PopGenTransformer

PopGenTransformer Manual

Beta version

Aug 2012

Fu-Hao Lu

Plant Resources Department

Kongju Natioanal University
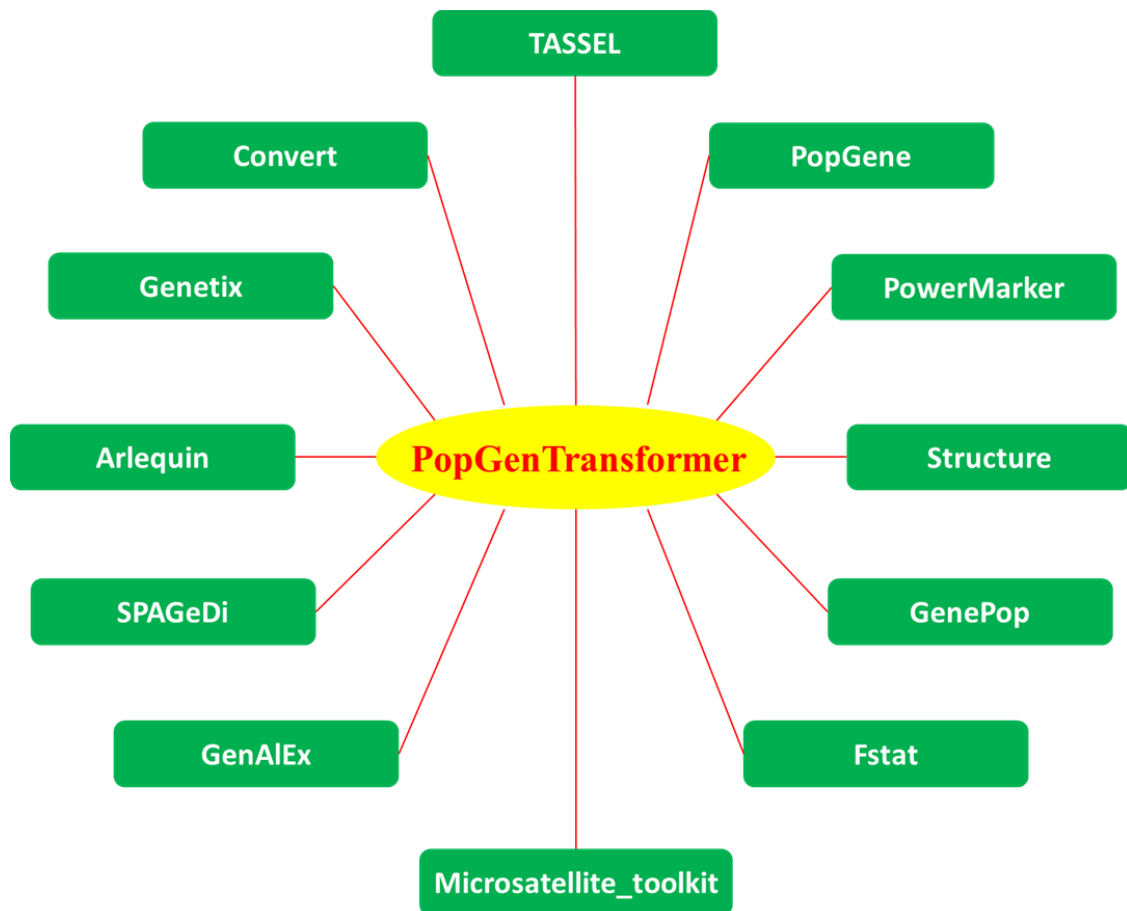
Yesan Campus, Yesan 340-702, R. Korea

1. Introduction

1.1. Brief background

If you are a geneticist working on population genetics, you must be tired of format preparation for different software and programs. Because each program may require specific input format, so you have to take some time to understand the format requirements for each. Finally you will find most of the time is spent on the format preparation. Every time when you get the genotyping data for further analysis, you must feel headache on the different format.

The situation is the same to me. So some day when I finally can't tolerate this kind of boring work, it comes to the decision that I have to solve that once for all. For the first version, I include around 10 formats for different programs. I really hope this script would be helpful for your studies. If you also like it as I do, please recommend this script to your friends, classmates, and colleagues working in the same field.

TASSEL

Convert

PopGene

Genetix

PowerMarker

Arlequin

PopGenTransformer

Structure

SPAGeDi

GenePop

GenAlEx

Fstat

Microsatellite_toolkit

1.2. Objective

This script is designed to convert your genotyping date into different input formats for Genetics software, currently supporting programs as follows:

Table 1 Program formats supported

No	Program		
1	PopGene v1.32	http://www.ualberta.ca/~fyeh/popgene.html	√
2	PowerMarker v3.25	http://statgen.ncsu.edu/powermarker/	√
3	Structure v2.3.4	http://pritch.bsd.uchicago.edu/structure.html	√
4	GenePop v4.14	http://kimura.univ-montp2.fr/~rousset/Genepop.htm	√
5	Fstat v2.93	http://www2.unil.ch/popgen/softwares/fstat.htm	√
6	Microsatellite_toolkit	http://animalgenomics.ucd.ie/sdepark/ms-toolkit/	√
		Link died	
7	GenAlEx v6.41	http://www.anu.edu.au/BoZo/GenAlEx/	√
8	SPAGeDi v1.3	http://ebe.ulb.ac.be/ebe/Software.html	√
9	Arlequin v3.513	http://popgen.unibe.ch/software/arlequin35/	√
10	Genetix v4.05	http://kimura.univ-montp2.fr/genetix/	N
11	Convert v1.31	http://www.agriculture.purdue.edu/fnr/html/faculty/ rhodes/students%20and%20staff/glaubitz/software. htm	√
12	TASSEL v2.1	http://www.maizegenetics.net/index.php?option=co m_content&task=view&id=89&Itemid=119	√

*Links were verified on 21Aug2012

1.3. System environment

- o Perl environment on Windows or Linux

- o No specific perl modules

1.4. Limitation

○ Currently only support two-column numeric diploid data

2. Input_format

You may prepare the input format **EXACTLY** in Excel or other similar

programs, such as LibreOffice_Calc, as follows:

Table 2 Format preparation for PopGenTransformer

Ind_name	Geo_Cat	Struct_Cat	MK01		Mk02		……
Ind01	CHN	pop1	187	187	129	156	
Ind02	KOR	pop2	169	169	158	158	
Ind03	USA	pop3	158	135	369	158	
Ind04	CHN	pop2	135	187	156	156	
Ind05	JPN	pop4	234	146	148	148	
Ind06	UK	pop1	158	158	135	156	
……							

1st column: individual name

2nd column: geographically categorical data, usually referring to the sample origin

3rd column: categorical data from STRUCTURE software or others.

4th column: marker name and **numeric** genotyping data (SSR, SNP or others).

Important Note:

√ **Except the blank cells between marker names, no other blank cells were allowed;**

√ **Missing data** can be indicated by "**?**" or "**.**" symbols (no quote mark)

√ For the 2nd column and 3rd column, if you do not have any idea or data, just input anything like space, numeric number, etc, with the same throughout each column. **Do not leave it blank.**

√ The data in 2nd column was set as the main categorical data for each program on default; if you do not want to include categorical data or population in your analysis, you may input the same symbol throughout 2nd column.

√ 3rd column were designed for PowerMarker analysis. If you want to use STRUCTURE results to infer different population, just exchange the 2nd column and 3rd colum.

3. Script running

(1) Copy the data you prepare in Excel to TextEditor (click one data cell,

CTRL+A, and CTRL+C); Do not include blank cells.

	A	B	C	D	E	F	G	H	I
1	Ind_name	Geo_Cat	Struct_Cat	MK01		Mk02		...	
2	Ind01	CHN	pop1	187	187	129	156		
3	Ind02	KOR	pop2	169	169	158	158		
4	Ind03	USA	pop3	158	135	369	158		
5	Ind04	CHN	pop2	135	187	156	156		
6	Ind05	JPN	pop4	234	146	148	148		
7	Ind06	UK	pop1	158	158	135	156		
8	...								
9									
10									

In TextEditor, it should look like this:

```
Ind_name→        Geo_Cat→        Struct_Cat→      MK01→  →        Mk02→  ↓
Ind01→ CHN→      pop1→  187→     187→   129→      156↓
Ind02→ KOR→      pop2→  169→     169→   158→      158↓
Ind03→ USA→      pop3→  158→     135→   369→      158↓
Ind04→ CHN→      pop2→  135→     187→   156→      156↓
Ind05→ JPN→      pop4→  234→     146→   148→      148↓
Ind06→ UK→       pop1→  158→     158→   135→      156↓
←
```

save the text file to the folder you want, for example, "C:\SoftRunData\";

and put this "PopGenTransformer.pl" script in the same folder.

Name	Date modified	Type	Size
PopGenTransformer.pl	8/21/2012 12:03 AM	PL File	18 KB
trans_input.txt	8/21/2012 5:10 PM	Text Document	1 KB

(2) Get command_prompt in Windows (Win+R, then input cmd, Enter),

or Terminal in Linux

(3) [Optional] Check your perl environment by:

perl –version, and Enter

then it appears:

```
C:\>perl -version

This is perl 5, version 14, subversion 2 (v5.14.2) built for MSWin32-x64-multi-t
hread
(with 1 registered patch, see perl -V for more detail)

Copyright 1987-2011, Larry Wall

Binary build 1402 [295342] provided by ActiveState http://www.ActiveState.com
Built Oct  7 2011 15:19:36

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl".  If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.
```

(3)Direct the to the folder your input located using "cd" command

   In Windows: cd C:\SoftRunData

(4) Run the script by typing "**perl script_name**";

```
C:\SoftRunData>perl PopGenTransformer.pl

*******************************************************************
*******************************************************************
*****This PopGenetConverter is used to convert your genotyping data*****
*****    into format for multiple programs in Population Genetics   *****
*******************************************************************
*******************************************************************

Please input the name of TEXT file you want to convert:
```

Type the TEXT name

Type the **total marker number, total individual number** and **current job name**, respectively.

```
Format supported

*********************************************************************

1-Popgene,              2-PowerMarker,  3-Structure,
4-GenePop,      5- Fstat,                       6-Microsatellite_toolkit,
7-GenAlEx,              8-SPAGeDi,                      9-Arlequin
10-Genetix,             11-Convert,             12-TASSEL,
13-(on request)
*********************************************************************

Please indicate the target format you want to convert: (1-11)

For example: 3(Enter for single convet)

Please input your choice:
```

After you input the job name, it appears the choice of different programs.
Input the number before the program you want and ENTER. **One time one**
number;

```
Please indicate the target format you want to convert: (1-11)

For example: 3(Enter for single convet)

Please input your choice:1
PopGene conversion start...
PopGene conversion finished
Do you want to convert to another format?
1 for YES
0 for No
```

For example, you want to convert into PopGene format, just type "1"

then ENTER. When you see "PopGene conversion finished…", it comes

the choice of another program conversion. Type "1" to continue and "0"

for exit.

```
1 for YES
0 for No
0

C:\SoftRunData>_
```

4. Plans in future

- o **Support one column data;**

- o **Support more programs, such as NTsys, mapmaker, joinmap;**

- o **Give more choices to customize your conversion output;**

5. Citation

This is just a small script that can save you lots of time. If you think it is helpful and want to cite our work, the citation may like this:

Lu Fu-Hao. (2012) PopGenTransformer manual. Department of Plant Resources, Kongju National University, Yesan, R. Korea. (http://code.google.com/p/popgentransformer)

6. Questions and answers

I will keep maintaining this script even after my graduation at the end of Aug 2012. If you have some problems when running this script, or have some good suggestions, please do not hesitate to E-mail me:

lufuhao@gmail.com

7. Acknowledgements

This script was finished during my Ph.D. course under the supervision of Professor Yong-Jin Park. Thanks should go to Professor Park, and my lab members. To those who make it true, I dedicate this script.