Beoung Hun Lee

CS 199

March 23, 2019

Edwin Solares

To compare and analyze AUGUSTUS and GeneMark, web version of AUGUSTUS and Genemark.hmm Eukaryotic have been used. Hedgehog(hh)(1), indy(2) and tinman(3) genes were used as the samples. Even though it was recommended to use Genemark.hmm prokaryotic with bash terminal, I used Genemark.hmm Eukaryotic from the web instead since these genes were from Drosophila melanogaster. For tinman gene, AUGUSTUS predicted that there is one gene with two exons and two introns and Genemark predicted that there are two different genes with two exons on the first gene and three exons on the second gene. The actual gene is closer to that of AUGUSTUS, as the original gene structure is one gene with three exons. For hh gene, AUGUSTUS predicted that there is one gene with three exons and two introns and Genemark predicted that there are two genes with four exons on the first one and two exons on the second one. Again, the actual gene structure is closer to that of AUGUSTUS, since it was one gene with three exons. Lastly, for indy gene, AUGUSTUS predicted two genes with three coding sequences on the first gene and eight coding sequences on the second gene and Genemark predicted one gene with ten exons. For indy gene, Genemark predicted better as the actual gene structure was one gene with eleven exons.

AUGUSTUS and Genemark differ in two main methods. First, AUGUSTUS makes usage of external hints, while Genemark only uses intrinsic method to predict gene structure. Genemark only uses hidden Markov Model and query sequence, while AUGUSTUS uses hints and constraints provided users, along with existing information about gene structures. Second, AUGUSTUS uses generalized hidden Markov Model while Genemark uses a regular hidden Markov Model. Algorithms used in the models differ as well, as Genemark uses forward algorithm and delete interpolation, while AUGUSTUS uses forward algorithm and sampling algorithm for extrinsic evidence. Overall, AUGUSTUS had a better prediction for hh and tinman gene, while Genemark performed better on indy gene. AUGUSTUS performed better for hh and tinman gene because both genes were included in 100 gene sequences that were included in web version of AUGUSTUS. These genes had hints to be compared with, while indy gene was not included in the hints, so prediction was not as accurate as other genes. Also, there were slight difference in gene prediction even with hints. This seems to be due to the fact that nucleotide sequences of genes get updated throughout the years. To conclude, when given hints, AUGUSTUS is more accurate on predicting gene structure, but Genemark is more accurate when hint is not available.

References.

1. URL. <https://www.ncbi.nlm.nih.gov/gene/42737>
2. URL. <https://www.ncbi.nlm.nih.gov/gene/40049>
3. URL. <https://www.ncbi.nlm.nih.gov/gene/42536>
4. URL. http://bioinf.uni-greifswald.de/augustus/submission.php
5. URL. http://exon.gatech.edu/GeneMark/gmhmme.cgi