

Créer des variétés résistantes à la sécheresse grâce à la data

Bioinformatique, sélection génomique et statistiques pour l'amélioration des plantes

Pinel Baudelaire T. DAHOUI

Data Scientist & IA | Ingénieur en Travaux Statistiques
tpineldahoui@gmail.com | github.com/Baudelaire12

22 Avril 2026

Série LinkedIn STEM au Quotidien — Jour 27/40

Plan de la présentation

- 1 Mise en contexte
- 2 Concepts clés
- 3 Développement mathématique
- 4 Cas pratique
- 5 Conclusion
- 6 Références

L'amélioration des plantes face au changement climatique

Chiffre clé

Le changement climatique pourrait réduire les rendements agricoles mondiaux de **2 à 6%** par décennie d'ici 2100 (GIEC, 2022). En Afrique subsaharienne, les pertes de rendement du maïs liées à la sécheresse atteignent déjà **20 à 40%** les années de El Niño.

- La sélection végétale traditionnelle prend **10 à 15 ans** pour développer une nouvelle variété. La sélection génomique (GS) peut réduire ce délai à **3 à 5 ans**.
- **WEMA** (Water Efficient Maize for Africa) a développé des variétés de maïs tolérant la sécheresse qui maintiennent des rendements **20 à 30%** supérieurs sous stress hydrique (CIMMYT, 2019).
- Le coût de séquençage d'un génome végétal est passé de **100 millions \$** en 2001.

Question centrale

Comment les statistiques génomiques et le machine learning permettent-ils d'identifier et de combiner des variants génétiques favorables pour créer des variétés plus productives et résilientes ?

Modèle infinitésimal de Fisher (1918)

Les traits quantitatifs complexes (rendement, tolérance à la sécheresse) sont contrôlés par un très grand nombre de loci génétiques, chacun ayant un effet individuel faible.

Étapes de la sélection génomique :

- ➊ **Phénotypage** : mesure du trait cible (rendement, stress hydrique) sur une population de référence
- ➋ **Génotypage** : séquençage SNP à haute densité (10^4 à 10^6 marqueurs par individu)
- ➌ **Calibration** : entraînement d'un modèle statistique reliant génotype → phénotype
- ➍ **Prédiction** : estimation de la valeur génomique de nouvelles lignées non phénotypées
- ➎ **Sélection** : choix des meilleures lignées pour la génération suivante

Modèle GBLUP — Genomic Best Linear Unbiased Predictor

Modèle mixte génomique (Meuwissen et al., 2001) :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

- $\mathbf{y}$: vecteur des phénotypes observés ($n \times 1$)
- $\mathbf{X}\boldsymbol{\beta}$: effets fixes (environnement, bloc, année)
- $\mathbf{u}$: valeurs génétiques additives $\sim \mathcal{N}(\mathbf{0}, \mathbf{G}\sigma_u^2)$
- $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma_e^2)$

Matrice de parenté génomique (VanRaden, 2008) :

$$\mathbf{G} = \frac{\mathbf{M}\mathbf{M}^T}{2 \sum_j p_j(1 - p_j)}$$

où $\mathbf{M}$ est la matrice de marqueurs centrés ($M_{ij} = m_{ij} - 2p_j$) et p_j la fréquence allélique du marqueur j .

Prédiction BLUP :

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^T (\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \lambda \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad \lambda = \frac{\sigma_e^2}{\sigma_u^2}$$

Machine learning pour la prédiction génomique

Régression Ridge (RR-BLUP) — équivalente à GBLUP :

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{M}\alpha\|^2 + \lambda \|\alpha\|^2$$

Bayesian LASSO — sélection de marqueurs (Park & Casella, 2008) :

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{M}\alpha\|^2 + \lambda \|\alpha\|_1$$

Deep Learning génomique — réseau convolutif 1D sur séquences SNP : Encode la séquence de SNPs comme une séquence 1D → CNN 1D capture les interactions épistatiques locales → couche dense → prédiction du phénotype.

Précision de prédiction comparative

Corrélation prédiction/observation : GBLUP $\approx$ 0.65, RF $\approx$ 0.68, CNN-1D $\approx$ 0.72 pour le rendement du maïs sous stress (Montesinos-López et al., 2021).

Cas pratique : WEMA — Maïs tolérant la sécheresse en Afrique

Projet WEMA (Water Efficient Maize for Africa)

Partenariat CIMMYT, IITA, AATF et 5 pays africains (Kenya, Tanzanie, Ouganda, Mozambique, Afrique du Sud). Objectif : développer des variétés de maïs hybrides tolérant la sécheresse adaptées à l'Afrique subsaharienne.

Résultats sur 10 ans de sélection génomique (CIMMYT, 2019) :

Indicateur	Variétés traditionnelles	WEMA génomique
Rendement moyen (t/ha)	1.2	2.1 (+75%)
Rendement sous stress	0.7	1.4 (+100%)
Cycle de sélection	7–10 ans	3–4 ans
Agriculteurs bénéficiaires	—	> 5 millions

Bénin : L'INRAB développe des variétés de maïs, sorgho et igname adaptées aux conditions locales. L'intégration de la sélection génomique accélérerait considérablement ces programmes.

Synthèse

La sélection génomique fusionne statistiques bayésiennes, algèbre linéaire et bioinformatique pour accélérer l'amélioration des plantes. Elle est particulièrement stratégique pour l'Afrique, où l'adaptation rapide des cultures au changement climatique est une question de survie alimentaire.

Perspectives béninoises :

- Programme de sélection génomique pour le maïs, l'igname et le manioc au Bénin (INRAB + CGIAR)
- Séquençage à bas coût des ressources génétiques phytogénétiques béninoises (banque de gènes)
- Partenariat avec Africa Biogenomics Initiative pour l'accès aux données génomiques africaines

« Le meilleur outil contre la faim n'est pas l'aide alimentaire — c'est une graine améliorée adaptée au sol africain. »

Références bibliographiques

-  Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829.
-  VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423.
-  CIMMYT. (2019). *WEMA Project : Water Efficient Maize for Africa — Final Report*. International Maize and Wheat Improvement Center.
-  Montesinos-López, O. A., et al. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, 22, 19.
-  Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681–686.
-  GIEC. (2022). *Sixième rapport d'évaluation — Impacts, adaptation et vulnérabilité*. IPCC Working Group II.

Merci pour votre attention

Pinel Baudelaire T. DAHOUI

Data Scientist & IA | Ingénieur en Travaux Statistiques

`tpineldahoui@gmail.com | github.com/Baudelaire12`

LinkedIn : @Pinel Baudelaire DAHOUI

Série STEM au Quotidien — 40 jours