We thank the recommender and two reviewers for providing their constructive comments. We have now revised our manuscript with an aim to address those comments, focusing on improving the connection between the introduction, and the rest of the manuscript with respect to ideas about parallel evolution. Responses to specific comments can be seen below. Recommender/ reviewer comments are shown in blue, followed by our responses in black.

Response to comments from Stephanie Bedhomme

There seems to be a disconnection between the title and the introduction which focus on parallel evolution and the results and discussion which focus on the factors affecting the probability of a gene to carry a mutation by the end of the experimental evolution. The introduction makes the reader expect that the methods developed is going to be able to determine to what extent parallel evolution is due to the probability of the mutation to happen and to selection. In other words, from the introduction, I expected the method to be able to discriminate cases where parallel evolution can be truly taken as a strong signal for adaptive mutations and cases where parallel evolution is due to neutral processes. The methods developed is not reaching this goal, at least not explicitly, and the added value of the methods does not appear clearly to the reader.

We think there is a valid point here. We recognize the potential for disconnection and we have now added an additional figure (Fig 4) and a section in the discussion (lines 450-460) to better connect the paper results and discussion to the idea of parallel evolution. We argue that our modeling approach makes testable predictions regarding the expected level of parallel evolution. We illustrate this on the dataset by contrasting predicted and observed degree of parallelism (estimated as the pairwise Jaccard Index, J).

More details should be given on the experimental design, in particular on the ploidy of the yeast and the reproduction mode they had during experimental evolution (see point 4 of MA).

This is now included on lines 156-159.

The authors rely on the hypothesis that synonymous mutations are neutral to selection, which is a classical one, but they write in their discussion (l 403): “Relying on the assumption that synonymous mutations are selectively neutral (which does appear to be the case for these data)” and I do not see where the neutrality is tested in the study. More importantly, for non-synonymous mutations, they fit a number of models to try and detect the effect of different genomic variables on the heterogeneity in the probability that a mutation rises. Among this genomic variables, some are likely to affect non-synonymous as well as synonymous mutations and their link to selection is not obvious and straightforward. The comment on GC content by MA is going in this direction and similar argument could be developed for CAI and recombination rate. As far as I understand the effect of these variables on the synonymous mutations has not been tested, so it cannot be claimed that they have an effect on NS mutation that they have not on S mutations.

We agree and we now explicitly test these assumptions. See lines 218-219 (methods), lines 296-301 (results), and supplementary info table S2.

All the manuscript is focused on SNP when high levels of parallelism have been found for IS and large duplications and deletions (see for example Tenaillon et al. 2012). I recognized that it is more difficult to derive a modelling framework for them and that the present one cannot be easily adapted but I think that these mutations have a strong impact on adaptation and would like to see some comments on them, at least, in the discussion.

This has now been added to the discussion on lines 444-449.

Response to comments from anonymous

Major comments:

1) GC content is included as a variable in non-synonymous mutation rates, but not as a variable in synonymous mutation rates. One could argue that the failure to detect substantial mutational heterogeneity between genes (i.e. the Poisson had a lower AIC than the negative binomial) implies that GC would not be a significant predictor of mutation rates. This may be correct; however, the significance of GC content in the non-synonymous models is most probably explained by the effect of GC content on mutation rate and not as a predictor of the strength of selection. If GC content does not significantly correlate with synonymous mutation counts, then this points to a difference in the power to detect mutational heterogeneity at non-synonymous and synonymous sites. This difference in power has implications for the interpretation of the results and should be addressed.

We agree that GC content of a gene is a genomic variable that could be a proxy for a variety of processes including mutation and (biased) gene conversion rate variation throughout the genome. We now report the effect of including this variable when modeling both synonymous and non synonymous mutation counts as well as additional tests. While the general point made by the reviewers is well taken, it turns out here that this covariate explains preciously little of the variation in mutation counts. We therefore chose to keep the discussion of this variable to a minimum (but see revised results lines 406-410).

2) It has been consistently found that some substantial fraction of mutations occur in complex events that alter many nearby nucleotides (multinucleotide mutations or MNM; Schrider 2011). This is problematic if the authors' method would tabulate a single MNM event as two or more parallel mutations. Additionally, because MNM events happen on very small scales, typically affecting adjacent nucleotides, they disproportionally cause adjacent non-synonymous changes rather than adjacent synonymous changes. This can be addressed by counting MNM events as single events.

We do not think this is a problem in our study. There are cases of multiple mutations occurring in the same gene within the same population in the data we analyze and so mutations are all >~1000 bps away from each other. We now mention this on lines 162-165.

3) A persistent challenge in experimental evolution is separating relaxed selection on a gene from adaptation. While relaxed selection is arguably a form of parallel evolution, the methods adopted by the authors could provide insight into separating these two forms of evolution. It would be an interesting addition to discuss this in some detail.

We are not sure we understand what the reviewer is suggesting here. However, it is not clear how we would separate relaxed selection from adaptation with our method. This might be feasible if one were explicitly fitting different fitness classes of mutations, but this approach is out of the scope of this study.

4) Because this paper analyzes data from a single experiment, more details on the conditions in that experiment should be included, particularly information on general growth conditions (batch size, frequency of transfer, volume transferred, etc), whether the yeast were grown as haploid or diploid, and whether they were given the opportunity to have sex. This information is crucial to determining the meaning of these results, and should be at least broadly summarized in this paper.

Additional information has been added on lines 156 – 159, as well as an additional reference that details the experimental conditions.

Minor comments:

Line 174 “essential genes” reads awkwardly in the list modifying “each gene.” Perhaps “essentiality of the gene.”

Changed, see line 190.

Line 199 reports a result in the methods section... omitted word “whether.”

There is actually no omitted word here.

Line 360: This observation would be much more interesting and informative if the authors had tested for an effect of r on synonymous mutation counts.

We now include this test. See supplementary table S2.

Line 366: Are the yeast growing as haploids or diploids? If they are growing as haploids w/o sex, then then there should be no opportunity for BGC to occur.

The yeast were grown as haploids without sex in the experiment that the data come from. Regarding the likelihood of BGC, we agree that scope for this process is probably more limited in haploids although yeast is known to undergo mitotic gene conversion (e.g. Bethke and Golin, 1994, Genetics) and there is certainly evidence from other haploid organisms, namely bacteria, that BGC can drive GC content in genomes (Lassalle et al. 2015. GC-content evolution in bacterial genomes: The biased gene conversion hypothesis expands. PLoS genetics).

Expression levels are sensitive to growth conditions. If available, the expression data from growth under experimental conditions should be used for all analyses.

Both the experimental evolution study and the expression study growth conditions are standard YPD growth media at 37 deg C. We now make this clear in the manuscript on lines 157, and lines 189-190.

I favor the definition of parallel evolution being used here, but quite a lot of confusion exists between the use of the terms parallel evolution and convergent evolution. Since both of these terms are used in this paper, it would be useful to clearly define the terms. I would recommend citing an authoritative usage of the term, such as Zhang and Kumar 1997.

We now define these terms and include a citation in the introduction (lines 55-58)

Response to comments from Bastien Boussau

This manuscript aims at understanding the variables that affect parallel evolution in an experiment conducted in yeast. It compares statistical models that include different variables and conclude that gene length or recombination rate affect the rate of mutation. I found this paper interesting and I think the model comparison approach is sound, but in the end I was a bit confused about what had really been achieved. The introduction focuses on parallel evolution, but looking at the methods, it seems like all mutations have been analyzed in the manuscript (lines 145-151, page 7), not only the mutations that occur in genes that have been hit multiple times. So in the end it is unclear to me why the results apply to parallel mutations and not to mutations in general. The authors analyze 414 substitutions in total, assuming that non-synonymous substitutions are under selection, and synonymous mutations are evolving neutrally. However it may be that not all non-synonymous substitutions are under selection. In the Lang paper where the sequencing was conducted, it is noted that some genes have been hit multiple times in the populations, and it is concluded that these genes are likely targets of selection. I think it would be interesting to analyze separately the subset of mutations occurring in those genes only (if there are enough), because several non-synonymous substitutions that the authors chose to analyze may in fact be neutral or nearly neutral. The other experiment I would be curious to see conducted is an analysis of the mutations with respect to the GC content of the arrival state. As I suggest below, GC-biased gene conversion may partly explain why there is a correlation between the number of mutations and the local recombination rate.

As mentioned above in our response to other reviewer comments, we have now added an additional figure (Fig 4) and a section in the discussion (lines 450-460) with an aim to better connect the paper results and discussion to the idea of parallel evolution. We argue that our modeling approach (including all the mutation data, both those that represent parallel evolution and those that do not) makes testable predictions regarding the expected level of parallel evolution. We also agree with the reviewer's point that several of the non-synonymous mutations in this data set may be neutral. In fact, this possibility fits well with our approach. In our statistical models, selective effects of non-synonymous mutations can range from deleterious, to neutral, to beneficial. We now state this explicitly on lines 125-127.

More specific comments:

p4 l66: "genes that exhibit a higher that expected number": than

Fixed.

"We used a codon table model with a fixed tree topology
165 (a comparison of AICs among alternative codon based models indicated this was the most appropriate
166 model for the data set).": this is not clear to me, I's prefer to see the name of the model according to PAML (e.g. M0, M1...).

We now include the following "We used a codon table model (i.e. seqtype = 1; CodonFreq = 3) with a fixed tree topology (i.e. runmode = 0)" on lines 180-181.

p10 l215: "permutation tests instead of relying on asymptotic distribution of the LRTs": it is not clear to me how the permutations were done. What variables were permuted, and how were they permuted?

We have expanded our explanation of the permutations on lines 233-237.

p12 l262 "and MS1: λS = constant*(Li)α": I think in other parts of the manuscript alpha was alpha*1.*

Changed.

p13 l277: "evenly loaded with a number genomic": number of

Fixed.

p13 l291: "that can significantly predict the distribution of mutations": I'm not sure what significantly predicting means.

We have changed the wording here to “significantly improve models predicting the distribution of mutations”.

p14 l307: "mutation counts from Lenksi's long term evolution experiment": Lenski's

Fixed.

p15 l343:" Further evidence that gene length acts as a summary variable comes from the M3 results (summarized in Table 3), where we see that gene length is no longer significant when other summary variables – the principal components – are included in the model.": I'm confused. M3 is a new notation, not found in table 3. If M3 is in fact MN.NBPC, then gene length is included in PC10 already, so I don't understand the argument.

This sentence was inadvertently left in the manuscript form a previous iteration of the models. It has now been removed.

p16 l355: what about the other correlations? Could the number of domains be another "summary variable"?

Yes, this is a valid point. We now point this out on lines 383-284.

p16 l362 "double strand breaks in substantially increases the frequency of nearby point mutations in nearby intervals": remove the first in, and too many "nearby"s.

Fixed.

p16 l365 "Another non exclusive possibility might be the fact that biased gene conversion might vary from gene to gene and also – like selection - affect the probability of detecting variants in evolve and re-sequence experiments": (a point is missing at the end of the sentence) Indeed, biased gene conversion behaves as selection in terms of its impact on the probability of fixation. In that case, wouldn't we expect the variants to be GC biased (cf https://www.ncbi.nlm.nih.gov/pubmed/23505044)? Would it be possible to check the GC content of those variants?

We checked, and saw that GC content is not significant in our model. We also see no bias towards GC in the variants. We now explain this on lines 406-410.

p18 l408 "and move closer the goal of predicting which genes": closer to

Fixed.

p25 table 1 "based growth assays of deletion strains.": based on

Fixed.