**TUTORIAL FOR POPSICLE ANALYSIS**

Written by Dr. Fumiaki Ihara

RIMD, Osaka Univ.

**INTRODUCTION**

POPSICLE is a software suite developed by Jahangheer S. Shaik, Asis Khan, and Michael E. Grigg at NIAID, designed to determine population structure and Ancestral Determinants of Phenotypes using Whole Genome Sequencing data. This tutorial explains how to run POPSICLE using major Type I, II, and III Toxoplasma genomes. For more information about POPSICLE, please visit <https://popsicle-admixture.sourceforge.io/AnalyticalPipelinePopulStr.html>.

**Before beginning this TUTORIAL, you must install Java and Circos on your computer. It is assumed that the tutorial will be run on Mac OS Sonoma.**

**All parameters in the scripts refer to the values used by the author in this paper.**

**CITATION:** Jahangheer S. Shaik, Asis Khan and Michael E. Grigg, POPSICLE: A Software Suite to Study Population Structure and Ancestral Determinants of Phenotypes using Whole Genome Sequencing Data, 2018, bioRxiv.

**FLOWCHART**



**TEST DATA**

9 bam files generated by BWA-MEM v0.7.17 mapping against ToxoDB-57 ME49.

|  |
| --- |
| # example  bwa mem -t {number of threads} {Path to the Reference} {filename}\_1.fastq.gz {filename}\_2.fastq.gz > {filename}.sam  samtools sort -@ 24 ${filename}.sam > ${filename}.bam |

Original fastq files are available from the following accession numbers.

RH-88\_SRR521957: GT1\_ DRR513067: TgDgCo17\_ SRR350734: ME49\_ DRR513065: PRU\_ SRR350739: B73\_SRR521556: CTG\_DRR513066: VEG\_ SRR516406: M7741\_SRR521653

You can download test data from <https://drive.google.com/file/d/1ZY_IKXrrAdIYXweY4d150Y8UqIeo5pse/view?usp=sharing>

**RUN**

1. Download LPDtools.jar, Script1.sh, Script2.sh, Script3.sh, and conf.

Download reference genome from ToxoDB.org, for example ToxoDB-57\_TgondiiME49\_Genome.fasta

1. Set “Workdir” where you put the files (**Line 4**).
2. Set block size, e.g., 10000 (**Line 21**).

|  |
| --- |
| #Set derectory  Workdir= ‘Path to Your directory with POPSICLE TUTORIAL’  **cd** ${Workdir}  # run name  RUN="POPSICLE\_TUTORIAL"  IDS="${RUN}.txt"  # directories  BIN="${Workdir}"  WD="${Workdir}"  RUND="${WD}/${RUN}\_analysis"  REF="${WD}/ref"  BAM="${WD}/bam"  mv ToxoDB-57\_TgondiiME49\_Genome.fasta ${REF} |

Place bam under /bam

|  |
| --- |
| bam folder contains HG1(RH-88, GT1, and TgDgCo17), HG2(ME49, PRU, and B73), and HG3(CTG, VEG, and M7741) |

1. Run Script1 on Terminal. This step takes a long time.

|  |
| --- |
| $ bash Script.1.sh |

1. See popsilce\_clusters to determine number of clusters. Typically, the Dunn index is the highest can be chosen as the number of clusters.

|  |
| --- |
| minClusters 2 maxClusters 4  0 1 1 1 1 0 0 1 1  0 1 2 1 2 0 0 1 2  0 1 2 1 2 3 0 1 2  Metadata  The Dunn index for 2 clusters is:1.539877239540446  The Dunn index for 3 clusters is:2.7005129959828236  The Dunn index for 4 clusters is:1.2961881828779294  Based on these indices, we recommend the number of clusters to be :3. See if you agree.  Clusters  3 |

1. Set the last line of the popsicle\_clusters to the number of clusters you want.

|  |
| --- |
| Clusters  3 → Change this number |

1. Run Script2

|  |
| --- |
| $ bash Script.2.sh |

1. Move 4 files in conf to circus\_10000.
2. Run Script3 and you will get POPSICLE plot as svg and png files.

|  |
| --- |
| $ bash Script.3.sh |

グラフ, サンバースト図

自動的に生成された説明**Output**