

Journées SUCCES – France Grilles
Institut de Physique du Globe de Paris
5 et 6 novembre 2015

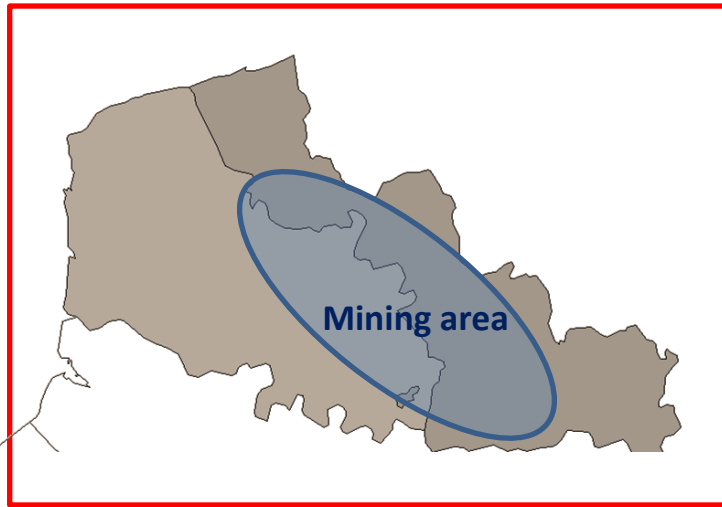


UN EXEMPLE D'INFÉRENCES BAYÉSIENNES EN GÉNÉTIQUE DES POPULATIONS, LE CAS DU CRAPAUD CALAMITE (*EPIDALEA CALAMITA*) DANS LE NORD DE LA FRANCE

Leslie Faucher, Sophie Gallina & Jean-François Arnaud
Université de Lille – Sciences & Technologies



North of France (Nord Pas De Calais district)



Three centuries (XVIIIth to XXth century) of coal extraction

➔ Contrasting landscape shaped by the human activity from the coastline to the mining area.



Ponds in the top of the slag heap or in its slope



By contrast, mining activity had created new open field habitats suitable for pioneering species

- Natterjack toad (*Epidalea calamita*)

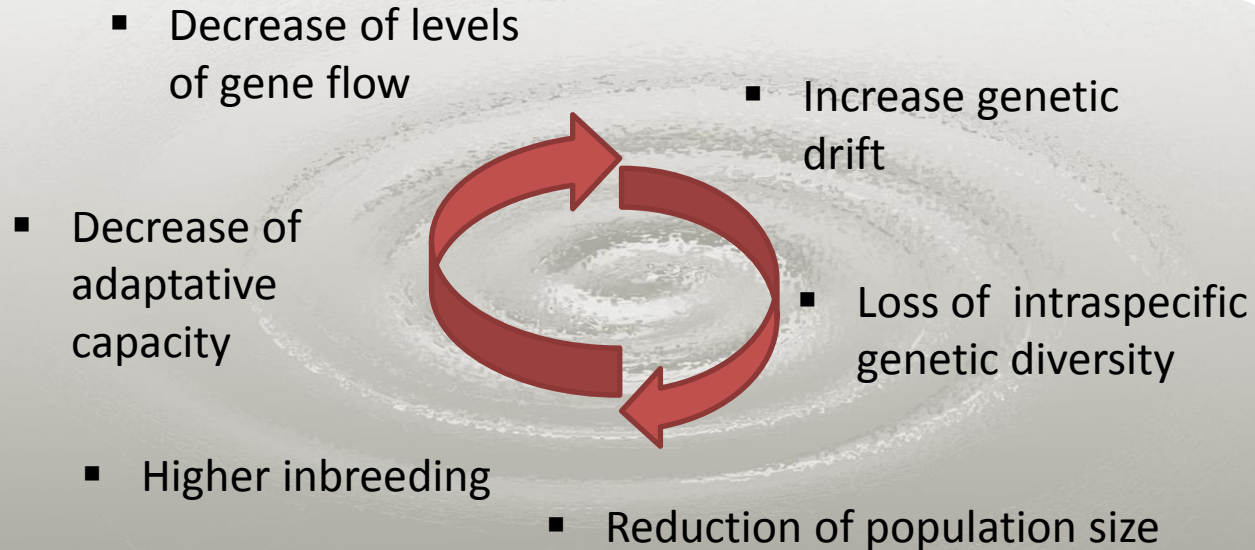
- Pioneer species
- High patrimonial interest
- Protection status



- Found in the « Nord Pas de Calais » district in coastal **natural but fragmented habitats** and more recently in **new habitats** within the mining area



Landscape fragmentation → isolation of populations

- 
- Decrease of levels of gene flow
 - Increase genetic drift
 - Decrease of adaptative capacity
 - Loss of intraspecific genetic diversity
 - Higher inbreeding
 - Reduction of population size

It questions the probability of species persistence

Fragmentation is a major factor implicated in population declines (*e.g.* Fischer & Lindenmayer 2007)



What is population genetics?

Genetic :

Study of the transmission of inherited information

Individual-level thinking: What gametes and offspring are produced, in what frequencies, from a given pair of parents?

$A_1A_2 \times A_1A_2$



	A_1	A_2
A_1	A_1A_1	A_1A_2
A_2	A_2A_1	A_2A_2

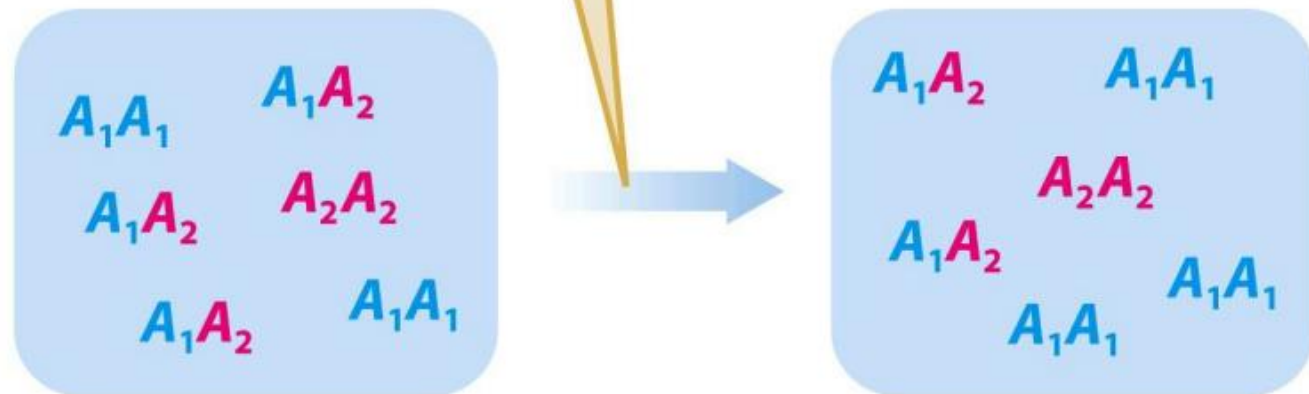
Tiré de Bergstrom C. T. & Dugatkin L. A. (2012) *Evolution*. W. W. Norton & Compagny Inc., New York, NY.

What is population genetics?

→ Population Genetics :

How and why genetic information evolve over time [and space] within species and populations ?

Population-level thinking: How do the characteristics of the population change over time as the result of evolutionary processes?

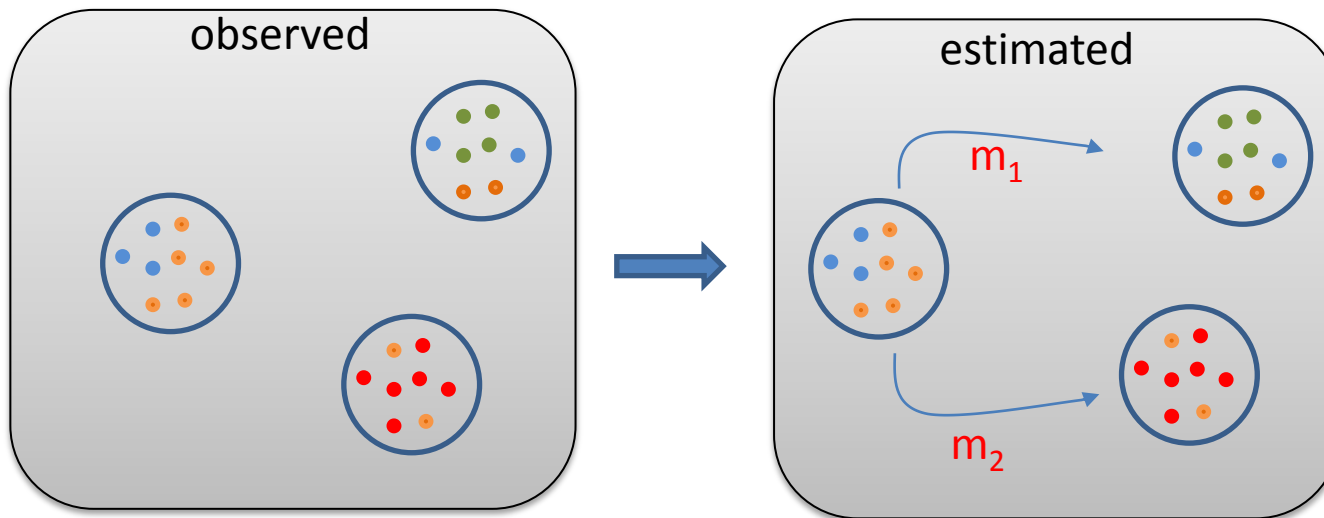


Tiré de Bergstrom C. T. & Dugatkin L. A. (2012) *Evolution*. W. W. Norton & Compagny Inc., New York, NY.

What is population genetics?

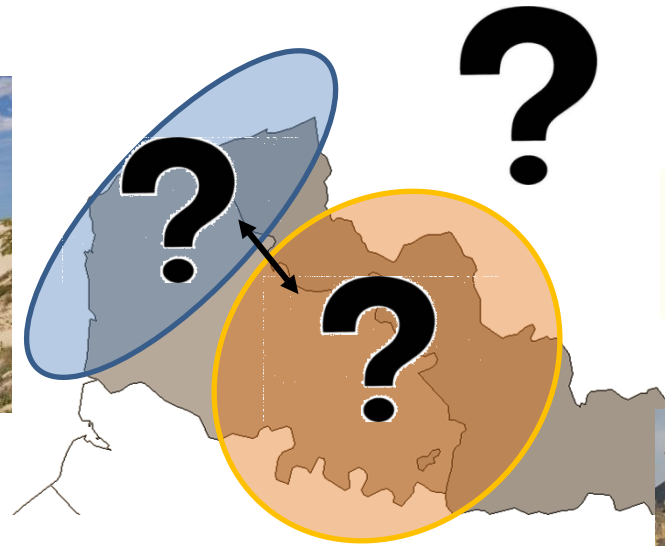
Main goal : estimating the reproductive system, migration rates ... from the information carried by neutral genetic diversity

In our case



Analysing dispersal movements over time and space through the spatial distribution of neutral genetic variability in populations

Main question : do coastal and inland populations differ in terms of genetic structuring?



- Are there contrasting patterns of inbreeding and/or level of genetic diversity between these two areas?
- Where do the mining area populations come from?
- Have these populations been founded by accidental human-mediated dispersal events during mining exploitation?

How do we proceed ?



SAMPLING



CHAÎNE DES TERRILS

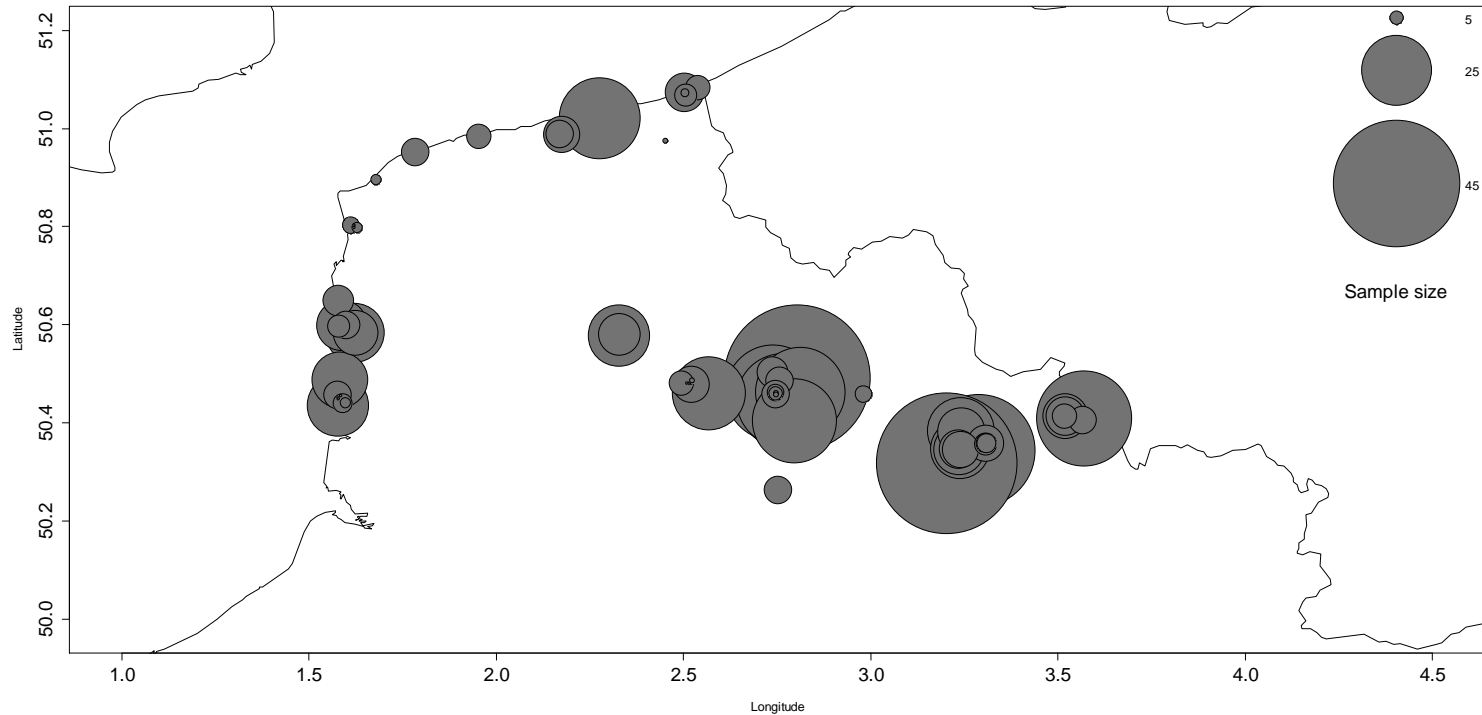


Conservatoire
d'espaces naturels
Nord - Pas-de-Calais



DNA extractions

SAMPLING



959 natterjack toads sampled for 72 ponds (29 along the coastline and 43 in the mining area)

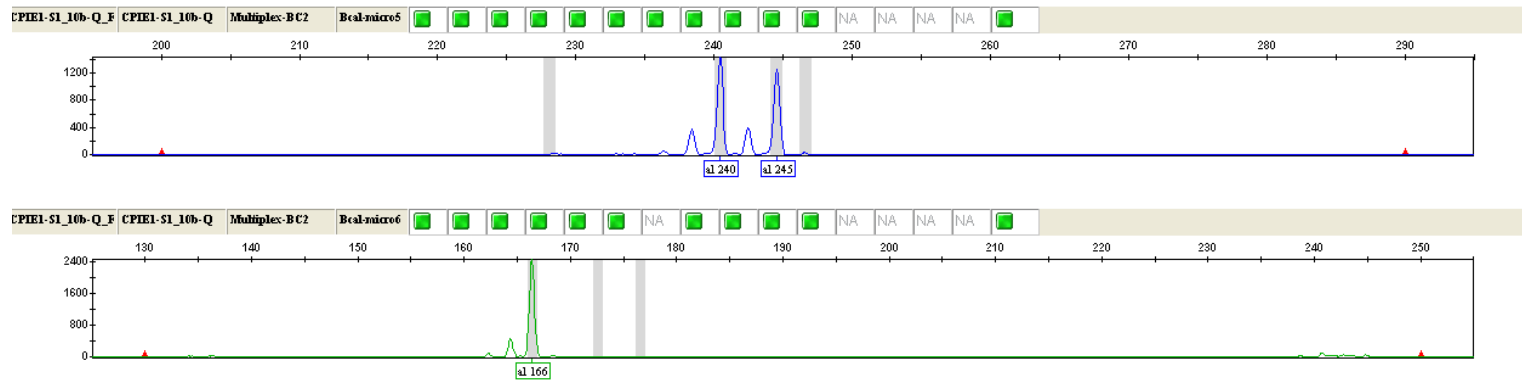
MOLECULAR MARKERS

- 37 nuclear microsatellite loci
(15 already published, 22 recently isolated)

Copyright 1998 Randy Glasbergen. www.glasbergen.com



"Looks aren't everything. It's what's inside you that really matters. A biology teacher told me that."



• CLUSTERING BAYÉSIEEN :

→ The method attempts to assign individuals to clusters on the basis of their multilocus genotype, while simultaneously estimating cluster allele frequencies

K cluster : each of which is characterized by a set of allele frequencies at each locus.

↳ Assignment of individuals to one cluster (or jointly to two or more clusters if their genotypes indicate that they are admixed).

✓ *Within clusters, the loci are at Hardy-Weinberg equilibrium, and linkage equilibrium.*

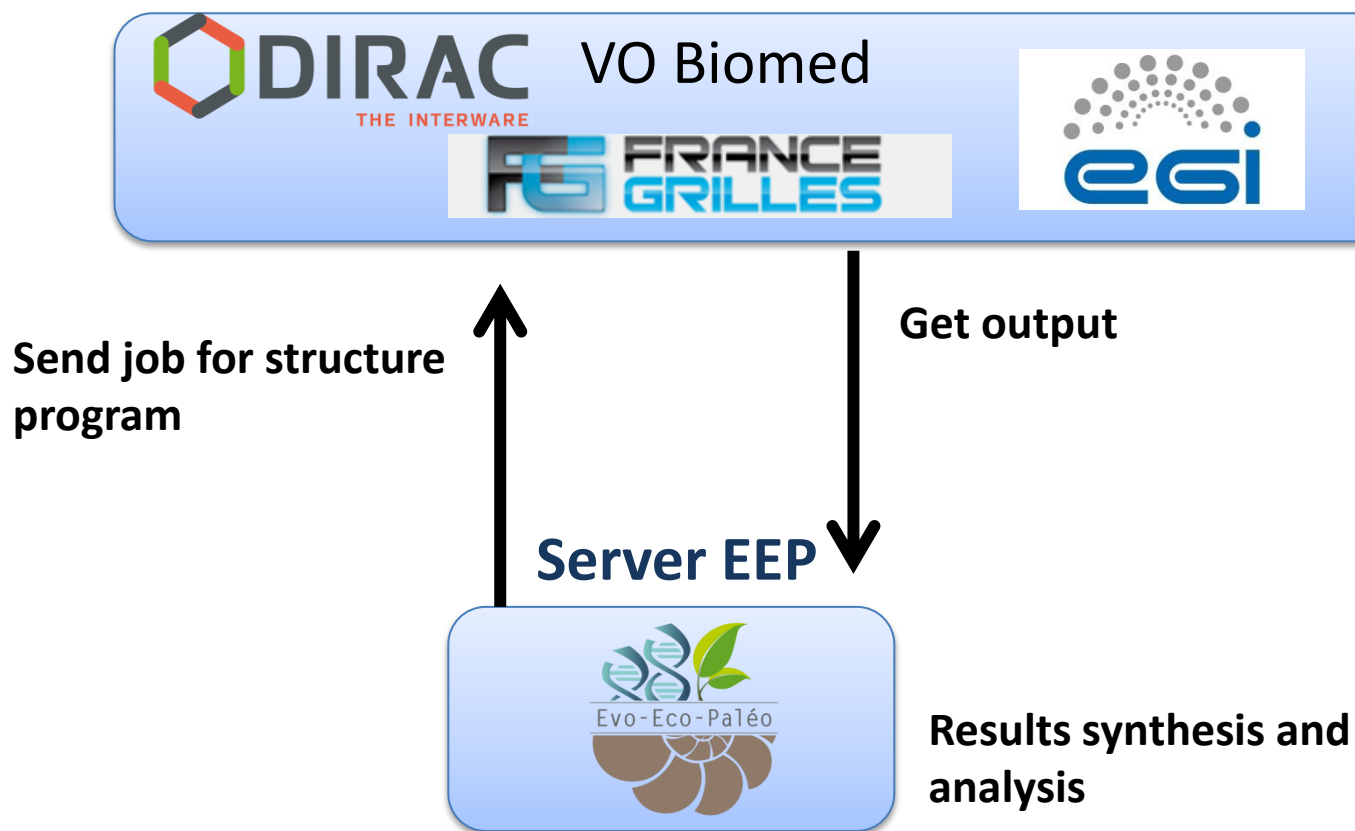
❖ **STRUCTURE** (Pritchard et al. 2000)

30 replicates

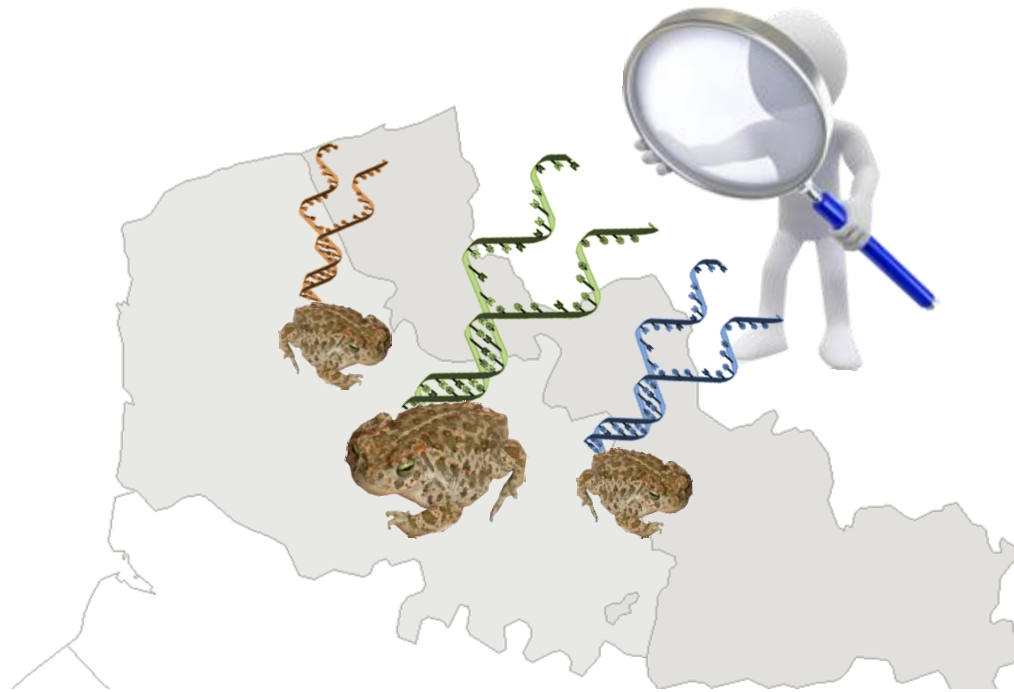
K = from 2 to total number of populations sampled

TECHNICAL DETAILS :

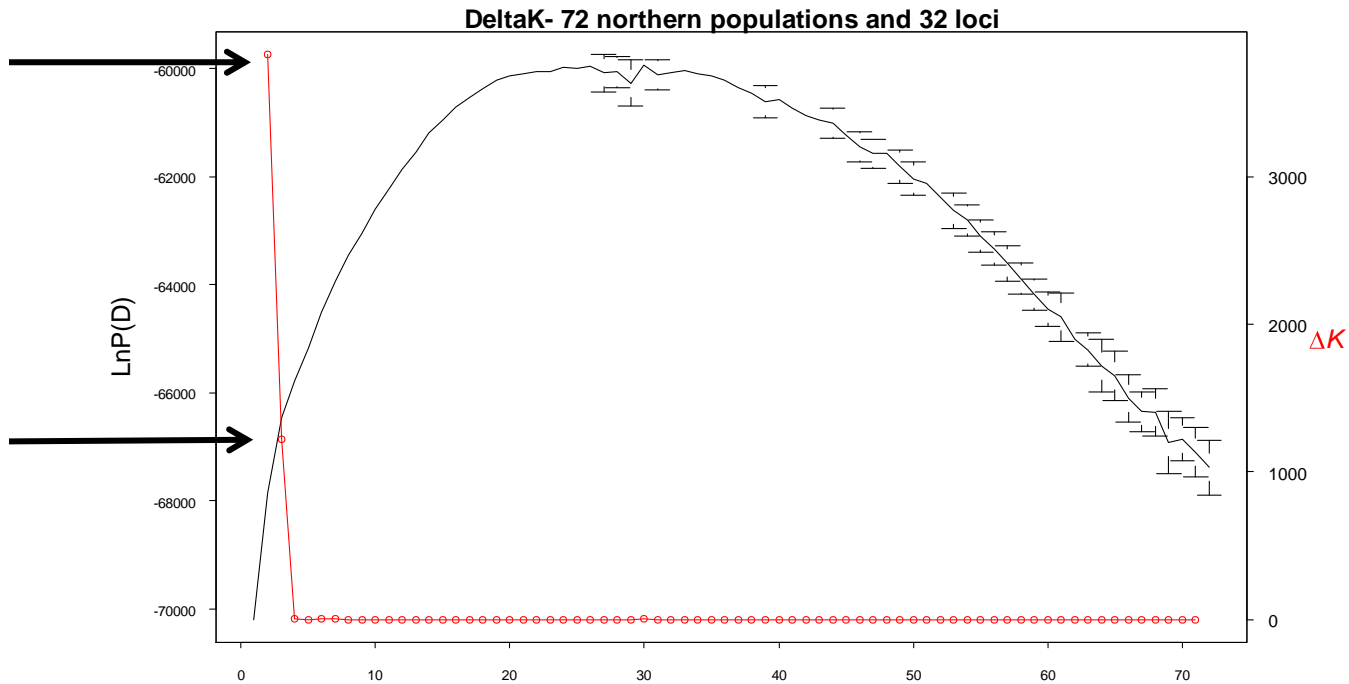
Grid Infrastructure



What did we observed ?

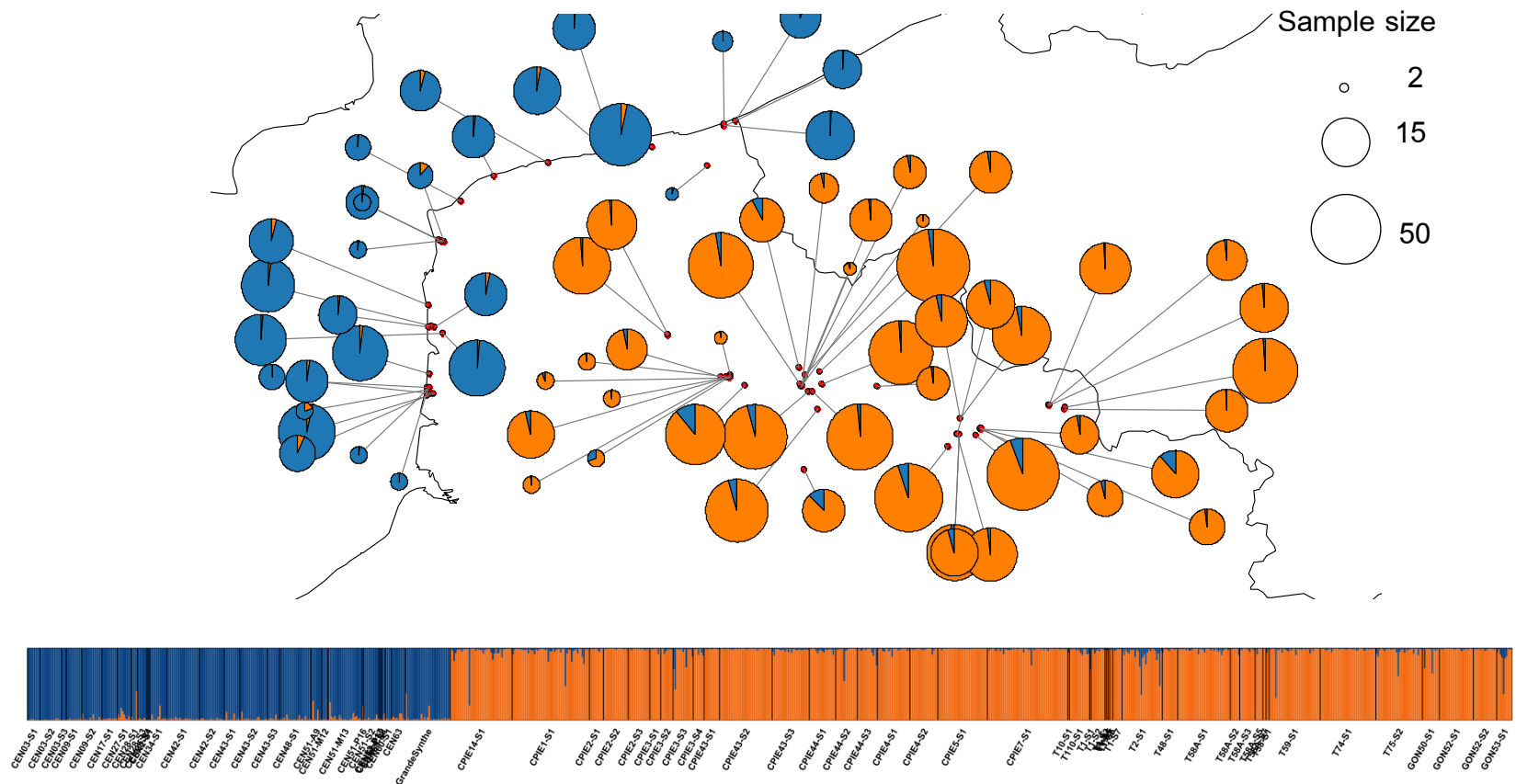


Bayesian clustering analysis (Pritchard et al. 2000)



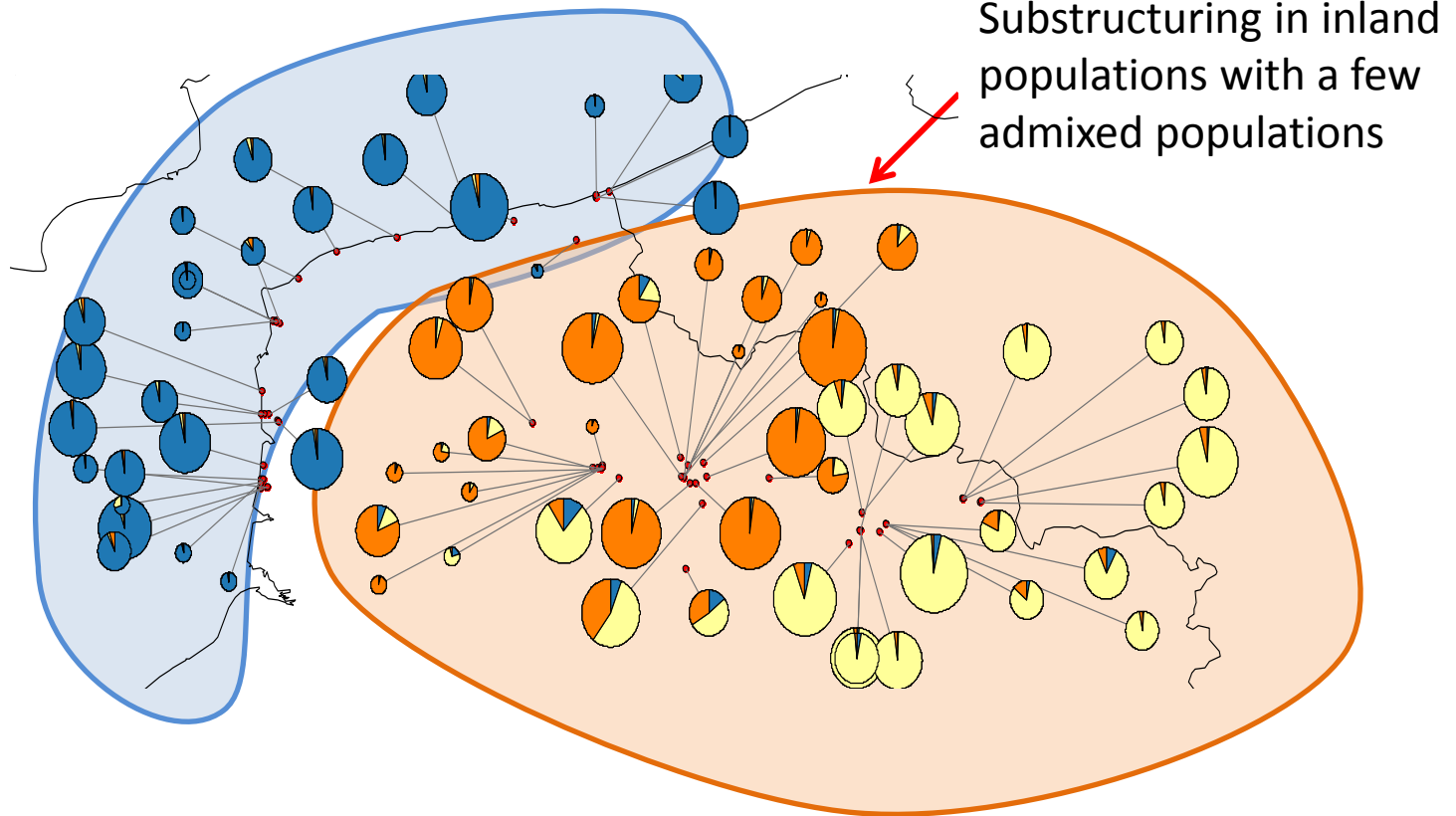
Mean population probabilities of membership to belong to K=2 genetic clusters

Bayesian clustering analysis (Pritchard et al. 2000)

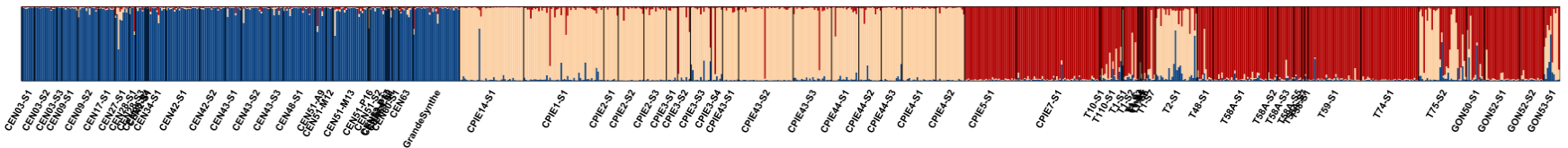


Mean population probabilities of membership to belong to $K=2$ genetic clusters

Bayesian clustering analysis (Pritchard et al. 2000)



➔ Clear genetic distinctiveness of coastline and inland populations



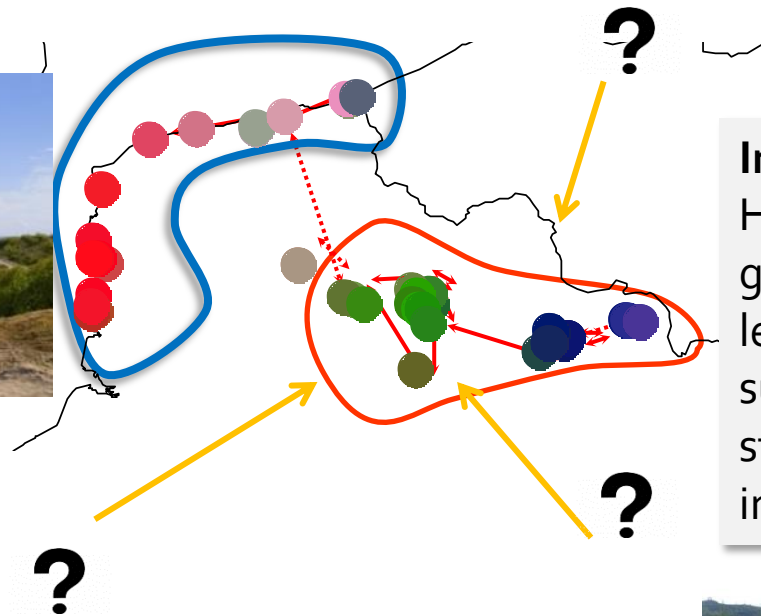
Bayesian clustering analysis (Pritchard et al. 2000)

	Loci	Kmax	rep1	rep2	rep3	jobs	hours	days	months	years
35locus_68pop-	35	68	40	35	30	2380	71400	2975	99.17	8.15
35locus_44pop-	35	44	40	35	30	1540	30800	1283.33	42.78	3.52
35locus_24pop	35	24	40	35	30	840	5880	245	8.17	0.67
26locus_TOTAL	26	85	40	35	30	2975	8925	371.87	12.40	1.02

➔ Need of high throughput computation

Coastal populations:

Isolation by distance through a classical stepping stone model following a scenario of post-glacial recolonisation

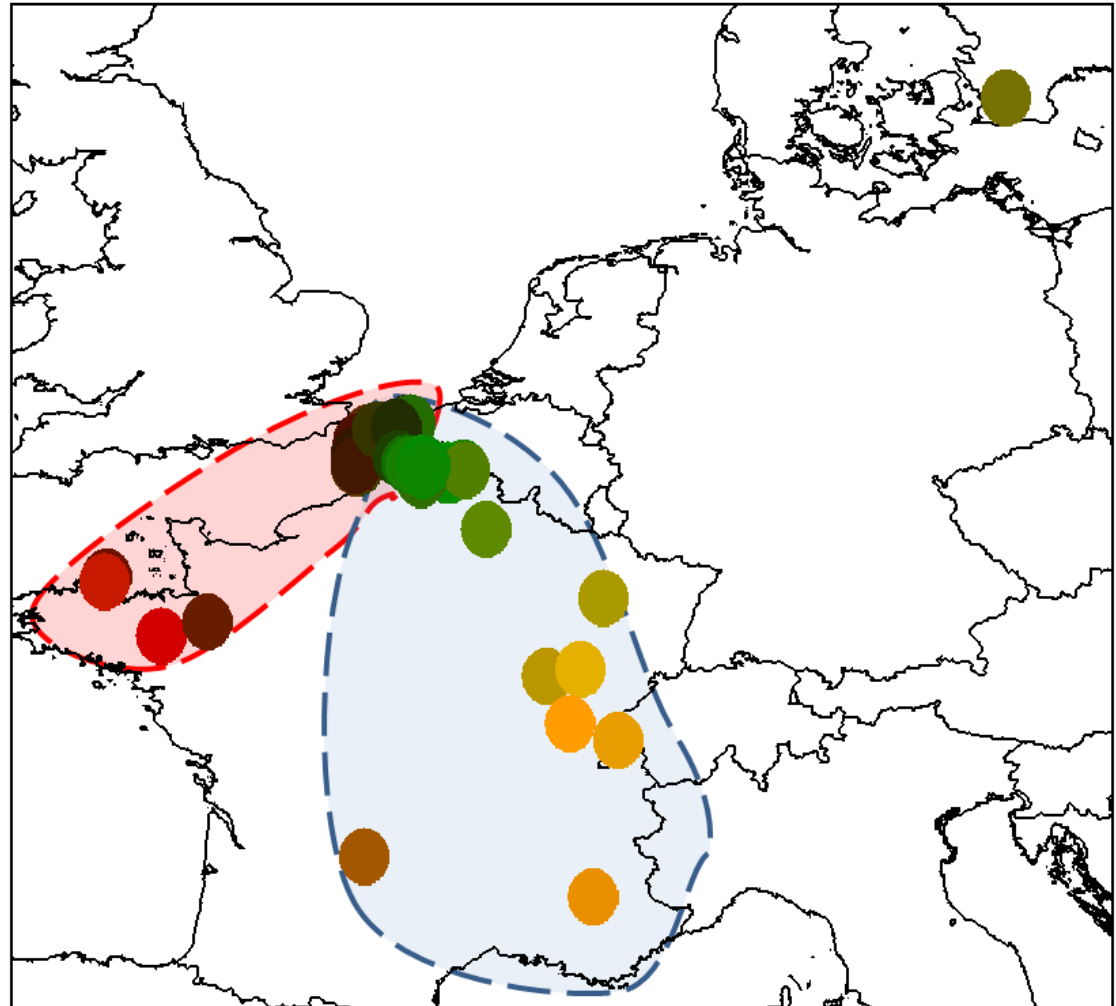


Inland populations:

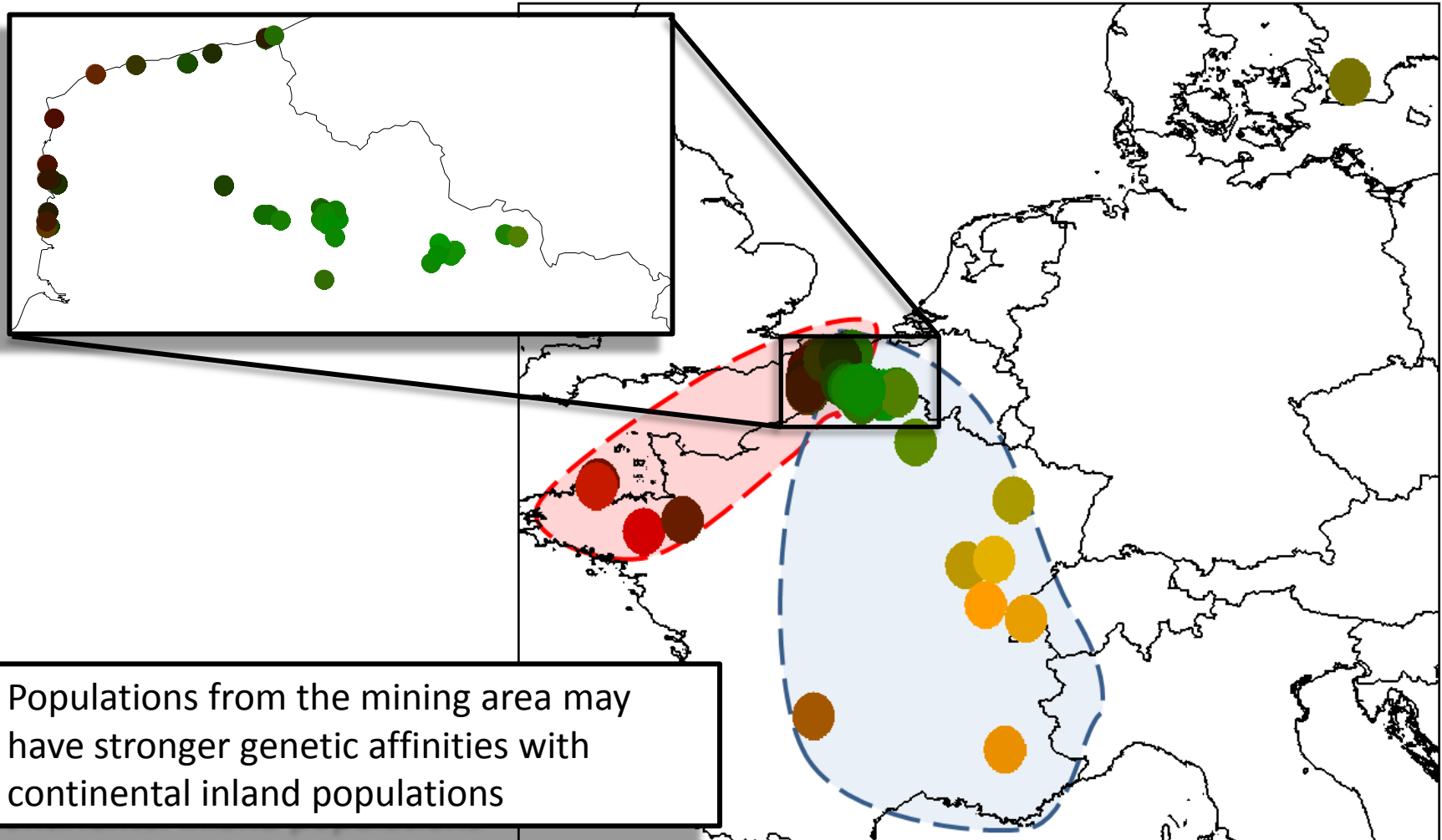
Higher observed levels of genetic admixture and levels of genetic diversity suggest a metapopulation structure with multiple introductions



Tracing back the history of colonization of the coal basin:
preliminary results...



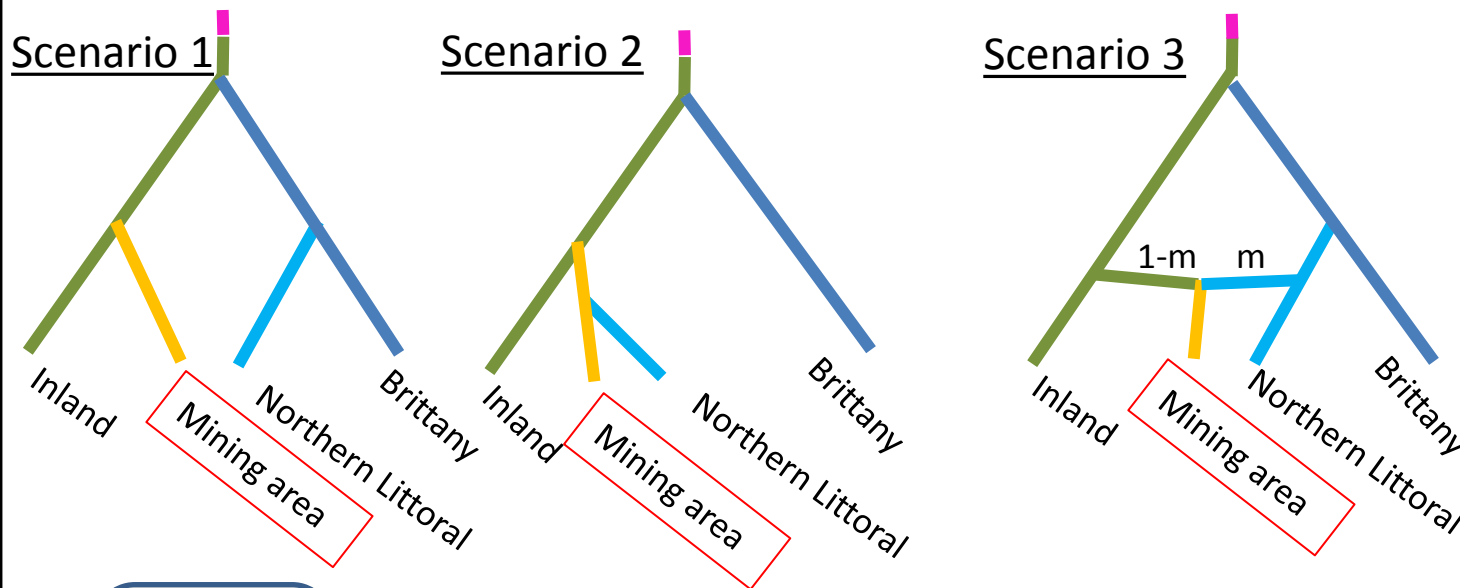
Tracing back the history of colonization of the coal basin: preliminary results...



Populations from the mining area may have stronger genetic affinities with continental inland populations

PERSPECTIVES:

→ Test different evolutionary scenarios :



End of the last
glaciation
(-11 000 BP)

t_3

t_2

t_1

t_0

Actual

Na
N1
N2
N3
N4
N5

Several parameters and even more scenarios

→ which one is the best ?

- We plan to proceed to Approximate Bayesian Computation (ABC) analysis

Ressources needed:

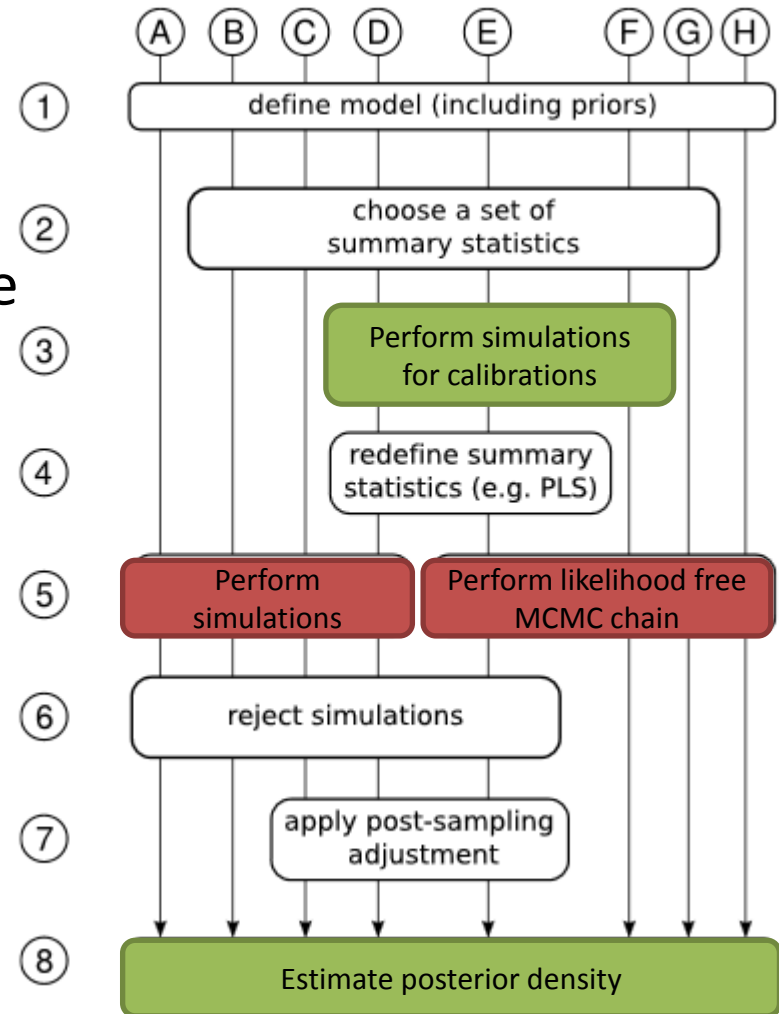
Pipeline using many tools,
with dependences for example
library python numpy

↳ very long task (~2-4 weeks)

➔ incompatible with grid usage

➔ need the use of virtual
machines on cloud

ABC analysis workflow



- We are currently setting up an appliance with pipeline
- Virtual Machines using this appliance will be deployed on three clouds
 - Institut Français de Bioinformatique (IFB)– Paris (3000 cores)
 - Univ Lille - IFB & France Grilles (FG) plateforme - Lille (320 cores)
 - **IPHC**, FG plateforme - Strasbourg (176 cores)



- It will allow us to deploy each different scenarios on one virtual machine



Thank for your attention !

Many thanks to all the people who participated to the sampling, Conservatoire d'Espaces Naturels du Nord-Pas de Calais, CPIE Chaine des terrils, Groupement ornithologique et naturaliste du Nord, Julie Jacquery (Université de Rennes), Baptiste Faure (Biotop), Conservatoire des Espaces Naturels de Lorraine, Bretagne Vivante, LPO Franche Comté, LPO Lot, Conservatoire d'Espaces Naturels de Bourgogne

Many thanks to Cécile Godé and Laura Henocq for help in genotyping

And many thanks to Geneviève Romier, Jérôme Pansanel at France Grilles, Christophe Blanchet at Institut Français de Bioinformatique and Matthieu Marquillie, Cyrille Toulet at Univ Lille for the computing infrastucture.