

# Développement d'un outil pour l'analyse de la co-localisation des systèmes de défense anti-phage et des prophages dans les génomes bactériens

## Contexte scientifique :

La coévolution entre les bactéries et les virus qui les infectent, les bactériophages (ou phages), est une dynamique fascinante de l'écologie microbienne. Les bactéries ont développé divers systèmes de défense pour se protéger contre les infections phagiques (CRISPR-Cas, Restriction/Modification, etc.). Certains phages peuvent intégrer leur ADN dans les génomes bactériens (formant ainsi des prophages), y rester dormants pendant un certain nombre de générations et finalement s'exciser pour aller coloniser de nouvelles cellules hôtes. De fait, les prophages constituent des régions mobiles du chromosome bactérien et peuvent apporter des nouvelles fonctions à leur hôte. Des analyses récentes ont révélé que, souvent, les prophages contribuent à la défense de leur hôte contre d'autres infections phagiques.

Une question scientifique centrale est de savoir si les systèmes de défense anti-phage détectés dans les génomes bactériens sont situés dans les parties non mobiles du chromosome bactérien ou dans les prophages. Cette information permet de déterminer l'impact des prophages dans le transfert horizontal et l'évolution des systèmes de défense et les interactions phages-bactéries.

Des outils ont récemment été développés pour identifier les systèmes de défense (comme DefenseFinder<sup>1</sup>(python)) et pour détecter les prophages (comme geNomad<sup>2</sup> ou Phastest<sup>3</sup> (python)). Cependant, aucun outil n'intègre ces deux analyses pour permettre de vérifier la colocalisation des systèmes de défense et des prophages.

## Objectif du projet :

Développer un outil bio-informatique permettant d'identifier simultanément les systèmes de défense anti-phage et les prophages dans les génomes bactériens, et d'étudier la colocalisation potentielle entre ces deux éléments.

L'outil devra fournir une analyse synthétique montrant si les systèmes de défense se trouvent dans des régions bactériennes ou prophagiques.

## Compétences mobilisées :

- Programmation en Python ou un langage similaire.
- Utilisation et intégration d'outils bio-informatiques existants.
- Analyse génomique et traitement de données biologiques.
- Visualisation de données.

## Déroulement du projet :

### 1. Revue des outils existants et conception de l'outil

- Analyse des fonctionnalités et des sorties de DefenseFinder et geNomad.
- Conception d'un pipeline bio-informatique combinant ces outils.
- Définition des formats d'entrée et de sortie de l'outil (par exemple, fichiers FASTA pour les génomes et fichiers tabulés ou graphiques pour les résultats).

## 2. Intégration technique des outils

- Création d'un script ou pipeline en Python pour :
  - Lancer un outil (DefenseFinder) pour identifier les systèmes de défense.
  - Lancer un outil (geNomad,..) pour détecter les prophages.
  - Croiser les résultats pour déterminer les régions colocalisées.
- Gestion des dépendances logicielles et installation des outils requis.

## 3. Analyse des données et visualisation

- Développement d'un module pour visualiser les résultats :
  - Localisation des systèmes de défense et des prophages sur une carte génomique.
  - Analyse des distances entre systèmes de défense et régions prophagiques.
- Présentation des résultats sous forme de tableau ou graphique (exemple : graphiques circulaires pour les génomes, histogrammes pour les distances).

## 4. Test

Analyse d'un jeu de données composé de génomes de souches cliniques de *Pseudomonas aeruginosa*. Ces souches ont été collectées chez des patients en cas d'impasse thérapeutique souhaitant bénéficier d'une phagethérapie. Les génomes ont été séquencés et assemblés au laboratoire.

### Livrables attendus :

1. Un outil bio-informatique fonctionnel (éventuellement sous forme de package Python ou d'un script exécutable).
2. Des jeux de données tests et des résultats d'analyse sur des génomes bactériens.
3. Une documentation complète pour installer et utiliser l'outil.

### Perspectives :

L'outil développé pourrait contribuer à des études de coévolution bactérie-phage et être élargi pour analyser d'autres types d'interactions génomiques. Une mise en ligne sur une plateforme publique (comme GitHub) permettrait de le partager avec la communauté scientifique.

### Références :

1. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022 May 10;13(1):2561. doi: 10.1038/s41467-022-30269-9.
2. Camargo, A.P., Roux, S., Schulz, F. et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol (2023). doi: 10.1038/s41587-023-01953-y
3. Wishart, D.S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J., Stothard, P., Gautam, V. (2023), PHASTEST: Faster than PHASTER, Better than PHAST, Nucleic Acids Research (Web Server Issue), /doi: 10.1093/nar/gkad382