

**Name:** Dr. Ahmed El Bakkali

**E-mail address:** ahmed\_elbakkali@yahoo.fr

**Host institute address:** Centre Régional de la Recherche Agronomique de Meknès (CRRRA-Meknès)  
BP 578, VN 50000, km 13 Route Haj Kaddour, Meknès, Morocco  
Tel: + 212 5 35 30 03 66, Fax: + 212 5 35 30 02 44

**Home institute address:** Institut National de la Recherche Agronomique (INRA)  
Avenue Ennasr, BP 415 RP, Rabat, Morocco  
Tél: +212 0537 77 26 54, Fax: +212 0537 77 00 49

---

## **Etude de la diversité génétique et construction de core collections en vue de la génétique d'association chez l'olivier (*Olea europaea* L.)**

Le déterminisme génétique des caractères d'intérêt agronomique et adaptatif peut être abordé par l'étude de populations en ségrégation mais aussi par l'étude de génotypes non-apparentés par la génétique d'association. Basée sur l'étendue du déséquilibre de liaison (DL), cette approche est prometteuse pour l'utilisation dans la génétique d'adaptation et l'amélioration des plantes. Cependant, quel matériel à étudier est une question centrale, en particulier pour des espèces pérennes telle que l'olivier méditerranéen (*Olea europaea* L.). Malgré l'importance socio-économique et la grande diversité de cette espèce, les études du déterminisme génétique des caractères adaptatifs et agronomiques demeurent toujours au stade préliminaire. En étudiant la diversité et la structure génétiques de l'olivier cultivé, cette thèse a pour objectif principal de définir les dispositifs génétiques adéquats pour mener des travaux à long-terme sur la génétique d'association gène/caractère chez l'olivier.

Nous avons examiné la diversité génétique chez l'olivier cultivé selon deux échelles géographiques : (i) méditerranéenne avec 557 accessions issues de 14 pays en collection mondiale à Marrakech (Maroc) ; et (ii) locale avec 121 accessions françaises en collection à Porquerolles (France). L'analyse à l'aide de 17 marqueurs microsatellites nucléaires, répartis sur le génome, et 39 marqueurs chloroplastiques montre une diversité génétique élevée dans les deux collections, soulignant l'importance de ces ressources génétiques pour des études d'association gènes/caractères. La collection méditerranéenne est structurée en trois *pools* génétiques reliés à la provenance des variétés et aux processus de domestication et de diversification de l'olivier : est, centre, et ouest de la Méditerranée. En revanche, les variétés françaises sont hautement admixées par les trois *pools* génétiques. Nous avons ensuite testé différentes méthodes et tailles d'échantillonnage en vue de construire des *core collections* de taille pertinente et adaptées à la génétique d'association. Nous proposons une nouvelle méthodologie d'échantillonnage en deux étapes permettant un compromis entre les différents paramètres génétiques examinés. La première étape consiste à construire une *core collection* primaire capturant le maximum de diversité avec une liste de génotypes définitive et incontournable pour une bonne représentativité des ressources génétiques. Dans la deuxième étape, la méthodologie permet d'avoir une flexibilité dans le choix des génotypes en vue de compléter la diversité génétique manquante. Les *core collection* finales (méditerranéenne  $CC_{94}$  et locale française  $LCC_{49}$ ) permettent un choix rationnel du matériel à étudier tout en assurant une représentativité de la diversité et de la structure génétique révélées dans les deux collections, mondiale et française. Les résultats d'analyse de  $CC_{94}$  et  $LCC_{49}$  montrent que la *core collection* méditerranéenne est structurée en deux *pools* génétiques : est et ouest/centre, tandis que celle locale, les génotypes montrent une forte *admixture*. Selon le coefficient d'apparentement, la majorité des génotypes des deux *core collections* proposées sont peu ou non-apparentés. En outre, d'après le paramètre multilocus  $\bar{r}_d$ , plus de DL est observé dans la *core collection* locale  $LCC_{49}$  que dans celle méditerranéenne  $CC_{94}$  qui vraisemblablement le résultat d'un

processus d'*admixture* récente chez le germoplasme français. En termes des analyses, deux dispositifs génétiques complémentaires sont proposées pour l'étude des caractères chez l'olivier : (i) génétique d'association en utilisant la *core collection* méditerranéenne ; et (ii) cartographie du mélange génétique, *admixture mapping*, en utilisant la *core collection* locale française.

La mise en place du matériel proposé et son évaluation selon des dispositifs expérimentaux randomisés en plusieurs sites, afin de prendre en compte les interactions génotype x environnement, permettra de décomposer les bases génétiques contrôlant les caractères d'intérêt pour une application dans des programmes futurs de sélection génomique.

**Mots Clés** : Olivier (*Olea europaea* L.), diversité génétique, structure génétique, *core collection*, génétique d'association, cartographie du mélange génétique

## Study of genetic diversity and construction of core collections for genetic association in olive (*Olea europaea* L.)

Genetic determinism of traits of agronomic and adaptative interest can be greatly helped by the application of linkage mapping using crossing populations, but also by using unrelated genotypes with association mapping approach. Based on linkage disequilibrium (LD), this approach is promising for adaptation genetic and breeding in plants. However, which material to study is a central question in occurrence for perennial fruit species such as Mediterranean olive tree (*Olea europaea* L.). Despite the socio-economical and the broad diversity of the species, studies on genetic determinism of traits of interest are preliminary. By studying genetic diversity and structure in cultivated olive, this thesis has a main objective to define suitable genetic designs to conduct gene/trait association studies for olive species.

We examined genetic diversity for cultivated olive at two geographical scales: (1) Mediterranean with 557 accessions belonging to 14 countries in the *ex situ* Olive Worldwide Germplasm Bank, Marrakech (Morocco, OWGB Marrakech); and (2) local with 121 French accessions in the *ex situ* French Olive Germplasm Bank, Porquerolles island (France, FOGB Porquerolles). Analysis using 17 nuclear microsatellites markers, spanning the genome, and 39 plastid loci reveals high genetic diversity in both collections, indicating therefore the importance of such genetic resources for genes/traits association studies. The Mediterranean collection showed to be structured into three gene pools, which strongly matched geographic areas of varieties and also domestication and diversification process of the species: east, centre, and west of Mediterranean. However, French varieties showed high admixture level between the three gene pools. We then tested different sampling methods and sizes in order to construct core collections suitable for association mapping with optimal size. We propose a new two-step based sampling methodology allowing a trade-off between different genetic parameters examined. The first step by establishing a primary core collection capturing the maximum diversity with a final and essential genotypes list for high representativeness of genetic resources. In the second step, the proposed methodology allows flexibility in the choice of material in order to capture remaining diversity. Final core collections (Mediterranean  $CC_{94}$  and local  $LCC_{49}$ ) enable experimental flexibility and rational choice of varieties to be studied, and reflect the genetic diversity and structure revealed in both original collections. Analysing both core collections,  $CC_{94}$  and  $LCC_{49}$ , reveals that the Mediterranean one is structured into two gene pools: eastern and western/central, whereas for the French local one, high admixture level was noted within genotypes. According to kinship coefficient, most genotypes in both core collections were revealed to be unrelated. More LD was revealed in local core collection  $LCC_{49}$  than in Mediterranean one  $CC_{94}$  according to the multilocus  $\bar{r}_d$  measure which likely resulted of recent admixture processes in French germplasm. In terms of analysis, two complementary genetic designs were proposed to link traits to genes/markers: (i) association mapping using Mediterranean core collection; and (ii) admixture mapping using French core collection.

The setup and phenotypic evaluation of proposed material using randomized multi-sites experimental designs, to take into account genotype-by-environment interactions, will certainly allow the dissection of the genetic basis controlling traits of interest for a long term application using genomic selection.

**Key words:** Olive (*Olea europaea* L.), genetic diversity, genetic structure, core collection, association mapping, admixture mapping