

Guiglielmoni et al. is an informative and thorough review about the approaches developed in the last years to sequence and assemble genomes with a specific focus on invertebrate genomes. I think both the structure and content (including the numerous suggestions for tools) of this review to be of great relevance and interest for the genomic community that deals with non-model organisms.

My comments are very minor, they are actually only suggestions to include some particular points/references or to rephrase small parts of the text.

Since there are no line numbers I tried to give clear indications about the text location (as clear as possible).

### Minor comments

**Page 2, end of 3<sup>rd</sup> paragraph:** The sentence "Many phyla with less direct human implications, however, do not even have a single good-quality genome assembly available to date (e.g., chaetognaths)." may be further supported by Hotaling et al. 2021 that explores (among other things) which phyla are lacking any type of genome assembly <https://doi.org/10.1073/pnas.2109019118>

**Figure 1:** consider to replace "," in the legend with "+" so to make clearer that "Short reads, long reads" means that a combination of those technologies was used to build the assembly.

**Page 5, end of page:** it would be important to highlight that the greater accuracy given by HiFi is obtained at the expense of the length of the reads themselves that must be shorter than the ones used for "regular" PacBio.

**Page 6, end of 2<sup>nd</sup> paragraph:** "In addition, secondary metabolites associated to DNA molecules, or branched DNA structures, can also disturb the sequencing reaction." This is an interesting point that I heard discussed many times and it would be nice to gather some references to support it if possible.

**Page 14, 3<sup>rd</sup> paragraph:** "To improve the contiguity of an assembly, contigs can be grouped, ordered and oriented into scaffolds." I think that the concept of contiguity should be reserved to unfragmented sequences (contigs) and that the process of scaffolding does not really improve the contiguity aspect since scaffolds are, by definition, clusters of sequences bridged by gaps (therefore fragmented). I agree that an assembly with thousands of little separated contigs is much worse than an assembly where these contigs are grouped into a bunch of scaffolds but still the contiguity would remain the same, what changes is the representation of the genome/chromosome structure. I would advise to rephrase the topic sentence of

this paragraph to have a less ambiguous meaning of “contiguity”. Also, I think that the authors used the equivalence “contiguity = measure of quality/fragmentation = contig N50” in the first part of the review (e.g., Figure 1) so I suggest to keep this one meaning throughout the paper.

**Page 15, middle paragraph:** regarding linked reads and the discontinuation of the 10X Genomics service, it can be added that there are at least two other replacing technologies: 1) TELL-seq <https://genome.cshlp.org/content/30/6/898.short> ; 2) haplotagging <https://www.pnas.org/content/118/25/e2015005118>  
The second was used already on invertebrates (butterflies) and TELL-seq seems to work with ultra-low DNA input.

**Page 16, end of first paragraph:** the last sentence of the paragraph could be slightly rephrased in a way that becomes 100% clear that using Omni-C can yield de novo genome assemblies. What I mean is something like this: “[...] such as Omni-C, therefore adequate for de novo genome assemblies.” This is just an example, no need to rephrase it exactly like this!

**Page 17, beginning of the 2<sup>nd</sup> paragraph:** specify which N50, I guess “Contig N50” (?)