Reviewer’s report

Title: Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences

Version: 0 Date: 26 Dec 2015

Reviewer: Thomas Keane

Reviewer's report:

Tamazian et al. introduce a tool for arranging assembled contigs or scaffolds into chromosomes using a closely related reference genome. There have been various tools published over the years for doing this, many referenced by the authors. As the authors say in the abstract, the process of generating ordered/orientated chromosomes is a 'crucial step of any genome study'. The consequences of introducing errors in this step of a genome project potentially affect many of the downstream analysis/interpretation steps.

The text of the paper is overall fine and well written, in terms of language and presentation.

I have one major concern though - I think the paper requires a much more substantial evaluation of the accuracy of the software (ideally to include more than just one other tool - several are cited). The only numerical measures of accuracy are identity, length, mismatches, and coverage. To me, the goal of the assessment should be to measure the structural accuracy of the chromosomes produced - from the local structure (few kbp range), longer range accuracy (e.g. tens to hundreds of kbp), and the overall chromosome organisation accuracy. Then the question of where are the mistakes being made, e.g. are the coding regions of the genome correct - for species with splicing, are the exons in the correct order+orientation? Are structural variations differences between the species being assembled and the reference genome represented correctly in the chromosomes? Especially transposable element differences, which can be difficult since they contain repetitive sequence.

I can think of a few ways to create this sort of evaluation 1) take a species with PE reads, and mate pair libraries of various sizes, create scaffolds with all but the largest mate pair library, run
chromosomer and other similar tools, and then use the largest library to see how many of those mate-pairs align in the correct orientation and in the expected size range. 2) Create simulated genomes with structural variations at different size ranges, create PE and mate pair libraries, run chromosomes (and other tools), and then look at the representation of the introduced simulated structural variants. 3) In the real data sets (could use the existing presented datasets), it would be very very useful to know the accuracy of the coding regions, e.g. exon order+orientation, as a way to compare the performance of different tools.

Minor things:

- the quality of the figures with the dotplots is not good. I can't read the legends and most of the dot plot is quite blurry.

Level of interest

Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English

Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?
6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

None

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal