What was the methodology for "machine learning?" One infers that "neural nets" were used to pick features and also to do classification, but these are only guesses. What was the larger list of features from which the neural net (if that's what was employed) picked its features? To what extent does the "power law" really apply?

Comment: we detailed our approach in the material and method section, page 15. We used a classical and straightforward Bayesian supervised learning method. We used Bayes theorem for predicting the posterior probability of a new combination of values of genotypic and genetic indices to belong to a class. The likelihood of a class was previously defined by fitting multidimensional Gaussian distribution on a training database (here all results from all simulations we made) in which the probability density distribution of each genotypic or genetic indices knowing a specific rate of clonality constitute one of its dimension. Please, see page 15: “we used the results obtained from simulations as classifiers to train a Bayesian supervised learning algorithm. We used simulation results to compute the approximated nonparametric probability distributions of genotypic and genetic descriptors (*i.e.*, the seven *features* $\varphi_7 = \left[R, \beta_p, \bar{r_d}, Mean[F_{IS}], Var[F_{IS}], Skew[F_{IS}], Kurt[F_{IS}]\right]$) under known rate of clonality.” We however missed to explicitly write down what were the twelve classes, even if we previously mentioned that our purpose was to infer a rate of clonality, and that mathematical formalizations answer on this point ($L\left(R, \beta_p, \bar{r_d}, Mean[F_{IS}], Var[F_{IS}], Skew[F_{IS}], Kurt[F_{IS}] \middle| N, c, \mu\right)$.

Action: To clarify, we modified as follows: “a *classifier* with 12 classes (one class for each rate of clonality to be inferred, c=0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99 and 1) referred to hereafter as C_{12}.”

Reviewed by David Macaya-Sanz, 2019-04-23 06:15
Download the review (PDF file)

The paper addresses the topic of the relationships between the rate of clonality (c) and some genotypic and genetic parameters often used to assess it and their empirical estimations basing on sampling. It is a very necessary study, as it is highlighted by the finding of somewhat counterintuitive results.
Probably, the parameter more generally used to assess clonality is genotypic richness (denoted by R) and this work shows how not only this parameter is not linearly dependent of c, but the relationship slightly depends on the population size. This is important, because to actually have a strong decline in the genotypic richness in a population, c has to be rather high.
I have found the paper very informative and an eye-opener. In general, I recommend its endorsement but I have some comments that I believe should be addressed in advance.

Some major concerns:

Why did you use different mutation rates for somatic mutations and for sexual events? Can you introduce your rationale in the text? If it is related with germinal cell lines, I am not how this approach could be extrapolated to plants. Have you made a sensitivity analysis to evaluate the impact on your results of choosing different mutation rates?

Comment: We chose those mutation rates by taking into account the possible mutagenic effect of recombination during meiosis (Webster & Hurst 2012), as reflected in the literature where estimates of range of rates have been proposed for several organisms. Without recombination (a situation that would apply to the simple mitosis at the origin of most clonal events), the high end of estimates of DNA polymerase fidelity is estimated around 10^{-8} and 10^{-9} mutation/bp/generation (McCulloch & Kunkel 2008). Such estimates match mutation rates estimated on coding genes in most studied eukaryotes (Agrafioti & Stumpf 2007). In turn, mutation rates estimated for sexual eukaryotes across generations for SNPs and microsatellites range from respectively 10^{-4} to 10^{-7} and 10^{-2} to 10^{-5} (Payseur & Cutter 2006). Our goal here was to choose mutations rates that would fall within these estimates and reflect the fact that sexual events including full meiosis should result in a higher rate of mutations than clonal reproduction which lacks the prophase I (Webster & Hurst 2012), suspected to be the most mutagenic one.

For example, in *Saccharomyces cerevisiae* (with the lowest mechanistically-due recombination rate known in eukaryotes), mitotic mutation rate are one to two orders of magnitude lower (10 to 100 times lower) than meiotic mutation on coding genes (Lang & Murray 2008, Rattray et al. 2015). So, as a two orders of magnitude between mitotic and meiosis mutation rates seems conservative on a unicellular eukaryote species and that the number of vegetative multiplication seems not to matter in plant (Watson et al. 2016), we chose mutation rates that sounded realistic and a good compromise to deliver results which interpretation may apply to both SNPs and microsatellites.

We agree that it would be interesting to vary mutation rates as well. This would however be an interesting exercise per se, in order to investigate the compared effects of drift and mutation. This would address a different question than the one we focus on in this work, and given the results obtained at different population sizes, we feel that unless extreme situation (like extremely high mutation rate), a reasonable variation around those chosen rates is unlikely to affect the relative incidence of clonality and drift reported here, or the accuracy of parameters and effect of sampling size to estimate the rates of clonality.

Action: we now improved the explanations for the choice of mutation rates using the information here above

Bibliography:

- Agrafioti I and Stumpf MPH (2007). SNPSTR: a database of compound microsatellite-SNP markers. Nucleic Acids Res. 2007 Jan; 35(Database issue): D71–D75. doi: 10.1093/nar/gkl806
- McCulloch SD, Kunkel TA (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases.Cell Res18(1):148–161.
- Lang GI, Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics 178: 67–82.
- Payseur BA & Cutter AD (2006). Integrating patterns of polymorphismat SNPs and STRs TRENDS in GeneticsVol.22 No.8 August 2006
- Rattray A, Santoyo G, Shafer B, Strathern JN (2015) Elevated Mutation Rate during Meiosis in Saccharomyces cerevisiae. PLoS Genet 11(1): e1004910. https://doi.org/10.1371/journal.pgen.1004910
- Matthew Watson, Alexander Platzer, Anita Kazda, Svetlana Akimcheva, Sona Valuchova, Viktoria Nizhynska, Magnus Nordborg, Karel Riha (2016). Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. Proceedings of the National Academy of Sciences Oct 2016, 113 (43) 12226-12231; DOI: 10.1073/pnas.1609686113
- Webster MT, Hurst LD. Direct and indirect consequences of meiotic recombination: Implications for genome evolution. Trends Genet. 2012;28(3):101–109.

Is selection modeled in any way in the simulations? My feeling is that it is not and it may be an important factor driving the demographics of PC populations. A little of discussion about how selection could affect your inferences would increase the completeness of the paper (though I know is not the focus of the paper).

Comment: We did not formalize in this model selection and changes in demography, as our purpose was to help future works using neutral markers to interpret their genetic diversity to assess the level of clonality. Using selected loci to assess the level of clonality would be hazardous without a clear knowledge of the joint effects of selection and partial clonality, and without having a trusty quantification of the strength of selection on each locus and allele/genotype analysed. Introducing i/ selection (and all their different type: heterosis, background selection, directional selection) with varying intensities and ii/ changes in demography would be two full separate works in themselves, and given the persisting difficulties to infer clonality rates using one-shoot sampling, it would be a premature hasty action to include such complexities. We thus believe that our work should not yet include changes in demography and selection, even if as you, we are eager to understand all the joint effects of partial clonality with other forces and release a general methodology including loci under selection and demographic changes to quantitatively estimate rates of clonality in populations.

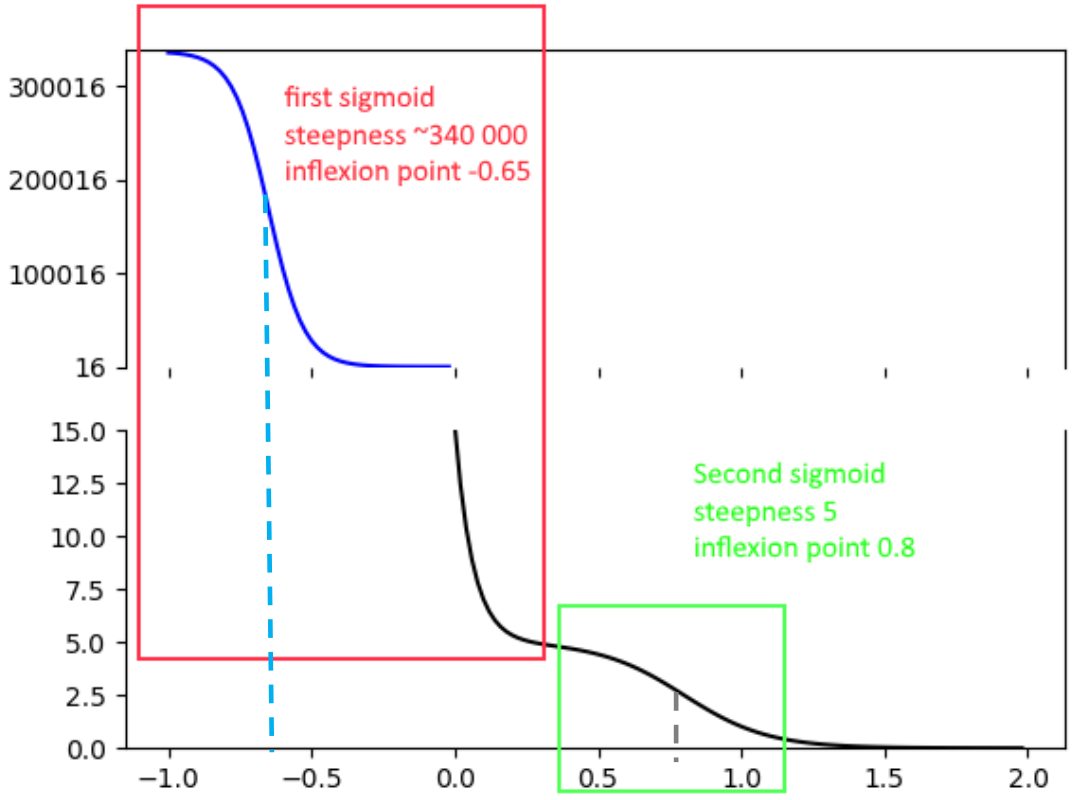
Related to this, in the simulations, how were the individuals to be reproduced clonally selected? Same for the sexual events? Were they selected completely randomly?

Action: we clarified by adding: "*All hermaphrodite individuals in each generation had identical probability of being parents, both in clonal and sexual events. The probability for a single parent to be drawn i) to birth a clonal descent and ii) to sire half a sexual descent followed a Bernoulli scheme, with respective probabilities* $P(clonal\ parent) = \frac{c}{N}$ *and* $P(sexual\ parent) = \frac{1-c}{2N}$ *where N is the population size.*"

The main result of the study, or at least the most stressed one in the discussion, is the effect of sampling size in the estimation of the values of the genotypic parameters. I think it would be interesting to see some more figures depicting this phenomenon from different perspectives. At least to me, it is not very intuitive, and some more figures would help to understand the mechanics of this bias. Maybe focusing in the case of the sample size of 10,000 out of 100,000. Some errors bars could provide more insights of the magnitude of this issue.
Comment: We now changed the figures 2 and 4 into Violin Plots picturing the full distributions of indices as function of sample sizes, number of generations after a first fully-randomly-drawn population at generation 0, and rates of clonality. We also introduced variations in y-axis to better picture on some indices the effects of partial clonality. We believe it would help depicting the studied phenomenon in its full perspectives.

Minor comments:
Format:
You may want to set clearer demarcations of paragraphs, such as first line indentation or separations between paragraphs. Use consistent format for your parameter symbols across the text (i.e. rd). Also use consistently either Pareto b or b Pareto (I prefer the first one). Place a space before and after the equal, the minor and the major signs.
Other:
P3L18: consider changing "based on" by "by" Done.
P3L18: e.g. instead of i.e. Done.
P4L4: Consider removing "but". Done.
P4L6: "Disabling the information" sounds weird to me. Please, consider rephrasing.
Action: we replaced "disabling" by "preventing access to"
P4L9: the value of c… Maybe: the extent of c…. Done.
P7L16: Not sure if the word "during" is appropriate here: In clonal reproduction events,… "during" changed by "in".
P9L6: It could be interesting to report the statistic that reflect the goodness of fit of the population empirical distribution with the theoretical power-law inverse cumulative distribution, if you have those data saved.
Comment: we haven't saved this information. We believe this point was yet addressed in previous papers and reviewed in Arnaud-Haond et al. 2007. But we may have missed the question and its relaying hypotheses motivating this minor comment? You would expect less goodness of fit in some situations?
P14L14: You mention that there are two inflection points in the curve, but I only can see one around c = 0.5
Comment: at N=100 000 individuals per generation, the two inflexion points are -.65 for the first sigmoid and 0.8 for the second (see analytical formulas, sigmoids are very readable parametric curves in which you will explicitly find those numbers). Moreover, the steepnesses of the two sigmoids are very different as attested by their numerator values (i.e., 337335 and 5). This relationship (sum of 2 sigmoids) may be more obvious with an enlarged view of it (we mean outranging the range imposed by the rates of clonality). Please find here a figure with an enlarged view of the mathematical formula set with two y-axis scales (due to the two very different steepnesses), the first having a step height of 337335 between its two horizontal asymptotes, and the second sigmoid having only 5 of step height between its two horizontal asymptotes. Of course, due to the two very different scaling, the curves in blue and in black only make one (We coloured them to picture each of the two sigmoids that are summed; the left part of the second black sigmoid should be in blue in the red rectangle). You may also better see the nearly horizontal linear in-between part (in the range of c=0.3 and c=0.7) specific to the

addition of 2 sigmoids that results to fail fitting a simple sigmoid to the relationship between Pareto β and c.
Whatever population size, the true general form of the relationship between *c* and the average Pareto β is a typical sum of sigmoids. This general relationship is may be more obvious for population sizes as low as N=1000 (see Figure S1.a) where the two inflexion points take place more frankly in the range of the plot.



P15: I would add some other reference to Figure 1 in the text. Done.
Figure 1: Include a note to stress that the x axes are not linear. Done
Figure 2: The meaning of the vertical lines should be stated. In addition, I find this figure difficult to interpret. Most lines overlap, and the plots are small. You may consider removing some c lines and maybe spliting the figure in several pages, removing some unnecessary labels and reducing subplot margins to increase the plotting area. Also, it would be nicer to have the Y axes lined up. Not sure if the text header matches with the plot either. It talks about 10000 generations, but in the plot the X axes finish at 500. It is difficult also to appreciate the relationship between the main text (P17) and the figure.
Action: we now changed this figure for better clarity.
Figure 3: Consider placing it in an edge (top or bottom) of the page to avoid splitting the main text. Done
P19L7: Looking at Figure 3, and not considering the subsampling bias, I feel that the parameter R is enough to evaluate c, being the other ones informative but redundant.
Comment: we already mention this point in the text: "Genotypic parameters evolve gradually with high accuracy of the estimated *c* based on *R*" but we agree that we haven't explicitly mention that those results can only be obtained if all individuals within populations would have been genotyped. **Action:** we added "knowing all genotypes of populations"
Figure 4: Consider here as well the issues raised for Figure 2.

Action: we now changed this figure for better clarity.
Figure 1 and 5: At which generation were those values drawn? **Action:** to clarify, we added "at equilibrium (10^4 generations of quantitatively-homogeneous evolution since the initial random population)"
P21 last lines, P22 first lines: Consider rephrasing. **Action:** we changed to "Nevertheless, this work also demonstrates that the most useful parameters, namely *R* and Pareto β describing genotypic diversity, are seriously affected by sampling density. It raises questions about our ability to detect clonality by sampling only once large populations ~~with available analytical tools~~, and casts even stronger doubts as to the possibility to quantitatively infer rates of clonality with genotypic diversity in large populations."
P22L13: consider rephrasing content within parentheses. **Action:** we modify by "*being slightly positive –although nearly null-, with a random error depending on the strength of genetic drift*"
P33L3: Capital letters. Done.
Disclaimers: Since English is a second language for me, I have some limitations when assessing the quality of the writing. My impression is that the overall quality is more than adequate for a scientific publication, but I may have overlooked some errors. In addition, I have no experience in the use of machine learning, so I can only scrutinize that analysis in a shallow manner.

Reviewed by Marcela Van Loo, 2019-04-30 19:48
Download the review (PDF file)

I have reviewed the preprint entitled" The discernible and hidden effects of clonality on the genotypic and genetic states of populations: improving our estimation of clonal rates (DOI https://arxiv.org/abs/1902.09365) by Solenn Stoeckel, Barbara Porro, Sophie Arnaud-Haond. This is an interesting manuscript, which I really enjoyed to read. I appreciate that authors present us individual-based simulations varying population sizes, rate of clonality, number of generations, subsampling etc. to give us insight how the clonality is theoretically associated with commonly used genotypic and genetic indices. The authors also tackled sample size, which effects the indices, and developed guidelines to infer clonality, which to my knowledge, biologists missed so far for population genetics analyses of clonal and partially clonal organisms.

Although the manuscript is overall well written I´m asking for some changes to be done. They will improve understandability of the preprint. Below, I go through these and other concerns:

Introduction:

Although the introduction is well written in the terms that the authors provide a good theoretic framework, I found confusing that a) authors refer in their study to partial clonality only, although for simulations they used the rate of clonality ranges from 0 to 1, representing thus also the strict sexual, and strict asexual reproduction.

Action: We better stated this point in the introduction (p.6): "*We used comprehensive forward individual-based simulations to obtain the theoretical distribution of genotypic (genotypic richness and size distribution of lineages) and genetic (departure from HWE and LD) parameters describing the population composition at increasing rates of clonality from 0 to 1, including all population with some clonal reproduction, hereafter denoted PC to be short,*

ranging from partial clonality (0<c<1) to strict clonality (c=1), and strict sexuality (c=0), denoted sexual populations."

I have also problem with the term "rate of clonality", which in my opinion, wasn´t explained and defined, but it is used throughout the entire manuscript. Do the authors understand the rate of clonality as a proportion of individuals in the population that do not undergo sexual reproduction (sensu Halkert et al. 2005)? If so, then they have to take into account that depending on the life cycle of the organism, the rate can be constant over time (as they used fixed rates in the simulations) or fluctuates owing to periodic events of sexual reproduction representing a sort of partial clonality (which however to my understanding was not simulated). Please provide clear definitions for both (partial clonality and rate of clonality you study), make it clear in the text and check this out throughout the text of introduction.

Comment: we agree this is an important point. Please see our Action answered to Myriam Heuertz and referee 1 on her similar comment, and the Box 1 we added to clarify the use of terms.

Page3, Line 5: I don't understand the following part of sentence: This mode of reproduction called partial clonality (PC) is particularly relevant for… ensuring human development… Is this linked to pathogenetic species? Unless the readers are skilled in prokaryotic, protozoan and other pathogens, and parasites, this is far to obvious.

Action: we now cut this sentence in two parts. "This mode of reproduction called partial clonality (PC) is used by many species strongly relevant for understanding ecosystems, life evolution and ensuring human development (Schön, Van Dijk, & Martens, 2009; Yu, Roiloa, & Alpert, 2016). PC is particularly important for understanding the success of pathogenic and invasive species, presents some importance in challenging environments or at the leading edges of distributions (Barrett, 2016; Barrett, 2015; Tibayrenc & Ayala, 2012; Yu, Roiloa, & Alpert, 2016)."

Page7, Lines with mutation rates for both, somatic and sexual reproduction. Please provide references for the rates and provide an explanation for selecting them.

Comment: please see our answer to David Macaya-Sanz on his similar comment.

Fig.1, Fig3. (and S2a, S2b, S2b; S5aS5b) Is it possible to select different scale on y-axis to increase the visibility and readability of those different colours?

Action: We completely changed those figures including variations in the axis scalings to better picture effects and identifiable signals. We hope it will help readability and future readers tackling with their complexities.

Discussion: Here in general, less detailed description of some examples could be used. I am positive that in the coming years a case-by-case re-evaluation of the clonal studies will happen (as the authors suggest), this is how the science usually works and improves.

Action: we now reduced the discussion which also helped avoiding details and repeatitions.

Pages 22-23, please avoid repeating of previous sentences and shorten these two pages.

Action: we reorganize and shorten this part. We now reduced it of ¼ pages over 2, and magnify the readability of the possible conjecture of our results on populations under selection, migration and on general polymorphism.

Pages 24-25, I would suggest to shorten the story (or completely skip) the story on Alexandrium minutum?

Page 26, Is the study on Aphis glyceris really appropriate to be stressed here, when in that study two-time steps sampling was used, which is not related to any simulation done by authors

Comment: Using those examples, we believe we move one step forward from our pure theoretical results by explaining some previously discussed anomalies (*sensus* Thomas Kuhn 1962) to concretize the interest of our results for the future applied studies that would interpret similar results.
In this sense, we really think those examples illustrate our theoretical findings and are enlightening on the difficulties to link congruent R and β values to clonality. *Alexandrium minutum* illustrates how genotypic indices may completely miss a massive clonal blossom due to the structural susceptibility of genotypic indices to sampling biases, and the fact that on haploid species (like algae, fungi, many eukaryote diseases like babesiosis, some trypanosomiases phases, and other protozoa) only trusting such indices by default would be problematic. Assessing clonality with such approaches may have deep consequences, as when studying newly emerging protozoan deceases in cattle for example with no or few biological knowledge on their life cycles.
Orantes *et al.* study aggregates all the contradictions often founded in large populations using cyclical parthenogenesis which leads them to discuss about 1/ the meaning of those indices, 2/ questioning the interest of population genetics studies in such species. In their study, genotypic indices are in total contradiction with our biological knowledge (and afterwards in many parthenogenetic species literature). With hindsight and our theoretical results, such results better seem to reflect the change in population sizes than in reproductive mode.
We chose those two examples because they are illustrative, and they were/are still discussed in conferences on partially clonality. Moreover, we benefit, beyond those observed values, of a great amount of biological knowledge that help understanding, together with our theoretical results, what biases could have explained such results. We believe that discussing those examples will help shedding light on those repeated observations and will help future applied and theoretical studies.