

Dynamique locale et globale du *watermelon mosaic virus*

Patrick Hoscheit, Imane Boualaoui, Hervé Lecoq, Cécile Desbiez

Journée du réseau ModStatSAP, 16 mai 2024

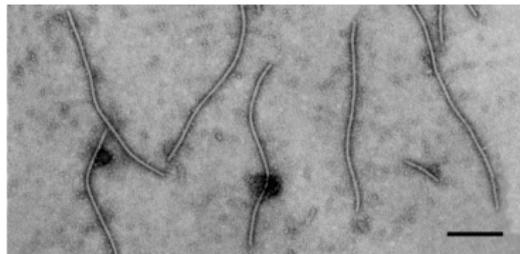


Watermelon mosaic virus

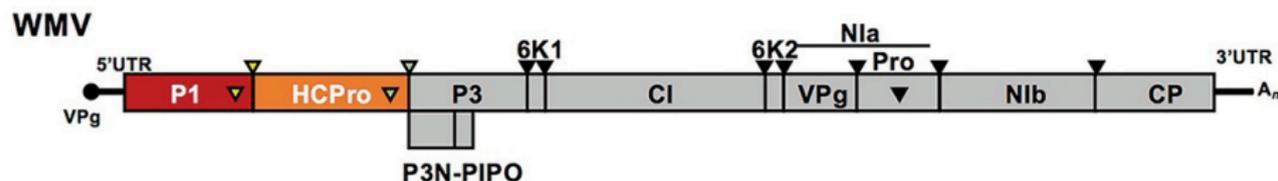
- Appartient au genre *Potyvirus* (190 espèces)
- Peut infecter la plupart des cucurbitacées et certaines légumineuses
- Transmission locale par pucerons ou par manipulation directe
- Impact économique important (30% de perte de rendement)

Endémique dans le sud de la France :

- Souches « classiques » présentes depuis plus de 30 ans
- Depuis début des années 2000, invasion par des souches « émergentes », plus virulentes



Génome de WMV



[Domingo-Calap et al. 2021]

Unique molécule d'ARN simple brin :

- Longueur $\sim 10^4$ nt
- Constituée de deux ORF
- Traduction de l'ORF majeur en une polyprotéine, clivée en 10 protéines, notamment :
 - ▶ Protéine CI (*Cylindrical Inclusion*), ~ 1900 nt
 - ▶ Protéine N1b (ARN-polymérase), ~ 1550 nt
 - ▶ Protéine CP (*Coat Protein*), ~ 830 nt
- Architecture générale partagée par tous les potyvirus

Phylogéographie continue

Données : Séquences de virus, avec date et lieu de collecte

- Reconstruction de phylogénies datées et spatialisées
- Hypothèse : lignées virales se propagent suivant un mouvement brownien avec un taux de diffusion dépendant de la branche
- Estimation dans un cadre bayésien :

$$P(\text{Genetic sequences} \mid \text{Genealogy} \text{ Demographic model} \text{ Substitution model} \text{ Molecular clock model}) = \frac{P(\text{Genealogy} \mid \text{Genetic sequences} \text{ Demographic model} \text{ Substitution model} \text{ Molecular clock model}) P(\text{Genetic sequences} \mid \text{Genealogy}) P(\text{Demographic model}) P(\text{Substitution model}) P(\text{Molecular clock model})}{P(\text{Genealogy})}$$

| Key: | | | | |
|---|---|---|---|---|
|  |  |  |  |  |
| Genetic sequences | Genealogy | Demographic model | Substitution model | Molecular clock model |

- Estimation jointe de l'arbre et des paramètres du modèle : horloge moléculaire, modèle de substitution, **diffusion spatiale**

Invasion par EM-1, EM-2, EM-4

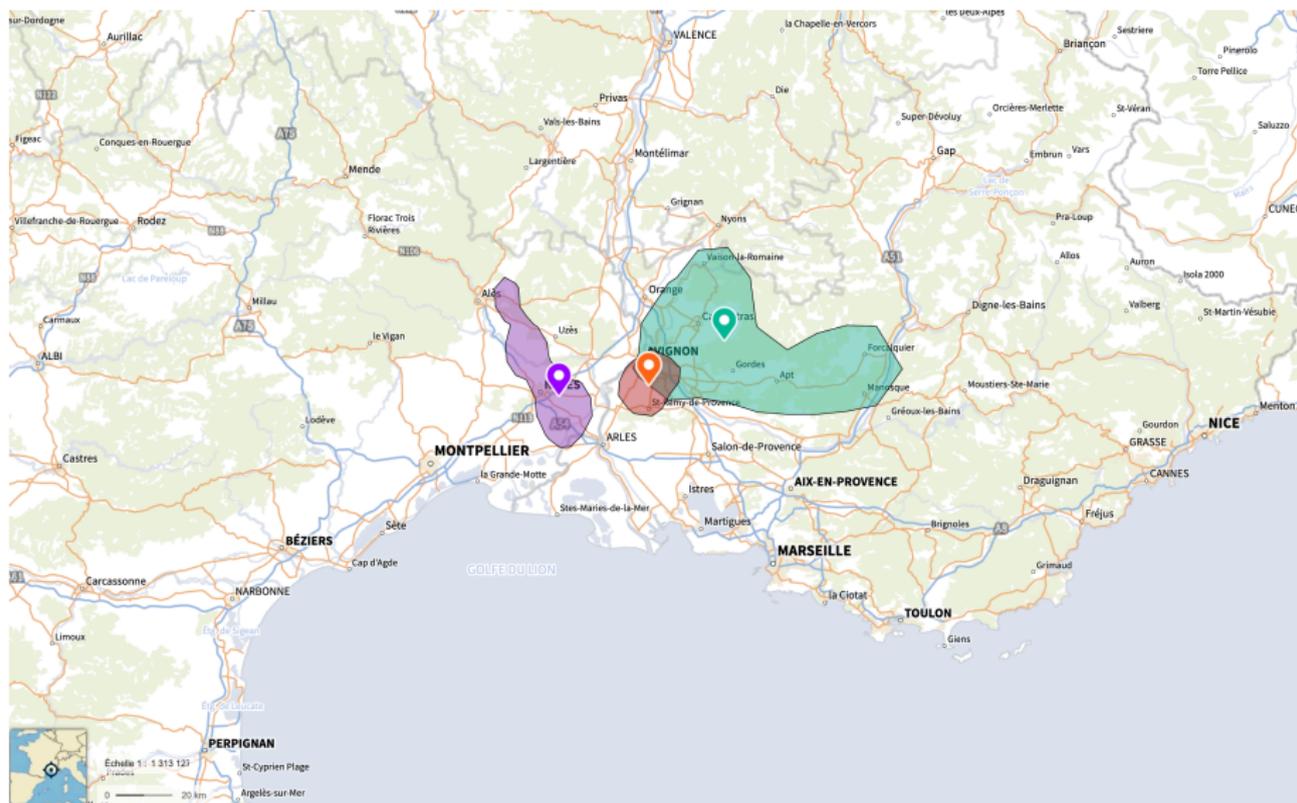
Analyse phylogéographique continue d'échantillons des gènes CI et CP

- Resp. 191 (EM-1), 79 (EM-2) et 147 (souche EM-4) séquences
- Longueurs variables, de 600 à 1000nt
- Prélèvements en 2004-2008 et en 2016-2017 dans le sud de la France
- Géolocalisations au niveau de la commune
- Modèle de diffusion inhomogène, prior uniforme sur les communes pour les localisations ([Nylinder et al. 2014, Dellicour et al. 2020])

Résultats :

- Introductions autour de 1995 pour les trois lignées
- Diffusion plus rapide pour EM-1 (6.9 km/an) et EM-4 (6.8 km/an) que pour EM-2 (5.05 km/an) ⇒ Fitness intrinsèque? Paysage?
- Extinction des lignées « classiques » en 2016-2017
- Résultats analogues à [Roques et al. 2021]

Lieux d'introduction



Diversité globale de WMV

Classification fondée sur le génotypage CP :

- Distinction traditionnelle entre souches classiques (G1 et G2) et émergentes (G3, subdivision en sous-groupes EM-1 à EM-4)
- Observation de nouvelles lignées en 2016-2017 + génomes complets
⇒ nouvelle classification requise

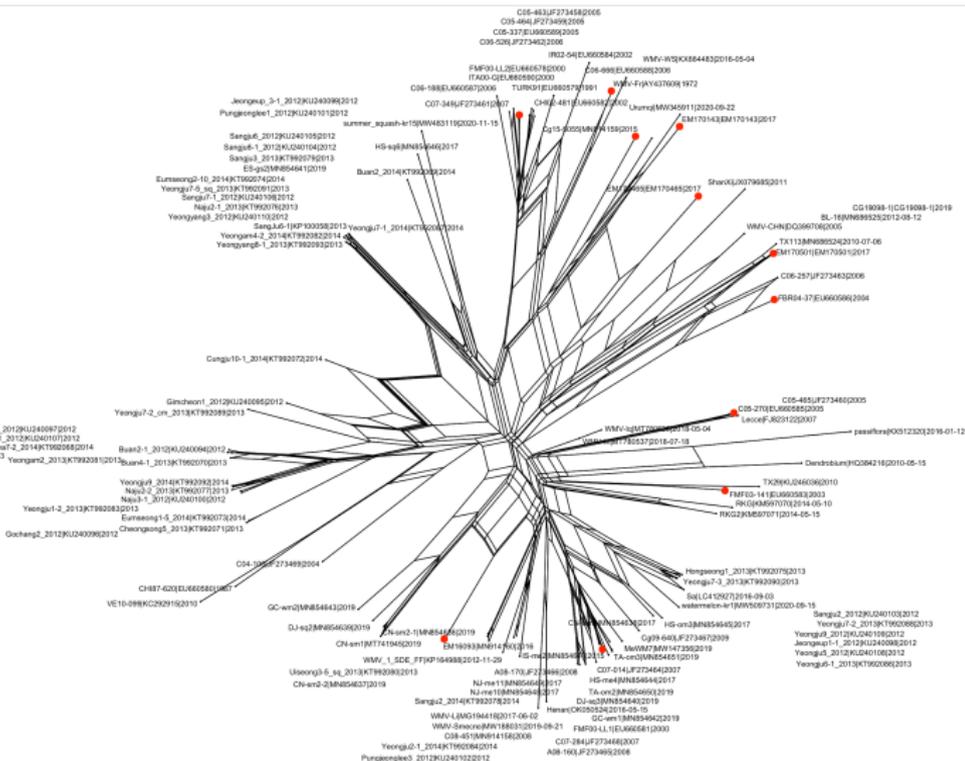
Données disponibles :

- 113 séquences complètes sur Genbank + 4 non-publiées = 117 échantillons datés et géolocalisés au niveau du pays
- Origine : France (28), Chine (9), Corée du Sud (64)
- 11 souches présentes en France (EM-1 à EM-4 + 7 nouvelles)

Principale difficulté : présence de très nombreuses recombinaisons

Représentation graphique

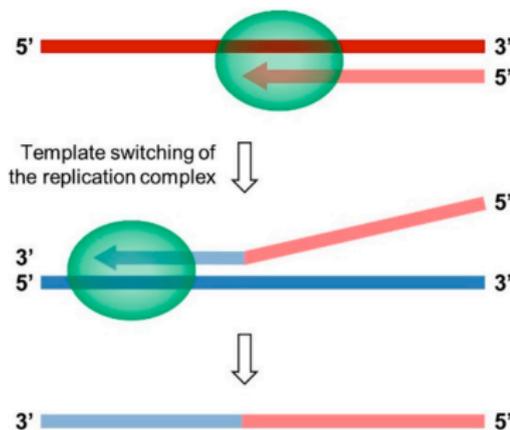
0 0.01 0.02 Fit: 99.15



Recombinaison des virus à ARN

Phénomène omniprésent dans les virus à ARN (surtout +ssRNA) :

- Recombinaison homologue par changement de brin modèle (*template switching*)
- Réassortiment pour virus à génomes segmentés
- Requiert la co-infection de l'hôte par deux ou plus pathogènes différents
- Moteur important de l'évolution



[Muslin et al. 2019]

Estimation bayésienne d'ARG

Utilisation du package `Recombination` ([Müller et al. 2022])

- Inférence sous le coalescent avec recombinaison ([Hudson, Kaplan 1988])
- Jeu de données réduit : sélection d'un sous-ensemble optimal conservant 95% de la diversité phylogénétique \Rightarrow 61 séquences
- Temps de calcul significatif (\sim 1 mois)
- Modèle d'horloge stricte à taux fixe 4×10^{-4} s/s/a
- Synthèse d'un échantillon *posterior* de réseaux phylogénétiques par maximisation de crédibilité de clade

Résultats de l'estimation :

- Grand MRCA autour de 1700
- Médiane de ~ 250 événements de recombinaison estimés
- Taux de recombinaison ~ 0.07 recombinaisons/lignée/an
- Diversité mondiale bien représentée en France
- Posterior correctement explorée par l'algorithme
- Forte incertitude topologique : division CL-EM mal supportée (posterior $< 5\%$ sur tout le génome)

Perspectives

Pour WMV :

- Nouvelles définitions de souches nécessaires
- Immenses lacunes dans les données (Moyen-Orient, Asie Centrale)

Méthodologie :

- Passage à l'échelle nécessaire, mais intractable actuellement
- Modèles phylogéographiques avec recombinaison (SCAR, ScoRe)
- Prise en compte de la recombinaison dans les analyses interspécifiques (intégration du modèle PoW?)
- Représentation et visualisation (nextstrain)
- Rôle important de recombinaison/réassortiment pour l'émergence virale (H5N1 bovine aux USA)

Prochaine rencontre :

- Quand? Jeudi 14 novembre 2024
- Où? Institut des sciences complexes, Paris (Ici)
- Qui? Toutes les personnes intéressées par approches moléculaires en épidémiologie, pour maladies des animaux et des plantes
- Et ensuite? *Workshop* le 15 novembre, thème à définir (contributions bienvenues)