

Prédiction génomique du potentiel adaptatif de populations dans un nouvel environnement : application aux espèces envahissantes



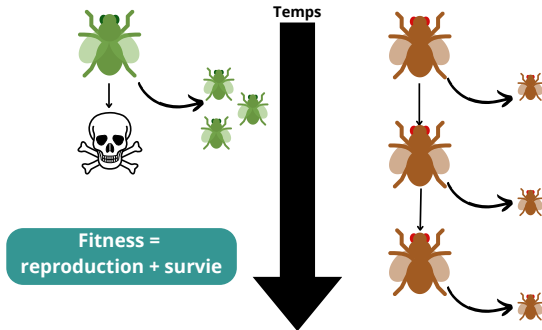
Louise Camus

Sous la direction de Simon Boitard et Mathieu Gautier

19 Décembre 2024

L'adaptation

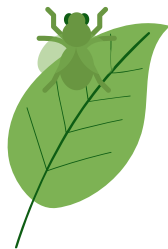
→ **sélection naturelle** = les individus ayant le meilleur succès reproducteur (*fitness*) transmettent plus leurs caractères (= **phénotype**) aux générations suivantes



Caractères totalement ou partiellement héréditaires = transmis par l'ADN

L'adaptation

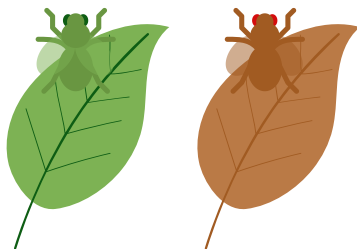
→ **sélection naturelle** = les individus ayant le meilleur succès reproducteur (*fitness*) transmettent plus leurs caractères (= **phénotype**) aux générations suivantes



→ adéquation entre le phénotype et l'environnement → **adaptation**

L'adaptation

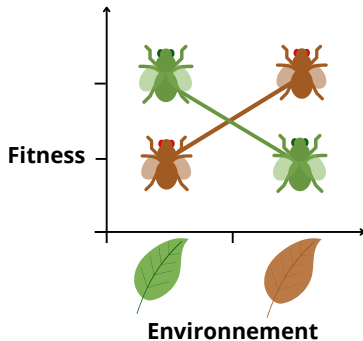
→ **sélection naturelle** = les individus ayant le meilleur succès reproducteur (*fitness*) transmettent plus leurs caractères (= **phénotype**) aux générations suivantes



→ adéquation entre le phénotype et l'environnement → **adaptation locale**

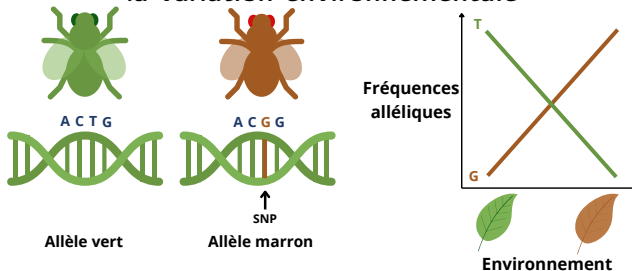
L'adaptation locale

Les populations développent des caractères leur conférant une fitness moyenne plus élevée dans leur environnement local que les populations provenant d'autres environnements



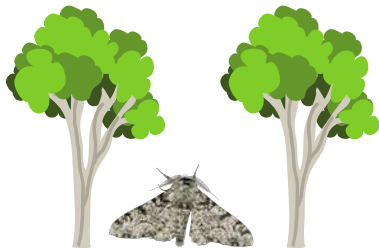
L'adaptation locale

Les populations développent des caractères leur conférant une valeur sélective moyenne plus élevée dans leur environnement local que les populations provenant d'autres environnements, **en raison de la correspondance spatiale entre la variation génétique adaptative et la variation environnementale**

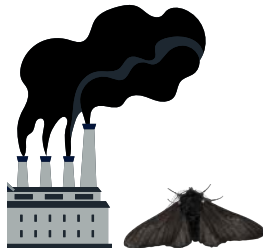


L'exemple de la phalène du bouleau

Campagne

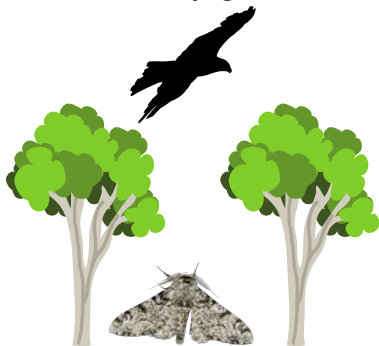


Zone industrialisée



L'exemple de la phalène du bouleau

Campagne



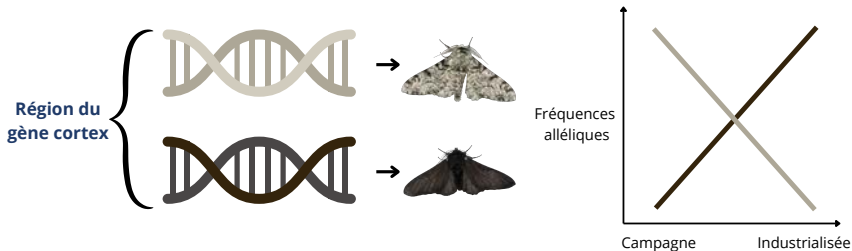
Zone industrialisée



→ Meilleure fitness dans l'environnement local

L'exemple de la phalène du bouleau

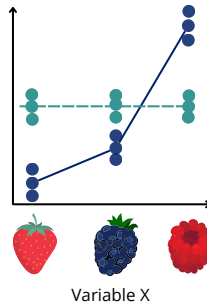
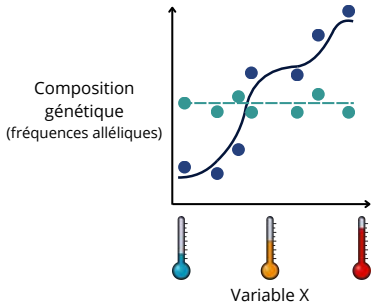
- Correspondance entre la variation génétique adaptative et la variation environnementale.



- Un exemple simplifié
- L'adaptation locale a été détectée chez de nombreuses espèces.
- Utilisation de l'information génétique

Détecter l'adaptation locale

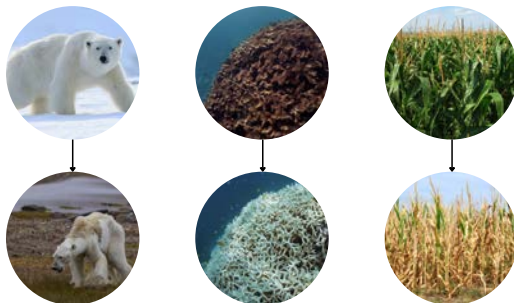
GEA = Genome Environment Association



- Identification de marqueurs/variables environnementales ayant un rôle dans l'adaptation

De l'adaptation locale à la maladaptation : quand l'environnement change

Maladaptation = Diminution de la *fitness* moyenne d'une population au cours du temps, en raison d'une rupture de l'adéquation entre son phénotype moyen et son environnement

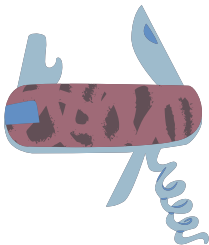


Le potentiel adaptatif

Comment répondre à un changement d'environnement ?

Le potentiel adaptatif d'une population = sa capacité à répondre génétiquement à la sélection

Diversité génétique

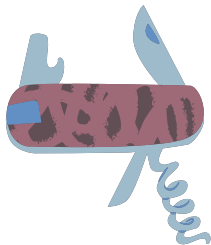


Le potentiel adaptatif

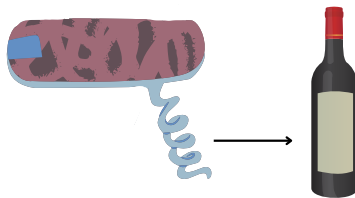
Comment répondre à un changement d'environnement ?

Le potentiel adaptatif d'une population = sa capacité à répondre génétiquement à la sélection

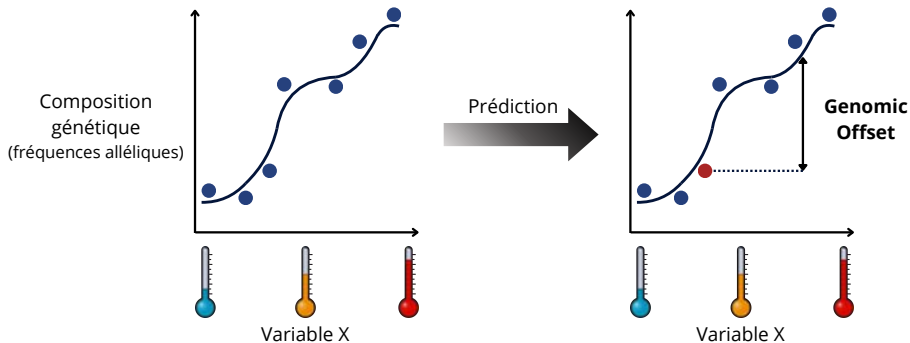
Diversité génétique



Relatif à l'environnement



Le Genomic Offset (Fitzpatrick et Keller, 2015)

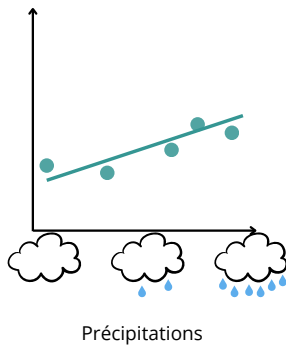
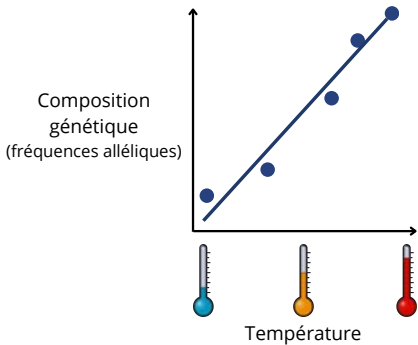


Genomic Offset (GO) = une mesure du risque de (mal)adaptation

- Temporel (conditions environnementales futures)
- Spatial (autres zones géographiques)

Une autre interprétation

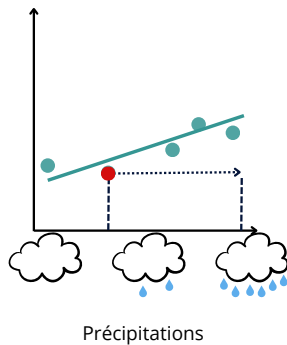
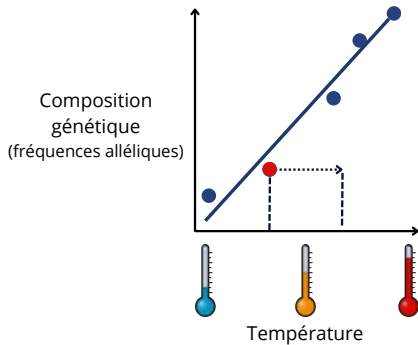
Genomic Offset (GO) = pondération de la distance environnementale par la génétique (Gain et al., 2023)



Distance environnementale < Distance génétique

Une autre interprétation

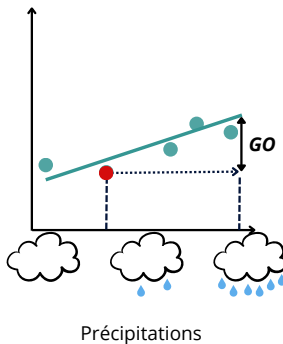
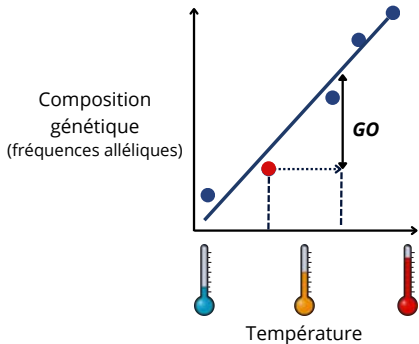
Genomic Offset (GO) = pondération de la distance environnementale par la génétique (Gain et al., 2023)



Distance environnementale < Distance génétique

Une autre interprétation

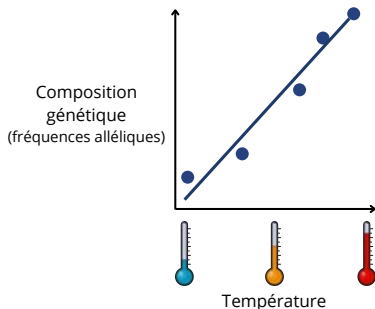
Genomic Offset (GO) = pondération de la distance environnementale par la génétique (Gain et al., 2023)



Distance Contribution au GO > Contribution au GO

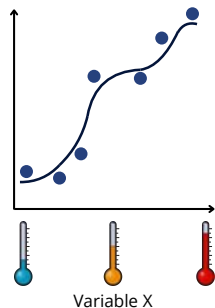
Deux types de méthodes

Linéaire



GO Géométrique (*Gain et al., 2023*)
RDA (*Capblancq et al., 2021*)
RONA (*Rellstab et al., 2016*)

Non linéaire



Gradient Forest (*Fitzpatrick et Keller, 2015*)
GDM (*Fitzpatrick et Keller, 2015*)
SPAG (*Rochat et al., 2021*)

Cependant ...

Plusieurs méthodes possibles

Des hypothèses fortes

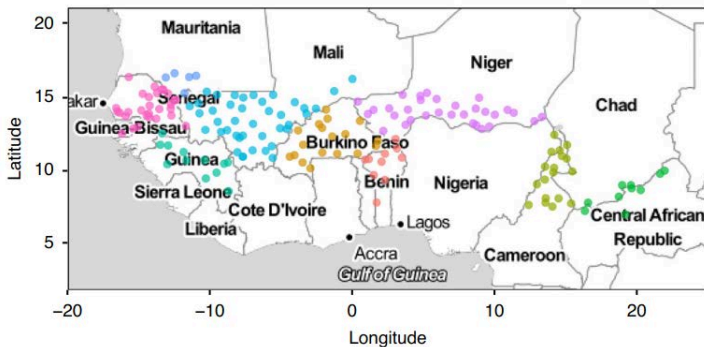
- Modélisation correcte des GEA
- Populations adaptées
- Adaptation *via* des variants pré-existants
- Relation génome ↔ environnement invariante

Le GO prédit-il le risque de (mal)adaptation ?

Lien entre GO et fitness

Simulation de données : Laruson *et al.* (2021), Gain *et al.* (2023)

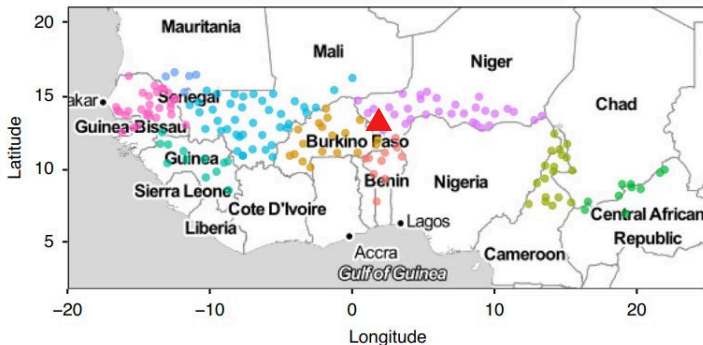
En jardin commun : Rhoné *et al.* (2020), Fitzpatrick *et al.* (2021)



Lien entre GO et fitness

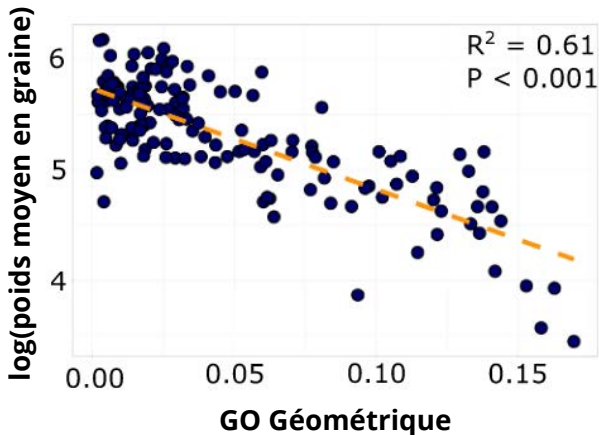
Simulation de données : Laruson *et al.* (2021), Gain *et al.* (2023)

En jardin commun : Rhoné *et al.* (2020), Fitzpatrick *et al.* (2021)



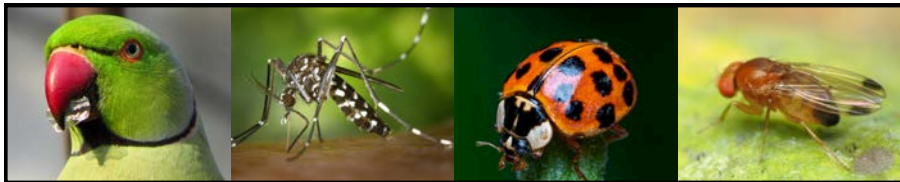
Lien entre GO et fitness

Simulation de données : Laruson *et al.* (2021), Gain *et al.* (2023)
En jardin commun : Rhoné *et al.* (2020), Fitzpatrick *et al.* (2021)



Espèces envahissantes

Invasions biologiques = un exemple de changement environnemental rapide



- Potentiellement nuisibles
- Augmentation de leur nombre ces dernières années du fait de la mondialisation (Seebens, 2021a, Hulme, 2021)
- Influence du dérèglement climatique ?
 - Comment expliquer et prédire leur succès dans de “nouveaux” environnements ?

Le paradoxe de l'invasion

- Nombre réduit d'individus fondateurs
 - Ex : frelon asiatique (Arca et al., 2015)
 - Réduction diversité génétique, consanguinité, ... → potentiel adaptatif réduit



- Adaptation à un environnement "inconnu" (pré-adaptation?)

Apport de la génomique

- Reconstruction des histoires d'invasions



- Identification de zones du génome sous sélection
- GO : évaluer la (mal)adaptation à des environnements futurs ou à d'autres zones géographiques → prédire le risque d'invasion

Objectifs de ma thèse

Etudier l'adaptation de populations envahissantes à de nouveaux environnements à travers la génomique

Partie 1 : Simulations : Le GO est-il corrélé à la probabilité d'établissement ?

Partie 2 : Compréhension du succès de l'espèce envahissante et ravageuse de cultures *Drosophila suzukii* à l'aide de méthodes génomiques.



Partie 3 : Adaptation génétique à la plante hôte chez *D. suzukii* : scans génétiques et prédiction génomique.

Objectifs de ma thèse

Etudier l'adaptation de populations envahissantes à de nouveaux environnements à travers la génomique

Partie 1 : Simulations : Le GO est-il corrélé à la probabilité d'établissement ?

Partie 2 : Compréhension du succès de l'espèce envahissante et ravageuse de cultures *Drosophila suzukii* à l'aide de méthodes génomiques.



Partie 3 : Adaptation génétique à la plante hôte chez *D. suzukii* : scans génétiques et prédiction génomique.

Processus d'invasion

Transport

Processus d'invasion

Transport



Introduction

Processus d'invasion

Transport



Introduction



Etablissement

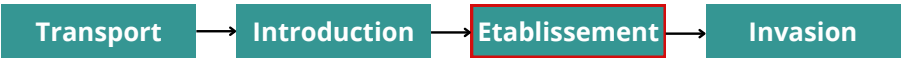
Survie et reproduction suffisantes pour établir une/des population(s)
stable(s) et autonome(s)

Processus d'invasion



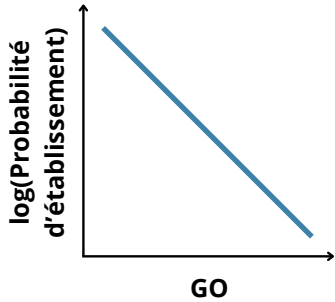
Espèce envahissante dont les individus se dispersent, survivent et se reproduisent sur de multiples sites

Processus d'invasion



Dernière étape avant qu'une espèce soit qualifiée d'envahissante
Fitness suffisante pour maintenir une/des population(s) stable(s)

Hypothèse :



Design des simulations

2 étapes (SLiM v4)

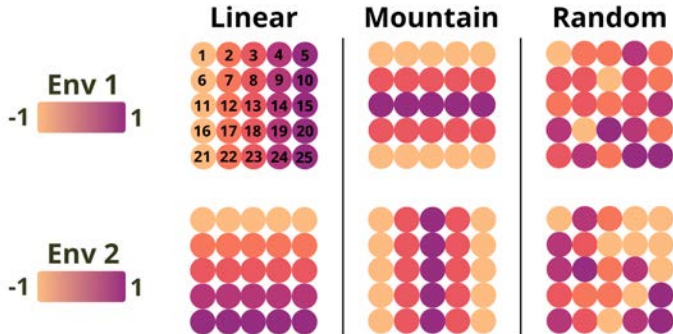
1. Aire native

- Grille de population (5x5) , stepping stone
- 1000 individus/population
- Fort (0.05) ou faible (0.005) taux de migration
- 3000 générations simulées



Design des simulations : aire native

- 2 variables environnementales reliées à l'adaptation par des QTLs (Quantitative Trait Loci)
 - Σ tailles d'effets = phénotype \leftrightarrow environnement = fitness
 - 3 différents types d'environnements (répliqués 10x chacun) :



Design des simulations : aire envahie

2. Aire envahie

- Population panmictique caractérisée par 2 var. environnementales
↪ **9 env. possibles à envahir** : $(-1,0,1) \times (-1,0,1)$

Design des simulations : aire envahie

2. Aire envahie

- Population panmictique caractérisée par 2 var. environnementales
↪ **9 env. possibles à envahir** : $(-1,0,1) \times (-1,0,1)$
- 10 ou 100 individus envahissants
↪ **3 populations source possibles** : -1/-1, 0/0, 1/1

Design des simulations : aire envahie

2. Aire envahie

- Population panmictique caractérisée par 2 var. environnementales
↪ **9 env. possibles à envahir** : $(-1,0,1) \times (-1,0,1)$
- 10 ou 100 individus envahissants
↪ **3 populations source possibles** : -1/-1, 0/0, 1/1
- Taille de population non contrainte ("Non-WF") :
 - Générations chevauchantes
 - Mort \neq Reproduction

Design des simulations : aire envahie

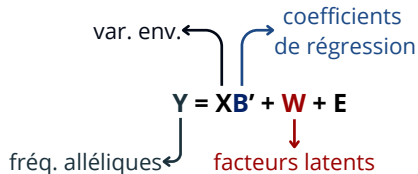
2. Aire envahie

- Population panmictique caractérisée par 2 var. environnementales
 ↪ **9 env. possibles à envahir** : $(-1,0,1) \times (-1,0,1)$
- 10 ou 100 individus envahissants
 ↪ **3 populations source possibles** : -1/-1, 0/0, 1/1
- Taille de population non contrainte ("Non-WF") :
 - Générations chevauchantes
 - Mort \neq Reproduction
- La simulation s'arrête quand la population :
 - s'éteint
 - **s'établit**
 - ↪ atteint 50 000 individus
 - ↪ survit pendant 100 générations

Méthodes de GO

GO Géométrique (Gain et al. 2023)

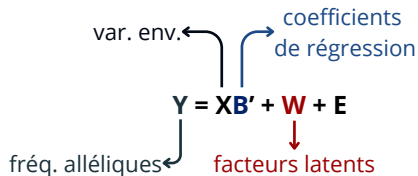
- Linéaire (régression)



Méthodes de GO

GO Géométrie (Gain et al. 2023)

- Linéaire (régression)

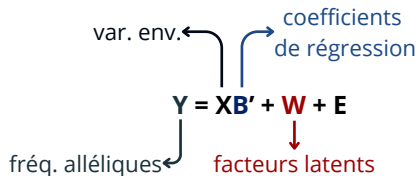


- Correction pour la structure neutre :
 - Facteurs latents (LFMM)
 - Covariance des fréquences alléliques (Baypass)

Méthodes de GO

GO Géométrique (Gain et al. 2023)

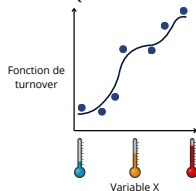
- Linéaire (régression)



- Correction pour la structure neutre :
 - Facteurs latents (*LFMM*)
 - Covariance des fréquences alléliques (*Baypass*)

Gradient Forest (Fitzpatrick & Keller, 2015)

- Non Linéaire (Random Forest)



- Fréquences alléliques corrigées par LFMM
 - Optimisé

Analyses

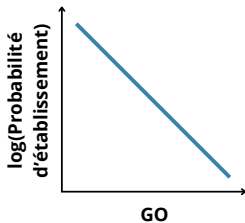
Le GO peut être calculé :

Variables

2 variables causales

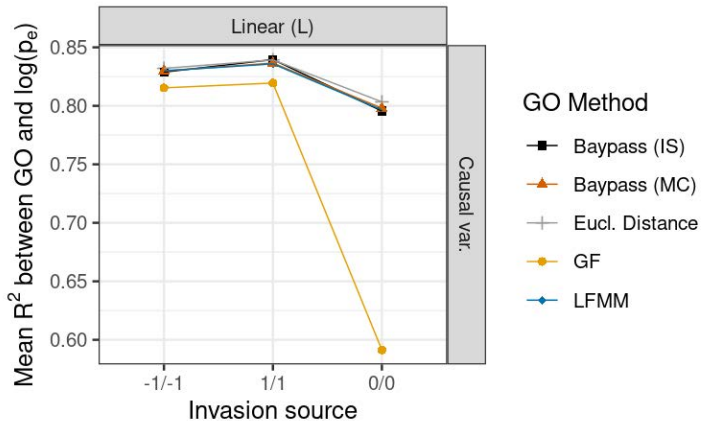
2 variables causales
+ 6 confondantes

Chaque invasion répétée 250 fois → proba. d'établissement (PE)



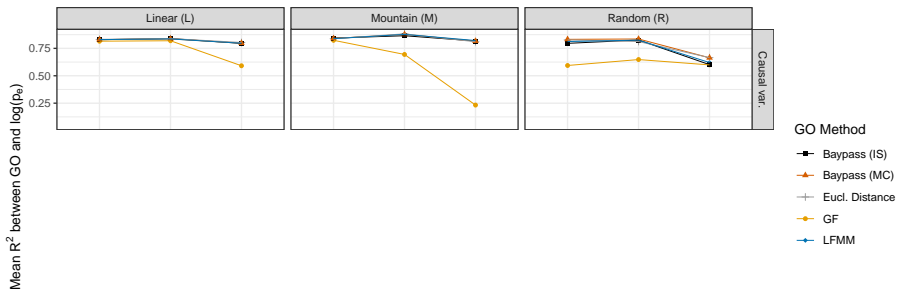
Résultats dans un cas "idéal"

R^2 moyen entre GO et PE (10 ind., structuration forte)



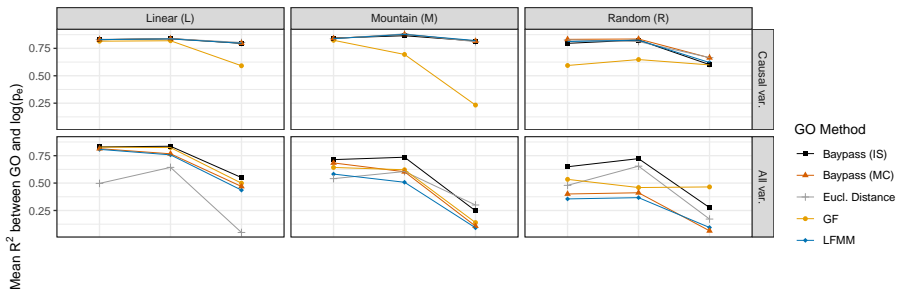
Bonnes performances avec les variables causales

R^2 moyen entre GO et EP (10 ind., structuration forte)



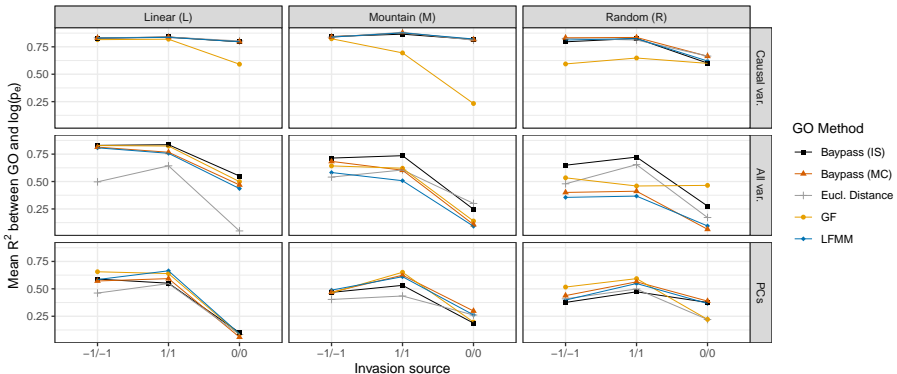
Variables confondantes : réduction des performances

R^2 moyen entre GO et EP (10 ind., structuration forte)



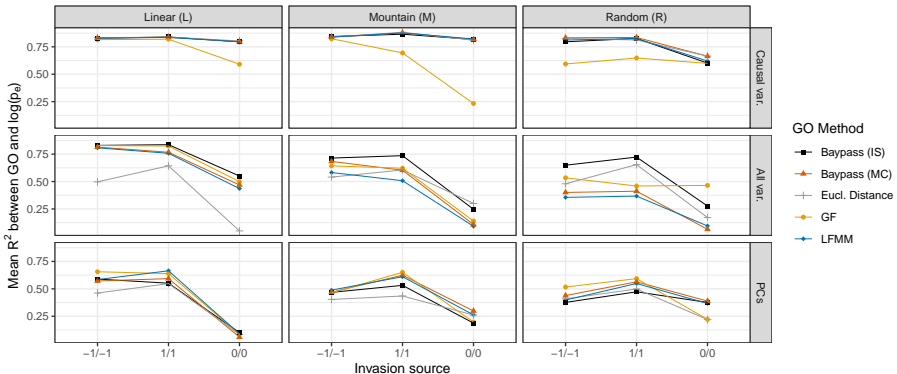
Les PCs comblent l'écart de performances entre méthodes

R^2 moyen entre GO et EP (10 ind., structuration forte)



Les PCs comblent l'écart de performances entre méthodes

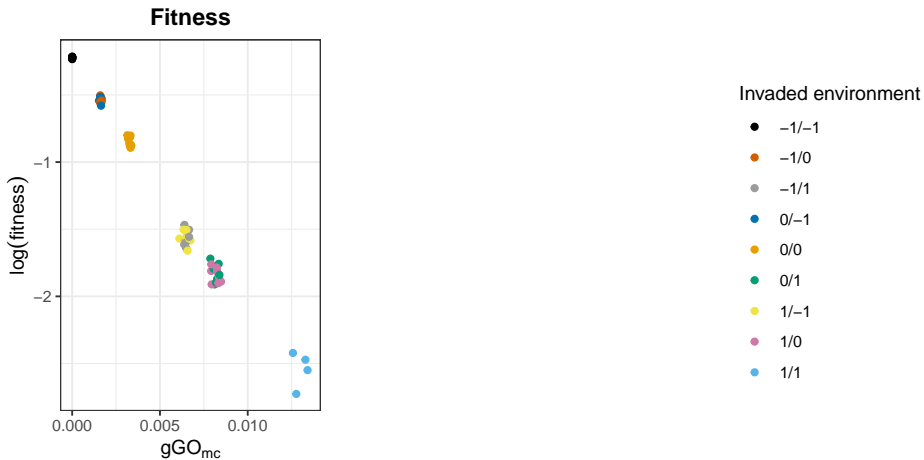
R^2 moyen entre GO et EP (10 ind., structuration forte)



Résultats similaires avec une structuration faible

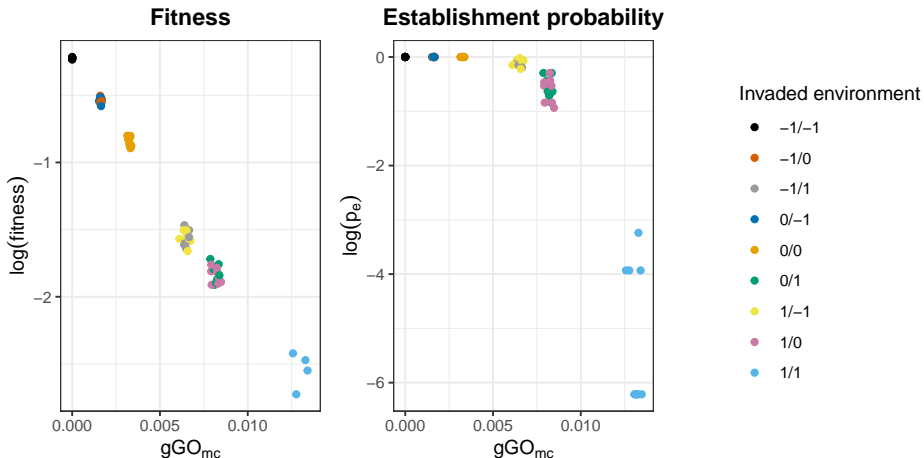
La question de l'interprétation des valeurs de GO

Avec 100 individus → lien fort avec la fitness



La question de l'interprétation des valeurs de GO

Avec 100 individus → lien fort avec la fitness
 Mais défi adaptatif "tamponné"



Conclusions

Lien entre GO et probabilité d'établissement ?

- Forte corrélation entre GO et probabilité d'établissement
- Robuste à la force de l'adaptation (=structuration) dans l'aire native
- Perturbation par la complexité environnementale et les variables confondantes

Conclusions

Lien entre GO et probabilité d'établissement ?

- Forte corrélation entre GO et probabilité d'établissement
- Robuste à la force de l'adaptation (=structuration) dans l'aire native
- Perturbation par la complexité environnementale et les variables confondantes

Performances des méthodes ?

- Utilisation de méthodes univariées ou de PCs recommandées
- Meilleures performances du GO géométrique

Conclusions

Lien entre GO et probabilité d'établissement ?

- Forte corrélation entre GO et probabilité d'établissement
- Robuste à la force de l'adaptation (=structuration) dans l'aire native
- Perturbation par la complexité environnementale et les variables confondantes

Performances des méthodes ?

- Utilisation de méthodes univariées ou de PCs recommandées
- Meilleures performances du GO géométrique

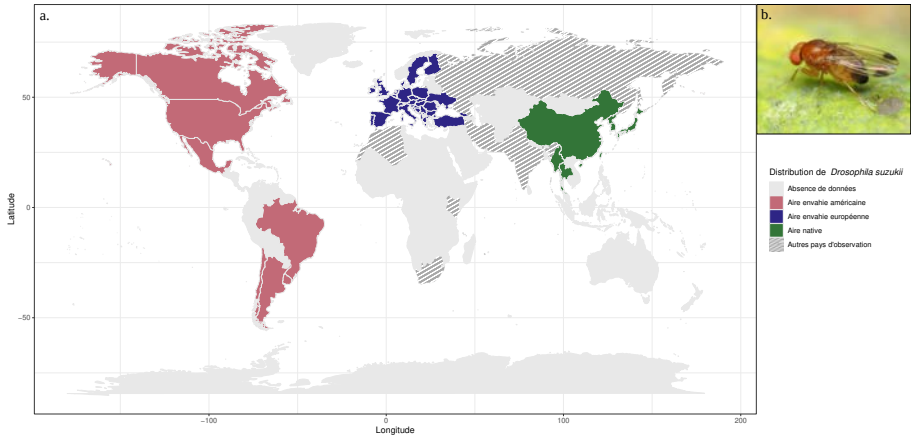
Certaines questions demeurent ...

- Nombre d'introductions successives ? Influence du mélange génétique ?
Interprétabilité des valeurs absolues de GO ?

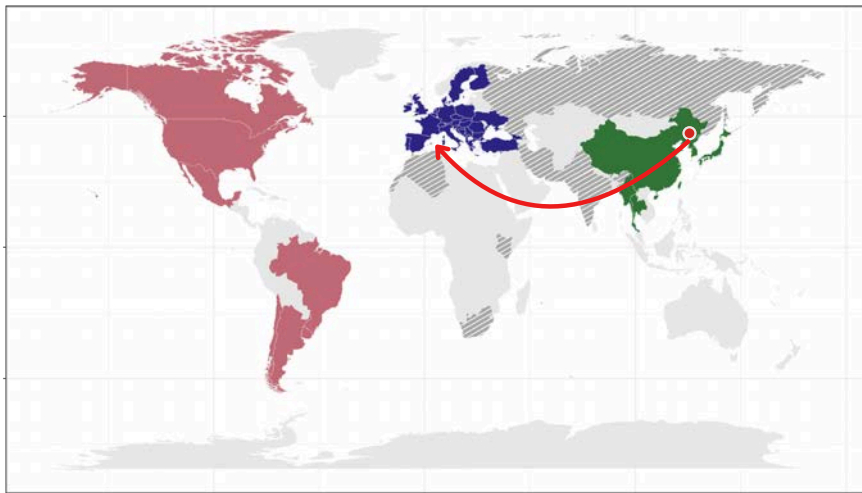
Partie 2

Le GO est corrélé à l'établissement de populations
→ Combinaison des approches de GEA et de GO pour explorer
l'adaptation aux environnements envahis chez *D. sukuzii*

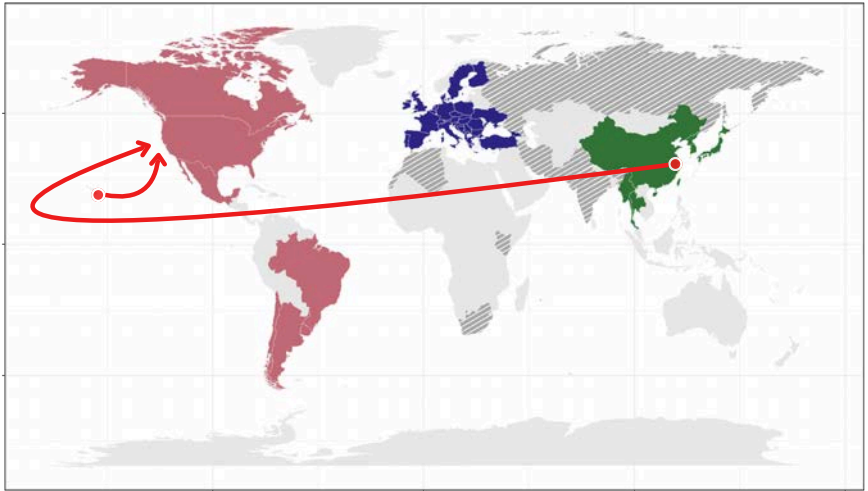
D. suzukii : une espèce envahissante...



... avec 2 invasions indépendantes... (Adrion et al., 2014, Fraimout et al., 2017)



... avec 2 invasions indépendantes... (Adrion et al., 2014, Fraimout et al., 2017)



...Et ravageuse de cultures



- Une espèce nuisible et généraliste
- Ovipositeur sclérifié : peut pondre des oeufs dans des fruits mûrs ou en train de mûrir
- Une large gamme de plantes hôtes (120 espèces reportées en Europe)

Une espèce avec une invasion rapide

Objectifs de recherche

- Comprendre l'histoire adaptative (**génétique**) de l'invasion de *D. sukukii*
- Tenter de **prédire** les régions présentant un risque d'invasion

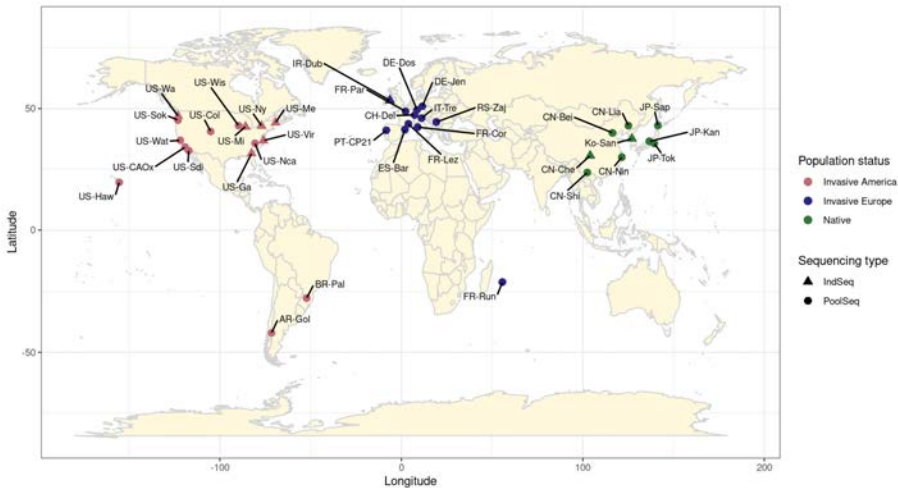
Une espèce largement étudiée

- 21 + 2 + 1 échantillons Pool-Seq (Olazcuaga et al. 2020, Feng et al. 2024, Sario et al. 2024)
- 8 échantillons Ind-Seq (Lewald et al. 2021)
- 4 échantillons Pool-Seq + 1 échantillon Ind-Seq (CBGP)

→ Analyses de données combinées avec Baypass

Nouvelles données

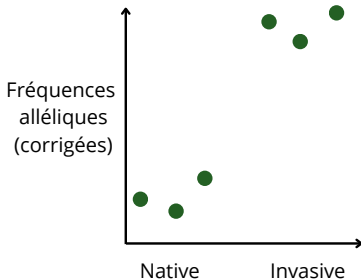
37 populations \approx 4 200 000 SNPs identifiés avec un nouvel assemblage



Analyses

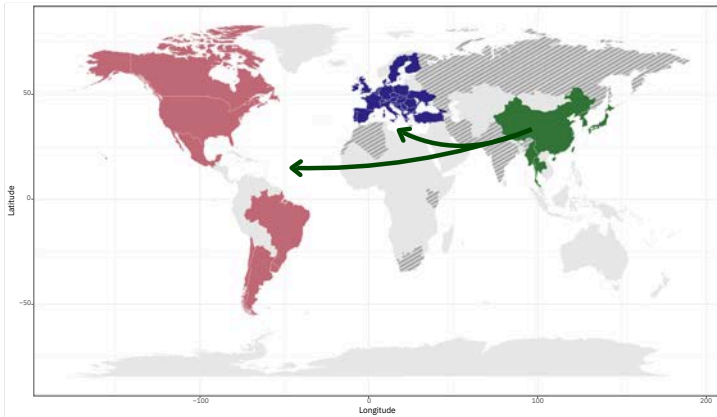
1. Identifier des régions génomiques associées au statut invasif

→ **Indice C_2** de Baypass (Olazcuaga et al., 2020)



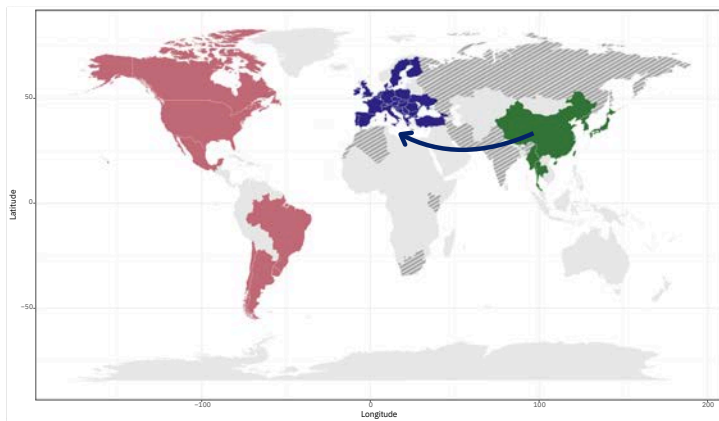
Analyses

1. Identifier des régions génomiques associées au statut invasif



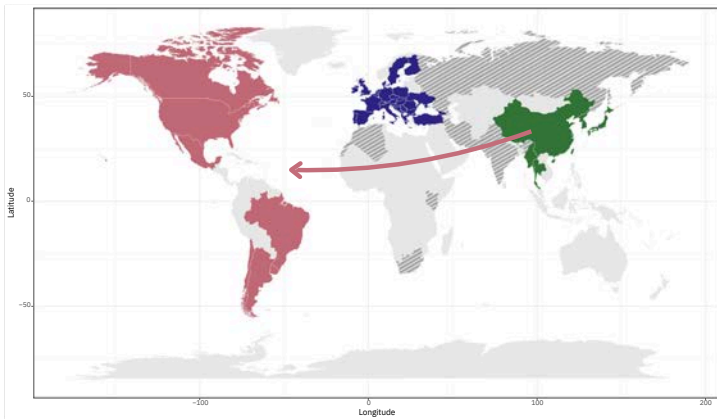
Analyses

1. Identifier des régions génomiques associées au statut invasif



Analyses

1. Identifier des régions génomiques associées au statut invasif



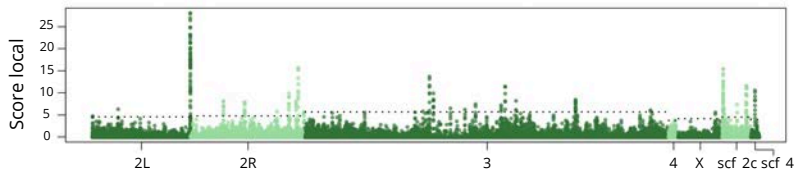
Analyses

1. Identifier des régions génomiques associées au statut invasif

- **Indice C_2** de Baypass (Olazcuaga et al., 2020) combiné avec l'approche du **score local** (Fariello et al., 2017) pour identifier les régions surdifférenciées.
- Identifier les gènes présents dans ces régions (annotation automatique du NCBI pour *D. suzukii*)
- Processus biologiques associés : gènes orthologues chez *Drosophila melanogaster*

Association avec le statut invasif

Global : 33 fenêtres



52 gènes

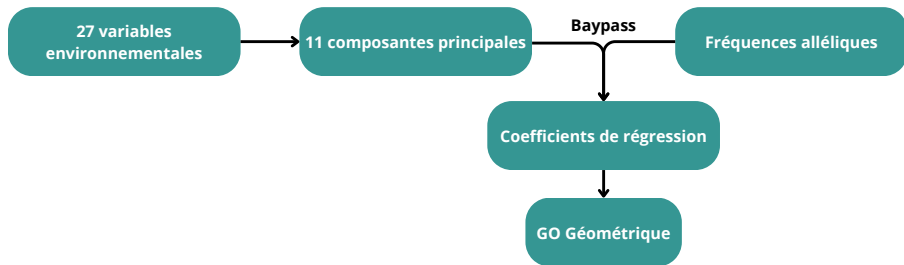
72 gènes candidats (Total des contrastes)

- Vps-13 : Résistance pesticide chez *Spodoptera litura* (Jia et al. , 2020)
- Vha36-1 : Défense immunitaire ingestion de bactéries (Vaibhvi et al. , 2022)
- Che42Bc : Perception phéromones → reproduction ? (Starostina et al. , 2009)

→ Des fonctions biologiques très variées

Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures



Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures

↔ Comprendre l'invasion passée : les aires envahies posaient-elles un défi adaptatif ?

Environnement source

Fraimout et al. (2017)

Liaoyuang, Chine



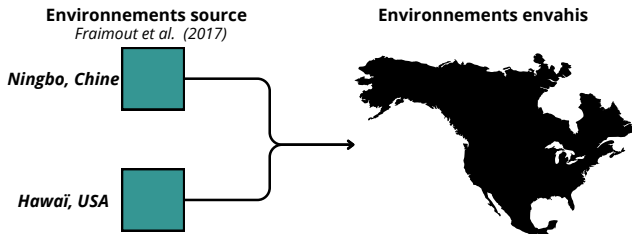
Environnements envahis



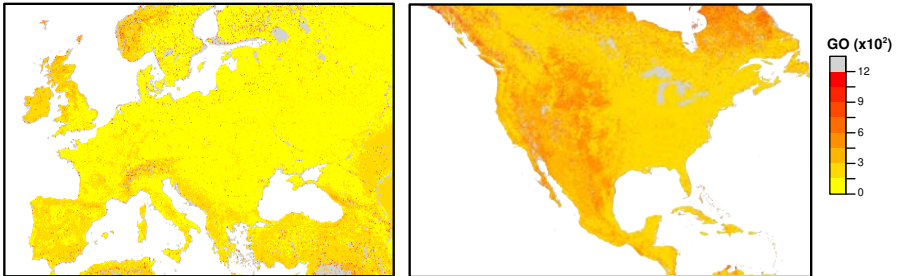
Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures

↪ Comprendre l'invasion passée : les aires envahies posaient-elles un défi adaptatif?



Invasion passée



→ Globalement, défi adaptatif faible et homogène en Europe et Amérique du Nord

Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures

↔ Prédire les invasions futures : quelles aires encore non envahies sont le plus susceptibles de le devenir ?

Environnement source potentiels

Little Rock, USA



Environnements non-envahis



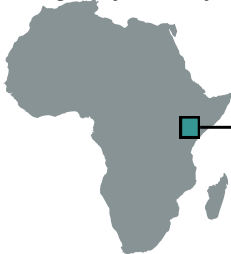
Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures

↪ Prédire les invasions futures : quelles populations sont le plus susceptibles d'envahir des régions à risque ?

Environnement source potentiels

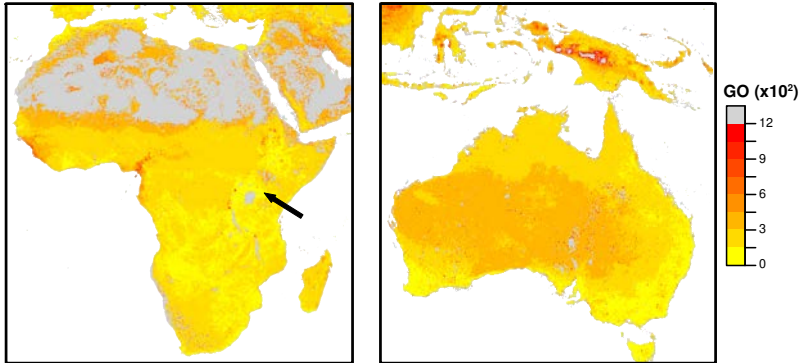
Longonot farm, Kenya



Environnements non-envahis



Future invasion



- Des zones présentant un risque de maladaptation faible
- Des valeurs cohérentes biologiquement

Analyses

1. Identifier des régions génomiques associées au statut invasif
2. GO pour comprendre l'invasion passée et prédire les invasions futures

↪ Prédire les invasions futures : quelles populations sont le plus susceptibles d'envahir des régions à risque ?

Environnements sources potentiels

USA

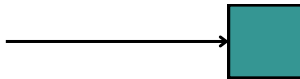


Thaïlande

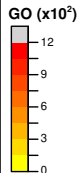
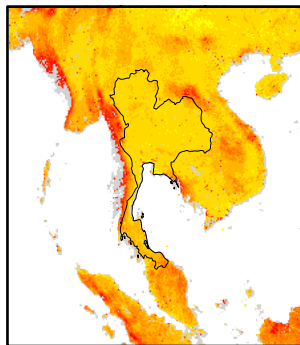
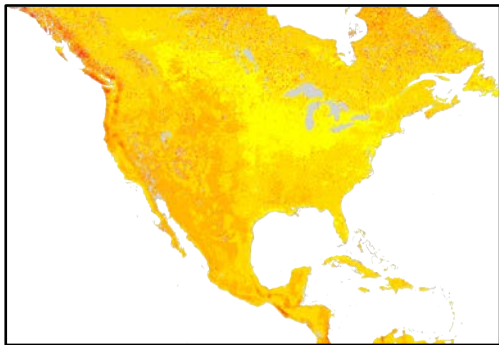


Environnement non-envahis

Port de Sidney, Australie

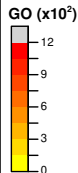
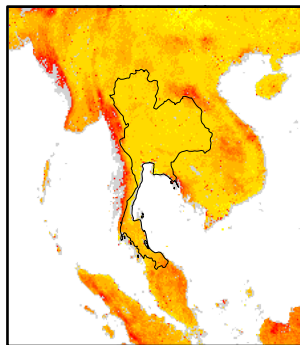
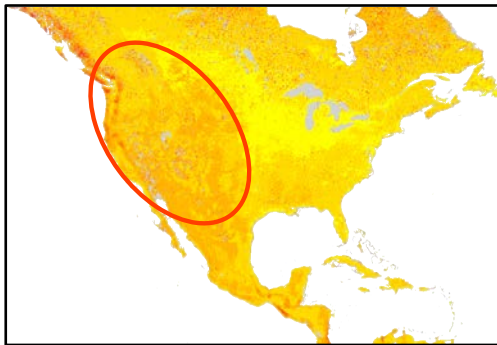


Populations sources



→ Des valeurs faibles à intermédiaires

Populations sources



- Des valeurs faibles à intermédiaires
- La Thaïlande présente un risque plus fort mais...
- Les fruits aux USA sont plus infestés

Conclusions

Une meilleure compréhension de l'invasion passée de *D. sukuzii*

- Plusieurs régions candidates associées au statut invasif → signal polygénique
- GO faibles et homogènes → pré-adaptation potentielle aux environnements envahis

Conclusions

Une meilleure compréhension de l'invasion passée de *D. sukuzii*

- Plusieurs régions candidates associées au statut invasif → signal polygénique
- GO faibles et homogènes → pré-adaptation potentielle aux environnements envahis

Le GO pour prédire les invasions futures

- Un modèle cohérent avec la biologie de l'espèce
- Des zones avec un risque de maladaptation faible

Conclusions

Une meilleure compréhension de l'invasion passée de *D. sukukii*

- Plusieurs régions candidates associées au statut invasif → signal polygénique
- GO faibles et homogènes → pré-adaptation potentielle aux environnements envahis

Le GO pour prédire les invasions futures

- Un modèle cohérent avec la biologie de l'espèce
- Des zones avec un risque de maladaptation faible

Néanmoins...

- Influence du choix de la population source ? Du mélange génétique ?
Du nombre d'individus ?

Etude du succès dans un nouvel environnement : l'exemple de *D. sukuzii*

Un succès lié à un défi adaptatif faible

- Pré-adaptation à l'environnement bioclimatique
- Pas d'adaptation génétique forte à la plante hôte (Partie 3)
→ Une absence de contrainte adaptative forte ayant pu freiner la propagation de *D. sukuzii*

Etude du succès dans un nouvel environnement : l'exemple de *D. sukukii*

Un succès lié à un défi adaptatif faible

- Pré-adaptation à l'environnement bioclimatique
- Pas d'adaptation génétique forte à la plante hôte (Partie 3)
→ Une absence de contrainte adaptative forte ayant pu freiner la propagation de *D. sukukii*

Une espèce présentant un potentiel risque d'établissement dans de nouvelles zones

Etude du succès dans un nouvel environnement : l'exemple de *D. sukukii*

Un succès lié à un défi adaptatif faible

- Pré-adaptation à l'environnement bioclimatique
- Pas d'adaptation génétique forte à la plante hôte (Partie 3)
→ Une absence de contrainte adaptative forte ayant pu freiner la propagation de *D. sukukii*

Une espèce présentant un potentiel risque d'établissement dans de nouvelles zones

Mais la génétique n'est pas le seul facteur à considérer !

- Méthodes de lutte : pesticides ?
- Disponibilité en ressources ?
- Survie des populations au long de l'année ?

Prédiction génomique et espèce envahissantes

Le GO est un outil prometteur pour étudier les invasions biologiques :

- Lien fort avec la probabilité d'établissement.
- De nombreuses utilisations possibles
- Une première étape : d'autres évaluations sont nécessaires

Prédiction génomique et espèce envahissantes

Le GO est un outil prometteur pour étudier les invasions biologiques :

- Lien fort avec la probabilité d'établissement.
- De nombreuses utilisations possibles
- Une première étape : d'autres évaluations sont nécessaires

Exploitation des simulations de la Partie 1

- Populations récemment introduites adaptées ?
- Influence du mélange génétique ?

Prédiction génomique et espèce envahissantes

Le GO est un outil prometteur pour étudier les invasions biologiques :

- Lien fort avec la probabilité d'établissement.
- De nombreuses utilisations possibles
- Une première étape : d'autres évaluations sont nécessaires

Exploitation des simulations de la Partie 1

- Populations récemment introduites adaptées ?
- Influence du mélange génétique ?

Evaluation expérimentale

- Données temporelles

Pour aller plus loin : vers une amélioration du GO

Des apports méthodologiques facilitant l'utilisation du GO

La limite de l'interprétabilité

- Comparaison relatives uniquement
- Pas d'interprétation biologique claire
- Pas de comparaison possible entre espèces

Pour aller plus loin : vers une amélioration du GO

Des apports méthodologiques facilitant l'utilisation du GO

La limite de l'interprétabilité

- Comparaison relatives uniquement
- Pas d'interprétation biologique claire
- Pas de comparaison possible entre espèces

Quelques pistes d'amélioration...

- Définition de seuil à ne pas dépasser (Lachmuth et al., 2023)
- Intervalle de crédibilité ?

Pour aller plus loin : vers une amélioration du GO

Des apports méthodologiques facilitant l'utilisation du GO

La limite de l'interprétabilité

- Comparaison relatives uniquement
- Pas d'interprétation biologique claire
- Pas de comparaison possible entre espèces

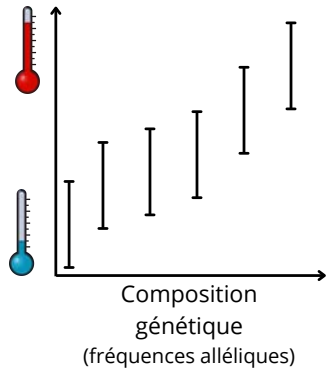
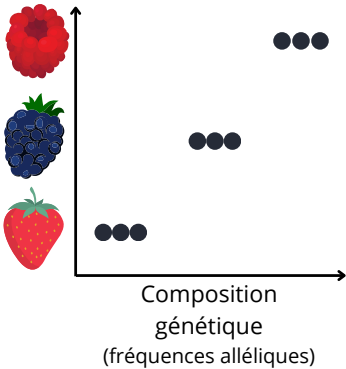
Quelques pistes d'amélioration...

- Définition de seuil à ne pas dépasser (Lachmuth et al., 2023)
- Intervalle de crédibilité ?

Le GO reste une méthode parmi d'autres !

- Prédiction de l'environnement adéquat

Pour aller plus loin : prédire l'environnement



Remerciements

Jury

Thibaut Capblancq, Frédérique Viard, Myriam Heuertz, Tristan Mary-Huard

Encadrants

Mathieu Gautier et Simon Boitard

Comité de thèse

Arnaud Estoup, Bénédicte Rhoné, Olivier François, Joëlle Ronfort, Renaud Vitalis

Merci de votre attention !

