Identifier les scénarios complexes de colonisation à l’échelle continentale des champignons pathogènes à l’aide des méthodes ABC

Benoit Barrès¹,²,³, Stéphanie Robert², Catherine Abadie², Jean Carlier², Cyril Dutech¹ et Virginie Ravigné²

¹ UMR BIOGECO  
² UMR BGPI  
³ UPR 31

Histoire de la colonisation et approche ABC chez les champignons pathogènes
Why reconstruct routes of emergences?

- Identify source populations

- Gain insights on dispersal mode

- Understand the process involved in emergence (admixture, bridgehead effect)

  ➔ Improve quarantine politics

  ➔ First step in the identification of evolutionary changes linked with emergence (comparison between source populations and emerging populations)

(Keller and Taylor, 2008)
Why use ABC methods?

- Assignment methodology (GeneClass2)
- Model-based coalescent method (MIGRATE, IM, …)

⇒ Need tractable Likelihoods

…but Likelihoods are not always tractable

(Hartig et al., 2011)
The ABC framework

3 steps:

1) generating (a lot of) simulated data sets

2) selecting simulated data sets closest to observed data set

3) estimating posterior distributions of parameters through a regression procedure

(Excoffier et al., 2005)

(Hartig et al., 2011)

Reviews: Beaumont 2010; Bertorelle et al., 2010; Csilléry et al., 2010
DIYABC software

- design and simulation of scenarios (admixture, unsampled population, divergence between populations)
- tools to perform tests on priors distribution
- evaluation of posterior distribution
- tools to assess confidence and goodness of fit of the model
- haploid and diploid data
- microsatellite and/or sequence data
Continental scale dispersal of *Microcyclus ulei*

Benoit Barrès, Jean Carlier, Marc Seguin, Catherine Fenouillet, Christian Cilas and Virginie Ravigné
Continental scale dispersal of *Microcyclus ulei*
Sampling design

- 4 populations
- 116 individuals
- 16 microsatellites markers

$F_{ST} = 0.70$
Genetic structure

Bayesian Clustering

PCA

K=2

K=3

K=4

Bayesian Clustering

PCA
Four independent foundation events
Three independent foundation events
Two independent foundation events

Class IV

Class V

Class VI

ancestral population

Original divergence leading to genetic structure (TO)

| : N  | : NO. |
| : NB. | : N.  |

Primary foundation events (TF)

Secondary foundation events (TS or TA)

Sampling
Continental scale dispersal of *Microcyclus ulei*

- Two independent foundation events
- An unknown common origin in western *Hevea* plantations
  → man-mediated dispersal

(Barrès *et al.*, submitted)
Introduction of *Cryphonectria parasitica* in North America

Cyril Dutech, Benoit Barrès, Julien Bridier, Cécile Robin, Michael Milgroom and Virginie Ravigné
Introduction of *Cryphonectria parasitica* in North America

- several hypotheses on the introduction in North America exist
- 10 scenarios tested (4 with simple and 6 with multiple introduction)
Introduction of *Cryphonectria parasitica* in North America

Posterior probabilities of best tested scenarios:

<table>
<thead>
<tr>
<th></th>
<th>Japan</th>
<th>Japan + China</th>
<th>Japan + Intermediate</th>
<th>Japan + Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYM</td>
<td>0.34 [0.31-0.38]</td>
<td>0.24 [0.22-0.27]</td>
<td>0.17 [0.15-0.19]</td>
<td>0.21 [0.19-0.24]</td>
</tr>
<tr>
<td>KEM</td>
<td>0.03 [0.03-0.04]</td>
<td>0.08 [0.07-0.10]</td>
<td>0.19 [0.17-0.22]</td>
<td>0.51 [0.47-0.54]</td>
</tr>
<tr>
<td>NHM</td>
<td>0.14 [0.12-0.16]</td>
<td>0.14 [0.12-0.16]</td>
<td>0.17 [0.14-0.19]</td>
<td>0.47 [0.43-0.50]</td>
</tr>
<tr>
<td>ONM</td>
<td>0.16 [0.14-0.19]</td>
<td>0.07 [0.06-0.08]</td>
<td>0.22 [0.20-0.25]</td>
<td>0.39 [0.35-0.42]</td>
</tr>
</tbody>
</table>

- Corroborates the Japanese origin of *C. parasitica*
- No (or undetectable) Chinese introduction
- A possible admixture with an unidentified population

(Dutech *et al.*, accepted)
Introduction of *Mycosphaerella fijiensis* in South America

Stéphanie Robert, Jean Carlier, Marie-Françoise Zapater, Catherine Abadie and Virginie Ravigné
Introduction of *Mycosphaerella fijiensis* in South America

Scenario assumed:

Other possible scenarios:
An introduction with admixture

- Introduction with admixture
- One unidentified source population
## Conclusion

### Method

ABC is a useful tool to decipher complex colonization scenario

- identification of admixture event

- importance of unsampled population

### Biology

- identify source population (confirm or infirm historical knowledge)

- multiple introductions

- underline lack of knowledge on native area of pathogen

### Some limitations

- panmictic population

- no regular gene flow between populations

- comparison between tested scenarios only
Thank you for your attention

UMR BIOGECO
Cyril Dutech
Cécile Robin
Julien Bridier

UMR BGPI
Jean Carlier
Virginie Ravigné
Elisabeth Fournier
Didier Tharreau
Daniel Bieysse

Cornell University
Michael Milgroom

UPR 31
Catherine Fenouillet
Christian Cilas
Jean Guyot
Franck Rivano

MICHELIN
Eric Cavaloc
Carlos Mattos
Milton Luiz

UMR DAP
Marc Seguin
Vincent Le Guen

Funding: CIRAD, MICHELIN, ANR EMERFUNDIS, ANR EMILE
Prior distribution of parameters

<table>
<thead>
<tr>
<th>Population effective sizes of the</th>
<th>Parameter name</th>
<th>distribution</th>
<th>Extremum values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ancestral population</td>
<td>N</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>population leading to Brazilian population foundation</td>
<td>NOb</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>population leading to Equatorian population foundation</td>
<td>NOe</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>population leading to Guatemalan population foundation</td>
<td>NOgt</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>population leading to French Guyanian population foundation</td>
<td>NOgy</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>population leading to unsampled population foundation</td>
<td>NOu</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>bottleneck during foundation of Brazilian population</td>
<td>NBB</td>
<td>log-uniform</td>
<td>{1- 100}</td>
</tr>
<tr>
<td>bottleneck during foundation of Equatorian population</td>
<td>NBE</td>
<td>log-uniform</td>
<td>{1- 100}</td>
</tr>
<tr>
<td>bottleneck during foundation of Guatemalan population</td>
<td>NBgt</td>
<td>log-uniform</td>
<td>{1- 100}</td>
</tr>
<tr>
<td>bottleneck during foundation of French Guyana population</td>
<td>NBgy</td>
<td>log-uniform</td>
<td>{1- 100}</td>
</tr>
<tr>
<td>bottleneck during foundation of unsampled population</td>
<td>NBu</td>
<td>log-uniform</td>
<td>{1- 100}</td>
</tr>
<tr>
<td>Brazilian population</td>
<td>Nb</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>Equatorian population</td>
<td>Ne</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>Guatemalan population</td>
<td>Ngt</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>French Guyana population</td>
<td>Ngy</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
<tr>
<td>Unsampled population</td>
<td>Nu</td>
<td>uniform</td>
<td>{10 – 100,000}</td>
</tr>
</tbody>
</table>

Time of the original divergence leading to genetic structure in source population

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>distribution</th>
<th>Extremum values</th>
</tr>
</thead>
<tbody>
<tr>
<td>TO</td>
<td>uniform</td>
<td>{600 – 100,000}</td>
</tr>
<tr>
<td>foundation of the Brazilian population</td>
<td>TFb</td>
<td>log-uniform</td>
</tr>
<tr>
<td>foundation of the Equatorian population</td>
<td>TFe</td>
<td>log-uniform</td>
</tr>
<tr>
<td>foundation of the Guatemalan population</td>
<td>TFGt</td>
<td>log-uniform</td>
</tr>
<tr>
<td>foundation of the French Guyana population</td>
<td>TFGy</td>
<td>log-uniform</td>
</tr>
<tr>
<td>foundation of the unsampled population</td>
<td>Tfu</td>
<td>log-uniform</td>
</tr>
<tr>
<td>secondary foundation of Brazilian population</td>
<td>TSb</td>
<td>log-uniform</td>
</tr>
<tr>
<td>secondary foundation of Equatorian population</td>
<td>TSe</td>
<td>log-uniform</td>
</tr>
<tr>
<td>secondary foundation of Guatemalan population</td>
<td>TSgt</td>
<td>log-uniform</td>
</tr>
<tr>
<td>secondary foundation of French Guyana population</td>
<td>TSgy</td>
<td>log-uniform</td>
</tr>
<tr>
<td>admixture event</td>
<td>TA</td>
<td>log-uniform</td>
</tr>
</tbody>
</table>

Population effective size :
- population UNIF {10-100,000}
- bottleneck LOG-U {1-100}

Time of divergence :
- ancestral UNIF {600-100,000}
- foundation pop LOG-U {16-500}

Admixture rate :
- UNIF {0.001-0.999}
The choice of a scenario and its consistency

- Goodness of fit evaluated by Model-checking

- Confidence in the choice of scenario evaluated with pseudo-observed datasets

α = 21%

β = 1.8%

P = 0.123
[0.082-0.154]

P = 0.877
[0.833-0.905]
Final scenario for *M. ulei*

Erreur de Type II = proportion de jeux de données simulés sous de mauvais scénarios et qui pointent vers le meilleur scénario  
Ici : 1.8 % !!

Erreur de Type I = proportion de jeux de données simulés sous le bon scénario et qui pointent vers un autre scénario  
Ici : 21 % (beurk !)

Toutes les erreurs sont entre ce scénario et le scénario V
Final scenario for *M. fijiensis*