Prédiction génomique de l'adaptation de Drosophila suzukii à différentes plantes hôtes

Simon Boitard, Arnaud Estoup, Mathieu Gautier, Nicolas Rode

▶ **To cite this version:**

Simon Boitard, Arnaud Estoup, Mathieu Gautier, Nicolas Rode. Prédiction génomique de l'adaptation de Drosophila suzukii à différentes plantes hôtes. Journée Intelligence Artificielle et Biodiversité, Labex Cemeb & NUMEV, Mar 2021, Montpellier, France. hal-03649525

HAL Id: hal-03649525
https://hal.inrae.fr/hal-03649525

Submitted on 22 Apr 2022

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.

Prédiction génomique de l'adaptation de *Drosophila suzukii* à différentes plantes hôtes

Simon Boitard, Arnaud Estoup, Mathieu Gautier, Nicolas Rode

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

Journée Intelligence Artificielle et Biodiversité
Montpellier, 30 mars 2021

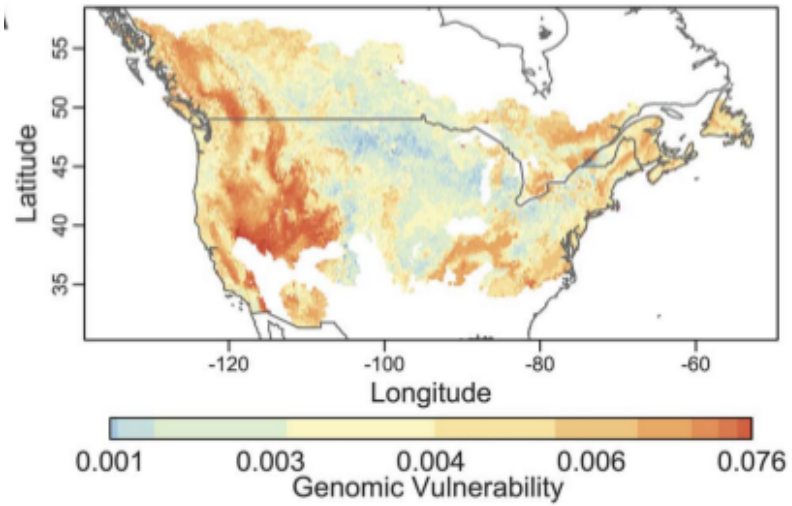
Contexte

- Beaucoup de caractères ont une base génétique.
- Prédire la valeur d'un caractère d'intérêt pour un individu à partir de son génome (élevage / médecine personnalisée).
- L'adaptation d'une population a son environnement a aussi une base génétique (détection de locus sous sélection).

$\Rightarrow$ **Prédire le potentiel adaptatif** d'une population dans un nouvel environement à partir de sa diversité génétique?

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

- Apprentissage: fréquences alléliques pour $\approx$ 100,000 SNPs et 21 populations naturelles évoluant dans conditions climatiques contrastées, supposées **bien adaptées**.
- Test: fréquences optimales pour de nouvelles conditions climatiques issues de projections $\rightarrow$ **Vulnérabilité**.



Cadre statistique

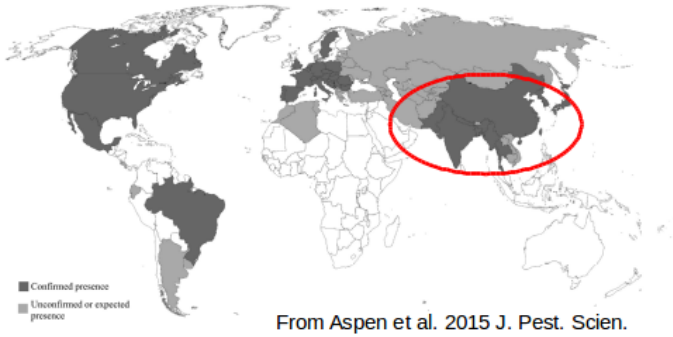
- Populations d'apprentissage $i = 1 \dots n$.
- $X_i = (X_{i,1}, \dots, X_{i,p})$ fréquences alléliques pour p marqueurs.
- $Y_i = (Y_{i,1}, \dots, Y_{i,q})$ co-variables environnementales.
- Estimer $X_{i,j} = g_j(Y_i)$, $\forall j$
- Prédire $X_{i+1,j} = g_j(Y_{i+1})$ pour un nouvel environnement.
- $\tilde{Y}_i$ climat prédit pour la population i.
- Vulnérabilité $v_i = \sum_j (g_j(\tilde{Y}_i) - X_{i,j})$.

Cadre statistique

- Populations d'apprentissage $i = 1 \dots n$.
- $X_i = (X_{i,1}, \dots, X_{i,p})$ fréquences alléliques pour p marqueurs.
- $Y_i = (Y_{i,1}, \dots, Y_{i,q})$ co-variables environnementales.
- Estimer $X_{i,j} = g_j(Y_i)$, $\forall j$
- Prédire $X_{i+1,j} = g_j(Y_{i+1})$ pour un nouvel environnement.
- $\tilde{Y}_i$ climat prédit pour la population i.
- Vulnérabilité $v_i = \sum_j (g_j(\tilde{Y}_i) - X_{i,j})$.

Adaptation de *Drosophila suzukii* à la plante hôte

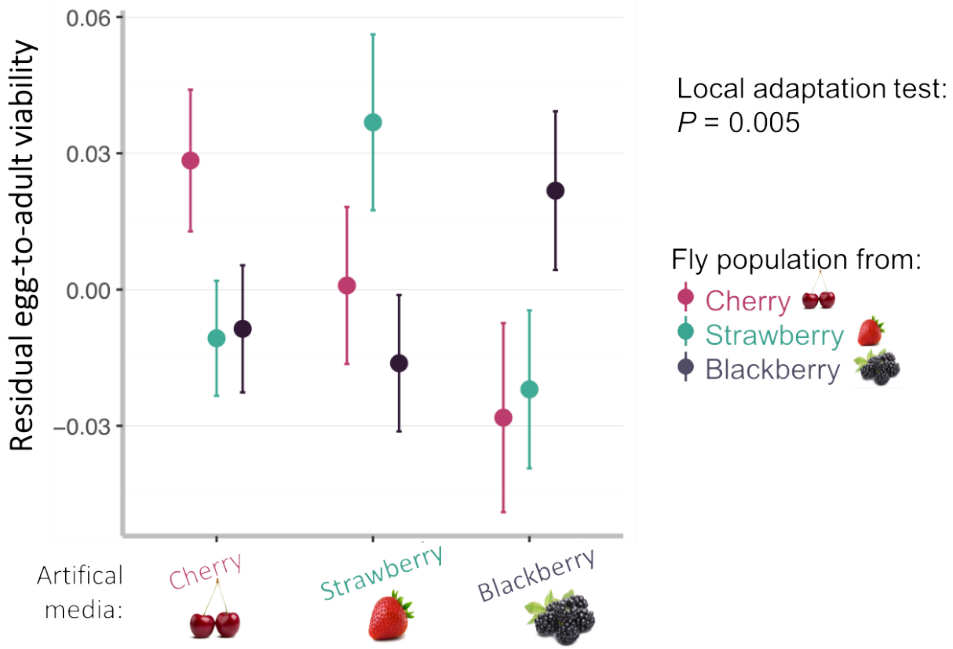
Espèce envahissante originaire d'Asie causant des dommages conséquents aux cultures fruitières.



From Aspen et al. 2015 J. Pest. Scien.

Adaptation de *Drosophila suzukii* à la plante hôte

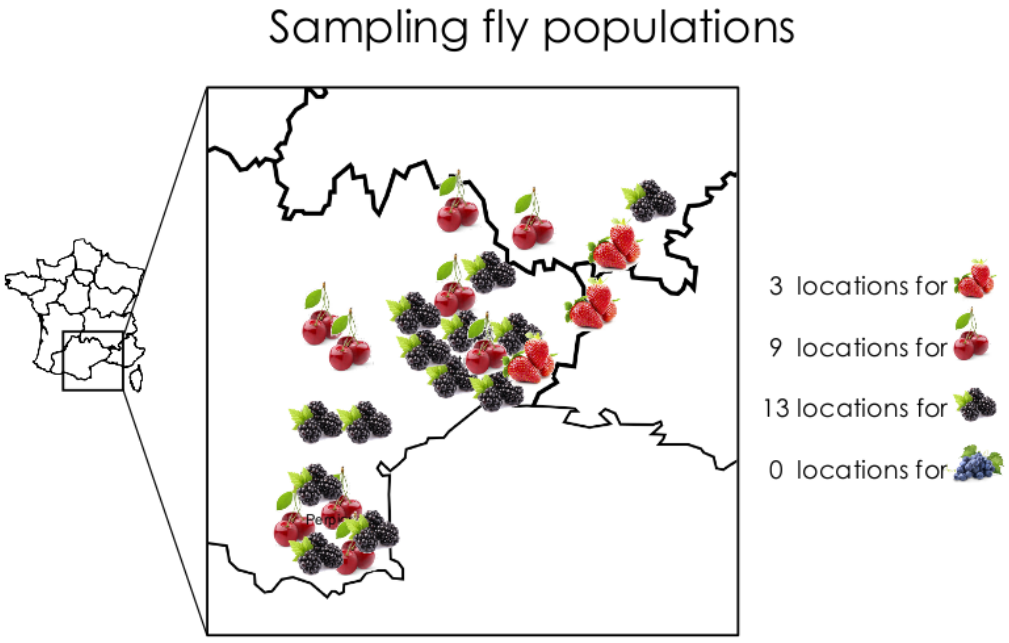
- Généraliste mais capable de s'adapter à un fruit hôte particulier en quelques générations.



- Prédire la valeur sélective moyenne d'une population sur un fruit donné à partir de ses fréquences alléliques.

Données disponibles

- 55 populations naturelles issues de 3 fruits (fraise 15, cerise 15, mûre 25) et deux zones d'invasion indépendantes (France et USA) en cours de sequençage.



- 11 lignées expérimentales maintenues sur 3 fruits (fraise, cerise, canneberge) et séquencées aux générations 0, 12, 29 (déjà fait) et 56 (à venir).

Méthodes envisagées

- **Gradient Forest** (Bay *et al*, 2018; Fitzpatrick et Keller, 2015).
  - Forêts aléatoires.
  - Lien $g()$ non linéaire.
  - Aspect multi-marqueur peu exploité.
- **BayPass** (Gautier, 2015)
  - Modèle Bayésien hiérarchique (MCMC).
  - Détection des locus adaptatifs (modèle prédictif à valider).
  - Prise en compte de la structure génétique neutre entre populations.
  - Lien $g()$ linéaire.
- **Autres méthodes d'AI**?

Feuille de route

- Comparaison de méthodes sur données simulées.
- Evaluation sur les données réelles disponibles (validation croisée).
- Développement de nouvelles méthodes.
- **Financement de thèse disponible pour septembre 2021**.