

# Can genomic data provide anything useful to help us understand the ecology of *Xylella fastidiosa* diseases?

Rodrigo P.P. Almeida



DEPARTMENT of ENVIRONMENTAL  
SCIENCE, POLICY, AND MANAGEMENT

**Berkeley**  
UNIVERSITY OF CALIFORNIA

College of  
Natural Resources

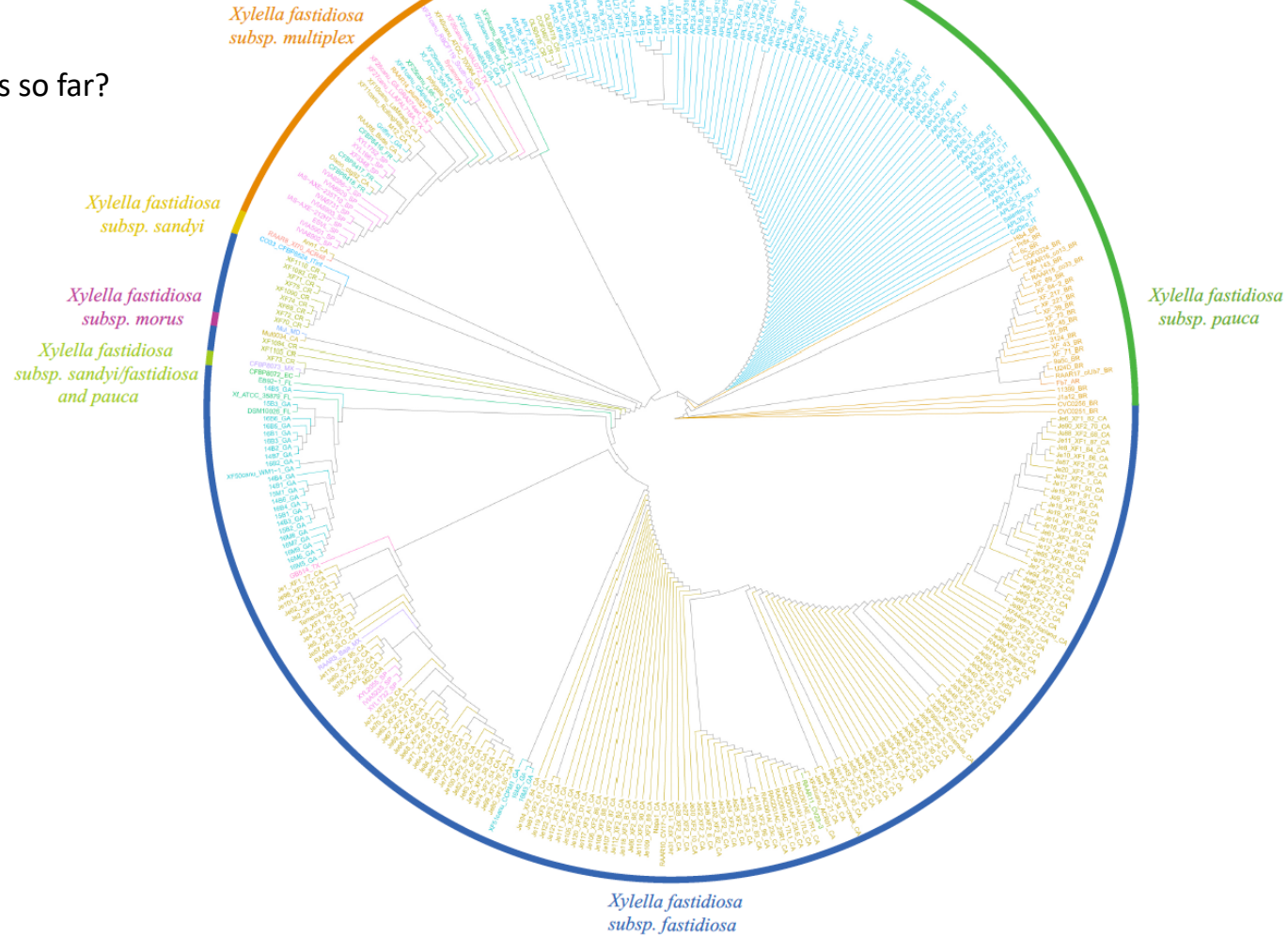
Anne Sicard  
Mathieu Vanhove  
Andreina Castillo Siri  
Alexandra Kahn

Drake Stenger  
Leo De La Fuente  
Maria Saponari  
Annalisa Giampetruzzi

Helvecio Coletta-Filho  
Carlos Chacon Diaz  
Blanca Landa  
and many others

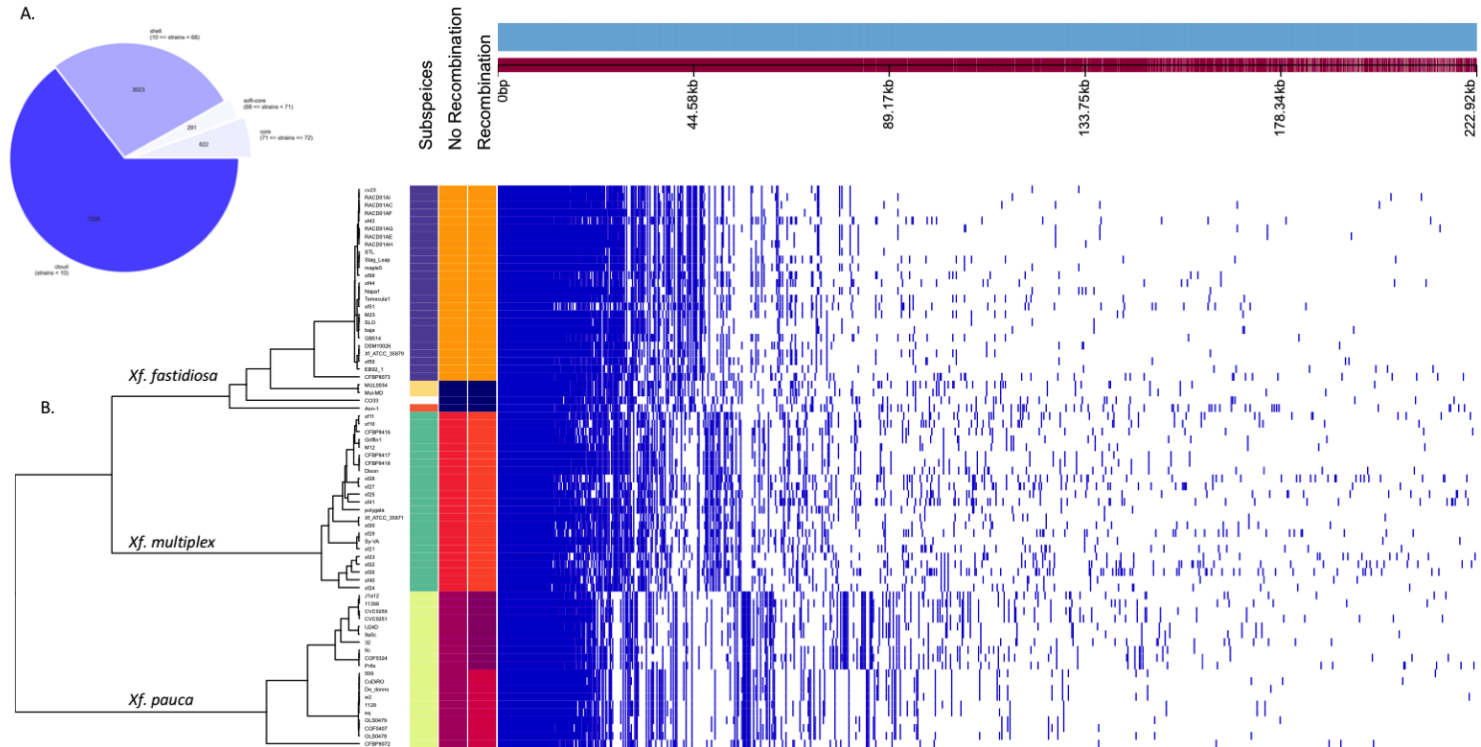


What have genomic data told us so far?

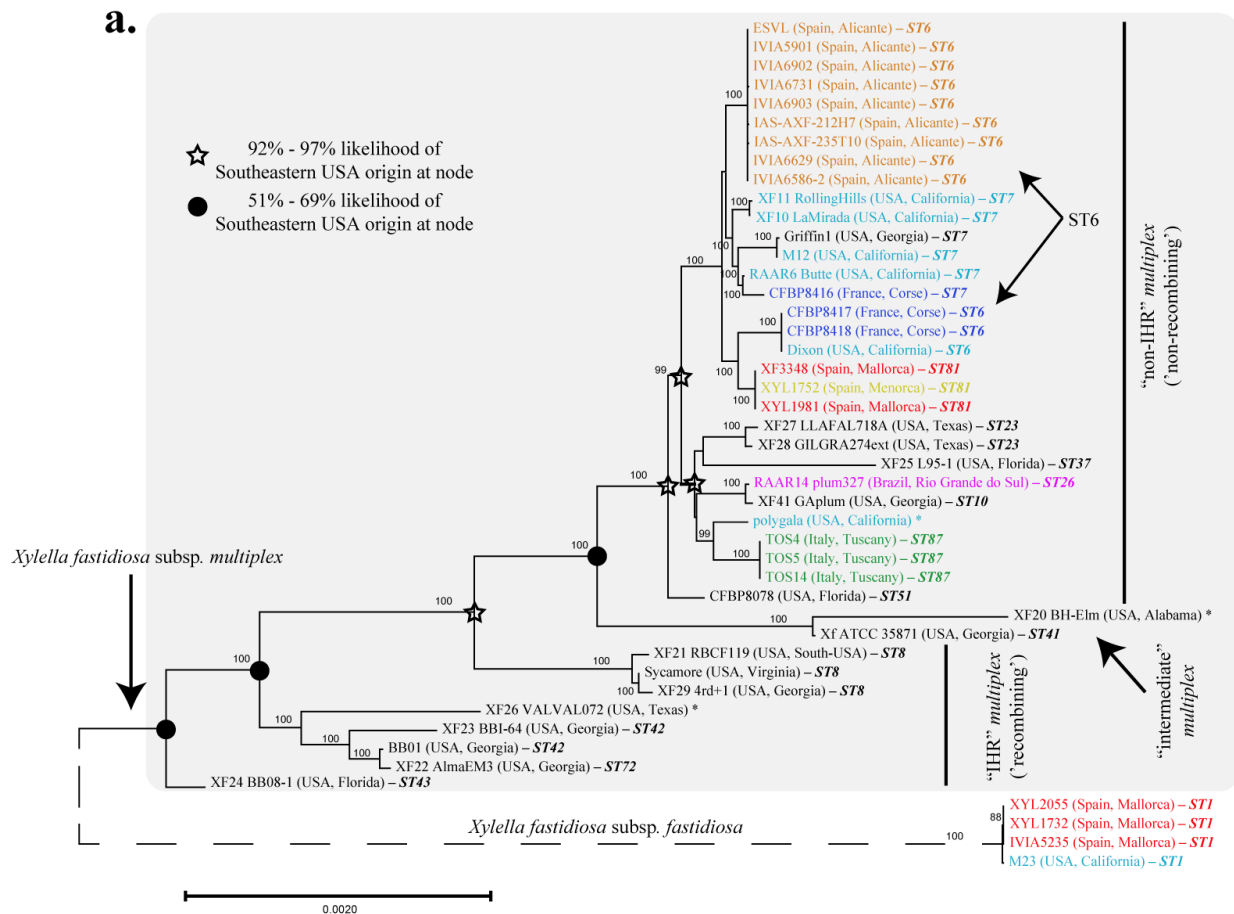




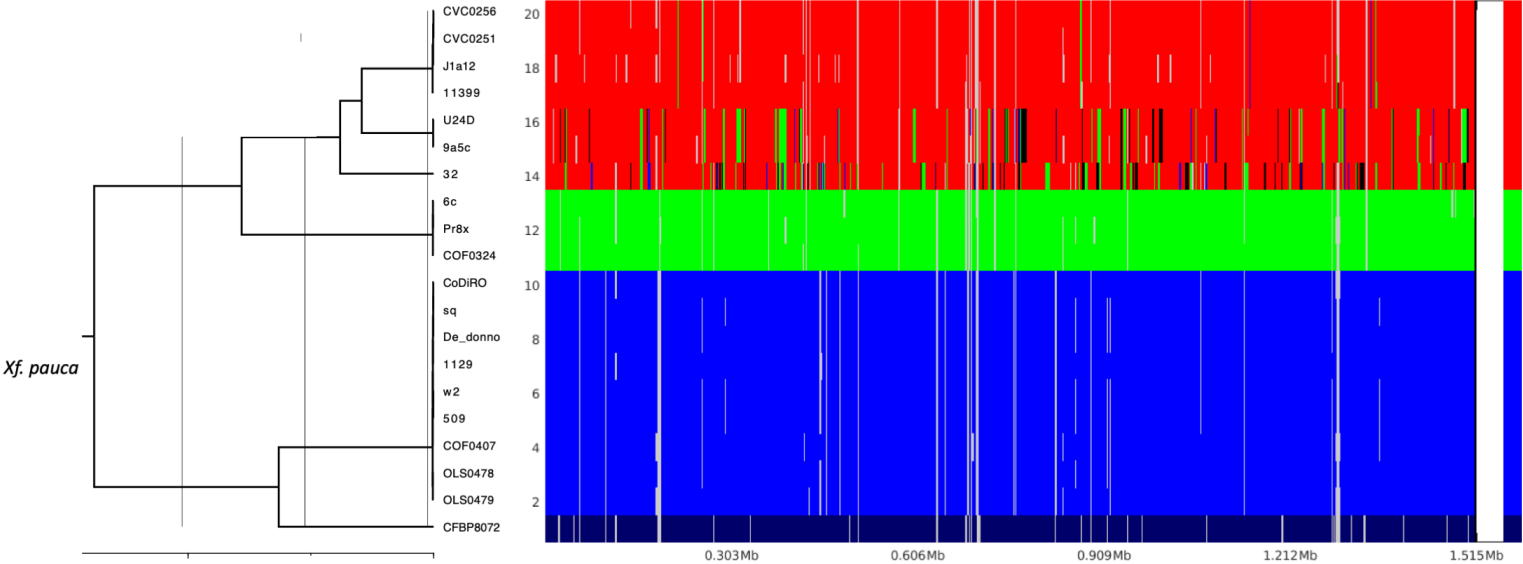
# *Xylella fastidiosa* is very diverse



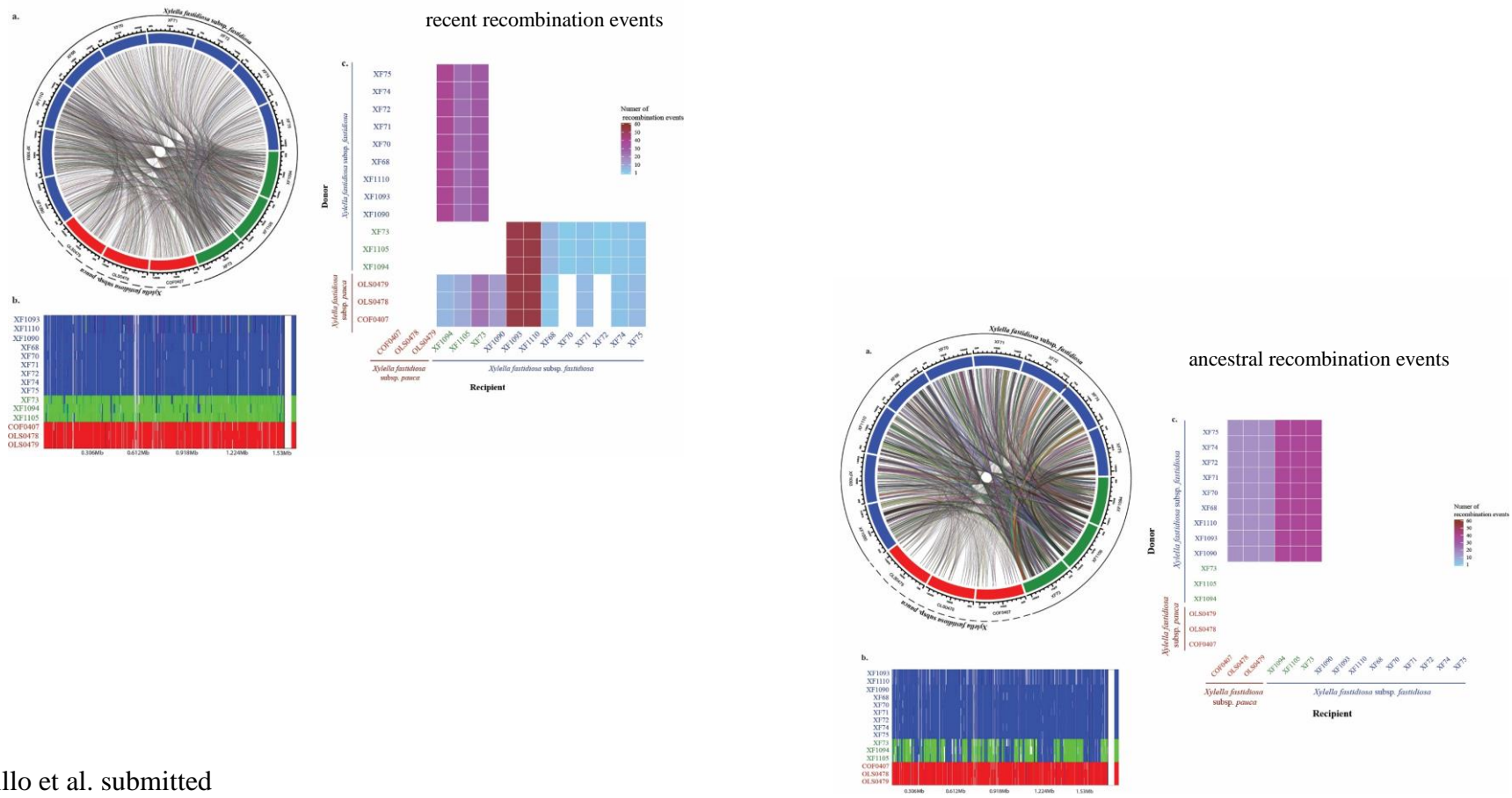
# Genomes provide useful phylogenetic resolution to study dispersal pathways



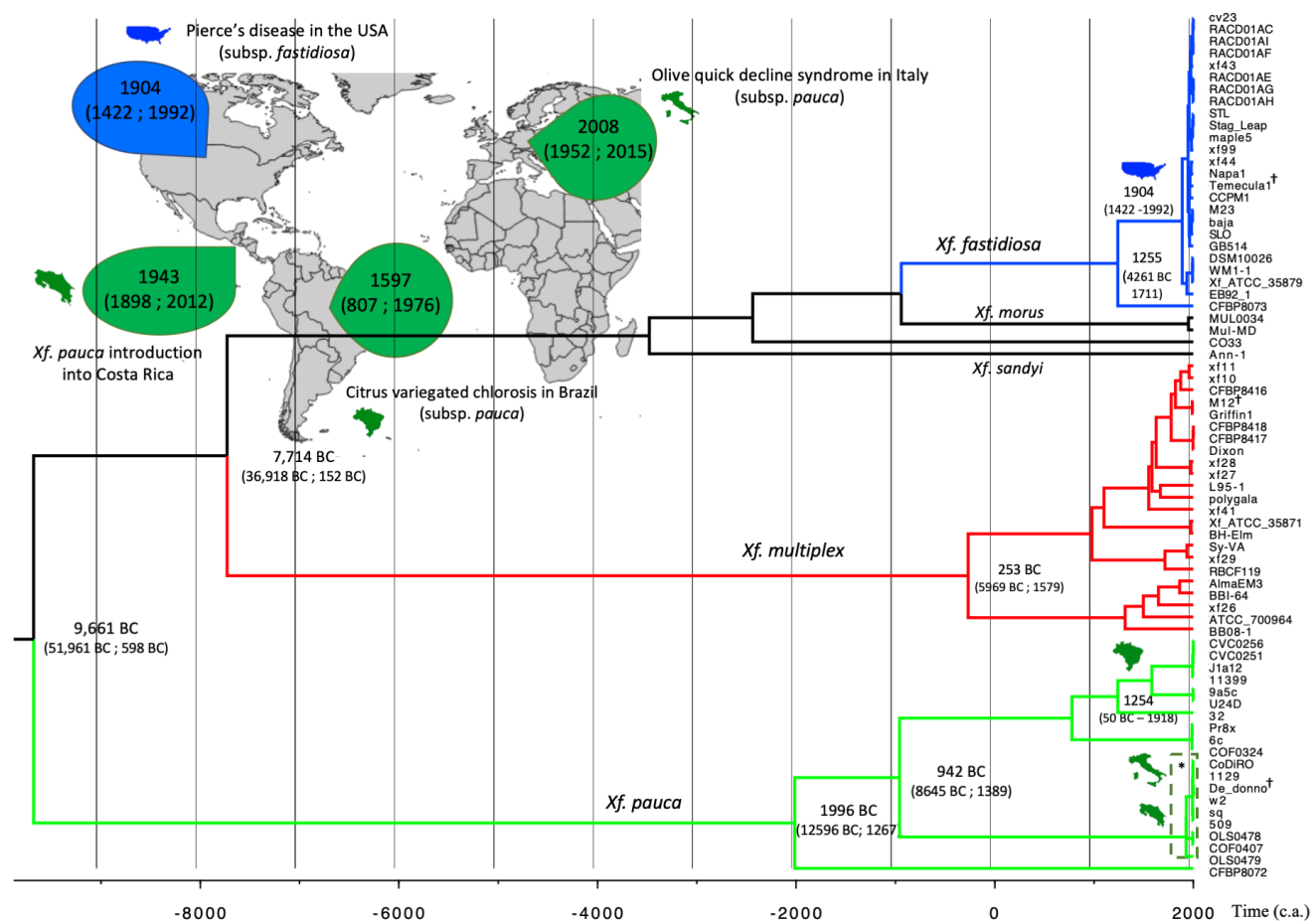
Evidence of virulence linked to clades with more recombination events



# Recombination patterns are not simple, and population history and ecology are important

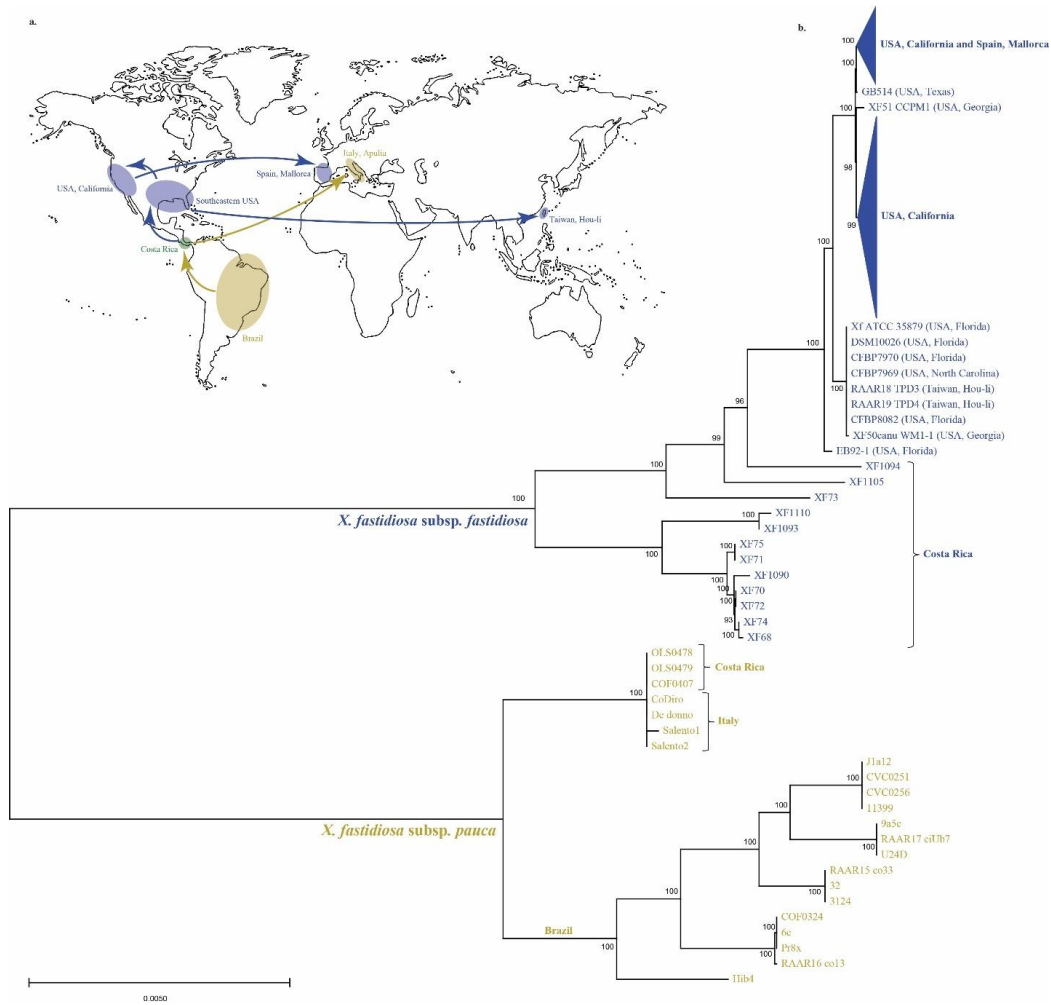


# Reconstructing epidemiological history – species level

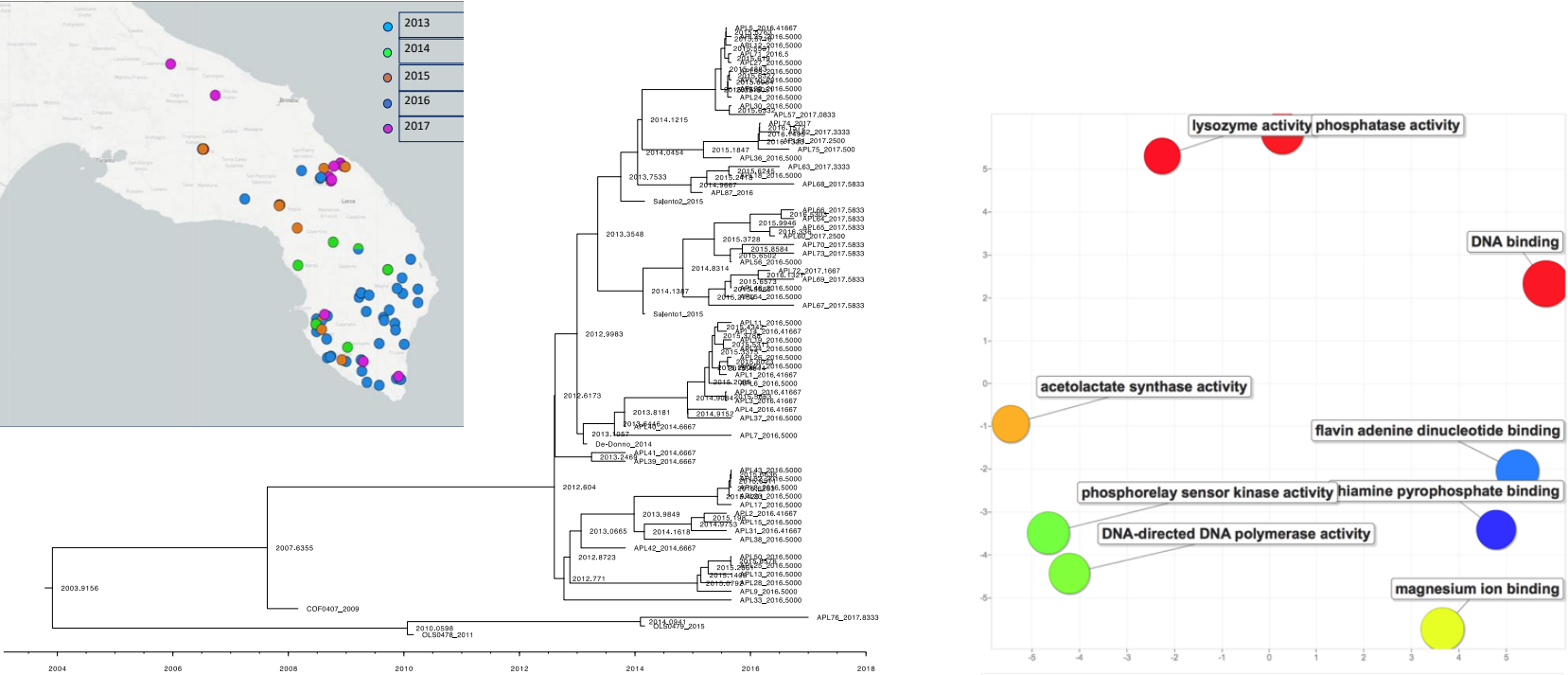




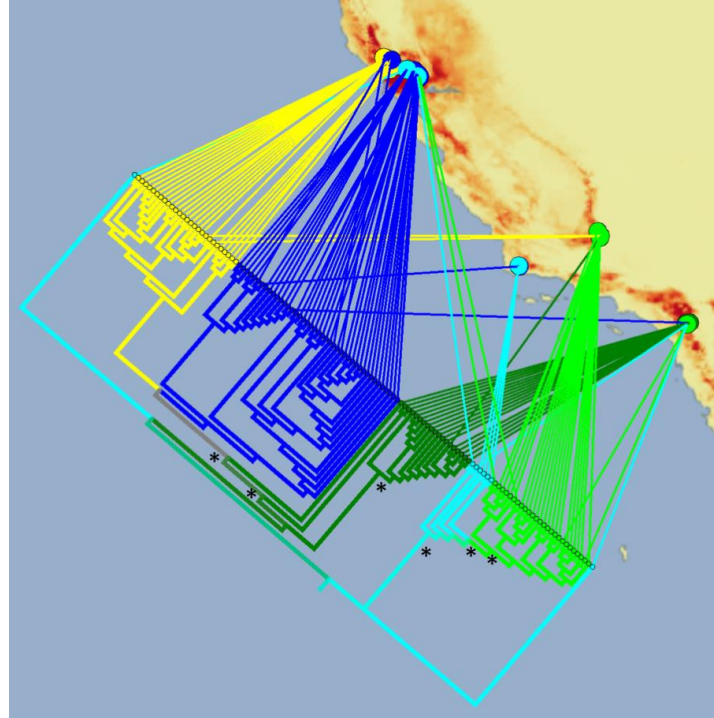
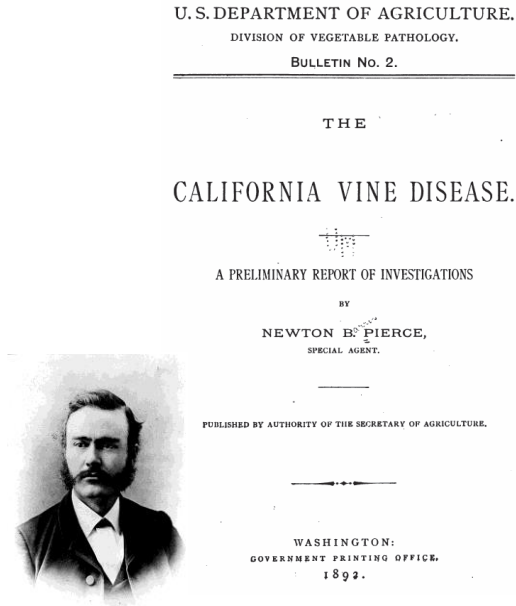
## Castillo et al. submitted



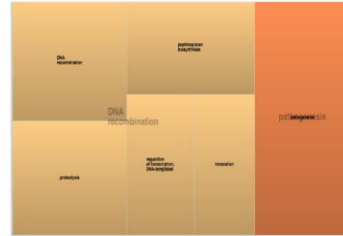
# Reconstructing epidemiological history – the case of Apulia



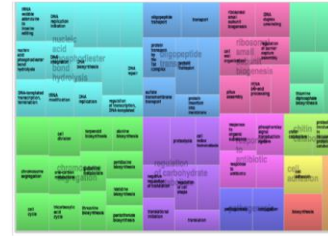
# Adaptation of a population after invasion - insights



# Different populations appear to be under different selective pressures



Common GO term



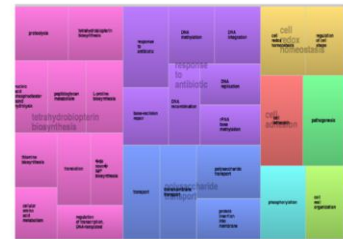
Cluster 1 GO term



Cluster 2 GO term



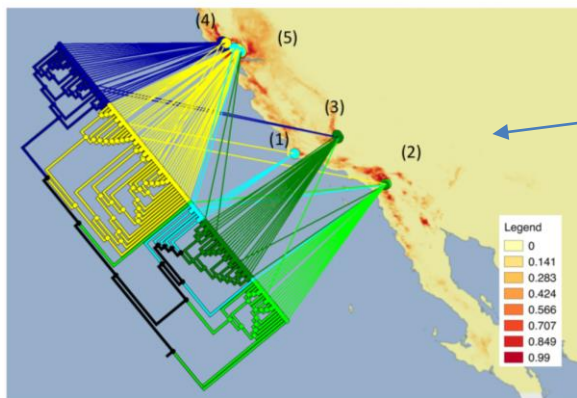
Cluster 3 GO term



Cluster 4 GO term

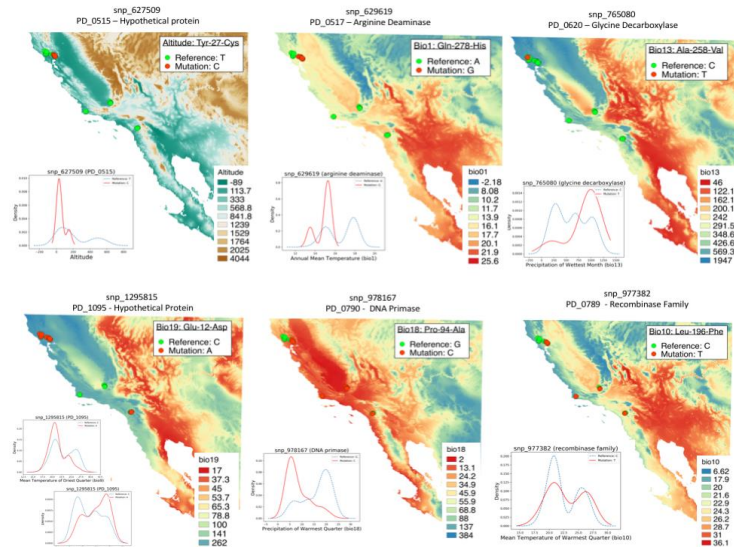


Cluster 5 GO term



What is responsible for the geographical structure?

snp	Protein Method	Gene	Protein	Proteomic Database	Protein	Variable	Ref Name	Ref Access
00000	RX	PD_0107	DUF325 domain-containing	MF & p*	bio6	Temperature		
00009	RX	PD_0120(gpr)	phage-related portal protein	MF & UPM	bio13	Precipitation	GO:0000048	virion assembly
00009	RX	PD_0001(sprk)	ABC transporter ATP-binding protein	MF & p*	bio10	Temperature	GO:0000007	ATPase activity
00009	RX, H <sub>2</sub> O, H <sub>2</sub>	PD_0115	Unknown	MF & UPM	bio6	Temperature	GO:0000021	integral component of membrane
00009	RX, H <sub>2</sub> O, H <sub>2</sub>	PD_0115	Unknown	p* & UPM	bio6	Temperature	GO:0000021	integral component of membrane
00009	RX, H <sub>2</sub> O, H <sub>2</sub>	PD_0115	Unknown	MF & UPM	Altitude	Altitude	GO:0000021	integral component of membrane
00009	H <sub>2</sub> O, H <sub>2</sub>	PD_0116	Unknown	MF & UPM	bio6	Temperature	GO:0000021	integral component of membrane
00009	RX	PD_0117	Arginine deiminase	MF & UPM	bio6	Temperature	GO:0000021	integral component of membrane
00009	RX	PD_0020(gpr)	Glycine decarboxylase	MF & p*	bio13	Precipitation	GO:0000046	glycine catabolic process
00009	RX	PD_0184(spr)	Phage-related integrase	p* & UPM	Altitude	Altitude	GO:0000010	DNA integration
00009	RX, H <sub>2</sub> O, H <sub>2</sub>	PD_0189	Recombinase family	MF & UPM	bio10	Temperature	GO:0000077, GO:0000025, GO:0000020	DNA recombination
00009	RX, H <sub>2</sub> O, H <sub>2</sub>	PD_0189	Recombinase family	MF & UPM & p*	bio10	Temperature	GO:0000077, GO:0000025, GO:0000020	DNA recombination
00009	RX	PD_0190(sprk)	DNA primase	p* & UPM	bio6	Precipitation		
00009	RX	PD_0008	DUF325 domain-containing	MF & p*	bio13	Precipitation	GO:0000016, GO:0000025, GO:0000048	carboxyl-binding, carboxylate-dependent protein kinase activity, protein phosphorylation
00009	RX	PD_1001	Unknown	MF & p*	Altitude	Altitude		
00009	RX	PD_1001	Unknown	p* & UPM	bio6, bio10, bio13, bio14, bio15	Temperature		
00009	RX	PD_1243	DUF325 domain-containing	UPM & p*	bio6	Temperature		
00009	RX	PD_1243	DUF325 domain-containing	UPM & p*	Altitude, bio6	Altitude, Temperature		





Have genomic data helped us so far?  
Yes.

But:

- . Most questions cannot be appropriately answered with the data available
- . Questions need to be generated first, then adequate data must be collected
- . Generating testable hypothesis from genomic data is not straightforward (at least for functional/mechanistic work)
- . Raw sequence data are key for future work
- . We are still at the very 'tip of the iceberg'

