Overview of my master thesis

Lab meeting at Duke University

03/19/2015

Alice Ledent, second year of the master’s degree in biology, University of Liège (ULG)
Overview of my master thesis

Lab meeting at Duke University

03/19/2015

Alice Ledent, second year of the master’s degree in biology, University of Liège (ULG)
Unravelling the Quaternary biogeography history of European bryophytes through Approximate Bayesian Statistics
Introduction
Introduction

- What’s the impact of global warming on species repartition?
  - → Studying past climate changes
    - (Petit et al. 2005)
Introduction

- What’s the impact of global warming on species repartition?
  - → studying past climate changes
    - (Petit et al. 2005)

- Which climatic phenomenon makes the actual repartition of species?
  - → the called “Quaternary” glacial periods
  - Last Glacial Maximum (LGM, 21.000 YBP) = the more virulent
    - (Hewitt 1999)

- Europe : stronger consequences than in North America or in the Southern hemisphere
  - Why? Harder climatic conditions and dispersal barriers W-E
    - (Hewitt 2000)
Introduction

- What’s the impact of global warming on species repartition?
  - → studying past climate changes
    - (Petit et al. 2005)

- Which climatic phenomenon makes the actual repartition of species?
  - → the called “Quaternary” glacial periods
  - Last Glacial Maximum (LGM, 21,000 YBP) = the more virulent
    - (Hewitt 1999)

- Europe: stronger consequences than in North America or in the Southern hemisphere
  - Why? Harder climatic conditions and dispersal barriers W-E
    - (Hewitt 2000)
Introduction

- Where to survive in Europe?
  - 1) Mediterranean refugia hypothesis (SRH)
    - species survived in 52 Mediterranean refugia
    - refugia = sources for continental recolonization
    - (Hewitt 1999) and (Médail & Diadema 2009)
Introduction

Where to survive in Europe?

1) Mediterranean refugia hypothesis (SRH)
   - species survived in 52 Mediterranean refugia
   - refugia = sources for continental recolonization
   - (Hewitt 1999) and (Médail & Diadema 2009)

2) Northern micro-refugia hypothesis (NRH)
   - South = isolated from the other parts of Europe
   - Southern refugia = sink, no source of biodiversity
   - Northern micro-refugia = sources for recolonization
   - (Petit et al. 2003) et (Petit et al. 2005)

Non-exclusives hypotheses!
   - depending on: physiological characteristics and biogeographical affinities
   - for « review »: (Bhagwat & Willis 2008)
Where to survive in Europe?

1) Mediterranean refugia hypothesis (SRH)
   - species survived in 52 Mediterranean refugia
   - refugia = sources for continental recolonization
   - (Hewitt 1999) and (Médail & Diadema 2009)

2) Northern micro-refugia hypothesis (NRH)
   - South = isolated from the other parts of Europe
   - Southern refugia = sink, no source of biodiversity
   - Northern micro-refugia = sources for recolonization
   - (Petit et al. 2003) et (Petit et al. 2005)

Non-exclusive hypotheses!
   - depending on: physiological characteristics and biogeographical affinities
   - for « review »: (Bhagwat & Willis 2008)
Introduction

- Biogeographical studies = vertebrates or angiosperms
  - what about bryophytes?
    - less tolerant → first to suffered from changes
    - higher dispersal capacities → able to cross seas and oceans
      - for « review »: (Lewis et al. 2014)
Introduction

- **Biogeographical studies = vertebrates or angiosperms**
  - → what about bryophytes?
    - less tolerant → first to suffered from changes
    - higher dispersal capacities → able to cross seas and oceans
      - for « review » : (Lewis et al. 2014)

- **New hypothesis**
  - E and W of Europe evolve separately
  - W colonized by Macaronesia or North America
  - (Désamoré et al. 2012) et (Stenøien et al. 2011)

- **Few studies on the biogeography of bryophytes**
  - → first meta-study in the domain (18 species)
  - as examples : (Désamoré et al. 2012), (Désamoré 2013), (Stenøien et al. 2011) et (Urmi & Schneller 2004)
Introduction

- Biogeographical studies = vertebrates or angiosperms
  - what about bryophytes?
    - less tolerant → first to suffered from changes
    - higher dispersal capacities → able to cross seas and oceans
      - for « review » : (Lewis et al. 2014)

- New hypothesis
  - E and W of Europe evolve separately
  - W colonized by Macaronesia or North America
    - (Désamoré et al. 2012) et (Stenøien et al. 2011)

- Few studies on the biogeography of bryophytes
  - first meta-study in the domain (18 species)
  - as examples: (Désamoré et al. 2012), (Désamoré 2013), (Stenøien et al. 2011) et (Urmi & Schneller 2004)
Objectives

- The principal goal of this study is the estimation, in Europe, of the impact of the last glacial maximum on bryophytes repartition.

- Especially,
  - (1) to contrast 5 biogeographical hypotheses for each studied species,
  - (2) to group species that present a consensual biogeographical response,
  - (3) to corroborate the groupings to the species ecological affinities.
Objectives

- The principal goal of this study is the estimation, in Europe, of the impact of the last glacial maximum on bryophytes repartition.

- Especially,
  
  - (1) to contrast 5 biogeographical hypotheses for each studied species,
  
  - (2) to group species that present a consensual biogeographical response,
  
  - (3) to corroborate the groupings to the species ecological affinities.
Objectives

- The principal goal of this study is the estimation, in Europe, of the impact of the last glacial maximum on bryophytes repartition.

- Especially,

  - (1) to contrast 5 biogeographical hypotheses for each studied species,
  - (2) to group species that present a consensual biogeographical response,
  - (3) to corroborate the groupings to the species ecological affinities.
Objectives

- The principal goal of this study is the estimation, in Europe, of the impact of the last glacial maximum on bryophytes repartition.

- Especially,
  - (1) to contrast 5 biogeographical hypotheses for each studied species,
  - (2) to group species that present a consensual biogeographical response,
  - (3) to corroborate the groupings to the species ecological affinities.
Biogeographical hypotheses

- **H0**: no impact of the LGM on European bryophytes repartition
- **H1**: N and S not separated but N dead during the LGM
  - → recolonization of the N by the S
- **H2**: N and S separated → independent evolution
- **H3a**: W and E not separated but W dead LGM
  - → recolonization of the W by the E
- **H3b**: W and E not separated but W dead LGM
  - → recolonization of the E by the W
- **H4**: W and E separated → independent evolution
- **H5**: Europe : all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Biogeographical hypotheses

- **H0**: no impact of the LGM on European bryophytes repartition
- **H1**: N and S not separated but N dead during the LGM
  - → recolonization of the N by the S
- **H2**: N and S separated → independent evolution
- **H3a**: W and E not separated but W dead LGM
  - → recolonization of the W by the E
- **H3b**: W and E not separated but W dead LGM
  - → recolonization of the E by the W
- **H4**: W and E separated → independent evolution
- **H5**: Europe: all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Biogeographical hypotheses

- **H0**: no impact of the LGM on European bryophytes repartition
- **H1**: N and S not separated but N dead during the LGM
  - → recolonization of the N by the S
- **H2**: N and S separated → independent evolution
- **H3a**: W and E not separated but W dead LGM
  - → recolonization of the W by the E
- **H3b**: W and E not separated but W dead LGM
  - → recolonization of the E by the W
- **H4**: W and E separated → independent evolution
- **H5**: Europe: all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Biogeographical hypotheses

- **H0**: no impact of the LGM on European bryophytes repartition
- **H1**: N and S not separated but N dead during the LGM → recolonization of the N by the S
- **H2**: N and S separated → independent evolution
- **H3a**: W and E not separated but W dead LGM → recolonization of the W by the E
- **H3b**: W and E not separated but W dead LGM → recolonization of the E by the W
- **H4**: W and E separated → independent evolution
- **H5**: Europe: all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Hypothèses biogéographiques

- **H0** : no impact of the LGM on European bryophytes repartition
- **H1** : N and S not separated but N dead during the LGM → recolonization of the N by the S
- **H2** : N and S separated → independent evolution
- **H3a** : W and E not separated but W dead during the LGM → recolonization of the W by the E
- **H3b** : W and E not separated but W dead during the LGM → recolonization of the E by the W
- **H4** : W and E separated → independent evolution
- **H5** : Europe : all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Hypothèses biogéographiques

- H0: no impact of the LGM on European bryophytes repartition
  - H1: N and S not separated by N dead during the LGM → recolonization of the N by the S
- H2: N and S separated → independent evolution
- H3a: W and E not separated but W dead LGM → recolonization of the W by the E
- H3b: W and E not separated but W dead LGM → recolonization of the E by the W
- H4: W and E separated → independent evolution
- H5: Europe: all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Hypothèses biogéographiques

- H0 : no impact of the LGM on European bryophytes repartition
- H1 : N and S not separated but N dead during the LGM
  - → recolonization of the N by the S
- H2 : N and S separated → independent evolution
- H3a : W and E not separated but W dead LGM
  - → recolonization of the W by the E
- H3b : W and E not separated but W dead LGM
  - → recolonization of the E by the W
- H4 : W and E separated → independent evolution
- H5 : Europe : all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Biogeographical hypotheses

- **H0**: no impact of the LGM on European bryophytes repartition
- **H1**: N and S not separated but N dead during the LGM
  - → recolonization of the N by the S
- **H2**: N and S separated → independent evolution
- **H3a**: W and E not separated but W dead LGM
  - → recolonization of the W by the E
- **H3b**: W and E not separated but W dead LGM
  - → recolonization of the E by the W
- **H4**: W and E separated → independent evolution
- **H5**: Europe: all disappear → recolonization by external inputs

+ for each scenario, test of the external inputs
Methods

- **Biogeographical studies = descriptive statistics**
  - To describe a biogeographical scenario on the basis of observed genetic data
    - not necessarily a link between genetic data and biogeographical scenario
    - confrontation of scenarios is not possible
    - for examples: (Bhagwat & Willis 2008), (Désamoré 2013), (Hewitt 1999), (Petit et al. 2003) et (Petit et al. 2005).

- **Méthode utilisée pour le mémoire : ABC**
  - Modéliser des scénarios biogéographiques → simuler des jeux de données correspondant
    - comparaison avec les données observées
    - lien direct entre le scénario biogéographique et les données génétiques
    - Confrontation de scénarios possibles
    - Pour « review » : (Csilléry et al. 2010)
Methods

- **Biogeographical studies = descriptive statistics**
  - To describe a biogeographical scenario on the basis of observed genetic data
    - not necessarily a link between genetic data and biogeographical scenario
    - confrontation of scenarios is not possible
    - for examples: (Bhagwat & Willis 2008), (Désamoré 2013), (Hewitt 1999), (Petit et al. 2003) et (Petit et al. 2005).

- **Method used in this study : ABC**
  - To model biogeographical scenarios → to create trees → to simulate correspondent genetic data
    - comparison with observed genetic data
    - direct link between the biogeographical scenario and the genetic data
    - confrontation of scenarios is possible
    - for « review » : (Csilléry et al. 2010)
Approximate Bayesian Computation

1) Create tree topology
   - For each scenario and each species, determine prior distribution of demographic parameters
     - absolute nucleotide substitution rates
     - migration rates between populations
     - effective population size (given by SDMs)
   - $10^6$ random draws of all the parameters $\Rightarrow 10^6$ trees for each scenario and each species

2) Create sequence matrices
   - Define the likelihood $\Rightarrow$ sequence mapping on the trees
   - 1 matrix for each trees $\Rightarrow 10^6$ matrices per scenario et species
Approximate Bayesian Computation

1) Create tree topology
   - For each scenario and each species, determine prior distribution of demographic parameters
     - absolute nucleotide substitution rates
     - migration rates between populations
     - effective population size (given by SDMs)
   - $10^6$ random draws of all the parameters $\Rightarrow$ $10^6$ trees for each scenario and each species

2) Create sequence matrices
   - Define the likelihood $\Rightarrow$ sequences mapping on the trees
   - 1 matrix for each tree $\Rightarrow$ $10^6$ matrices per scenario and species
Approximate Bayesian Computation

3) Compare with observed sequence matrix
   - Choose descriptive statistics to resume matrices
   - Compute Euclidian distance to
     - resume all the statistics of a matrix
     - determine distance between each simulated matrix and the observed one

4) choose the best scenario
   - Sort the distances by ascending order
   - Take the 1000 first distances
   - Obtain the posterior distribution of demographic parameters
     - mean, median,…
   - Calculate the percentage of each scenario in the 1000 best coalescence simulations

The best scenario is the one that has the highest %
Approximate Bayesian Computation

3) Compare with observed sequence matrix
   - Choose descriptive statistics to resume matrices
   - Compute Euclidian distance to
     - resume all the statistics of a matrix
     - determine distance between each simulated matrix and the observed one

4) Selection of the best scenario
   - Sort the distances by ascending order
   - Take the first 1000 distances
   - Compute the percentage of each scenario among the 1000 best coalescence simulations
     The best scenario is the one that has the highest %
   - Obtain the posterior distribution of demographic parameters
     - mean, median,…)
Biogeographical hypothesis (example)
Species Distribution Modeling

- Models → simplification
- SDMs → prediction of suitability for the development of species
- Statistical or mathematical association between dependent (data on distribution of species) and independent variables (environmental factors) → extrapolation to the whole study area

![Maps and graphs showing species distribution models](Calypogeia muelleriana)
Independent variables

- **Type**
  - Climatic variables: Worldclim 1.4, present, past and future
  - Soil, lithology and geology
  - Elevation and derived variables
  - Variables obtained through remote sensing
  - Human demography and land use

- Different scales

- **Raster**
  - each pixel = a value of the climatic variable
Independent variables

- **Type**
  - Climatic variables: Worldclim 1.4, present, past and future
  - Soil, lithology and geology
  - Elevation and derived variables
  - Variables obtained through remote sensing
  - Human demography and land use

- **Different scales**

- **Raster**
  - each pixel = a value of the climatic variable
Dependent variables

- Occurrence points
- Point shapefile

Two principal types of model

- Presence-only
  - Descriptive, describe a climatic envelop
  - Backgrounds don’t change anything
- Presence-absence/pseudo-absence
  - Probabilistic, distribution of probability of presence
  - Pseudo-absence = when no absence data are available
    → randomly distributed into the background
  - Backgrounds change the model
Dependent variables

- Occurrence points
- Points shapefile

Two principal types of model

- Presence-only
  - Descriptive, describe a climatic envelop
  - Backgrounds don’t change anything

- Presence-absence/pseudo-absence
  - Probabilistic, distribution of probability of presence
  - Pseudo-absence = when no absence data are available
    → randomly distributed into the background
  - Backgrounds change the model
Model

- **Raster**
  - Each pixel is associated with a value = index of suitability

- **Can be projected**
  - Into another area
  - In the past
  - In the future

- **Binarisation (optional)**
  - Threshold → if $\geq T \ 1$ ; if $< T \ 0$
  - Area of suitability for the species

- **MESS analyze (optional)**
  - Define a envelop of analogous climate
  - Model can projected on analogous region only
How to get effective population size?

- Define 100% of probability of presence \( (1) = \min X ; \max Y \) individuals
  - The range usually taken is min 1 and max 50 individuals

- Sum all of the pixels \( \rightarrow N_e_{\text{min}} \)
  - if binarised model \( \rightarrow \) sum all the value 1
  - If not \( \rightarrow \) standardize the values by divided each value by the maximum value + sum the pixels

- Multiply the sum by Y \( \rightarrow N_e_{\text{max}} \)

- This range is use in the ABC as a prior parameter
  - One range for each species and each period (if projections)

- It’s possible to cut the model into different regions \( \rightarrow N_eA, N_eB \)
Example in our study

- Model cut into 6 parts
  - North-East of Europe
  - North-West of Europe
  - South-East of Europe
  - South-West of Europe
  - North America
  - Macaronesia

Range of Ne for each region

- To get Ne East => Ne N-E + Ne S-E
- To get Ne South => Ne S-E + Ne S-W
- ...
Applications

- **Biogeography**: Study the effects of global change and distribution of species in the past
- **Biodiversity**
  - Search for new populations of endangered organisms
  - Selecting areas for reintroduction
  - Biodiversity patterns and hot spots
  - Reserve design
  - Basis of the IUCN classification of endangered species
Limitations

- Pseudoequilibrium hypothesis: we assume that the population is in balance or pseudoequilibrium with environmental conditions.

- Biotic interactions

- Uncertainty of the input data and computations

- No space component: biogeographical barriers, limitations in the dispersal capacity of the species, and so on.

  Bryophytes are good model because of their high dispersal capacities.
Dynamic models

Actuel  2050  2090

Potentially suitable area

Engler et al. (2012)

Dispersion parameters

Potentially colonizable area

Distances ?
Dynamic models

Distances?

Estimations

Short distance

Long distance

Spores size

Dispersion distance
Dynamic models

- Green-red gradient = gradient of suitability
- Pink = suitable but not colonizable area
Improved dynamic models

Dispersion capacities

- Short distance
- Long distance

Experimental approach:
  [Direct]
  - Dispersion corridors

Genetic approach:
  [Indirect]
  - Populations structure

Scenario 1
Scenario 2
Orthotrichum stramineum

Thank you for your attention!
Bibliography


Bibliography


