Musa bioinformatics course (2013/11/20)

Practical session 3

Gene annotation workflow

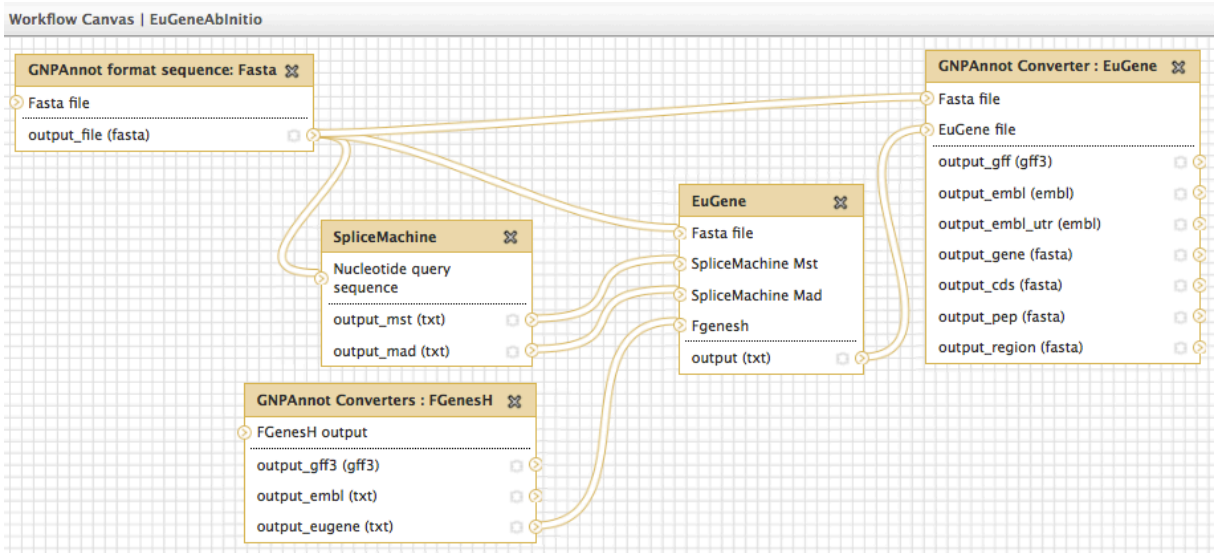
Goal

Eugene combiner allows to predict protein coding genes on a genomic eukaryotic sequence. We are going to use it on a BAC clone genomic sequence of the commercial triploid *Musa acuminata* subgroup Cavendish cv. Grande Naine that should contain genes coding for pectin methylesterase (PME) (Mbeguie, et al., 2009).

*NB: The workflow **EugeneAbInitio** allows an automatic prediction of monocotyledon genes encoding polypeptides along a given genomic region. For this purpose, it combines the results of three ab initio or intrinsic methods requiring a learning phase.*
***EugeneIMM** http://eugene.toulouse.inra.fr/ is a statistical search method for genes encoding eukaryotic polypeptides, by content, i.e. based on Interpolated Markov Models (IMM) to discriminate the coding regions from the non-coding regions of a DNA sequence.*
***Splicemachine** http://bioinformatics.psb.ugent.be/webtools/splicemachine/ is a method of gene prediction by signal, i.e. identification of initiation and termination translation codons and splicing sites of introns. It employs Linear Support Vector Machines (LSVM) for classifying the real from the pseudo- splicing sites.*
***FGENESH** http://www.softberry.com/berry.phtml is a gene finder by content, based on Hidden Markov Models (HMM) with a supervised learning phase.*

Task

In order to annotate the **MaC088K20.fna** BAC clone sequence, we are going to run the **EugeneAbInitio** combiner included in the Galaxy workflow manager of the South Green bioinformatics platform http://gohelle.cirad.fr/galaxy/



Mbeguie, A.M.D., *et al.* (2009) Expression patterns of cell wall-modifying genes from banana during fruit ripening and in relationship with finger drop, *J Exp Bot*, **60**, 2021-2034.