

## Associate Editor

Comments to the Author:

Hadjivasiliou et al. investigate the evolution of mating type switching using a simulation model. They show that mating type switching can evolve in response to stochastic distortions in the distribution of mating types arising during asexual proliferation. The reviewers agreed that the work has the potential to make an interesting contribution to the field but raised some concerns. I agree with this assessment and believe that the paper is, in principle, suitable for publication in *Evolution*.

We thank the Associate Editor for his positive remarks. We think that Dr Azevedo and both reviewers have raised some important issues and we have revised our manuscript and analysis accordingly. We believe that our MS has been substantially improved as a result. We provide a point-by-point reply to the Editor and reviewers' comments below (our response in blue). Page and line numbers refer to the revised manuscript.

### Specific comments

1. A word on terminology. Ciliate biologists don't refer to the stochastic mating type determination in *Tetrahymena thermophila* as mating type "switching." (I confirmed this with my colleague Becky Zufall.)

Thanks for pointing this out. We have used both phrases in the paper, and indicate at a suitable early point in the Introduction (page 3, line 41) and Discussion (page 16, line 376) that we use the two interchangeably, as both phenomena are in need of an evolutionary explanation that can be provided by our model.

2. Fig. 2B shows that switching is increasingly favored as  $g$  increases and  $N$  decreases. However, even at the lowest number of rounds of vegetative growth ( $g = 10$ ) the largest populations ( $N = 5000$ ) show  $q_{\text{fix}} \sim 10\%$ , approximately twice the neutral expectation ( $q_0 = 5\%$ ). Is  $q_{\text{fix}} \sim 5\%$  when  $g = 1$ , or does an advantage of switching persist? What about when  $g = 0$ ?

This is a good suggestion, thanks. We have run our simulation for  $g = 0, 1$  and  $5$  (not considered previously) and we have included the relevant points in a revised Fig. 2B. As the figure now shows,  $q_{\text{fix}}$  is approximately  $0.05$  for  $g = 0$  (as expected, there is no advantage for switching if reproduction is only sexual). When  $g$  increases to  $1$ ,  $q_{\text{fix}}$  increases to roughly  $0.08$ , so there is a small advantage relative to drift even for  $g = 1$ . This is in line with Fig. 2A showing that the mating type ratio becomes distorted even with a single vegetative round. The value of  $q_{\text{fix}}$  increases further to approximately  $0.095$  for  $g = 5$  and then to  $0.01$  for  $g = 10$  (as discussed in the MS).

3. The two-fold advantage of switching when  $g = 10$  and  $N = 5000$  appears to be approximately independent of  $N$ . Why is that?

It is true that the advantage of switching varies amongst different population sizes only for larger  $g$  (the distinction is just about obvious for  $g = 10$  but vanishes for  $g < 10$ ; see new Fig. 2B). We think that is a reflection of the difference in the mating type distortion in populations of varying  $N$  increasing as  $g$  increases (see Fig. 2A). The difference in the mating type ratio distortion for smaller  $g$  is smaller (i.e., the frequency ratio of rare/common lies within the limited range 0.7-0.9 for  $g = 10$ ). Hence, even in a population of say, size  $N = 100$ , selection for switching is relatively weak. At the same time, smaller populations experience more drift and so the switching gene is more likely to be lost during the vegetative growth before selection can “see” it. This results in the observed neutralization of the role of  $N$  for smaller  $g$ . We discuss these effects in light of Fig. 2A (page 8, lines 182-187).

4. In the legend of Fig. 5 you state that the red lines were "estimated from the simulations C-F." Please provide details.

The red lines in C-F were obtained by running our model for a non-switching mutant that bears the same cost of switching (defined by the value of  $c$ ) as its switching counterparts. We now briefly mention this in the legend. We also provide more details for all simulations in the online Supplementary File “Computational Methods”.

5. It should be "two mating type alleles" not "loci" (l. 126)

Thanks, corrected!

6. Typos in ll. 140, 184 and 512.

Corrected, thank you!

Dr. Ricardo Azevedo

## Reviewer 1

### Comments to the Author

Mating type switching is an interesting aspect of the genetics of sexual behaviour, which is arguably under-studied. Hadjivasiliou et al. used simulation models to examine the role of several relevant ecological and genetic factors in the evolution of mating type switching. Their methods are accessible, and the results make sense on their own, and in light of other observations. Although they are in some cases unsurprising, I think these findings will help to generate interest in the topic.

Thank you for the summary and positive review!

An issue the authors do not address, which I think bears further investigation, or at least discussion, is that of spatial population structure. The authors assume that mating occurs at random apart from the requirement of dissimilar mating types (line 124). But in many systems mother and daughter cells will tend to be adjacent following asexual reproduction, and potential mating partners must also be physically adjacent, so that potential mating partners will tend to be more closely related than expected under random mating. In the absence of mating type switching, relatives (or adjacent cells) will have the same mating type. It seems to me that this proximity of relatives might be quite relevant here, e.g. in determining the optimal switching rate. I don't think the treatment with a cost of mating, which is interesting and effectively reduces the number of possible mates in a different way, quite addresses this issue.

An extreme example would be that a haploid genotype that is the sole colonist of a new patch would be unable to persist in the context described here, without some of its descendants switching mating type during asexual growth. Such bottleneck events would presumably generate strong selection for switching.

It may or may not be straightforward to create a spatially-explicit version of the model (although such models are not uncommon), but I think this issue should at least be discussed in the manuscript.

We agree that the evolution of mating-type switching in spatially explicit settings is an interesting and important question. However, addressing this question requires substantial additional modeling effort. Various alternative implementations are possible; for example we could have studied a model of colonization by a small number of individuals followed by expansion (equivalent to the bottleneck events mentioned above). We expect that drift would affect the mating type ratio but also the strength of selection on switching alleles as in the current model. An alternative would be to run a

more explicit agent-based model where individuals occupy specific spatial positions but this again is not a trivial extension. These developments seem worthwhile but are unlikely to change the general conclusions of our work, and feel like being too much for this paper. Our analysis suggests that in such a case we expect higher switching rates to evolve. We have followed the referee's advice and added a paragraph in our conclusion discussing these issues (page 15, lines 366-374):

"Our model does not explicitly incorporate a spatial component. For example, some yeasts and other fungi are restricted in their movement and so are likely to find themselves amongst descendants that would share their mating type in the absence of switching. Alternatively when colonies are formed from a single or few individual founders, local mating type distortion is likely to be extreme. Such discrepancies in the mating type ratio are likely to be similar to those caused by the vegetative period in our model. They would cause drift in the ratio of different mating types and in the switching locus itself, and so are likely to favor switching. The examination of a spatially explicit extension of our model is needed to fully elucidate how spatial structure influences the evolution of the switching rate."

Minor comments:

It does not seem fair to imply that diploidy is "essential to adult life" in budding yeasts (line 77), which seem to grow well as haploids when mating type switching is prevented.

Agreed. Budding yeasts indeed happily grow as haploid. They do however have a preference to be diploid and compatible haploid spores will mate when mixed unless they are prevented from doing so. We have changed the end of the pertinent sentence which now reads: "[...] restore the diploid condition which is the preferred adult state [...]" (page 4, line 81).

The use of the term 'protist' may be confusing, as there seem to be multiple definitions, and I'm not sure that such a term is necessary here.

We used the term "protist" to refer to simple eukaryotes with isogamy. We agree that the term can be ambiguous (e.g. some protists are anisogamous) and have changed all references to "protists" in the text accordingly.

Check for typos, lines 21, 113.

Thank you, corrected.

## Reviewer 2

The paper by Hadjivasiliou et al. deals with the evolution of mating type switching in a species with both asexual and sexual reproduction. The work is only based on simulations results. They showed that under a variety of conditions, mating types switching is expected to evolve, because of mating type ratio discrepancy due to random drift.

The paper is really interesting, well-written and well-documented, the question is original. The basic results are not surprising but they are worth to be published since to my knowledge this is the first time it is clearly shown that mating type switching can evolve because of such a simple mechanism. I am thus really excited by the paper. However, I have several concerns about the paper (detailed below). First, much more detailed of the simulations, the model and the algorithm is needed. Many informations are lacking. Second, and most importantly, some results seem flawed to me. I might be wrong and if it is the case, I would be happy to learn something. If I am not wrong, the authors have to correct and verify their results before publication.

Thank you for the summary and positive remarks. We appreciate the detailed review and we think it has helped us to greatly improve our MS. We agree that further details of our methodology and simulations are important. We have produced a new supplemental document that provides details of the general model, methodology and simulations used to produce all figures. We also appreciate that our presentation of Fig. 2A (and similarly Fig. 7A) was insufficient and we regret that it caused confusion. The problem was that we plotted both Figures starting from  $g = 1$  instead of  $g = 0$  and so what appeared like the initial mating type ratio was in fact the mating type ratio after a single round of vegetative growth. We have now corrected this in both Fig. 2A and Fig. 7A.

- My main concern: there might be a big flaw in the computer simulations. As I understand, Fig. 2A shows the ratio of frequencies between the commonest vs. the rarest mating type as a function of  $g$ , the number of clonal reproduction between each event of sexual mating. In the case of two mating types, at the end of the sexual phase, the ratio should be exactly 1 since a zygote is necessarily M1M2. However, in Figure 2A, when there is no clonal reproduction  $g = 0$ , we should thus expect the ratio to be exactly 1, which is by far not the case here. Hence, there are three possibilities: either I did not understand properly how the simulations were performed, in which case, simulations should be detailed more precisely. Either there is big flaw in the simulations code, in which case it is possible that the whole results are wrong. Either this is a simple graphics

problem, in which case it can be certainly fixed very quickly. Since the results make sense and are mostly intuitive, I am expecting that there is no big flaws. However, it is impossible in the present state to be entirely confident on the results in the paper. I therefore reviewed the rest of the paper making the hypothesis there is no big flaws, but with an suspicious feeling.

Thank you for flagging this mistake. We appreciate your open-mindedness in reviewing the paper besides this confusing figure. Like we mentioned above Fig. 2A (and Fig. 7A) were misleading because we mistakenly plotted the mating type ratio starting at  $g = 1$  instead of  $g = 0$ . We have now corrected both figures. Note that the same mistake was not carried forward with our old version of Fig. 3B confirming that this was an error in plotting Fig. 2A and Fig. 7A alone.

- L.42 : this is only true in *Tetrahymena thermophila* and not for the whole genus (see Phadke and Zufall 2009).

Clarification added in all our references to this study.

- L.131 : “the population is allowed to grow back to carrying capacity” is not clear: population grows clonally, randomly, who reproduces, stochastic growth, etc. ?

If the population size is below carrying capacity after sex we sample with replacement until the population reaches carrying capacity again. We have added a clarification in the main text “We assume that the carrying capacity of the population is fixed and sample without replacement to return the population to  $N$ . ” (page 6, line 124) and repeat this in the supplemental “Computational Methods” where our model and simulation methods are outlined in more detail.

- In the section “model”, a paragraph should be devoted to initial conditions, and how simulations were run, very important especially in works only by computer simulations: criterion for stopping simulations, etc. This paragraph should also include information about the locus “S” and how the mutated allele “S2” was managed : how it is introduced in the population, at which frequency, etc. Since the paper is centered on this “S2” allele, the authors must explain precisely what was done with it : how  $q_{fix}$  was estimated, how many independent repetitions, etc. How are the ratio of frequencies rarest/commonest estimated ? How many repetitions, how many generations, etc. How did you manage the extinctions? We

have far too few informations about this.

We regret that our presentation of the simulation methods was inadequate and thank the referee for flagging this down. We have now added a supplementary text (“Computational Methods”) where we provide details of our simulations including our initial conditions, implementation of mutation, method for estimating  $q_{fix}$  etc. This also includes details of how simulations are terminated, and details of our continuum-of-alleles implementation (relevant to Fig. 5 and 6). These details could not be included in the main text due to space restrictions but we point the reader to the supplementary text at the end of the model section (page 7, lines 140-142):

“ Further details about initial conditions and simulation routines are provided in a supplemental text file (Computational Methods). ”

The number of independent repetitions is stated in each figure legend.

- L.169 “the fixation probability is highest with  $g = 200$  independent of  $p_s$ ” does not seem to be exact: the authors do not show results for higher  $g$  values, hence “highest” is not appropriate here.

Fair enough. We have changed this to “higher” (page 8, line 172).

- L.174-184: The balance drift-selection is basically the same processes here, and the authors are actually repeating twice the same thing. I would suggest to merge those two paragraphs.

We agree that the underlying forces are the same here. However the first paragraph deals with the effects of  $g$  and the second with those of  $N$  (population size). Although this is a bit repetitive we think it is important to treat  $g$  and  $N$  separately so as to clarify why they each have the observed impact. To emphasize that we are discussing similar effects we now begin the second paragraph by saying: “A similar argument can also explain why varying the population size...”, page 8, line 182.

- L.199: How the cost is taken into account is by far not enough detailed. How exactly is it implemented in the simulation algorithm? How does it affect the survival of individuals? What is the probability of an individual to asexually reproduce from generation  $g$  to  $g+1$  ?

The cost function  $f$  defines the survival probability of a cell at each vegetative step: all cells divide (so the population grows from  $N$  to  $2N$ ), then we sample from  $2N$  to return to  $N$ . When switching is costly we introduced biased sampling so that  $f$  determines the probability of an individual cell being sampled from generation  $g$  to  $g+1$ . We have added a clarification in the main text (shown in italics here, page 9, line 202-204): “The cost is applied at each asexual growth round: *all cells divide to produce a daughter cell leading to a population of size  $2N$ . Cells are then sampled with a probability defined by  $f$  until the population size returns to  $N$ .*” We have also included a detailed description in the SI where we provide the methods used to generate Fig. 4 and Fig. S3.

- L.264-277: I think this paragraph is not appropriate: the result is interesting but really not surprising as the authors say it here. It is intuitive, not to say trivial, that mating type switching will be less advantageous with  $m > 2$  because the proportion of compatible partner is larger for a given mating type. This a phenomenon well known: negative frequency dependent selection. I would suggest to largely reduce and neutralize the whole paragraph.

We agree with the reviewer that negative frequency dependent selection reduces the advantage of having additional mating types as the number of mating types increases. At the same time, the more mating types there are in a population, the more “uneven” the mating type distribution (rarest vs. commonest) will be over increasing rounds of vegetative growth (see Fig. 7A, B and a new figure, Fig. S6). This observation could lead to the expectation that more mating types result in stronger selection for switching (hence our use of “surprising” here). But as the reviewer points out, with more mating types, the most common type tends to be at a lower frequency to start with (at equilibrium =  $1/m$ ), and there are more compatible individuals with which it could mate, so the proportion of unmated cells declines with  $m$  (Fig. S5). This leads to weaker selection for switching when  $m$  increases. In accordance with the referee’s suggestion, we have avoided describing our results as “surprising” here, but we think discussing these different effects is important and so have not reduced the paragraph significantly.

Note that our model also suggests that the impact of an increasing number of mating types depends on the mating kinetics (see next reply).

- I think what would be interesting is to allow mutation to introduce regularly new mating types and look at the joint distribution between the mating type switching probability and the number of mating types in a population.

We agree that the evolution of the number of mating types is an interesting topic. We are currently working on another MS on that topic and this is not a trivial question! We



think it would be interesting to combine the two works and allow both switching and the number of mating types to evolve but this would require a lot more work and we think deserves to be studied in a new article. We now mention this briefly at the end of the article (page 15, lines 363-365):

“It would be interesting to further explore the evolution of the number of mating types in populations that sustain switching, but this is outside the interest of the current work.”

L.286-290 “As anticipated,  $q_{fix}$  now increases with the number of mating types (Fig. 7C)”  
: I do not agree at all with these statements, and I do not understand at all these results. I do not think that “anticipated” is appropriate because the work by Iwasa and Sasaki (1987) showed that a higher number of mating types is favored with a cost to mate finding because it increases the finding rate of a compatible partner. Following Iwasa and Sasaki (1987), it is expected to be easier to find a compatible mate, and thus the cost of not finding a mate is avoided, when the number of mating types is higher. If this is true, then I would expect that switching would be even less favored: if there are plenty of different mating types in the population, there is a high probability to find a compatible partner, even without switching. We can see that looking at an extreme case: let the number of mating types be infinite, then switching would be neutral! This seems not to be the case in your figure 7C as the probability of fixation increases with  $m$ ! I am thus totally unconvinced by your explanations, and now, really surprised by your findings, contrarily to the previous paragraph. Two possible things: either there is a very strange and unexpected mechanism that would explain your findings, and I would be really happy to learn more about that, either there is a flaw somewhere.

We appreciate the referee’s concerns here. We think that exactly the same forces that lead to selection for further mating types in the Iwasa and Sasaki paper also lead to stronger selection for switching in our work.

When we impose more strict mating kinetics (speedy mating scenario) so that cells are only given one chance to mate (cells of the same mating type that are sampled together are removed from the mating pool) we see an increase in the fixation probability for switching. As Fig. 7A and B and (a new Fig. S6) show, the more mating types there are in a population, the more uneven the mating type distribution during asexual growth. In this situation, more common mating types are more likely to be sampled together earlier on and so lose the opportunity to mate, whereas rarer types will almost certainly mate if sampled. Selection for speedy mating leads to selection for traits that ensure quick mating, be it the evolution of additional mating types or the capacity to switching from a more common to a rarer type. We have added a few sentences in the last paragraph of this section to clarify these points (page 12-13, lines 287-295):

“In this case,  $q_{fix}$  increases with the number of mating types (Fig. 7C). This is because more common mating types suffer a greater disadvantage when the mating type distribution is more uneven as they are more likely to encounter a cell with the same mating type and so be removed from the mating pool. This effect becomes weaker as the number of mating types increases because the relative frequency of the commonest type decreases with  $m$ . It leads to a plateau rather than decline in  $q_{fix}$  (Fig. 7C) because as the number of mating types increases so does the rate at which mating types are lost during vegetative growth (Fig. S6). ”

We have also extended our x-axis on Fig. 7C to address the referee’s concerns about very large mating types. As stated above, the number of mating types increases we observe a plateau in the fixation probability. As the number of mating types increases the frequency of the most common allele decreases and so the incremental increase in  $q_{fix}$  as  $m$  increases decreases. This does not begin to decline as  $m$  tends to infinity because as the number of mating types increases so does the rate at which mating types are lost in the vegetative round (Fig. S6).