Project

September 13, 2016

Name: Martin Basterrechea

Field: Bioinformatics

Credits: 7.5 cr.

Duration: 2016-09-12–10-23, second part on half time

Draft assembly, gene prediction and annotation of a species from genus Leucocytozoon

Background

The genome of a *Leucocytozoon* species has been sequenced using the Illumina sequencer. This species is one of many avian malaria parasites [1]. The aim of this project is to construct a draft assembly, make a gene prediction and annotate the genes.

For the first part of the project an assembler, like Spades, is used. An extensive list of assemblers is found at
http://omictools.com/genome-assembly-category
The assemblies can be evaluated with GAGE [2]
(http://gage.cbcb.umd.edu/)
After deciding on which assembly to use, gene prediction with different sofware, like GeneMark, Augustus and FGENESH, should be performed. An extensive list is found at
http://omictools.com/protein-coding-gene-prediction-category)
How gene predictions are evaluated is unclear but probably the predictor with a combination of few and long transcripts should be chosen. The genes from "the best" gene prediction should be annotated with protein description, go terms, pfam domains and subcellular location. If there is time a genome browser should be created and some comparative genomics conducted.

References

[1] S. L. Perkins. Malaria's many mates: past, present, and future of the systematics of the order Haemosporida. *J. Parasitol.*, 100(1):11–25, Feb 2014.

[2] S. L. Salzberg, A. M. Phillippy, A. Zimin, D. Puiu, T. Magoc, S. Koren, T. J. Treangen, M. C. Schatz, A. L. Delcher, M. Roberts, G. Marcais, M. Pop, and J. A. Yorke. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res.*, 22(3):557–567, Mar 2012.