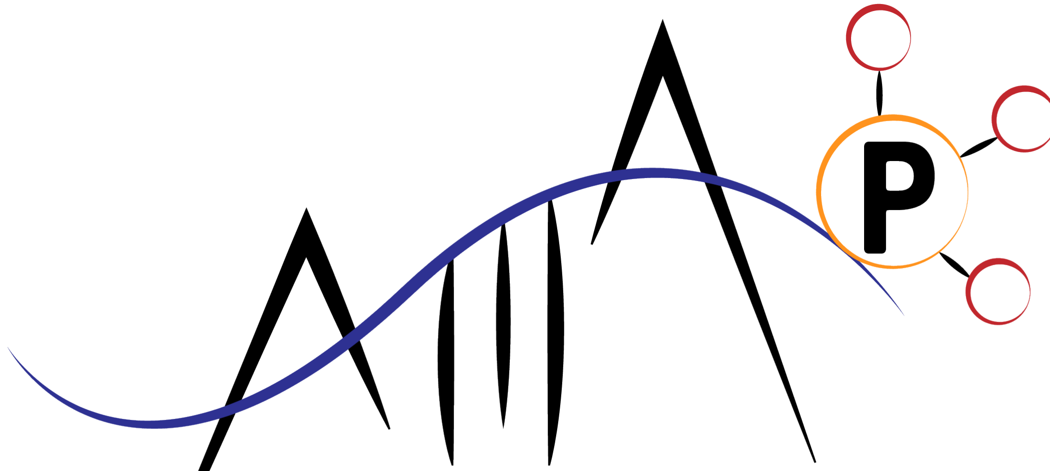

Developed by

ADITYA SINGH

Ph.D. Scholar

Haematology/Oncology Unit

Dept. of Paediatrics, Advanced Paediatric Centre

Postgraduate Institute of Medical Education and Research

Chandigarh, India

aditya.onco@gmail.com

adityas@live.in

Welcome to AMAP Guide!

1. Introduction

AMAP is a code written in Python 3.x language. It incorporates various other programs to achieve its task. AMAP accepts AB1 chromatogram files in pairs (Forward & Reverse), reference sequence as template in FASTA format and generates final alignment. The program converts AB1 files into FASTQ, reverse complements reverse sequence, trims low quality ends of the sequences, aligns and creates a consensus sequence of these two forward and reverse sequence, aligns with the template sequence and displays the alignment. Also, the program saves all these files for future reference and study. Please note that the program will only display the alignment if not multiplexed, i.e. if only one pair of sequence is given to the program. In case of multiplexed runs, the program only saves the files which can be viewed easily in any alignment viewer program. Optionally, if provided with a FASTA file of exonic region and a reference amino acid file, AMAP also extracts the exonic region from the input sequence, translates it into +1, +2 & +3 frames and aligns it with the test reference amino acid sequence.

2. Installation

Open the code in text viewer (nano/ emacs etc.) on your system and edit the lines 15 & 16 as below:

seqtk = "Path to SEQTK executable"
seaview = "Path to SeqView executable"

AMAP can be executed directly without any installation. However, for easy usage, make the code executable using "chmod +x AMAP" and copy it to your system's PATH for invoking it from any folder.

3. Usage

Automated Mutation Analysis Pipeline

For the impatient:

Typical usage

```
> AMAP F/R/FR/FE/RE/FRE Forward_seq.ab1 Reverse_seq.ab1 Exons.fasta
AminoAcidRef.fasta Reference_seq.fasta
```

F: All sequences are forward.
R: All sequences are reverse.
FR: Sequences are in pairs, Forward and Reverse per sample.
Adding 'E' to the argument will enable exon extraction based on supplied exon file, its translation and alignment with reference amino acid fasta file.

For multiplex usage for *n* number of samples:

Example:
n = 2; That is, Sample A and B.

Typical Usage for such a scenario:

```
> AMAP F/R/FR/FE/RE/FRE Forward_seq_Sample_A.ab1
Forward_seq_Sample_B.ab1 Reverse_seq_Sample_A.ab1
Reverse_seq_Sample_B.ab1 Reference_Seq.fasta
```

Kindly note that SeaView will not be launched automatically in case of multiplexed runs. You can manually open the *._result.aln file to visualize the alignment.

4. Requirements:

Linux/Unix System with EMBOSS, SEQTK, Python 3.x, BioPython & CLUSTALW2 installed.
Made and tested on Mac OS X 10.11.4 with Pyhton 3.4, BioPython 1.66, EMBOSS 6.5.7, SEQTK, CLUSTALW2 2.1
You can download these freely available softwares from the below given links:
EMBOSS : http://emboss.sourceforge.net/download/
SEQTK : https://github.com/lh3/seqtk
BioPython : http://biopython.org/wiki/Main_Page
SeaView : http://doua.prabi.fr/software/seaview
NCBI BLAST+: *ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/*

5. Description of results:

*._Forward_Sequence_QC.fastq : QC Trimmed Forward Sequence in FASTQ format.
*._Reverse_Sequence_QC.fastq : QC Trimmed Reverse complemented reverse sequence in FASTQ format.
*._consensus.fasta : Consensus sequence in FASTA format.
*._result.aln : Alignemnet of consensus with reference sequecne in ALN format.
*._result.dnd : Guide tree file for alignment in DND format.
*._result.fasta_report.txt : ClustalW2 alignment report.
*.consensus_BLAST.txt: Simplified BLAST result between Exonic sequence and extracted exon.
*.EXON_*: Files related with Exon extraction.

Automated Mutation Analysis Pipeline

6. Troubleshooting

A. If any of the required softwares aren't installed/ not accessible to AMAP, it will specifically point them out. Install them and make sure they are in your system's PATH. Also, make sure you followed the installation instructions carefully.
B. Very poor alignment:
    a. Make sure you gave the sequences to AMAP in proper order. All forward sequences for all samples first, followed by all reverse sequences of all samples second and finally the reference sequence.

    b. Make sure you/ sequencing provider did not reverse complement the reverse sequences.
    c. Although, quality trimming is performed but still if the alignment is not at all right and all above mentioned troubleshooting has been followed, manually see the chromatogram files to check for background noise.

C. I installed all the requirements but still AMAP is unable to use them!

    a. Make sure you have installed the requirements like BioPython with Python 3. If you have installed it specifically with a sub version of Python 3 like Python 3.5 and the program is not working, then kindly edit the first line of the code using the same method described in installation instruction as:
    `“#!/usr/bin/env python3”` to `“#!/usr/bin/env` python3.5” or any other version that you installed your BioPython and other dependencies with.
    Also, make sure you do not use Python 2.x. The program will only work with Python 3.x.

7. Citation

If you use the AMAP in any of your work, then kindly cite the following:

Singh, A., Bhatia, P. 2016. Automated Mutation Analysis Pipeline (AMAP): a tool for rapidly analyzing Sanger sequencing data with minimum user interference. (Under submission process)

8. Contact

For any suggestions/ bugs/ complaints, please feel free to contact me at:

ADITYA SINGH
Ph.D. Scholar
Haematology/Oncology Unit
Dept. of Paediatrics, Advanced Paediatric Centre
Postgraduate Institute of Medical Education and Research
Chandigarh, India
aditya.onco@gmail.com
adityas@live.in

Automated Mutation Analysis Pipeline

Thank you!

Hope AMAP helps you in achieving things faster and better.