Prepared by Khoa:

Date: 2013-12-06

After our discussion today, we came up with the preliminary plan for the project and the flow chart for the implementation of the noise reduction algorithm:

The folder MAP/data/appbio11/asymmetric\_0.5 has 300 file containing noisy alignments and one file containing a tree (reference tree). So are the rest of the folders in MAP/data/appbio11/

**Preliminary plan:**

1. Implementation of noise reduction algorithm according to the project’s description at MAP/doc/ project\_description\_appbio13.htm (our algorithm)
2. Running our algorithm on all noisy alignments to get un-noisy alignment set A
3. Running Gblocks algorithm on all noisy alignments to get un-noisy alignment set B
4. Running TrimAI algorithm on all noisy alignments to get un-noisy alignment set C
5. Running fastprot+fnj on set A to get the neighbor tree TA
6. Running fastprot+fnj on set B to get the neighbor tree TB
7. Running fastprot+fnj on set C to get the neighbor tree TC
8. Compare tree TA, TB, TC with the reference tree in each folder

Noisy alignments

Un-noisy alignment

set A

Un-noisy alignment

set B

Un-noisy alignment

set C

Tree TA

Tree TB

Tree TC

Our Algorithm

GBlocks

TrimAI

fastprot/fnj

fastprot/fnj

fastprot/fnj

Reference trees

compare

compare

compare

**Flow chart for the implementation of the noise reduction algorithm**

Read

Noisy alignment (fasta format)

/data/appbio11/asymmetric\_0.5/S001.align.1.msl

First Column

>50% indels?

At least 50% amino acids are unique?

No amino acid appears more than twice?

Remove column i because it’s noisy

Yes

Yes

Yes

No

No

No

Next column