



HAL
open science

Prédiction génomique du potentiel adaptatif dans un nouvel environnement

Louise Camus, Simon Boitard, Mathieu Gautier

► **To cite this version:**

Louise Camus, Simon Boitard, Mathieu Gautier. Prédiction génomique du potentiel adaptatif dans un nouvel environnement. Journée d'animation INRAE SPE et Numérique, May 2022, Lyon, France. hal-04659137

HAL Id: hal-04659137

<https://hal.inrae.fr/hal-04659137v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Prédiction génomique du potentiel adaptatif dans un nouvel environnement

Louise Camus, Simon Boitard, Mathieu Gautier

INRAE, Centre de Biologie et de Gestion des Populations (CBGP), Montpellier

Journée d'animation SPE et numérique
Lyon, 31 mai 2022

1 Contexte

2 Méthodes

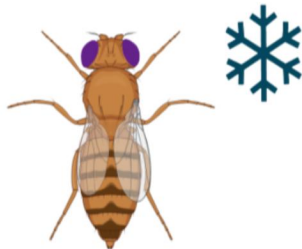
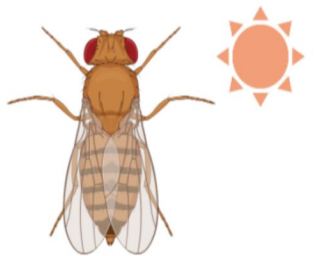
3 Applications

1 Contexte

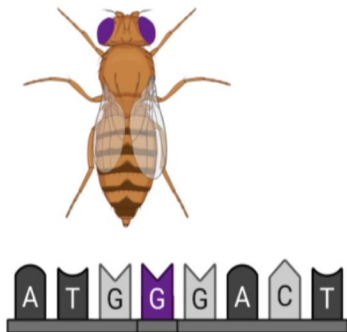
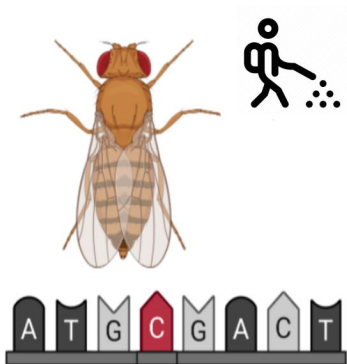
2 Méthodes

3 Applications

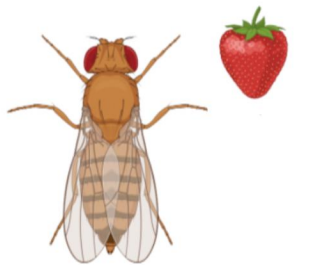
Adaptation génétique à l'environnement



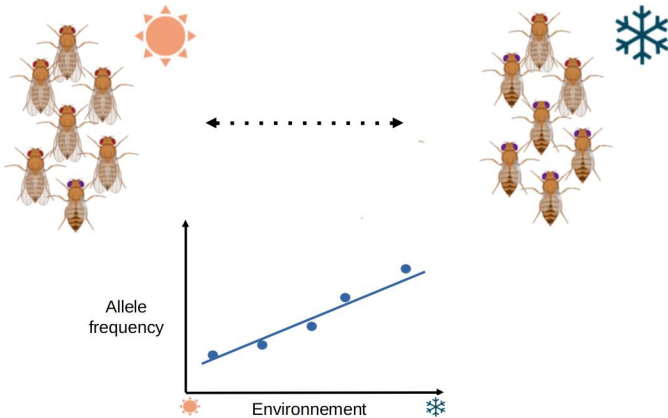
Adaptation génétique à l'environnement



Adaptation génétique à l'environnement

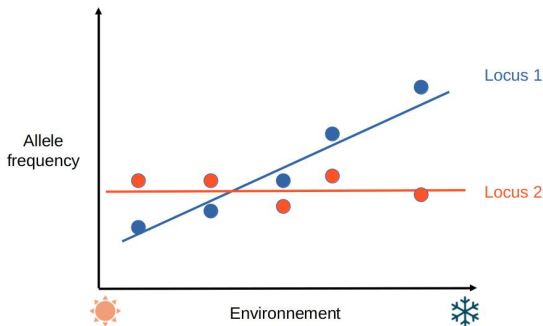


Impact sur la diversité génétique



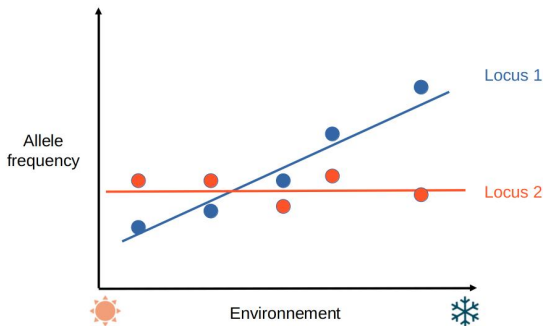
Questions de recherche

- Détecter les locus permettant de s'adapter : *GEA (Genotype Environment Association)*.
- Identifier les variables importantes pour l'adaptation.
- Prédire les fréquences alléliques optimales dans un nouvel environnement : **vulnérabilité génomique / genomic offset**.
- Prédire le phénotype ou l'environnement associé à un fond génétique : **prédiction génomique**.



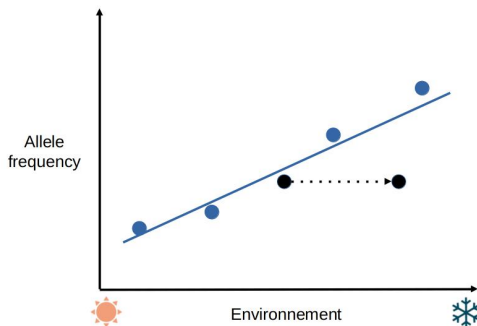
Questions de recherche

- Détecter les locus permettant de s'adapter : *GEA* (*Genotype Environment Association*).
- Identifier les variables importantes pour l'adaptation.
- Prédire les fréquences alléliques optimales dans un nouvel environnement : **vulnérabilité génomique / genomic offset**.
- Prédire le phénotype ou l'environnement associé à un fond génétique : **prédiction génomique**.



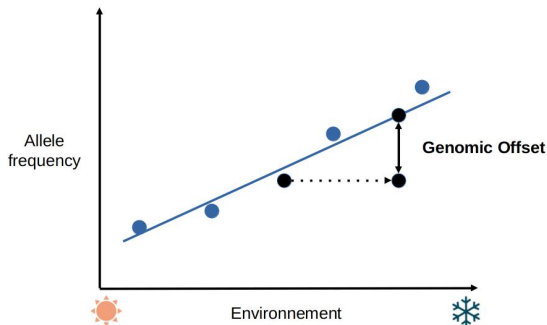
Questions de recherche

- Détecter les locus permettant de s'adapter : *GEA (Genotype Environment Association)*.
- Identifier les variables importantes pour l'adaptation.
- Prédire les fréquences alléliques optimales dans un nouvel environnement : **vulnérabilité génomique / genomic offset**.
- Prédire le phénotype ou l'environnement associé à un fond génétique : **prédiction génomique**.



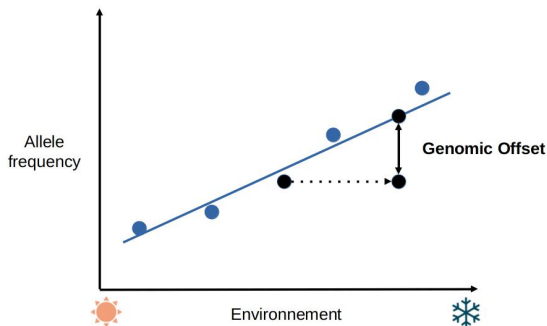
Questions de recherche

- Détecter les locus permettant de s'adapter : *GEA (Genotype Environment Association)*.
- Identifier les variables importantes pour l'adaptation.
- Prédire les fréquences alléliques optimales dans un nouvel environnement : **vulnérabilité génomique / genomic offset**.
- Prédire le phénotype ou l'environnement associé à un fond génétique : **prédiction génomique**.



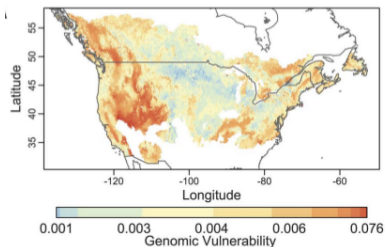
Questions de recherche

- Détecter les locus permettant de s'adapter : *GEA* (*Genotype Environment Association*).
- Identifier les variables importantes pour l'adaptation.
- Prédire les fréquences alléliques optimales dans un nouvel environnement : **vulnérabilité génomique / genomic offset**.
- Prédire le phénotype ou l'environnement associé à un fond génétique : **prédiction génomique**.



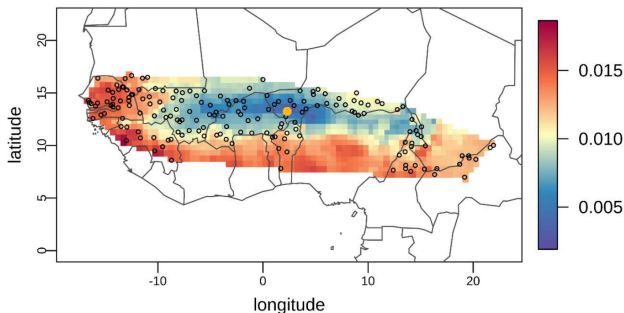
Vulnérabilité au changement climatique (Bay *et al*, 2018)

- Apprentissage : 21 populations naturelles genotypées pour $\approx 100,000$ SNPs, 19 variables climatiques (WorldClim).
- **Projection temporelle:** nouvel environnement = climat futur des localités échantillonnées.



Déplacement de populations (Rhoné *et al*, 2020)

- Apprentissage : 173 variétés de millet perlé genotypées pour $\approx 140,000$ SNPs, 157 variables climatiques (CMIP5).
- **Projection spatiale:** nouvel environnement = localité différente.



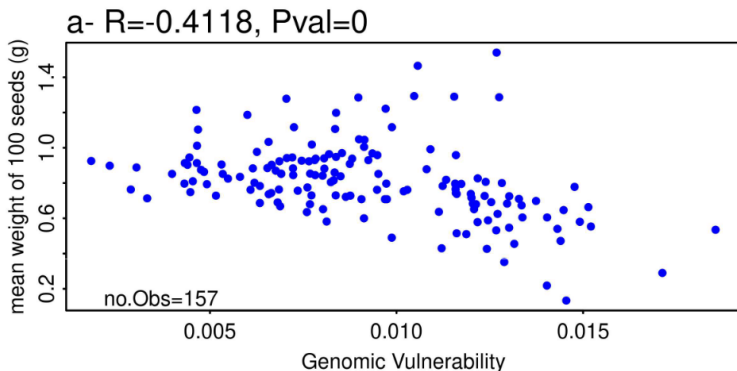
⇒ Application aux **espèces envahissantes**.

Pertinence des prédictions?

- **Hypothèses** fortes :
 - Populations d'apprentissage localement adaptées.
 - Adaptation future via diversité existante.
 - Populations statiques.
- **Evaluation** des prédictions :
 - Souvent absente ou très indirecte ...
 - Vulnérabilité génomique vs performance (e.g. Rhoné *et al*, 2020).
 - Simulations (e.g. Laruson *et al*, 2022).
 - Validation croisée?

Vulnérabilité génomique vs performance (Rhoné *et al*, 2020)

- Prendre une population dans l'environnement A.
- Evaluer sa performance dans l'environnement B (jardin commun).
- Comparer avec la vulnérabilité génomique entre A et B.



- Evaluer la qualité des prédictions génomiques à l'aide de simulations.
- Comparer différentes méthodes de prédiction.
- (Développer de nouvelles méthodes de prédiction).
- Evaluer l'intérêt de ces approches pour l'étude des espèces envahissantes ou nuisibles.

Drosophila. suzukii : C. Deschamp, A. Estoup, A. Loiseau, L. Olazcuaga, N. Rode (CBGP).

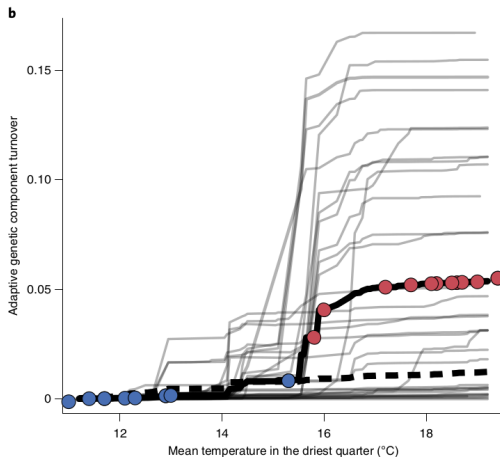
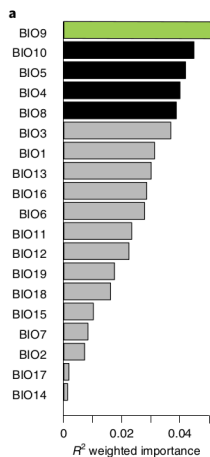
1 Contexte

2 Méthodes

3 Applications

Gradient Forest (Fitzpatrick et Keller, 2015)

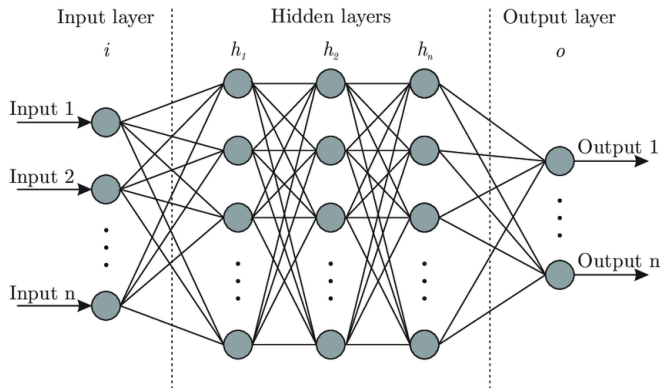
- $Freq_{i,j} = g_j(Env_i), \forall \text{ SNP } j$.
- Forêt aléatoire \rightarrow lien $g_j()$ non linéaire.
- Corrélations entre SNPs et structure génétique neutre **non prises en compte**.



- $Freq_{i,j} = g_j(Env_i) + \epsilon_i, \forall \text{ SNP } j$.
- Modèle bayésien hiérarchique, GEA.
- Structure génétique neutre **modélisée**.
- $g_j()$ linéaire.
- Potentiel pour la prédiction et environnement multi-dimensionnel à confirmer.

Réseaux de neurones profonds

- $Env_i = g(Freq_i)$.
- Corrélations entre SNPs et entre variables environnementales.
- Lien $g()$ non linéaire.
- Structure génétique neutre? Quantité de données?



1 Contexte

2 Méthodes

3 Applications

Drosophila suzukii

Espèce envahissante originaire d'Asie, dommages importants pour cultures fruitières.



- Séquençage tout génome (*WGS*) (individuel ou en pool) de ≈ 45 populations naturelles en Asie, Europe et Amérique (Olazcuaga *et al*, 2020; Lewald *et al*, 2021).
- 19 variables climatiques (WorldClim).
- Variables déterminantes pour l'adaptation aux aires envahies?
- La vulnérabilité génétique est elle un bon indicateur du potentiel invasif?

Adaptation à la plante hôte

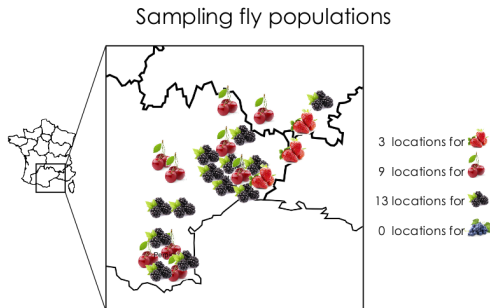
- Généraliste mais capable de s'adapter à un fruit hôte particulier en quelques générations (Olazcuaga *et al*, BioRxiv 2022).



- Prédire la performance relative sur un fruit donné à partir du génome?

Adaptation à la plante hôte

- 55 populations naturelles issues de 3 fruits (France et USA), données WGS en pool.

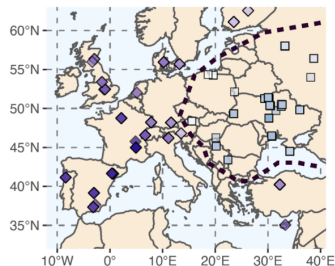
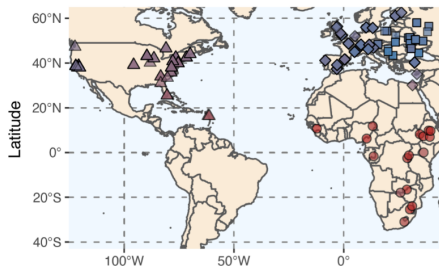


- 11 lignées expérimentales sur 3 fruits, données WGS en pool pour générations 0, 12, 29 and 56.

Adaptation saisonnière chez *D. melanogaster*

Projet DEST (Kapun *et al*, 2021):

- Données WGS en pool pour 270 échantillons, 55 populations échantillonnées à différentes saisons.
- Prédiction génomique de la saison d'un échantillon, réseaux de neurones.
- M2 Louis NGuyen.



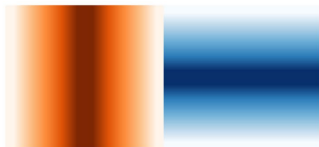
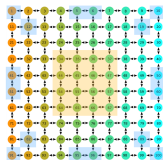
1 Contexte

2 Méthodes

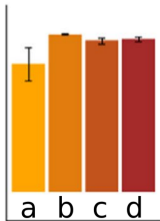
3 Applications

Simulations (Laruson *et al*, 2022)

Polygenic stabilizing selection for two traits, distinct local optima.

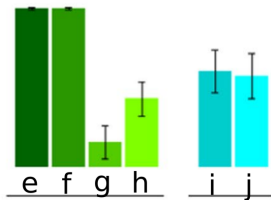


Case 2 Core



Case 3 Core

|



- a** GF off. genome, all env.
- b** GF off. genome, causal env.
- c** GF off. causal, all env.
- d** GF off. causal, causal env.

- e** Causal Env. dist. (ED)
- f** Causal Env. dist. (MD)
- g** All Env. dist. (ED)
- h** All Env. dist. (MD)

- i** FST genome
- j** FST causal