

# Stochastic Multiple Markov Blankets with Ant Colony Optimization (SMMB-ACO) documentation

---

SMMB-ACO is a program permitting to find epistasy patterns using Multiple Markov Blankets generated by ant colony optimization.

## Compilation instructions

---

In order to compile this method a makefile was given at the root of SMMB-ACO's folder. Please notice that a general makefile was generated at the project's root permitting to produce executable of both methods.

## Preparation of pre requisit

In the makefile please change BOOST\_FOLDER value with path of the installed boost library on current workstation.

Moreover, please check if current g++ version is compatible with C++11 functionalities (version  $\geq 5.0$ ). Please also check boost library. Authors cannot guaranty compatibility with other boost version even if boost version seems to be compatible. For this method version 1.61.0 was employed.

## Compilation

To compile this method, please call those lines a the root of the project:

```
cd smmb_aco
make
```

If a recompilation is needed use:

```
cd smmb_aco
make clean
make
```

Execute program with:

```
./smmb_aco.exe <path_to_genotype_dataset> <path_to_phenotype_dataset> <path_to_para
```

## Summary of files of this project

| Module   | Files              | Usability  | Inputs   | Outputs   |
|----------|--------------------|--|--|---|
| Parsing  | file_parsing       | Allows to parse files containing data set                    | Files names given as argument in program's call                          | One boost matrix containing all genotype data<br>One boost vector containing all phenotype data<br>boost vector with SNPs IDs |
| Parsing  | parameters_parsing | Allows to parse parameters and save them in parameters class | Will use parameters.txt  | Class object with all parameters as class variables   |
| Smmb aco | smmb_aco           | Core of this project allows to run smmb-aco's algorithm      | Parameters object<br>Matrix of genotype data<br>Vector of phenotype data | A set of different Markov Blanket found during the run  |
|          |                    |  |  |   |

|                |                |  |  |  |
|----------------|----------------|--|--|--|
| Statistics     | statistics     | Used to do g2 conditional test of independence   | Matrix column of genotype (subset of genotype matrix with only tested SNPs)<br>Vector of phenotype<br>Indexes of tested SNPs   | g2 score and associated p-value<br>Number of cell considered as non reliable because $n < 5$ |
| Statistics     | contingencies  | Build observed and expected contingencies table, some functions useful to work on contingencies tables | Matrix column of genotype (subset of genotype matrix with only tested SNPs)<br>Vector of phenotype<br>Indexes of tested SNPs<br>Number of observation in tested subset | One observed contingency table<br>One expected contingency table                             |
| Output writing | output_writing | Can save algorithm's results in file   | Markov blanket and associated score  | Final result file  |
| Tools          | tools          | Used to generate random discrete distribution used to pick SNPs when building subset                   | Weight vector size of subset in construction<br>Random seed  | Subset of SNPs for every ant   |

# Parameters

---

Parameters of this program are stored in parameters folder into a file called parameters.txt. This file is commented to permit an easy setup for particular needs. This allows to tweaks many parameters like output directory or output prefix and many parameters specific to SMMB-ACO.

## Launch the analysis of the toy data set

---

With this project is provided a python script allowing a simple launch on every platform equipped with python. This script will launch program on a naive data set with default configuration given by authors.

To execute this script please call:

```
./launch_smbaco_toy_example.py
```

## Expected output

---

This method give as output a text file gathering all results. Here is an example of expected output:

```
# Result from SMMB-ACO
# Pattern || Occurences || G2-score || p-value || unreliable case
{M0P7,M0P8} || 254 || 64.744194 || 4.87e-12 || 0
{N9,M0P8} || 3 || 41.084591 || 2.79e-07 || 0
{M0P8} || 49 || 25.260149 || 3.27e-06 || 0
{M0P7} || 78 || 23.079254 || 9.74e-06 || 0
{N0,N18} || 3 || 19.843737 || 2.95e-03 || 0
{N6,N17} || 2 || 18.021011 || 6.18e-03 || 1
{N4,N7} || 1 || 17.862349 || 6.59e-03 || 1
{N9} || 13 || 8.291620 || 1.58e-02 || 0
{N2,N15} || 2 || 13.636168 || 3.40e-02 || 0
{N6} || 21 || 6.601450 || 3.69e-02 || 0

# Time of execution: 4234 milliseconds
### SMMB_ACO has finished please see results in:
./smbb_aco/toy_example/toy_results/result_genotypes_toy_dataset_smbb_aco.txt
```

Those results can be checked using our evaluation tool given in evaluation folder.

Using toy data set:

```
Filename,TP,FP,FN,recall,precision,f_measure,power
```

```
genotypes_toy_dataset.txt,100,0,0,1.0,1.0,0.5,1.0
```

## Built With

---

- [BOOST](#) - peer-reviewed portable C++ source libraries

## Authors

---

Tanguy Lallemand -  
Jonathan Cruard