

Programme de Sélection

Jacques David, Philippe Brabant et Timothée Flutre

17/02/2016

Contents

1	Contexte	2
2	Informations biologiques	2
2.1	<i>Apimeta simulans</i> , une espèce pleine d'avenir	2
2.2	Biologie de l'espèce	2
2.3	Génome et ressources génomiques	2
2.4	Ressources génétiques et données disponibles	2
3	Moyens expérimentaux	3
3.1	Réseau d'expérimentation	3
3.2	Serre	3
3.3	Laboratoire de biotechnologie	3
3.4	Récapitulatif des coûts	3
4	Epreuve finale	3
5	Déroulement d'une campagne	4
5.1	Implantation des essais	4
5.2	Création de matériel	4
5.2.1	Croisements	4
5.2.2	F2 et SSD	4
5.2.3	Haploides doublés	4
5.2.4	Génotypage	5
5.3	Exemple de programme	5
5.3.1	0h00	5
5.3.2	0h30	6
6	Données disponibles initialement	6
7	Références	7
8	Annexe	7

1 Contexte

Ce document fait partie de l’atelier “Prédiction Génomique” organisé et animé par Jacques David et Timothée Flutre en 2016, avec l’aide de Julie Fiévet et Philippe Brabant, à [Montpellier SupAgro](#) dans le cadre de l’option [APIMET](#) (Amélioration des Plantes et Ingénierie végétale Méditerranéennes et Tropicales) couplée à la spécialité SEPMET (Semences Et Plants Méditerranéens Et Tropicaux) du [Master 3A](#) (Agronomie et Agroalimentaire), et de la spécialisation [PIST](#) du [Cursus Ingénieur d’AgroparisTech](#).

Ce document a pour but d’explorer par simulation les stratégies possibles d’un sélectionneur en prenant un cas *très simplifié* de plante annuelle. Il est recommandé d’avoir déjà lu et suivi les documents “Premiers pas” et “Prédiction génomique” de l’atelier.

Concrètement, cela prend la forme d’un “jeu sérieux” (*serious game*) qui implique plusieurs équipes. L’objectif de chacune est de parvenir à réaliser du progrès génétique aboutissant à l’inscription d’une nouvelle variété qui satisfait aux critères usuels de DHS et VATE. L’évaluation des variétés proposées à l’inscription par chaque équipe aura lieu à la fin d’une période de sélection s’étalant sur plusieurs générations.

2 Informations biologiques

2.1 *Apimeta simulans*, une espèce pleine d’avenir

Découverte récemment aux confins de la vallée supérieure de l’Aghromonpe, *Apimeta simulans* appartient au genre des *Statisticaceae*. Elle produit des fleurs qui contiennent un composé alcaloïde, la *sepmetine*, consommée par les étudiants pour éviter les maux de tête lors des efforts intellectuels trop intenses. Le marché est donc très important et se développe rapidement. *Apimeta* peut être sensible à quelques maladies, dont la redoutable rouille fluorescente, *Putrida psychedelica*. Les producteurs sont payés à la quantité produite (les rendements sont de l’ordre de 40 kg de fleurs par ha), mais les transformateurs ont réussi à exiger que la teneur moyenne en sepmetine des lots commerciaux soit au dessus de 14 pour mille.

2.2 Biologie de l’espèce

Apimeta simulans est hermaphrodite et autogame. Elle est produite sous forme de lignées pures, et se croise facilement. On peut produire en serre jusqu’à deux générations par an pour accélérer la fixation et le retour à l’homozygotie. Il est également possible de produire des haploïdes doublés. Le taux de multiplication de l’espèce est très élevé, chaque plante étant capable de produire plus de 1000 graines.

2.3 Génome et ressources génomiques

L’espèce est diploïde. Elle possède 10 chromosomes, tous de même taille. Une carte physique est d’ailleurs disponible. Deux puces ont été construites à partir du séquençage *de novo* de 20 individus: une puce à haute densité de 10 000 marqueurs SNPs, le génotypage coûtant 50 Mendels par individu, et une puce basse densité de 5000 marqueurs dont le génotypage coûte 17 Mendels. Un génotypage de type KASPar peut également être développé pour des SNPs uniques, au prix d’1 Mendel par point.

2.4 Ressources génétiques et données disponibles

Environ 800 lignées sont disponibles. Le lieu d’expérimentation est composé de 300 parcelles. Chaque année pendant dix ans, 150 lignées ont été plantées, en deux répétitions chacune (sans bloc). De plus, chaque lignée a été expérimentée deux années successives. Chaque année, l’essai comporte donc 75 lignées déjà expérimentées l’année précédente, et 75 nouvelles lignées.

Les données relevées sont la production de fleurs (kg/ha), la teneur en sepmatine (g/kg) et la présence de symptômes causés par *Putrida psychedelica*. Pour les 150 lignées expérimentées la dernière année, les données génotypiques sur la puce à haute densité sont déjà disponibles.

3 Moyens expérimentaux

Chaque équipe dispose d'un capital maximum de 10 000 Mendels.

3.1 Réseau d'expérimentation

Il y a un site expérimental possible, Agrom-sur-Lez (AZ), qui a 300 parcelles. L'implantation d'une parcelle nécessite environ 500 graines. Le coût unitaire d'une parcelle expérimentale (semis, phénotypage des trois caractères et récolte), 1 PE, est de 50 Mendels.

3.2 Serre

La serre peut servir à effectuer les croisements et les générations accélérées. Il est possible de faire 2 générations par an. Le coût d'un croisement (quelques semences F1 par croisement) est de 1/2 PE, soit 25 Mendels. La culture et récolte d'une plante individuelle est de 1/25 PE, soit 2 Mendels.

3.3 Laboratoire de biotechnologie

La production d'une graine haploïde fonctionne comme sur le maïs, la pollinisation de la plante donneuse se fait par le pollen d'un génotype inducteur de gynogénèse. Les graines haploïdes se repèrent facilement. Le doublement du stock chromosomique de la plante haploïde prend une génération également. Au final, l'obtention de la récolte de 1000 graines d'un haploïde doublé prend un an et coûte 1/5 PE, soit 10 Mendels.

3.4 Récapitulatif des coûts

Prestation	Coût (Mendels)	Coût (PE)
parcelle	50	1
croisement	25	1/2
autofécondation serre	2	1/25
haploïde doublé serre	10	1/5
génotypage haute densité	50	1
génotypage basse densité	17	1/3
génotypage simple-SNP	1	1/50

4 Epreuve finale

Chaque groupe devra proposer à l'inscription une ou plusieurs variétés à la fin du programme (vendredi midi). Le coût de dépôt à l'inscription est de 200 Mendels par variété.

Elles devront satisfaire aux critères DHS qui seront évalués essentiellement sur le taux d'hétérozygotie de la variété proposée: $\leq 3\%$.

Elles devront aussi satisfaire aux critères VATE correspondant à un minimum de 103 % de la productivité brute moyenne des 5 meilleurs lignées disponibles pour tous en début de programme:

Coll0640 Coll0249 Coll0005 Coll0602 Coll0186. Les variétés en-dessous du seuil de 14 pour mille de sepmetine seront éliminées. Les variétés résistantes auront un bonus.

La disponibilité en semence devra être suffisante. La variété proposée devra donc avoir été expérimentée au moins une fois en parcelle pour disposer de suffisamment de semences à envoyer aux évaluateurs.

5 Déroutement d'une campagne

Une campagne se déroule sur une année, ce qui équivaut à une heure pleine.

5.1 Implantation des essais

Ces implantations ne sont possibles que toutes les heures pleines. Vous devez envoyer aux formateurs un fichier de type "phénotypage". Ce fichier contient la liste des individus que vous voulez phénotyper, et le nombre de parcelles. Voir le format ci-dessous. Vous recevez les résultats à la demi-heure d'après.

5.2 Création de matériel

Attention aux identifiants des individus: c'est vous qui les choisissez, mais ils doivent être uniques (au sein de chaque équipe) et sont définitifs !!

5.2.1 Croisements

Tous les croisements se font en serre. Vous pouvez en commander 2 fois par campagne, en envoyant aux régulateurs un fichier de type "croisement". Il faut autant de lignes que de descendants demandés pour une croisement. Voir le format ci-dessous. Les individus résultant sont disponibles dans la demi-heure d'après.

Par exemple, 62-87.1 est l'identifiant pour un hybride F1 entre les lignées Coll062 et Coll187 de la collection.

5.2.2 F2 et SSD

Ces demandes se font dans le même fichier "croisement". Toutes les autofécondations se font en serre. Pour chacune des générations que vous voulez avancer, envoyer le fichier des identifiants de vos plantes-mères de la génération n-1 et les identifiants que vous souhaitez pour chaque descendant. Il faut autant de lignes que de descendants demandés pour une plante-mère. Voir le format ci-dessous. Les individus résultant sont disponibles dans la demi-heure d'après.

Par exemple, 62-87.1.1, 62-87.1.2 et 62-87.1.3 sont les identifiants de trois plantes F2 issues de l'hybride précédent.

Idem, 62-87.1.3.1 est l'identifiant d'une plante F3 issue de 62-87.1.3.

5.2.3 Haploïdes doublés

Ces demandes se font dans le même fichier "croisement". Voir le format ci-dessous. Les individus résultant sont disponibles dans la demi-heure d'après.

Par exemple, 62-87.1.HD1 et 62-87.1.HD2 sont les identifiant de 2 haploïdes doublés à partir de la plante hybride F1 62-87.1.

5.2.4 Génotypage

Vous pouvez demander le génotypage selon les trois techniques, et vous pouvez soumettre aux heures pleines et aux demi-heures. Vous soumettez un fichier avec les identifiants des individus concernés dans un fichier en précisant la méthode. Pour la méthode “simple marqueur”, il vous faut préciser le code du SNP désiré. Voir le format ci-dessous. Vous recevez vos données dès qu’elles auront pu être générées.

5.3 Exemple de programme

Timothée et Jacques créent une entreprise, la Tim&Jack Breeding Corp (dit *La Firme* dans le milieu). Ils ont eu le temps d’analyser les données, et ont élaboré une stratégie dont voici les premières étapes.

5.3.1 0h00

Ils décident de phénotyper 250 lignées de la collection, non expérimentées la dernière année, 50 sur deux parcelles (par exemple l’individu Coll1342), les autres sur une seule (par exemple l’individu Coll1013).

Leur fichier ressemble à ça:

```
(dat <- data.frame(ind=c("Coll1342", "Coll1013", "..."),
                  task=c("pheno", "pheno", "pheno"),
                  details=c("2", "1", "..."),
                  stringsAsFactors=FALSE))
```

```
##      ind  task details
## 1 Coll1342 pheno      2
## 2 Coll1013 pheno      1
## 3      ... pheno     ...
```

Ils décident aussi de réaliser des croisements:

- 25 croisements correspondant à différentes combinaisons des 30 lignées ayant montré le meilleur classement sur l’ensemble des données disponibles dont 50% sans symptôme de maladie;
- 25 croisements entre une très bonne lignée (top 10) avec les lignées les plus distantes.

Leur fichier ressemble à ça:

```
(dat <- data.frame(parent1=c("Coll1002", "Coll1766", "..."),
                  parent2=c("Coll1013", "Coll1025", "..."),
                  child=c("2-13.1", "766-25.1", "..."),
                  stringsAsFactors=FALSE))
```

```
##  parent1 parent2  child
## 1 Coll1002 Coll1013 2-13.1
## 2 Coll1766 Coll1025 766-25.1
## 3      ...      ...     ...
```

Ils décident enfin de génotyper certains individus, par exemple l’individu Coll1013 sur la puce haute-densité, l’individu Coll1125 sur la puce basse-densité et l’individu Coll1025 pour le SNP 33602.

Leur fichier ressemblent à ça:

```
(dat <- data.frame(ind=c("Coll013", "Coll025", "Coll125", "..."),
  task=c("geno", "geno", "geno", "..."),
  details=c("hd", "snp33602", "ld", "..."),
  stringsAsFactors=FALSE))
```

```
##      ind task details
## 1 Coll013 geno      hd
## 2 Coll025 geno snp33602
## 3 Coll125 geno      ld
## 4      ...   ...     ...
```

5.3.2 0h30

Les phénotypes des essais sont délivrés, dans le même format que les données initiales de la collection:

```
##      ind year plot pathogen trait1 trait2 trait3
## 1 Coll342 2015   1   FALSE  43.98  15.77      0
## 2 Coll342 2015   2   FALSE  39.82  16.53      0
## 3 Coll013 2015   3   FALSE  39.54  16.98      0
## 4      ...   ...   ...     ...     ...     ...
```

Les génotypes issus des puces sont reçus dans le même format que les données initiales, un fichier par puce avec tous les individus ensemble:

```
##      snp07454 snp14386 snp21483
## Coll013      1       0       2
## ...      ...     ...     ...
```

Pour les simples marqueurs, le format est le suivant:

```
##      ind      snp geno
## 1 Coll025 snp33602    1
## 2      ...     ...   ...
```

Et le reste de leur stratégie, et bien, cela ne vous regarde pas... ;)

6 Données disponibles initialement

A votre arrivée dans le programme, nous vous donnons les fichiers contenant les phénotypes de quelques essais et les génotypes de plusieurs variétés.

Voici les phénotypes:

```
phenos <- read.table(file="phenos_coll.txt.gz", header=TRUE, stringsAsFactors=FALSE)
class(phenos)
dim(phenos)
str(phenos)
head(phenos)
summary(phenos)
```

Et voici les génotypes:

```
genos <- as.matrix(read.table(file="genos_coll-2014.txt.gz", header=TRUE))
class(genos)
dim(genos)
str(genos)
genos[1:4,1:6]
summary(genos)
```

7 Références

- Massman, J. M., Jung, H.-J. G., Bernardo, R., 2013. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Science* 53 (1), 58+. [URL](#)
- Riedelsheimer, C., Melchinger, A. E., Nov. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theoretical and Applied Genetics* 126 (11), 2835-2848. [URL](#)
- Longin, C. F., Mi, X., Würschum, T., 2015. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theoretical and Applied Genetics* 128 (7), 1297-1306. [URL](#)
- Duvick, D. N., Smith, J. S. C. & Cooper, M. Long-Term selection in a commercial hybrid maize breeding program. *Plant Breeding Reviews* 24, 109-151 (2004). [URL](#)
- Mackay, I. et al. Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. *Theoretical and Applied Genetics* 122, 225-238 (2011). [URL](#)
- Laidig, F., Piepho, H.-P., Drobek, T. & Meyer, U. Genetic and non-genetic long-term trends of 12 different crops in german official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics* 127, 2599-2617 (2014). [URL](#)
- Moreau, L., Lemarié, S., Charcosset, A. & Gallais, A. Economic efficiency of one cycle of Marker-Assisted selection. *Crop Science* 40, 329+ (2000). [URL](#)
- Oury, F.-X., Godin, C., 2007. Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? *Euphytica* 157 (1-2), 45-57. [URL](#)

8 Annexe

```
print(sessionInfo(), locale=FALSE)
```

```
## R version 3.2.2 (2015-08-14)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.4 LTS
##
## attached base packages:
## [1] stats      graphics  grDevices utils      datasets  methods   base
##
```

```
## other attached packages:
## [1] knitr_1.12.3    rmarkdown_0.9.2
##
## loaded via a namespace (and not attached):
## [1] magrittr_1.5    formatR_1.2.1  tools_3.2.2    htmltools_0.3  yaml_2.1.13
## [6] stringi_1.0-1  stringr_1.0.0  digest_0.6.9   evaluate_0.8
```