**Contrôle continu de**

**« Génomique et génétique statistiques » (EM7BBSBM)**

**9 novembre 2011**

**Problème 1: Génétique des populations**

Le pois est une plante diploïde annuelle utilisée pour l’alimentation du bétail. Une exploitation agricole compte sur une production importante et la moins coûteuse possible de cette plante. Un phénotype de pois sauvage **[s]** présente une viabilité à l’âge adulte deux fois plus importante que le phénotype de pois habituellement cultivé **[c]**, en condition de réduction d’apport hydrique. L’exploitation décide d’effectuer un croisement en serre des deux phénotypes (un parent **[s]** x un parent **[c]**), puis de semer les graines en parcelles expérimentales afin de suivre l’évolution de la descendance.

**Question 1** (2 points) :

Dans la parcelle expérimentale 1 qui est irriguée en conditions normales, on observe à l’âge adulte 50% de phénotype sauvage [s] et 50% de phénotype [c] dans la descendance F1 du croisement. Qu’en déduisez-vous sur les bases génétiques du caractère de résistance au stress hydrique: Dominance ou non ? Monogénique ou polygénique ? Quel était le génotype des parents au(x) locus ? Justifiez votre réponse.

**Question 2**(3 points) :

Dans la parcelle expérimentale 1, on laisse la descendance du croisement initial arriver à floraison, et la reproduction se faire. Entre temps, une analyse de cartographie a permis de localiser le locus et il est maintenant possible de génotyper les individus pour ce locus. L’année suivante sur 100 plantes (F2) issues de ce deuxième croisement, on observe 10 ss, 30 sc, 60 cc.

* La population est-elle panmictique ? Utilisez le test du χ2 sur les effectifs (la valeur du χ2 théorique au seuil α = 0.05, pour 1 degré de liberté, est de 3.84) et concluez au seuil de 5%. On donne χ2obs = Σ ( (Nobs – Nexp)2/ N\_exp) (Σ = somme sur les génotypes)
* Le génotypage confirme par ailleurs qu’il y avait 50% d’hétérozygotes dans la F1. Quel est selon vous le mode principal de reproduction du pois en conditions « naturelles » (ici en parcelle) :

A) Autofécondation

B) Allofécondation (panmixie)

C) Autofécondation plus une petite proportion d’allofécondation (panmixie)?

D) Allofécondation (panmixie) plus une petite proportion significative d’autofécondation ?

**Question 3** (6 points) :

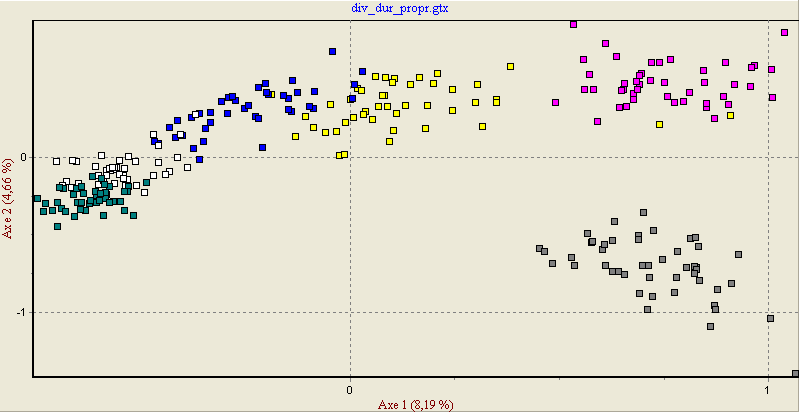
Dans la parcelle expérimentale 2, 100 graines de F2 sont semées et on laisse la population évoluer en condition de réduction d’apport hydrique. On admet que la population va se reproduire en allofécondation complète (panmixie).

* Donner la valeur adaptative de chaque génotype dans cet environnement.
* Quelle sera la structure génotypique, et phénotypique à la génération suivante (F3) à l’âge adulte, après une génération de sélection?
* Quelle sera la fréquence des allèles **s** et **c** à l’équilibre, et concluez. On donne la variation de fréquence Δp: Δp = pq [(w1-w2)p + (w2 –w3)q] / wbar.

**Problème 2: Génétique des populations, populations structurées**

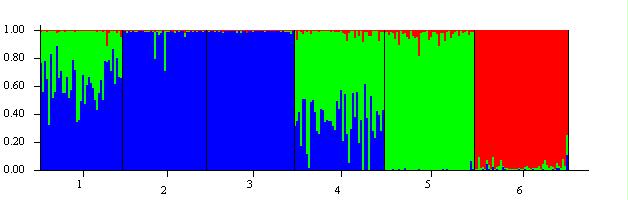
On veut étudier la diversité génétique de 6 populations d’une même espèce, géographiquement proches. Pour cela, on dispose de données de génotypage sur une vingtaine d’individus par population environ, typés sur 38 marqueurs multialléliques.

On effectue une Analyse en Composantes Principales sur le tableau de distance IBS (nombre d’allèles partagés), c’est-à-dire un Positionnement Multidimensionnel, entre tous les individus. Les résultats sur le premier plan principal sont ceux-ci, avec une couleur par race :



Les individus de la Pop1 sont en bleu, Pop2 en blanc, Pop3 en vert, Pop4 en jaune, Pop5 en rose, et Pop6 en gris. On nous dit que Pop2 et Pop3 ont des caractéristiques phénotypiques quasiment identiques, et que les Pop1 et Pop4 sont issues de croisements entre Pop2/Pop3 d’une part, et Pop5 d’autre part.

On effectue ensuite une analyse à l’aide du logiciel Structure, avec le modèle « admixture » et fréquences alléliques corrélées. On lance, pour K variant de 2 à 6, dix chaines MCMC indépendantes. Chaque chaine est de longueur 50 000, pour une période de « chauffe » de 10 000. Le K optimal (graphique non donné) semble être 3. Dans le graphique suivant, pour K=3, les individus sont classés par population de départ, de 1 à 6 (de Pop1 à Pop6). Les 10 chaines fournissent toutes la sortie suivante:



**Question 1** (1.5 point). Quelles interprétations tirez-vous de l’ACP ?

**Question 2** (0.5 points). Que pensez-vous de la convergence des chaines pour K=3 en sortie du logiciel Structure, et pourquoi ?

**Question 3** (1.5 points). Quelle interprétation tirez-vous de la sortie pour K=3 ?

**Question 4** (0.5 point). Quelles conclusions tirez-vous de ces analyses ? Quel regard critique y portez-vous ?