

## Formation

# « Analyses statistiques de données génomiques pour l'amélioration génétique : théorie et application sous R »

\*\*\*

### **OBJECTIFS :**

Cette formation proposée par le Cirad vise au développement des capacités de chercheurs et doctorants de l'Université Yaoundé 1, afin qu'ils puissent :

- appliquer des méthodes statistiques actuellement parmi les plus innovantes pour l'**analyse de données génomiques pour l'amélioration génétique** (approches QTL, GWAS et sélection génomique)
- utiliser le langage **R pour les analyses génomiques** d'une manière générale (c-à-d approches statistiques citées précédemment mais aussi autres analyses pour lesquelles des packages sont disponibles librement pour R : études de diversité, simulation de génomes, etc).

Les personnes intéressées seront invitées à remplir un formulaire de candidature, à partir duquel 30 participants seront retenus, sur la base de leurs connaissances préalables et de leurs besoins dans ce domaine.

### **FORMATEURS :**

Dr. David CROS, généticien, UMR AGAP, CIRAD (Université Yaoundé 1)

Dr. Marie DENIS, mathématicienne, UMR AGAP, CIRAD (Montpellier, France)

Dr. Sébastien TISNE, généticien, UMR AGAP, CIRAD (Montpellier, France)

**DATES :** mardi 11 octobre à vendredi 14 octobre (4 jours)

**LIEU :** Université Yaoundé 1, ENSP, bloc pédagogique, 3ème étage (bâtiment en face de la nouvelle bibliothèque)

*(Voir ci-dessous le planning détaillé et une liste de publications des formateurs)*

## PLANNING DE LA FORMATION :

(4 jours / nombre total d'heures : 29h)

### JOUR 1

---

#### 9h00-9h30 – Introduction

- Présentation des formateurs et des participants, des objectifs et du programme ; accès internet

#### 9h30-10h30 / 10h45-12h - Concepts de base en génétique quantitative et génomique - D. Cros

- modèle de génétique quantitative, effet des gènes, apparentements, héritabilité et précision de sélection, réponse à la sélection, déséquilibre de liaison (DL), différenciation génétique entre populations (Fst), etc

12h00-13h30 - Pause déjeuner

#### 13h30-18h00 - R pour les débutants (TP) - D. Cros (+ M. Denis et S. Tisné)

- installation de R et RStudio,
- le langage et les commandes de R : objets, opérations, fonctions, graphiques, utilisation des matrices et dataframes, lire / exporter des données, if / for, packages, etc
- Exercices : régression linéaire, ANOVA, graphiques multivariés, ... et, en parallèle, vérification des packages

### JOUR 2

---

#### 9h00-10h00 / 10h15-11h15 - Le modèle linéaire mixte - M. Denis

- Théorie du modèle linéaire mixte : définition du modèle et présentation des méthodes d'inférence
- Prise en main du modèle linéaire mixte avec le logiciel R sur des exemples simples (un effet fixe et un effet aléatoire), dont TP avec *lme4* (R)

#### 11h15-12h00 - Concepts de base en génétique quantitative et génomique - D. Cros

- modèle linéaire mixte en génétique, gènes, marqueurs SNPs, génotypage (puces SNPs et génotypage par séquençage)

12h00-13h30 - Pause déjeuner

#### 13h30-15h30 - Principe des approches Bayésiennes - M. Denis

- Introduction à l'inférence bayésienne (théorème de Bayes)
- Présentation des méthodes de Monte Carlo par chaînes de Markov (MCMC)

#### 15h45-17h00 - R pour les débutants (TP) - D. Cros (+M. Denis et S. Tisné) (fin des exercices de la veille)

17h00 - 17h30 - Questions des participants / discussion

### JOUR 3

---

9h00-10h15 / 10h30-12h - **R pour les analyses préliminaires de génomique (TP)** - **D. Cros**, (+**M. Denis** et **S. Tisné**)

- Présentation du jeu de données (maïs),
- lecture des données moléculaires et représentation d'une carte,
- étude de la structure des populations (visualisation par ACP et calcul du Fst),
- visualisation du DL,
- calcul des apparentements

12h00-13h30 - Pause déjeuner

13h30-15h30 - **Approches QTL** - **S. Tisné**

- Principes de la cartographie QTL (variation quantitative, population de cartographie)
- Construction de carte génétique (marquage moléculaire, taux de recombinaison, distance génétique)
- Relation phénotype-information génétique (modélisation et tests statistiques, épistasie, interaction QTL\*environnement)

15h45-17h45 - **Approches QTL avec R/qtl (TP)** - **S. Tisné** (+**M. Denis** et **D. Cros**)

- Présentation du jeu de données (Arabidopsis)
- Analyses de détection de QTLs et interprétation des résultats

## **JOUR 4**

---

9h00-11h00 - **Genomewide association studies (GWAS)** - **M. Denis**

- Présentation des approches univariées
- Introduction aux méthodes pas à pas de sélection de variables
- Mise en œuvre d'une méthode univariée avec le logiciel R
- Interprétation des résultats

11h15-12h00 - **La sélection génomique** - **D. Cros**

- principe, précision
- exemple du palmier à huile

12h00-13h30 - Pause déjeuner

13h30-14h45 - **La sélection génomique** - **D. Cros**

- modèles et méthodes statistiques,

15h00-17h00 - **La sélection génomique (TP sur données précédentes de maïs)** - **D. Cros** (+**M. Denis** et **S. Tisné**)

- avec R : installation des packages, prédictions génomiques, effet du nombre de SNPs, ...

17h00-17h30 - **Questions des participants / discussion**

17h30-18h00 - **Evaluation** (échange oral + questionnaire écrit)

**Exemples de publications récentes des formateurs :**

- *Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests*, **2016**. Marchal A., Legarra A., **Tisne S.**, Carasco-Lacombe C., Manez A., Suryana E., Omoré A., Nouy B., Durand-Gasselin T., Sanchez L., Bouvet J.M., **Cros D.** *Molecular Breeding*, 36 : 15 p.
- *Mixed model approach for IBD-based QTL mapping in a complex oil palm pedigree*, **2015**. **Tisne S.**, **Denis M.**, **Cros D.**, Pomies V., Riou V., Syahputra I., Omoré A., Durand-Gasselin T., Bouvet J.M., Cochard B. *BMC Genomics*, 16 (798) : 12 p.
- *Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.)*, **2015**. **Cros D.**, **Denis M.**, Sanchez L., Cochard B., Flori A., Durand-Gasselin T., Nouy B., Omoré A., Pomies V., Riou V., Suryana E., Bouvet J.M. *Theoretical and Applied Genetics*, 128 (3) : p. 397-410.
- *Pedigree-based linkage map in two genetic groups of oil palm*, **2015**. Cochard B., Carasco-Lacombe C., Pomies V., Dufayard J.F., Suryana E., Omoré A., Durand-Gasselin T., **Tisne S.** *Tree Genetics and Genomes*, 11 (4) : 12 p.
- *Long-term genomic selection for heterosis without dominance in multiplicative traits: Case study of bunch production in oil palm*, **2015**. **Cros D.**, **Denis M.**, Bouvet J.M., Sanchez L.. 2015. *BMC Genomics*, 16 (651) : 16 p.
- *Placental genome and maternal-placental genetic interactions: A genome-wide and candidate gene association study of placental abruption*, **2014**. **Denis M.**, Salazar M., Williams M.A., Enquobahrie D.A., Tadesse M.G., Sanchez S.E., Ananth C.V. *PloS One*, 9 (12) : 24 p.
- *Evaluation of hierarchical models for integrative genomic analyses*, **2015**. **Denis M.** and Tadesse G.M. *Bioinformatics*. doi: 10.1093/bioinformatics/btv653