The manuscript "Revisiting pangenome openness with k-mers" describes a method to estimate how "open" a pan-genome is, that is, to estimate whether the number of novel sequences expected as more genomes are sequence will keep growing to infinity or is bounded. The method proposed uses k-mers rather than say genes or open reading frames, and is shown to have good correlation with previous methods.

Although the computational method to efficiently compute the openness (the parameter α), the mathematical background is still flawed, and the extra text added to the manuscript since the first revision does not properly address that issue.

Major issues

1. Adding that "the traditional concept of open and closed pa-ngenome may be mathematically flawed" (line 171, page 5), does not properly address the issue. It is well understood that models are only approximation of the phenomena they represent, nevertheless these models need to be at least intrinsically coherent to draw any conclusion. Presenting a mathematically flawed model with an admission that the model is flawed is not a way to resolve the issue or design software methods.

I would agree that the description in Tettelin et al., on which this manuscript is based, is also (very) confusing. Regardless, the method in this manuscript should be correct. As it is described currently, the condition $\alpha > 1$ implies that f_{tot} is decreasing with a limit of 0. This cannot model the size of a union of sets of elements as the union is necessarily increasing in size. As such, the distinction between open and closed genomes is vacuous as no pan-genome can satisfy the close definition.

Maybe surprisingly, it is only the presentation of the model that needs fixing, the method itself seems correct.

As I understand it, the definitions are as follow. It is the growth of the number of elements (be it k-mers, genes, etc.) that follows a power law. The number of elements is a power law plus a possible constant.

(In the following, all the additive constants are named C, even though they might not be all equal. There actual values are not important for the exposition). That is:

- For open genomes, $f_o(m) = C + K_1 m^\gamma$, with $\gamma > 0$. I.e., a constant plus an increasing power law. The derivative is $f'_o(m) = \gamma K_1 m^{\gamma-1}$, and it is positive for all m. $f_o(m)$ grows to infinity as m grows.

- For close genomes, $f_c(m) = C - K_1 m^\gamma$, with $\gamma < 0$. I.e., a constant minus a decreasing power law. The derivative is $f'_c(m) = -\gamma K_1 m^{\gamma-1}$, and it is also positive for all m. $f_c(m)$ grows to C as m grows.

Both derivative have the form $K_2 m^{-\alpha}$ with $K_2 > 0$ and $\alpha = 1 - \gamma$. And the value of the exponent α (i.e., < 1 or > 1) determines the openness of the pan-genome.

Conversely, define: $f_{tot}(m) = \int_{m_0}^{m} K_2 x^{-\alpha} dx = C + K_2 m^{1-\alpha}/(1-\alpha)$. Then, if $\alpha < 1$, f_{tot} has the same form as f_o (i.e., a constant plus an increasing power law), and if $\alpha > 1$, f_{tot} has the same form as f_c (a constant minus a decreasing power law).

Note that the constant C in f_{tot} depends on the starting point m_0, which properly models what is actually done in practice in this method (e.g., line 264, page 10).

2. There are no examples of closed genomes in the evaluation. There should be.