Manual of YTool

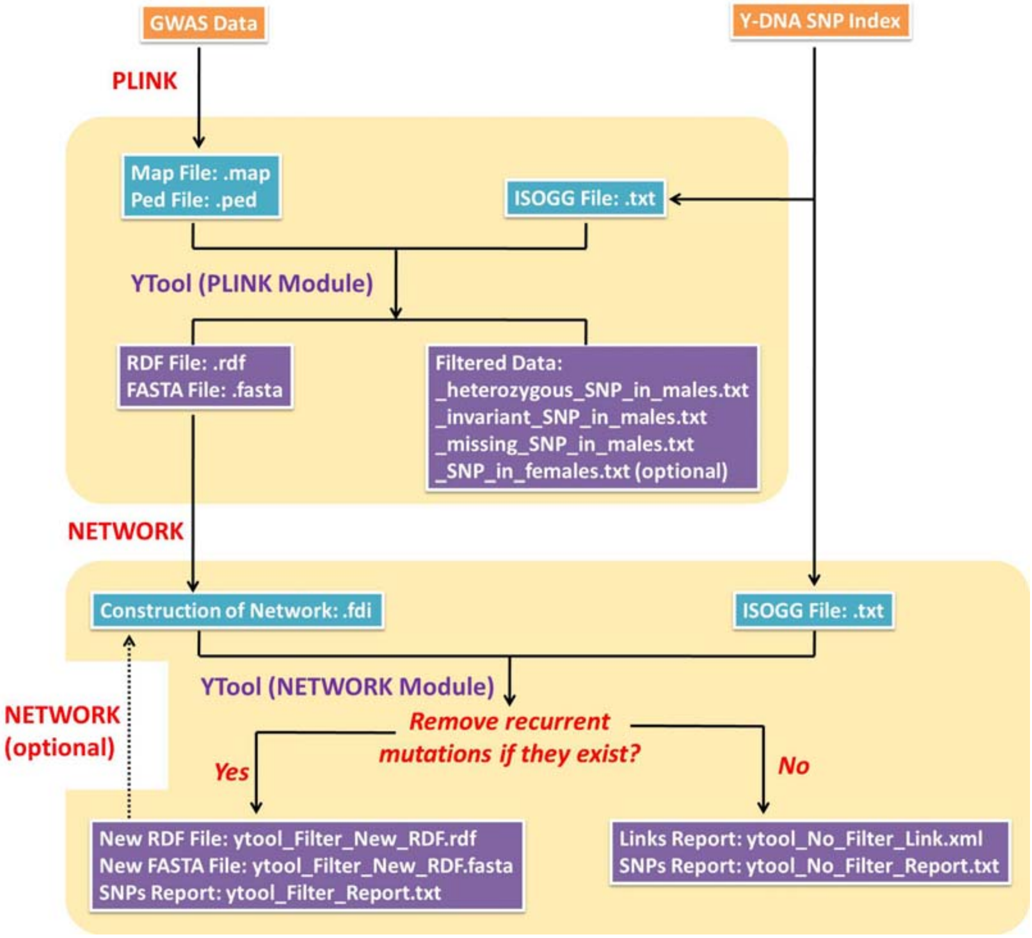
Version 1.0

Download: http://mitotool.org/ytool/

1. What is YTool?

YTool is a free stand-alone software designed for extracting Y-SNPs information from genome-wide SNPs data in order to perform phylogenetic analyses. It is programmed by using C++, and is available for WinXP/Vista/Win7/Win8.

2. Overview


GWAS Data
Y-DNA SNP Index
PLINK
Map File: .map
Ped File: .ped
ISOGG File: .txt
YTool (PLINK Module)
RDF File: .rdf
FASTA File: .fasta
Filtered Data:
_heterozygous_SNP_in_males.txt
_invariant_SNP_in_males.txt
_missing_SNP_in_males.txt
_SNP_in_females.txt (optional)
NETWORK
Construction of Network: .fdi
ISOGG File: .txt
YTool (NETWORK Module)
NETWORK
(optional)
Remove recurrent
mutations if they exist?
Yes
No
New RDF File: ytool_Filter_New_RDF.rdf
New FASTA File: ytool_Filter_New_RDF.fasta
SNPs Report: ytool_Filter_Report.txt
Links Report: ytool_No_Filter_Link.xml
SNPs Report: ytool_No_Filter_Report.txt

For detailed manipulation of YTool, please watch our demo video:

http://mitotool.org/ytool/YTool_Demo.mp4.

Note: we take the example of 117 male Burmese here (i.e. folder of 117MalesBurmese)!

3. Input files

(1) MAP and PED files

Now the GWAS data with PLINK format (including MAP and PED files) gets supported both by Affymetrix and Illumina platforms. For PLINK format, extracting Y-SNPs as well as other basic uses, please refer to PLINK website (http://pngu.mgh.harvard.edu/~purcell/plink/data.shtml) and Anderson, et al. (2010) for details. After extraction, we recommend users to check the proportions of missing genotypes for each of males by using "--missing" in PLINK. The individuals with high proportions of missing genotypes should be removed.

Example: burmaMalesYFWD.map & burmaMalesYFWD.ped

(2) ISOGG file used for annotation

Data source of ISOGG file: http://www.isogg.org/tree/ISOGG_YDNA_SNP_Index.html

Here we provide the version 8.20 (Date: 26 February 2013) of ISOGG files. Please note the Y chromosome positions in ISOGG and MAP files must refer to the same Build (37 or 36) of human reference genome!

ISOGG file format: three columns need to be separated by a semicolon, so each line of the file looks like: <SNP>;<Y-position>;<Haplogroup>.

Example: ISOGG_GRCh37_v8.20.txt

(3) NGS data file used for annotation (OPTIONAL)

We provide the NGS data retrieved from Wei, et al. (2013).

Please note the Y chromosome positions in NGS data and MAP files must refer to the same Build (37 or 36) of human reference genome!

Annotation file format: the first line contains column titles and each column needs to be separated by a semicolon, meanwhile the first column is the position of Y-SNP.

Example: NGS_Wei_GRCh37.txt

4. PLINK module

(1) Function of PLINK module

This module aims to filter unqualified Y-SNPs. It will delete:

(i) Heterozygous Y-SNPs in males;
(ii) Y-SNPs with missing rate in males higher than a preset threshold (Default: 0.05);
(iii) Invariant Y-SNPs in all of the males;
(iv) Y-SNPs detected in females (OPTIONAL)

When input data contain female samples, Y-SNPs genotyped in females can be treated as false positive signals. Y-SNPs with calling rate in females higher than a preset threshold can be deleted by using optional function (iv) mentioned above.

(2) Output files of PLINK module

(i) RDF file is for inputting of software Network (***Example: untitled.rdf***);

(ii) FASTA file can be used for many softwares, such as MEGA, and Y-SNPs are aligned according to the Y-positions (***Example: untitled.fasta***);

(iii) Deletion report for Y-SNPs genotyped with heterozygous alleles in males and the annotation information (***Example: untitled_heterozygous_SNP_in_males.txt***);

(iv) Deletion report for Y-SNPs with missing rates above threshold and the annotation information (***Example: untitled_missing_SNP_in_males.txt***);

(v) Deletion report for invariants in males and the annotation information (***Example: untitled_invariant_SNP_in_males.txt***);

(vi) Deletion report for Y-SNPs genotyped in females and the annotation information (**OPTIONAL**).

5. Network construction

The median-joining network based on RDF file (***Example: untitled.rdf***) can be constructed by using NETWORK software (http://www.fluxus-engineering.com/sharenet.htm). The files of network (FDI file; ***Example: untitled.fdi***) and recurrent mutations (STA file; ***Example: Noname.sta***) are generated. For employing other network construction, please refer to the website of NETWORK and its manual for details.

For this example, we performed actions as following:
Calculate Network -> Network Calculations -> Median Joining -> File -> Open (***untitled.rdf***) -> Calculate network -> Save ***untitled.out*** -> Draw Network -> File -> Open (***untitled.out***) -> Finalise -> Statistics, Export ***Noname.sta*** -> Save ***untitled.fdi***

6. NETWORK module

Note: we suggest depositing the results into different path or subfolder!

(1) Function of NETWORK module

This module aims to generate reports based on phylogenetic analyses with NETWORK software.

(2) Input files of NETWORK module

FDI file (***Example: untitled.fdi***).

(3) Output files of NETWORK module

(i) The annotation for each of links in network FDI file (***Example: ytool_No_Filter_Link.xml, subfolder of NoFiltering***);

(ii) The annotation for Y-SNPs in network FDI file (***Example: ytool_No_Filter_Report.txt, subfolder of NoFiltering***).

(3) Remove recurrent mutations (OPTIONAL)

Input files:

(i) RDF file the same as used in network construction (***Example: untitled.rdf***);

(ii) STA file (***Example: Noname.sta***);

Output files:

(i) Deletion report for recurrent mutations and the annotation information (***Example: ytool_Filter_Report.txt, subfolder of Filtering***);

(ii) The new RDF file without recurrent mutations for NETWORK software (***Example: ytool_Filter_New_RDF.rdf, subfolder of Filtering***);

(iii) The new FASTA file without recurrent mutations for other phylogenetic softwares (***Example: ytool_Filter_New_RDF.fasta, subfolder of Filtering***).

Users need to re-run NETWORK software as described previously to get the FDI file (***Example: ytool_Filter_New_RDF.fdi***).

Re-run YTool for the new FDI file to get:

(i) The annotation for each of links in network FDI file (***Example: ytool_No_Filter_Link.xml, subfolder of Filtering***);

(ii) The annotation for Y-SNPs in network FDI file (***Example: ytool_No_Filter_Report.txt, subfolder of Filtering***).

7. Introduction to parameters

(1) Missing Rate In Male Samples (Default 0.05)

If Y-SNPs were genotyped with missing/unknown genotypes in **≥ 5%** male samples, Y-SNPs would be removed.

(2) Calling Rate Threshold In Female Samples (Default 0.5)

When sex information was not available, the samples with **≥ 50%** missing/unknown genotypes would be “proposed” as females.

(3) Female Threshold (Default 0.9)

If Y-SNPs were genotyped in **≥ 90%** female or “proposed female” samples, Y-SNPs would be removed.

References

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. 2010. Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564-1573.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.

International Society of Genetic Genealogy. 2013. Y-DNA Haplogroup Tree 2013. Version: 8.20, Date: 26 February 2013. http://www.isogg.org/tree/. Accessed: 28 Februray 2013.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.

Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C. 2013. A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res* 23:388–295.