

POY Documentation

Louise Crowley

November 4, 2011

1 Major additions to POY 5.0

1.1 Likelihood (Nick, Ward, John)

I don't know jack about this.

1.2 Chromosomal Characters

Questions: Vinh's Code (Lin/Ward)

1. Are we going to make any reference to Vinh's original code?
2. Are any of his options still in use?
3. Will this be referenced in the documentation?

Questions: MAUVE (Lin/Ward)

1. Which part of MAUVE are we using *exactly*? It will be necessary to compare Mauve Original and Progressive Mauve to see which parts of which we are using (Lin). I know we are mostly using the Original algorithm, but I think there are some parts, which are in the Progressive algorithm, that we are using.
2. It would be useful (for me anyway) to have a flow chart of how we handle these data types, from the initial reading in of a fasta file to the final output. At what stage in the analysis is Mauve involved? Use of Fixed States? How the Newkonn algorithm is involved etc:
3. Did 'we' ever consider using multi-LAGAN, which is better at handling more divergent genomes and better suited for interspecies comparison?

4. In the initial stage of analysis where Mauve is aligning the sequences, what about the unaligned regions that MAUVE doesn't align?
5. LCBs, and seeds will all need to be explained. Given that LCBs are a part of multi-MUMs, will this have to be explained, or is this not an option that the user defines? Is a string the same as a seed? Do we have spaced seeds? They are palindromic, correct?

Questions: Diagnosis file (Lin/Ward)

1. How is the diagnose file outputted – what are all the options?
2. We can output a synteny block image, correct? Is Mauve called for this part?

2 Documentation

1. Change 4.0 to 5.0
2. Typos (Louise)
3. All of the documentation needs to be checked for "currently not supported", to see whether this is still true.
4. The code from 4.0 and 5.0 needs to be compared to see what's new (especially if Andres put in more things after the last release and these have not been documented) (Nick).
5. When to invoke the long sequences option when compiling - for 32 bit and 64 bit computer architecture.

3 Tutorials

1. Add additional tutorials, e.g Timed Search, Bootstrap, an example of drift (Louise).
2. Redo all Chromosomal Tutorials (4.6-4.9). The current tutorials state that they are only for demonstration purposes – these are inadequate for the new code (Louise, Megan, Lin).
3. Examples for every new option should be included – these will go in the *Transform* section (Louise, Megan, Lin).

4. Dummy files for the chromosomal data types, as well as cost matrices, will have to be included (Megan, Louise).
5. Add a tutorial about pounding or chopping data to reduce the amount of missing data – this is a constant struggle for many people starting out with POY

4 Additional Questions/Points

1. Megan and I feel that it is a misnomer to refer to *Locus Costs*, given that it is not referring to a functional gene. Perhaps in the documentation we should refer to the synteny block content, hereafter locus cost”. This can be included as an operational definition of a locus in a note.
2. Redo all Chromosomal tutorials (4.6-4.9). The current tutorials state that they are only for demonstration purposes – these are inadequate for the new code (Louise, Megan, Lin).
3. Examples for every new option should be included – these will go in the *Transform* section (Louise, Megan, Lin).
4. Are we going to have a POY4.0 command index just like we had a POY3.0 command index?