Review of: "Mutational selection: fragile sites, replicative stress, and genome evolution"

Milind Watve

Potential competing interests: No potential competing interests to declare.

The manuscript is a “stress-test” for the reviewer and I hope I am at least at the minimum passing score.

The paper is a collection of several distinct but interrelated ideas that are fascinating but not articulated too well making the job of a reader difficult. I had to read the paper multiple times to understand the arguments and I am not sure I still have sufficient clarity about some of them. Nevertheless the paper triggered a lot of thinking, realization, striking new possibilities as well as doubts and more questions.

I will first try to place this work in the broader picture of evolutionary biology, then I will add some novel possibilities about what the author calls “mutational selection” and lastly make some suggestions for ease and clarity of the reader.

1. Evolutionary theory has three distinct phases. The first one was pre-Darwinian, Darwinian and a little bit of post-Darwinian radical but qualitative thinking phase that laid the foundations of the field. The second phase consisted of rather abstract and oversimplified but mathematically sound pursuit of core principles. This phase made simplistic assumptions such as one gene one character, modularity of traits; it treated inheritance, mutations as black boxes and avoided getting into the molecular details. The simplicity of this phase allowed identifying the core principles such as spontaneous mutations, types of selection, socio-biology, life history and aging theories, optimality theory, sex and sexual selection, neutral evolution and molecular clock and many others. But simultaneously the simplification left a big gap between theory and reality.

I would like to call the third phase as “Devil’s phase” because devil lies in the details!! What we need in the ongoing and future third phase is to get to know how evolution works at the level of complex reality, particularly at the level of molecules, signalling pathways, genome organization, systems network etc. All that we have at the molecular level at present is molecular phylogeny. But the question ‘why’ remains largely unattended at the level of genome organization, multi-omics network structures, pleiotropy, epistasis and so on. For example antagonistic pleiotropy is very much used in phase 2 but the question why it is there at all, is not addressed. Most of the assumptions made in phase 2 need to be addressed with a ‘why’ question in the devil’s phase now. Some contributions to the devil’s phase have been happening in the last 2-3 decades but rather as isolated fragmental pieces of work and lack comprehensive articulation. David Haig’s paper is a bold step going from phase 2 to 3 and therefore is extremely important.

2. The complex dynamics of selection acting at the gamete stage, a part of which is captured in this paper, raises even more diverse possibilities. Reading this paper provoked my imagination to capture potential phenomena that are either

implicit in this manuscript, and need to be made explicit or David has not (yet) thought of these possibilities clearly.

i. Evolution of pleiotropy: Most of the arguments about mutational selection would be strong if the same set of genes are relevant in the gamete stage as well as post zygotic developmental and adult stages. If different sets of genes regulate gamete phase and adult phase, the arguments in this paper become weak or redundant. For genes and molecules involved in basic cell survival and replication, we can safely assume that the same set of genes operates. But that may not be the case with regulatory and signalling pathways that have different tasks in different phases. We can perceive a possibility that pleiotropy in these genes also evolves as a result of mutational selection. The joint dynamics of selection at haploid and diploid state will favour pleiotropy between regulatory and signalling pathways too. Such a possibility receives a passing mention in the paper and is perhaps implicit in the logic. But this is a possibility with very important consequences and needs to be dealt with elaborately. Indeed such pleiotropy is extremely common. For example many of the cancer related genes have important roles in gametogenesis as well as in early embryonic development. That includes P53, EGF and EGFR, RB, notch and so on. Civetta and Ranz (2019) (https://pubmed.ncbi.nlm.nih.gov/31572439/) list genes that influence sperm competition in mice. Interestingly most of them have some or the other role in cancers as well. This pleiotropy would result into mutational selection becoming a mechanism by which cancer causing mutations are prevented from getting inherited. By classical cancer models, cancer needs accumulation of multiple mutations and individual mutations alone were considered neutral. If there is pleiotropy, they will not be neutral and gamete or embryo level selection will wipe out such mutants. I am currently working on this concept, an earlier version of this work is in a preprint (https://www.preprints.org/manuscript/202303.0431/v1) and a more detailed version is under preparation. Certainly your concept of mutational selection is helpful in uncovering some aspects of the dynamics of cancers.
ii. Selection on copy numbers: The paper talks about copy numbers, length of introns and length of DNA in some context. There are selfish elements and other mechanisms for increasing the copy numbers. There should be something that limits copy numbers too. I thing mutational selection has a role here. The fundamental question what is rate limiting in cell replication is important here. The length of the genome decides the time required for replication. For cell divisions in certain contexts, chromosomal replication time need not be rate limiting. Only when rapid multiplication to generate large number of progeny matters, it can become limiting. In bacteria, this kind of selection is likely to be important as a selective force in plasmid copy number selection (Watve and Watve 2010) (https://pubmed.ncbi.nlm.nih.gov/20195362/). Oogenesis and spermatogenesis will differ dramatically in this respect. Since sperms need to be generated in large numbers over a short time, replication time can be rate limiting. Therefore in spermatogenesis selection can act against large copy numbers and excessive intron lengths. This may not happen in oogenesis or in adult stem cells or germline cells. Ultimately the dynamics of selfishly replicating segments and selection for shorter length during spermatogenesis would shape the copy numbers and sequence lengths. This is a multi-level selection problem that needs to be modelled carefully.

I feel more possibilities are raised by the dynamics of mutational selection and we need to be open for further exploration.

3. Now from the readers' perspective: I asked myself why I needed to read multiple times to get the central idea. I presume I am not an outright dumb reader, but certainly an average reader. What would have made my job easier? Here

is what I think.

a. Most evolutionary biologists and population geneticists work with typical higher sexual diploid organisms and assume that this is THE general theory of evolution. Majority of life on earth is haploid and/or asexual. The dynamics of selection can be quite different in the two. It is necessary to state clearly in the beginning that these arguments are specific to such and such class or organisms.
b. Right in the beginning the purpose should be made clear. Which questions are being addressed or which possibilities being explored and what is new in it was what I grappled with until quite late in the manuscript. The manuscript cites earlier work but what were prior ideas and what is newly being proposed is often not crystal clear.
c. The jargon appears unnecessarily complex. My feeling is that all the ideas can be described quite well just by describing the dynamics between selection in the haploid state versus that in the diploid state of life cycle. The intended meaning of "mutational selection" is defined in the text clearly but just going by the word it is confusing because selection acts on mutations at so many stages and in so many ways. Why only one of them is being called mutational selection is not intuitive.
d. The introduction was more confusing than enlightening. I started finding some meaning and relevance from page 3 onwards where the actual logic begins. Prior to that I felt lost.
e. The logic why haploid phase selection would evolve fragile rather than robust genes appears to be left for the reader to work out. It is easy to work out but not very explicitly stated. Because it is not about selection for or against the mutant itself, this might make it appear like a group selectionist idea, while it is not.
f. Synonymous constraints is an interesting riddle no doubt, but it appeared disjunct from other problems. While how haploid stage selection acts wrt the other problems was clear, how it is connected to synonymous constraints is still not clear to me. Does it need to be treated as a separate problem, not in this paper?
g. I could not perceive what was 'competence' and 'difficulty' when it is described first. Only much later, after reading about WWOX and BRCA1, I wondered whether this is what is meant by competence and difficulty. But no real life example is cited when these words are introduced and after specific examples are discussed, they are not clearly referred back to the terms used earlier. The same applies to stress test hypothesis. It took me some time to figure out the exact meaning. Defining a word when it appears first along with a real life example should be a better practice.
h. Some schematic explanatory diagrams would certainly help, as suggested by other reviewers as well.

These are only suggestions. The author ultimately decides how he wants it to be. In the Qeios philosophy, there is no acceptance-rejection. So the author need not do anything to satisfy the reviewer/editor !!

Further, an important question is how one would ultimately test these hypotheses? Can any testable predictions be made? What more patterns we expect to see if the hypotheses are true? This is something that would add substantial quality to the ideas being explored. At this stage exploration is important and we don't expect a 'proof' but any vision on how to test the ideas would be valuable.

I would certainly place the ideas in this paper a big step towards the devil's phase, but I think it should become more reader friendly. I can understand and also have experienced why and how it is difficult to articulate complex ideas that

have been developing in a mind over a long time. We typically tend to take the early appeared components of these ideas for granted and therefore may not feel like making them explicit, but for the reader everything could be new.

Let my excitement after the third reading be found by most readers after the first one itself.