Finding a proper paper with the proper information

When trying to find your next paper to analyze, it’s not always as easy as it seems so here are a few things that I look for in each paper.

1. Check that the year is 2010 or newer, anything before that is usually not useful for this project.
2. Make sure that the species they are working with is either saccharomyces cerevisiae or Escherichia coli.
3. Now check if the proper information is there:
   1. I start by going to the supplementary information section of the paper and look at the data they provided.
   2. I check to see if they have any files with a list of mutations, which genes occurred in those mutations, which Population these genes were in, and if they have frequencies, that’s a bonus. If they do not have frequencies, then it usually means those genes listed were collected at a certain time point, indicating their presence.
   3. If the SI doesn’t seem to have any data that works, then look through the paper, sometimes (not usually) they put useful tables in the paper.
   4. If you’re missing any data, i.e, the mutations don’t mention which genes occurred in them, then it might be in a different file/in the paper somewhere, or we will have to email the author in hopes that they reply and provide the missing information.
4. Now that you have your information, we have to check how accurate it is.
5. I finally read the paper to make sure it’s actually what we want, it does whole genome sequencing, has more than 2 populations, it’s not a dissertation or a thesis paper and its microbial (a.k.a not viral)
6. At this point I look in the paper to see if the following information matches the information provided in the data:
   * 1. Number of mutations
     2. Number of generations (if any)
     3. Number of populations
     4. Number of clones (if any)
     5. Any other information that is listed in the paper
7. If the data matches everything the paper says, great, conduct the analysis. If not, try to figure out why it doesn’t add up.
8. Sometimes you simply won’t find an answer and it might be due to a mistake made by the author. Other times it’s a lack of information by the author. As in they might not count the synonymous mutations in their mutation count but won’t say that in their paper.
9. Overall do as much sleuthing as you can, in the end of the day if you can’t figure out why it doesn’t match, you can either get another set of eyes to see if they can figure it out or email the author with your very specific questions.