



**HAL**  
open science

## Dynamique locale et globale du watermelon mosaic virus

Patrick Hoscheit, Imane Boualaoui, Hervé Lecoq, Cécile Desbiez

► **To cite this version:**

Patrick Hoscheit, Imane Boualaoui, Hervé Lecoq, Cécile Desbiez. Dynamique locale et globale du watermelon mosaic virus. Rencontres du GdR Phylodynamique, Groupe de Recherche (GDR) de phylodynamique, Jul 2023, Lyon, France. hal-04227983

**HAL Id: hal-04227983**

**<https://hal.inrae.fr/hal-04227983>**

Submitted on 4 Oct 2023

**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.

# Dynamique locale et globale du *watermelon mosaic virus*

Patrick Hoscheit, Imane Boualaoui, Hervé Lecoq, Cécile Desbiez

Journées du RT Phylodynamique

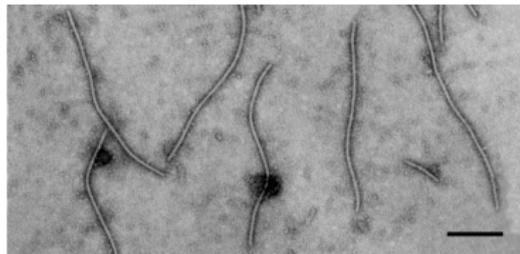


## Watermelon mosaic virus

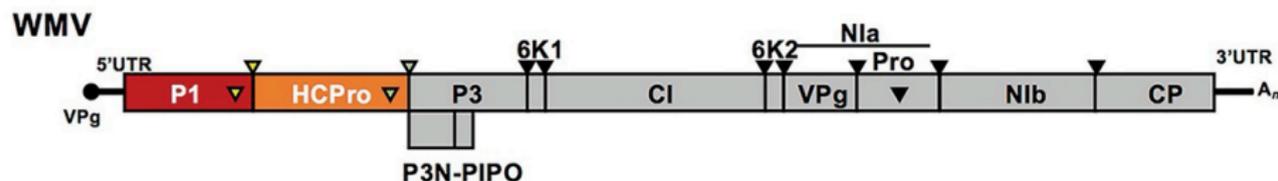
- Appartient au genre *Potyvirus* (190 espèces)
- Peut infecter la plupart des cucurbitacées et certaines légumineuses
- Transmission locale par pucerons ou par manipulation directe
- Impact économique important (30% de perte de rendement)

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*),  $\sim 1900$ nt
  - ▶ Protéine NIB (ARN-polymérase),  $\sim 1550$ nt
  - ▶ Protéine CP (*Coat Protein*),  $\sim 830$ nt
- Architecture générale partagée par tous les potyvirus

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

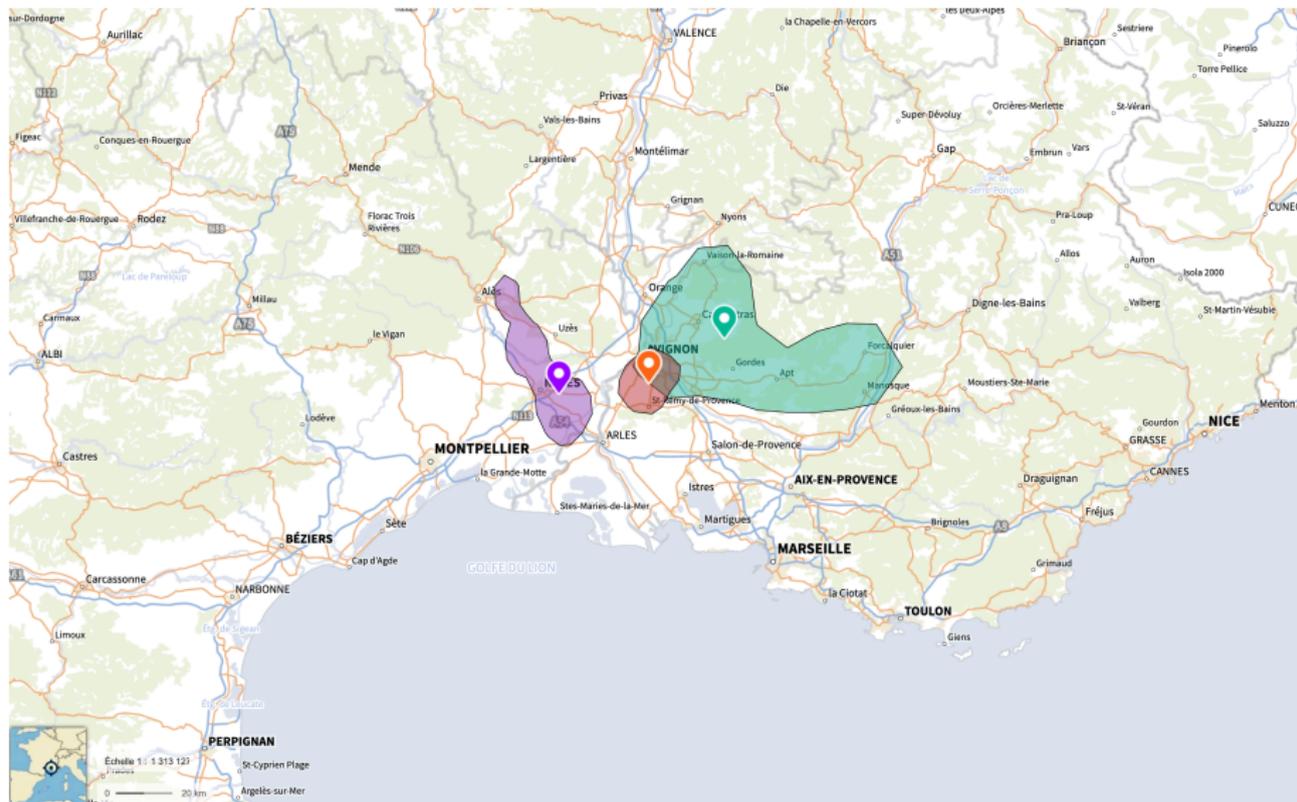
Analyse phylogéographique continue d'échantillons des gènes CI et CP

- 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)

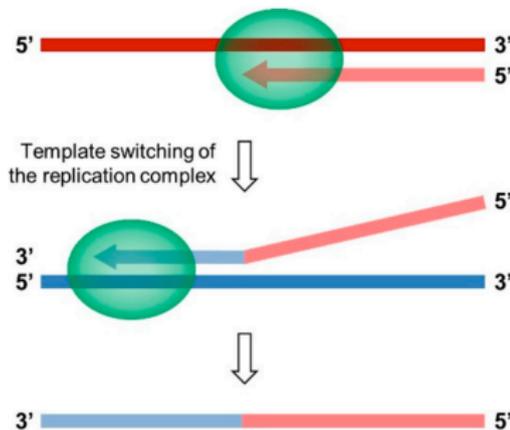
Principale difficulté : présence de très nombreuses recombinaisons



# 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
- WMV lui-même issu d'une recombinaison interspécifique



[Muslin et al. 2019]

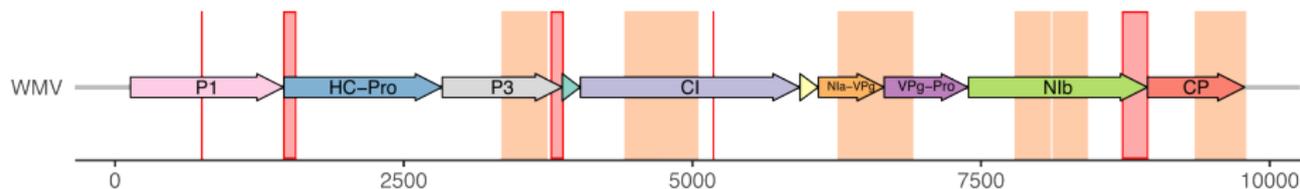
# Approche classique

## Détection de recombinaison par méthodes statistiques

- Littérature bien établie : 3SEQ, RDP, GENECONV, GARD,...
- Méthodes regroupées dans RDP5 [[Martin et al. 2021](#)]
- Sélection d'événements détectés par  $> 4$  méthodes différentes
- Permet de supprimer les séquences recombinantes ou de délimiter des régions du génome sans recombinaison

## Résultats pour les 117 séquences complètes de WMV :

- 98 événements distincts
- Seulement 2 séquences identifiées comme non-recombinantes
- 8 régions non-recombinantes  $> 250$ nt



# Approche classique

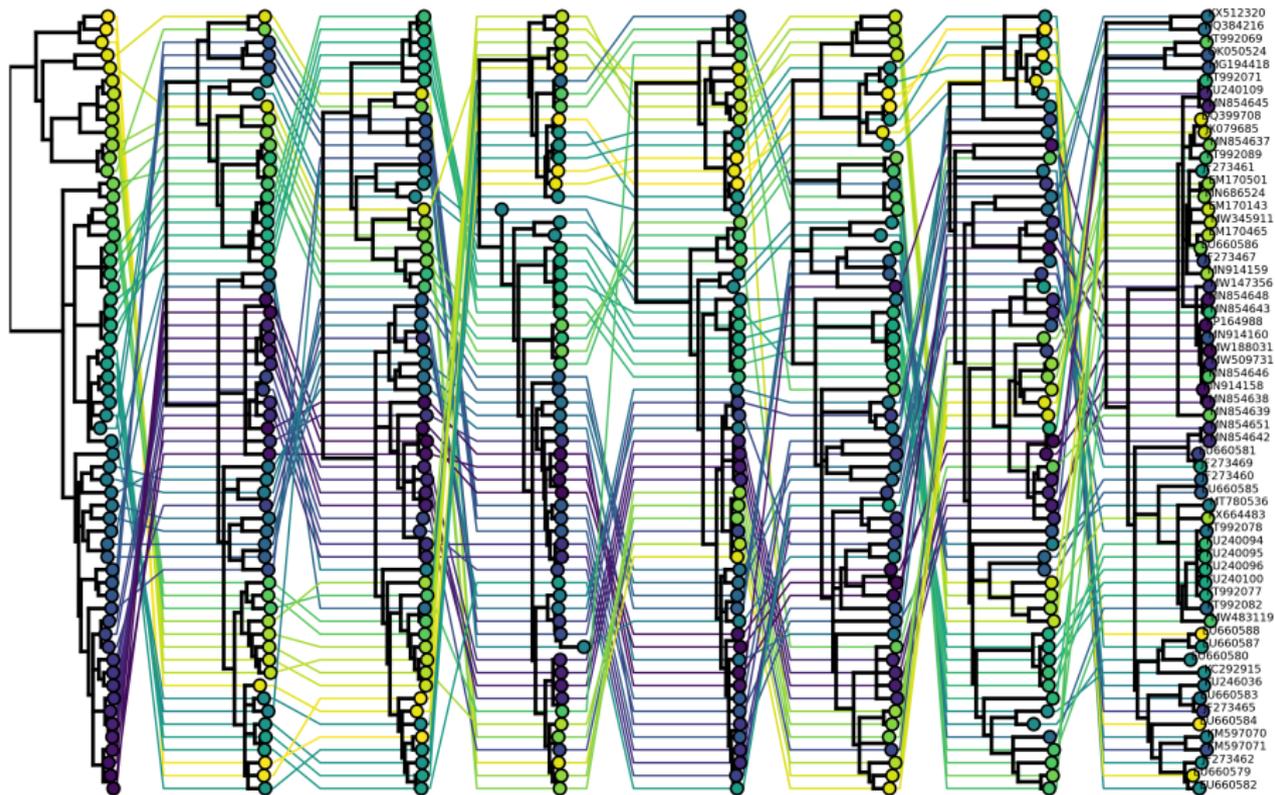
Estimation d'arbres pour chacun des 8 segments :

- Modèle d'horloge (relâchée) couplée :  $3.81 \times 10^{-4}$  s/s/y
- MRCA locaux estimés entre 1757 et 1854

Inconvénients :

- Ne couvre qu'une partie limitée ( $\sim 30\%$ ) du génome
- Signal phylogénétique limité sur chaque segment
- Ne permet pas d'étudier directement les événements de recombinaison

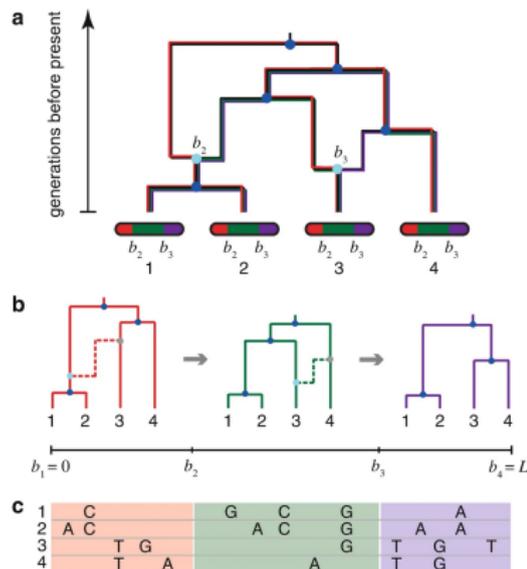
# Enchevêtrogramme



# Reconstruction d'ARG

Graphes de recombinaison ancestral (ARG) :

- Ensemble d'arbres locaux non-congruents (point de vue de [Wiuf, Hein 1999])
- Techniques issues de la génétique humaine et microbienne
- Nombreux algorithmes, en particulier argweaver [Rasmussen et al. 2014] et tsinfer [Kelleher et al. 2019]
- Ne prennent pas en compte la structure temporelle des échantillons, ni des modèles sophistiqués d'horloge et de substitution



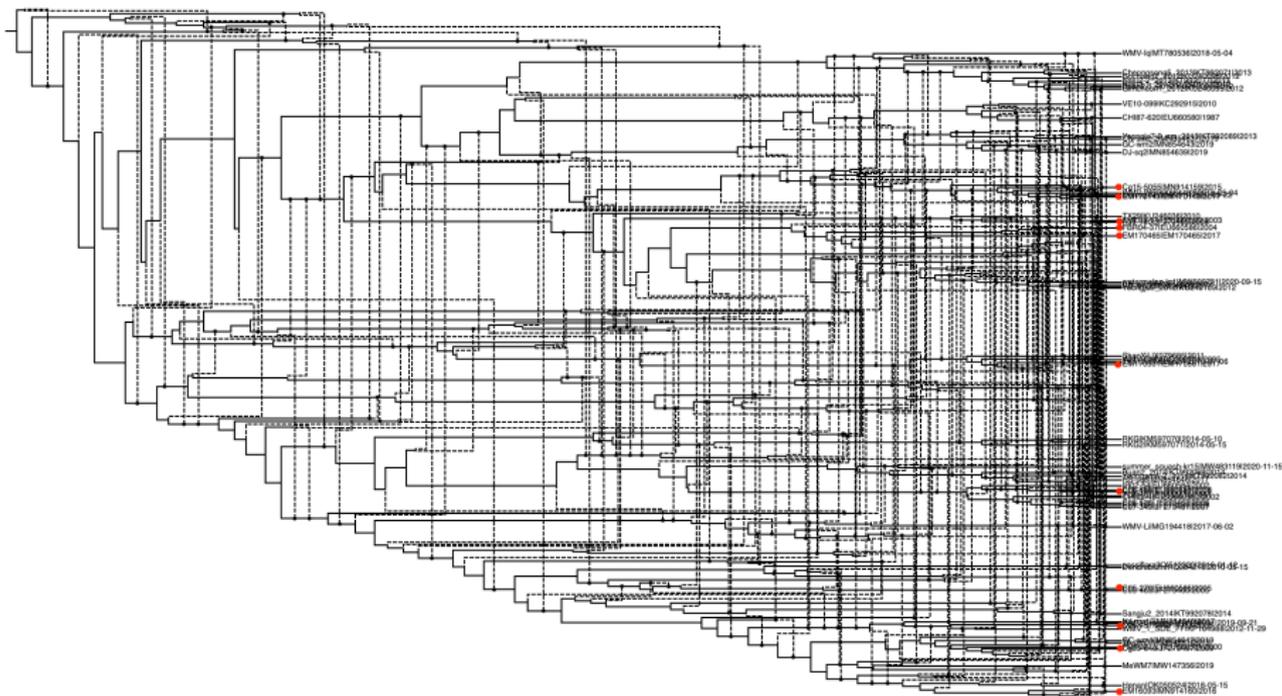
[Hubisz, Siepel 2020]

# 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  $4 \times 10^{-4}$  s/s/y
- Synthèse d'un échantillon *posterior* de réseaux phylogénétiques par maximisation de crédibilité de clade

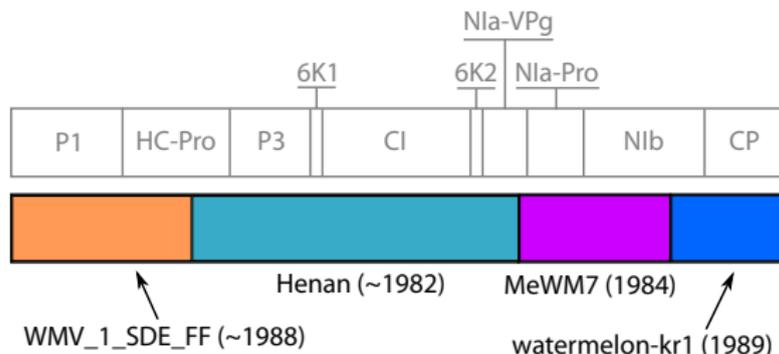
# Réseau MCC



# Synthèse

## Résultats de l'estimation :

- Grand MRCA autour de 1700
- ~ 250 événements de recombinaison estimés
- Permet l'identification de lignées non échantillonnées
- Décomposition des séquences contemporaines en fonction de la plus proche séquence locale
- Par exemple, probable origine asiatique de EM160093 (France, 2016) :



- Diversité mondiale bien représentée en France

# Perspectives

Pour WMV :

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

Méthodologie :

- Modèles phylogéographiques avec recombinaison (SCAR, ScoRe)
- Intégration de covariables (réseaux commerciaux? paysage?)
- Prise en compte de la recombinaison dans les analyses interspécifiques (intégration du modèle PoW?)
- Moteur d'estimation ad hoc (non-BEAST)
- Représentation et visualisation (nextstrain)

# Perspectives

Pour WMV :

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

Méthodologie :

- Modèles phylogéographiques avec recombinaison (SCAR, ScoRe)
- Intégration de covariables (réseaux commerciaux? paysage?)
- Prise en compte de la recombinaison dans les analyses interspécifiques (intégration du modèle PoW?)
- Moteur d'estimation ad hoc (non-BEAST)
- Représentation et visualisation (nextstrain)

Prochaine rencontre du groupe PhyloMAP : mardi 7 novembre 2023,  
Institut des sciences complexes, Paris