

Genetic characterization of 16 barley accessions (*Hordeum vulgare* L.) using SSR markers

Caractérisation génétique d'une collection de 16 accessions d'orge (*Hordeum vulgare* L.) à l'aide des marqueurs moléculaires de type SSR

SALEM MARZOUGUI^{1,2*}

¹Pôle Régional de Recherche Développement Agricoles du Nord-Ouest semi-aride à El Kef, Institution de la Recherche et de l'Enseignement Supérieur Agricoles (IRESA), Tunisie

²Laboratoire des grandes cultures, INRAT, Tunisie

*Corresponding author: salem.marzougui@iresa.agrinet.tn

Abstract – In barley breeding programs, information on genetic dissimilarity and population structure is very important for the conservation of genetic diversity and therefore the development of new cultivars. This study aimed to assess the genetic variation of Tunisian barley (*Hordeum vulgare* L.) accessions using SSR-type molecular markers. A total of 30 alleles were detected in 11 SSR markers. The number of alleles per locus ranged from 2 to 4 with an average of 2.7 alleles per locus detected from the 16 barley accessions and the average polymorphic information content (PIC) value was 0.54. Analysis of the population structure with the STRUCTURE software identified 4 subpopulations with allele frequencies ranging from 0.0119 to 0.0597 and a fixation index (Fst) ranging from 0.29 to 0.43. The higher the allelic frequency, the more there is a differentiation between individuals of the same subpopulation. Analysis of the genetic diversity of the barley collection will facilitate cultivar development and efficient use of genetic resources.

Keywords: Barley, Genetic diversity, Molecular Markers, Alleles, Fixation Index

Resumé - Dans les programmes de sélection de l'orge, les informations sur la dissimilarité génétique et la structure de la population sont très importantes pour la conservation de la diversité génétique et donc le développement de nouveaux cultivars. Cette étude visait à évaluer la variation génétique des accessions d'orge tunisienne (*Hordeum vulgare* L.) en utilisant des marqueurs moléculaires de type SSR. Un total de 30 allèles a été détecté dans 11 marqueurs SSR. Le nombre d'allèles par locus variait de 2 à 4 avec une moyenne de 2.7 allèles par locus détectés à partir des 16 accessions d'orge et la valeur moyenne du contenu d'information polymorphe (PIC) était de 0,54. L'analyse de la structure de la population avec le logiciel STRUCTURE a identifié 4 sous population ayant des fréquences alléliques allant de 0.0119 à 0.0597 et un indice de fixation (Fst) variant de 0.29 à 0.43. Plus, la fréquence allélique est importante, plus il y a une différenciation entre les individus de la même sous population. L'analyse de la diversité génétique de la récolte d'orge facilitera le développement des cultivars et l'utilisation efficace des ressources génétiques.

Mots clés: Orge, Diversité génétique, Marqueurs moléculaires, Allèles, Indice de fixation

1. Introduction

L'orge (*Hordeum vulgare* L.), l'une des premières et des plus anciennes cultures domestiquées par l'homme, est une céréale majeure cultivée dans les climats tempérés du monde entier. C'est l'une des cultures les plus anciennes au monde et se classe quatrième après le blé, le riz et le maïs (Poets et al. 2015). En Tunisie, l'orge est principalement cultivée en climat aride et semi-aride dans des zones de moins de 400 mm de précipitations annuelles. Dans les pays méditerranéens moins développés comme la Tunisie, l'orge joue un rôle clé car ses céréales et ses pailles sont le principal aliment du bétail. Les petits ruminants tels que les ovins et les caprins sont le principal bétail en Tunisie et représentent une contribution alimentaire précieuse dans les zones rurales ainsi que la principale production économique (Sana et al. 2012). L'amélioration génétique pour répondre à l'augmentation du rendement est en cours dans le programme de sélection Tunisien. Le développement de variétés à haut rendement adaptées aux



conditions locales dépend de la compréhension de la variabilité existante et de la relation génétique entre les accessions d'orge. Par conséquent, l'évaluation de la diversité génétique des lignées d'orge à l'aide de marqueurs moléculaires est importante dans la sélection d'orge pour une exploration réussie, une stabilité génétique et une conservation efficace, car les caractères morphologiques sont limités en nombre et instables (Sahranavard et al. 2015). Plusieurs types de marqueurs moléculaires tels que le polymorphisme de longueur de fragment amplifié (AFLP), l'ADN polymorphe amplifié au hasard (RAPD), les répétitions de séquences simples (SSR), la technologie de matrice de diversité (DArT) et le polymorphisme nucléotidique unique (SNP) ont été utilisés pour étudier la diversité génétique et la structure des populations (Esther et al. 2019 ; Ernest et al. 2020). La diversité génétique a été estimée par la richesse allélique, c'est-à-dire nombre d'allèles détectés, la diversité génétique (Nei 1973) et l'occurrence d'allèles uniques et hétérogènes. Les marqueurs SSR ont été largement utilisés dans l'analyse de la diversité génétique en raison de leur disponibilité, hautement informatifs et leur distribution dans tout le génome (Varshney et al. 2005). L'objectif principal de cette étude est d'analyser la diversité génétique existant parmi les 16 accessions d'orge ancienne Tunisiennes à l'aide de 11 marqueurs moléculaires. Cette étude facilitera efficacement le développement de cultivars et l'utilisation des ressources génétiques.

2. Matériels et Méthodes

2.1 Matériels végétales et extraction d'ADN

L'analyse de la diversité génétique a été réalisée sur 16 accessions d'orge ancienne Tunisienne issu de la banque des gènes international (U.S. National Plant Germplasm System) (Tableau 1). La collection consiste en 11 accessions de race locale (Landrace), une variété et quatre sont incertain. Huit graines de chaque accession ont germé et les feuilles ont été récoltées au stade de trois feuilles après 15 jours de plantation. L'ADN génomique a été extrait selon le kit d'extraction GRS Genomic DNA (Grisp, Portugal). La qualité et la quantité d'ADN ont été déterminées en utilisant un spectrophotomètre UV-Vis et une comparaison visuelle de l'électrophorèse sur gel d'agarose à 2%.

Tableau 1. Liste des accessions d'orge génotypés par les 11 marqueurs SSR

	Nom de l'accession	Origine	Statut
G1	Martin	Tunisienne	Cultivar
G2	4a	Tunisienne	Incertain
G3	1356-33	Tunisienne	Landrace
G4	3380-35	Tunisienne	Landrace
G5	3362-81	Tunisienne	Landrace
G6	Arbi	Tunisienne	Incertain
G7	Staf	Tunisienne	Landrace
G8	Ariana	Tunisienne	Landrace
T1	Revil No.1	Tunisienne	Incertain
T2	1144-87	Tunisienne	Landrace
T3	3452-114	Tunisienne	Landrace
T4	Arbi	Tunisienne	Incertain
T5	Jbali	Tunisienne	Landrace
T6	Dinar	Tunisienne	Landrace
T7	Besert 13	Tunisienne	Landrace
T8	186	Tunisienne	Landrace

2.2 Marqueurs moléculaires et amplification

Les microsatellites ou SSR sont des marqueurs moléculaires basés sur la PCR (Temnykh, S. et al. 2001). Les microsatellites sont constitués de séquences de di-, tri- ou tétranucléotides répétés en tandem (toujours dans le même sens, par opposition à répétitions inversées répétées). Ces éléments sont uniformément répartis en plusieurs exemplaires sur l'ensemble du génome d'une espèce et présentent un taux de polymorphisme élevé. Ce polymorphisme repose sur la variation du nombre d'unités de répétition constituant le microsatellite. Dans cette étude on a utilisé 11 marqueurs de type SSR qui sont distribués tout au long des 7 chromosomes de l'orge (Tableau 2) (Figure 1).

Tableau 2. Les 11 marqueurs moléculaires utilisés pour la caractérisation génétique de la collection d'orge

Marquers	Nombre d'allèles	HObs	HExp	PIC
EBmac0701	2	0	0.315	0.258
EBmac415	2	0	0.226	0.195
Bmag120	3	0	0.605	0.52
Bmag0496	4	0.063	0.607	0.544
Bmac0206	2	0.063	0.466	0.349
Bmac0134	3	0.563	0.647	0.551
EBmac0708	3	0.688	0.546	0.419
Bmag382	3	0.438	0.567	0.452
Bmac0213	3	0.438	0.454	0.393
Bmag0125	3	0.125	0.387	0.305
Scssr07759	3	0.188	0.454	0.393
Moyenne	2.72727273	0.2332727	0.479455	0.398091

Hobs : Hétérozygotie observé ; HExp : Hétérozygotie attendue ; PIC : Contenu en information de polymorphisme

L'amplification par PCR a été réalisée dans un volume total de 10 pl consistant en 6 pl de GRS Hotstart Taq Mastermix (Grisp, Portugal), 0,25 pl de chaque marqueur SSR (10 uM) et 1 pl d'ADN (50 ng). Le produit PCR a ensuite été analysé sur gel d'agarose à 2%. L'amplification de l'ADN a été réalisée dans FastGene Ultra Cycler (96 puits) (Nippon Genetics, Allemagne). La PCR a ensuite été soumise aux conditions suivantes: Dénaturation initiale à 95 °C pendant 5 minutes suivie de 35 cycles de dénaturation à 95 °C pendant 30 secondes, hybridation de 52 à 62 °C pendant 30 secondes et extension finale à 72 °C pendant 30 secondes.

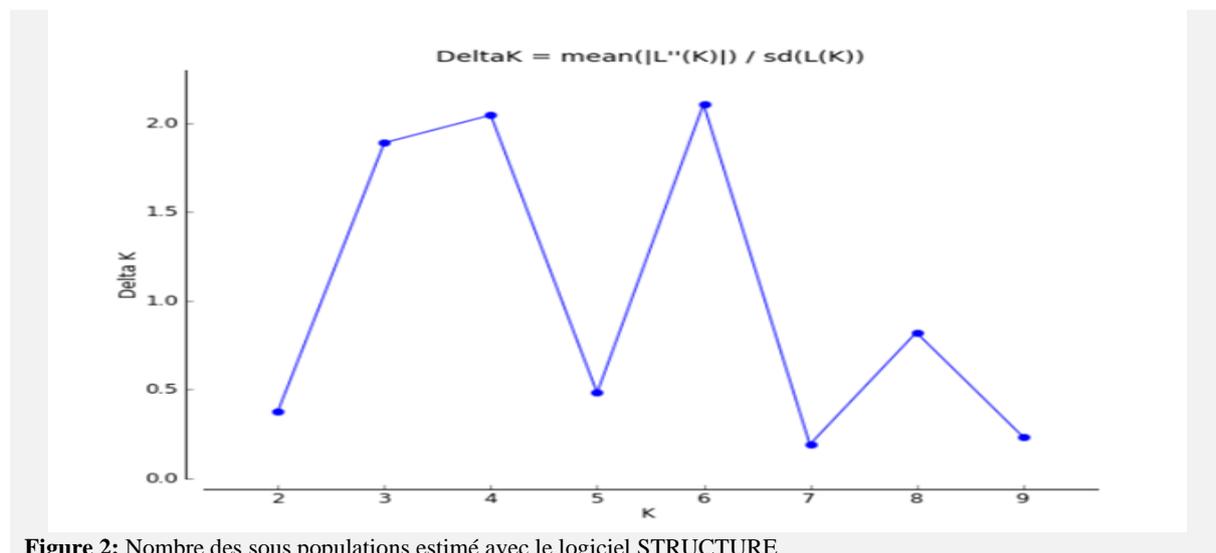


Figure 2: Nombre des sous-populations estimé avec le logiciel STRUCTURE

3.3 Outils d'analyse de la diversité génétique

Le nombre d'allèles, l'hétérozygotie observée (Hobs), l'hétérozygotie attendue (HExp) et le contenu de l'information polymorphe des locus (PIC) ont été déterminés à l'aide du logiciel Cervus version 3.0.7 (Kalinowski et al. 2007). L'analyse des relations entre les accessions sur la base des données des marqueurs SSR a été réalisée avec des méthodes de WARD utilisant DARWin 6.0 (Perrier et Jacquemond-Collet 2006). Les résultats du génotypage des marqueurs SSR ont été utilisés pour estimer la structure de la population des 16 accessions d'orge à l'aide du logiciel Structure. La distribution des valeurs ΔK a été déterminée en évaluant la vraisemblance logarithmique $[L(K)]$ (Evanno et al. 2005). Pour déterminer la structure de la population des accessions étudiées, les données de génotypage ont été traitées avec le logiciel STRUCTURE 2.3.4, qui met en œuvre une analyse de cluster bayésienne basée sur un modèle (Pritchard et al., 2000). Un nombre putatif de sous-populations allant de $k = 1$ à 10 a été évalué en utilisant 50 000 itérations de rodage suivies de 50 000 itérations de chaîne de Markov enregistrées. Pour estimer la variance d'échantillonnage de la structure de la population inférée, 10 essais indépendants ont été effectués pour chaque k . Le nombre réel de sous-populations a été déterminé par le logarithme de similarité pour chaque K ; $\ln P(D) = L(K)$ et la valeur optimale ΔK a été obtenue par $\Delta K = [L''(K)] / \text{Stdev}$ selon Evanno (2005) pour déterminer le nombre le plus probable de groupes. Sur la base des sous-populations inférées par l'analyse structurale, nous avons effectué une analyse de la variance moléculaire (AMOVA) pour évaluer la différenciation de la population en utilisant GenAlEx version 6.5 (Peakall et Smouse 2012) avec 999 fois le boost-strapping.

3 Résultats et discussion

3.1 Analyse de la diversité allélique

Cette étude a été réalisée sur 16 accessions d'orge tunisien par 11 marqueur moléculaire de type SSR distribuer sur les 7 chromosomes d'orge. Le tableau 2 résume le nombre d'allèles, l'indice de la diversité génétique et le PIC, calculés pour les 11 amorces SSR à l'aide du logiciel Cervus. La richesse allélique d'une population, définie comme le nombre d'allèles présents à un locus donné, est connue pour dépendre de la taille de l'échantillon (Foulley et al. 2006).

Le nombre moyenne d'allèles (A) traduit la richesse en allèles d'une population il est calculé selon la formule suivante:

$$A = \frac{1}{l} \sum a$$

où "a" représente le nombre d'allèles à un locus et l le nombre de loci étudié. Dans ce travail le nombre d'allèles totaux est 30, la distribution allélique varie de 2 loci pour EBmac0701, EBmac415, et Bmag0206 à 4 loci pour Bmag0496 avec une valeur moyenne de 2,72 d'allèle par locus.

3.2. Analyse de Contenu en informations de polymorphisme (PIC)

La Pic a varié de 0,195 pour EBmac415 à 0,544 pour Bmag0496, tous les loci sont polymorphes avec une valeur moyenne de 0,398. Parmi ces marqueurs, les Bmag120, Bmag0496, Bmac0134 et Bmag382 sont les plus informatifs alors que l'EBmac415 a le pouvoir discriminant la plus faible.

Ces valeurs indiquent le pouvoir des marqueurs SSR utilisés pour analyser la diversité génétique de notre collection d'orge.

3.3. Analyse d'hétérozygotie observée et d'hétérozygotie attendu

Le taux d'hétérozygotie observé (HO) est la proportion d'individus hétérozygotes au locus K comme dans la formule:

$$H_{OK} = \sum_{i,j=1}^{a_k} P_{ij} (i \neq j)$$

où Pij est l'estimation de la fréquence du génotype ij au locus k et a le nombre d'allèles au locus k.

Le taux d'hétérozygotie attendu (He) peut être calculé, sous l'hypothèse d'équilibre de Hardy-Weinberg, à partir des fréquences alléliques déterminées pour chaque locus à l'aide de la formule:

$$H_E = 1 - \sum p_i^2$$

où pi est la fréquence du ième allèle à ce locus.

Afin d'estimer l'importance de polymorphisme génétique, les taux d'hétérozygotie observés et attendus ont été calculés pour chaque locus sous l'hypothèse d'équilibre de Hardy-Weinberg (Tableau 2) et sont comparés. Parmi les loci polymorphes le locus Bmac0134 présente le taux d'hétérozygotie attendu (He) le plus élevé 0,647 et le locus EBmac0708 présente le taux d'hétérozygotie observé (Ho) le plus élevé 0,688. Les moyennes de He et Ho pour les 16 accessions sont respectivement 0,479 et 0,233. L'hétérozygotie attendue est supérieure à celle observée, traduisant un écart positif suggérant un déficit d'hétérozygotie dans la population. Cela veut dire qu'il existe une similitude génétique entre les individus testés.

3.4 Analyse de la structure de la population

La structure génétique des populations de l'orge, a été réalisée à l'aide du logiciel *STRUCTURE* 2.3.4. Les résultats obtenus sont traités à l'aide de site web *STRUCTURE HARVESTER*, un programme python qui analyse et récapitule les données de *STRUCTURE* et calcule la valeur de vraisemblance (ΔK). Le nombre de sous population (K), est déterminé à partir une estimation des probabilités postérieurs des données obtenus Pr (X | K) (Pritchard et al. 2000), cette valeur est appelée Ln P (D) ou aussi L(K). K prend la valeur maximale de L (K). L'analyse des données montre que K prend la valeur 4 (Figure 2).

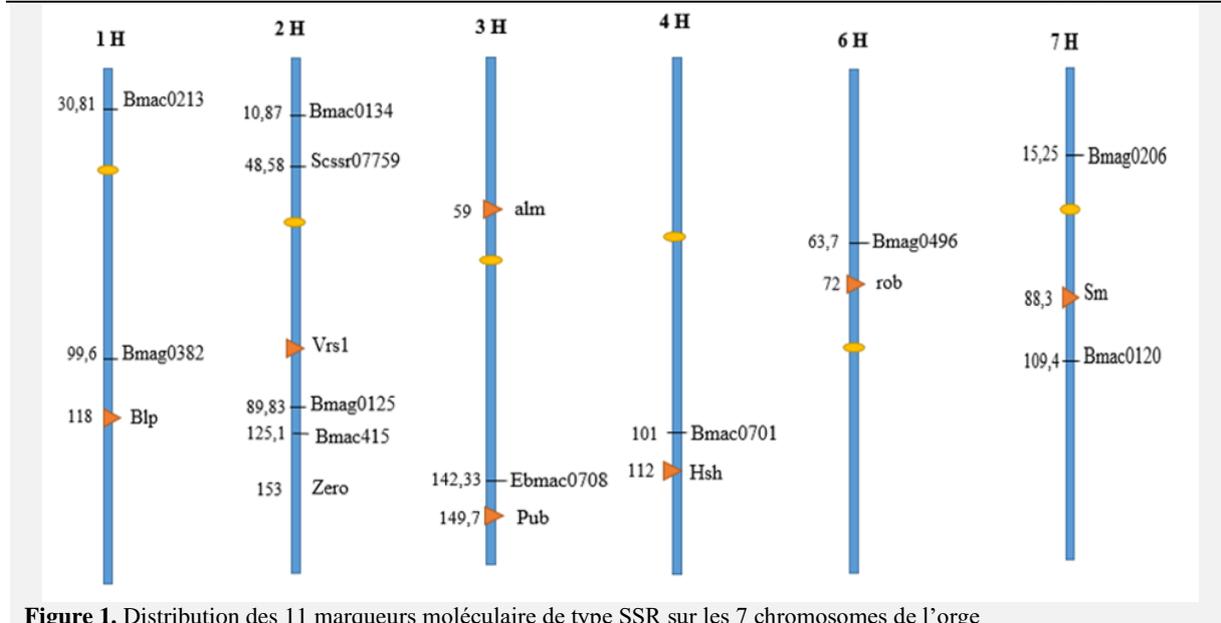


Figure 1. Distribution des 11 marqueurs moléculaires de type SSR sur les 7 chromosomes de l'orge

Analyse de la fréquence allélique de la population:

Les valeurs de fréquence allélique (Tableau 3) varient entre 0,0119 et 0,0597. La plus petite fréquence est observée entre les populations 1 et 4, cela suggère que la population 1 et la population 4 sont très proches l'une de l'autre autrement dit pas d'accumulation génétique et donc une similitude entre les pools génétiques. La valeur la plus élevée est observée entre les populations 2 et 3 (0,0597), cela signifie qu'il y a une divergence moyenne entre les deux populations, d'où une légère différence entre les deux pools génétiques des populations.

Tableau 3. Fréquences alléliques générés par STRUCTURE.

	1	2	3	4
1	-	0,0170	0,0320	0,0119
2	0,0170	-	0,0597	0,0516
3	0,0320	0,0597	-	0,0166
4	0,0119	0,0516	0,0166	-

Analyse de l'indice de diversification (Fst)

L'indice de diversification génétique F_{st} informe sur le taux de différenciation entre les populations, autrement dit l'absence ou la présence d'un flux de gènes entre les différentes populations:

$$F_{ST} = (HT - HS) / HT$$

Si la F_{st} prend une valeur proche de 0 cela signifie qu'il n'y a pas une différenciation génétique dans ce cas on parle d'une population panmictique. Dans le cas contraire, si la F_{st} est inférieure ou égale à 1, cela signifie qu'on a une forte différenciation entre les populations. En observant la moyenne des F_{st} pour chaque population (Tableau 4), on remarque que les 4 populations entre elles ont un indice F_{st} supérieur à 0,25. L'indice F_{st} est pratiquement significatif pour les 4 populations, la valeur la plus élevée est observée chez la deuxième population 0,4378. Cette valeur indique une différenciation très importante entre les accessions de la deuxième population, et un flux des gènes peu important. La valeur la plus petite est observée chez la quatrième population, elle est de 0,2991, suggérant ainsi une différenciation importante au sein de la quatrième population ($0,2991 > 0,25$).

Tableau 4. Indices de diversification (F_{st}) par population

	1	2	3	4
F_{st}	0,3538	0,4378	0,3247	0,2991

4. Conclusion

L'analyse de la diversité génétique existant parmi les 16 accessions d'orge ancienne Tunisiennes à l'aide de 11 marqueurs moléculaires nous a permis d'identifier 4 sous-populations ayant des fréquences alléliques allant de 0,0119 à 0,0597 et un indice de fixation (F_{st}) variant de 0,29 à 0,43. La différenciation entre les individus de la même sous-population est importante puisque la fréquence allélique est importante. L'analyse de la diversité génétique de la collection d'orge facilitera le développement des cultivars et l'utilisation efficace des ressources génétiques.

5. Références

- Nei M (1973)** Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70: 3321-23.
- Poets AM, Fang Z, Clegg MT, Morrell PL (2015)** Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biology 16: 173
- Sana M, Mouldi E, Mohamed E (2012)** Barley breeding for quality improvement in Tunisia. AFRICAN JOURNAL OF BIOTECHNOLOGY 11(89): 15516-15522
- Sahranavard AFR, Darvishzadel M, Ghadimzadel H, Azizi Z Aboulghasemi (2015)** Identification of SSR loci related to some important agro morphological traits in different oil sunflower (*Helianthus annuus* L.) lines using association mapping. J Plant Gene Res 2: 15-32
- Ernest CB, Salem M, Esther M, Choi W, Lee MC (2020)** Genetic diversity analysis and population structure of some African and Asian Finger Millet (*Eleusine coracana* L.) accessions using Expressed Sequence Tags -Simple Sequence Repeat (EST- SSR) markers. International Journal of Environment, Agriculture and Biotechnology 5(3): 643-646
- Esther W Mwangi, Salem Marzougui, Jung Suk Sung, Ernest C.Bwalya, Yu-Mi Choi, Myung-Chul Lee. (2019)** Assessment of Genetic Diversity and Population Structure on Kenyan Sunflower (*Helianthus annuus* L.) breeding Lines by SSR Markers. Korean J Plant Res 32(3): 244-253
- Varshney RK, Graner A, Sorrells ME (2005)** Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23: 48-55
- Temnykh, S, Declerk, G, Lukashova A, Lipovich L, Cartinhour S, Mc Couch S (2001)** Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon association and genetic marker potential. Genome Research 11: 1441-1452.
- Evanno G, Regnaut S, Goudet J (2005)** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14: 2611-2620
- Perrier X, Jacquamoud-collect JP (2006)** DARWin software <http://Darwin.cirad.fr/>.
- Kalinowski ST, Taper ML, Marshall TC (2007)** Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16(5): 1099–1106
- Peakall R, Smouse PE (2012)** GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-An update. Bioinformatics 28: 2537-2539
- Foulley JL, van Schriek MGM, Alderson L et al (2006)** Genetic diversity analysis using lowly polymorphic dominant markers: the example of AFLP in pigs. Journal of Heredity 97: 244–252.
- Pritchard JK, Stephens M, Donnelly P (2000)** Inference of population structure using multi locus genotype data. Genetics 155: 945-959