



```
name: dummy  
file: mapArglogo  
state: unknown
```

juillet 2011 :: version 0.8

Dernière modification: lundi 5 septembre 2011

Tables des matières

1	Introduction	1
2	Installation de FastSimCoal	1
3	Utilisation de FastSimCoal	2
4	Aperçu général de Sample	3
5	Changements récents	3
6	Compilation	3
7	Ligne de commande et options disponibles	4
8	Fichiers de sortie	5

1 Introduction

[Sample](#) a pour objet de convertir la sortie du programme [FastSimCoal](#) de Laurent Excoffier en un fichier lisible par [MapArg](#). Ce court manuel explique la logique du programme, et donne un exemple.

2 Installation de [FastSimCoal](#)

Le programme [FastSimCoal](#) est disponible sur une page d'ici la page: [FastSimCoal](#). Sur cette page, téléchargez la version MacOSX 64 bits. La distribution contient un manuel très bien fait qui décrit son usage. Pour se faciliter la vie, on l'installera dans le système pour pouvoir l'utiliser de n'importe quelle place de travail. Les étapes à suivre:

1. création du répertoire `.bin` sur votre compte (ce répertoire contiendra tous vos exécutables, qui seront trouvés par le système d'als que vous appeler un des programmes présents dans ce répertoire):

```
mkdir ~/.bin
```

2. En supposant que [FastSimCoal](#) se trouve téléchargé dans le répertoire `fsc_mac64` du répertoire `Downloads` (par défaut, ça devrait être comme cela !)

```
cp ~/Downloads/fsc_mac64/fastsimcoal ~/.bin
```

3. pour que le système trouve les programmes du répertoire `.bin`, il faut l'ajouter au `PATH`:
 - a. Éditer le fichier `~/.bash_profile` avec un éditeur de texte;
 - b. Rajouter à la fin la ligne: `PATH="~/.bin:${PATH}"`;
 - c. Fermer le shell actuel, ou faire `source ~/.bash_profile`.

3 Utilisation de FastSimCoal

Pour utiliser `FastSimCoal` il faut un fichier paramètre. Des exemples sont dans la distribution et dans le manuel. Voici un exemple avec le fichier `PopAEch1.par`:

```
1 //Number of population samples (demes)
2 1
3 //Population effective sizes (number of genes)
4 20000
5 //Sample sizes
6 1000
7 //Growth rates : negative growth implies population expansion
8 0
9 //Number of migration matrices : 0 implies no migration between demes
10 0
11 //historical event: time, source, sink, migrants, new size, new growth rate, migr. matrix
12 0 historical event
13 //Number of independent loci [chromosome]
14 1 0
15 //Per chromosome: Number of linkage blocks
16 1
17 //per Block: data type, num loci, rec. rate and mut rate + optional parameters
18 SNP 150 0.000001 0.00000002 0.01
```

Brièvement, on demande la simulation d'un échantillon de 1000 haplotypes d'une population de taille effective 20 000. On demande la simulation de 150 SNPs tel que le taux de recombinaison entre 2 SNPs est 0.000001, et le taux de mutation 0.00000002. Cela nous fait un taux de recombinaison global de:

$$\rho = 4 \times Ne \times (L - 1) \times r = 4 \times 20000 \times (150 - 1) \times 0.000001$$

La commande typique ressemble à ceci:

```
fastsimcoal -i PopAEch1.par -n 1 -g -p -I
```

Vous pourrez voir les autres options et les détails dans le manuel de `FastSimCoal`. Les options `-g` et `-p` sont importantes, elles demandent une sortie de fichiers avec des diplotypes phasés.

`FastSimCoal` va automatiquement créer un répertoire de nom `PopAEch1` (le nom du fichier paramètre), et créer le fichier `PopAEch1_1_1.arp` qui contient les diplotypes: c'est de fichier qui nous intéresse.

4 Aperçu général de Sample

Ce programme, en partant d'une population¹, tire un échantillon aléatoire (stratifié ou aléatoire simple), applique un modèle de pénétrance, puis écrit divers fichiers utilisables dans MapArg.

5 Changements récents

Voir la section 7 sur les nouvelles options.

6 Compilation

Le projet XCode est fourni, c'est donc la façon la plus simple de le compiler.

¹ on abuse car il s'agit à proprement parlé d'un échantillon et non pas d'une population

7 Ligne de commande et options disponibles

La ligne de commande typique ressemble ÅĖ:

```
./Sample -i PopAEch1.par -d 0.1 -s 100 100 -F 0.01 0.01 0.90 -o toto -n 1
```

Les options sont:

-o <i>abcd</i>	fichier de sortie;
-d <i>x</i>	ξ , frÅĖquence du TIM;
-i <i>abcd</i>	fichier d'entrÅĖe;
-r <i>x</i>	racine du gÅĖnÅĖrateur alÅĖatoire;
-s <i>n</i> _{cas} <i>n</i> _{con}	sÅĖlectionne <i>n</i> _{cas} individus atteints, et <i>n</i> _{con} individus non atteints;
-S <i>x</i>	sÅĖlectionne <i>x</i> individus au hasard dans la population, sans regard ÅĖ leur statut cas/contrÅĖtes;
-F <i>f</i> ₀ <i>f</i> ₁ <i>f</i> ₂	pÅĖnÅĖtrances dans le cas diploÅĖde;
-T	pour garder obligatoirement le TIM dans les marqueurs.
-n <i>x</i>	nombres de TIMs (1 par dÅĖfaut)
-M <i>abcd</i>	si plusieurs TIMs, modÅĖle gÅĖnÅĖtique pour crÅĖler la maladie («rare», ou «aumoins1»).
-y <i>abcd</i>	oÅĖz <i>abcd</i> = { <i>FSC</i> , <i>MS</i> } pour indiquer que le fichier de donnÅĖes provient de FastSimCoal (dÅĖfaut) ou de ms respectivement.

La nouvelle option (expÅĖrimentale) permet de simuler l'effet de plusieurs TIMs. Actuellement, seulement deux modÅĖles sont disponibles, mais il est trÅĖs facile d'en crÅĖler de nouveaux;

- «rare»: pour chacun des TIMs, un individu est malade avec les pÅĖnÅĖtrances $F = (f_0, f_1, f_2)$. Il est ensuite considÅĖrer comme malade si il malade ÅĖ un des TIMs.
- «aumoins1»: on commencer ÅĖ regarder si l'individu si il possÅĖde au moins une allÅĖle mutante pour un des TIMs: si oui, il est malade avec les pÅĖnÅĖtrances f_0, f_1 .

Deux nouvelles options (expÅĖrimentales) ont ÅĖtÅĖ mises en place:

-E enrichir les cas si nÅĖcessaire;

-P x cr  le une population de x diplo  des.

L’option **-P** x , cr  le une population de x diplo  des    partir de l’  chantillon de **FastSimCoal** ; pour cela, pour chaque individu que l’on veut cr  ler, on tire au hasard de la population d’haplo  des (on ne tient pas compte des diplotypes d’  j   form  s donc) deux haplotypes que l’on met ensemble pour former un diplotype. Ce que l’on fait est de cr  ler de nouveaux “enfants” d’une population de “parents” ou il n’y aurait dans le processus ni recombinaison ni mutation (mais notons qu’il ne s’agit alors que d’une seule g  n  ration). La taille du nouvel   chantillon (x) est limit  e par la taille de la m  moire (1 million facilement, mais plus... pas facile !).

L’option **-E** (sans aucun param  tre) permet, si cela est n  cessaire, de g  n  rer plus de cas que l’  chantillon de **FastSimCoal** en poss  de. Le principe est le m  me qu’avec l’option **-P** mis    part qu’ici on ne garde que les cas. R  capitulons: si le nombre de cas demand  s dans l’option **-s** est plus grand que le nombre de cas disponible dans l’  chantillon **FastSimCoal** alors si l’option **-E** est pr  sente, on va g  n  rer autant de cas que demand  s; sur le principe d’  crit ci dessus, on g  n  re des diplotypes, puis on attribue au hasard selon $F = (f_0, f_1, f_2)$ un statut cas/contr  te au diplotype g  n  r  ; si c’est un contr  te, on l’oublie, si c’est un cas, on le garde. Comme on ne garde que les cas (contrairement    l’option **-P**), on peut g  n  rer un nombre de cas beaucoup plus grand.

Attention: ces deux options ne sont en place qu’avec un seul TIM (bien que ce serait possible de le faire pour plus de TIMs).

8 Fichiers de sortie

Sample cr  le automatiquement un certain nombre de fichiers de sortie; le nom des fichiers est bas   sur la cha   ne de caract  re suivant l’option **-o** dans la ligne de commande. Ainsi, pour **-o toto**, on a

   **toto-info.txt**: contient les informations sur l’ex  cution du programme.

De plus, les fichiers suivants sont cr  es (ils contiennent tous des diplotypes):

- **totoGD.dat**: contient les séquences génomiques; la première colonne contient un identificateur (la position originale dans le fichier provenant de **fsc**). La seconde colonne contient le phénotype, créé à partir des proportions f_0 et f_1 , et enfin les autres colonnes contiennent les marqueurs.
- **totoGD.par**: contient les paramètres (coordonnées, taux de recombinaison, mutation, etc).
- **totoHD.dat**: même fichier que **totoGD.dat**, mais à part que ce sont des diploypes phasés.
- **totoHD.par**: même fichier que **totoGD.par**, mais à part que les paramètres concernent des diploypes phasés.

∴ fin ∴