



USER MANUAL VERSION 2.1

JOHN R. GARBE¹ AND YANG DA²

¹MINNESOTA SUPERCOMPUTER INSTITUTE

²DEPARTMENT OF ANIMAL SCIENCE

UNIVERSITY OF MINNESOTA

August 6, 2013

1	INTRODUCTION	3
2	GETTING STARTED.....	3
2.1.1	<i>On Windows.....</i>	3
2.1.2	<i>On Linux.....</i>	3
2.1.3	<i>On Apple Mac OS X.....</i>	3
2.2	ERRORS	3
2.3	DOMINANT/RECESSIVE LOCI	4
2.4	SEX-INFLUENCED LOCI	4
2.5	SEX-LINKED LOCI	4
2.6	IMPRINTED LOCI	4
3	FILE FORMATS.....	4
3.1	PARAMETER FILE.....	5
3.1.1	<i>Parameters</i>	5
3.2	POPULATION DATA FILE.....	6
3.2.1	<i>Crimap Format.....</i>	6
3.2.2	<i>Locusmap Format.....</i>	6
3.3	LOCUSNAME FILE	7
3.4	PAIRWISE.....	8
3.5	LOCUS_ORDER	8
3.6	LINKAGE_MAP	8
3.7	LOCUS_INFO.....	8
3.8	DATA_ERRORS	8

1 Introduction

Locusmap is a computer package designed for rapid linkage analysis and map construction of loci with a variety of inheritance modes. The current version offers linkage analysis between codominant loci, between dominant loci, between codominant and dominant loci, and between codominant and sex-influenced, sex-linked, and imprinted loci. Using genotypic data supplied by the user, Locusmap calculates two-point recombination frequencies and LOD scores and then identifies linkage groups and orders loci within each linkage group automatically. Locusmap checks for genotypic and pedigree errors, and fills in missing genotypes when possible, and currently is a command-line program available for Windows, Linux, and Apple Mac OS X. Analyzing data sets with 150 loci and a few hundred individuals takes about two minutes on a 2-GHz computer.

2 Getting Started

To run Locusmap, open a MS/DOS, Linux or Mac OS X terminal and move to the directory where the Locusmap executable is located. Create or edit the locusmap.par parameter file if needed. The parameter file specifies the input and output files, as well as program options. If the parameter file cannot be read default values will be used. The population data filename may be specified on the command line. Type Locusmap at the prompt to run the program. Diagnostic information will be printed to the screen and the results of the analysis will be located in the output files specified in the parameter file.

2.1.1 On Windows

```
>chdir c:\locusmap_location  
>locusmap [datafilename]
```

2.1.2 On Linux

```
$cd /home/yourid/locusmap_location  
$./locusmap [datafilename]
```

2.1.3 On Apple Mac OS X

```
$cd /Users/yourid/locusmap_location  
$./locusmap [datafilename]
```

2.2 Errors

In the event of a program error the program will halt and print out an error message. There are two types of errors: fatal errors and fatal internal errors. Fatal errors are the result of user error, which means you did something wrong so you should fix it and run the program again. Fatal internal errors are the results of bugs in the program. If you

encounter an internal error please send a bug report to yda@umn.edu. If the program ever stops without an explanation or has a segmentation fault, please send a bug report to yda@umn.edu.

2.3 Dominant/recessive loci

Dominant/recessive loci must be coded. Dominant alleles should be coded 999 and recessive alleles should be coded 888. Both alleles at a heterozygous locus should be coded as dominant. These codings may be redefined in the parameter file. Note that the dominant/recessive coding numbers (888 and 999 by default) may not be used to code non-dominant/recessive loci.

2.4 Sex-influenced loci

Sex-influenced loci must be coded. An individual that expresses a male dominant sex-influenced trait should be coded 9000 9000 at the sex-influenced locus, and unexpressed individuals should be coded 8000 8000. For a female dominant trait reverse the codings - expressed should be 8000 8000 and unexpressed 9000 9000. These codings may be redefined in the parameter file. Note that the sex-influenced coding numbers (8000 and 9000 by default) as well as the numbers plus one (8001 and 9001 by default) may not be used to code non-sex-influenced alleles.

2.5 Sex-linked Loci

Each locus is automatically tested for sex-linkage. If a locus tests positive the log likelihood that the locus is sex-linked is printed to the screen. Currently sex-linked loci are identified solely by the absence of heterozygous males at the sex-linked locus.

2.6 Imprinted Loci

Imprinted loci must be coded. An individual that expresses a female imprinted trait should be coded 950 950, and individuals not expressing the female imprinted trait should be coded 850 850. A female-imprinted gene is a gene where only the allele received from the male parent is expressed. To encode a male-imprinted gene simply swap the genders of all the individuals in the pedigree (or you could just swap the definitions of male and female in the parameter file).

3 FILE FORMATS

Locusmap uses three input files: the parameter file, the population data file, and the locusname file. Locusmap produces five different output files: locus_info, pairwise, data_errors, locus_order, and linkage_map. The names and locations of these files are specified in the parameter file.

3.1 Parameter File

The parameter file allows the user to pass information to the program. The parameter file should be named locusmap.par and located in the same directory as the Locusmap executable. If the parameter file cannot be found default values for all parameter will be used. The file consists of name and value pairs. Each name and value pair must be on a separate line. Lines beginning with # are ignored, as are blank lines. All accepted parameters are listed here. Note that all input and output files (including any data file given on the command line) must be located in the directory specified by the directory parameter (the current directory by default).

3.1.1 Parameters

Name	Default value	Description
data_file	data_file.dat	Name of the population data file
loci_names	loci_names.dat	Name of the loci names file
loci_info	loci_info.out	Name of the loci info file
pairwise	pairwise.out	Name of the main output file
linkage_map	linkage_map.out	Name of the linkage output file
locus_order	locus_order.out	Name of the locus order output file
directory	.	Name of the directory where the input and output files are located
lod_threshold	0	Minimum LOD score for locus pair to be printed to the output file
cutoff	1	Minimum number of useful individuals in a phase-unknown family required for information to be used
brute_limit	10	Maximum number of loci to be ordered by the brute force method
num_best_orders	5	Number of orders to print for each linkage group. Setting this option to 0 will skip the linkage group ordering step
map_function	haldane	Map function used to convert recombination frequency to centiMorgans. Choose between haldane and kosombi.
locus_output_type	Locusnumber	How loci are output: with numbers or names. Choose between locusnumber and locusname.
dominant	999	Special coding number for a dominant allele
recessive	888	Special coding number for a recessive allele
expressed	9000	Special coding number for an expressed allele at a sex-influenced locus
unexpressed	8000	Special coding number for an unexpressed allele at a sex-influenced locus

male_def	1	Arbitrary string of up to 20 characters (without embedded blanks) identifying an individual in the population data file as male
female_def	2	Arbitrary string of characters identifying an individual in the population datafile as female
unknown_def	3	Arbitrary string of characters identifying an individual in the population data file as having an unknown gender
ask_sexlinked	true	When set to true the program will ask whether loci identified as sex-linked should actually be analyzed as sex-linked. Otherwise all sex-linked loci are analyzed as sex-linked.

3.2 Population Data File

The population data file contains all the genotypic data to be analyzed. The filename may be specified in the parameter file or on the command line. A filename given on the command line overrides any filename listed in the parameter file. Two different population data file formats are allowed: the format used by Crimap's ".gen" files or the Locusmap format (see below). Crimap formatted files must have filenames ending in ".cri". Locusmap formatted files may have any name provided they do not end with ".cri". Filenames must not begin with a number.

3.2.1 Crimap Format

Locusmap can accept files formatted according to the ".gen" format described in the Crimap documentation with the following exceptions:

- Note that any male_def, female_def, and unknown_def settings in the parameter file are ignored (Crimap values are used as defined in the documentation)
- Individual ID's may be any string of characters (without embedded blanks), not just integers

3.2.2 Locusmap Format

In this format each line in the datafile contains the information for one individual. Individuals are grouped by family. A family is a group of related individuals which may be either offspring, parents, or grandparents. An offspring is an individual with no offspring (children). A parent is an individual whose offspring do not have offspring. A grandparent is an individual whose offspring have offspring. All offspring with the same sire should, along with their dames, be listed in the same family (i.e. half-sib families should be listed together in one family).

The data items in this file are as follows:

For each individual:

{family ID} {ID} {parent1's ID} {parent2's ID} {sex: 2 if female, 1 if male, 3 if unknown} {locus1 allele1}{locus1 allele2}{locus2 allele1}{locus2 allele2} . . .

- Each item must be separated by a comma and/or one or more spaces and tabs
- The family ID and individual ID's may be any string of characters (without embedded blanks)
- The definition of male, female, and unknown can be redefined in the parameter file
- Alleles must be represented by integers, with missing alleles scored as 0.
- Each individual must be listed on a separate line
- The datafile may contain blank lines
- If an individual has no (i.e. unknown) parent then the parent ID should be coded as 0
- All children (individuals with no offspring) must have parents (non-zero parent ID's)
- An individual ID must be unique within a family
- A family ID must be unique within the datafile
- Individuals in the same family must be listed together
- individuals must be grouped by families
- individuals may be listed in any order within a family
- Individuals cannot have parents in a different family

3.3 Locusname File

The locusname file allows the user to assign names to the loci in the datafile, so that loci can be referred to by name instead of number in the output files. The user may also specify which chromosome each locus is on, and the coordinate (in basepairs) of the locus. A chromosome number of 0 indicates an unknown chromosome. Each line in the file contains a locus name and a chromosome number. The first name listed in the file refers to the first locus listed in the data file, the second name refers to the second locus, and so on. The following is an example of the locusname file:

Loc1	2	5328058
Loc2	17	9816323
Loc3	1	8473829
Loc4	5	4738495
Loc5	23	2140907
Loc6	0	
Loc7	0	

3.4 Pairwise

The pairwise file lists all the locus pairs and the gender-specific recombination frequencies (x and y), gender-averaged recombination frequencies (theta), and gender-averaged LOD score. Only the locus pairs who's LOD is larger than lod_threshold are listed.

3.5 locus_order

The locus_order file lists the most likely orderings of each linkage group and the likelihood of the order, as well as all the loci that could not be placed in a linkage group.

3.6 linkage_map

The linkage_map file lists the best order for each linkage group and map distances between the loci.

3.7 locus_info

The locus_info file contains information about the informativeness of each locus: the percentage of heterozygous sires, percentage of heterozygous dams, percentage of individuals with either one or two informative meioses (PIC), and 2 times the percentage of informative meioses (offspring meioses for which allele transmission can be determined either directly or indirectly).

3.8 data_errors

The data_errors file contains a list of all the population datafile errors, including noninheritance errors.

REFERENCES

- Da, Y., J. Garbe, N. London, and J. Xu. 2002. Linkage analysis using direct and indirect counting and relative efficiencies for codominant and dominant loci. *J. Animal Sci.* 2002, 80(10):2528.
- London, N.R., J.R. Garbe, S.M. Schmutz, M.S. Abrahamsen, and Y. Da. 2003. Linkage analysis for mapping genes of sex-influenced traits. *Mammalian Genome* 2003, 14(4):261-267.

ACKNOWLEDGEMENTS

Locusmap version 2.1 was supported by USDA National Institute of Food and Agriculture Grant no. 2011-67015-30333 and by project MN-16-043 of the Agricultural Experiment Station at the University of Minnesota, and Locusmap version 1.1 was supported by Cargill, Inc. Supercomputer computing time was provided by the Minnesota Supercomputer Institute at the University of Minnesota.