PMSeeker INSTRUCTION

INSTRUCTION FOR EXHAUSTIVE ALGORITHM

1. PRINCIPLE

To put it simply, the basic principle of PMS screening by the exhaustive algorithm is to take into account all of the genotypes that various parent combinations can theoretically produce, as well as all of the marker combinations with different sizes ranging from small to large, in order to determine whether progeny with any genotype can be completely traced back to the single parent pair that produced them.

2. TYPE OF MARKERS TO USE

This algorithm is applicable to all molecular markers, including codominant markers (SSR, SNP, MH, etc.) and dominant markers (RAPD, ISSR, etc.).

3. INPUT, OUTPUT AND PARAMETERS

INPUT FILE 1：Parents' genotype file (CSV file separated by ",")

The format is similar to the CERVUS input file format. The first line is basic information of markers, including "ID" and the IDs of each marker (the two genotypes of each marker are distinguished by ".A "and".B "). From the second line, specific genotypes of each parent are listed.

```
ID,M1.A,M1.B,M2.A,M2.B,M3.A,M3.B
2jing,0,0,0,1,0,1
F40,1,2,2,3,0,0
1jing,0,0,4,3,2,0
F13,0,0,2,1,2,0
M10,0,0,0,1,0,0
```

INPUT FILE 2：Information for parent pairs (TXT file separated by "\t")

There is no need to distinguish between male and female parents. For mixed groups, just note that individuals of the same gender are in the same column, and different genders can be distinguished by "\t". The first line can be arbitrarily marked with "female" and "male", also separated by "\t" (the upper figure). For populations with known parent pairs, ONE ROW corresponds to ONE DYAD, and the dyad is

separated by "\t" (the lower figure). You can also run the "Unknown Gender" scenario without providing the parent pair information.

```
female  male
F13     M10
F40     1jing
2jing
```

```
F13     M10
F40     1jing
F13     2jing
F40     M10
```

OUTPUT FILE：${prefix}.txt

Record all solutions that meet the requirements. When there is no solution, "Full Marker:" will be displayed in the output file. When there is a result, it will output "A valid combination:", where each line is a feasible solution.

```
A valid combination: CI01000046_472435_472569, CI01000358_825506_825552, CI01000026_3732966_3733034
A valid combination: CI01000046_472435_472569, CI01000026_3732966_3733034, CI01000358_825743_825793
A valid combination: CI01000358_825506_825552, CI01000026_3732966_3733034, CI01000358_825743_825793
```

Parameters

Basic parameters		
-h, --help	show this help message and exit	[None]
Required parameters		
-i, --input	Specify the genotypes of each individual for each marker (in the format of CSV)	*STR*
-o, --output	Output information of final results	*STR*
Optional parameters		
-b, --block	Parents' gender information or group information	*STR*, [None]
-f, --fam_file	A file specify the families (parent pairs), which contains one or more parent pairs (name pair or line number pair, like F1, M1 or 1, 17) in each line, multiple pairs should be separated by semicolon ";".	*STR*, [None]
--first-comb	Only output the first PMS.	*BOOL*, [False]
--all-size	Only output PMSs with fewest markers.	*BOOL*, [False]
--min-marker	Minimum number of markers in a PMS.	*INT*, [1]
--max-marker	Maximum number of markers in a PMS.	*INT*, [the number of all markers]

Note: Once "--fam-file" is provided, -b (--block) will not be used for making parental pairs.

INSTRUCTION FOR EXHAUSTIVE ALGORITHM

1. PRINCIPLE

Briefly speaking, The program screens markers by creating a multiway tree. All candidate parent pairs make up the root node. With the first layer of the tree, different markers can be used to produce the nodes of the next layer of the tree. The optimal next layer is chosen as the new layer with the lowest PDI, and the associated marker is added to the final PMS. By looping in this manner until all child nodes of some layer are leaf nodes, the whole multiway tree is retrieved, and the optimum marker set is obtained.

2. TYPE OF MARKERS TO USE

This algorithm is applicable to all molecular markers, including codominant markers (SSR, SNP, MH, etc.) and dominant markers (RAPD, ISSR, etc.).

3. INPUT, OUTPUT AND PARAMETERS

The greedy method input file is consistent with the exhaustive method, so I won't repeat it here. If there is no PMS, no results will be output, and "Fail to get the markers." will be displayed in the log file.

OUTPUT FILE 1：${prefix}.mm

Display the output PMS, which each row corresponding one layer, and if there are multiple marks in the same row, these marks have the same effect and can be replaced with each other.

```
['CI01000256_143531_143714']
['CI01087614_4687_4837', 'CI01000301_745642_745844', 'CI01000004_16955361_1
6955582', 'CI01000327_2473867_2474034', 'CI01000400_41055_41317', 'CI010002
99_5418306_5418563', 'CI01000321_5113983_5114147', 'CI01000321_5113617_5113
816', 'CI01000316_1054723_1054954', 'CI01086133_7968_8179', 'CI01000299_541
7935_5418090', 'CI01086133_7255_7466', 'CI01000299_5417656_5417899', 'CI010
00030_8452565_8452808', 'CI01000000_3227733_3227959', 'CI01000321_4747626_4
747801', 'CI01000030_10281333_10281519', 'CI01000304_14288520_14288779', 'C
```

OUTPUT FILE 2：${prefix}.txt

Record the genotypes of each parent under the optimal solution markers (OPTIMAL PMS). For replaceable markers, only the genotype corresponding to the first marker is displayed.

```
ID,CI01000256_143531_143714.A,CI01000256_143531_143714.B,CI01087614_4687_48
37.A,CI01087614_4687_4837.B
2jing,6,4,2,2
F13,2,2,2,2
F40,3,5,3,0
1jing,3,1,2,1
M10,4,0,2,2
```

Parameters:

Basic parameters		
-h, --help	Show this help message and exit	[None]
--PRINCIPLE	Display the principle of the algorithm.	[None]
Required parameters		
-i，--input	Specify the genotypes of each individual with each marker (in the format of CSV)	*STR*
-o, --output	Prefix of final results of the PTMS.	*STR*
Optional parameters		
--trim	Specify the parents, from a pool of parents, that areneeded to be analyzed, seperated by ','.	*STR*, [None]
--rlt_num	Expected number of PMSs. If the final results can't meet the requirement, the results that have been obtained will be listed out in the output file.	*INT*，[1]
--time	Maximum number of runs allowed for one PMS	*INT*, [1]
--block	If the gender information is known, specify the file containing the information.	*STR*, [None]
--pair	If parent pair are specified with one row per pair using '\t' to separate two parents, then combinations of parents are to specified.	*STR*, [None]
--rm_1st	Markers to be excluded for subsequent PMSs. Before screening any subsequent PMS, some markers (all['False'] or only the first one['True']) in each of previous PMSs will be excluded from the candidate marker set in order to obtain a distinct PMS.	*BOOL*, [False]
--PDI	Whether to only output first-layer PDI for all candidate markers	*BOOL*, [False]

Note: '--pair' and '--block' should not be set simultaneously.