Homework for June 26, 2014

Zhiyang Ong*

Texas A&M University

June 26, 2014

Abstract

This is the homework for June 26, 2014. It contains a description of the comparison of different SAM files. It also contains hypotheses for why the given SRA82660 sequences do not map to chromosome 7 (linkage group VII) of the *Neurospora crassa* species.

1 Acknowledgements*

I acknowledge that I have collaborated with my project teammates Diana Medina and Selene Howe to complete this homework. Specifically, we discussed some hypotheses.

I also acknowledge the help that I have received from the course instructor, Prof. Rodolfo Aramayo, in getting back to my queries about the source of the reference sequence for the *Neurospora crassa* species that was provided to us on Tuesday.

Lastly, I want to thank Ms. Tess L. Pham from the Tutor Zone at Evans Library, which is run by the Academic Success Center at Texas A&M University. Ms. Pham has tutored me about the basics of genetics and gave me feedback about the hypotheses which I proposed. I also want to thank Ms. Priyadharshini Venkat, my classmate for this course, for giving me feedback about my hypotheses and helping me to find alternate golden models for the *Neurospora crassa* species.

2 Comparison of different SAM files

Different types of SAM files have to be compared. Their differences are only based on the different combinations of `-v` and `-m` selected, or different combinations of `-n` and `-m` selected for *-Bowtie* [2].

The different combinations of these `-v`, `-n`, and `-m` options of *-Bowtie* are:
1. -v1 -m1
2. -v1 -m100
3. -v3 -m1
4. -v3 -m100
5. -n1 -m1
6. -n1 -m100

---

*Email correspondence to: ✉ ongz@acm.org

7. -n3 -m1

8. -n3 -m100

The `-v` option is used to determine the number of alignments that have no more than "v" mismatches. This addresses global alignments, where a sequence is matched with another. The `-m` option reports alignments with a maximum of "m" repeats. The `-n` option reports alignments with a maximum of "m" repeats. This addresses local alignments, where a region in a sequence is matched with another region in the sequence [1].

3 Hypothesis: SRA82660 sequence is not part of the *Neurospora crassa* genome

A hypothesis is that the SRA82660 sequence is not part of the *Neurospora crassa* genome. That is, it belongs to another species. Proceedure 3

TEST IF SRA82660 SEQUENCE IS PART OF THE *Neurospora crassa* GENOME(*Golden/ReferenceModel*)

    // *Golden/Reference Model*

1 **if** [given SRA82660 subsequence does not map to golden model]

2   Conclusion: The given SRA82660 subsequence is not part of the *Neurospora crassa* species.

3 **elseif** [given SRA82660 subsequence maps to golden model]

4   Conclusion: The given SRA82660 subsequence is part of the genome.

5

6 **return**

References

[1] Rodolfo Aramayo. Small reads genome mapping 101. Available online from the course Wiki of BIOL 689 Digital Biology at: https://geiger.tamu.edu/gitlab/raramayo/digitalbiology/wikis/05Lecture/SRGenome_Mapping_101.md; June 24, 2014 was the last accessed date, June 24 2014.

[2] Ben Langmead and Cole Trapnell. Bowtie: An ultrafast memory-efficient short read aligner. Available online from SourceForge at: http://bowtie-bio.sourceforge.net/index.shtml; April 2, 2014 was the last accessed date, March 14 2014.