MinimalMarker
Installation Guide and Manual for Mac
Version 4

July 22,2019

FUJII, Hiroshi

hfujii@affrc.go.jp

Institute of Fruit Tree & Tea Science,

The National Agriculture and Food Research Organization (NARO), Japan

Contents

- Overview 2
- Operating environment 3
- Executable file 4
- Install 5
- Command 6
- Sample input file & sample out file 7
- Input file format 8
- Output file format 9
- Calculation option 10
- Disclaimer 11
- Copyright 12
- Citation 13
- History of MinimalMarker Installation Guide and Manual 14

Overview

MinimalMarker is a computer program that detects marker sets with the smallest number of markers to identify all cultivars from polymorphism data obtained by applying co-dominant markers or dominant markers to many cultivars. Refer to the paper for details.

Fujii et al. (2013) Minimal marker: an algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient variety identification. Journal of Bioinformatics and Computational Biology 11(2):1250022.
DOI: 10.1142/S0219720012500229

Operating environment

- Mac OSX or Linux or Windows
  - Windows users can use another MinimalMarker version with GUI.
- Perl environment
  - Perl environment exists on Mac OSX and Linux even if not installed.
  - Windows user should Install Perl environment such as ActivePerl.
    - https://www.activestate.com/products/activeperl/

Executable file

MinimalMarker1.04.pl

Install

- Put MinimalMarker 1.04.pl at any directory.
- Set the path if necessary.

Command

Execute the following command from the terminal of Mac.

Format

```
>perl minimalmarker1.04.pl [option switch] -i[input filename(s)] -o[output filename(s)]
```

Input example

```
>perl minimalmarker1.04.pl -p1 -w -g -m1 -b1 -s3 -n1 -n2 -x3 -x4 -itestin.csv -otestout.txt
```

Sample input file & sample output file

- Sample input file
  - sample_data.csv
    - Refer: Yamamoto et al. (2003) Breed. Sci. 53:35-40.

- Sample output file
  - sample_data_output.txt

Input file format

A

```
Variety,V1,V2,V3,V4,V5
M1,aa,ab,aa,ab,aa
M2,aa,ab,ab,aa,bb
M3,aa,aa,aa,aa,bb
M4,aa,aa,ab,ab,aa
M5,ab,ab,bb,ab,aa
```

B

```
Variety,V1,V2,V3,V4,V5
M1,+,+,+,+,+
M2,+,-,-,-,+
M3,+,+,+,+,-
M4,+,-,+,+,-
M5,-,+,-,-,+
```

C

```
Variety,V1,V2,V3,V4,V5
M1,ab,ad,ae,bf,cg
M2,aa,bb,cc,dd,ee
M3,ab,cd,ef,gh,hi
M4,bc,cd,ac,de,dd
M5,ab,ab,bb,ab,aa
```

D

```
Variety,V1,V2,V3,V4,V5
M1,80/84,80/84,84/84,80/84,80/84
M2,78/94,78/94,78/88,88/94,80/94
M3,193/197,193/197,195/197,195/201,193/197
M4,177/195,195/195,177/195,177/177,177/195
M5,136/136,136/136,132/136,136/136,136/136
```

- MinimalMarker requires input files in comma separated value (CSV) format.
- First term of the data should be ‘Variety’ in the first line and following it names of cultivar must be written in first line. After second line, names of markers must be written in first column and following maker types.
- Any words are permitted in the cultivar names and the marker names instead of ‘ V1’ or ‘ M1’, such as (A)
- Character notation of co-dominant marker of diploid and it is the input example of sample data shown in (A)
- (B) +/- notation of dominant marker.
- (C) co-dominant marker with any number of alleles
- (D) Numerical notation of SSR marker of diploid. 2 figures must be separated by delimiter such as ‘/’.

Outputfile format

```
number of markers:   17
number of varieties: 17
option switches:
-p: 0, -w: OFF, -g: OFF, -m: 1, -b: ON
Akatsuki x Gyosei
No marker(s) can discriminate between varieties listed above
number of marker sets found: 6
list of marker sets found:
M1a,M4c,MA023a
M4c,MA007a,MA023a
M4c,MA007a,MA035a
M4c,MA014a,MA023a
M4c,MA015a,MaA027a
M4c,MaA027a,MA035a
elapsed time: 0 sec.
```

- Sample output data.
- Peach data was used as sample data (Yamamoto et al., 2003, Breed Sci 2003, 53:35-40.).
- The program outputted that SSR genotypes of 'Akatsuki' and 'Gyosei' are identical, and computation was continued.
- Six minimal marker sets which consist of three markers were outputted.

Calculation options in MinimalMarker

Option	Argument	Definition
-t		Display all identified minimal marker set(s) on the screen.
-w[1/2/n]	1: (default) 2: ("-w" is the same as "-w2") n: (n>2)	Search for the minimum marker set(s) necessary to identify all two cultivar combinations with at least n markers. Even if this option is specified, redundant marker sets which just added another marker to the minimum marker set will not be detected
-p[0/1]	0: Not output the table (default) 1: Output the table but not search	Output a table indicating whether or not to distinguish between the two cultivars with each marker, as in Table 2 (Fujii et al. 2013)
-m[0/1/2/n]	0: Enumaration metod 1: Branch-and-bound method (default) 2: Heuristic method n: Heuristic method with n power (n>2)	Enumeration method and branch-and-bound method can achieve the same result. By using the branch-and-bound method, the calculation speed can be increased more than the enumeration method. The heuristic method can dramatically reduce the calculation time when the number of markers and the number of varieties are large. However, we do not guarantee that it is a minimal marker set and that it will detect all minimal marker sets. Increasing n increases the power of detection but slows the calculation time.
-b[0/1]	0: Arithmetic operation 1: Bit operation (default)	When the calculation time is expected to be long, computation can be accelerated by use of the bit operation. Bit operations give the same results as arithmetic operatinos.
-k[character or symbol]	(e.g. -kZ -k#)	Specify a character or symbol indicating missing data (case sensitive)
-l[0/n]	0: Search all minimal marker sets (default) n: Search for a specified number (n) of marker sets and stop (n>0)	Specifies the number of minimal marker sets to search. Use when searching for all minimal marker sets may take time.
-v[n]	n: Column number of cultivar in given table Multiple can be specified (e.g. -v1 -v2)	When searching for a minimal marker set to distinguish one/some cultivar(s) from all other cultivars, the cultivar(s) is/are designated
-n[n]	n: Row number of marker in given table Multiple can be specified (e.g. -n1 -n2)	Specify marker(s) that must be included in the minimum marker set.
-x[n]	n: Row number of marker in given table Multiple can be specified (e.g. -x1 -x2)	Designate marker(s) to be excluded from the minimum marker set.
-e[1/n]	1: Stop searching when all minimum marker sets are found (default) n: Continue searching to the specified number of markers even if all the minimum marker sets are detected (n>1)	Specify the number of markers to stop the search (with or without minimum marker set detection).
-s[1/n]	1: Start search from 1 marker (default) n: Number of markers to start the search (n>1)	Specify the number of markers to start the search. In the case of dominant markers, at least 7 markers are required to identify 100 cultivars. To speed up such searches, use the -s7 option.

Disclaimer

- MinimalMarker provided by National Agriculture and Food Research Organization of Japan (NARO) may be freely distributed, provided that no charge above the cost of distribution is levied, and that the disclaimer below is always attached to it.
- MinimalMarker is provided as is without any guarantees or warranty. Although the author has attempted to find and correct any bugs in the program on MinimalMarker, the author is not responsible for any damage or losses of any kind caused by the use or misuse of MinimalMarker. The author is under no obligation to provide support, service, corrections, or upgrades of MinimalMarker.

Copyright

- The computer program on MinimalMarker is protected by copyright. Except as specifically permitted, no portion of this software may be distributed or reproduced by any means, or in any form, without NARO's prior written permission.

- COPYRIGHT NOTICE
  © 2006-2011 NARO, Japan. All rights reserved.

Citation

- When publishing an academic paper using the results obtained using the MinimalMarker, please describe the use of MinimalMaker in the paper and cite the following literature.

- Fujii H., T. Ogata, T. Shimada, T. Endo, H. Iketani, T. Shimizu, T. Yamamoto and M. Omura (2013) Minimal marker: an algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient variety identification. Journal of Bioinformatics and Computational Biology 11(2):1250022.
  DOI: 10.1142/S0219720012500229

History of MinimalMarker Installation Guide and Manual

- Version 1 June 11 2011
- Version 2 August 16 2011
- Version 3 September 6 2011
- Version 4 July 22 2019